STRUCTURE-AWARE DICTIONARY LEARNING WITH HARMONIC ATOMS

Ken O'Hanlon and Mark D.Plumbley

Queen Mary University of London
Centre for Digital Music
Mile End Road, London E1 4NS, United Kingdom
email: {ken.hanlon, mark.plumbley}@eecs.qmul.ac.uk

ABSTRACT

Non-negative blind signal decomposition methods are widely used for musical signal processing tasks, such as automatic transcription and source separation. A spectrogram can be decomposed into a dictionary of full spectrum basis atoms and their corresponding time activation vectors using methods such as Non-negative Matrix Factorisation (NMF) and Non-negative K-SVD (NN-K-SVD). These methods are constrained by their learning order and problems posed by overlapping sources in the time and frequency domains of the source spectrogram. We consider that it may be possible to improve on current results by providing prior knowledge on the number of sources in a given spectrogram and on the individual structure of the basis atoms, an approach we refer to as *structure-aware dictionary learning*. In this work we consider dictionary recoverability of harmonic atoms, as harmonicity is a common structure in music signals. We present results showing improvements in recoverability using structure-aware decomposition methods, based on NN-K-SVD and NMF. Finally we propose an alternative structure-aware dictionary learning algorithm incorporating the advantages of NMF and NN-K-SVD.

1. INTRODUCTION

Signals are often represented as a collection of atoms in some signal transform:

$$y = \sum_{\lambda} a_{\lambda} x_{\lambda} \tag{1}$$

where y is the signal, $A = \{a_{\lambda}\}$ is a dictionary of atoms, x is a coefficient vector or matrix, and $\lambda \in \Delta$, where Δ is the index set of coefficients. Sparse coding seeks to represent the signal with few non-zero coefficients. This may be achieved by using overcomplete dictionaries consisting of different bases which better represent different signal elements. Alternatively dictionaries of basis atoms can be used. Sparse coding requires solution of the following:

$$\min \|x\|_0 \quad s.t \quad y = Ax. \tag{2}$$

This problem is NP-hard and many different strategies have been proposed for the purpose of sparse coding including greedy methods such as Matching Pursuit (MP) [8] and global optimization methods, such as Basis Pursuit (BP). Sparse representations have been seen to be useful for many audio applications, such as source separation and coding [9].

Blind signal decomposition methods can be used to learn basis atoms from a given signal. Given a magnitude spectrogram S, we seek a non-negative matrix decomposition such that:

$$S \approx DT \tag{3}$$

where $S \in \Re_{+}^{M \times N}$, $D \in \Re_{+}^{M \times K}$, $T \in \Re_{+}^{K \times N}$, M is the number of frequency bins, N is the number of time bins and K is the order of the decomposition. Here D is a dictionary matrix with a spectrum basis atom in each column, while the rows of T contain the time support vectors for the atoms.

The most popular method for performing the decomposition in (3) is Non-negative Matrix factorisation NMF [7]. An alternative method is NN-K-SVD [1], a non-negative variation on the popular K-SVD sparse dictionary learning method. These two methods were compared in [2] for the task of automatic music transcription, with NMF outperforming NN-K-SVD. A similar result was demonstrated in [10] for the task of harmonic dictionary recovery.

In transcription tasks using blind signal decomposition the approximation that each atom contains sound from one source is used. This is often not the case due to temporal and frequency overlap in the spectrogram and are constrained by the predetermined number of atoms to be learnt, or learning order. When the learning order is small, we observe atoms containing two pitched sources. When the learning order is large, atoms in which the energy is focussed in a single spike may appear, due to overlap in the frequency domain of sources of different pitches. Many musical sources display harmonicity and several works aim to incorporate this structure with sparse coding or blind signal decompositions.

Gribonval and Bacry present Harmonic Matching Pursuit (HMP), which performs sparse coding on a signal over a dictionary of harmonic atoms in [5]. Carabias-Orti et al [4] extend HMP, using spectral smoothness constraints and clustering of extracted harmonic atoms, in order to learn a dictionary of harmonic atoms from the signal, which is used to perform polyphonic transcription. Vincent et al introduced Harmonic NMF in [11], utilising a harmonic filtering on each of the set of basis atoms after each iteration of NMF. In [10], a variable length harmonic dictionary learning method was proposed, which learned new atoms one at a time from a signal. The atoms were filtered with harmonic combs tuned to their estimated pitch.

In this paper we consider the task of harmonic dictionary recovery. We show the use of Non-Negative Order Recursive Matching Pursuit (NN-ORMP) for the sparse coding step in NN-K-SVD improves the performance. We implement structure-aware versions of NN-K-SVD and NMF. We show that structure-aware dictionary learning significantly improves the results for dictionary recovery. Finally, we note the relative advantages of structure-aware NN-K-SVD and NMF and propose a method that incorporates the advantages of both.

This research is supported by ESPRC Leadership Fellowship EP/G007144/1 and EU FET-Open Project FP7-ICT-225913 “SMALL”.

2. BACKGROUND

2.1 Non-Negative Blind Decompositions

NMF[7] seeks the approximation in (3) by minimising the error on the relationship between the data matrix and its reconstruction. The Frobenius norm is commonly used to measure the error, but other cost functions can be used. A typical algorithm for NMF is an iterative algorithm which alternately updates the dictionary and the coefficient matrix using multiplicative updates, which ensure non-negativity.

NN-K-SVD [1] applies the added constraint of sparsity to the approximation in (3). This is achieved through a two-step iterative algorithm that alternatively seeks to minimize the reconstruction error norm of the matrix S and the norm of T. Minimisation of the norm of T is performed by sparse coding. Any non-negative sparse coder can be used. The other step of the algorithm is the dictionary update stage, which differs from NMF in that each atom is updated individually. Each atom is selected, and a rank one singular value decomposition (SVD) is performed on the sum of the reconstruction error and the atom contribution, over that atom's support in the time domain.

2.2 Non-Negative Sparse Coding

In [1] Non-negative Basis Pursuit (NN-BP) is proposed which is based on the non-negative sparse coding scheme in [6]. Given D, NN-BP performs several iterations of a multiplicative update to estimate T. Then the L atoms with the largest non-zero coefficients in each column of T are selected to give an L-sparse support set. The coefficients of the support set are then calculated by minimization of the least-squares error.

Non-Negative Orthogonal Matching Pursuit (NN-OMP) is proposed in [3]. This is a variation of the greedy iterative OMP algorithm. Similar to OMP, at each iteration the atom that is most correlated with the residual signal is selected and added to the support. Then the coefficients of all support atoms are calculated through least-squares analysis. The residual signal is updated by subtracting the signal contribution of the selected atoms from the original signal.

2.3 Harmonic Atoms

Harmonic atoms are introduced in [5]. A harmonic atom is a group of harmonically related Gabor atoms, whose correlation with a signal r is given by:

$$|h(f_0), r| = \sum_{c=1}^{C} \max_{f_c : |f_c - cf_0| \leq \frac{ca}{2}} |\langle g(f_c), r \rangle|^2 \qquad (4)$$

where h is a harmonic atom with fundamental frequency f_0, a is the frequency resolution, C is the number of harmonics considered and $g(f)$ is a Gabor atom of frequency f. The harmonic atom is the set of local maximums, f_c which contribute to the expression in (4). In practice, this group is extended to include the sidelobes of the peaks, which are shown here with a width of 1:

$$h = \{f_c + l, \quad \forall \{f_c, l\} \mid l \in \{-1, 0, 1\}\}. \qquad (5)$$

Harmonic atoms have an implicit fundamental frequency, $f_0(h)$, and can be used for pitch estimation. However, typical pitch estimation problems such as octave jumping may be encountered, and HMP [5] suffers from overlapping partials being assigned to one atom.

3. PROPOSED METHOD

3.1 Harmonic NN-K-SVD

We propose a harmonic variation NN-K-SVD, using the assumption that each note or pitch is represented by one atom. We assume for this work a prior harmonic analysis providing knowledge of the learning order, K, and a binary harmonic structure matrix, I, which is used to constrain the subsequent learning. With a signal that is known or estimated to contain a group of pitches or notes and with knowledge of the members of the set of corresponding harmonic atoms $H = \{h\}$, we define I as follows:

$$I_{f,k} = \begin{cases} 1, \text{ where} f \in h_k \\ 0, \text{ otherwise .} \end{cases} \qquad (6)$$

The proposed algorithm is presented in *Algorithm 1*. It differs from NN-K-SVD only through the input of the indicator matrix and in the final dictionary update step when the learned atom d_k is filtered by its corresponding indicator vector i_k.

Algorithm 1 Harmonic NN-K-SVD

```
Input
  $I \in \{0,1\}^{M \times K}$ ; $S \in \Re_+^{N \times M}$
Initialise
  $T^0 = 0$; $D \in \Re_+^{M \times K}$; $D \longleftarrow D \circ I$
repeat
  Sparse code using pursuit algorithm
    $\min_T \{\|S - DT\|_2^2\} \quad s.t. \ \|t_n\|_0 = L$
  Update Dictionary
  for $k = \{1, 2.....K\}$ do
    $w_k = \{n \mid t_k(n) > 0\}$
    $E_k^{w_k} = (S - (DT - d_k t_k))^{w_k}$
    $E_k^{w_k} = U \Delta V^T$
    $t_k = V$;
    $d_k \longleftarrow U \otimes i_k$
  end for
until stopping condition met
```

3.2 Non-Negative Order Recursive Matching

In this non-negative framework, we modified Order Recursive Matching Pursuit (ORMP) to be non-negative using the *lsqnonneg* function in Matlab. NN-ORMP differs from NN-OMP[3] only in that it selects the candidate atom which when added to the support set minimises the least-squares error over the spectrogram S while NN-OMP[3] selects the atom which minimises the error over the residual error. This can be seen by replacing the second and third lines in the repeat loop in *Algorithm 2* with:

$$E_k = min_k \|d_k t_k - R^{i-1}\|_2^2 \qquad (7)$$

where R^i is the residual at iteration i. to implement NN-OMP.

We note the heavy use of non-negative least-squares in NN-ORMP. This function is also used in NN-OMP and in NN-BP. In this work we replaced the Matlab *lsqnonneg* function with a multiplicative update estimate of the non-negative least squares, which we observed to provide a speedup without detriment to the results.

```
Algorithm 2 NN-ORMP
  Input
    $D \in \Re_+^{M\times K}$; $S \in \Re_+^{M\times N}$
  Initialise
    $i = 0$; $r^0 = S$; $T^0 = 0$; $\Gamma^0 = \{\}$;
  repeat
    $i = i + 1$
    $\Gamma_k = \Gamma^{i-1} \cup k \, , \forall k \notin \Gamma^{i-1}$
    $E_k = \min_T \|D_{\Gamma_k} T_{\Gamma_k} - S\|_2^2$
    $k' = \arg\min_k E_k$
    $\Gamma^i = \Gamma^{i-1} \cup k'$
    $T^i = min_T \|D_{\Gamma^i} T_{\Gamma^i} - S\|_2^2; \; T \geq 0$
  until stopping condition met
```

3.3 Harmonic NMF

Harmonic NMF (H-NMF) was used in [11] for the purpose of polyphonic transcription. We implement our version of this algorithm which, similar to H-NN-K-SVD (*Alg. 1*), incorporates prior information and is constrained in its learning through the use of the binary structure indicator matrix, *I*. The algorithm is described in *Algorithm 3*, where the learning iterates through two multiplicative updates, with *I* incorporated in the second of these.

```
Algorithm 3 Harmonic NMF
  Input
    $I \in \{0,1\}^{M\times K}, S \in \Re_+^{M\times N}$
  Initialise
    $T^0 \in \Re_+^{K\times N}$; $D^0 \in \Re_+^{M\times K}$
  repeat
    $T \longleftarrow T \circ D'S \oslash D'DT$
    $D \longleftarrow D \circ I \circ TS' \oslash DTT'$
  until stopping condition met
```

4. EXPERIMENTS

We synthesise a set of spectrograms from an input dictionary and a coefficient matrix, to compare the algorithms. Given the spectrogram, the decompostion algorithms seek to find the original input dictionary. We decompose the spectrograms using NMF, NN-K-SVD, H-NMF and H-NN-KSVD. For the NN-K-SVD based algorithms we perform the decompositions using both NN-BP and NN-ORMP, separately. We provide a binary indicator matrix (6), derived from the input dictionary for the harmonic versions of all algorithms. All algorithms are run for the same number of iterations, which we set to 100.

We measure the success of the dictionary recovery by correlating the learned dictionary with the synthesised input dictionary using atomwise similarity. We consider two unit norm atoms having an inner product of greater than 0.9 as being strongly correlated. For each atom in the original dictionary, for which there exists a strongly correlated atom in the learned dictionary we assign a hit. We define *Accuracy = average hits per run* as the main measure of performance. We also record the average maximum correlation measure. For each atom, a_k in the input dictionary A, we consider only its largest correlation $r_k = \max_i \langle a_k d_i \rangle$ with the dictionary atoms d_i. We define the correlation between the input and learned dictionaries:

$$\rho = \frac{\sum_{k=1}^{K} r_k}{K} \tag{8}$$

and record *Correlation* as the mean of ρ over all runs.

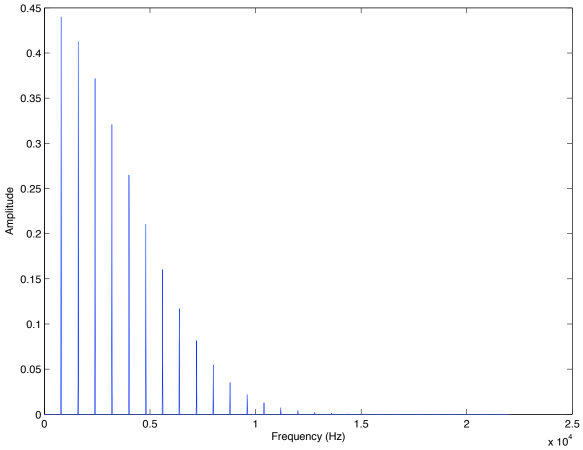

Figure 1: *Example harmonic atom*

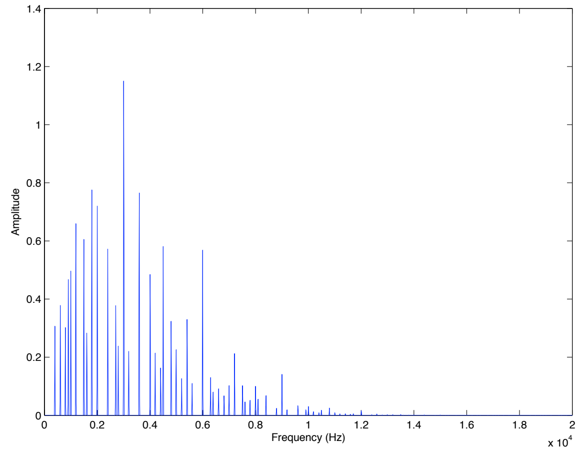

Figure 2: *Example datapoint with five harmonic atoms*

4.1 Experiment 1

We use a similar experimental setup as that in [10]. First K synthetic harmonic atoms, see (*Fig. 1*) are created for a range of pitches with a fixed spectral envelope used to define the relative amplitudes of the harmonic partials. The sidelobes consist of one frequency bin on either side of the main harmonic partial whose amplitudes are defined by a Gaussian centred on the main partial. A coefficient matrix is synthesised by randomly selecting L atoms at each of N time bins and setting their coefficient values to one. The spectrogram is calculated from the product of the dictionary and the coefficient matrix. A sample time bin from one of the spectrograms is shown in *Fig. 2*. For this experiment here we set $K = 10$, $L = 5$ and $N = 100$.

Algorithm	Accuracy	Correlation
NMF	6.95	0.92
NN-KSVD(BP)	6.65	0.95
NN-KSVD(ORMP)	7.75	0.95
H-NMF	9.99	0.99
H-NN-KSVD(BP)	10.0	1.00
H-NN-KSVD(ORMP)	9.99	0.99

Table 1: Results from *Experiment 1*

Algorithm	Accuracy	Correlation
NMF	9.55	0.98
NN-KSVD(BP)	8.78	0.97
NN-KSVD(ORMP)	8.71	0.97
H-NMF	10.0	0.99
H-NN-KSVD(BP)	10.0	0.99
H-NN-KSVD(ORMP)	10.0	0.99

Table 2: Results from *Experiment 2*

4.2 Experiment 2

We changed the parameters of the synthetic dictionary by altering the atom shapes. In *Experiment 1* the atom coefficients were set by a fixed spectral envelope. In *Experiment 2* we set the atom coefficients randomly, and filter to maintain the harmonic structure. We use the same values of *K, N* and *L* as those used in *Experiment 1*.

4.3 Experiment 3

For this experiment, we used the same atoms as used in *Experiment 1* but decreased the support to $L = 4$ at each time bin and skewed the distribution of the atom supports in the spectrogram, contrary to the random distributions used in the other experiments. The atoms were randomly ordered and the following support set cardinality applied:

$$|\Gamma_{k'}| = \{90, 80, 70, 70, 60, 10, 8, 6, 4, 2\}. \quad (9)$$

5. RESULTS

We observe the results of *Experiment 1* in *Table 1*. We see in the first part of the table that the use of NN-ORMP causes a large change in the performance of NN-K-SVD, surpassing that of NMF. The second part of the table presents the results of the structure-aware methods. We can see that all methods perform well, with almost complete recovery.

In *Table 2* we see again the that the structure-aware versions of the algorithms achieve high accuracy, with hits for all atoms. All the unstructured algorithms perform better than in *Experiment 1* suggesting that the fixed spectral envelope causes some problems for unconstrained learning. NMF performs better than NN-K-SVD, regardless of the sparse coding algorithm used. In this experiment NN-ORMP and NN-BP perform similarly.

Table 3 shows the results from *Experiment 3*, which contains the spectrograms with skewed atom support size. We can see that the results for all methods degrade relative to previous experiments. Having atoms with a small support set presents a greater problem to all algorithms. These results are similar to those in *Experiment 1* in that NMF performs better than NN-K-SVD(BP) but worse than NN-K-SVD(ORMP).

Algorithm	Accuracy	Correlation
NMF	5.75	0.88
NN-K-SVD(BP)	5.45	0.88
NN-K-SVD(ORMP)	6.50	0.92
H-NMF	7.99	0.93
H-NN-KSVD(BP)	9.20	0.98
H-NN-KSVD(ORMP)	8.02	0.94

Table 3: Results from *Experiment 3*

There is a notable difference in the performance of H-NN-K-SVD with relation to the sparse coder used, with NN-BP providing superior performance. We observe that NN-ORMP sometimes lost support for some atoms in the coefficient matrix, contributing to the error found with the algorithm.

The results in general suggest that dictionary recoverability is greatly enhanced by the incorporation of structure awareness, although this is expected due to the prior information supplied. Another feature of these experiments was the speed with which the learned dictionaries converged towards the input dictionaries when using the structure-aware methods for the first two experiments. Often less than twenty iterations were required to recover all atoms using the accuracy criteria set above, in particular when using the NN-K-SVD with NN-ORMP, in contrast to the performance of the algorithm in *Experiment 3*. We note that NMF is significantly faster per iteration than the NN-K-SVD.

6. UPDATED METHOD

We can see from the results that H-NMF and H-NN-KSVD both perform almost perfectly in the first two experiments. However, we notice, in general a divergence between the NMF and NN-K-SVD performances. NMF performs better in *Experiment 2* and seems more robust to variability in the atom shape, while H-NN-K-SVD is more robust to unevenness in the distribution of atom support size. It is presumed that the sparse coding enables the NN-K-SVD to quicker approximate the support of the dictionary in the spectrogram, which is a strength of the algorithm, while the relative increase in performance of the NMF when faced with variable shaped atoms suggests that the multiplicative update may update the dictionary more efficiently.

Based on this observation, we propose a Structure-Aware Non-Negative Sparse Multiplicative Update Dictionary Learning (SA-NN-S-MUDL), which is described in *Algorithm 4* and uses elements from both NMF and NN-K-SVD. It is a two-step iterative algorithm which updates the coefficient matrix by sparse coding, and uses the global multiplicative update stage from NMF to update the dictionary.

Using this method we decompose the spectrograms in *Experiment 2* and *Experiment 3*. We use NN-BP and NN-ORMP, separately. The results from these experiments are displayed in *Table 4*. Here we can see that the proposed method finds hits for all atoms in *Experiment 2*, similar to NMF and NN-K-SVD. However, the proposed method improves on all other methods for *Experiment 3*. We can see here again a better performance while using the NN-BP for this experiment.

Algorithm 4 SA-NN-S-MUDL Algorithm

```
Input
    $I \in \{0,1\}^{M \times K}, S \in \Re_{+}^{M \times N}$
Initialise
    $T^0 = 0$; $D \in \Re_{+}^{M \times K}$
repeat
    Sparse code using pursuit algorithm
        $\min_T \{\|S - DT\|_2^2\} \quad s.t. \ \|t_n\|_0 = L$
    Update Dictionary
        $D \longleftarrow D \circ I \circ TS' \oslash DTT'$
until stopping condition met
```

Algorithm	Accuracy	Correlation
Experiment 2		
SA-NN-S-MUDL(BP)	10.0	1.0
SA-NN-S-MUDL(ORMP)	10.0	1.0
Experiment 3		
SANNSMUDL-BP	9.86	0.99
SANNSMUDL-ORMP	9.17	0.97

Table 4: Results with updated method on spectrograms from *Experiment 2* and *Experiment 3*

7. CONCLUSIONS AND FURTHER WORK

We have shown the advantages of structure-aware dictionary learning, and have derived a structure-aware non-negative dictionary learning method, which improves the dictionary recovery for the learning of harmonic atoms. We believe this improvement is due to sparsity reducing the noise in the coefficient matrix, while the global multiplicative update for the dictionary is smoother than the K-SVD when there is large overlap in the atom supports, as the error is treated globally, countering the interaction element of K-SVD. We intend to further explore the use of this method, incorporating other metrics, testing on noisy datasets and using structures other harmonic. Similar work which incorporates sparsity and NMF, such as [6], will be compared.

Further investigation may be required on the use of non-negative sparse coding. While NN-ORMP seems effective in atom selection, even when handling largely overlapping atoms, it is computationally demanding, due to the least-squares analysis for each candidate atom being performed over multiple atoms.

The work shown in this paper is part of a larger work of building a transcription system using signal-adaptive sparse non-negative dictionary learning, and some of the findings here are consistent with our experience, such as the disappearance of atoms with a small support set. We built these noiseless experiments in order to test the methods, and investigate their limitiations for tasks such as transcription.

REFERENCES

[1] M. Aharon, M. Elad and A.M. Bruckstein, “K-SVD and its non-negative variant for dictionary design,” in *Proc. SPIE conference wavelets*,Vol 5914, pp. 327-339, 2005.

[2] N. Bertin, R. Badeau and G. Richard, “Blind signal decompositions for automatic transcription of polyphonic music: NMF and K-SVD on the benchmark,” in *IEEE Int. Conf. on Acoustics Speech and Signal Processing 2007 (ICASSP)*, pp. I-65 I-68, 2007.

[3] A.M. Bruckstein, M. Elad and M. Zibulevsky, “A non-negative and sparse enough solution of an underdetermined linear system of equations is unique,” *IEEE Trans. on Information Theory*, Vol. 54, pp. 4813-4820, 2008.

[4] J.J. Carabias-Orti, P. Vera-Candeas, F.J. Canadas-Quesada and N. Ruiz-Reyes, “Music scene-adaptive harmonic dictionary for unsupervised note-event detection,” *IEEE Trans on Audio, Speech and Language Processing*, Vol. 18, 2010.

[5] R. Gribonval and E. Bacry, “Harmonic decomposition of audio signals with matching pursuit,” *IEEE Trans. in Signal Processing*, Vol 51, pp. 101-111, 2003.

[6] P.O. Hoyer, “Non-negative sparse coding,” in *Neural Networks for Signal Processing XII (Proc. IEEE Workshop on Neural Networks for Signal Processing)*, pp. 557-565, 2002.

[7] D.D. Lee and H.S.Seung, “Algorithms for non-negative matrix factorization,” in *Advances in Neural Information Processing Systems*, Vol.13, pp. 556-562, 2001.

[8] S. Mallat and Z. Zhang, “Matching Pursuit in a time-frequency dictionary,” *IEEE Trans. in Signal Processing*,Vol.41, pp.3397-3415, 1993.

[9] M.D. Plumbley, T. Blumensath, T, L. Daudet, R. Gribonval and M.E. Davies, “Sparse representations in audio and music: From coding to source separation,” *Proceedings of the IEEE*,Vol. 98, pp. 995-1005, 2010.

[10] S.K. Tjoa, S.K. Stamm, M.C. Lin, W.S. Liu and K.J. Ray, “Harmonic variable-size dictionary learning for music source separation,” in *IEEE Int. Conf. on Acoustics Speech and Signal Processing 2010*, p. 413-416, 2010.

[11] E. Vincent, N. Bertin, R. Badeau, “Harmonic and inharmonic non-negative matrix factorization for polyphonic pitch transcription,” in *IEEE Int. Conf. on Acoustics Speech and Signal Processing 2008*, pp. 109 112.