SEPARATION OF MUSICAL NOTES WITH HIGHLY OVERLAPPING PARTIALS USING PHASE AND TEMPORAL CONSTRAINED COMPLEX MATRIX FACTORIZATION

Yi-Ju Lin, Yu-Lin Wang, Alvin W.Y. Su,
SCREAM Lab., Department of CSIE,
National Cheng-Kung University,
Tainan, Taiwan
lyjca.cs96@g2.nctu.edu.tw

Li Su,
Research Center for Information Technology
Innovation, Academia Sinica,
Taipei, Taiwan
li.sowaterking@gmail.com

ABSTRACT

In note separation of polyphonic music, how to separate the overlapping partials is an important and difficult problem. Fifths and octaves, as the most challenging ones, are, however, usually seen in many cases. Non-negative matrix factorization (NMF) employs the constraints of energy and harmonic ratio to tackle this problem. Recently, complex matrix factorization (CMF) is proposed by combining the phase information in source separation problem. However, temporal magnitude modulation is still serious in the situation of fifths and octaves, when CMF is applied. In this work, we investigate the temporal smoothness model based on CMF approach. The temporal ac-tivation coefficient of a preceding note is constrained when the succeeding notes appear. Compare to the unconstraint CMF, the magnitude modulation are greatly reduced in our computer simulation. Performance indices including source-to-interference ratio (SIR), source-to-artifacts ratio (SAR), source-to-distortion ratio (SDR), as well as modulation error ratio (MER) are given.

1. INTRODUCTION

Musical note separation (MNS) is the extension of musical source separation (MSS). MNS means to extract every note from a mixture source (e.g., the fuzzy interference [1],and the regular regression [2–4]). The fuzzy interference [1] separates the notes by estimating the harmonic rate of each partial; the regular regression [2–4] separates recursively and time-varyingly. On the other hand, MSS means to separate several sources from a mixture (e.g., the statistics [5], the sparse de-composition [6, 7], non-negative matrix factorization (NMF) [8–14], and the complex matrix factorization (CMF) [15–18]). In music, fifths and octaves may co-exist. Both cases produce the severe overlapping partial problems; therefore, how to separate the notes becomes an important and difficult issue. Some separation methods (e.g., NMF) focus on analysing the temporal activation coefficient of each note, but without considering the phase information. Recently, the CMF is proposed to separate the source by investigating the phase information [15], where each note could be analysed with the corresponding source. Usually, the CMF-separated magnitude parts of over-lapping partials are affected by their corresponding phases [15–18]. CMF also estimates most of the likely variations and dis-continuities of phases [18]. Hence, CMF is found more appropriate to separate the notes from different instruments. However, in practice, when fifths and octaves are played in a *interleaving* (i.e. notes appearing one after another) way, there exists a serious effect of *temporal magnitude modulation* in the separated notes. Such modulation is fairly audible in the sustain part of a note, where a listener can easily heard the preceding note is played one more time. The more overlapping partials there are, the more serious the modulation there is, as we will see in Figure 2 in the later section. Even the temporal sparsity and phase evolution constraints have been proposed in [18], the problems of overlapping partials still exist. Virtanen [13] investigates the temporal continuity to constrain the NMF on separation and achieve good results. Applying this constraint is valid when the temporal activation coefficients of each note would not change frequently in a short time. In order to mitigate the effect of temporal magnitude modulation in CMF-based note separation, the temporal smoothness constraint is investigated in this work. The constraint is applied to the preceding note before the succeeding notes come in for keeping the variation of the activation coefficient better, if the succeeding notes are fifths and/or octaves. The modulation is reduced dramatically and the phase continuity is kept, although the performance of the overall separation is still unsatisfactory. In addition to SIR, SAR, and SDR [19], MER proposed in [20] are used to evaluate the performance. All the test files are clipped from the RWC database [21]. The paper is organized as follows. Section 2 describes the CMF algorithm with phase constraint, and section 3 incorporate the temporal smoothness constrains into the CMF for reduction of temporal magnitude modulation. In section 4, the experiment results are shown and compared to the prior method [18]. Finally, conclusions and future works are given in section 5.

2. BACKGROUND

In this paper, the phase constraint of CMF is investigated for improving the note separation. The methodology of CMF [15] and its phase constraint [18] are introduced in this section.

2.1. Complex Non-Negative Matrix Factorization

CMF is involved the phase information to the matrix decomposition. This factorization supports the analysis of audio signal. Given an $N \times M$ complex-valued short-time Fourier transform (STFT) matrix $X \in \mathbb{C}^{N\times M}$. X could be decomposed into a basis matrix $W \in \mathbb{C}^{N\times K}$ satisfying $\sum_n W_{n,k} = 1, \forall k = 1, ..., K$,an activation matrix $H \in \mathbb{R}^{K\times M}_{\geq 0}$, and a tensor with phase information which can be represented as $Phi \in \mathbb{R}^{K\times M}$. The objective function of CMF is written as follows:

$$D_X = \sum_{n,m} |X_{n,m} - \hat{X}_{n,m}| \quad (1)$$

$$D_{\texttt{TSparsity}} = \sum_{k,m} |Hk, m|^g \quad (2)$$

$$\text{Minimize} : D_{\texttt{CMF}} = \frac{1}{2} D_X + \lambda D_{\texttt{sparsity}} \quad (3)$$

where $\hat{X}_{n,m} = W_{n,k} H_{k,m} \exp(i\Phi_{n,k,m})$.Here X is the spectrogram of the audio signal, W consists of the K spectra of K notes, H is the K temporal activation coefficients correspond to the each spectrum in different time, λ is the temporal sparsity parameter to penalize the objective function, and g is a parameter for descripting the shape of the sparse distribution. The only one constraint, here, is the temporal sparse constraint.

2.2. CMF under Phase Evolution constraints

Phase is a physical quantity which evolves regularly for most musical signals. Therefore, J. Bronson et al. [18] proposed an additional constraint of phase evolution in CMF to separate overlapping partials. This constraint is based on several assumptions: first, the pitches of each source should be known; second, there is no spreading energy of each partial bin; third, all the notes are played by the same instrument; and finally, each source can be represented by the combination of sinusoidal functions as eq. (4):

$$X_k = \sum_{p=1}^{P_k} A_{k,p} \exp[(\pi f_{0_k} pT + \phi_{0_k}, p)] \quad (4)$$

The cost function is based on these assumptions, written as follows:

$$D_{\texttt{Phase}} = \sum_{n,k,p,m} \mathbb{1}\mathcal{N}_{k,p} | \exp(i\Phi_{n,k,m}) - \exp(i\Phi_{n,k,m-1}) \exp(i2\pi f_{0_k} pLT)|^2 \quad (5)$$

where L is the frame shift in samples, T is the sampling period, and the $\mathbb{1}\mathcal{N}_{k,p}$ is the set of the membership function between the bins of the kth fundamental frequency and partial frequency. The cost function constrains each phase of the cur-rent frame should approximate to the estimated value from the earlier phase information. The objective function is shown as below:

$$\text{Minimize} : D_{\texttt{CMF}_\texttt{p}} = \frac{1}{2} D_X + \lambda D_{\texttt{sparsity}} + \sigma D_{\texttt{Phase}} \quad (6)$$

where σ is the phase continuity parameter for increase the rate of convergence. The optimization formula is derived in [18].

3. TEMPORAL SMOOTHNESS CONSTRAINT

In addition to the phase evolution constraint, the proposed temporal smoothness constraint is based on two assumptions: first, the temporal activation coefficients vary slowly unless encountering an onset. Second, the pitch f_0 and onset timing m_0 of each note are known before processing.

3.1. Temporal Smoothness Cost Function

The temporal smoothness cost function [2, 4, 13] is stated as follows:

$$D_{\texttt{TSmoothness}} = \sum_{k,m} |H_{k,m} - H_{k,m-1}|^2 \quad (7)$$

Eq. (7) implies that for the musical signal under analysis, $H_{k,m}$ should be close to $H_{k,m-1}$. This constraint forces the temporal activation coefficients vary slow with time, thereby mitigate the temporal magnitude modulation effect of the separated notes.

3.2. CMF under Phase Evolution and Temporal Smoothness Constraints

Combining the CMF algorithm with both the phase evolution constraint and the temporal smoothness constraint, the objective function is given as below:

$$\text{Minimize} : D_{\texttt{CMF}_\texttt{pt}} = \frac{1}{2} D_X + \lambda D_{\texttt{sparsity}} + \sigma D_{\texttt{Phase}} + \gamma D_{\texttt{TSmoothness}} \quad (8)$$

where γ is the parameter regularizing temporal smoothness. We modify the basic cost function of CMF, the phase evolution and temporal smoothness constraints. The temporal activation coefficients are limited with the onset-timing of each note. The phase evolution constraint is necessary for keeping continuity of the phase. For each note, the temporal smoothness constraint is applied from 100 ms before the note onset the instance of note offset. In our preliminary study we found 100 ms gives satisfactory result in general. Figure 1 shows the constraint is applied to the frames within the dash line. The CMF with Phase Evolution and temporal smoothness constraints is described in Algorithm 1.

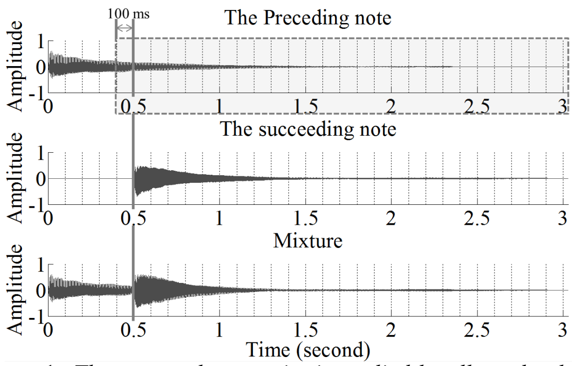

Figure 1: *The temporal constraint is applied locally and only on the preceding note starting from 100 ms before the on-set of the succeeding note till the end of signal.*

4. EXPERIMENTS AND RESULTS

In the following experiments, the proposed note separation method is evaluated on a set of mixture samples with two notes constituting intervals of perfect fifth, octave, and tritave (e.g., an octave plus a perfect fifth). Every mixture sample is made by adding the signals of the two notes together where the higher note is lagged by 0.5 second. For example, a mixture "A3+A4" contains a note A3 whose onset is at 0 second and a note A4 whose onset is at 0.5 second. In this way we can observe the temporal modulation effect of the lower note when the higher note joins in at 0.5 second. The notes are played with the same musical instrument for every sample. All sources are obtained from the part of piano, violin and guitar in the RWC database [21]. The sampling rate is 44.1 kHz. In computing the short-time Fourier transform (STFT) representation, we use a window with length of 4,096 samples and the hop size is 256 samples. The constraints of sparsity and phase have been discuss in the prior works [15, 18], we simply set the sparsity parameter $\lambda = 0.001$, and the phase continuity parameter $\sigma = 0.1$ according to the previous result. To show the temporal modulation

Algorithm 1: CMF with Phase Evolution and temporal smoothness constraints

Input: $X \in \mathbb{C}^{N \times M}$, $K \in \mathbb{N}$, $f_{0_k}(k = 1, ..., K)$ and $m_{0_k}(k = 1, ...K)$
Output: W, H, and Φ s.t.
$X_{n,m} \approx \sum_{k=1}^{K} W_{n,k} H_{k,m} \text{exp}(i\Phi_{n,k,m})$
$W \in \mathbb{C}^{N \times M}$, $H \in \mathbb{R}_{\geq 0}^{N \times M}$, $\Phi \in \mathbb{R}^{N \times K \times M}$
Initialized;
while *stopping criteria not met* **do**
Compute β
$\beta_{n,k,m} = \frac{W_{n,k} H_{k,m}}{\sum_k W_{n,k} H_{k,m}}$
Compute $\hat{X}$
$\hat{X}_{n,k,m} = W_{n,k} H_{k,m} \text{exp}(i\Phi_{n,k,m}) + \beta_{n,k,m}(X_{n,m} - \sum_k \hat{X}_{n,k,m})$
Compute $\hat{H}$
$\hat{H}_{k,m} = H_{k,m}$
Compute Φ
$\Phi_{n,k,m} = \text{Arg}\{\frac{\hat{X}_{n,k,m}}{\beta_{n,k,m}} W_{n,k} H_{k,m} + \sigma \sum_p \mathbb{1}\mathcal{N}_{k,p}[\text{exp}(i\Phi_{n,k,m-1})\text{exp}(i2\pi f_{0_k} rLT) + \text{exp}(i\Phi_{n,k,m+1})\text{exp}(-i2\pi f_{0_k} rLT)]\}$
Compute W
$W_{n,k} = \frac{\sum_m H_{k,m} \Re[(\frac{\hat{X}_{n,k,m}}{\beta_{n,k,m}})\text{exp}(-i\Phi_{n,k,m})]}{\sum_m \frac{H_{k,m}^2}{\beta_{n,k,m}}}$
Compute H
if $m_{0_{k+1}} + 100 < m \leq m_k$ **then**
$H_{k,m} = \frac{\sum_m W_{n,k} \Re[(\frac{\hat{X}_{n,k,m}}{\beta_{n,k,m}})\text{exp}(-i\Phi_{n,k,m})]}{\sum_m \frac{W_{n,k}^2}{\beta_{n,k,m}} + \lambda g(\hat{H}_{k,m})^{g-2}}$
else if $m < m_{0_{k+1}} + 100$ **then**
$H_{k,m} = \frac{\gamma H_{k,m-1} + \sum_m W_{n,k} \Re[(\frac{\hat{X}_{n,k,m}}{\beta_{n,k,m}})\text{exp}(-i\Phi_{n,k,m})]}{\sum_m \frac{W_{n,k}^2}{\beta_{n,k,m}} + \lambda g(\hat{H}_{k,m})^{g-2} + \gamma}$
else
$H_{k,m} = 0$
end
Project H onto non-negative orthant
iter=iter+1
end

effect and the separation performance using the smoothness constraint, the activation coefficient of each separated note with the lower pitch is displayed from 0.4 to 0.8 second. The magnitude of the higher pitch is not shown because it does not undergo the temporal magnitude modulation effect. We also compute and show dB-scaled SIR, SAR, SDR, and MER, where the first three are usually used to evaluate the separated sources in the audio source separation [19], and the last one is used to evaluate digital signal transmitter or receiver in a communications system, as formulated by [20]

$$\text{MER} := 10\log_{10}(\frac{P_{\texttt{signal}}}{P_{\texttt{error}}}) \tag{9}$$

where $P_{\texttt{signal}}$ is the RMS power of original signal and $P_{\texttt{error}}$ is the RMS power of the error signal (e.g., the RMS power of the difference between the original and the separated signal). Here, we use MER to compare the temporal activation coefficients of the original and separated notes.

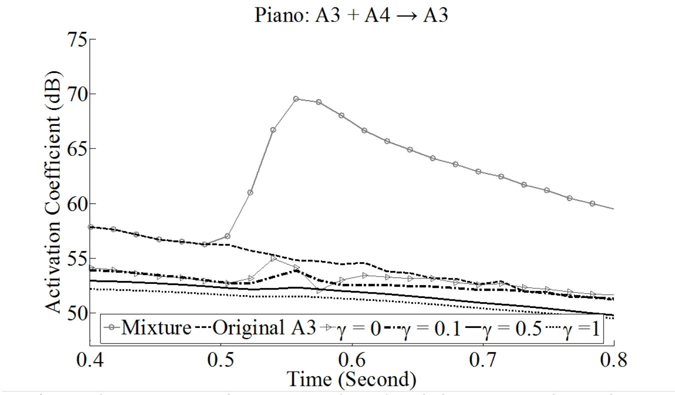

Figure 2: *An example (piano A3+A4) of the temporal envelopes of the preceding note (A3) with and without temporal smoothness constraint. The onset of the succeeding note (A4) is at about 0.5 second.*

4.1. Temporal Smoothness Parameter γ

We firstly investigate a case of A3+A4 in piano to observe the behaviors of the temporal magnitude modulation effect on the separated notes under different temporal smoothness parameters $\gamma = 0$ (no temporal smoothness constraint; the same case as in [18]), 0.1, 0.5 and 1. Figure 2 shows the temporal magnitudes of the mixture (A3+A4) before separation, and of the lower note (A3) before and after separation under different γ. Obviously, the magnitude of the separated A3 note without the temporal constraint (e.g., $\gamma = 0$) undergoes serious modulation after 0.5 second. For $\gamma > 0$, the temporal smoothness constraint suppresses the unwanted modulation. When γ increases, suppression becomes better while sacrificing overall note energy at the same time; the power of the separated A3 is by from 5 to 10 dB smaller than the original A3. In the following experiments on three different classes of instruments (i.e., piano, guitar and violin), we will compare only the case $\gamma = 0.5$ to the case $\gamma = 0$.

4.2. Piano

To give a more systematic evaluation, we consider three different note pairs (i.e., interval) and each with three distinctive pitch combinations: perfect fifth (i.e. A3+E4, E3+B3, and G3+D4), octave (i.e. A3+A4, E3+E4, G3+G4) and tritave (i.e. A3+E5, E3+B5, G3+D5), all of which are piano sounds. Figure 3 shows the temporal envelopes of the separated notes under three selected cases, and Table 1 shows the evaluation results. Similar to figure 2, from figure 3 we observe clearly that the temporal envelope becomes smoother after employing the temporal smoothness constraint. For the cases of fifth and tritave, there is less difference between $\gamma = 0$ and $\gamma = 0.5$ because there are less overlapping partials for the pairs. If there are more overlapping partials, such as the case of the octave where all the partials of the higher note are overlapped, the temporal smoothness constraint reduces the modulation a lot. However, from Table 1 we found no significant improvement for $\gamma = 0.5$ in terms of SIR, SAR, SDR and MER results, possibly due to the fact that the performance of piano note separation is intrinsically better: This can be seen from the case of perfect fifth and tritave, where the SDR with no temporal smoothness constraint is 9.2 and 9.9 dB, both of which are sufficiently high. Notably, the SDRs of the octave are only 1.3 and 1.6 dB for $\gamma = 0$

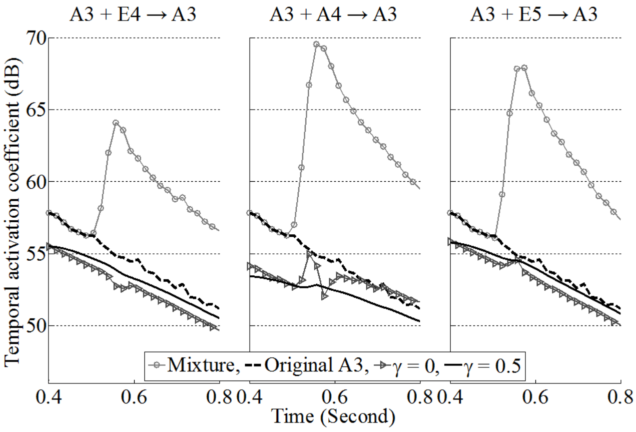

Figure 3: *Three cases of piano. Left: perfect fifth. Middle: octave. Right: tritave (circle: mixture, dashed line: Original A3, triangle:* $\gamma = 0$*, solid line:* $\gamma = 0.5$

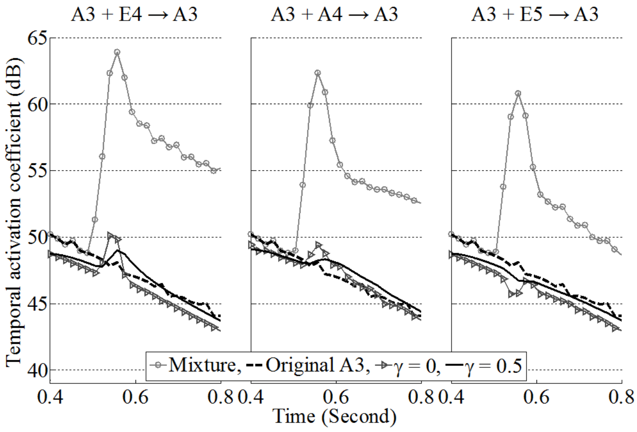

Figure 4: *Three cases of guitar. Left: perfect fifth. Middle: octave. Right: tritave (circle: mixture, dashed line: Original A3, triangle:* $\gamma = 0$*, solid line:* $\gamma = 0.5$

and 0.5, respectively, showing the challenge of separating octave notes.

Table 1: *The piano cases: The evaluation*

		Fifth		Octave		Tritave	
γ		0	0.5	0	0.5	0	0.5
Preceding note	SIR	14.7	14.7	4.5	5.2	17.2	17.0
	SAR	11.0	11.0	6.4	6.2	11.0	10.9
	SDR	9.2	9.2	1.3	1.6	9.9	9.8
	MER	13.8	14.0	9.1	9.3	13.6	13.3
Succeeding note	SIR	23.8	23.5	13.3	13.7	22.7	22.9
	SAR	15.0	15.0	12.9	12.8	15.3	15.2
	SDR	14.3	14.3	9.9	10.0	14.3	14.3
	MER	8.4	8.4	8.2	8.2	7.0	7.0

4.3. Guitar

We also present the cases of guitar sounds. Similarly, three cases with three sets of pitch combinations are considered: perfect fifth (i.e. A3+E4, E3+B3, and G3+D4), octave (i.e. A3+A4, E3+E4, G3+G4) and tritave (i.e. A3+E5, E3+B5, G3+D5). Figure 4 shows the temporal envelopes of the separated notes and Table 2 show the evaluation results. We also observe the suppression of temporal magnitude modulation, and a slightly improvement in terms of SIR, SAR, SDR and MER, in general. For example, in the case of perfect fifth, the SDR of preceding notes increases from 8.3 to 10.4 dB, SIR from 14.0 to 17.4 dB, and MER from 13.9 to 14.2 dB.

4.4. Violin

Finally, for violin, we consider the following cases: perfect fifth (i.e. A3+E4, B3+F4, and G3+D4), octave (i.e. A3+A4, B3+B4, G3+G4) and tritave (i.e. A3+E5, B3+F5, G3+D5). The experiment result is presented in Figure 5 and Table 3. In comparison to the cases of piano and guitar, the proposed temporal smoothness constraint is found more effective for violin. Not only can we see the suppression of the magnitude modulation by the higher note, but we also see general improvement of SIR, SAR, SDR and MER. Specifically, SIR is improved the most: in the case of tritave, the SIR of preceding notes is improved by 4.7 dB, and SDR by 4 dB. In all, the SIRs of the three cases are improved by at least 1.5 dB to 4.7 dB.

Table 2: *The guitar cases: The evaluation*

		Fifth		Octave		Tritave	
γ		0	0.5	0	0.5	0	0.5
Preceding note	SIR	14.0	17.4	12.5	13.6	16.7	18.4
	SAR	12.2	12.1	11.4	11.2	13.3	12.4
	SDR	8.3	10.4	7.0	7.8	9.7	10.3
	MER	13.9	14.2	17.0	17.0	17.5	17.0
Succeeding note	SIR	27.5	28.2	24.4	24.4	24.9	25.2
	SAR	13.6	13.9	13.9	13.8	11.9	12.2
	SDR	12.9	13.3	12.6	12.5	11.2	11.4
	MER	4.0	4.1	4.5	4.5	2.9	2.9

4.5. Discussion

First, according to the experimental results, the proposed temporal constraint approach greatly reduces the temporal magnitude modulation when separating notes with highly overlapping partials using CMF. In some cases we further observe improvement of the objective figures of merit. In particular, the SIR is increased by more than 4 dB in the cases for long-sustain notes like violin. However, for stuck-string or plucked-string instruments like piano or guitar, the objective performance indices are just marginally improved or unimproved. It is noted that the temporal constraint parameter shouldn't be too large and should be chosen carefully. Empirically, the parameter should be chosen between 0 and 1. Finally, in conventional CMF algorithm [15–18], the template matrix representing the frequency domain characteristics is fixed for all input features in the whole music piece. This is certainly unreasonable in most cases. As a fixed set of template is unsuitable for the modeling of vibrato, which is usually seen in violin notes. Since the objective measures such as SDR and SIR may not well reflect whether we have properly taken care of the modulation effect, in

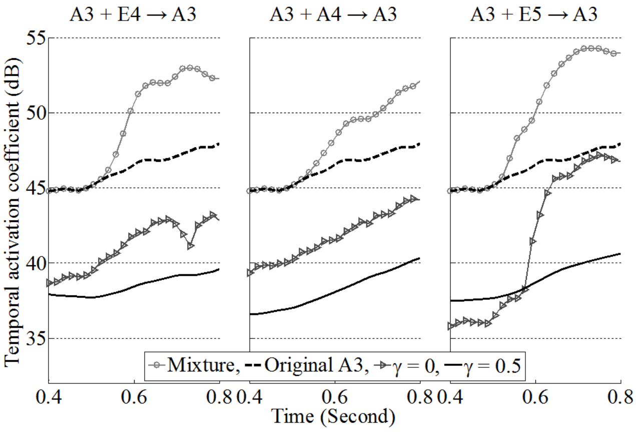

Figure 5: *Three cases of violin. Left: perfect fifth. Middle: octave. Right: tritave (circle: mixture, dashed line: Original A3, triangle:* $\gamma = 0$*, solid line:* $\gamma = 0.5$

Table 3: *The violin cases: The evaluation*

		Fifth		Octave		Tritave	
γ		0	0.5	0	0.5	0	0.5
Preceding note	SIR	10.2	11.6	8.7	12.6	17.2	21.9
	SAR	7.3	6.2	6.3	5.1	16.0	13.3
	SDR	4.6	4.6	3.3	4.0	9.0	12.0
	MER	4.0	3.2	4.4	3.3	2.4	3.8
Succeeding note	SIR	10.9	9.7	18.4	17.8	-0.6	0.4
	SAR	14.1	14.6	16.9	16.7	9.1	9.7
	SDR	8.9	8.3	10.1	9.5	-2.6	-1.7
	MER	6.4	7.6	10.3	12.0	5.8	10.5

the future we will also provide audio examples for subjective evaluation [22].

5. CONCLUSION AND FUTUREWORKS

Though CMF outperforms NMF in note separation applications, severe temporal magnitude modulation is presented when the notes have highly overlapping parts, especially in the cases of octaves. In this paper, the conventional CMF is combined with a temporal smoothness constraint, which not only reduces the temporal magnitude modulation but also improve the performance figures of merit including SIR, SDR, and SAR, and MER. Although the proposed method improves the original CMF note separation, the overall results are still unsatisfactory. Particularly, in the case of octave, all the objective figures of merit shown in this paper are still poor in comparison to other cases. Moreover, when processing vibrato notes, the situation is even worse, perhaps because that the template matrix corresponding to the frequency domain characteristics is fixed in the time domain during the separation process. This suggests a future work of developing a better solution such that the template matrix can be adaptable to the temporal of the notes.

6. ACKNOWLEDGMENT

The authors would like to thank the National Science Council, ROC, for its financial support of this work, under Contract No.MOST 103-2221-E-006-140-MY3.

7. REFERENCES

[1] G. Monti and M. B. Sandler, "Automatic polyphonic piano note extraction using fuzzy logic in a blackboard system," in *Proc. of the 5th Int. Conference on Digital Audio Effects (DAFx-02)*, 2002, number OCTOBER 2002, pp. 39–44.

[2] Y. J. Lin, T. M. Wang, T. C. Chen, Y. L. Chen, W. C. Chang, and A. W. Y. Su, "Musical note analysis of solo violin recordings using recursive regularization," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 25, pp. 1–13, 2014.

[3] Y. J. Lin, W. C. Chang, and T. M. Wang, "Timbre-constrained recursive time-varying analysis for musical note separation," in *Proc. of the 16th ...*, 2013, pp. 3–7.

[4] T. M. Wang, T. C. Chen, Y. L. Chen, and A. W. Y. Su, "Time-Dependent Recursive Regularization for Sound Source Separation," in *Audio, Language and Image Processing (ICALIP), 2012 International Conference on*, 2012, pp. 235–240.

[5] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[6] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary.," *Neural computation*, vol. 13, no. 4, pp. 863–882, 2001.

[7] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, 2001.

[8] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in neural information processing*, , no. 1, 2000.

[9] P. Smaragdis, "Non-negative matrix factorization for polyphonic music transcription," in *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on*, 2003, number 1, pp. 177–180.

[10] P. Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs," *Independent Component Analysis and Blind Signal Separation*, , no. 2, pp. 494–499, 2004.

[11] A. Cichocki, R. Zdunek, and S. Amari, "New algorithms for non-negative matrix factorization in applications to blind source separation," in *Acoustics, Speech and Signal Processing (ICASSP), 2006 IEEE International Conference on*, 2006, pp. 621–624.

[12] M. N. Schmidt and R. K. Olsson, "Single-channel speech separation using sparse non-negative matrix factorization," in *Spoken Language Processing ISCA International Conference on INTERSPEECH*, 2006, vol. 2, pp. 2–5.

[13] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1066–1074, 2007.

[14] R. Hennequin, R. Badeau, and B. David, "Time-dependent parametric and harmonic templates in non-negative matrix factorization," in *Proc. of the 13th International Conference on Digital Audio Effects (DAFx)*, 2010, number 1, pp. 1–8.

[15] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF: A new sparse representation for acoustic signals," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2009, vol. 1, pp. 2353–2356.

[16] B. J. King and L. Atlas, "Single-Channel Source Separation Using Complex Matrix Factorization," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 8, pp. 2591–2597, 2011.

[17] B. J. King, *New methods of complex matrix factorization for single-channel source separation and analysis*, Ph.D. thesis, University of Washington, 2012.

[18] J. Bronson and P. Depalle, "Phase constrained complex NMF: Separating overlapping partials in mixtures of harmonic musical sources," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2014, pp. 7475–7479.

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

[20] ETSI TR101290, "Digital video broadcasting (DVB); measurement guidelines for DVB systems," 2001.

[21] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: Music genre database and musical instrument sound database," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2003, number October, pp. 229–230.

[22] Y. J. Lin, "Music Files of SEPARATION OF MUSICAL NOTES WITH HIGHLY OVERLAPPING PARTIALS USING PHASE AND TEMPORAL CONSTRAINED COMPLEX MATRIX FACTORIZATION," 2015.