TIMBRE-CONSTRAINED RECURSIVE TIME-VARYING ANALYSIS FOR MUSICAL NOTE SEPARATION

Yiju Lin, Wei-Chen Chang, Tien-Ming Wang, Alvin W.Y. Su,
SCREAM Lab., Department of CSIE,
National Cheng-Kung University,
Tainan, Taiwan
lyjca.cs96@g2.nctu.edu.tw

Wei-Hsiang Liao,
Analysis/Synthesis Group, IRCAM,
Paris, France
wliao@ircam.fr

ABSTRACT

Note separation in music signal processing becomes difficult when there are overlapping partials from co-existing notes produced by either the same or different musical instruments. In order to deal with this problem, it is necessary to involve certain invariant features of musical instrument sounds into the separation processing. For example, the timbre of a note of a musical instrument may be used as one possible invariant feature. In this paper, a timbre estimate is used to represent this feature such that it becomes a constraint when note separation is performed on a mixture signal. To demonstrate the proposed method, a time-dependent recursive regularization analysis is employed. Spectral envelopes of different notes are estimated and a modified parameter update strategy is applied to the recursive regularization process. The experiment results show that the flaws due to the overlapping partial problem can be effectively reduced through the proposed approach.

1. INTRODUCTION

Audio source separation attracts more and more attentions from researchers in the last decade. One major reason is that lots of signal decomposition techniques have been well developed both in theoretical and practical sides. Especially, Nonnegative Matrix Factorization (NMF) with carefully-designed constraints shows great potential to deal with spectral data decomposition problems[1][2]. In practical, conventional NMF decompose the magnitude spectrogram of a given signal into a set of template (column) vectors and intensity (row) vectors and usually suffers from two problems. First, there is no guarantee that NMF can always converge to the same answer every time when it is performed on the same signal. Secondly, NMF is usually applied to lots of frames of data at a time such that it is less suitable for time varying signals. Therefore, NMF needs to add specific constraints for musical source separation. For example, Romain and et td-al. presented a parametric model called time-dependent NMF (TD-NMF) to limit template vectors by harmonic combs[1]. The constraint allows only solutions that are valid within the model and offers a high degree of robustness. To focus on the local characteristics of the notes to be separated, our previous work used a time-dependent recursive regularization (TD-RR) analysis in [3]. The matrix inversion operation is almost eliminated to bring down the computational complexity to the level of NMF based methods.

However, when decomposing musical notes with overlapping partials from an audio mixture, one always encounters the problem of how to determine the energy ratios of those overlapping partials belonging to the co-existing notes. A direct and quick solution is to have a prior musical instrument models [4][5]. However, the assumption that the specific musical instrument models are known is only under some particular recording circumstances. In most real-world applications, for example, to extract violin solo part from live violin concerto recordings, such assumption cannot be applied.

In general, musical signals are characterized by the sounding mechanism of a specific musical instrument which has very diverse components such like, strings, bridges, reeds, resonant vibrators, and etc [6]. In a linear system point of view, the musical signal of a specific timbre is produced by passing a simple excitation through a system (or a filter) consisting of its physical components. Generally speaking, a timbre feature may have two aspects. The first aspect is a certain fixed presentation resulted from the musical instrument's physical mechanism. The other is its dynamic temporal evolution due to the continuous excitation to the mechanism when the musical instrument is played. For example, the timbre of a musical instrument tends to vary smoothly and slowly in a certain period of time. In such a sense, it may be distinguished from the other instruments.

Timbre as one of the most important features in human aural perception is discussed and modelled in many musical applications, such as musical signal analysis/synthesis [7], musical instrument recognition [8], and music retrieval [9]. In this paper, we propose a timbre constraint to guide the note separation process when there are overlapping partials coming from different notes. By limiting the energy ratios of the overlapping partials, musical instruments' timbre features take effects in the note separation procedure. In particular, estimated spectral envelops of notes are used as our timbre constraint. Specifically, clips of the 1959 recording of Beethoven violin concerto played by David Oistrakh with Andre Cluytens conducting the French National Radio Orchestra [10] are used to demonstrate our algorithm in this paper. More musical note separation results can be heard at our website [11].

The rest of the paper is organized as follows. More about the timbre feature are discussed and the idea of timbre function is described in Section 2. Formulations of TD-RR and some background techniques are described in Section 3. The timbre constraint and the modified procedure are shown in Section 4. Experiments and results are given in Section 5. Finally, conclusions are drawn in Section 6.

2. TIMBRE ESTIMATE

In [1], Hennequin described a model to determine a set of partial magnitudes produced by a harmonic musical instrument. The model assumes the relationship of partial magnitudes of a note is fixed throughout the entire analysis period. This approach was capable of dealing with the overlapping partial problem if there existed sufficient number of frames within which there were non-overlapping notes. It may also solve the first aspect we had discussed in Section 1. That is the musical instrument had its fixed physical mechanism and produced its sound with a fixed spectral presentation. However, it didn't address the second aspect that the relationship of partial magnitudes can't be fixed for many musical instruments, such as violin, or for special performing techniques, such as vibrato.

Although the timbre is intuitive to human aural perception and understanding, it is not so obvious to observe such a feature using just one analysis frame. To be specific, if we can locate the fundamental frequency and its partials in the spectrum of a musical note, we can easily estimate a smooth spectral envelope from the amplitudes of its partials by methods such as Linear Prediction(LP)[12].

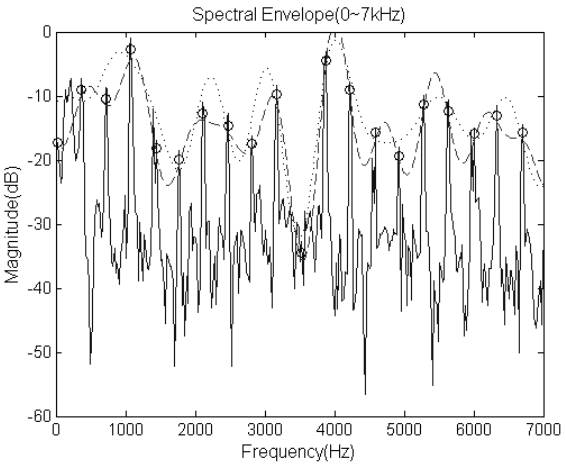

Figure 1: *Two possible spectral envelopes estimated from a guitar note (dashed lines: order-14 and -16 LPC estimation results, solid line: spectral magnitudes, circle: harmonic partials).*

For example, we took a spectrum of a guitar note and estimated two smooth spectral envelopes with the LP analysis method by using two different numbers of orders. The results are shown in Fig. 1. The two spectral envelopes both satisfied this harmonic set which is a single observation of the timbre of the note. Therefore, it is hard to say which one is more suitable to characterize the timbre. Thus, one needs more observations to determine what the true timbre may be. In this paper, it is preferred to estimate a timbre function in a small number of analysis frames in order to capture the local characteristics. Without loss of generality, we will derive our formulas based on the following assumption. For a harmonic musical instrument, the timbre of a note doesn't change much in a short duration throughout the temporal evolution of a note in the nearly stationary period.

We first consider the timbre function of a specific note *i* of a musical instrument *j* in a short duration of time, denoted as $T_{ij}(f)$, where *f* is the frequency index. If there are *I* notes of *J* musical instruments sounding together in a period of time, the amplitude of frequency index *f* should be equal to $T_{ij}(f)$ if there is no overlapping partials. That is, the energy of frequency index *f* belongs to note *i* of musical instrument *j* alone. Otherwise, it is equal to $\sum_{i\in I,\, j\in J} T_{ij}(f)$ because the energy of frequency index *f* comes from several different tones. It is noted that the phase information is omitted to keep the problem formulation simple. If the partials can be preliminarily separated from the mixture signal through source separation processing, a timbre constraint can be consequently applied onto the estimation of the amplitudes of all overlapping partials for each note of each musical instrument. In practice, difficulties usually occur. We will leave the details in section 4.

3. TIME DEPENDENT RECURSIVE REGULARIZATION ANALYSIS

Before introducing our timbre constraint, we need to describe TD-RR method in advance. Given the magnitude spectrogram of a mixture signal $V \in \Re_{\geq 0}^{M\times N}$ and the number of tone models $R \in \aleph$, classical NMF methods derive two non-negative matrices $W \in \Re_{\geq 0}^{R\times N}$ and $H \in \Re_{\geq 0}^{M\times R}$ such that a distance function $D(V, HW)$ is minimized:

$$V \approx \tilde{V} = H \times W . \tag{1}$$

In [3], the cost function consisted of additional penalty terms C_H and C_W shown in equation (2) can evaluate how well the multiplication of *H* and *W* can approximate *V*.

$$D = \|V - H\times W\|^2 + \lambda\|W - C_W\|^2 + \gamma\|H - C_H\|^2 , \tag{2}$$

where λ and γ are the corresponding regularization parameters. The template matrix *W* and the intensity matrix *H* can be obtained as

$$W = \left(H^T \times H + \lambda\cdot I\right)^{-1}\left(H^T \times V + \lambda\cdot C_W\right). \tag{3}$$

$$H^T = \left(W\times W^T + \gamma\cdot I\right)^{-1}\left(W\times V^T + \gamma\cdot C_H^T\right). \tag{4}$$

Unlike NMF, the above factorization of a nonnegative matrix may not always produce two nonnegative matrices. An empirical solution to keep the nonnegative property is to set the negative elements of W and H to zeros and re-evaluates these two equations until the nonnegative results are finally obtained.

Let the *R*-by-*N* template matrix for the *l*-th input frame be denoted as $W(l)$. The corresponding input matrix and intensity matrix are denoted as $V(l)$ and $H(l)$. According to the derivation in [3], a set of recursive frame-wise regularization equations can be acquired as the following equations.

$$W(l) = \left(H^T(l)\times H(l) + \lambda\cdot I\right)^{-1}\times \left(H^T(l)\times V(l) + \lambda\cdot W(l-1)\right) \tag{5}$$

$$H^T(l) = \left(W(l)\times W^T(l) + \gamma\cdot I\right)^{-1}\times \left(W(l)\times V^T(l) + \gamma\cdot H^T(l-1)\right) \tag{6}$$

In equation (5) and (6), $C_W(l)$ and $C_H^T(l)$ are set to $W(l-1)$ and $H^T(l-1)$ because it is assumed that the decomposing atoms and their intensities shall not change abruptly. The matrix inversion operation is eliminated to reduce the computational complexity. Then, the time-varying template matrix and the corresponding intensity matrix would be calculated iteratively when a new input frame is provided and the earliest frame is excluded.

4. TIMBRE CONSTRAINT TD-RR

The penalty term C_W in equation (2) originally refers to the harmonicity constraint of a note based on its fundamental frequency. As shown in equation (7), u_1 is the reference of the guard template (noise template) and $u_n(n>1)$ are the reference templates of notes.

$$C_W = [u_1\, u_2 \cdots u_N]^T . \tag{7}$$

For each reference template, it is constructed by using equation (8) which is the sum of a series of bell-shape functions, for example, Gaussian functions, based on the note's fundamental frequency and harmonics. In equation (8), $g_{n,p}$, the gain factor typically related to the previous estimated template, is applied to each Gaussian function G for enhancing the constraint. Such a method was adopted in both [1] and [3]. Empirically speaking, σ in equation (8) is chosen to make the bell-shape curve to cover a small frequency range around the harmonics.

$$u_n = \sum_p g_{n,p} G(pf_0, \sigma) . \tag{8}$$

To force a timbre constraint on these interested harmonic positions, a new update rule for gain factors is introduced. Suppose the amplitude of the *p*th partial of fundamental frequency $f_{0,t}$ for note *i* of musical instrument *j* in the instant time *t* is defined as $a_{p,t} = T_{ij}(pf_{0,t})$ if it isn't an overlapping partial. In particular, the partials will only reveal a sampled version of the instrument's resonance characteristic. When there is a small variation in both fundamental frequency and amplitude, we have a group of observations in a short analysis period, defined as $(F, A)_{i,j} = (pf_{0,t}, a_{p,t})_{i,j}, \forall p,t,i,j$.

Following the discussion in Section 2, $T_{ij}(f,t)$ varies little within a short period of time, i.e. $T_{ij}(f,t) = T_{ij}(f)$. Because $T_{ij}(f)$ is a spectral envelope, it is a non-negative function. Hence, its polynomial regressive approximation can be calculated in the TD-RR iterative update process based on the observed group in one template vector. Such an approximation of a timbre function of instrument *j* is denoted as $\hat{T}_{ij}(f)$. That is

$$\hat{T}_{ij}(f) = \sum_k a_k f^k + \varepsilon . \tag{9}$$

This regression model consists of a polynomial parameter a_k and an error term ε. Furthermore, it can be expressed in a matrix form in terms of an amplitude vector *A*, a partial's frequency vector *F*, a parameter vector α, and a random error vector *E*.

$$A = F\alpha + E . \tag{10}$$

The parameter vector α is then estimated in the least square sense:

$$\alpha = (F^T F)^{-1} F^T A . \tag{11}$$

After the timbre function is regressively determined, a modified update rule for u_n is then given by

$$u_n = \sum_p \hat{T}_{ij}(pf_0) G(pf_0, \sigma) . \tag{12}$$

The template-dependent gain factors defined in equation (8) are now determined by the estimated timbre function. An illustration indicated the modified C_W update procedure is shown in Fig. 2. To focus on timbre evaluation, we only showed the partial positions in the estimated W. When the estimated W is iteratively calculated by equation (5), it is used to regressively estimate a new timbre function. This new timbre function constructs a new C_W by equation (7) and (12). This update procedure is incorporated with the analysis process of TD-RR described in Section 3.

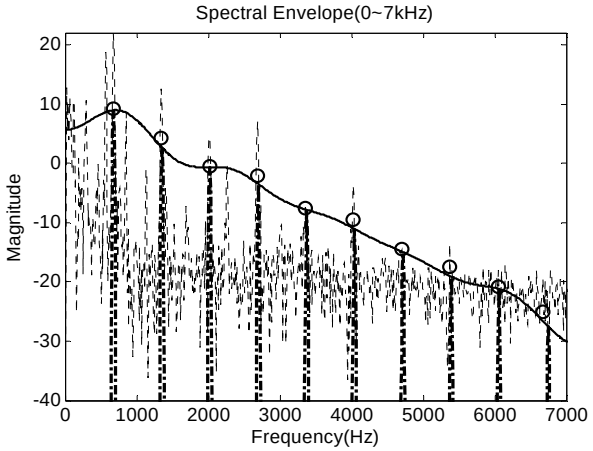

Figure 2: *C_W with timbre constraint (dashed line: spectrum of original polyphonic signal, solid circle: partial positions of the estimated W, solid line: estimated timbre function, bold dash-dot line: new estimated C_W).*

Although the distribution of the energy of an overlapping partial to different notes wasn't discussed and implemented in a particular processing, it is done through the competition among different template vectors, i.e. different notes, in the TD-RR procedure. Since the proposed timbre constraint has already restricted the penalty terms for corresponding template vectors, the separated note can keep a similar and smoothly-changed timbre when there are co-existing notes with overlapping partials.

5. EXPERIMENTS

We evaluated the proposed method with three artificial cases: one non-overlapping partial case and two overlapping partial cases. Each of the three cases is combined with two single notes to represent the specific situations. All notes are chosen from RWC Musical Instrument Sound Database [13]. The four test notes, C4, D4#, G4 and C5, are violin (I151) sounds with normal playing styles and the volume is at medium level. The proposed method is also tested using a commercial acoustic recording,

Beethoven violin concerto played by David Oistrakh [10]. The details will be described later.

As a control case, the non-overlapping case in Fig. 3 shows the comparable qualities for both results of TD-NMF and TD-RR with timbre constraint. In overlapping partial cases, we tried to demonstrate the effectiveness of the proposed timbre constraint design. In Fig. 4 and 5, the overlapping partials appeared in the second harmonic position and in the third harmonic position respectively. The results of TD-RR with timbre constraint had sharper and clearer harmonic partials when compared to the results of TD-NMF, especially in the high frequency range.

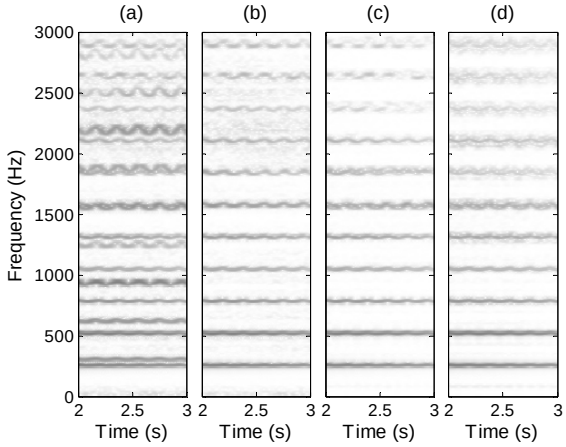

Figure 3: *Non-overlapping partial case: (a) original mixture (C4+D4#), (b) original C4, (c) C4 extracted by TD-RR with timbre constraint, (d) C4 extracted by TD-NMF.*

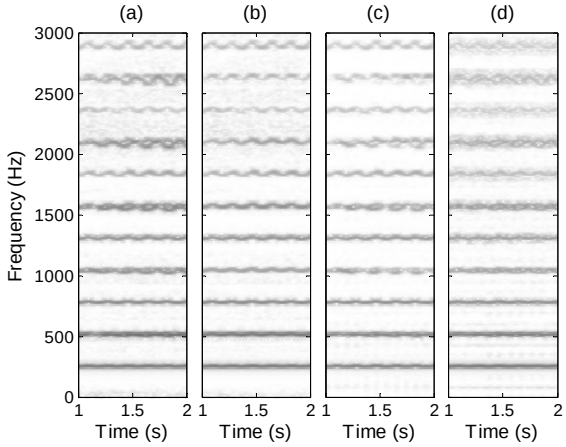

Figure 4: *Overlapping partial case 1 - Octave: (a) original mixture (C4+C5), (b) original C4, (c) C4 extracted by TD-RR with timbre constraint, (d) C4 extracted by TD-NMF.*

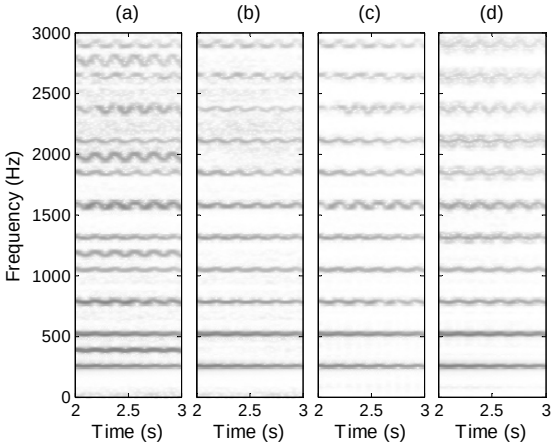

Figure 5: *Overlapping partial case 2 - Quint: (a) original mixture (C4+G4), (b) original C4, (c) C4 extracted by TD-RR with timbre constraint, (d) C4 extracted by TD-NMF.*

Two special real life performance test cases are also demonstrated as follows. They are extracted from the 1959 recording of Beethoven violin concerto played by David Oistrakh with Andre Cluytens conducting the French National Radio Orchestra [10]. The first one is a trill clip which appears in the 143rd bar of the 1st movement. As shown in Fig. 6, the result of TD-RR with timbre constraint shows clear start points and stop points where two notes take turns, especially in partials higher than the fifth one. The second one is a vibrato clip which appears in the 92nd bar of the 3rd movement. In Fig. 7, one can observe strong accompaniment musical instruments played in the background. The result of TD-RR with timbre constraint resists more interference and shows sharper partials than that of TD-NMF. Here, TD-NMF result has some band-limited artifact. It might result from its small harmonic bandwidth configuration for the bell-shape functions used in equation (8). A large harmonic bandwidth setup will probably improve the result. However, these comparisons are based on the same harmonic bandwidth configuration for both TD-NMF and TD-RR with timbre constraint.

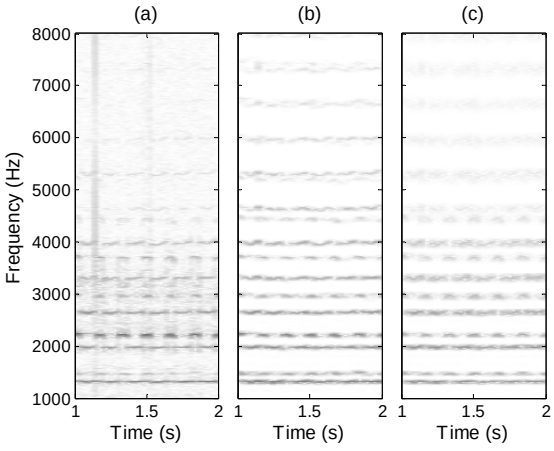

Figure 6: *Beethoven violin concerto played by David Oistrakh – trill: (a) original trill sound (between E5 and F5#), (b) trill sound extracted by TD-RR with timbre constraint, (c) trill sound extracted by TD-NMF*

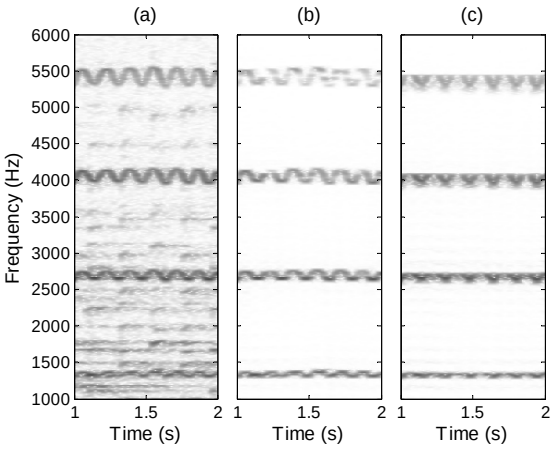

Figure 7: *Beethoven violin concerto played by David Oistrakh - vibrato: (a) original vibrato sound (around E6), (b) vibrato sound extracted by TD-RR with timbre constraint, (c) vibrato sound extracted by TD-NMF.*

6. CONCLUSIONS

A new musical note separation method for polyphonic recordings is presented. In this paper, we have proposed a modified TD-RR analysis incorporated with timbre constraints to determine the energy ratios of overlapping partials of simultaneous musical notes. When the parameters of TD-RR were updated, several timbre functions of corresponding specified templates were estimated as the upper bounds of their partials' amplitudes and were used to redistribute the overlapping partials' energies. A commercial acoustic recording of Beethoven's violin concerto is included in the experiments. As shown in experimental results, the proposed method achieved better results than TD-NMF. The separated results have appropriately preserved the desired timbre and have less interference with the subsequent notes.

The techniques introduced in this paper showed its potential in music signal analysis. More experiments will be arranged to improve its robustness. One future work is essentially related to the timbre feature extraction and aim at developing a robust parametric model for timbre re-synthesis or transformation. The sound examples can be heard at our website [11].

7. ACKNOWLEDGEMENT

The authors would like to thank the National Science Council, ROC, for its financial support of this work, under Contract No.NSC 100-2221-E-006-247-MY3.

8. REFERENCES

[1] R. Hennequin, R. Badeau, and B. David, "Time-dependent parametric and harmonic templates in non-negative matrix factorization," in *Proc. of the 13th Int. Conference on Digital Audio Effects*, Graz, Austria, 2010.

[2] Tuomas Virtanen, "Monaural sound source separation by nonnegative matrix factorizationwith temporal continuity," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1066–1074, March 2007.

[3] T.-M. Wang, T.-C. Chen, Y.-L. Chen, Alvin W.Y. Su, "Time-dependent recursive regularization for sound source separation," in *Proc. of the 3rd International Conference on Audio, Language and Image Processing* (*ICALIP2012*), Shanghai, China, Jul. 16-18, 2012.

[4] E. Vincent, "Musical source separation using time-frequency source priors," *IEEE Trans. Audio Speech Language Process.*, vol. 14, no. 1, pp. 91-98, 2006.

[5] M. Bay and J. W. Beauchamp, "Harmonic source separation using prestored spectra," in *Proc. ICA*, pp. 561-568., 2006.

[6] Neville Horner Fletcher, Thomas D. Rossing. *The Physics of Musical Instruments*,2nd ed., New York, Springer, 1998.

[7] H. Hahn, A. Röbel, J. J. Burred, and S. Weinzierl, "Source-filter model for quasi-harmonic instruments," in *13th International Conference on Digital Audio Effects*, September 2010.

[8] J.J. Burred, A. Röbel, and T. Sikora, "Dynamic spectral envelope modeling for timbre analysis of musical instrument sound," *IEEE Transactions on Audio, Speech and Language Processing*, March 2010.

[9] Aucouturier, J.-J., Pachet, F. and Sandler, M., "The Way It Sounds: timbre models for analysis and retrieval of polyphonic music signals," *IEEE Transactions of Multimedia*, 7(6):1028-1035, 2005.

[10] David Oistrakh, *Beethoven violin concerto in D major*, op.61, SXLP 30128, OC 047 ◦ 90905, EMI Records Ltd., 1959.

[11] Yiji Lin, "Timbre-constrained Recursive Time-varying Analysis." Available at: http://screamlab-ncku-2008.blogspot.tw/2013/04/music-files-of-timbre-constrained.html, Accessed April 14, 2013.

[12] J. Makhoul, "Linear Prediction: a tutorial review," *Proceedings of the IEEE*, vol. 63, pp. 561-580, 1975.

[13] Masataka Goto, "Rwc music database: Music genre database and musical instrument sound database," in *Proc. of the 4th International Conference on Music Information Retrieval* (*ISMIR 2003*) , pp. 229–230, Baltimore, Maryland, USA, October 27-30 2003.