Quantitative Evaluation of Violin Solo Performance

Yiju Lin, Wei-Chen Chang and Alvin W.Y. Su

SCREAM Lab., Department of Computer Science and Engineering, National Cheng-Kung University, Tainan, Taiwan, R.O.C.
E-mail: alvinsu@mail.ncku.edu.tw Tel: +886-6-2757575-62537

***Abstract*— Evaluation of performances of musical instruments is usually subjective. It may be easier for keyboard instruments. For bowed-string instruments such as a violin, delicate articulations are required in its performance such that there exhibit much complexities in its sound, making the evaluation more difficult. In this paper, a note separation algorithm based on spectral domain factorization is used to extract the notes from recordings of violin solo performances. Each note can then be quantitatively evaluated based on a set of metrics that is designed to provide various aspects of violin performances including pitch accuracy, bowing steadiness, vibrato depth/rate, bowing intensity, tempo, and timbre characteristics and so on. The tools should be useful in musical instrument performance education.**

I. INTRODUCTION

Music performance review and criticism have been a difficult and subjective task [1]-[4]. It may be arguable that objective evaluation of truly artistic performances is appropriate. However, it is definitely useful that certain physical aspects of sounds of music performances can be accurately delineated if a player wants to know exactly how he/she plays the instruments. Furthermore, it is possible to disclose the differences among violinists in order to provide more objective description of styles.

In the research areas of music information retrieval, well-developed technologies such as onset detection [5], pitch estimation [6], score alignment [7], beat tracking [8] and so on can be applied to extract physical features related to musical sounds. Quantitative indices can be derived based on the features presented in [9], too. In this paper, violin solo performances are studied and some useful quantitative indices will be created to describe both the sound and the performance. Because of the sustaining driven characteristics of violins, properties such as pitch, timbre, and volume keep changing during the play [10]. Therefore, it is also desired that the indices should be time varying, too.

A major difficulty of analysis of violin solo performances is to separate the violin part from the accompaniment part of acoustic recordings. Then, it is also necessary to divide the violin solo part into isolated notes. Audio source separation has been a very popular topic in the last few years [11]-[14]. In [15], a NMF (Nonnegative Matrix factorization) based method is proposed for the analysis of bowed-string instruments. In [16], a time dependent recursive regularization (TD-RR) approach for musical note separation is proposed. In the above methods, the magnitude spectrum of the signal is divided into two nonnegative matrices which represent the spectral template information and the associate intensity information, respectively. We will use the template matrix and the intensity matrix obtained by using TD-RR method to generate the quantitative indices of violin performances.

The rest of the paper is organized as follows. Section 2 reviews the note isolation algorithm employed in this paper. Section 3 presents the quantitative indices used to describe violin solo performances. In Section 4, experiments using different commercial recordings of Bach's violin solo work are presented. Conclusion and future works are in section 5.

II. NOTE SEPARATION OF MUSICAL RECORDINGS

In [16], for the magnitude spectrogram of a mixture signal $V \in \Re_{>0}^{M\times N}$ and the number of tone models $R \in \aleph$, factorization of V generates two nonnegative matrices, the spectral template matrix $W \in \Re_{>0}^{R\times N}$ and the intensity matrix $H \in \Re_{>0}^{M\times R}$, as shown in (1). M is the number of frequency bins and N is the number of audio frames involved in the analysis.

$$V \approx \widetilde{V} = H \times W . \quad (1)$$

The cost function in (2) contains the distance between V and $H \times W$ in the mean square sense and two additional regularization terms.

$$D = \|V - H\times W\|^2 + \lambda\|W - C_W\|^2 + \gamma\|H - C_H\|^2 . \quad (2)$$

With the above cost function minimized, one can obtain W and H by

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

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

In (2), the two regularization terms are used to constraint the form of the target matrices. Unlike conventional NMF methods, equation (3) and (4) may not always produce nonnegative results. By setting C_W and C_H as nonnegative matrices, the results tend to be nonnegative too. For temporal smoothness, it is desired that both template matrix and intensity matrix don't vary too much in a short period of time. Therefore, one can set C_W and C_H to be the analysis results of the previous time instant. Moreover, if the timbre of the target musical instrument is known, one could constraint the template matrix by using the timbre information as follows. Let the reference template matrix be

$$C_W = [u_1 \; u_2 \cdots u_r \cdots u_R]^T, \tag{5}$$

where u_r is used to represent the constraint template column vector for the *r-th* tone model. In [14] and [16], u_r can be obtained as the weighted sum of bell-shape functions,

$$u_r = \sum_p g_{r,p} G(pf_r, \sigma). \tag{6}$$

In (6), G is Gaussian function, f_r is the fundamental frequency of the *r-th* tone model, p is the partial index, and σ is used to control the width of Gaussian function. $g_{r,p}$ is the parameter to control the intensity of the *p-th* partial of the *r-th* tone model. If the timbre of the tone model is known in advance, $g_{r,p}$ can be a fixed parameter. Otherwise, one has to estimate these parameters during the analysis process using the method proposed in [17]. Hence, one could combine several constraints as

$$\begin{aligned} u_r(t) = \alpha \cdot W_r(t-1) + \\ (1-\alpha) \cdot \sum_p g_{r,p}(t-1) G\big(pf_r(t-1), \sigma\big), \end{aligned} \tag{7}$$

where $W_r(t-1)$ is the template column vector of the r-th tone model obtained in the previous time instant. In (7), we employ $W_r(t-1)$, the estimated fundamental frequency and the estimated timbre obtained at time t-1 to have the constraint template column vector in order to calculate $W_r(t)$ in (3).

Fig. 1 shows the intensity matrix derived from the analysis result of a short clip of Bach Sonata No.3 in C major, BWV 1005, Adagio, the 14-th measure, played by Hillary Hahn. There are 7 notes in the clip. Each note is separated to give a clear view of how the note was played. The template matrix is obtained without the knowledge of timbre information. It shows how the timbre changes in the period of a note. One can also observe how vibrato was executed. The intensity matrix shows the onset time, the offset time and how loud the note was played. We will describe how the information can be converted into the quantitative indices in the later section.

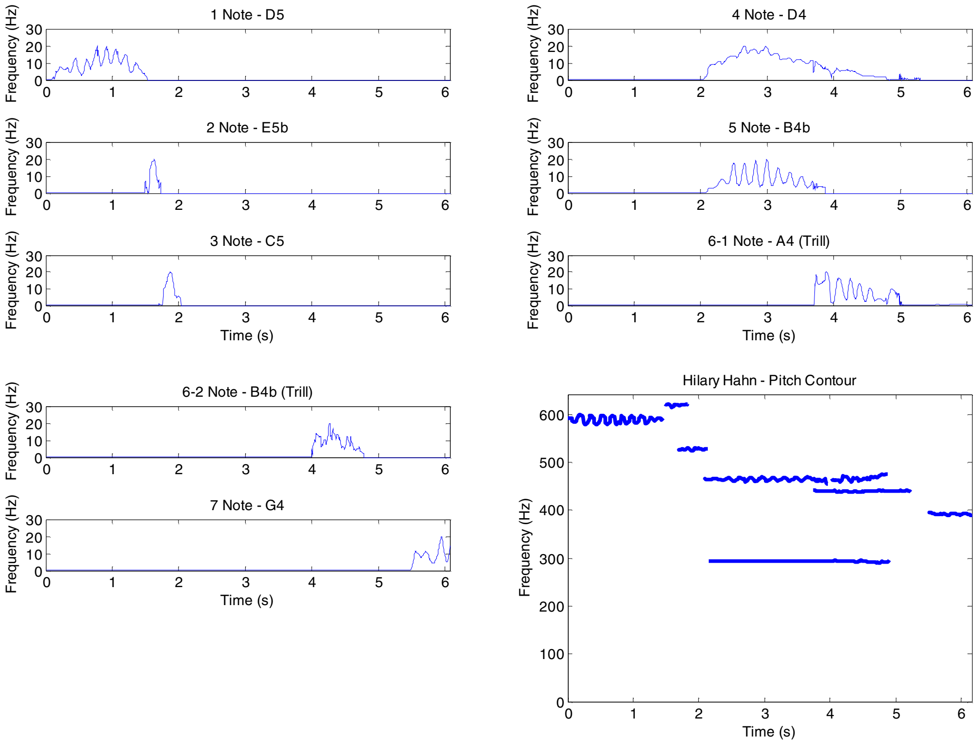

Fig. 1 Separated results from a clip of Bach Sonata No.3 in C major, BWV 1005, Adagio, the 14-th measure, played by Hillary Hahn. There are seven notes and their note names are shown correspondingly. The first 3 sub-figures show the intensity matrix. The fourth sub-figure shows the derived pitch contour.

III. OBJECTIVE INDICES OF VIOLIN PERFORMANCE

Many performances can be characterized by pitch and intensity evolution of their sounding notes. In this paper, we propose a new quantitative description, called the steadiness index, which indicates how stable the sounding note is over time when the performer is playing the instrument. It is also designed to provide an overall rank in the period of a note. To be specific, the pitch contour $P(\cdot)$ and the intensity contour $I(\cdot)$ can be fundamental start points for our quantitative description of a violin note. Since a note is said sounded or voiced, both contours are expected to be smooth and continuous curves over a limited time period T. Both contours behave specific characteristics in different sounding modes. For example, a violin note can be either a vibrato note or a non-vibrato note, so we need to evaluate a note in two different ways. In Figure 1, the converted pitch contour which is derived from the intensity matrix and template matrix is shown. One is a vibrato note and the other one isn't. In this paper, steadiness measure is used as the performance index. For a non-vibrato segment of a note whose duration is $T_s = T/N$, we estimate the pitch mean $\overline{P}(n)$ and variance $\sigma_p^2(n)$ as well as the intensity mean $\overline{I}(n)$ and variance $\sigma_I^2(n)$. The steadiness measure of a non-vibrato segment is defined as

$$s_{nv} = \frac{1}{N}\sum_{n=1}^{N}\frac{2\sigma_p}{\overline{P}(n)\times 2^{1/48}}\times\frac{\sigma_I}{\overline{I}(n)\times 20\%}. \quad (8)$$

If the pitch variation is smaller than a quarter of semitone and the intensity variation is smaller than 20 percentage of the intensity average, the steadiness value should be smaller than one, indicating that the note is steady enough.

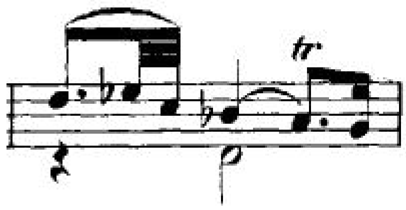

Fig. 2 Bach Sonata No.3 in C major, BWV 1005, Adagio, the 14-th measure.

Similarly, for a vibrato segment of a note, the vibrato rate $V(\cdot)$ and the vibrato depth $D(\cdot)$ are first estimated. Then, the vibrato rate mean $\overline{V}(n)$ and variance $\sigma_V^2(n)$ are consequently calculated. Similarly, we also have the vibrato depth mean $\overline{D}(n)$ and variance $\sigma_D^2(n)$. The steadiness measure of a vibrato segment is calculated as

$$s_v = \frac{1}{N}\sum_{n=1}^{N}\frac{\sigma_V}{\overline{V}(n)\times 20\%}\times\frac{\sigma_D}{\overline{D}(n)\times 20\%}. \quad (9)$$

If variations of the vibrato depth and the vibrato rate are smaller than 20 percentages of their averages, the steadiness value should be smaller than one, indicating that the vibrato note is steady enough.

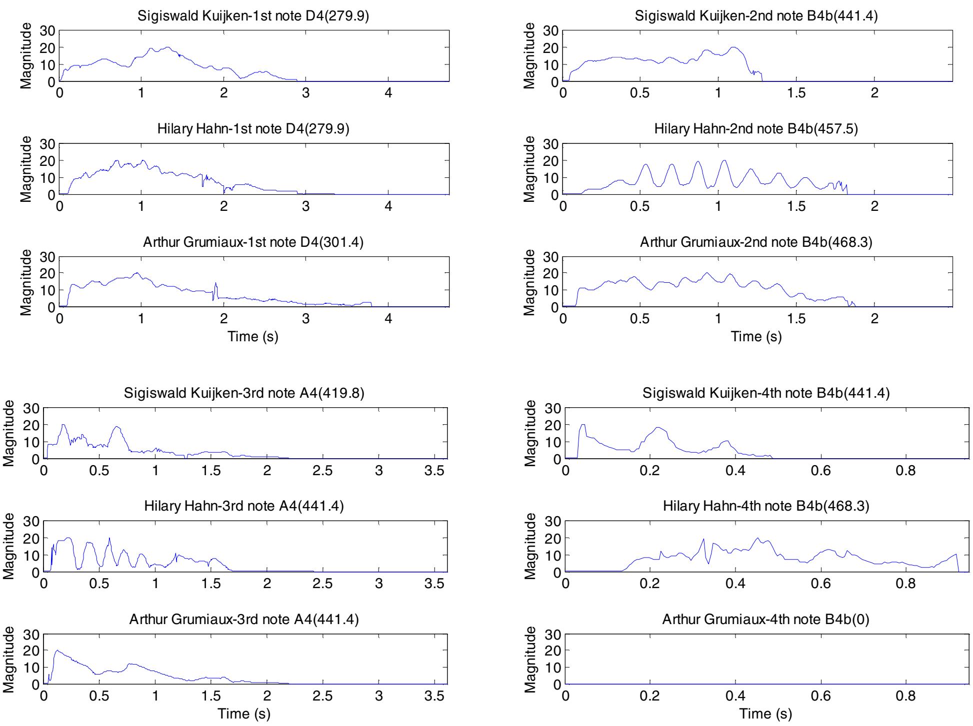

Fig. 3 Intensity matrices of 4 different notes played by three violinists.

IV. ANALYSIS RESULTS OF PERFORMANCES OF BACH'S VIOLIN SOLO WORK

In the experiments of this paper, we use commercial recordings of Bach Sonata No.3 in C major, BWV 1005, Adagio, the 14-th measure shown in Fig. 2, recorded by different famous violinists, Sigiswald Kuijkan [18], Hillary Hahn [19] and Arthur Grumiaux [20].

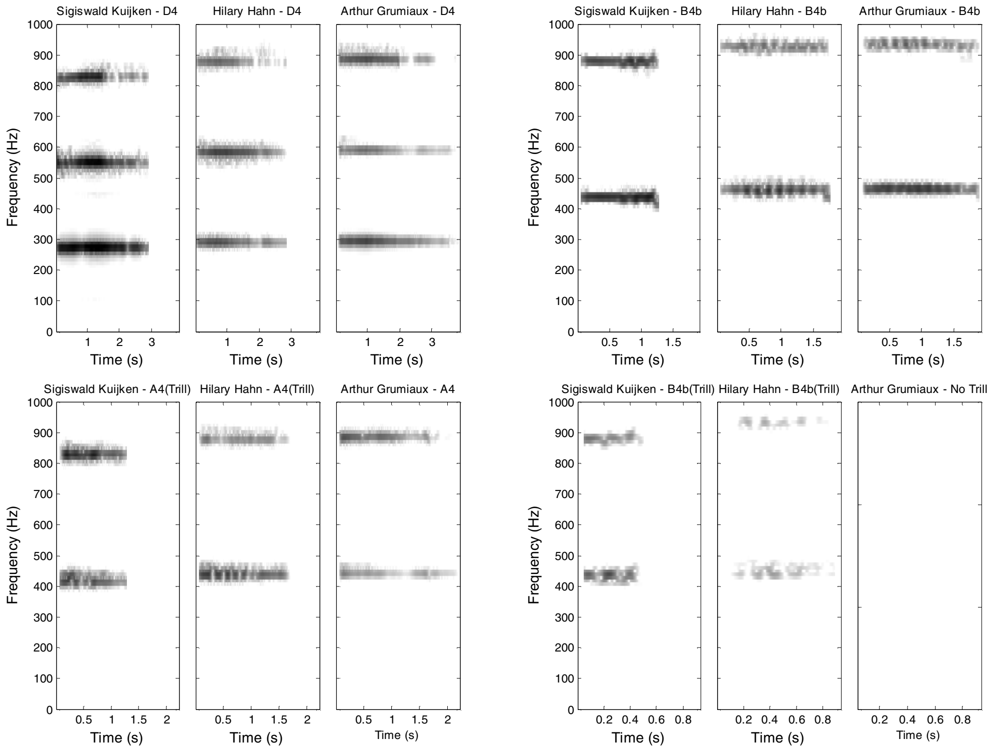

Fig. 4 Template matrices of 4 different notes played by three violinists.

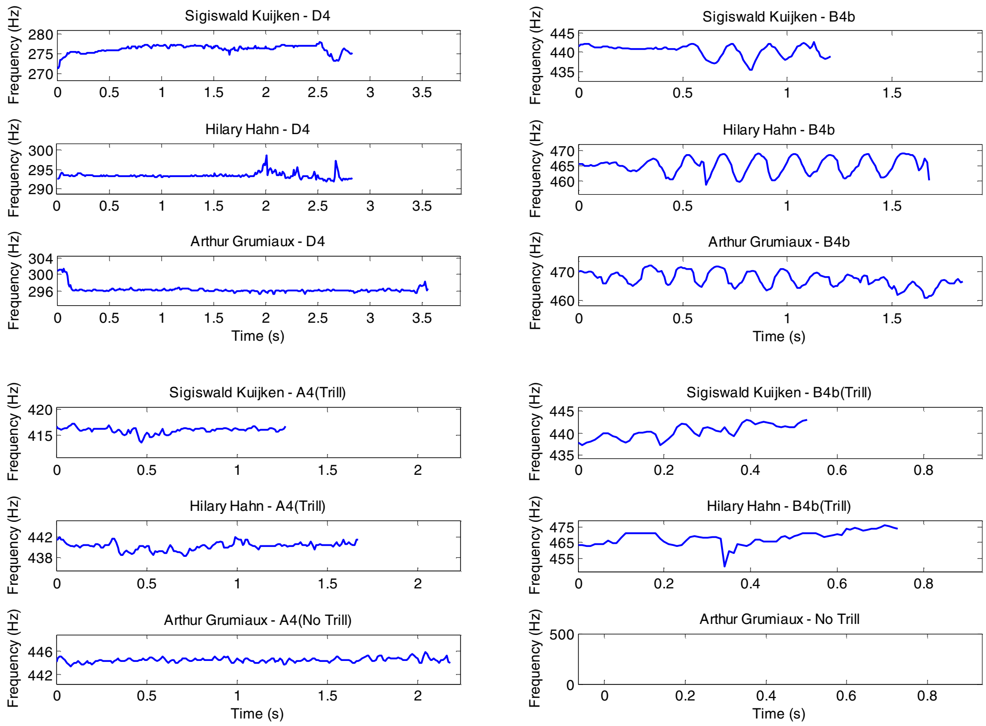

Fig. 5 Pitch contours of 4 different notes played by three violinists.

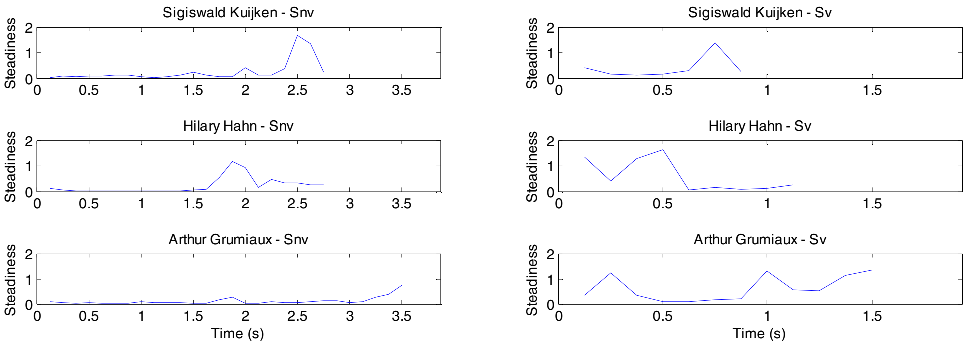

Fig. 6 Steadiness indices of the non-vibrato(D4) and vibrato(B4b) cases played by three violinists.

The respective template matrices and intensity matrices after the note separation processing can be obtained. We show only four notes, D4, B4b, A4 and another B4b. The first two notes are played at the same time though D4 note was played with an open string and started a little bit earlier. The last 2 notes are played at the same time using a technique called *trill*, as indicated in Fig. 3. One can compare the differences of their performances of the same piece of music. Fig. 3 shows the intensity matrices, Fig. 4 shows the template matrices and Fig. 5 shows the pitch contours, for all three players.

By observing the above figures, we found lots of differences. For D4 note, Kuijkan did not use vibrato, but the other two players used vibrato. For B4b note, all three players used vibrato, but Kuijkan's vibrato started at 0.5 second after B4b note was played. For A4 and B4b notes, because trill was used and the trill speed was high, it is difficult to use steadiness measure. However, we can see that Grumiaux didn't perform the trill and the B4b note was not played. Finally, Kuijkan played one semitone lower than the other two players.

Fig. 6 shows the steadiness indices of 0.25-second segments of two sample notes. There is a 50% overlap between two adjacent segments. Table 1 shows the average steadiness values of non-vibrato and vibrato notes played by three violinists. The clips are available at our web site so that one can listen and compare their performances.

TABLE I
AVERAGE STEADINESS VALUES OF THE NON-VIBRATO(D4) AND VIBRATO(B4B) CASES PLAYED BY THREE VIOLINISTS

Steadiness	Violinist Name		
	Sigiswald Kuigken	Hilary Hahn	Arthur Grumiaux
s_{nv} (D4)	0.302885	0.255376	0.258296
s_v (B4b)	0.529272	0.590483	0.605051

V. CONCLUSION AND FUTURE WORKS

In this paper, a note separation algorithm is presented and the quantitative evaluation approach of violin solo performances is also proposed. Commercial recordings of Bach's violin solo works played by various violinists are used to demonstrate our work. Though one may be able to tell the differences of the performances by observing the template matrix and the intensity matrix obtained in the note separation process, quantitative indices are still necessary for those who are not familiar with signal processing technologies. In this paper, steadiness indices are proposed to delineate the characteristics of violin solo performances in the more conclusive way such that it is easy for most people to understand and objectively describe performances. It is noted that the proposed work is not intended to touch the artistic part of performances which should still remain in the conventional domain of music performance assessment. In the future, we wish to improve our note separation algorithm and provide more types of quantitative indices. We also want to consult violin teachers to validate the proposed quantitative indices. Finally, the techniques used in this paper can be extended to the evaluation of performances of other musical instruments.

REFERENCES

[1] B. H. Repp, "Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's Träumerei," *Journal of the Acoustical Society of America*, 92, pp. 2546-2568, 1992.

[2] B. H. Repp, "Patterns of note onset asynchronies in expressive piano performance," *Journal of the Acoustic Society of America*, 1oo, pp. 3917-3932, 1996.

[3] C. S. Sapp, "Comparative analysis of multiple musical performances," in *Proc. of the 8th International Conference on Music Information Retrieval*, Vienna, Austria, pp. 497–500, 2007.

[4] C. S. Sapp, "Hybrid numeric/rank similarity metrics for musical performance analysis. " in *Proc. of 9th International Conference on Music Information Retrieval* (ISMIR 2008), Vienna, Austria, pp. 501–506, 2008.

[5] J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, "A tutorial on onset detection in musical signals," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 5, pp. 1035–1047, 2005.

[6] A. de Cheveigne and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music,” *Journal of the Acoustical Society of America*, vol. 111, pp. 1917–1930, 2002.
[7] N. Orio, S. Lemouton, and D. Schwarz, “Score Following: State of the Art and New Developments,” in *Proc. of the International Conference on New Interfaces for Musical Expression* (NIME), Montreal, May 2003.
[8] M. E. P. Davies and M. D. Plumbley, “Context-Dependent Beat Tracking of Musical Audio,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, 1009-1020, 2007.
[9] G. Peeters, “A large set of audio features for sound description (similarity and classification) in the CUIDADO project,” Cuidado project report, Institut de Recherche et de Coordination Acoustique Musique (IRCAM), 2004.
[10] Y.-S. Siao, W.-C. Chang, and Alvin W. Y. Su, “Pitch Detection/Tracking Strategy for Musical Recordings of Solo Bowed-String and Wind Instruments,” *Journal of Information Science and Engineering* 25, pp.1239-1253, 2009.
[11] D. D. Lee and H. S. Seung, “Learning the parts of objects by nonnegative matrix factorization,” *Nature*, vol. 401, pp. 788-791, 1999.
[12] B. Wang and M. D. Plumbley, “Musical audio stream separation by non-negative matrix factorization,” in *Proc. DMRN Summer Conference*, Glasgow, 2005.
[13] T. Virtanen, “Monaural sound source separation by nonnegative matrix factorization with temporal continuity,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1066–1074, March 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 Int. Conference on Digital Audio Effects,* Graz, Austria, 2010.
[15] Y.-L. Chen, T.-M. Wang, W.-H. Liao, and Alvin W.Y. Su, “Analysis and Trans-synthesis of Acoustic Bowed-String Instrument Recordings – A Case Study Using Bach Cello Suites,” in *Proc. of the 14th International Conference on Digital Audio Effects* (DAFx-11), Paris, France, Sep. 19-23, 2011
[16] T.-M. Wang, T.-C. Chen, Y.-L. Chen, and 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.
[17] Yiju Lin, W.-C. Chang, T.-M. Wang, Alvin W. Y. Su and W.-H. Liao, ” Timbre-Constrained Recursive Time-Varying Analysis for Musical Note Separation,” in *Proc. of the 16th International Conference on Digital Audio Effects* (DAFx-13), Maynooth, Ireland, Sep. 2-6, 2013
[18] Sigiswald Kuijkan, "Bach: Sonatas & Partitas, BWV 1001-1006," Deutsche Harmonia Mundi Records, CD1, 1990. Barcode: 0035627704321.
[19] Hillary Hahn, "Hilary Hahn Plays Bach," Sony Bmg Music Entertainment, 2005. Barcode: 074646279328.
[20] Arthur Grumiaux, "Bach: Sonatas & Partitas for solo violin," Decca Music Group Limited, CD2, 2006. Barcode: 028947575528.