2014 RISP International Workshop on Nonlinear Circuits,
Communications and Signal Processing
NCSP'14, Honolulu, Hawaii, USA. February 28 - March 3, 2014

Study on Analyzing Individuality of Instrument Sounds Using Non-negative Matrix Factorization

Keisuke Kobayashi, Daisuke Morikawa and Masato Akagi

School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292 Japan Phone:81-0761-51-1391
FAX:+81-0761-51-1149
E-mail: {k.kobayashi, morikawa, akagi}@jaist.ac.jp

Abstract

Nowadays, sound production softwares are poplularly used in computer music. However, to express individuality of instrument sounds, especially piano sounds, is still a challenging problem. Though individuality of a piano sound exists in its harmonic structures, the temporal variation of the harmonic structures has not been taken into consideration in the previous researches. This paper aims to study the individuality of piano sounds in their harmonic structures and temporal variation. Analyses are conducted using Non-negative Matrix Factorization (NMF). The results show that the temporal variation of high frequency components above 4 kHz is important to distinguish upright pianos and grand pianos. In addition, grand piano sounds are divided into 4 types by individuality, and individuality of upright piano sounds is observed by NMF analysis with fixing parts of basis matrix.

1. Introduction

Sound produnction softwares (e.g. MIDI sound source) are widely used in computer music. Due to this, it is easy for people who are not familiar with composing music to compose music by themselves. However, the softwares only provide typical musical instrument sounds. For this reason, people cannot use musical instrument sound with individuality with such softwares. In order to use musical instrument sounds with individuality instead of sound production softwares, huge single tone databases are required. Moreover, such databases are expensive and complicated to be constructed. Therefore, it is difficult to use instrument sounds with individuality in practice.

There have been researches investigating physical models of musical instruments considering sound production mechanisms. Nevertheless, in the case of pianos, no model is successful with pianos sound because it is difficult to express a desirable piano sound due to its complexity.

Yamaya *et al.* found that individuality of piano sounds is included in their harmonic structures and waveforms [1]. However, harmonic structures change their shapes in time [2]. Therefore, this work aims to study individuality of piano sounds by analyzing their harmonic structure and temporal variation correspondingly.

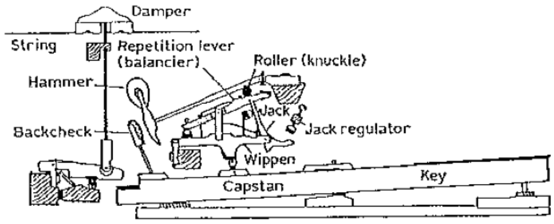


Figure 1: Key action model of piano [4]

2. Sound production mechanism of pianos

Sound production mechanism of a piano consists of four steps. Firstly, a player hit the key. Secondly, the hammer strikes the strings by a complex mechanism called “action”, as described in Fig. 1. After that, the vibrations of the strings occur. Finally, string vibrations propagate to the sound board through a bridge.

The strings in each piano tone are tuned differntly. In the middle tone, there are three strings: upper string, center string and lower string, in which the tuning of the upper string is a little bit higher than that of center string and the lower string is tuned a little bit lower than that of center. Therefore, this tuning causes inharmonicity when the hammer strikes the strings.

3. Analysis scheme

To analyze individuality of piano sounds, Non-negative Matrix Factorization (NMF) is used [3]. Before applying NMF to piano sounds, sound signals are pre-processed as described in Fig 2.

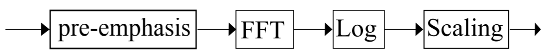


Figure 2: Scheme of pre-processing before applying NMF

3.1 Non-negative Matrix Factorization

NMF decomposes one matrix into two differnt matrices as follows:

$$Y_{\omega,t} \approx \hat{Y}_{\omega,t} = \sum_{k}^{K} U_{\omega,k} V_{k,t}. \quad (1)$$

where Y, U and V are respecticely $\Omega \times N$,$\Omega \times K$ and $K \times N$ non-negative matrices.

In the audio application, Y is corresponding to magnitude or power spectrogram. Ω is the number of frequency bins and N is the number of frames. U called basis matrix expresses spectra of K sources. V called activation matrix expresses variation of each vector in U.

3.2 Pre-emphasis

Preliminary experiment showed that the effects of components in high frequency area could not be observed without pre-emphasis because the effects of component in low frequency area were so dominant. To reduce this bias, pre-emphasis with a high pass filter (HPF) is performed before making spectrograms. The HPF is designed to cancel the spectral tilt of piano sounds. The average of spectral tilts given by regression line approximation is 6.6 dB/oct.

3.3 Log power spectrogram analysis with NMF

In this study, log power spectrograms are used as Y. There are two reasons to use them. The first is that log power spectrograms are better than linear power spectrograms in terms of timbre perception. The other is the sound generation mechanism of a piano.

As we describe in section 2, there are 4 steps in sound production mechanism. If each step is considered as a filter, this process can be regarded as a series of filters. NMF expresses the matrix as a sum of each component. Thus, if relation of each component is series, taking logarithm is necessary to express original matrix X as a sum of components A, B, C and D, as follows:

$$\begin{aligned} X &= \log(A \times B \times C \times D) \\ &= \log A + \log B + \log C + \log D. \end{aligned} \quad (2)$$

3.4 Scaling log power spectrogram

NMF can only deal with non-negative values. Therefore, transforming the range of log power spectrograms to non-negative scale is required. If zeros exist in the power spectrograms, negative infinity exists by taking logarithm. This prevents to transform all values to non-negative values. Therefore, zeros in power spectrograms are replaced to minimum value except 0 of power spectrograms before taking logarithm. After taking logarithm, all value in log power spectrograms are subtracted from the minimum value of log power spectrograms. This process enables to scaling log power spectrograms to non-negative value.

4. Experiment

4.1 Experiment Condition

In this experiment, five upright piano sounds, four grand piano sounds and one MIDI piano sound are used. One upright piano sound source is from MIDI Aligned Piano Sounds (MAPS) and two grand piano sounds are from RWC-DB (RWC1 and RWC3) [5, 6]. The other piano sounds, four upright piano sounds named as UP1 to UP4 and two grand piano sounds named as GP1 and GP2, were recorded in music practice room in ISHIKAWA ONGAKUDO. Though the positions of microphone are almost as those of MAPS and RWC-DB, the size of room which the upright pianos were in was smaller than that of MAPS. The MIDI piano sound was provided by Kawai Musical Instruments Manufacturing Co., Ltd.

The tone of each piano sound is A3 (220Hz). All sounds were recorded with 16-bit rate at a samping frequency of 44.1 kHz. Short Time Fourie Transform (STFT) was used to calculate sound spectrograms. The hanning window was used with a window length of 2048 points and a window shift length of 128 points.

4.2 Results of analysis using NMF

Some representative results of analysis piano sounds using NMF are shown from Fig. 3, 4 and 5. These results are corresponded to MIDI piano sounds, upright piano sounds and grand piano sounds respctively. Panels (a) and (b) in each figure are respectively the basis matrix and the activation matrix.

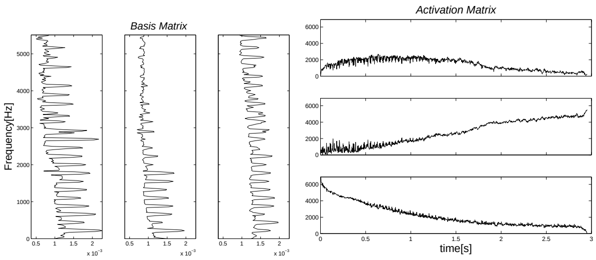
The activation vectors are divided into contour of stationary, attack and attenuation in Fig. 3, 4 and 5. Consequently, the basis vectors are named as stationary vector, attack vector and attenuation vector respectively. When the results of each basis vector are corresponded to the process of the piano sound generation mechanism, the stationary vector corresponds to sound board vibration and the attenuation vector corresponds to the string vibration.

When the correlation coefficients between these vectors are calculated, there is bias in stationary vector. The average of correlation coefficients between upright piano sounds or grand piano sounds are almost 0.78. However, the average of correlation coefficients between grand piano sounds and upright piano sounds is 0.62. The average of correation coefficients between MIDI and upright piano sounds is 0.67, while that of the MIDI piano sound and grand piano sounds is 0.73.

5. Discussion

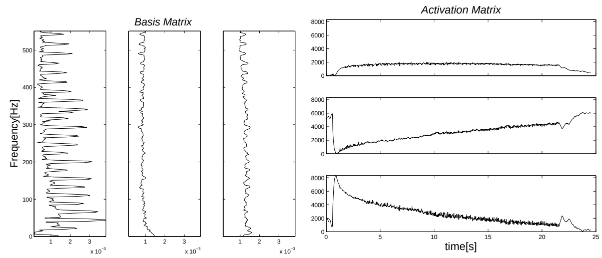
5.1 Difference among the three types of piano sounds

Correlation coefficients among same groups are higher than those among different groups. For this reason, stationary vector is important to distinguish grand piano sounds and



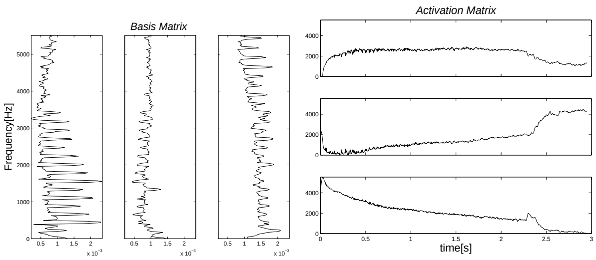
(a) (b)

Figure 3: Results of analyzing MIDI piano sound



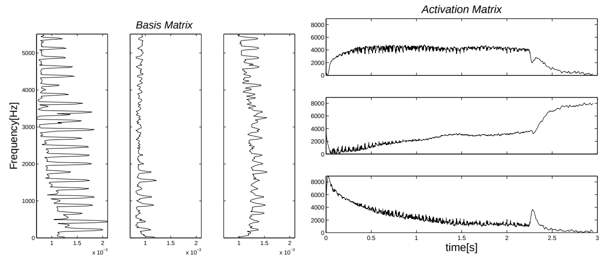
(a) (b)

Figure 5: Results of analyzing grand piano sound (RWC3)



(a) (b)

Figure 4: Results of analyzing upright piano sound (UP2)



(a) (b)

Figure 6: Result of analyzing GP2 using NMF

upright piano sounds. Furthermore, by calculation of coefficient between the MIDI piano sound and that of others, the MIDI piano sound is neither similar to grand piano sounds nor upright piano sounds. However the correlation coefficient between MIDI and grand piano sounds are higher than that between MIDI and upright piano sounds. As the MIDI piano sound considers grand piano sounds, this gap is occured. The shape of contour of the attenuation part in Fig. 3 looks like an exponential function, and is not observed fast decay caused by coupled oscillation such as the attenuation vector in Fig 5. For these results, MIDI piano sounds are differnt from real instrument sounds.

We futher analyze the characterisics stationary vectors because there is bias in their correlation coefficients. The differnce between the stationary vectors of upright piano sounds and grand piano sounds is the component above 4 kHz. There are some peaks above 4 kHz in stationary vector in grand piano sounds, while these peaks do not appear in upright piano sounds. On the other hand, they are observed in attenuation vector of upright piano sounds. Therefore, the temporal variation in high frequency area above 4 kHz causes the main difference between upright piano sounds and grand piano sounds. As a result, it should be the keypoint to distinguish them.

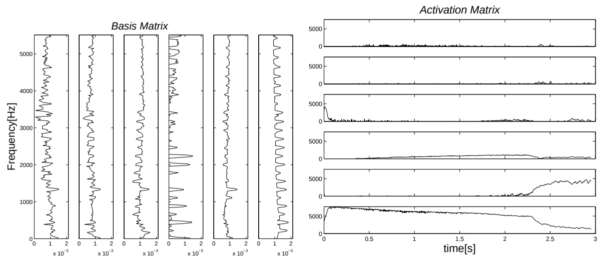
5.2 Individuality of grand piano sounds

Analyzing results of grand piano sounds are also shown in Fig 6. Since the correlation coefficient between the stationary vectors and the coefficient between the attenuation vectors are high, we focus on each vector. Based on the stationary vector, the piano sounds are divided into two groups. The first group includes the sounds (GP1 and RWC3) whose fundamental frequency(F_0) to the first harmonic ratio is lower than 0.5 dB (H_1/F_0) such as a vector in Fig. 5. On the other hand, in the other group which includes GP2 and RWC1, the ratio is higher than 0.5 dB such as shown in Fig. 6. The attenuation vectors are also divided into two groups. One of them including GP1 and GP2 has a attenuation vector with harmonic peaks (Fig 6). The other one which includes RWC1 and RWC3 has attenuation vectors with few harmonic peaks (Fig 5).

Therefore, grand piano sounds are divided into 4 types. This characteristics are related to individuality of grand piano sounds.

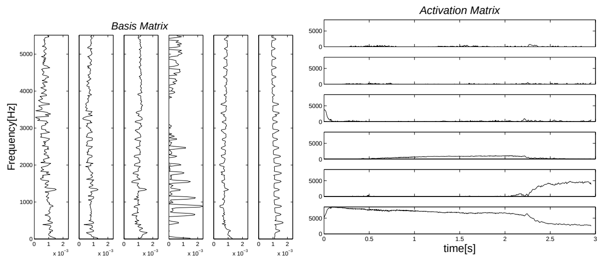
5.3 Individuality of upright piano sounds

Though upright piano sounds are divided into 2 groups by F0 to H1 power ration in the stationary vector, they are not divided by attenuation vector. Thus, to extract individuality from sounds, analysis using NMF with fixed basis matrix are applied to upright piano sounds. By fixing a part of basis matrix with common characteristics of upright pianos, the common components appear in activation matrix corresponding to fixed matrix, and individuality of upright piano sounds appear in both non-fixed basis and activation matrices. The first to third vectors of basis matrix are fixed and the other vectors which include all vectors in activation matrix are obtained by NMF. The fixed basis matrix is the minimum value in each



(a) (b)

Figure 7: Result of NMF with fixing basis matrix (UP1)



(a) (b)

Figure 8: Result of NMF with fixing basis matrix (UP3)

basis vector of upright piano sounds. Figure. 7 to 9 represent the results of this analysis. Panels (a) and (b) in each figure are respectively the basis matrix and the activation matrix.

Since the fourth to sixth vectors in each matrices are obtained from NMF iteration, they are not common parts of upright piano sounds. From the contours of activation vectors, the fourth vectors of Fig. 7 to 9 correspond to stationary part, the fifth vectors correspond noises and the sixth vectors correspond attenuation part.

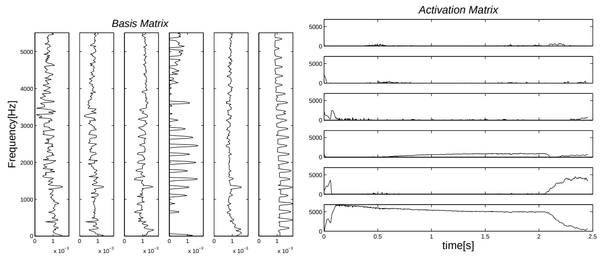
In the non fixed basis vectors, the contours of fourth vectors are different. In Fig. 7, the even peaks is higher than odd peaks. In Fig 8 , there is no peak nearby 3.2 kHz. However, in Fig. 9, there are some peaks nearby 3.2 kHz.

Moreover, the fourth activation vectors in figures slightly varies from 0.5 second to 2 or 2.2 seconds. For this results, they are considerd to correspond to the sound board.

6. Conclusion

In this study, individuality of piano sounds is analyzed using NMF. The difference between grand piano sounds and upright piano sounds appeared in temporal variations of high frequency components above 4 kHz.

There are four types of grand piano sounds distinguished by power ratio of F0 to H1 and harmonics of attenuation vector. Furthermore, the difference of harmonic structure from 0.5 kHz to 3 kHz are observed as individuality of upright piano sounds.



(a) (b)

Figure 9: Result of NMF with fixing basis matrix (MAPS)

Acknowldgement

This study was supported by the Grant-in-Aid for Scientific Research (A) (No. 25240026) and the Fostering ICT Global Leader Program.

References

[1] K.Yamaya, N.Ogawa, Y.Yamamoto and I.Tokuhiro, “Relation between Sound of Steinway&Sons or Bösendorfer Piano and Player’s Impression”, J.Acoust. Soc. Jpn,1,pp.823–824 2003.

[2] C.T.Lee, Y.H.Yang and H.H.Chen, ”Multipitich Estimation of Piano Music by Examplar-Based Sparse Representation”, IEEE Transactions on Multimedia, 14, 608–618, 2012.

[3] D.D.Lee and H.S.Seung, ”Learning the parts from object by non-negative matrix factorization”, Nature, 401, 788–791, 1999.

[4] I.Nakamura and H.Suzuki, “The sound generation mechanism and timbre of the piano”, J.Acoust. Soc. Jpn, 49, 3, 178–183, 1993.

[5] V. Emiya, R. Badeau and B. David,”Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle”, IEEE Transactions on Audio, Speech and Language Processing, 18, 1643–1654, 2010.

[6] M.Goto, H.Hashiguchi, T.Nishimura and R.Oka,”RWC Music Database: Music Genre Database and Musical Instrument Sound Database”, Proceedings of the 4th International Conference on Music Information Retrieval , 229-230, 2003.