Audio Engineering Society

Conference Paper

Presented at the Conference on
Semantic Audio
2017 June 22 – 24, Erlangen, Germany

Note Intensity Estimation of Piano Recordings by Score-informed NMF

Dasaem Jeong and Juhan Nam

Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology

Correspondence should be addressed to Juhan Nam (`juhannam@kaist.ac.kr`)

ABSTRACT

While dynamics is an important characteristic in music performance, it has been rarely researched in automatic music transcription. We propose a method to estimate individual note intensities from a piano recording given pre-aligned score data of the recording. To this end, we use non-negative matrix factorization in a score-informed framework, where the basis vectors and weights are constrained to estimate harmonic note spectra and corresponding intensities, respectively. We examine various choices in the learning process including the use of synthesized note scale for initialization, type of spectrum, and temporal constraint. We evaluate our method using Saarland Music Data (SMD) and estimate note intensities in MIDI velocity. The results show that the proposed method makes an improvement compared to previous work.

1 Introduction

A piece of music is played in different ways depending on the style and interpretation of the player. The differences can be computationally analyzed by extracting note information from the performance using an automatic music transcription (AMT) system and comparing them to the sheet music score. However, the majority of AMT systems handle only the presence of notes, that is, onset and duration in time, ignoring other performance information. Among others, dynamics is important along with the temporal information of notes in analyzing music performance [1]. In particular, transcribing it at the note level, or evaluating the individual note strength, allows to encode the performance as complete MIDI note messages as well as a detailed analysis such as comparing dynamics of melody lines and accompaniment parts [2].

Estimating individual note intensity in polyphonic piano recordings has been conducted by few work. Ewert and Müller challenged the task using a parametric model that represents audio spectra with note activity, spectral envelope, tuning and energy distribution over partials [3]. They searched the parameter values that explain the spectrogram most properly after aligning notes onset and duration. While they showed significant improvement compared to a simple energy picking method, the optimiztion process of the model is nontrivial. Szeto and Wong proposed a sinusoidal model to separate chords tones into individual piano tones and estimated the note intensity as part of the source separation task [4]. However, the evaluation was conducted only for simple chords rather than actual piano performance recordings. Also they used a training set that consists of isolated notes with exactly the same sound that the evaluation set has.

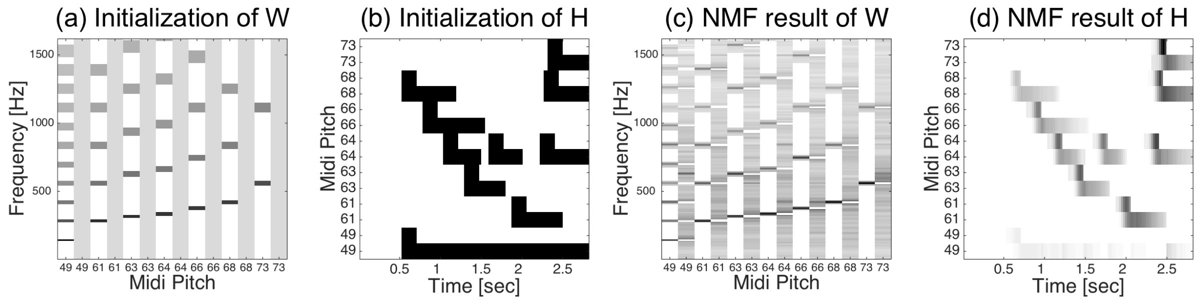

Fig. 1: Score-informed NMF with the harmonic constraint in the harmonic-percussive model (Bach BWV841-1).

In this paper, we present a method to automatically estimate note intensity from polyphonic piano recordings using Non-negative Matrix Factorization (NMF). By learning spectral template of each piano key from a training set or a piano synthesizer, we can factorize the spectrogram of the target recording by the basis vectors and their activations over time. Estimating the intensity of sound sources in a mixture using NMF was previously investigated in the context of general sound recognition [5]. We, however, estimate the intensity for polyphonic piano music where individual sources are more homogeneous and correlated to each other in time and frequency. Also we use a score-informed NMF framework where the symbolic representations of music are given as prior information [6]. We examine learning strategies in the NMF framework and other choices to improve the estimation accuracy including fractional power spectrum and temporal constraint in the framework. We evaluate our method using Saarland Music Data (SMD) used in [3] and estimate note intensities in MIDI velocity. The results show that the proposed method makes an improvement compared to previous work.

2 Methods

2.1 NMF Modeling of Polyphonic Piano Music

Our method is based on NMF, which decomposes a non-negative input $\mathbf{V}$ into two non-negative matrices $\mathbf{W}$ and $\mathbf{H}$ as $\mathbf{V} \approx \mathbf{WH}$. There are various algorithms to approximate the input $\mathbf{V}$ to $\hat{\mathbf{V}} = \mathbf{WH}$. We use the Kullback–Leibler (KL) divergence $D(\mathbf{V}||\hat{\mathbf{V}})$ and minimize it by the multiplicative update rule:

$$\mathbf{W} \leftarrow \mathbf{W} \otimes \frac{(\mathbf{V} \otimes \hat{\mathbf{V}}^{-1})\mathbf{H}^T}{\mathbf{1}\mathbf{H}^T} \tag{1}$$

$$\mathbf{H} \leftarrow \mathbf{H} \otimes \frac{\mathbf{W}^T(\mathbf{V} \otimes \hat{\mathbf{V}}^{-1})}{\mathbf{W}^T\mathbf{1}} \tag{2}$$

where $\mathbf{1}$ is an all-one matrix of the same size as $\mathbf{V}$, $\otimes$ means element-wise matrix product, and all divisions and inverts are also element-wise.

Applying NMF to audio spectrogram returns a set of spectral vectors $\mathbf{W}$ and their activations over time $\mathbf{H}$. Since this explains the spectrogram in a compositional manner based on the non-negativity and additivity, NMF has been widely used to model sound mixtures [7].

When NMF is applied to music signals, prior knowledge is often employed to obtain musically meaningful results. One of the implementations is to initialize each column vector in the $\mathbf{W}$ matrix such that only spectral bins classified as corresponding fundamental frequency or its harmonics contain non-zero entries [6]. This enforces harmonicity for the basis vectors and prevents them from learning spectral distributions of other notes because the zero entries of the matrix remains zeros during the multiplicative update of NMF. Specifically, we used the harmonic constraint proposed by [6], which sets non-zero entries with the bandwidth parameter $\phi = \pm\ 0.5$ semitone for each note.

Our system assumes that pre-aligned music score in MIDI is provided. Using this infomation, we add another constraint to the NMF. The MIDI data indiates which note is played at a certain time and how long it is sustained. We encode this to a piano-roll format and initialize temporal activation matrix $\mathbf{H}$ by setting the note bars to ones and the other entries of $\mathbf{H}$ to zero.

We adopt this score-informed NMF framework in two different settings. In the first setting, one column of the basis matrix $\mathbf{W}$ accounts for the spectral distribution

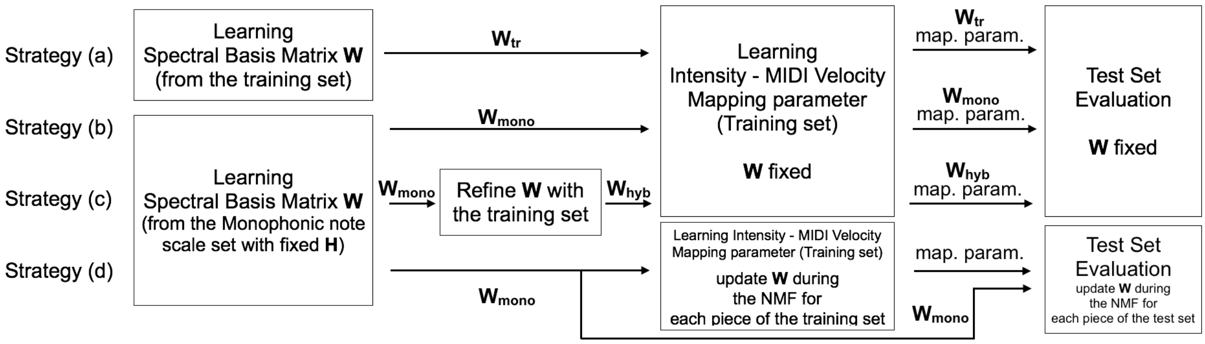

Fig. 2: Four strategies to learn basis matrix and estimate the activations

of a single piano note exclusively. Then, the corresponding row of $\mathbf{H}$ yields the temporal activities of the single piano notes. In this setting, we measure the note intensity as follows:

$$I(n) = \max_{t_n \le j \le t_n + r} \mathbf{H}_{p_n,j} \tag{3}$$

where p_n is the index of the basis vector for a note n, t_{n} is the onset frame of the note provided from the score and r denotes a search range parameter that can be adjusted. We call this setting *harmonic model*. In the second setting, we use two columns of $\mathbf{W}$ to explain a single piano note following [6]; one is for the harmonic distribution of a single piano tone and the other is for the remaining non-harmonic (or percussive) distribution. In this case, The harmonic and score-informed constaints are applied to the two types differently. The harmonic constraint is applied only on the harmonic basis vectors so that they effectively capture the harmonic distribution for a single note. The score-informed constraint is set to the whole duration of note for harmonic basis vectors whereas it is to a short attack time (the length is tunable) for percussive basis vector. This initialization scheme is illustrated in Figure 1 (a) and (b) and the corresponding NMF result is shown in Figure 1 (c) and (d). In order to estimate the intensity, we use activations from the harmonic basis vector because we observe that they are more proportional to MIDI velocity than those from the percussive basis vector. We term this second setting *harmonic-percussive model*.

2.2 Basis Matrix Learning Strategy

We can have multiple choices of learning the basis matrix and accordingly estimating the activations, depending on which stage (i.e. training or test) the basis matrix and activations are updated in and what datasets are used in each stage. Considering this, we design four different strategies to learn basis matrix as described in Figure 2. A notable part is that we use a monophnic note scale dataset apart from the polyphonic music dataset. Although we use harmonic constraint so that one or two basis vectors exclusively account for a single note, learning $\mathbf{W}$ directly from polyphonic music still does not guarantee clean harmonic distributions for each note because harmonically close notes are played toghether very often in tonal piano music. In order to improve the exclusiveness, we use a monophonic note scale set recored from a high-quality piano synthesizer. Since the target piano tone will be different from the synthesized one in terms of spectral characteristics, however, the basis matrix learned from the monophonic scale set are not fixed but rather used to initialize $\mathbf{W}$ before updated with the polyphonic music sets.

Figure 3 compares the results from each strategy in the harmonic-percussive model. We can see that spectral distributions of the training set and the monophonic note scale set are different from each other by comparing Figure 3(a) and (b), which means that the timbre and tuning of piano is quite different in the two data sets. But in Figure 3(b), the harmonic basis vector and the percussive basis vector are well separated whereas, in Figure 3(a), the two basis vectors are not perfectly separated in the first harmonic and fifth harmonic. Figure 3(c) shows the advantage of using both datasets: proper harmonic characteristics, and separation between harmonic and percussive basis vectors. Figure 3(d) also shows that the fourth strategy can achieve the similar advantage, though this strategy refines spectral basis using only a piece from the test set, rather than the whole training set.

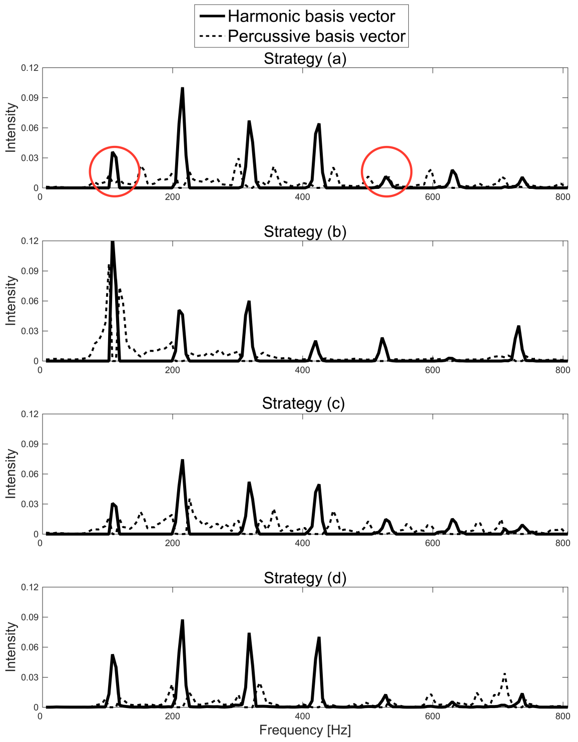

Fig. 3: Comparison of the basis vector in the different learning strategies (A♭2, MIDI pitch 44). The circles indicates the partials where the harmonic basis vector and percussive basis vector interfere with each other.

We observed that iterating the update rule too many times does not necessarily guarantees the best results in strategy (c) and (d) as found in [8]. The optimal number of iteration was experimentally searched. For example, when trained with the training set in strategy (c), we iterated the updating rule ten times. When the note intensity is estimated in strategy (d), we iterated it five times.

2.3 Fractional Power Spectrum

The principle of NMF assumes additivity based on non-negative basis vectors and weights. Physically, however, a mixture of sounds satisfies the additivity only when their phases are exactly identical. Otherwise, the amplitude of the sum will be smaller than the sum of the individual amplitudes. Since our task is very sensitive to the amplitude, this additivity is an important issue. We address this problem by using

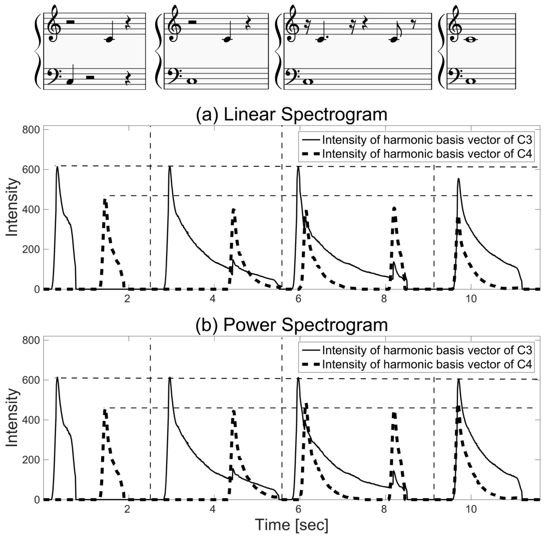

Fig. 4: Comparison between linear spectrogram and power spectrogram. The activation of each note in the power spectrogram case is constant regardless of whether the notes are overlapped.

fractional power spectrum. In [9], the fractional power spectrum is defined as α-spectrogram, p^{α} where p is the magnitude spectrum and α is a continuous value. We attempted to use a set of α (1.0, 1.3, 1.5, 1.8, 2.0). We observed that the intensity estimation worked best when α is 2, that is, with power spectrogram. This seemed opposite to the result in the context of source separation [9] where $\alpha = 1.0$ is optimal. This might be because our task, which estimates note intensities from the peak activations, is different from the source separation tasks, which is mainly concerned with reducing average divergence between the input signal and its model.

Figure 4 shows the effect of different fractional power spectrum for simple examples where two notes in octave unison (C3 and C4) are played individually or simultaneously with constant MIDI velocities. Ideally, one note should maintain the same level of intensity regardless of the existence of the other note. However, when linear spectrum is used, the peak activations of the C4 note decrease if it is played while the C3 note is sustained or both are played at the same time. On the other hand, when power spectrum is used, the peak activations of the C4 note remains in the same level.

2.4 Partial Temporal Constraint

Yet another option to the NMF framework is temporal regularization. Previous research showed that adding temporal continuity by using parameters obtained from sources [5] or by reducing the the difference between adjacent entry [10] in **H** matrix can improve the separation quality. Since piano notes gradually decrease in the sustained part and the activations will follow the pattern, we also apply the temporal continuity constraint. Referring to [10], we use the following cost function c_t:

$$c_t(\mathbf{H}) = \sum_{i=1}^{I} \frac{1}{\sigma^2} \sum_{t=2}^{T} (h_{i,t} - h_{i,t-1})^2 \tag{4}$$

where $\sigma = \sqrt{(1/T)\sum_{t=1}^{T} h_{i,t}{}^2}$. We incorporate this cost by adding a gradient of the cost with a scaling factor to the gradient of KL divergence of **V** and **WH**.

Employing this constraint to the whole **H**, however, restrains the activation from rising at the attack part of note events since it smooths the sudden increase of basis activation. Since we pick the peak activation at the attack part to identify the intensity of the note, this can be critical to our task. Therefore, we modify the temporal cost function c_t so that the function is applied only when notes are sustained. This is possible because we already have pre-aligned score information. In the harmonic-percussive model, the constraint is applied to only the activation of harmonic basis vectors.

Figure 5 shows the effect of the partial temporal constraint. As we can see from the difference between (a) and (b), the temporal constraint makes the activation of harmonic basis vector of each note smooth as indicated in the sustaining part of notes marked with the red circles. Also, the attack parts of notes are not smoothed as we intended. In Figure 5 (c), we can see that spurious peaks near note offset are suppressed by the temporal constraint. The spurious peaks typically occur when two or more notes share many harmonics, for example, notes in octave and one note ends earlier than others in score. In this case, since piano notes are usually sustained even after the offset time in score, especially when pedals are pressed, the basis vector for one note can explain the spectrum of other notes and this results in the sudden jump around note offsets. In Figure 5 (c), the offset of note A♭1 (MIDI 32) is one frame earlier than A♭2 (MIDI 44). Thus, without the temporal constraint, the harmonic basis vector of A♭2 note explains sustaining part of both A♭1 and A♭2 notes. The temporal constraint reduces this type of error and, in turn, helps to learn better basis vectors.

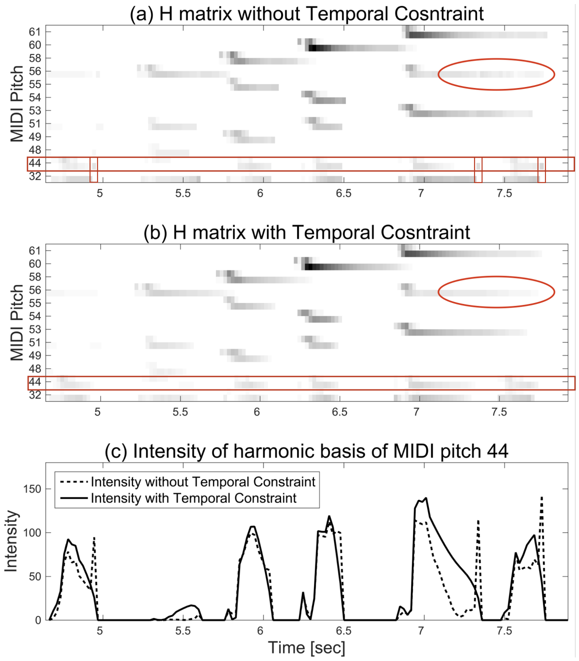

Fig. 5: Effect of partial temporal constraint (Beethoven Op. 27-1, 2nd mov.). The circles show the smoothing effect of the constraint in the sustaining part of the note. The activations marked with red bars are re-displyed in the bottom.

3 Experiments

3.1 Datasets

To evaluate our result, we employed Saarland Music Data (SMD) [11]. SMD consists of piano recordings in both audio and MIDI recorded by Disklavier. Disklavier records every mechanical movement of piano keys and pedal during the performance. Therefore, we can obtain ground truth of intensity of each note in the form of MIDI.

Since the ground truth for note intensity is in MIDI velocity unit and the result of estimation is from the activation **H**, we need to have a scale conversion so that two data types are comparable. The relation between MIDI velocity and the intensity was investigated in

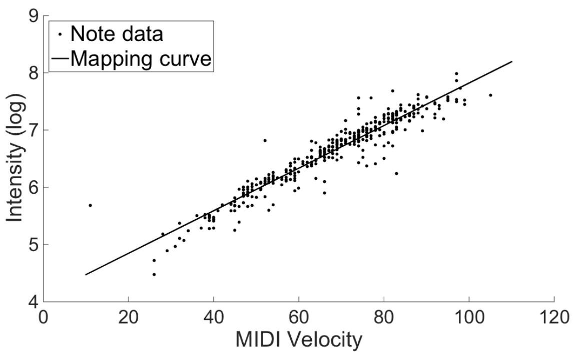

Fig. 6: Relation between MIDI velocity and note intensity (E♭ 5 note)

[12]. The research showed that the relation between velocity and intensity is different depending on synthesizer models. Also, the relation is affected by the recording environment. Thus, a training set to learn the mapping between note velocity and intensity should be recorded in the same environment with the same piano. Considering this issue, we grouped the dataset into three subsets according to the recording date [1]. For each subset, we have conducted three-fold cross validation. While one fold is used as a test set, the other two sets are used to learn the basis vectors and the mapping parameters between ground truth MIDI velocity and estimated note intensity. Note that although we have different experimental settings from those in [3], the sum of the subset covers the whole test set used in [3].

In addition to SMD, we used a monophonic note scale data for basis vector learning. The scale covers 88 keys with 12 different MIDI velocity from 10 to 120 for each key. The data was synthesized with Synthogy Ivory II Yamaha C7 samples in 440Hz-stretched tuning.

3.2 Evaluation Methods

We first learn the basis matrix by one of the four strategies. Then, we performed the note intensity estimation for the training set and obtained the mapping parameters for each pitch by comparing estimated note intensities and the ground truth MIDI velocity values as shown in Figure 6. Then we performed the note intensity estimation for the test set and convert them to MIDI velocities using the intensity-to-MIDI velocity mapping. We kept the same NMF settings in both training and test stages.

[1] 6 Bach recordings in 20090916, 13 Chopin recordings in 20100611 and the 15 other recordings in 20090916 where the eight digit numbers mean a recording date in form of yyyymmdd (year-month-day).

We used two types of metrics to measure the accuracy. First, we calculated absolute difference in MIDI velocity between ground truth value and mapped note intensities

$$AE := |V_{est}(n) - V_{\mathrm{GT}}(n)| \tag{5}$$

where $V_{est}(n)$ is estimated MIDI velocity of nth note.

The other metric was mapping the ground truth MIDI velocity back to the note intensity using the same mapping parameters and then calculating relative error of intensity as follows:

$$RE := 100(\%) \cdot \left| \frac{I^{0.3}(n) - I^{0.3}_{\text{mapped GT}}(n)}{I^{0.3}_{\text{mapped GT}}(n)} \right| \tag{6}$$

We used this metric to make our result comparable to [3]. They scaled the note intensity, considering psychoacoustic perception of loudness. We followed the same metric: spectral energy sum of power spectrogram to the power of 0.3, or linear spectrogram to the power of 0.6.

4 Results and Discussion

Figure 7 shows the absolute error of four different strategies for learning basis vectors. In strategy (a), the harmonic-percussive model made a large error than harmonic model if there was no harmonic constraint. But when the harmonic constraint was applied, the error decreased drastically. This means that it is not easy to learn proper basis vectors in the harmonic-percussive model if there is no harmonic constraint. Applying harmonic constraint to the harmonic model made a negative effect, because this makes the basis matrix ignore the inharmonicity of the sound. Strategy (b) made the largest error except for when using the harmonic-percussive model without harmonic constraint. We assume that the reason for high error is due to the difference in spectral characteristic of the monophonic note scale and the test set, for example, different tuning or timbre.

The best performance was made by strategy (c), which learns the basis vectors from the monophonic note scale and refines it with the training set. We can see that the

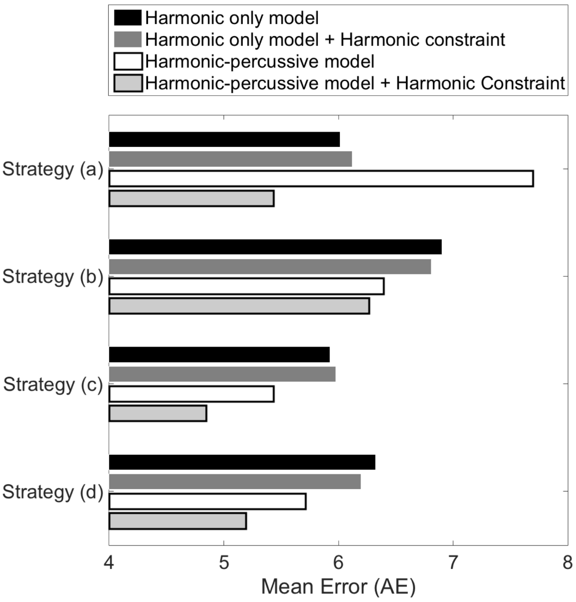

Fig. 7: Result of note intensity estimation according to the four basis vectors learning strategies and the NMF modeling.

difference between strategy (a) and (c) using harmonic model is slight. This means that the advantage of employing the monophonic note scale is only limited to the harmonic-percussive model. Strategy (d) showed a less precise performance than strategy (c), but better than the others. This shows that strategy (d) can be an alternative solution if there is no available training set to update spectral basis vectors.

Figure 8 shows the effect of harmonic constraint, power spectrogram and partial temporal constraint for the two best NMF settings, strategy (c) and (d). This indicates

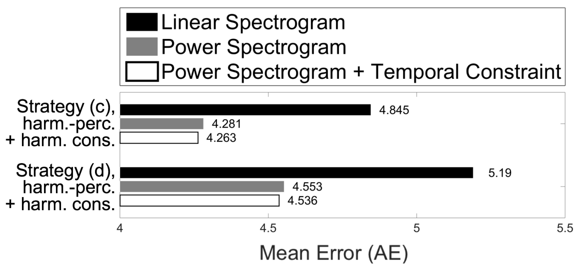

Fig. 8: Result of note intensity estimation when applying the power spectrogram and temporal constraint.

Composer	Piece	Proposed				Ewert	
		Mean Abs	STD Abs	Mean Rel%	STD Rel%	Mean (%)	STD (%)
Bach	BWV849-01*	2.6	3.3	6.3	8.2	9.3	5.3
Bach	BWV849-02	2.1	2.7	5.5	7.5	9.3	5.5
Bach	BWV871-01	1.5	2.1	3.9	6.9	11	6.2
Bach	BWV871-02	1.9	2.6	4.4	5.1	7.7	5.1
Bach	BWV875-01	1.9	2.5	4.3	4.7	13.9	6.7
Bach	BWV875-02	1.8	2.5	4.6	8.7	8.3	4.9
Beethoven	Op027No1-01	4.4	4.9	11.0	17.1	12.5	7.1
Beethoven	Op027No1-02*	4.3	4.8	11.0	16.2	10.3	6.5
Beethoven	Op027No1-03	4.0	6.2	12.3	41.6	13.6	7.3
Beethoven	Op031No2-01	4.5	5.5	12.0	16.6	16.1	8.7
Beethoven	Op031No2-02	4.3	4.7	10.3	12.1	27.2	14.5
Beethoven	Op031No2-03	2.6	3.3	7.2	13.3	13.2	8.1
Brahms	Op010No1*	5.6	6.7	13.8	22.9	13.8	7.3
Brahms	Op010No2	5.5	6.7	12.8	23.0	13.6	7.9
Chopin	Op010-03	5.0	5.3	14.6	20.5	25.2	13
Chopin	Op010-04	3.7	4.6	11.8	30.5	25	13.2
Chopin	Op026No1	4.1	5.4	14.6	39.1	22.6	13.2
Chopin	Op026No2	7.4	7.6	23.6	38.0	23.6	14.2
Chopin	Op028-01	3.7	4.2	11.7	19.0	22.9	11.4
Chopin	Op028-03	3.1	2.9	9.5	11.5	19	12.2
Chopin	Op028-04	4.7	4.0	15.3	14.4	19.5	11.6
Chopin	Op028-11	3.7	3.9	10.8	12.3	18.8	9.1
Chopin	Op028-15	5.0	4.6	13.9	17.6	18	9.2
Chopin	Op028-17	6.6	6.9	24.5	47.4	22.1	10.7
Chopin	Op029*	4.8	5.1	14.4	19.1	20.1	11.6
Chopin	Op048No1	6.4	6.9	20.7	37.3	26	11.2
Chopin	Op066	4.5	4.4	13.3	15.7	22.4	13.5
Haydn	Hob017No4*	2.9	4.4	8.5	24.8	14.8	8.1
Rachman.	Op039No1	4.9	6.8	14.2	30.1	15.5	9
Skryabin	Op008No8	3.7	4.1	9.9	12.7	10.1	5.6
Liszt	LectureDante	7.4	9.5	23.0	50.7	-	-
Liszt	S. 179	7.3	9.6	24.0	57.8	-	-
Liszt	S. 144-2	4.4	7.3	15.8	50.4	-	-
Ravel	Valse Nobles	5.1	7.0	15.5	39.0	-	-
Average		4.3	5.1	12.6	23.3	16.8	9.3

Table 1: Result of note estimation of individual pieces. The stars indicate pieces that have been used to refine Ewert's note intensity dictionary.

that using power spectrogram reduces the error and applying partial temporal constraint had a slight but positive effect.

Table 1 compares our results to those in [3] using the relative error of intensity. Overall, our proposed method shows higher accuracy but lower precision than those in [3]. The main reason for low precision is due to the neglect of pedal effect. We did not consider the pedal usage in the recording for initializing **H**. Therefore, the piece with constant pedal usage like some of Chopin's works made a large error, while the piece with less pedal usage such as Bach's pieces made a less error. Furthermore, sometimes soft notes were screened by the preceding notes because of the pedal effect. There are some notes with extremely low velocity, e.g. less than 5 in MIDI velocity, which can make more than 1000 % relative error when combined with the pedal effect. For example, in Chopin's Op. 28-17 of SMD, there are eleven note events that have MIDI velocity lower than 5, among 2072 entire note events. If we exclude these notes, the standard deviation decreases from 47.4 to 34.4.

We should note that since our mapping parameters solely depend on the result of the same note intensity estimation procedure on the training set, there are some possibilities of making a bias. On the other hand, [3] made a note intensity dictionary for mapping the ground truth MIDI velocity to note intensity. The dictionary was made by employing a training set that contains single note events in several velocity values and refine it with the five pieces in the test set, which are marked in Table 1. Since SMD is published by one of the authors, it is probable that they could reproduce single note events in the same recording condition with SMD. Therefore, we needs to be careful when comparing two results directly. The definition of ground truth and standard methods for the evaluation are necessary to tackle this task further, which was actually mentioned in [3] as well. Also, we should admit that our system is limited because pre-aligned MIDI data must be provided. On the other hand, Ewert's parameteric model is more robust to misalignment between score and audio.

5 Summary

In this paper, we have presented a novel method for estimating note intensities from piano recordings. It is based on a NMF framework for learning basis matrix from a training set and estimating note intensities in performance audio. We have examined various methods such as the harmonic-percussive modeling, the harmonic constraint, employing a piano synthesizer for learning basis matrix, the use of power spectrogram and the partial temporal constraint. The evaluation results show an improvement compared to the previous research. In the future, we plan to make our system allow MIDI score that are not aligned to audio. Also we will update the NMF model considering the pedal effect.

References

[1] Widmer, G., Dixon, S., Goebl, W., Pampalk, E., and Tobudic, A., "In search of the Horowitz factor," *AI Magazine*, 24(3), p. 111, 2003.

[2] Repp, B. H., "The dynamics of expressive piano performance: Schumann's "Träumerei"revisited," *The Journal of the Acoustical Society of America*, 100(1), pp. 641–650, 1996.

[3] Ewert, S. and Müller, M., "Estimating note intensities in music recordings," in *Proc. of 2011 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 385–388, 2011.

[4] Szeto, W. and Wong, K., "Source separation and analysis of piano music signals using instrument-specific sinusoidal model," in *Proc. 16th Int. Conf. Digital Audio Effects*, 2013.

[5] Nam, J., Mysore, G. J., and Smaragdis, P., "Sound Recognition in Mixtures." in *Proc. of 10th Int. Conf. on Latent Variable Analysis and Signal Separation*, pp. 405–413, 2012.

[6] Ewert, S. and Müller, M., "Using score-informed constraints for NMF-based source separation," in *Proc. of 2012 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 129–132, 2012.

[7] Virtanen, T., Gemmeke, J., Raj, B., and Smaragdis, P., "Compositional Models for Audio Processing: Uncovering the structure of sound mixtures," *IEEE Signal Processing Magazine*, 32, pp. 125 – 144, 2015.

[8] Germain, F. G. and Mysore, G. J., "Stopping criteria for non-negative matrix factorization based supervised and semi-supervised source separation," *IEEE Signal Processing Letters*, 21(10), pp. 1284–1288, 2014.

[9] Liutkus, A. and Badeau, R., "Generalized Wiener filtering with fractional power spectrograms," in *Proc. of 2015 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 266–270, IEEE, 2015.

[10] Virtanen, T., "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3), pp. 1066–1074, 2007.

[11] Müller, M., Konz, V., Bogler, W., and Arifi-Müller, V., "Saarland Music Data (SMD)," in *Late-Breaking and Demo Session of the 12th Int. Conf. on Music Information Retrieval (ISMIR)*, 2011.

[12] Dannenberg, R. B., "The Interpretation of MIDI Velocity," in *Proc. of Int. Computer Music Conf.(ICMC)*, pp. 193–196, 1996.