Automatic evaluation for a scale performance by the MIDI-Piano

Heng-Shao Chen, Yu-Chen Chang, and Wen-Whei Chang

Institute of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan
email: henrychen0702@gmail.com, kieran1230@gmail.com, wwchang@cc.nctu.edu.tw

***Abstract*[+]—Many schools today require students to practice basic piano performance involving from scale fingering to finger legato at the beginning stage. This paper focuses on the proficiency estimation for a scale performance within one octave by the MIDI-piano. Given a piano performance, goal of our system is to assess the performer's ability in terms of scale fingering and assign a rating score that emulates the teacher's subjective preference. The system starts by extracting basic features concerning onset time, tempo, velocity, duration and articulation. Based on their deviations from constant standard, we apply the spline curve fitting technique to extract the parameters representative of the performance tendency. With respect to the automatic evaluation, machine learning algorithms based on decision tree are employed to estimate proficiency scores of testing data. Experiments on 11 subjects demonstrate the validity of the proposed system for automatic evaluation of scale performances.**

***Keywords*—scale performance, MIDI, decision tree**

I. Introduction

Piano is a popular recreational pastime in East Asia. On learning playing the piano, there are various kinds of stages, having various levels of difficulties to master. At the beginning stage, students may learn some skills, such as basic movements of fingers, tempo, articulation and rhythm. Without proper instruction by experts, students will not learn much and will struggle with long-term training. However, experts are not always available and modern business life makes hard for people to make time for scheduled musical lessons. This implies that some kind of supporting systems would be well-suited for self-learners at the beginning stage. An ideal system would have to be able to track progress of its users and also adapt its courses depending on one's level of skill. One of the first systems to this application has been the Piano Tutor [1]; however, the goal of the project was not performance evaluation but to devise an expert system that embodies knowledge about teaching the piano. The system combines multimedia and score following technology to synchronize an accompaniment with performance and to turn pages automatically. Also, it analyzes the student's performance, and based on which it coordinates the presentation of necessary lessons to strengthen his playing skill. pianoFORTE [2] is an interactive system that can capture user's piano playing over MIDI interface. It tracks several basic aspects of playing, including tempo, articulation and synchronization of music events. After playing, system gives feedback in graphical form on what was played correctly and where mistakes were made. Another recently released mobile application for music education is SmartMusic [3]. The system gives users the feedback about wrongly played notes, but does not make use of score following to show which part they are playing.

Even though prior works employ various data to assess the user's ability in terms of technical accuracy, they cannot evaluate appropriately the proficiency of piano performances. This is because that estimation of the proficiency requires several aspects of a given performance; however, the appropriate aspects of the performance have not been clear. The performance task used in this study was the playing of a scale within one octave on the MIDI piano. Scales are a staple of pianists' practice regimen, often played over and over for extended periods of time at various speeds. On playing the scale at a beginning step, students often play with listening to click sounds generated by a metronome. In such cases students are instructed that deviations from metronome are not preferable and to play exactly synchronized to the metronome is the objective of the practice. In addition, they are instructed that volume level and duration of each note should be common among all notes. However, such ideal performances may lead to mechanical sounding music. In fact, Kitamura [4] has found that it is not necessary for piano players to play mechanically on realizing a good performance from a musical viewpoint. Experiments on various kinds of scale performances demonstrated that recorded data by expert pianists shows some deviations from an ideal performance as well as those by novice players. Even though a diversity of scale performances among players is shown, there are still some common features representing the global tendency of the player's skill. It is because the players need to make a strategy how to play it based on his experience in a certain degree. In this study, actual performance was modeled as the sum of a tendency curve and deviation from it, followed by spline interpolation. Also proposed is an evaluation

[+]The work was partially supported by the National Science Council of Taiwan, Republic of China under research contract MOST105-2221-E-009-068

method based on onset time, velocity, duration, tempo and articulation for the MIDI data, in which several parameters used to evaluate the proficiency of a given performance were introduced.

II. MIDI DATA OF SCALE PERFORMANCES

Piano students are often instructed to play a scale over and over for extended periods of time at various speeds. Features of an ideal performance are thought to be as follows: timing is exactly synchronized to metronome, volume and duration for each note should be common among all notes. However, Kitamura [4] has found that it is not necessary for piano players to play mechanically on realizing a good performance from a musical viewpoint. Since the players tend to perform with different degree of skills, the deviations from the musical standard is expected to be useful for performance evaluation. The performance task used in this study was the playing of a scale within one octave in both directions, upward and downward with the right or left hand. Subjects were asked to play in non-legato style and the tempo was standardized at 120 beats per minute for a quarter note. To illustrate this, Fig. 1 shows the musical score for a one octave scale according to the regular C major key and the number below each note represents fingering on the right hand. Moreover, fingerings corresponding to five different keys for left and right hands are shown in Table 1. In Fig. 1, a turning point indicates where the keying changes the direction of the motion, and a crossing point means wrist moving according to the fingerings listed below. In this study, a group of 11 elementary school students participated in the test. They are asked to play a one octave scale in five different keys using left or right hand. There are 10 recorded performances for each player and at the end, three expert pianists are asked to evaluate the proficiency of each performance on a scale from 1 to 5. All the performances are recorded by MIDI piano YAMAHA P105 and a PC-based MIDI sequencer.

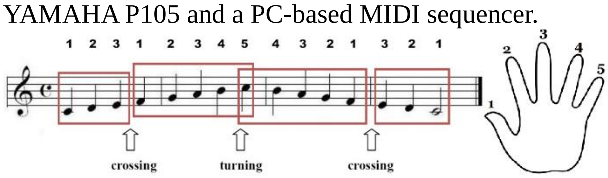

Figure 1.Regular C major key and the number below each note represents fingering on the right hand.

TABLE 1. Fingerings corresponding to five different keys for left and right hands (↑:crossing, ↓:turning).

	Major key	Fingering
Right hand	C、D、G、A	1 2 3 ↑ 1 2 3 4 5↓ 4 3 2 1 ↑ 3 2
Right hand	F	1 2 3 4 ↑ 1 2 3 4↓ 3 2 1 ↑ 4 3 2 1
Left hand	A、C、D、G、F	5 4 3 2 1 ↑ 3 2 1↓ 2 3 ↑ 1 2 3 4 5

Musical Instrument Digital Interface (MIDI) is a standard digital protocol for the communication of music events among a wide variety of electronic instruments, computers and other related devices. In recent years, the low-priced MIDI piano has lead to the development of many supporting systems in music education. MIDI technology facilitates a faithful recording of a performance on an electric piano whose keys are weighted to provide the same physical sensations as an acoustic piano. One of the most virtues of MIDI files is that they store music data as event messages instead of recorded sound. A MIDI file always starts with a header chunk, and is followed by one or more track chunks. The header chunk consists of a literal string denoting the header, a length indicator, the format of the MIDI file, the number of tracks, and a timing value specifying delta time units. A track chunk consists of a literal identifier string, the size of the track, and actual event data making up the track. When a scale performance is recorded using MIDI piano and a MIDI sequencer, key onset time, velocity and duration for each note are obtained automatically. Based on these basic features, we also compute the tempo and key overlap time (KOT). The KOT is often used to measure the degree of legato articulation and is defined as the time interval between the onset of key depression for one note and the key release for the preceding note. For the $\mathrm{j-th}$ note in a scale, let v_j , d_j , o_j , g_j and t_j represent the MIDI-velocity, duration, onset time, KOT and tempo, respectively. To simplify the notation, MIDI data of scale performance will be represented as $\{x_j, 1 \leq \mathrm{j} \leq \mathrm{J}\}$, where $x \in \{\mathrm{v,d,o,g,t}\}$ and J=15 is the number of played notes on a scale performance. Note that for an ideal performance, students are often instructed to play exactly synchronized to the metronome and volume and duration of each note should not have any deviations. In such cases, each feature of MIDI data will assume a constant standard, such as 64 for MIDI-velocity, 0.5 for duration, 0 for KOT, and 120 for tempo. In case of onset time, we chose to use $0.5 * (\mathrm{j}-1)$ as a simple standard for the $\mathrm{j-th}$ note. After obtaining these basic features, we compute the difference of MIDI data from constant standard, labeled as $\{x'_j, 1 \leq \mathrm{j} \leq \mathrm{J}\}$. The sequence of x'_j are expected to be useful for performance evaluation, because the degree of deviation is mainly affected by the player's skill level.

III. FEATURE EXTRACTION

Prior works employ various MIDI data to assess the player's ability in terms of technical accuracy, but they cannot evaluate appropriately the proficiency of piano performances. Moreover, on using the MIDI piano, we only have limited information for the piano performance, such as onset time, velocity and duration. In this study, the parameters to be applied for evaluation are those expressing the tendencies of a given performance. After obtaining the recorded performance, MIDI data x_j are subtracted by the constant standard to obtain the deviation x_j'. For the sequence of x_j', we employ the spline interpolation on it to obtain a tendency curve for a given performance. In order to apply spline curve fitting, a prerequisite is to determine the representative points to be passed by the spline curve. By observing two crossing points and one turning point of the scale fingering, we chose to use 6 representative points by dividing a sequence of 15 notes into four clusters, each comprised of several notes. Then, the center in each cluster is regarded as the point representative of each cluster. The basic strategy behind this is to propose a method that exploits the movements of fingers, which is also an important skill to master.

To illustrate this, we consider an example of scale fingering with C major key. As shown Fig. 1, one crossing is between note 3 and note 4, one crossing between note 12 and note 13, and one turning on note 8. Therefore, we divide 15 notes into 4 clusters and then select 6 representative points, labeled as $\{m_i, 1 \leq \mathrm{i} \leq 6\}$. Among them, $m_1 = x_1'$, $m_6 = x_{15}'$, and the other four points are computed by taking the average of x_j' in each cluster. It is important to note that both the crossing and turning points differ according to the hand and key used to play the scale performance, as listed in Table 1. With these 6 representative points to pass, we apply the spline curve fitting technique to obtain a tendency curve and then by sampling it to obtain the performance data $\{\widehat{x'}_j, 1 \leq \mathrm{j} \leq \mathrm{J}\}$. Using the sequence of $\widehat{x'}_j$, we extract a total of 15 parameters for use in automatic evaluation. They are labeled as $f_{x,k}$, where the suffix k distinguishes the range of spline curve $(\mathrm{k} = 1)$, the rms difference of spline curve between adjacent notes $(\mathrm{k} = 2)$, and the sum of spline curve from standard $(\mathrm{k} = 3)$. Specifically, they are given by

$$f_{x,1} = \max(\hat{x}') - \min(\hat{x}'), \tag{1}$$

$$f_{x,2} = \sqrt{\sum_{j=2}^{n}\left(\hat{x}_j' - \hat{x}_{j-1}'\right)^2}, \tag{2}$$

$$f_{x,3} = \sum_{j=1}^{n} \hat{x}_j'. \tag{3}$$

For notational convenience, the resulting set of 15 parameters for automatic evaluation is represented as a vector $\mathbf{P} = \{p_j, 1 \leq \mathrm{j} \leq 15\}$.

IV. SYSTEM IMPLEMENTATION

Given a piano performance, the aim of our system is to evaluate the performer's ability in terms of scale fingering and assign a rating score for proficiency. The flow of the proposed method for automatic evaluation is shown in Fig. 2. Given the recorded performance, it first extracts performance deviations concerning onset time, velocity, duration, articulation, and tempo for the MIDI representation. Next, it computes performance data $\widehat{x'}_j$ on the basis of a spline curve representing the global tendency of given performance. Afterwards, a set of 15 evaluation parameters is extracted and is represented as a vector. Finally, the decision tree algorithm is applied to obtain an estimation score for the scale performance.

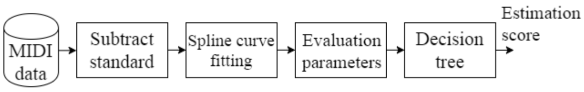

Figure 2. The proposed method for performance evaluation.

The decision tree algorithm [5] is a top-down recursive greedy algorithm and is widely used in data mining applications. Its popularity is due to several advantages: 1) It has a simple structure in need of little background knowledge; 2) Its high efficiency and low complexity are suitable for use in large-sized training dataset; 3) It has high classification accuracy. To begin, we assume that the training data consists of N_T recorded performances. For the $\mathrm{i-th}$ performance, let S_i represent the average subjective score by three expert pianists and let P_i represent the evaluation vector. By sorting the $\mathrm{j-th}$ parameter of N_T elevation vectors in ascending order, we obtain an attribute vector $\mathbf{A}_j = (a_{j,1}, \ldots, a_{j,N_T})$. This implies that we have $N_T - 1$ different ways to split with the split node given by

$$L_{j,k} = \frac{a_{j,k} + a_{j,k+1}}{2}, 1 \leq \mathrm{k} \leq N_T - 1. \tag{4}$$

Proceeding in this way, the data set π_T can be divided into $\pi_{L,j,k}$ and $\pi_{R,j,k}$, representing the data set of left subtree and right subtree, respectively. Let S_T, $S_{R,j,k}$ and $S_{L,j,k}$ denote the average subjective score of scale performances in π_T, $\pi_{R,j,k}$ and $\pi_{L,j,k}$, respectively.

The construction of decision tree can be formulated as the primal quadratic optimization problem. To begin, we compute the information gain as follows:

$$G_{j,k} = SS_T - \left(SS_{R,j,k} + SS_{L,j,k}\right),$$

where:

$$SS_T = \sum_{j \in \pi_T}\left(S_j - \overline{S_T}\right)^2,$$

$$SS_{R,j,k} = \sum_{c \in \pi_{R,j,k}}(S_c - \overline{S_{R,J,k}})^2,$$

$$SS_{L,j,k} = \sum_{c \in \pi_{L,j,k}}(S_c - \overline{S_{L,J,k}})^2. \tag{5}$$

A common approach is to select the one with the maximum information gain to split. To accomplish this, we calculate the information gain for the $\mathrm{j-th}$ evaluation parameter as follows:

$$G_j = \max\left(G_{j,1}, \ldots, G_{j,N_T-1}\right),$$

$$G = \max(G_1, \ldots, G_l) \triangleq SS_T - (SS_R + SS_L),$$

where SS_R corresponds to

$$SS_{R,j,k} \text{and} SS_L \text{corresponds to} SS_{L,j,k}. \tag{6}$$

After selecting the most suitable attribute to split on, we split the decision node into several branches by taking values on the splitting attribute. The above steps are repeated iteratively until the improvement is smaller than the predefined threshold, where the improvement is defined as

$$\mathrm{H} = \frac{G}{SS_T} = 1 - \frac{SS_R + SS_L}{SS_T}. \quad (7)$$

Finally, we calculate the average subjective score of each split note to obtain an estimation score. If the test data q lies below the split note β, its estimation score is computed as follows:

$$S_{DT}(q) = \frac{1}{N_{\mu_\beta}} \sum_{i \in \mu_\beta} S_i,$$

where μ_β is the data set of the split note and N_{μ_β} is the capacity of this data set. (8)

V. Experimental Results

Computer simulations were conducted to compare the estimation score and the subjective scores by three expert pianists. The performance of decision tree is examined by the leave-one-out cross-validation, where one of the N MIDI music files is selected for testing, and the remaining N-1 files are used for training. This process is repeated N times until every music file has been tested. The system performance is evaluated in terms of the correlation coefficient between estimation score and the subjective score by three expert pianists. A group of 11 elementary school students participated in the test. They are asked to play a one octave scale in five different keys using left or right hand. Table 2 shows the correlation coefficients computed with respect to each individual expert pianists and to all of them. The results clearly demonstrate that the proposed system can obtain an estimation score that emulates the experts' subjective preference.

TABLE 2. Correlation coefficients between estimation score and subjective score.

	Expert pianist 1	*Expert pianist 2*	*Expert pianist 3*	*Overall*
Correlation coefficients	0.6003	0.6349	0.6619	0.7625

For further investigation, we also show the decision tree structure in Fig. 3, where n represents the number of MIDI files belonging to each leaf node. The decision tree has a simple structure consisted of 15 nodes and among them, 8 leaf nodes. According to the 8 paths of tree structure, we also show in Table 3 the IF-THEN rules used to evaluate the proficiency of piano performances. The results indicate that only 5 out of 15 evaluation parameters are sufficient to distinguish the 110 recorded performances. They are summarized as follows:

(1) $f_{v,3}$ is the sum of spline curve from standard for MIDI-velocity

(2) $f_{o,2}$ is the difference of spline curve between adjacent nodes for key onset time.

(3) $f_{v,2}$ is the difference of spline curve between adjacent nodes for MIDI-velocity

(4) $f_{g,2}$ is the difference of spline curve between adjacent nodes for KOT.

(5) $f_{g,3}$ is the sum of spline curve from standard for KOT.

By observing the attributes of node 1, 9 and 13, we have found that students of high performance scores can master their skills with high $f_{v,3}$ and low $f_{o,2}$. On the other hand, scale performances of lower scores are often characterized by low $f_{v,3}$ and high $f_{v,2}$ by observing the attributes of node 1 and 2. Support for such observations can be found by listening to the recorded MIDI performances. It was found that student of higher scores tend to play the piano with high volume level and stable key onset time. By contrast, performances of lower scores have in common low volume and unstable onset time. Furthermore, node 1 is the first split node, indicating that its attribute $f_{v,3}$ is the most important parameter for proficiency evaluation, that is the sum of spline curve from constant standard for MIDI-velocity. To illustrate this, we further compare the tendency curves of MIDI-velocity using two recorded performances, one of estimation score 1, and the other of score 5. The results are shown in Fig. 4 and 5. As should be expected, the velocity tendency curve resulting from low score data exhibits high variation and most of them are negative. These results are in accord with subjective listening tests by three expert pianists.

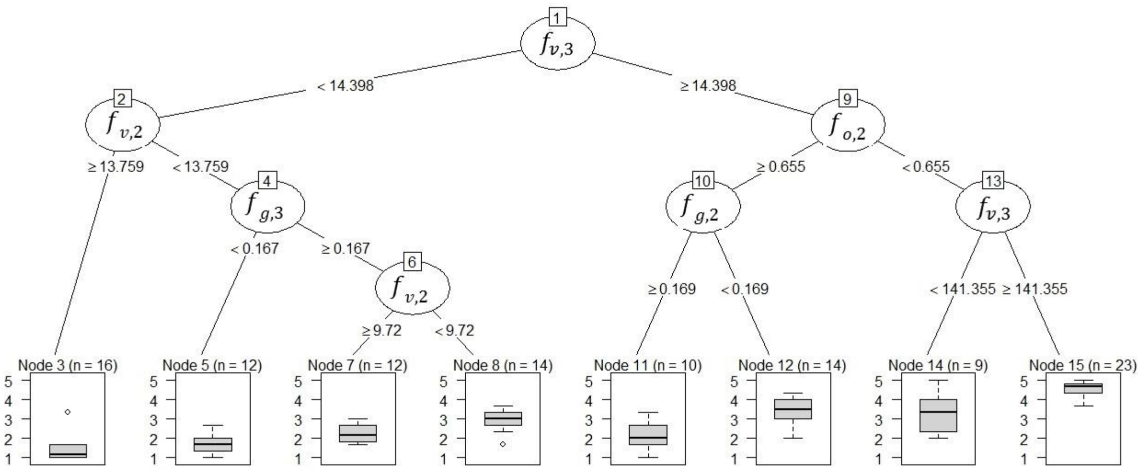

Figure 3. Decision tree for performance evaluation.

TABLE 3. IF-THEN rules derived from decision tree.

Node	*Condition*	*Estimation score*
3	$f_{v,3}$ < 14.3975 $f_{v,2}$ ≥ 13.7585	1.375
5	$f_{v,2}$ < 13.7585 $f_{g,3}$ < 0.16724	1.694
7	$f_{v,3}$ < 14.3975 $f_{g,3}$ ≥ 0.16724 $f_{v,2}$ ≥ 9.7198	2.250
8	$f_{v,3}$ < 14.3975 $f_{g,3}$ ≥ 0.16724 $f_{v,2}$ < 9.7198	2.857
11	$f_{v,3}$ ≥ 14.3975 $f_{o,2}$ ≥ 0.65507 $f_{g,2}$ ≥ 0.16918	2.067
12	$f_{v,3}$ ≥ 14.3975 $f_{o,2}$ ≥ 0.65507 $f_{g,2}$ < 0.16918	3.381
14	$f_{v,3}$ ≥ 14.3975 $f_{o,2}$ < 0.65507 $f_{v,3}$ < 141.355	3.222
15	$f_{v,3}$ ≥ 14.3975 $f_{o,2}$ < 0.65507 $f_{v,3}$ ≥ 141.355	4.536

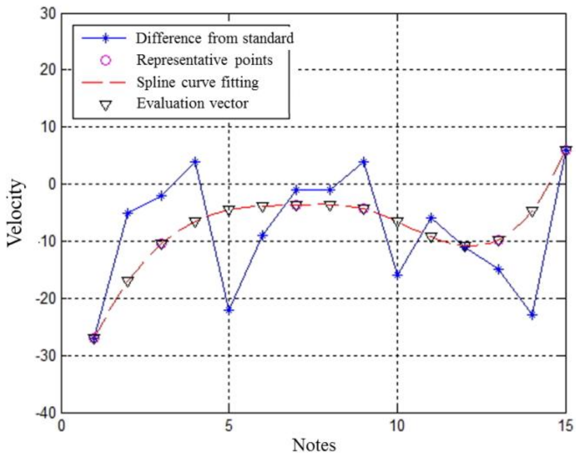

Figure 4.MIDI-velocity curve for a scale performance of score 1.

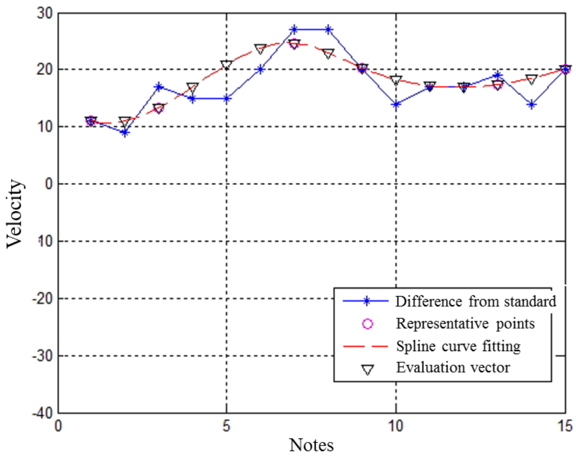

Figure 5. MIDI-velocity curve for a scale performance of score 5.

VI. CONCLUSION

The increasing use of MIDI piano leads the development of supporting systems for self-learners. In this study, we proposed a method of estimating proficiency for a scale performance within one octave. We began by extracting five basic features concerning onset time, velocity, duration, articulation and tempo for the MIDI representation. Based on their deviations from constant standards, we then apply the spline curve fitting to extract the parameters representative of the scale performance. With respect to the automatic evaluation, machine learning algorithms based on decision tree are employed to estimate a rating score that emulates the teacher's subjective preference. Experiments on 11 subjects show that correlation coefficients between estimation scores and subjective scores by experts were 0.76.

REFERENCES

[1] R.B. Dannenberg, M. Sanchez, and A. Joseph, “A computer-based multi-media tutor for beginning piano students,” Interface-Journal of New Music Research, 19(2-3), pp. 155-173, 1990.

[2] Smoliar, S.W., Walterworth, J.A., and Kellock, P.R., “pianoFORTE: a system for piano education beyond notation literacy,” ACM Multimedia 95, San Francisco, CA, 1995.

[3] http://www.smartmusic.com/

[4] Kitamura, T, and M. Miura, “Constructing a Support System for Self-learning Playing the Piano at the Beginning Stage,” Proceedings of International Conference on Music Perception and Cognition, pp. 258-262, 2006.

[5] J. Ross Quinlan, “Induction of decision trees,” 1, 81-106, 1986.