Apollon13: A Training System for Emergency Situations in a Piano Performance

Yuki Yokoyama and Kazushi Nishimoto

Japan Advanced Institute of Science and Technology
1-1, Asahidai, Nomi, Ishikawa, 923-1292, Japan
{y_yoko,knishi}@jaist.ac.jp

Abstract. During a piano performance, there is always the possibility that the musician will cease playing on account of an unexpected mistake. In a concert, such a situation amounts to an emergency state in the piano performance. Therefore, we propose a system named "Apollon13" that simulates emergency states by replacing particular notes with different ones, in the manner of miss-touches, by referring to the performer's degree of proficiency as determined by a performance estimation algorithm. From the results of user studies, we confirmed that Apollon13 is basically effective as a training system for handling emergency states. However, the estimation algorithm could not precisely identify the note-replacement points where the subjects become upset. Accordingly, we evaluated the estimation algorithm by comparing it with the player's subjective assessment based on the data of an experiment. As a result, we found a clear relationship between the subjective assessment and the points, obtained by experiment, at which players become upset. This result suggests that an algorithm could gain the ability to detect the "upset points" by approximating a human's subjective assessment.

Keywords: emergency training, performance estimation, piano performance, note-replacement

1 Introduction

This paper proposes a novel piano-performance training system named "Apollon13." This system aims to foster the ability to avoid performance cessation caused by unexpected mistakes such as miss-touches. Performance cessation, where the performer "freezes up," is a "fatal situation" in a piano concert. Therefore, the performer must avoid such a situation by any means, and the performance must go on despite whether mistakes occur. However, no countermeasures to this situation have been taught in conventional piano lessons, and there is no active training methodology for avoiding performance cessation.

A piano lesson usually consists of several steps. The first step is basic training. In basic training, an educand learns the method of reading scores and trains in fingerings using some etude (e.g., HANON). The second step is building a repertoire. This step is further divided into two sub-steps. The first is partial exercise and the second is full exercise. In this step, the educand learns musical performance and musical expression. Although the educand can build a repertoire in these steps, the educand is not yet able to train for a piano concert. Generally, the way to train one's skills toward a piano concert is simply to repeat a full performance again and again after memorizing the score and fingering. However, this way of training cannot develop the educand's ability to cope with an unexpected accident: The only way to accomplish this is to actually perform in concerts. Obviously, it is impossible for typical educands to use concerts for training.

Various piano-performance training systems have been developed [1][2][3]. However, these systems have only supported the users in becoming able to perform a musical piece accurately in accordance with its score. The problem of performance cessation during a concert has been completely out of the scope of such systems. Consequently, there has been no active ways to train performers in avoiding performance cessation.

In the aerospace field, astronauts and pilots spend much time in training. Of course they learn how to control aircraft and spaceships under normal conditions. However, to accomplish a mission safely, it is much more important to know how to deal with abnormal emergency situations quickly and effectively. For this purpose, in this field, they conduct training for emergency situations using simulators.

We introduce such a situational training concept to piano-performance training. Apollon13 simulates unexpected mistakes as emergency situations. Using Apollon13 in the final stage of exercises before a concert, it is expected that the educand can acquire the ability to avoid the worst result, i.e. performance cessation. There has been no training method or training system against performance cessation up to now. Therefore, we believe that our attempt has very high levels of novelty and utility.

2 How to Simulate Emergency Situations

How to simulate emergency states was important in designing Apollon13. While there are many causes of emergency states, we focused on miss-touches in performance. A miss-touch results in an unexpected sound, which makes the player upset and, in the worst case, leads to performance cessation. To induce a similar situation, Apollon13 replaces a few of the performed notes with different notes. By trying to keep playing even when the output notes are different from his/her intended notes, the player would be able to learn how to recover from miss-touches without falling into performance cessation.

It's important to understand that the note replacement function should be used only in the final stage where the player is repeating the full exercise, in contrast conventional piano-lesson support systems are used in the initial stage. Musicians use various feedbacks in playing musical instruments. In particular, they are alert to auditory feedback. The proposed system's note replacement intentionally breaks our auditory sense. In the initial stage of a piano lesson, however, auditory feedback is a fundamental element. Therefore, the note-replacement function must not be used in the initial stage of a piano lesson.

Previous literature [4] demonstrated that note replacement has the effect of disorienting piano performance. However, although a keyboard with the note-replacement function is used in this research, the objective of this research is to formulate a kind of stuttering model. Therefore, the way of note replacement in the earlier work is factitious, since such miss-touches never happen in real piano performances. To adopt note replacement in piano practice, a note-replacement method that simulates realistic miss-touches is required. To simulate such realistic miss-touches, there are two factors that should be considered: which performed note should be replaced, and by which note. In section 3, we describe the employed simulation method.

3 System Setup

3.1 Overview

Apollon13 is a MIDI (Musical Instrument Digital Interface) based system that consists of a MIDI-keyboard, a personal computer, and a MIDI sound module. Apollon13 has two operation modes: a practice-monitoring mode and a rehearsal mode (Table 1).

In the practice-monitoring mode, the system tracks and records the user's full piano performances. In using this mode, the user repeats the full performance of a musical piece many times. A score-tracking function (described in 3.2) compares each performance with the score and records how accurately it is performed.

When the practice-monitoring mode is finished, the system decides which notes should be replaced. Too many replacements would become an excess burden for the user. Therefore, the system finds only a few notes where the user would surely become upset by note replacement based on the performance estimation results using the recorded tracking data (described in 3.3). We call such a selected note a "replacing-point" hereafter.

In the rehearsal mode, the system tracks the user's performance again. When the user performs the replacing-point, the system replaces this note with another note neighboring the correct note. This is done because an actual miss-touch in piano performances follows such a pattern.

Table 1. Operation mode of proposed system

	Practice monitoring mode	→	Rehearsal mode
System	Score tracking Recording performance	Decision of note-replacement part	Score tracking Note replacement
User	Repeat of full performance		Continue performing even if miss-touches are simulated

3.2 Score tracking

A score-tracking technology is necessary to obtain performance data for performance estimation. Apollon13 utilizes the score-tracking function of "Family Ensemble" (FE) [5]. FE is a piano-duo support system for a novice child and his/her parent who is an amateur at piano. Since FE's score-tracking function is robust, it is applicable to tracking performances that include mistakes.

We modified FE's score-tracking function in two points. First, FE's original score-tracking function tracks only the highest notes at each place. We modified it to polyphony compatible by simply summing all note numbers of the notes in the chord and regarding the sum as the note number of the chord. Second, FE outputs three tracking data: performed position in the score, whether the performed note is correct or incorrect, and timestamp of each performed note. We further added velocity data to represent the loudness of each note for the performance estimation.

3.3 Performance estimation

The aim of performance estimation is to find where the user would surely become upset by note replacement. The performance-estimation algorithm classifies each score-event (i.e., (a) note-on event(s) at the same instant in FE's score data) into four categories. The criterion of estimation is performance stability. If the performance of a score-event is highly stable throughout all performances in the practice-monitoring mode, the score-event is estimated as "skillful." If the performance of a score-event is not so stable, it is estimated as "poor." If the performance of a score-event becomes stable, it is estimated as "improved." The other score-events are estimated as "other."

3.3.1 Factors used for performance estimation

Previous related studies [6] used three factors for the performance estimation, i.e., IOI (Inter Onset Interval), duration, and velocity. On the other hand, we use three factors obtained from the score-tracking function, i.e., IOI (calculated by the received timestamps of the performed score-events), velocity of each score-event, and data on whether performed score-event is correct or erroneous (CE-data, hereafter). Mukai et al. used deviation of IOI for estimating the performance: If the deviation value of the same fingering pattern is large, this pattern is estimated as poorly stable [7]. We also use deviation of IOI as well as that of velocity. However, we calculate the deviations of each score-event in all of the full performances, while Mukai et al. calculated those of the same fingering patterns.

Fluctuation in the overall tempo of each performance influences the deviation of tempo at each score-event. To cancel this effect, we calculated normalized local tempo at each score-event. First, average tempo of each entire performance is calculated. Then, the normalized local tempo is calculated by dividing each local tempo at each score-event by the average tempo of the performance. Here, the note value of each score-event is necessary to calculate the normalized local tempo; therefore, we added the note value data to the score data of FE.

3.3.2 Classification of each score-event

The performance estimation requires at least three sessions, each of which should include at least ten full performances. This algorithm classifies each score-event based on the deviations (stability) with progress of the sessions as follows:

1. Calculating "coarse score"
   A) Calculating "tempo score": First, the deviation of normalized local tempo at each score-event for all performances in *all* practice sessions is calculated. Then, all of the score-events are sorted based on their deviation value. Finally, the 30% of score-events with the smallest deviation values score 2 points, the 30% of score-events with the largest deviation values score 0 point, and the remaining score-events with moderate deviation values score 1 point.
   B) Calculating "velocity score": First, the deviation of velocity at each score-event for all performances in *all* practice sessions is calculated. Then, all score-events are sorted based on their deviation value. Finally, the 30% of score-events with the smallest deviation values score 2 points, the 30% of score-events with the largest deviation values score 0 point, and the remaining score-events with moderate deviation values score 1 point.
   C) Calculating "accuracy score": First, the accuracy rate of each score-event for each practice session is calculated based on CE-data. Then, the transition of accuracy rate for each score-event through all practice sessions is obtained from the regression line of the accuracy rates. Finally, one-third of the score-events, having the highest gradient values of the regression lines, score 2 points, one-third of the score-events, having the lowest gradient values, score 0 point, and the remaining one-third of score-events with moderate gradient values score 1 point.
   D) Coarse score is calculated by the following equation:

$$\textit{Coarse score} = \textit{tempo score} * 5 + \textit{velocity score} * 3 + \textit{accuracy score} * 2 \quad \textbf{(1)}$$

2. Calculating "adjustment score"
   A) Calculating "fine tempo score": First, the deviation of normalized local tempo at each score-event for all performances in *each* practice session is calculated. Then, the transition of deviation for each score-event through all practice sessions is obtained from the regression line of the tempo deviations. Finally, one-third of the score-events with the lowest gradient values of the regression lines score 1 point, one-third of the score-events with the highest gradient values score -1 point, and the remaining one-third of score-events with moderate gradient values score 0 point.
   B) Calculating "fine velocity score": First, the deviation of velocity at each score-event for all performances in *each* practice session is calculated. Then, the transition of deviation for each score-event through all practice sessions is obtained from the regression line of the velocity deviations. Finally, one-third of the score-events with the lowest gradient values of the regression lines score 1 point, one-third of the score-events with the

highest gradient values score -1 point, and the remaining one-third of score-events with moderate gradient values score 0 point.

C) Adjustment score is calculated by the following equation:

$$Adjustment\ score = fine\ tempo\ score + fine\ velocity\ score \quad \mathbf{(2)}$$

3. Classifying each note into one of four categories (skillful, improved, poor and other) based on the coarse score and the adjustment score. Table 2 shows the classifying rules.

Table 2. Classifying rule of performance estimation value

Skillful part	coarse score >= 15
Improved part	coarse score < 15 and adjustment score > 0 or coarse score < 5 and adjustment score = 2
Poor part	coarse score < 5 or coarse score < 15 and adjustment score < 0
Other part	coarse score < 15 and adjustment score = 0

4 Experiments

After users continuously use Apollon13, if they eventually lose the tendency to become upset when certain notes are suddenly replaced by incorrect ones, we can say that it is an effective training system for emergency situations in piano performance. In this experiment, we investigate the effects of training with Apollon13 by analyzing the users' subjective assessments and their performance data.

4.1 Experimental settings and procedures

We conducted experiments with three subjects, who were 23–24-year-old males. They have 18–20 years experience of playing the piano. We prepared a compulsory musical piece "Merry Christmas Mr. Lawrence" composed by Ryuichi Sakamoto, which has only a two-page score. It takes about two minutes to perform it. We selected this piece since it is not too difficult but not so easy to perform by one hand. The subjects received the score one week before the experiments. We asked them to practice one week freely to finish the partial exercise stage. We confirmed that the subjects could play through it before the experiments. Table 3 shows the equipment used in the experiments.

The experimental period was five days, and two sessions were held each day: ten sessions in total for each subject. A session takes about thirty minutes. In a session, each subject was required to perform the compulsory piece ten times or more. The interval between the sessions was at least five hours. The first five sessions were assigned as "practice sessions." In these sessions, Apollon13 works in practice-monitoring mode. The remaining five sessions were assigned as "rehearsal sessions."

Table 3. Equipment used in experiments

MIDI-keyboard	YAMAHA grand piano C5L + silent ensemble professional model
MIDI sound source	YAMAHA MU128
MIDI-IO	Midiman MIDISPORT2×2
PC	Notebook type, CPU: Core2Duo T7250 2.00 GHz, memory: 1.0 GB

In these sessions, Apollon13 works in rehearsal mode. In one rehearsal session, we enabled Apollon13's note-replacement function in about five randomly selected performances.

The number of replacing-points in one performance was four, and these were selected according to the results of the performance estimation obtained in the practice sessions. At present, although the performance estimation algorithm works, it cannot decide in which category the users definitely become upset. Therefore, in this experiment, the system chooses one replacing-point for each category classified by the performance-estimation algorithm (four points in total) to collect data for validating the performance-estimation algorithm. At the end of each practice session, we asked the subjects to indicate, note by note, where they could skillfully perform and where they could not. At the end of each rehearsal session, we asked the subjects where they became upset.

4.2 Results

Figure 1 shows the transition of the ratio of the number of replacing-points where each subject became upset to the number of all replacing-points in each session. The horizontal axis indicates the rehearsal sessions and the vertical axis indicates the ratio. Thus, as the sessions progressed, the subjects gradually came to avoid getting upset by the note replacement.

To investigate the effect of note replacement in detail, we analyzed fluctuations in the performances before and after the replacing-points. For this analysis, we first prepared the target performances (with note replacement) and the baseline performances (without note replacement). We employed the performances in the 4th and 5th practice sessions as the baseline performances: average IOI and velocity of each score-event of the baseline performances were calculated as the baseline data. On the other hand, we prepared two target performances: "R1-3" target performances consist of the performances where the note-replacing function was activated in the 1st to 3rd rehearsal sessions, and "R3-5" target performances consist of those in the 3rd to 5th rehearsal sessions. We also calculated average IOI and velocity of each score-event of R1-3 and R3-5. Finally, the difference values of seven score-events before and after each replacing-point were calculated (namely, the difference values at the 15 points in total, including the replacing-point, were obtained).

Figure 2 shows an example of the obtained average difference values between subject A's R1-3 performances and his baseline performances at a certain replacing-point. The horizontal axis indicates the score events. The 8th event corresponds to the

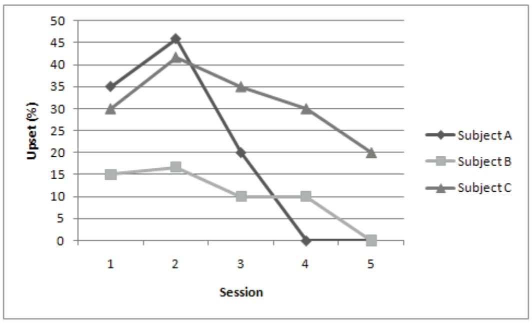

Fig. 1. Transition of percentage of upset points

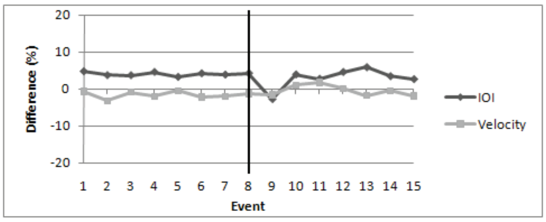

Fig. 2. Difference in performance around a replacing-point

replacing-point. The vertical axis indicates the average difference. If the performance becomes disordered by the replaced note at the 8th event, the graph becomes undulant after there. Therefore, we compared deviation of the performance data before and after the 8th score-event by F-test. As a result, we found a significant difference in IOI of subject A's R1-3 ($p<0.05$) and a marginal difference in IOI of his R3-5 ($p<0.1$). We could not find differences in IOI of subjects B and C or in velocity of any subject.

5 Validation of Performance-Estimation Algorithm

From the results of the experiment described in 4.2, continual use of the system made the subject imperturbable despite the note replacement. However, in the experiments, we selected the replacing-points from all four categories, i.e. skillful, improved, poor and other, as estimated by the proposed algorithm described in 3.3.2. In order to more accurately and dependably simulate emergency situations by note replacement, it is necessary to find which category is the most effective as well as to validate the proposed algorithm's classification performance.

5.1 Algorithm-estimated categories

Figure 3 shows the results of classification by the estimation algorithm. We designed this algorithm to classify all of the score-events into four categories as evenly as possible. Similar to 4.2, Table 4 shows the comparison results of the deviation of the performance data in IOI and velocity before and after the replacing-points by F-test for each subject and each category. In Table 4, "**" indicates a significant difference ($p<0.05$) and "*" indicates a marginal difference ($p<0.1$).

The results in Table 4 show that the replacing-points at which the subjects can become upset are distributed into multiple categories. Thus, it is actually difficult to isolate one category that can effectively make the subjects upset based on the algorithm-estimated categories.

5.2 Subject-estimation-based categories

At the end of each practice session, we asked each subject to categorize his own performance into one of two categories (skillful or poor). Based on the estimation results, we translated them into four categories corresponding to those of the algorithm estimation as follows. First, we gave estimation scores 1 and -1 to skillful score-events and poor score-events, respectively, while we gave estimation score 0 to the score-events where the subjects did not classify performance to either category; this was done for all of the estimation results of all practice sessions. Second, we

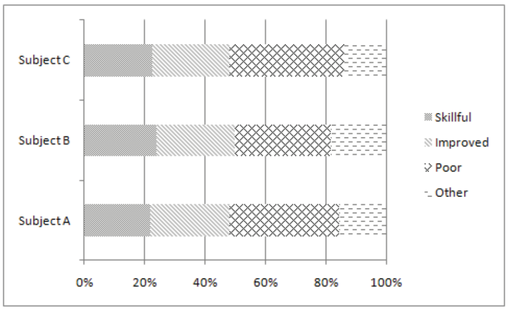

Fig. 3. Results of estimation based on algorithm

Table 4. Comparison results of the deviation of the performance data in IOI and velocity before and after replacing-points by F-test for each subject and each category

Category	Session	Subject A		Subject B		Subject C	
		IOI	Vel	IOI	Vel	IOI	Vel
Skillful	R1-3	**					
	R3-5						**
Improved	R1-3	*		*	**		
	R3-5			*	*		
Poor	R1-3			*			
	R3-5				*		
Other	R1-3	**				**	
	R3-5					**	

calculated regression lines of the estimation scores for each score-event: the x-axis corresponds to the number of the practice session and the y-axis corresponds to the estimation value. Third, we looked for the score-events whose regression line's gradient is positive: we classified these score-events into the "improved" category. Finally, the score-events that were not classified into the improved category were classified into "skillful" if their estimation score of the final (5^{th}) practice session was 1, "poor" if it was -1, and "other" if it was 0.

Figure 4 shows the results of subject-estimation-based classification into four categories. Each subject has different estimation criteria. The concordance rate between algorithm classification and subject-estimation-based classification is about 25% for each subject.

Similar to 5.1, Table 5 shows the comparison results of the deviation of the performance data in IOI and velocity before and after the 8^{th} score-event by F-test for each subject and each category obtained by the subject-estimation-based classification. In Table 5, "**" indicates a significant difference (p<0.05) and "*" indicates a marginal difference (p<0.1).

The results in Table 5 show that the replacing-points at which the subjects become upset are concentrated in specific categories, although the specific categories depend on the subjects. For subject A, the disorder of IOI gathers in the "Other" category and the disorder of velocity gathers in the "Improved" category. For subject B, IOI disorder gathers in the "Poor" category and velocity disorder gathers in the

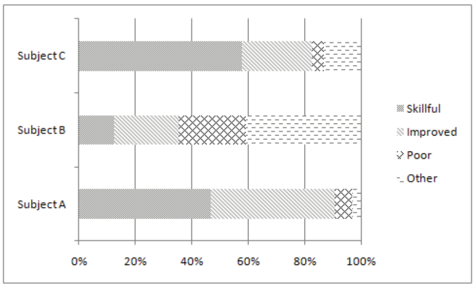

Fig 4. Results of subject-estimation-based classification into four categories

Table 5. Comparison results of the deviation of the performance data in IOI and velocity before and after replacing-points by F-test for each subject and each category obtained by the subject-estimation-based classification

Category	Session	Subject A		Subject B		Subject C	
		IOI	Vel	IOI	Vel	IOI	Vel
Skillful	R1-3						
	R3-5						
Improved	R1-3		**			*	
	R3-5		**		**	**	
Poor	R1-3			**			
	R3-5			*			
Other	R1-3	**				**	
	R3-5	**					*

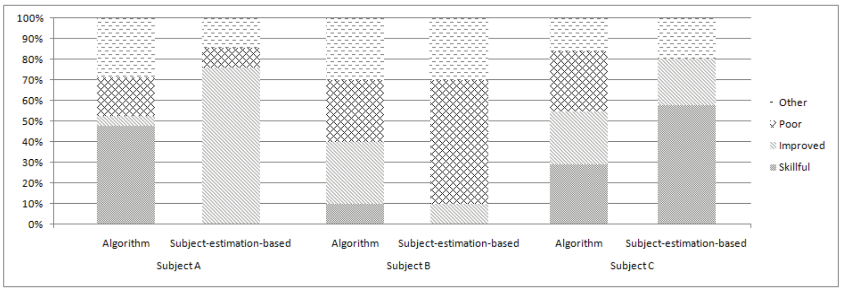

Fig. 5. Classification results of subjectively upset replacing points.

"Improved" category. For subject C, IOI disorder gathers in the "Improved" and "Other" categories, and velocity disorder gathers in the "Other" category.

5.3 Relations between subjectively selected upset-points and categories

We classified the replacing-points where the subjects became upset (this data was obtained from the interview results after each rehearsal session) into algorithm-estimated categories and subject-estimation-based categories. Figure 5 shows the results. These results show that the replacing-points where the subjects became upset also gather into specific subject-estimation-based categories rather than into specific algorithm-estimated categories. Based on the subject-estimation-based categories, 76% of the upset-points gather in the "Improved" categories for subject A, 60% gather in the "Poor" category for subject B, and 58% gather in the "Skillful" category for subject C. However, based on the algorithm-based categories, the upset-points are distributed among all four categories for all subjects. Consequently, we can find fairly strong evidence of a relationship between the subject-estimation-based categories and the upset-points, although this relationship differs depending on the subject.

6 Discussion

From the results shown in Fig. 1, although all of the subjects became upset by many replacing-points in the early rehearsal sessions, they gradually gained imperturbability through training with Apollon13. Furthermore, from the comparison results between R1-3 and R3-5 for subject A, the objective degree of performance disorder in IOI decreased. Consequently, these results show that Apollon13 has a certain amount of efficacy.

We further investigated the relationships between subjective/objective upset-points and algorithm-estimated categories/subject-estimation-based categories (5.1–5.3). We found that both objective and subjective upset-points gather in a specific subject-estimation-based category, depending on the subject, while both objective and subjective upset-points were distributed into many algorithm-estimated categories. This result suggests that if we could classify the score-events into four categories similar to the subject-estimation-based categories, we would be able to more reliably make the users upset by a note-replacing technique. Accordingly, a more practical simulator of emergency situations in piano performance can be achieved.

7 Conclusion

In this paper, we proposed Apollon13, which is a training system for emergency situations in a piano performance. Apollon13 simulates "miss-touches" by using a note-replacing function. The user can gain the ability to overcome unexpected accidents by continuing the performance even when an unexpected note replacement happens while using Apollon13. Therefore, he/she can train to avoid the worst situation of a piano performance in a concert, i.e., performance cessation.

From the results of experiments using three subjects, we confirmed that Apollon13 has a certain amount of efficacy as a training simulator of emergency situations in piano performance. However, the algorithm for classifying the score-events into four categories, i.e. skillful, improved, poor, and other, is still inadequate. As a result, the present system cannot dependably make the user upset by note replacement. To more reliably simulate emergency situations, it is necessary to develop a classification algorithm that can classify the score-events in the manner of subjective human judgment. Accordingly, we intend to tackle this problem in the future.

References

1. Dannenberg, R.B., Sanchez, M., Joseph, A., Joseph, R., Saul, R., and Capell, P.: Results from the Piano Tutor Project. in Proceedings of the Fourth Biennial Arts and Technology Symposium, pp. 143--150(1993).
2. Shirmohammadi, S., Khanafar, A., and Comeau, G.: MIDIATOR: A Tool for Analyzing Students' Piano Performance. In: Revue de recherche en education musicale vol.24(2006).
3. Smoliar, S.W., Waterworth, J.A., and Kellock, P.R.: pianoFORTE: A System for Piano Education Beyond Notation Literacy. Proceedings of the third ACM international conference on Multimedia, pp. 457--465(1995).
4. Takahashi, N.: Effect of auditory feedback disturbance in keyboard performance - Evaluation of possibility as a model system for stuttering-. IEICE Technical Report Vol. 106 No. 501 pp. 87--92 (2007) (in Japanese).
5. Oshima, C., Nishimoto, K., and Hagita, N.: A Piano Duo Support System for Parents to Lead Children to Practice Musical Performances. ACM Transactions on Multimedia Computing, Communications and Applications (ACM TOMCCAP), Vol. 3, Issue 2, Article 9, 2007.
6. Akinaga, S., Miura, M., Emura, N., and Yanagida, M.: Toward realizing automatic evaluation of playing scales on the piano. Proc. of ICMPC9, pp. 1843--1847 (2006).
7. Mukai, M., Emura, N., Miura M., and Yanagida, M.: Generation of suitable phrases for basic training to overcome weak points in playing the piano. MUS-07-018 (2007).