TRANSFERRING PIANO PERFORMANCE CONTROL ACROSS ENVIRONMENTS

Maoran Xu[⋆†] *Ziyu Wang*[⋆] *Gus G. Xia*[⋆]

[†] University of Florida [⋆]New York University Shanghai

ABSTRACT

Player pianos driven by computers are able to record and reproduce various performance control parameters, including pitch, timing, velocity and pedaling. However, the resulting sound of performance is not 100% reproducible in a new environment due to the difference in room acoustics and physical properties of the piano. Inspired by the Psychoacoustic studies which showed that human pianists adjust their controls in new environments for better performances, we have developed a system that automatically transfers performance control across environments in order to make the reproduced sound as similar as the original one. In specific, our work includes (1) a systematic measurement of the control-sound relationship of player pianos under different environments, and (2) a novel algorithm to adjust the control parameters through interpolating the measured control-sound functions. We evaluated the effectiveness of our method by conducting a listening test. Experimental results show that our algorithm outperforms the baseline significantly.

Index Terms— Expressive performance, player piano, performance transfer, audio similarity.

1. INTRODUCTION

It has been discovered by both musicians and scientists [1–3] that professional pianists adapt their performance controls such as timing and velocity nuances to different room acoustics and pianos, sometimes even unconsciously, so that the sound effect is consistent across environments. This phenomenon motivates us to design a corresponding mechanism for computer-driven performance systems. In this paper, we present such a system that transfers the control parameters of player pianos across environments so that the difference of perceived sound effect is minimized. The whole procedure is illustrated in Fig. 1. The control C_1 in the source environment E_1 generates Sound 1, while the control C_2 in the target environment E_2 generates Sound 2. We regard the environment $E(\cdot)$ as a function that unifies the factors of room and piano acoustics, mapping the control parameters to actual sound. The goal is to find the optimal C_2 (transferred from C_1) that minimizes the "distance" between Sound 1 and Sound 2. Formally:

$$C_2^* = \underset{C_2}{\arg\min}\{dist[E_1(C_1) - E_2(C_2)]\}. \tag{1}$$

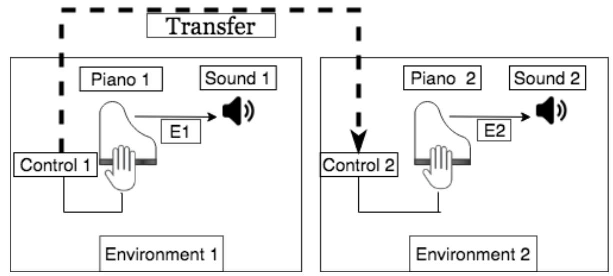

Fig. 1. A system diagram of the performance transfer system.

We first measured the performance of player pianos in two environments via recording single tones under different performance controls. In particular, we encoded the controls of each key using the MIDI protocol [4–6] and quantified the actual sound of each pitch by *sound intensity* and *sound duration*. Through signal analysis, we have found that sound intensity and sound duration have monotonic, nearly piecewise linear dependency on MIDI velocity and MIDI duration respectively, but the behavior on each pitch varies a lot with no obvious patterns. Moreover, MIDI velocity and MIDI duration have joint effects on the two sound features. Based on these observations, we designed an iterative interpolation algorithm for velocity and duration control transfer. As for sustain pedal control, we analyzed the effective ranges of sustain pedal, and designed two algorithms for pedal control transform: a note-extension method and a range-transfer method.

This is the first study on the topic of performance control transfer to our knowledge (based on a comprehensive discussion of different types of music style transfer in [7]). To test the effectiveness of the proposed method, we applied it on several music excerpts and conducted a listening test. Experiments show that, in a new environment, the transferred control generates significantly better performances compared to the baseline (original) controls .

2. DATA REPRESENTATION

The performance control for any piece of music can be encoded with a MIDI file. For each note, it is sufficient to represent the control with $C = (p, v, o, d, s)$[4], where p denotes *pitch*, v denotes *velocity* (speed with which the key is hit), o denotes *onset*, d denotes *duration*, and s denotes the control of sustain pedal. To be specific, p and v are integers ranging from 0 to 127, o and d are floating point numbers in seconds,

Room	Size	Piano Model	Features
Aud	$100\ m^2 \times 3m$	Disklavier DGCIE3	A spacious hall with long reverberation time.
Lab	$20\ m^2 \times 2.5m$	Disklavier DYUS1 ENST	A small, soundproof studio with short reverberation time.

Table 1. Room conditions for the experiments.

and s is an integer array that describes the controlled depth of the pedal over the duration of note.

We represented sound S by frequency F, intensity I, note onset O, and sound duration D. By [8] and [9], the intensity perceived by human is the average intensity of the first phase of decay from the peak. Therefore we measured the intensity in decibel by calculating root mean square amplitude of the first 10 ms after the peak. Formally,

$$I = 20 \log \frac{P_{rms}}{P_{ref}}, \tag{2}$$

$$P_{rms} = \sqrt{\frac{1}{\Delta t} \int_{t_{peak}}^{t_{peak}+\Delta t} 10A^2(t)\mathrm{d}t}, \tag{3}$$

where $A(t)$ denotes the amplitude at time t, Δt= 10 ms and the constant reference threshold $P_{ref} = 1 \times 10^{-4}$.

The sound duration of each note is computed by sliding a 5 ms-length window from the peak of the notes until the average intensity of the window falls under a certain threshold. Formally,

$$D = \arg\min_t \{I_t < I_{ref}\} - t_{peak}, \tag{4}$$

where I_t is the average intensity of the t-th window.

To conclude, the input and output of the function $E(\cdot)$ are the control parameters $C = (p, v, o, d, s)$ and the sound parameters $S = (F, I, O, D)$. Note that $E(\cdot)$ could vary a lot for different environments, causing the same C to trigger very different S. The idea of performance control transfer is to adjust C in a new environment so that the new S remains as similar as the original one.

In general, we can assume that the same MIDI pitch always generates the same frequency and the sound starts right at the MIDI onset. That is, we can keep (p, o) of a single note and only transfer (v, d, s) in a new environment based on the relationship between (v, d, s) and (I, D). In the rest of the paper, we first consider the transfer for (v, d) from section 3 to section 4.1 and then consider the transfer for s in section 4.2.

3. A MEASUREMENT OF PIANO PERFORMANCES

We conducted a systematic measurement in order to explore the relationship between performance control and sound effect of play pianos under different environments. We describe the recording setup in section 3.1 and discuss the control-sound relationship in section 3.2.

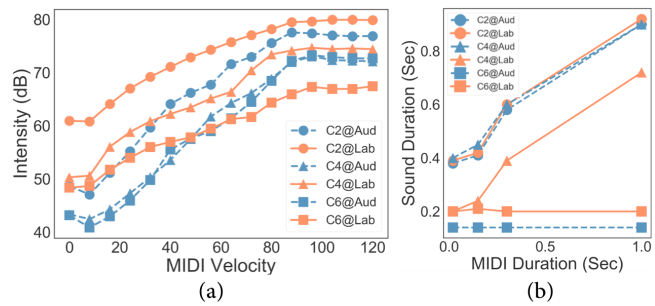

Fig. 2. An illustration of control-sound relationship across pitches in different environments: (a) v-I plot for fixed d = 0.7 sec, (b) d-D plot for fixed v = 79.

3.1. Recording Setup

Recording was carried out on two Disklavier pianos placed in two rooms, both were tuned right before the recording sessions. The source piano was placed in a soundproof music lab (henceforth referred to as Lab), while the target piano was placed in a spacious auditorium (henceforth referred to as Aud). More details are described in Table 1. We used a TASCAM DR-100 digital recorder, which was placed 20 inches away from the keyboard acting as the player's ears and was kept still during the whole recording procedure. Double channels with a sampling rate of 48.0 kHz and a resolution of 24 bits were used in the recording.

3.2. The Control-Sound Relationship

We recorded the samples of all possible p (88 keys), each associated with sixteen levels of v ranging from *ppp* (1 MIDI velocity unit) to *fff* (127 MIDI velocity units) and four levels of d ranging from *staccato* (0.02 sec) to *legato* (1.0 sec). In total, we recorded $88 \times 16 \times 4 = 5632$ samples of each pitch (for both pianos) and computed their I and D by Equation 2 and 4. In this section, we first analyze the v-I relationship and d-D relationship and then discuss the joint effect of d and v.

Fig. 2 shows the v-I and d-D relationship of three selected pitches (C2, C4, and C6) in the two different environments. The curves of different environments are quite different from each other (e.g., the curves for C6 in (a) have completely different slopes). This fact supports the necessity of a control transfer method. In addition, the curves of different pitches varies a lot with no obvious regulation. For example, the curve for C2 at Lab is always above the curve at Aud. However, the curves for C6 intersect near $v = 50$. In general, we found no dependency between the differences and the pitch values, which motivates us to build a tailored control transfer model for each individual pitch. v and d have joint effects on both I and D. When MIDI duration increases, the v-I

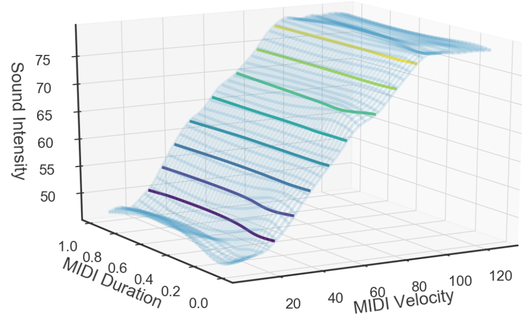

Fig. 3. An illustration of the joint effect of (v, d) on I with contours, measured in the target environment.

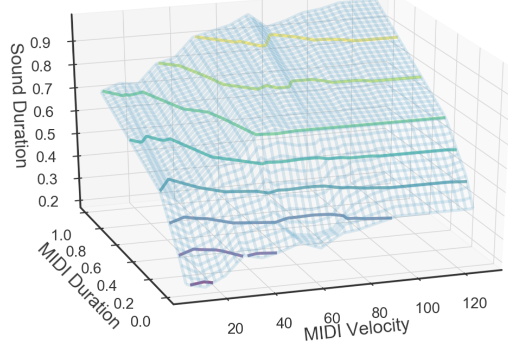

Fig. 4. An illustration of the joint effect of (v, d) on D with contours, measured in the target environment.

relationship also fluctuates, as shown in Fig. 3. When MIDI velocity increases, it takes longer time for a note to fall below 10 dB and causing longer sound duration, as shown in Fig. 4. We conducted repeated measurement and found the fluctuation was not measurement error. By [10], the fluctuation mainly comes from the mechanical deviation in player pianos, which is also part of the physical qualities of pianos that should be considered by the transfer algorithm.

4. THE PERFORMANCE TRANSFER ALGORITHM

Based on the control-sound relationship discovered in section 3.2, we introduce *PETA* (the PErformance Transfer Algorithm) in section 4.1. It solves the velocity and duration transfer problem for each pitch. As for the pedal control transfer, we proposed two ways including a note-extension method and a range-transfer method, and describe how to incorporate pedal transfer into the scheme in section 4.2.

4.1. Algorithm for Velocity and Duration Transfer

By section 3.2, (v, d) has a joint effect on (I, D). Therefore, when a note of known pitch and onset is played with no sustain pedal, we can simplify the notation of environment functions as

$$\begin{aligned} E_i(v_i, d_i) &= \Big(f_i(v_i, d_i), g_i(v_i, d_i)\Big) \\ &= (I_i, D_i), \text{ for } i = 1, 2. \end{aligned} \tag{5}$$

Here, i is the index of the environment, f_i is the mapping $(v_i, d_i) \to I_i$ and g_i is the mapping $(v_i, d_i) \to D_i$. Since we recorded pairs of control and sound as grid points in Fig. 3 and 4, we extended both f_i and g_i to continuous functions by linear interpolation [11] so that any value of control vector has a corresponding sound intensity and duration. Note that f_i and g_i are different for different pitches. For notation simplicity, we omit the pitch parameter.

Given a control vector (v_1, d_1) in E_1, the expected sound $(I,\, D)$ can be computed by f_1 and g_1. The optimal control (v_2^*, d_2^*) in E_2 is then computed by

$$\begin{aligned} (v_2^*, d_2^*) &= \arg\min_{v,d}[dist\{E_2(v, d), (I, D)\}] \\ &= (\arg\min_{v} |f_2(v, d_2^*) - I|, \arg\min_{d} |g_2(v_2^*, d) - D|). \end{aligned} \tag{6}$$

We proposed an iterative coordinate-search algorithm to find the optimal control. To be specific, the initial vector is set to be $(v_2^{(0)}, d_2^{(0)}) = (v_1, d_1)$. At k-th iteration, we search along the v-axis to find a point $v_2^{(k)}$ that minimizes $|f_2(\cdot) - I|$; then we search along the d-axis to find a point $d_2^{(k)}$ that minimizes $|g_2(\cdot) - D|$. The search procedure is conducted iteratively until convergence, the detailed of which is shown in Algorithm 4.1. Theoretically, if both of the implicit functions $f_2 = I$ and $g_2 = D$ have proper Lipschitz-continuous properties, the *Banach Fixed Point Theory* [12] ensures the algorithm to converge to the optimal solution. In practice, we set the thresholds $\delta_v = 1$ and $\delta_t = 0.02$, and the algorithm converges within 3 to 4 iterations.

Algorithm 1 Calculate v_2^* and d_2^* when v_1 and d_1 is played on the source piano with pitch p and onset o, without sustain pedal ($s = 0$).

$I \leftarrow f_1(v_1, d_1), D \leftarrow g_1(v_1, d_1)$
$v_2^{(0)} \leftarrow v_1$
$d_2^{(0)} \leftarrow d_1$
do:
  $v_2^{(k+1)} \leftarrow \arg\min_v |f_2(v, d_2^{(k)}) - I|$
  $d_2^{(k+1)} \leftarrow \arg\min_d |g_2(v_2^{(k+1)}, d) - D|$
while $v_2^{(k+1)} - v_2^{(k)} > \delta_v$ **and** $d_2^{(k+1)} - d_2^{(k)} > \delta_d$
$(v_2^*, d_2^*) \leftarrow (v_2^{(k+1)}, d_2^{(k+1)})$

4.2. Transfer Methods of the Sustain Pedal

When the sustain pedal is activated, the damper is lifted and every string makes free vibration. We proposed two methods to transfer the pedal control and incorporate them into PETA respectively as follows.

PETA1: The function of the sustain pedal was simplified as merely extending the duration of notes. We applied "hold-on note" method by first removing the pedal information in the original MIDI files and extending the duration of each note till the ending time of the corresponding pedal. Then, we applied PETA on the modified control.

PETA2: We conducted a range-transfer on the depth of the pedal. The MIDI pedal depth is encoded by integers ranging from 0 to 127. However, the pedal has no effect when stepped too shallow and no longer changes the effect when stepped deeper than certain threshold. Therefore, we measured the "effective range" as the depth range on which the sound duration changes with pedaling. Formally, if the effective range of the piano is $[s_{m1}, s_{M1}]$ in E_1 and $[s_{m2}, s_{M2}]$ in E_2 given the sustain pedal depth in E_1 as s_1, the transferred s_2 is computed by

$$s_2 = \frac{s_1 - s_{1m}}{s_{1M} - s_{1m}}(s_{2M} - s_{2m}) + s_{2m}, \tag{7}$$

and other control parameters are still transferred by PETA.

5. EXPERIMENT

5.1. The Listening Test

We recorded three classical pieces: Debussy's Gradus Ad Parnassum L. 119, Chopin's Nocturne Op. 2 No. 9 and Mozart's sonata Kv. 545 under the two environments described in Table 1. PETA1 and PETA2 were applied on all three pieces, considering the Lab as the source environment and auditorium as the target environment. All of the pieces were recorded from the same microphone position as used for measuring the control-sound relationship.

Fig. 5 shows an example which illustrates the objective difference between the original and the transferred performance controls. Here, (a) is the score, (b) is the piano roll representation of the original performance control version played on the piano in the source environment and (c) is the piano roll of the transferred version in the target environment.

We evaluated the performance of our algorithm through a listening test among 20 pianists ranging from 20 to 31 years old. The participants were asked to listen to one original version from the source environment and three performance versions from the target environment of each piece. The performance versions are played with control parameters that are (1) the same MIDI as in Lab without transfer, (2) the transferred MIDI with PETA1 and (3) the transferred MIDI with PETA2. The order of the versions was counterbalanced. After each piece, participants were asked to rate the similarity of the performance versions to the recorded version. The grading was on a scale from 1 (not similar) to 5 (completely the same).

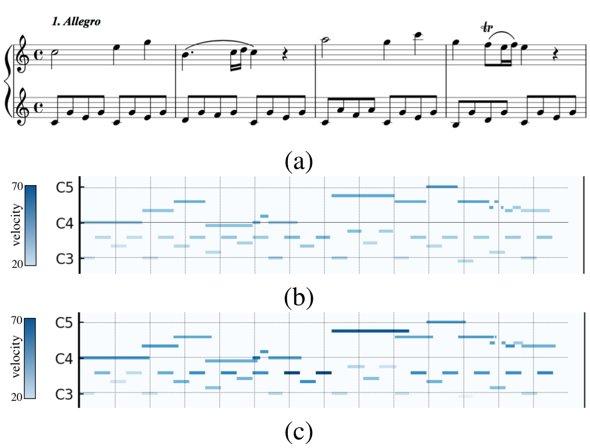

Fig. 5. An illustration of a transferred performance using PETA on Mozart's Piano Sonata No.16, Kv. 545.

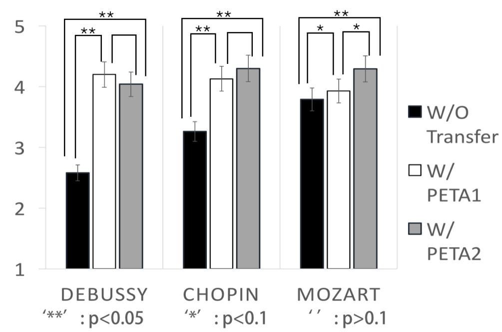

Fig. 6. The subjective evaluation results of the transfer methods on three music excerpts.

5.2. The Experiment Results

A Wilcoxon signed-rank test was conducted to explore the significance of differences among 3 versions. The test compares each pair of scores to identify whether their population mean differ. Evaluation results are shown in Fig. 6. For each piece of music, the transferred ones were rated significantly higher. Though PETA2 is slightly better than PETA1 overall, difference between the two transfer methods was quite marginal. The gaps between the baseline and PETA are large for Debussy and Chopin, but smaller for Mozart. This is mainly because pedals are less used in Mozart.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we present a first method that transfers control parameters on computer-driven player pianos to cope with the deviation caused by environments. Data were collected by recording single tones in different environments and were analyzed by signal computing. The results of the listening test reveal success in increasing audio similarity across different environments.

There are still several challenges to be solved in the future. For example, the experimental results have revealed an interesting phenomenon that the audience can not tell the difference between the extended-note-simulated sustain pedal and the real one. This phenomenon needs to be studied more so that we can conduct the pedaling transfer better. Furthermore, our method can be developed into an online learning system. The computer-driven pianos may receive the signals and analyze them while playing, and then automatically transfer according to the environment. Thus, they can adjust to the environment like professional pianists and require no pre-training.

7. REFERENCES

[1] Alexander Galembo, "Perception of musical instrument by performer and listener (with application to the piano)," in *Proceedings of the International Workshop on Human Supervision and Control in Engineering and Music, Kassel, Germany*, 2001, pp. 257–266.

[2] S. Bolzinger, O. Warusfel, and E. Kahle, "A study of the influence of room acoustics on piano performance," *Journal De Physique IV*, vol. 4, no. C5, pp. 617–620, 1992.

[3] Zora Schärer Kalkandjiev, *The influence of room acoustics on solo music performances: an empirical investigation*, Logos Verlag Berlin GmbH, 2015.

[4] Otto Ortmann, *The physical basis of piano touch and tone*, Read Books Ltd, 2013.

[5] Werner Goebl, Roberto Bresin, and Ichiro Fujinaga, "Perception of touch quality in piano tones," *The Journal of the Acoustical Society of America*, vol. 136, no. 5, pp. 2839–2850, 2014.

[6] Werner Goebl, Roberto Bresin, and Alexander Galembo, "Once again: The perception of piano touch and tone. can touch audibly change piano sound independently of intensity," in *Proceedings of the International Symposium on Musical Acoustics*, 2004, pp. 332–335.

[7] Shuqi Dai, Zheng Zhang, and Gus G. Xia, "Music style transfer: A position paper," 2018.

[8] Alexander Adli, Zensho Nakao, and Yanunori Nagata, "Calculating the expected sound intensity level of solo piano sound in midi file," *Scis & Isis*, vol. 2006, pp. 731–736, 2008.

[9] Caroline Palmer and Judith C Brown, "Investigations in the amplitude of sounded piano tones," *The Journal of the Acoustical Society of America*, vol. 90, no. 1, pp. 60–66, 1991.

[10] Werner Goebl and Roberto Bresin, "Are computer-controlled pianos a reliable tool in music performance research? recording and reproduction precision of a yamaha disklavier grand piano," in *Proceedings of the 2001 Workshop on Current Research Directions in Computer Music*, 2001, pp. 45–50.

[11] Philip J Davis, *Interpolation and approximation*, Courier Corporation, 1975.

[12] Andrzej Granas and James Dugundji, *Fixed point theory /*, PWN-Polish Scientific Publishers,, 1982.