

MirrorFugue III: Conjuring the Recorded Pianist

Xiao Xiao, Anna Pereira and Hiroshi Ishii

MIT Media Lab

75 Amherst St.

Cambridge, MA 02142

[x_x, anna_p, ishii]@media.mit.edu

ABSTRACT

The body channels rich layers of information when playing music, from intricate manipulations of the instrument to vivid personifications of expression. But when music is captured and replayed across distance and time, the performer's body is too often trapped behind a small screen or absent entirely.

This paper introduces MirrorFugue III, an interface to conjure the recorded performer by combining the moving keys of a player piano with life-sized projection of the pianist's hands and upper body. Inspired by reflections on a lacquered grand piano, our interface evokes the sense that the virtual pianist is playing the physically moving keys.

Through MirrorFugue III, we explore the question of how to viscerally simulate a performer's presence to create immersive experiences. We discuss design choices, outline a space of usage scenarios and report reactions from users.

Keywords

piano performance, musical expressivity, body language, recorded music, player piano, augmented reality, embodiment

1. INTRODUCTION

For much of human history, music could only be heard live; audiences gathered to watch performers of physical instruments at a specific location and time. The communication of musical ideas was intricately and inextricably tied to the body of the performer. The development of audio recording in the late 19th century untethered music from the physical world of performer, instrument and concert hall [16]. Technology enables music to transcend the constraints of space and time—people can now listen to music anywhere at any time, even play with one another over distance [3].

But music loses its tight coupling with physically present performers when recorded and replayed across distance and time. Video may capture performers' movements and expressions, but the images only exist behind screens, not present in the world. More often, the performer's body is absent entirely, completely divorced from the resulting music. A musical performance does not need to be seen to be understood, and music extricated from the physical trappings of live performance affords a purer type of listening

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

NIME'13, May 27 – 30, 2013, KAIST, Daejeon, Korea.

Copyright remains with the author(s).



Figure 1: Photo of our interface. The hands and body of a projected virtual pianist are synchronized with moving keys of a player piano.

with its own advantages [15]. However, the presence of the live performer adds layers of richness to the musical experience, from displays of technical virtuosity to expression projected through the face and body [6].

This paper explores the relationship between music and the performer's body in the experience of recorded piano performances. A central question of our work is how to simulate a *visceral* feeling of the pianist's presence by piecing together information streams current technology can readily capture. Snibbe et al. describe a *visceral* mode of perception as an understanding of reality based in the body that precedes analytic, symbolic and linguistic thought [20]. During a live performance, the performer engages the audience physically and emotionally on a visceral level. Our work seeks not only to make available and legible streams of captured information but also to present them so as to engage the audience as in a live performance. In essence, we seek to conjure the recorded pianist.

We introduce, MirrorFugue III, an interface for playback of captured piano performances that combines life-sized video projection of the pianist's hands and body with the actuated keys of a Yamaha Disklavier grand piano [23], evoking the impression that the virtual pianist is playing the physically moving keys. This paper begins with a review of techniques for conveying human presence followed by the rationale and implementation of our system.

Being able to feel the presence of a performer outside of live performance enriches the space of interactions with recorded and remote playing. We present a set of scenarios illustrating our system's potential as an expressive medium for experiencing music as well as some usages for pedagogy. We then discuss results from a preliminary evaluation in support of our core claims.

2. RELATED WORK

We first survey two approaches of depicting the presence of a human being, figurative and disembodied representations. We then look at ways video has been combined with the player piano.

2.1 Figurative Representations

A common way to depict human presence is by rendering the figure through video. Interfaces for remote collaboration often feature a video stream of a remote user on a screen. Extensive research has revealed ways in which choice of framing, size and placement of the video effect empathy and the feeling of co-presence [13, 8, 9]. Studies have shown that including the upper-body in the video drastically improves collaborators' empathy over a head-only display [13] and that approximately life-sized video is more effective at simulating presence than smaller sizes [8].

Several researchers have explored how the arrangement of video screens can simulate real-life spatial relationships between distant users. Ishii's ClearBoard employs the metaphor of speaking through a glass pane [9], and Buxton's Hydra situated video surrogates of remote users around a table to simulate real-life meetings.

Artists have projected video of human figures directly into the physical world to create powerful statements engaging the viewer in contemplative experiences. Naimark's Displacements [12] and Wodiczko's projections on public architecture [19] both play on the concept of human presence.

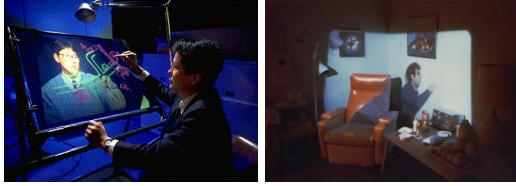


Figure 2: Examples of using video to convey presence: Clearboard (left) and Displacements (right)

2.2 Disembodied Representations

Another approach abandons the image in favor of animating objects in the world to suggest the presence of an invisible being. Objects may move as if under the direct touch of an invisible person [4], or they may appear to be infused with an invisible spirit [10]. While the movement of animated physical objects carry significant meaning [7], it is difficult to imagine identity of the ghostly presence from seeing the objects alone without explanation.

A classic example of disembodied presence is the player piano. Its modern incarnation, the Yamaha Disklavier, captures and re-renders a pianist's touch and tone in near perfect detail. A recent concert featured the late Marvin Hamlisch on a Disklavier posthumously performing a duet with the violinist Joshua Bell [2].

2.3 Video and the Player Piano

The Disklavier has been used in conjunction with video for both practical and artistic applications. The Sync-a-Vision by PianoDisc features a built-in screen at the music stand which plays video of performances synchronized with the moving keys [17]. Both remote lessons and concerts have employed the Disklavier to transmit playing over distance along with a video feed shown on a separate screen [24, 1].

Theatrical concerts have projected video of the live pianist onto the Disklavier's surroundings such as Sakamoto and Iwai's Parabola [18].



Figure 3: Death and the Powers (left). The disembodied main character is represented by movement of the chandelier and visuals on the walls. Parabola (right). A copy of the live pianist is projected on the wall next to the Disklavier.

3. DESIGN

Our work aims to combine the immediacy of the moving keys with the visual identity of the pianist such that the interface is perceived not as streams of synchronized information but as a coherent, engaging experience. Our design is inspired by the ghostly reflections on the surface of a lacquered grand piano. Guided by this metaphor, we discuss choices in framing, size and placement of video in relation to the piano.

We project life-sized video streams of the pianist's hands, face and upper torso onto surfaces of the piano. Video of the hands include the full keyboard recorded from above, which is projection mapped exactly onto the physical keyboard. The piano's glossy surface naturally mirrors the physical keyboard as well as the projection. Video of the face and upper torso with a black background is projected on the piano's music stand where a live pianist would be reflected. Though most of the virtual pianist is not explicitly shown, the synchronized hands, body and keys suggest the form and movement of the invisible.

We also designed an alternate variation where the virtual hands and keyboard are aligned but not coincident with the physical keys, building on a prior interface for remote lessons [22]. Although this interface departs from the reflection metaphor, it may offer advantages when used for pedagogy. We consider differences between the interface configurations in our evaluation.



Figure 4: Alternate Interface

4. PROTOTYPE

Our prototype builds on a Yamaha Disklavier baby grand piano, which communicates with the computer through a MIDI to USB cable. We installed a short-throw projector 5.5 feet overhead the piano bench, which covers both the keyboard and vertical surfaces without blockage from a person seated normally on the bench. For the upper body video, we constructed a custom display surface from plywood treated with projection screen paint, which fits over the existing music stand of the piano. For the alternative

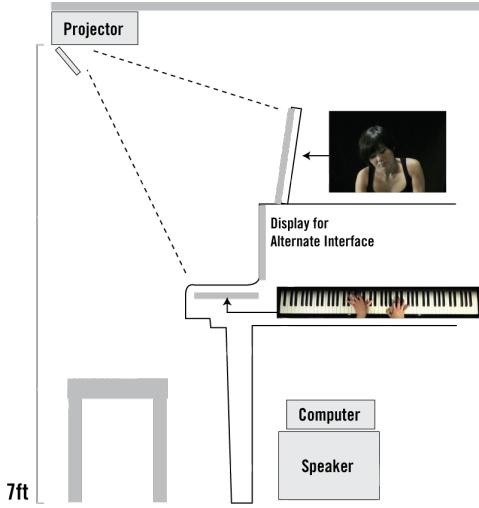


Figure 5: System Diagram

interface configuration, we constructed a similar display for the hands which replaced the vertical surface of the original keyboard cover.

Each recording for MirrorFugue III is captured on two cameras, a MIDI sequencer and a high quality audio recorder. We mounted one camera on a truss above the keyboard and another at the end of the piano to capture the pianist's face and upper torso. MIDI signal from the Disklavier is captured by Reaper on a computer. Both cameras and Reaper also capture audio, which is used to synchronize the streams when editing. The final product for each recording consists of the MIDI file and two video files rendered at 720p.

A custom Java program controls the entire interface. It composites the video streams, computes projection mappings in real time and synchronizes playback.

While a recording is playing on the piano, keys that are not moving function normally, and a user can layer live playing with the recording.

5. INTERACTION SCENARIOS

MirrorFugue III could be used to interact with content over distance and time as well as to visualize musical structure. We detail applications in each of these categories. Though applications themselves may not be novel, we articulate how our interfaces could enhance the user's experience.

5.1 Over Distance

As it is currently possible to transmit synchronized video, audio and MIDI signal over distance with a small fixed delay [24], MirrorFugue III could be used when connecting two pianos in real time over distance.

5.1.1 Concert

People can experience live concerts right in their living room with a view closer than a seat in the concert hall. Audiences gain an intimate view of the performer's facial expressions, and they can see how the hands are playing. Audiences can sit on the piano bench to experience the concert from the performer's point of view and can shadow the performer's hands on the keyboard.

5.1.2 Lesson

Remote lessons and master classes can also be conducted through MirrorFugue III. The Disklavier accompanied by a video feed has been used for remote lessons [1], but the

video has always been shown on a display separate from the piano. Our interface gives the student and teacher the illusion of a shared keyboard, as if they were sitting next to each other at a typical lesson. This makes it easier for the student to observe the teacher's demonstrations and for the teacher to watch the student's technique. Being able to view each other's faces also allows better communication. The alternate interface with vertical projection may be used to better see the hands, and another video feed can be added to view the wrists and arms.

5.1.3 Duet

MirrorFugue III can also be used for 2-way playing over distance. During ensemble playing, musicians are peripherally aware of each other's presence and use physical cues for synchronization. Distant pianists can feel each other's presence when they are playing together on our interface.

5.2 Across Time

MirrorFugue III is a new medium through which to interact with recordings from the past.

5.2.1 Archival References

People can watch recorded performances of great pianists on our interface and compare how different people play the same piece. In addition to listening to the sound, people can observe how different pianists channel different styles of playing.

5.2.2 Tutorial

How-to videos such as those on Youtube are popular among aspiring musicians. Tutorials on MirrorFugue III can demonstrate playing on the physical piano. They could also include exercises allowing the student to play along with the recorded teacher.

5.2.3 Self-Reflection

Pianists can also record their own playing on MirrorFugue III. These recordings can be a record of progress. The pianist can also record part of a piece and play the rest of it as a duet with their virtual self.

A recording on MirrorFugue III captures a moment in time. Replaying a recording momentarily offers a glimpse of that moment, creating the illusion of time-travel where one confronts selves from the past. A pianist can play a duet with a much younger version of the self. Grandchildren can play with versions of their grandparents at their age. Our interface can create intimate, reflective experiences on the passing of time.

5.3 Musical Structure

Much of music consists of multiple layers or lines [21]. Canons and fugues typical of baroque music are classic examples of multi-layered pieces for the keyboard. Music of other styles also feature multiple layers with different roles such as melody and bass line. In an ensemble, the timbre of different instruments help clarify the different lines, and in live performance lines are personified by different musicians. In solo piano music, every line is played on the same instrument by the same performer, and it can be difficult to fully appreciate the intricacies of the sound.

MirrorFugue III could be used to clarify the layers of a piano composition by personifying each voice with a virtual pianist. The same pianist could be duplicated multiple times, or a different person could represent each voice. Each virtual pianist can project a different character, adding depth to the performance especially when voices exhibit different characters (e.g. a stately bass line with a soaring melody).



Figure 6: Some scenarios: A student learns to play from a master pianist (left). A child plays a duet with herself (center). Chopin’s Opus 25. No. 7 Etude [5] expanded into tenor, harmony and soprano lines (right).

Expanding a performance into multiple layers can also serve as a useful practice tool, allowing the student to take over and play along one layer at a time. Practicing this way enables the student to experience the whole piece while working on it portions at a time. This may allow the student to gain a deeper understanding of how the piece fits together.

6. EVALUATION

The described scenarios are largely based on the hypothesis that spectators perceived our interface as an immersive whole and can viscerally feel the presence of the virtual pianist. We conducted a preliminary evaluation to investigate this hypothesis through user feedback.

Through the study we also wished to gain insight into user perception of what makes the interface viscerally engaging. Specifically, we wanted to understand the contribution of each visual channel (moving keys, projections of hands and body) and how varying their arrangement affects the experience.

We devised five interface configurations (see Table 1) to present to study participants based on three parameters:

- Music generated from player piano’s **moving keys** or as **sound only**
- **Projection** of pianist or **no projection**
- Projection of keys in **original** or **alternate** configuration

6.1 Method

A convenience sample of 15 subjects (8 male, 7 female; Aged 23-49(28)) with varying piano experience watched recordings of 5 different performances by the same pianist on the 5 interfaces. Each recording was approximately one minute long, and the order of interfaces and recordings was randomized. After watching all recordings, subjects ranked the interfaces from 1-5 with 1 being most preferred and completed a questionnaire on their preferences.

6.1.1 Statistical Analysis

Differences in ratings between interface configurations were evaluated using Friedman’s matched group analysis of variance test with the Nemenyi multiple comparison test.

6.2 Results

The average ranking of each interface configuration is summarized in Table 1. Our original interface (I) had the most preferred average ranking and was significantly more preferred than the configurations that lacked either the projection or moving keys (III, IV, V). The alternate interface (II) had the second most preferred average ranking and

was significantly more preferred than the two configurations without the moving keys. The alternate interface without moving keys (IV) had the worst preference ranking.

6.2.1 Immersion and Presence

Eleven of the 15 participants felt that the original interface enhanced the presence of the recorded pianist. They spoke of the interface’s immersive experience, and their language indicated their perception of the projection and moving keys as an integrated whole. Some participants described the player as “virtually in the piano”, and others felt “as if the player were sitting besides me playing for me”. A few even said that the interface “tricked my brain completely into thinking that I was playing.”

6.2.2 Alternate Interface Arrangement

The alternate configuration was four people’s top preference, and they tended to speak of the interface in terms of its separate information streams. Each had a different reason for preferring the hands projection not on the keyboard. One was an experienced pianist and liked how the vertical projection gave a clear view of the hands, explaining that “watching and learning are blurred for me as a former player.” Another experienced pianist had difficulty reconciling the 2D projection “playing” the 3D keys in the original interface but said that he could imagine preferring it once he “gets used to the dimensionality.” Others felt that the hands on the keys was distracting and preferred seeing them more peripherally.

When asked which interface they would use if they were to learn to play, 12 out of 15 chose some variant of the alternate configuration (II: 9, IV: 3). These participants felt that it was easier to follow the vertically projected hands because “it is not blocked by your hands.” Some described direct projection on keys as “too cluttered” and “distracting” for a tutorial interface.

6.2.3 Role of Moving Keys and Projection

When asked about the effect of the player piano’s moving keys, 12 of the 15 participants mentioned that it made the performance seem more real, live and present. For example one participant said that it “gives real energy to the music and made the performance seem more immediate and in the moment.”

Fourteen of the participants mentioned that the projection makes the performance more human and emotionally engaging. For example, one participant said that seeing the movement and expression of the performer was “very much like a traditional performance but directly in front of you.” Some people reported paying more attention to the “nuance of facial expressions that enhanced the character of the

Configurations	Original (I)	Alternate (II)	III	IV	V	
Moving keys	yes	yes	yes	no	no	
Hand projection	on keys	vertical	none	vertical	on keys	
Average rankings	1.4 ^{a, b, c}	2.3 ^{d, e}	3.3 ^a	4.3 ^{b, d}	3.7 ^{c, e}	p-value¹
Std. deviations	0.7	1.2	1.2	0.9	1.0	$p < 0.01$

¹Friedman’s test and Nemenyi follow up

Values with similar superscripts are significantly different.

Table 1: Participants ranked their preferences for configurations from 1 (most favorite) to 5 (least favorite).

piece”. Others paid more attention to the hand movements to better understand the “flow of rhythm and phrasing.”

Though almost all participants enjoyed seeing the moving keys and the projection, the interfaces with only projection or moving keys had the lowest preference averages. This suggests that the positive effects of the player piano and projection are enhanced when the two appear in combination. Some users explicitly clarified that they found the moving keys “ghostly and unnerving” on their own but really enjoyed it with projection. Others noted that “projection on the keys without the player piano was awkward.”

6.3 Study Summary

After one minute of watching a piano performance on our original interface, more than two-thirds of the study participants described feeling the impression of a present pianist. This result supports our core claim of our interface’s immersive, engaging qualities. Based on responses from study participants, we also summarize some observations on the perception of our interface.

Combining modalities creates richer experience-

The projection and moving keys each served a different role and complemented each other in evoking the pianist.

Configuration matters- Different arrangements of the same information streams yield vastly different experiences for the spectator. In our interface, the coincidence of projecting hands with moving keys created the impression of causality [11], which added to the illusion of a present virtual pianist. Though the alternate configuration featured the same video, its arrangement made it more suitable as a useful reference.

A convincing illusion does not require perfection-

When projecting directly on the piano keyboard, the image is only visible on the white keys. Despite not being able to see fingers on the black keys, people still saw the virtual hands as “playing” the physical piano. Similarly, though the lower torso, arms and wrists of the virtual pianist cannot be seen, people still saw the virtual pianist as being there.

7. FUTURE WORK

This paper presented a brief evaluation whose results suggested issues of perception that would like to further investigate in the future. We plan on conducting more quantitative studies on how users perceive our interface, backing our claims through collected data such as eye gaze and affective

measures. We also plan to evaluate our interface grounded in more specific scenarios.

We will continue to examine possible pedagogical uses for our interface suggested by this paper. Although study participants pointed out the potential of using the alternate configuration for learning to play pieces, we would like to pursue how our interface can go beyond a useful reference for learning notes [22]. Of particular interest is how the illusion of the physically present virtual pianist can help students learn certain advanced techniques, embody a range of characters and promote engagement during practice.

7.1 Weight Touch Technique

Two professional pianists who saw our interface commented that the original configuration can be useful for seeing the weight-touch technique of the recorded pianist. Weight touch is an advanced technique highly prized in classical playing where the pianist uses the weight of the arms to control tone [14]. We would like to study how the combination of projected hands and moving keys can help students understand and acquire this technique.

7.2 Embodying Characters

Musicians display vastly different body language when playing different styles of music. For example, they may sway slowly when playing a lyrical, expressive piece or dance to the beat when playing strong rhythms. We are interested in how students can learn to embody different characters in a variety of musical styles by mirroring the body language of the virtual pianist.

7.3 Longterm Engagement

An oft-lamented fact of music practice is that the musician is always alone. This is especially true for the pianist who has fewer opportunities for ensemble rehearsals than other instrumentalists. Consistent, convincing presence of a virtual pianist could engage students on a personal level and keep practice from being a dull, mechanical exercise. We would like to explore if and how our interface could promote this sort of longterm engagement and how frequency of use influences students to play more expressively in performances.

8. CONCLUSION

This paper explores how visual cues of the performer’s physical presence enhances the experience of a musical performance. Surveying a range of work in both research and the arts, we articulated strategies for representing human presence, either figuratively through video or through the

disembodied movements of physical objects. We then introduced MirrorFugue III, an interface for experiencing piano performances that combines video with the movement of a player piano's keys. We described seven categories of scenarios of how the interface can enrich interactions with recorded music, discussing its use over distance and time as well as its use to visualize musical structure. In an evaluation, we showed that most spectators can feel the presence of the virtual pianist, perceiving the multiple information streams as an integrated experience.

MirrorFugue III reconsiders the musical performance not just as abstract ideas from the composer rendered accurately in time but also as the expressive gesture of a performer who lives, breathes and embodies the music. Whether the performer is live or conjured on our interface, the music becomes a visceral communicative experience between the performer and the audience, an experience that is deeply and fundamentally human.

9. ACKNOWLEDGMENTS

We are grateful to Yamaha and Hitachi America for their generous donations that enabled this research. Many thanks to the Tangible Media Group and the MIT Media Lab for their continual support.

10. REFERENCES

- [1] Musicpath: Networking people and music.
<http://musicpath.acadiau.ca/main.htm>.
- [2] New york philharmonic's new year's eve concert.
<http://www.yamaha.com/yasi/newsdetail.html?CNTID=5816177&ref>.
- [3] A. Barbosa. Displaced soundscapes: A survey of network systems for music and sonic art creation. *Leonardo Music Journal*, 13:53–59, 2003.
- [4] S. Brave, H. Ishii, and A. Dahley. Tangible interfaces for remote collaboration and communication. In *Proc. CSCW '98*, pages 169–178. ACM, 1998.
- [5] F. Chopin. *Etudes*. G. Henle Verlag, Munich, Germany, 2006.
- [6] S. Dahl, F. Bevilacqua, R. Bresin, M. Clayton, L. Leante, I. Poggi, and N. Rasamimanana. Gestures in performance. In M. Leman and R. I. Godøy, editors, *Musical Gestures. Sound, Movement, and meaning*, pages 36–68. Routledge, 2009.
- [7] T. Djajadiningrat, B. Matthews, and M. Stienstra. Easy doesn't do it: skill and expression in tangible aesthetics. *Personal Ubiquitous Comput.*, 11(8):657–676, Dec. 2007.
- [8] Y. Ichikawa, K.-i. Okada, G. Jeong, S. Tanaka, and Y. Matsushita. Majic videoconferencing system: experiments, evaluation and improvement. In *Proc. ECSCW '95*, pages 279–292. Kluwer Academic Publishers, 1995.
- [9] H. Ishii, M. Kobayashi, and J. Grudin. Integration of inter-personal space and shared workspace: Clearboard design and experiments. In *Proc. CSCW '92*, pages 33–42. ACM, 1992.
- [10] E. Jessop, P. Torpey, and B. Bloomberg. Music and technology in death and the powers. In A. R. Jensenius, A. Tveit, R. I. Godøy, and D. Overholt, editors, *Proc. NIME '11*, pages 349–354, Oslo, Norway, 2011.
- [11] A. Michotte. *The perception of causality*. Basic Books, Oxford, England, 1963.
- [12] M. Naimark. Displacements. <http://www.naimark.net/projects/displacements.html>.
- [13] D. T. Nguyen and J. Canny. More than face-to-face: empathy effects of video framing. In *Proc. CHI '09*, pages 423–432. ACM, 2009.
- [14] O. Ortmann. *The Physiological Mechanics of Piano Technique*. Dutton, New York, NY, USA, 1962.
- [15] T. Page, editor. *The Glenn Gould Reader*. Alfred A. Knopf, New York, 1984.
- [16] R. Philip. *Performing Music in the Age of Recording*. Yale University Press, New Haven, CT, USA, 2004.
- [17] PianoDisc. Sync-a-vision.
<http://www.pianodisc.com/>.
- [18] R. Sakamoto and T. Iwai. Music plays images x images play music, 1997.
http://www.aec.at/fleshfactor/art_events.html.
- [19] K. Shulman. Art; a monument to mothers and lost children. *The New York Times*, September 1998.
<http://www.nytimes.com/1998/09/20/arts/art-a-monument-to-mothers-and-lost-children.html>.
- [20] S. S. Snibbe and H. S. Raffle. Social immersive media: pursuing best practices for multi-user interactive camera/projector exhibits. In *Proc. CHI '09*, CHI '09, pages 1447–1456. ACM, 2009.
- [21] D. F. Tovey. *The Forms of Music*. Meridian Books, New York, NY, USA, 1956.
- [22] X. Xiao and H. Ishii. Mirrorfugue: communicating hand gesture in remote piano collaboration. In *Prof. TEI '11*, pages 13–20. ACM, 2011.
- [23] Yamaha. Disklavier.
<http://www.yamaha.com/disklavieroffer/>.
- [24] Yamaha. Remotelive. <http://www.yamaha.com/usa/remotelive/home.html>.