Musical Piano Performance by the ACT Hand

Ada Zhang[1,2], Mark Malhotra[3], Yoky Matsuoka[3]

Department of Bioengineering[1], Department of Music[2], Department of Computer Science and Engineering[3]
University of Washington, Seattle, WA, USA

Abstract— In the past, the music community conducted research on what makes music more musical or expressive. Much of this work has focused on the manipulation of phrasing, articulation and rubato to make music more expressive. However, it has been difficult to study neuromuscular control used by experts to create such musical music. This paper took a first step toward this effort by using the Anatomically Correct Testbed (ACT) Robotic Hand to mimic the way expert humans play when they are instructed to perform "musically" or "robotically." Results from 22 human subjects showed that musical expression contained a larger range of dynamics and different articulation than robotic expression, while there was no difference in the use of rubato. The ACT Hand was controlled to the level of precision that allowed the replication of expert expressive performance. Its performance was then rated by 17 human listeners against music played by a human expert to show that the ACT Hand could play as musically as an expert human. Furthermore, articulation, phrasing, and rubato were tested in isolation to determine the importance of articulation over phrasing and rubato. This type of study will lead to understanding how to implement future robots to perform musically without preprogramming them, finding ways to teach novice pianists strategies in controlling their muscles to become expressive musicians more quickly, and understanding why humans feel expressiveness or even emotion in music.

I. INTRODUCTION

Playing piano musically is a highly dexterous skill that humans spend years developing. Many factors contribute to musicality, including *phrasing* – deviations in dynamic level (note volume), *articulation* – attack and duration of a note, and *rubato* – small deviations in tempo (speed of performance). Precise fingertip velocity control is required to subtly control dynamics on the piano and thereby perform musical phrasing. Control of the timing of key-strikes and releases and coordination between fingers is necessary to perform rubato and produce different types of articulation.

The Anatomically Correct Testbed (ACT) Hand is a biomimetic robotic hand (Fig. 1) [9], [2]. By mimicking the structure and biomechanics of the human hand, it allows physical simulation of muscle movements. It also provides a testbed for studying neuromuscular control of the hand. We are interested in how pianists control their fingers to perform "musically" as well as how to implement this control in the ACT Hand. The robot can then serve as a tool to understand (1) what kind of muscle control is used by experts but not by novices and (2) whether there are better training paradigms to help novices reach an expert musical level.

In the 1980s, the WABOT-2 robot, which was capable of playing the organ, was developed by Waseda University [5]. This anthropomorphic robot used fingers to depress the keys
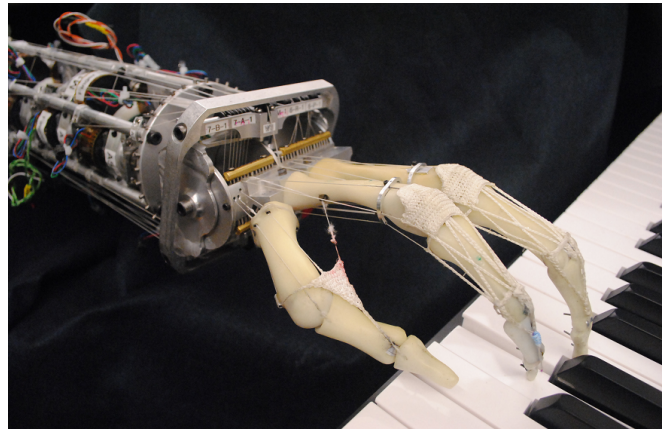

Fig. 1. The Anatomically Correct Testbed Hand

and legs to depress the pedals. It was capable of reading a score, conversing with and accompanying a singer. This robot is certainly remarkable in its ability to perform many of the tasks related to playing organ. Because the ACT Hand is not customized to this task, but matches human anatomy in flexibility and complexity, it allows us to explore the subtleties of control that are required for humans to play the piano musically.

Gabrielsson [3] and Palmer [7] provide two large-scale reviews of music performance research. Many have explored the concept of musicality and expressivity in music by examining performances by expert pianists [6], [4] and developing computational models [1], [10]. The difficulty lies in the large number of parameters that affect musicality and the variability within those parameters. Therefore, we have limited our scope to musical phrasing, articulation and rubato. Once the ACT Hand is capable of performing like an expert pianist, we can implement the models mentioned above, allowing it to play any piece musically when only given the score.

This paper explores how closely the ACT Hand can match performances by expert human pianists. Section II shows experiments conducted with expert, intermediate and novice pianists to compare the differences in phrasing, articulation and rubato when they are asked to play a piece musically versus robotically. Section III discusses the ACT Hand hardware and the control strategy, which allows it to perform like an expert pianist. Finally, Section IV compares human ratings of different performances by the ACT Hand and by a human as a kind of Turing Test. We intend this work to play
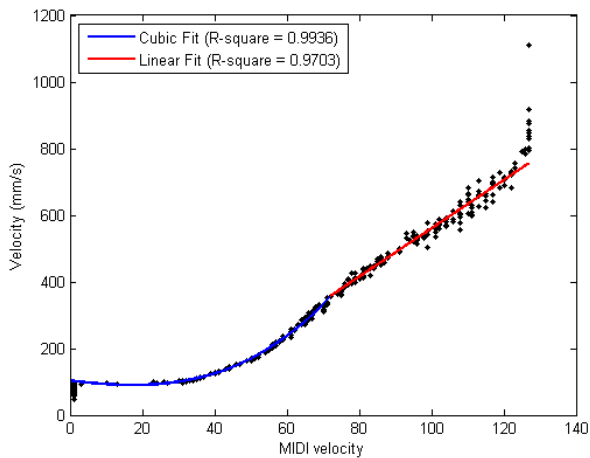
Fig. 2.   Calibration curve for MIDI velocity to mm/s

Fig. 3.   Example keystroke velocities for an expert playing Schumann

a first step toward understanding neural control of human expression.

II. HUMAN PERFORMANCE

A. Setup

1) MIDI keyboard: Human trials of musical performance were conducted on a Korg SP-170 digital keyboard. This keyboard had natural weighted hammer action keys, realistically mimicking the feel of an acoustic piano. Data was collected in MIDI format, which contains the time stamp, command type (note on or off, e.g.), note identifying number, and keystroke velocity for each command.

MIDI keystroke velocities are average velocities that are mapped to a value between 1 and 127. In order to calibrate these values to SI units, a PHANTOM Premium Haptic Device from SensAble Techologies was used. It was coupled with a key on the piano and recorded position values as the key was pressed at varying velocities. Average velocities were computed for each keystroke, plotted against the MIDI velocity, and fitted with a multi-part function consisting of a cubic function for MIDI velocities ≤ 72 and a linear function for those > 72 (Fig. 2). MIDI values of 1 and 127 were excluded in the fits since these were the limits of MIDI keystroke velocities. This mapping allows the keyboard to be used as a velocity sensor (Fig. 3).

In order to compute the average velocity of a keystroke, MIDI begins a timer when the key that the user is striking reaches some displacement x_0 from the rest position of the key. MIDI stops the timer once the user has reached the key bed x_f of the piano, divides this defined distance by the time, and maps it to a value between 1 and 127. Due to the acceleration inherent in the beginning of a keystroke, calibration required that we first determine the displacement x_0 from the rest position of the key at which MIDI begins its timer. It was found that $x_0 = 3.25$mm since this resulted in the least amount of spread in the data.
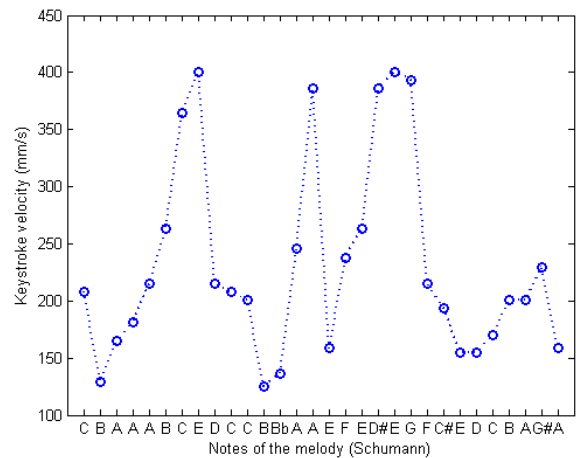
2) Experimental Protocol: Twenty-two pianists of varying levels participated in the experiment – 7 expert, 10 intermediate, 5 novice. To determine the experience level, the pianists were shown measures 12-19 of the Schumann *Piano Concerto in A minor* (Op. 54) and asked how long it would take them to learn it. The three answer choices were "can sightread the music," "can learn within three hours," and "would consider this piece quite difficult to learn" and the pianists thus self-selected into the "expert," "intermediate," or "novice" category.

The pianists were asked to perform two different melodies: "Mary Had a Little Lamb" (see Fig. 4) and the melody from the aforementioned Schumann excerpt (subsequently referred to as "Schumann"). They were asked to perform each melody a total of six times, three times as "musically" as possible and three times as "robotically" as possible. The definition of "musically" and "robotically" was left up to the pianists' interpretation. The purpose for these two performance styles was to help determine the importance of various parameters for creating musical performances. Novice pianists, in general, found Schumann too difficult to play. Therefore, this melody was not analyzed for novices.

Four hypotheses were made concerning differences between musical and robotic performance:

1) The range of keystroke velocities would be larger for musical versus robotic performance.
2) Experts would use more consistent keystroke velocities than less experienced pianists.
3) Musical performances would demonstrate use of *legato* (smoothly connected) articulation while robotic performances would use detached articulation.
4) Rubato would be used for musical performances, resulting in an inconsistency in inter-onset intervals (IOI) between musical and robotic performances.

B. Results

1) Range of Keystroke Velocities: For each performance, data was collected on the difference between the highest
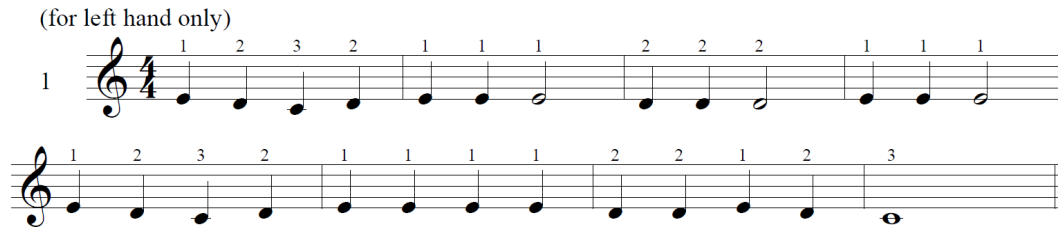
Fig. 4. Score of "Mary Had a Little Lamb" given to pianists

TABLE I

MEAN KEYSTROKE VELOCITY DIFFERENCES (MM/S) AND P-VALUE FOR A T-TEST ON THE DIFFERENCE IN MEANS

	Schumann			Mary Had a Little Lamb		
	Musical	Robotic	p-value	Musical	Robotic	p-value
Expert	308.1	239.6	0.0019	177.6	123.2	0.0043
Intermediate	261.1	183.1	0.0033	148.4	96.8	0.0021
Beginner	N/A	N/A	N/A	132.7	125.1	0.7090

TABLE II

MEAN AND MEDIAN OF AVERAGE CV (%) (FOR SETS OF THREE PERFORMANCES)

	Schumann		Mary Had a Little Lamb	
	Mean	Median	Mean	Median
Expert	18.23	18.99	26.74	24.99
Intermediate	27.67	27.42	22.13	20.04
Beginner	N/A	N/A	26.46	29.21

TABLE III

AVERAGE NUMBER OF LEGATO NOTES WITHIN A SINGLE PERFORMANCE AND P-VALUE FOR A T-TEST ON THE DIFFERENCE IN MEANS

	Schumann (30 notes total)			Mary Had a Little Lamb (26 notes total)		
	Musical	Robotic	p-value	Musical	Robotic	p-value
Expert	14.85	9.619	0.0017	12.05	5.714	0.000019
Intermediate	8.480	8.565	0.73	5.962	1.160	0.00020
Beginner	N/A	N/A	N/A	7.833	0.3333	0.000011

and lowest keystroke velocity. These values were averaged for robotic and musical performances from each group of pianists (expert, intermediate and novice). A t-test was then performed to determine if the means were significantly different. As shown in Table I, ranges in keystroke velocity differences were larger for musical versus robotic performances, and the averages were statistically different for nearly all performance groups. Therefore, the ACT Hand should be capable of keystroke velocities ranging from ~125-425 mm/s (Fig. 3) in order to perform musically.

2) Consistency of Keystroke Velocities: For the three musical performances by each of the pianists, a coefficient of variance (CV) was taken for the mean keystroke velocity of each note in melody. The CV values were then averaged to produce an average CV for each pianist. The mean and median of these average CV values for each group of pianists is shown in Table II. These results give the level of precision required for the ACT Hand, which is a CV level no greater than 18.23%.

3) Articulation: The number of legato notes within each performance were calculated and then averaged across all robotic and all musical performances within a pianist group.

The means were then compared with a t-test to determine if they were significantly different. As shown in Table III, there were more legato notes in musical versus robotic performances, and all but one of the differences in means were statistically significant. Therefore, in order to perform piano musically, the ACT Hand must be capable of legato and detached articulation.

4) Consistency of IOI: The IOI was defined as the time between adjacent "note on" commands, calculated as a percentage of the total time of the performance. The IOIs of each note were averaged over all musical performances and over all robotic performances of a given pianist group. The set of IOIs for these average performances (musical and robotic) were then correlated to each other, as shown in Table IV.

The correlation coefficients were all very high, implying that pianists do not consider IOI variation to be important for creating musical versus robotic performances of these two melodies. Fig. 5(a) shows the average IOI values for expert musical and robotic performances of Mary Had a Little Lamb, and the two performances indeed seem very similar. However, it seems that experts played the quarter notes (IOI

TABLE IV
IOI CORRELATION COEFFICIENT OF ROBOTIC TO MUSICAL PERFORMANCES

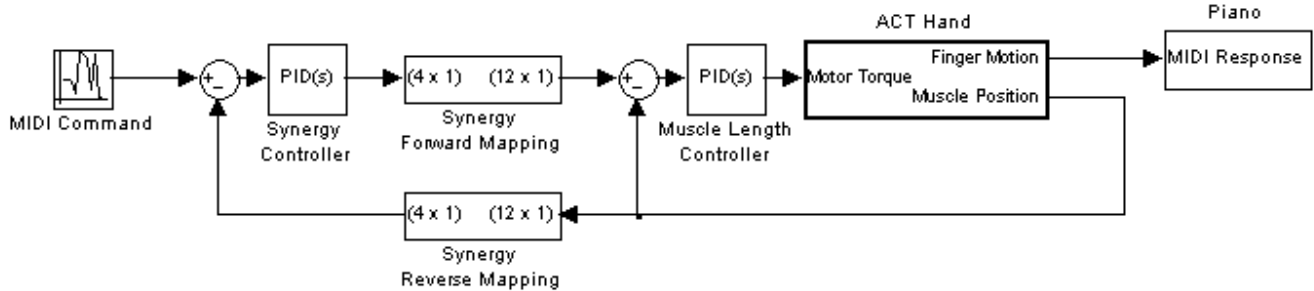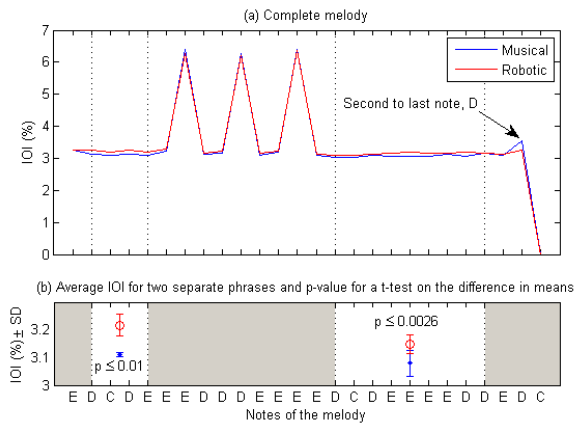|	Schumann	Mary Had a Little Lamb
Expert	0.9968	0.9978
Intermediate	0.9963	0.9958
Beginner	N/A	0.9982

Fig. 6.   Controller Diagram

Fig. 5.   Average expert IOI for musical and robotic performances of Mary Had a Little Lamb

$\approx 3\%$) slightly faster in musical performances compared to robotic performances. That is, they slightly rushed this part of the melody. This is highlighted in Fig. 5(b), where the average IOI for subsets of the quarter notes are plotted. Note that the p-value for a t-test of the difference of means between robotic and musical performances shows that the means are different with 95% confidence. Furthermore, experts played the second-to-last note (D) slightly longer in the musical performances, creating a slight *ritardando*. The p-value for a t-test on the difference of means between robotic and musical performances of this note was $p \leq 0.00043$, again showing that the means are different with 95% confidence. Although the IOI values of the complete robotic and musical performances were highly correlated, certain subsets of these performances showed significant differences in musical versus robotic performance. This implied that subtlety in keystroke timing would be important for musical performances by the ACT Hand.

III. ACT HAND PERFORMANCE

A. Setup

1) ACT Hand Hardware: The ACT Hand is designed to mimic the biomechanics of the human hand and to provide control at the speed and strength required for most dexterous tasks. In this paper, we consider control of the index and middle fingers. Six motor-driven tendons are merged into a tendon hood that terminates in anatomically correct locations on the finger bones and is similar in shape to a human tendon hood. The motors are equipped with high-resolution encoders that allow precise control of tendon excursion at 200 Hz.

2) ACT Hand Controller: Synergies are a proposed explanation for neuromuscular control's capability to coordinate activations of muscles that are highly coupled and redundant. For a specific task like playing piano, the small subspace of muscle activations required may be expressed as a lower-dimensional linear combination of muscle activations. Tendon excursions are recorded during a human-guided demonstration of each the index finger and the middle finger pushing down a key on the piano, as well as centering over the key. Principal Component Analysis (PCA) is then used to determine the basis vectors of tendon excursions that best capture movement. These synergies act as a proxy for both functional measurement and control, such that moving along a specific synergy can produce a piano key strike or centering over the key, just as shown by human example. Fig. 6 shows the hierarchical control scheme used to track a trajectory as specified by MIDI. More details on this type of controller can be found in Rombokas, et al. [8].

A PID controller is tuned for each synergy to provide speed of response, precision, and steady-state error rejection as needed. Some problems encountered while tuning the PID controller were instability, delay, overshoot, and slow key-strike velocity. To resolve these problems, the PID controller was carefully tuned with different gains during the two main phases of motion – strike and release – to optimize each

Keystroke Velocity (mm)		Keynote Length (s)		IOI (s)	
Mean	CV (%)	Mean	CV (%)	Mean	CV (%)
162.8452	8.71	1.7034	2.93	5.0013	0.07
148.5636	10.30	1.7488	2.36	5.0027	0.18
138.9577	10.51	1.7305	3.33	5.0031	0.07

phase's relevant performance measures. For example, during keystrike, striking velocity is more important than accurately meeting the set-point position. However, while preparing to strike, maintaining a certain height above the key is more important. With this PID controller, the ACT Hand remained stable and could track high velocity command signals accurately: for example, when the synergy velocity command was to move from 0.15 to 0.6 (synergy space units) in 0.0173s the ACT Hand moved from 0.15 to 0.6 in 0.0165s.

B. Results

Table V gives mean and coefficient of variation values for the keystroke velocity, keynote length and IOI of three performances by the ACT Hand. The CV values for the keystroke velocities (Table II) are all lower than that of experts (18.23%). The CV values for keynote length and IOI were all small, implying high consistency for the timing of striking and releasing a key. Thus, the control strategy for the ACT Hand seems capable of producing musical phrasing, articulation and rubato of expert-like quality.

IV. TURING TEST

A. Setup

Phrasing, articulation and rubato can be independently employed, since phrasing concerns only the keystroke velocities, articulation concerns only the timing of the note off command relative to the next note on command, and rubato concerns only the timing of the note on commands relative to that which is dictated by the rhythm. Therefore, these three parameters are independent of each other, allowing eight different MIDI files using the "Mary Had a Little Lamb" (see Fig. 4) melody to be generated:

1) Non-musical (no phrasing, articulation, or rubato)
2) Only phrasing
3) Only articulation
4) Only rubato
5) Phrasing and articulation
6) Phrasing and rubato
7) Articulation and rubato
8) Musical (phrasing, articulation, and rubato)

Finger command trajectories were then created from these MIDI files, resulting in eight different performances by the ACT Hand. Since control of only the index and middle fingers were considered, the middle finger substituted for the thumb as well. The notes played by the middle finger (C) were later changed to the correct note (E), however, the note on/off command timings and the keystroke velocity were left unchanged. Therefore, changing the pitch of the note did not alter its phrasing, articulation, or rubato.

Described below are the definitions – based on the musical taste of the experimenters – for the use and lack of phrasing, articulation and rubato in the MIDI files from which the ACT Hand command trajectories were created. These definitions were influenced by the results described in Section II-B. The tempo for all MIDI files was set at 80 beats per minute (bpm). The authors would like to stress that the purpose of this work was not to develop an algorithm for musicality, but rather to explore the neuromuscular control required of pianists to perform musically. The aim of the Turing Test is not to demonstrate the musicality of the ACT Hand, but rather to demonstrate the finesse of control that we have achieved with the ACT Hand's finger.

1) Phrasing: Lack of phrasing was defined using a MIDI velocity of 53 (187 mm/s) for each of the notes. To create musical phrasing, the keystroke velocities for each note were made similar to those of an expert pianists' performance taken from the recordings made in Section II.

2) Articulation: Lack of articulation was defined by making each note off command occur approximately one eighth note (0.375s at 80 bpm) before the next note on command. This produced a detached articulation. Musical articulation was defined by making each note off command occur approximately one sixteenth note (0.1875s) after the next note on command (excluding repeated notes, which had the note off command occur one sixteenth note before the next note on command). This produced a largely legato articulation. Therefore, ACT Hand performances using musical articulation employed legato as opposed to detached articulation, reflecting the results found in Section II-B.3 and shown in Table III

3) Rubato: Lack of rubato was defined by making each note on command occur exactly as specified by the rhythm dictated by the score. Musical rubato was created by slightly rushing and dragging certain portions of the melody, as discussed in Section II-B.4 and shown in Fig 5.

The eight ACT Hand performances, along with two expert human performances from those collected for Section II (one musical and one non-musical) were used in a survey. They were given to the participants in triplicate, resulting in a total of 30 questions, and presented in a randomized order. The 17 participants, none of whom were musicians, rated each one for its level of musicality on a scale of one to ten.

B. Results

Fig. 7 shows the results from the survey. A two-factor ANOVA was used to assess the effects of performer (human versus ACT Hand) and style (musical versus non-musical), followed by a *post hoc*, paired Tukey simultaneous test to
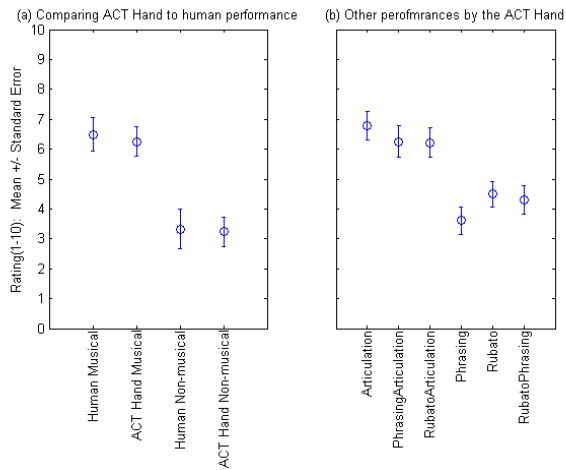
Fig. 7. Average rated musicality from 17 participants

further assess the difference between the musical and non-musical performance ratings. Analyses were completed using Minitab Statistical Software (Minitab Inc., State college, PA).

The two-factor ANOVA showed a statistically significant effect of style (musical versus non-musical) on the musicality rating of the 17 participants ($p < 0.001$), but no effect of the performer (human versus ACT Hand) on the rating ($p = 0.716$). The *post hoc* Tukey simultaneous test showed that the ratings for the musical performances were significantly higher ($p < 0.001$) than the non-musical performances. The effect of the interaction between style and performer was not significant ($p = 0.881$). These results imply that the musical and non-musical performances were significantly different, but that the survey participants could not distinguish between the musicality of the human and ACT Hand performances.

Furthermore, the more musical performances all used musical articulation, implying that legato articulation is highly important for a musical performance of Mary Had a Little Lamb. Note that articulation was the least subtle factor among the three factors that were used. It is interesting that the performance using only phrasing was rated as one of the least musical performances. This may be due to a difficulty for non-musicians to perceive phrasing when isolated from other musical factors. One should also remark that these results are specific to the melody used in the survey.

V. CONCLUSIONS

Our measurements of 22 human subjects show that musical piano performance requires a high level of keystroke velocity control, timing, and coordination between fingers. Adopting a synergy control scheme for the ACT Hand enables such quality of control: its keystroke velocity is more consistent than that of the expert pianists, and it is capable of highly precise finger strike and release timing. When phrasing, articulation, and rubato are defined, the ACT Hand can perform as musically or as robotically as an expert pianist.

These initial results have set down a promising foundation for exploring further into the neuromuscular control required for musical piano performance. Future work will incorporate wrist and arm movement into the ACT Hand's piano performance, thus increasing its musical capabilities. It is likely that this work will allow robots to perform musically without prior teaching and expand the understanding of expressivity in music. Deeper knowledge of expert pianist muscle control schemes may lead to improved pedagogical methods for the development of novice pianists' musicality.

VI. ACKNOWLEDGMENT

REFERENCES

[1] R. Bresin and A. Friberg, "Emotional coloring of computer-controlled music performances," *Computer Music Journal*, vol. 24, no. 4, pp. 44–63, 2000.

[2] L. Chang and Y. Matsuoka, "A kinematic thumb model for the ACT hand," in *Proceedings IEEE International Conference on Robotics and Automation*, May 2006, pp. 1000 –1005.

[3] A. Gabrielsson, "Music performance research at the millennium," *Psychology of Music*, vol. 31, no. 3, pp. 221–272, 2003.

[4] C. M. Johnson, "Effect of adding interpretive elements to a musical performance on the rhythmic and dynamic variations," in *Bulletin of the Council for Research in Music Education*, no. 147. University of Illinois Press, 2000, pp. 91–96.

[5] I. Kato, S. Ohteru, K. Shirai, S. Narita, S. Sugano, T. Matsushima, T. Kobayashi, and E. Fujisawa, "The robot musician 'WABOT-2' (WAseda roBOT-2)," *Robotics*, vol. 3, pp. 143–155, 1987.

[6] J. Langner and W. Goebl, "Visualizing expressive performance in tempo-loudness space," *Computer Music Journal*, vol. 27, no. 4, pp. 69–83, 2003.

[7] C. Palmer, "Music performance," *Annnual Review of Psychology*, vol. 48, pp. 115–138, 1997.

[8] E. Rombokas, M. Malhotra, and Y. Matsuoka, "Task-specific demonstration and practiced synergies for writing with the act hand," in *Proceedings of the IEEE International Conference on Robotics & Automation*, 2011 (To appear).

[9] M. V. Weghe, M. Rogers, M. Weissert, and Y. Matsuoka, "The ACT Hand: Design of the skeletal structure," in *Proceedings of the IEEE International Conference on Robotics & Automation*, 2004.

[10] G. Widmer and W. Goebl, "Computational models of expressive music performance: The state of the art," *Journal of New Music Research*, vol. 33, no. 3, pp. 203–216, 2004.