Infrared vs. Ultrasonic Finger Detection on a Virtual Piano Keyboard

Yuri De Pra
Fausto Spoto
Università di Verona
Department of Computer Science
Strada le Grazie 15
Verona 37134, Italy
yuridepra@libero.it

Federico Fontana
Università di Udine
Department of Mathematics
and Computer Science
via delle Scienze 206
Udine 33100, Italy
federico.fontana@uniud.it

Linmi Tao
Tsinghua University
Department of Computer Science
and Technology
Beijing 100084, P.R. China
linmi@tsinghua.edu.cn

ABSTRACT

An immaterial digital keyboard is presented, aiming at testing possibilities to substitute physical with augmented piano keys during the performance. The main effort has been made around the realization of an accurate meanwhile fast detection of the hands movement. To achieve this goal we have tested low-cost infrared as well as ultrasonic capture devices, whose current pros and cons are presented in either cases. Multimodal feedback has been realized by filming the hands' action with the rear camera of a consumer's tablet PC, and then projecting this action on its screen; furthermore this projection has been layered over the image of a piano keyboard reacting to the hands' action. Especially in connection with a Leap Motion system in charge of doing the infrared-based detection, and other light hardware for the sonic and vibrotactile rendering of the feedback, the proposed prototype promises potential application as an inexpensive mobile music interface transforming a normal table in an augmented reality scenario, where a pianist can perform simple musical tasks by relying on reasonably accurate and realistic feedback even in absence of a hardware keyboard.

1. INTRODUCTION

Recent virtual musical instrument interfaces try to make use of multiple sensory channels to add consistency to the performing experience over a simulated instrument. Indeed, guaranteeing the perceptual unity of the musical feedback is not obvious once multimodality is targeted: specifically, the temporal window in which the visual, auditory and somatosensory cues are allowed to stay in order to provide perceptual coherency of a single event is tightly constrained. Experimental studies of applied perception exist which have concluded, for instance, that a temporal window of 100 ms can report multimodal feedback about a single event if the respective cues are presented in the following order: tactile, auditory, visual [1, 2, 3, 4].

On the other hand, not all the input and output devices that can be found in the market are ready to meet the needed temporal constraints, irrespectively of their cost as well as detection or actuation accuracy. In the case of portable keyboards, reasonably accurate piano reproductions capturing the gestures of the fingers with sufficiently low latency still need to resort to hardware peripherals based on dedicated USB-MIDI protocols. In this sense, all the recent efforts made so far to get rid of the physical interface have resulted in mobile prototypes that yet do not reach the goal of providing sufficiently realistic rendering of an immaterial keyboard.

In spite of such difficulties, which currently limit the diffusion of realistic piano simulators based on light mobile hardware devices such as tablet PC's and smartphones, quite promising works have been proposed pioneering the disappearance of the piano as a solid-body instrument, such as Mike Heavers' Air Piano, capturing on-air hand gestures through fast camera tracking [5]. In parallel, significant research has been conducted supporting the software design leading to these works [6, 7], and their applicability to educational contexts [8, 9].

In this scenario, augmented reality provides an ideal conceptual framework since the designer can choose to reproduce images, vibrations and sounds by superimposing synthetic information over existing physical components, best if they are found among objects and furniture that are already present in one's everyday setting: for instance, we can think of drawing piano keys over a table and then play them through an Android smartphone like in (Augmented) Piano by Amit Ishai & Moshe Liran Gannon [1], meanwhile attaching vibrotactile exciters that make the own table vibrate and thus emit sounds; alternatively we can visualize interactive portions of a piano keyboard over a mobile touch screen, allowing a performer to press keys so to generate corresponding notes: this paradigm has led to more than 250 virtual piano realizations, including Piano 3D (mobileagency.com.au), Real Piano 3D (imudra tech.), Play Piano (android technologies), Learn Piano, My Piano and Grand Piano, to cite some.

Touch screens have excellent spatial resolution, furthermore they can transmit vibrations to the fingers in response to a touch event. On the other hand finger contact detection is affected by initial latency, which increases along the hardware/software path until becoming perceivable at the visual, auditory and vibratory actuation; moreover, touch

[1] https://sites.google.com/site/pianoreality/

screens are not sensitive to variable force. In an effort to overcome the variable force detection problem we moved the input interface elsewhere: more precisely, below the screen.

The prototyped concept is depicted in Figure 1: it consists

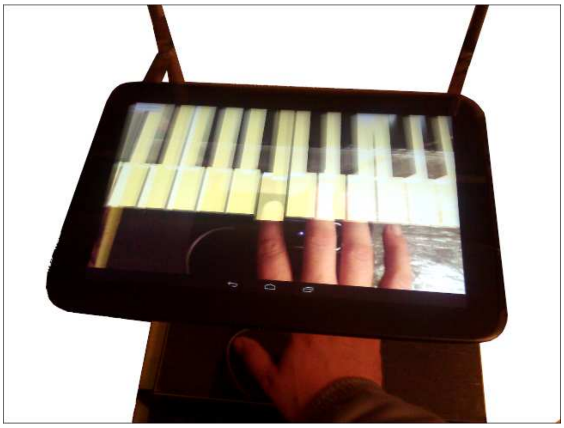

Figure 1. On-screen projection of the augmented piano keyboard in action.

of a tablet showing the performer's hands on the screen, and furthermore augmenting the playing scenario by an image of a piano keyboard actuated by their action, appearing on the same screen. Moving images of the hands are captured using the web camera behind the tablet; in parallel, the piano keyboard is superimposed as a semi-transparent image layer. Sound, as we will see later, is generated depending on the moving speed of the fingers. Currently we are using a Google Nexus 10: a ten-inch tablet operated by Android 4.4.

The research work on this concept does not aim at challenging the sensations coming from playing a real piano. Rather, it is centered around the search for a reasonably acceptable, well-balanced multimodal experience for the performer. In fact, the absence of a real keyboard does not prevent from prompt somatosensory feedback as the performer taps, for instance, on a rigid table; however, the precise but inherently poor haptic experience needs to be supported, and possibly improved by precisely space-aligned, accurately synchronized and musically realistic synthetic cues.

To date the sound synthesis problem has largely been solved, as the generation of convincing real-time sounds is made possible by the existence of fast and accurate sample-based or physically-informed sound models. Among the various sonic reproduction techniques, these sounds can be radiated via proper vibrotactile exciters located directly where the performer taps (e.g., the table): this solution provides an auditory impression similar to staying in front of the piano soundboard.

Concerning the display, relatively small delays affecting the visual flow from the screen become tolerable once the sequence tactile–auditory–visual is respected. In our system, however, subjective experiments should be made to quantify the perceptual unity of the resulting multimodal feedback [4].

In this paper we report results of our design and realization activity, aiming at achieving coherency of the multimodal feedback from the interface of Figure 1 and, ultimately, an overall feeling of natural interaction with the system. This activity so far has focused on the detection part of the system: once integrated with a powerful sound synthesis engine it has potential to lead to a portable, low cost augmented keyboard of reasonable quality supporting the activity of composers, early students and practitioners once having their mobile device at hand.

2. HANDS' ACTIVITY DETECTION

Web cameras aboard current mobile devices perform accurate meanwhile too slow visual captures, especially if the auto-focus function is implemented by the camera and cannot be disabled. As explained in the introduction, the latency of the video may be tolerable for the pianist but certainly prevents from using these captured data for triggering the sonic and vibrational feedback [10]. Hence, alternative detection strategies must be investigated.

2.1 Infrared detection system

Recent infra-red visual tracking systems, such as the Microsoft Kinect™ and Leap Motion™ cameras, have gained significant success as tools capable of detecting the moving speed of recognized objects. Concerning their specific use for finger tracking, Silva et al. found that the performances of the Leap Motion system were not impressive as expected, nor was the hand articulation detection algorithm particularly robust [11]. This behavior was confirmed in our study, as also reported in the last part of this paper.

The Leap Motion controller is specifically designed for hand and finger tracking. It comes with native support to SDK and several programming languages (Python, C++, Java, etc.) and promises to identify hand and finger motions with sub-millimeter precision at high frame rate (up to 300 fps), thus creating the conditions for the recognition of a number of gestures.

We have developed a software interface on an Intel Core i3 laptop PC that receives the data from the Leap Motion driver, and then sends the spatial coordinates of the fingers to an Android application running on the tablet. This application can also use such coordinates to visualize markers that follow the fingers along their movement (see Figure 2).

Due to the characteristics of its recognition system, the Leap Motion must stay in front of the palm or back of the hands. For this reason, in our system configuration the performer taps over a transparent surface that is positioned above the camera, which in its turn is oriented toward the ceiling. Recognizing the vertical movement of the fingers, and hence their impact velocity against the keys, from stereoscopic cues of depth is not particularly efficient.

We made use of the gesture "key pressed" provided by the SDK; then, by reading the finger velocities directly from the native API, we sent MIDI data of 'note on' along with corresponding velocities to Android. Unfortunately, the captured number of false positive and false negative hits

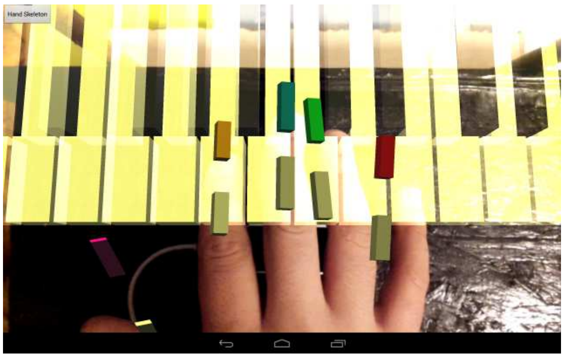

Figure 2. Detection using Leap Motion, with marked fingers.

using this procedure was annoying, as well as the latency exceedingly high.

2.2 Ultrasound detection system

An alternative, less researched detection technique involves the use of ultrasound [12] and laser [13] devices, containing both the ray emitter and the receiver. Ray reflections are captured and then used to estimate the position and motion of a reflecting body—in this case the finger. To our knowledge no studies have been conducted so far, investigating on the applications of ultrasonic and laser beams for realizing piano keyboard interfaces.

We set up another system consisting of a narrow matrix of eight HC-SR04 ultrasound sensors (see Figure 3). These sensors perform best and at highest rate (i.e. 200 fps) if tuned to detect objects between 20 and 30 millimeters: for these distances in fact the sound emission-reflection mechanism, based on pulse-width modulation (PWM), runs optimally.

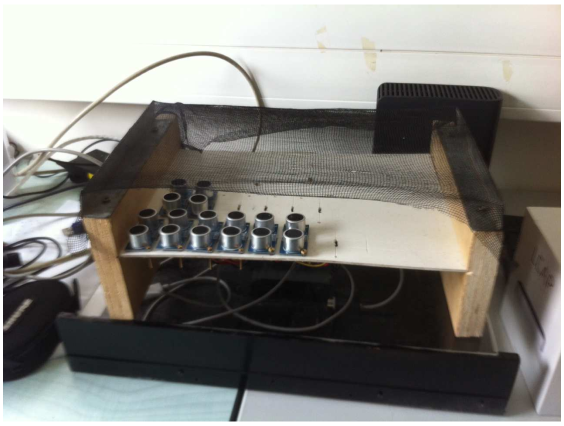

Figure 3. Ultrasound matrix, for eight keys.

The matrix was driven by an Arduino board. More in detail, the Arduino pings each sensor by regularly sending a trigger signal through the corresponding pin, then waits for the response from the sensors; the response times are proportional to the distance of the objects from the corresponding sensors. These actions are realized by a standard library called newPing [2] .

The detection procedure considers three finger-key distance thresholds, based on the positional information that is transmitted by each sensor at highest rate. The procedure uses the timestamp that labels every frame, to estimate the moving speed of the fingers in correspondence of such thresholds. Besides the small required computational effort, this procedure in particular allows for the prediction of the final velocity of the finger occurring in correspondence of the key stop position: if carefully tuned, as we want to do in a future work, a prediction strategy can compensate, all or in part, the latency that is caused by the various components of the system.

The serial communication between the Arduino and the Android application residing on the tablet was made via Internet UDP (see Figure 4); this communication adds small meanwhile unpredictable latency. We are currently overcoming this issue by substituting UDP with wired serial communication, via micro-USB, between the new Arduino ADK board and the tablet; the wired connection in fact ensures low and constant communication time.

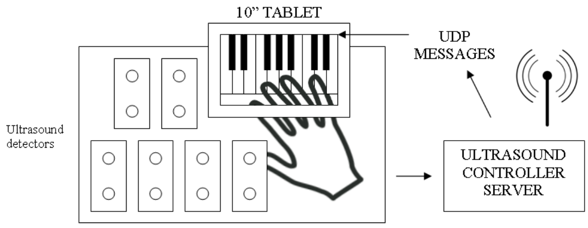

Figure 4. Augmented reality system schema with ultrasonic hit detection.

3. FEEDBACK

As previously said, the hand was visualized on the screen. Since we used the camera behind the laptop to capture this image, the visualization was affected by some latency. The result is nevertheless pleasing, as far as the image augmentation presents a piano keyboard that responds to the action of the pianist consistently. Alternatively, the touch screen can be substituted by unfolding a carpet, or even drawing a piano keyboard on the table: all such options should be tested by rigorous user experiments.

Audio is not a problem at this stage of the research: suffice to import a reasonably good digital piano sound bank in the system. Conversely, the haptic modality to date is rather unexplored, and much research has to be done to render 3D somatosensory cues representative of the keyboard: below, we report the experiences we made concerning this issue.

3.1 Tactile Feedback

Reproducing, at least partially, the consistency of a piano key by virtual means is a hard task. Technologies based on ultrasounds [14, 15], air vortex generation [16], magnetic repulsion [17] have been developed to reproduce sense of

[2] http://playground.arduino.cc/Code/NewPing

materiality of interactive virtual objects, with encouraging results. We have drawn from these technologies in an effort to generate virtual somatosensory cues of piano keys, while keeping an eye open on the applicability of the resulting system to mobile contexts.

In the case of visual capture, currently an ultrasound-transparent surface is needed where the fingers can tap. Such a solution has been proposed, for instance, in Cristal Piano [11].

In the case of ultrasound detection, we could add an elastic mesh about 33 mm above the receivers (see Figure 3), without introducing interferences in the reflected ultrasounds. This solution creates an interesting somatosensory effect of increasing feedback force with dipping of the finger inside the net.

In alternative, we asked pianists to wear gloves mounting small magnets at the fingertips (see Figure 5); then,

Figure 5. Magnetic glove.

we covered a section of the table with a surface exposing an opposed magnetic field. The force feedback resulting from the consequent magnetic repulsion presents surprisingly realistic aspects; unfortunately, pianists reported discomfort when tapping over the table or on the touch-screen meanwhile wearing these gloves.

Finally, early attempts to concentrate the beam of five high-energy ultrasonic emitters at a point located few centimeters over the table, where the pianist taps, have led to no interesting results so far.

4. DISCUSSION

A qualitative comparison between the two tracking systems is summarized in Table 1. Although not presenting precise figures, which would have required the use of oscilloscope, synchronized camera tracking and the like, it can give an idea of existing differences in performances and computational costs. The Leap Motion controller provides an almost continuous and spatially precise detection. On the other hand, the ultrasonic matrix needs one sensor per key.

At the moment, the hand recognition algorithm coming along with the Leap Motion is not yet suitable for tracking vertical movements of the fingers: too often it happens that fingers are not recognized by the system, especially after rapid actions. Hand distances from the camera which are lower than 7 centimeters are not detected, preventing from building a slim prototype. Furthermore, the detection suffers from solar light. Finally, the current drivers of the Leap Motion require lots of CPU time to identify the hands. For our purpose it would probably be much more efficient to track the pianist's fingers from the frontal position. If such drivers will feature this functionality, or at least will put raw image data available to the software environment, then the Leap Motion may become a challenging competitor in the field of music performance tracking.

Property	**Leap Motion**	**Ultrasounds**
Detected Points	continuous detection	one sensor per key
Hardware requirements	Android via laptop	Android via Arduino
fps	30∼300	200
False positive	too many	few
False negative	few	none
Latency	acceptable	excellent
Tactile feedback	discrete-space (surface tapping)	continuous (dipping on mesh)
Subjects' tolerance	low (gloves)	good (mesh)
Full Android integration	not yet	through Arduino ADK

Table 1. Comparison between Leap Motion and ultrasonic matrix detection.

The experiences using the ultrasonic detection revealed more robustness and predictable behavior. In fact, the data coming from the Arduino allowed for designing a fast and simple yet accurate distance estimation algorithm. The ultrasonic devices also offered the interesting possibility to add an elastic mesh in between the hands and the devices themselves. Conversely it is not possible to put reflective surfaces above them, as they would be mistakenly detected as additional bodies on top of the hands. Relief from this side-effect comes by slightly tilting the tablet while mounting it on the stand. Last, but not least, the ultrasound-based approach inevitably points to a more encumbering, not necessarily portable user interface: in this sense, our experience with this approach aimed at understanding the usability of the ultrasonic devices in the musical interface context rather than promising a commercially viable design strategy.

Informal use of the prototype suggests that both systems, in front of the execution of a note, generate haptic, auditory and visual cues that elicit sensation of a single event. Sometimes during this use we encountered random latencies that almost certainly depend on the wireless transmission via UDP; this issue will be soon solved by switching to USB communication between the detection device and the tablet PC.

Both the infrared- and ultrasound-based settings are quite cheap: currently, the cost of both prototypes in terms of hardware amounts to about one hundred Euros.

5. CONCLUSIONS

A mobile augmented reality test environment has been realized, and two detection systems have been compared to realize an immaterial piano keyboard interface. Preliminary tests show that immaterial keyboards have chance to be realized furthermore working in real time, yet several issues remain to be solved. The hardest issue is the generation of realistic haptic feedback. Future work will concern the customization at low-level of the Leap Motion detection algorithm, possibly making use of infrared markers for the pianist's fingertips once frontal positioning of the camera is made possible. Moreover, we will further research materials and techniques improving our current solutions for the presentation of haptic feedback.

Acknowledgments

The authors acknowledge partial support from the project PiaNo - *Piano from Nothing*, funded by Intel Corporation – Santa Clara (CA) through Tsinghua University in Beijing.

6. REFERENCES

[1] D. J. Levitin, K. Maclean, M. Mathews, L. Chu, and E. Jensen, "The perception of cross-modal simultaneity (or "the Greenwich Observatory Problem" revisited)," D. M. Dubois, Ed., vol. 517, no. 1. AIP, 2000, pp. 323–329. [Online]. Available: http://scitation.aip.org/getabs/servlet/GetabsServlet?prog=normal&id=APCPCS000517000001000323000001&idtype=cvips&gifs=yes

[2] N. Miner and T. Caudell, "Computational requirements and synchronization issues for virtual acoustic displays," *Presence: Teleoper. Virtual Environ.*, vol. 7, no. 4, pp. 396–409, Aug. 1998. [Online]. Available: http://dx.doi.org/10.1162/105474698565802

[3] T. Mäki-patola and P. Hämäläinen, "Latency tolerance for gesture controlled continuous sound instrument without tactile feedback," in *Proc. International Computer Music Conference (ICMC)*, 2004, pp. 1–5.

[4] J. M. Silva, M. Orozco, J. Cha, A. E. Saddik, and E. M. Petriu, "Human perception of haptic-to-video and haptic-to-audio skew in multimedia applications," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 9, no. 2, pp. 9:1–9:16, May 2013. [Online]. Available: http://doi.acm.org/10.1145/2457450.2457451

[5] Leap motion air piano. [Online]. Available: http://mikeheavers.com/main/portfolio-item/leap_motion_air_piano

[6] J. L. Crowley, F. Bérard, and J. Coutaz, "Finger tracking as an input device for augmented reality," in *Workshop on Automatic Face and Gesture Recognition,Zurich, Switzerland*, 1995, pp. 195–200.

[7] J. Letessier and F. Bérard, "Visual tracking of bare fingers for interactive surfaces," in *Proceedings of the 17th annual ACM symposium on User interface software and technology (UIST 04)*. Santa Fe, NM, USA: ACM Press, 2004, pp. 119–122.

[8] A. Goodwin and R. D. Green, "Key detection for a virtual piano teacher," in *IVCNZ*, 2013, pp. 282–287.

[9] F. Huang, Y. Zhou, Y. Yu, Z. Wang, and S. Du, "Piano ar: A markerless augmented reality based piano teaching system," in *2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics*, U. IEEE Computer Society Washington, DC, Ed.

[10] T. Kaaresoja and S. A. Brewster, "Feedback is... late: measuring multimodal delays in mobile device touchscreen interaction," in *ICMI-MLMI*, 2010, p. 2.

[11] E. S. Silva, J. A. O. de Abreu, J. H. P. de Almeida, V. Teichrieb, and G. L. Ramalho, "A preliminary evaluation of the leap motion sensor as controller of new digital musical instruments," in *14th SBCM - Brazilian Symposium on Computer Music*, 2013.

[12] M. Reunanen, K. T. Palovuori, T. Ilmonen, and W. Mäkelä, "Näprä - affordable fingertip tracking with ultrasound," in *IPT/EGVE*, E. Kjems and R. Blach, Eds. Eurographics Association, 2005, pp. 51–58. [Online]. Available: http://dblp.uni-trier.de/db/conf/egve/egve2005.html#ReunanenPIM05

[13] S. Perrin, A. Cassinelli, and M. Ishikawa, "Laser-based finger tracking system suitable for moems integration," in *Proc. IVCNZ03, New Zealand*, 2003, pp. 131–136.

[14] T. Hoshi, M. Takahashi, T. Iwamoto, and H. Shinoda, "Noncontact tactile display based on radiation pressure of airborne ultrasound," *IEEE Trans. Haptics*, vol. 3, no. 3, pp. 155–165, Jul. 2010. [Online]. Available: http://dx.doi.org/10.1109/TOH.2010.4

[15] T. Carter, S. A. Seah, B. Long, B. Drinkwater, and S. Subramanian, "Ultrahaptics: multi-point mid-air haptic feedback for touch surfaces." in *UIST*, S. Izadi, A. J. Quigley, I. Poupyrev, and T. Igarashi, Eds. ACM, 2013, pp. 505–514. [Online]. Available: http://dblp.uni-trier.de/db/conf/uist/uist2013.html#CarterSLDS13

[16] S. Gupta, D. Morris, S. N. Patel, and D. Tan, "Airwave: Non-contact haptic feedback using air vortex rings," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '13. New York, NY, USA: ACM, 2013, pp. 419–428. [Online]. Available: http://doi.acm.org/10.1145/2493432.2493463

[17] K. Karunanayaka, S. Siriwardana, C. Edirisinghe, R. Nakatsu, and P. Gopalakrishnakone, "Haptic mouse, enabling near surface haptics in pointing interfaces," in *Proc. ACHI 2013 - Conference on Advances in Computer-Human Interactions*, 2013.