Air Violin: A Body-centric Style Musical Instrument

Xin Fan
Computer Science & Engineering Division
University of Michigan
2260 Hayward Ave
Ann Arbor, MI 48109-2121
fanxin@umich.edu

Georg Essl
Electrical Engineering & Computer Science and Music
University of Michigan
2260 Hayward Ave
Ann Arbor, MI 48109-2121
gessl@eecs.umich.edu

ABSTRACT

We show how body-centric sensing can be integrated in musical interface to enable more flexible gestural control. We present a barehanded body-centric interaction paradigm where users are able to interact in a spontaneous way through performing gestures. The paradigm employs a wearable camera and see-through display to enable flexible interaction in the 3D space. We designed and implemented a prototype called Air Violin, a virtual musical instrument using depth camera, to demonstrate the proposed interaction paradigm. We described the design and implementation details.

Keywords

NIME, musical instrument, interaction, gesture, Kinect

1. INTRODUCTION

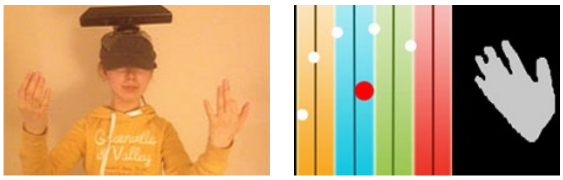
Humans perceive their own interactions from a body-centric perspective. We look outward into the world and our primary means of acting on it, our arms reach into our visual space. This makes this perspective on interaction in some sense "natural". The body-centric interaction space moves with us and provides a contextual situatedness of the interaction in a given setting.

The goal of this work is to exploring this notion of an interactive space directly by designing a prototype of a body-centric air violin performance. Evolutions in current sensing technology has solved a critical problem in finding positions in space. Depth cameras such as the Kinect and the emerging Leap Motion allow tracking of position in three-space within a visual field of view that is sensibly precise to allow for interesting interaction design.

In order to pre-figure what true body-centric sensing might look like, we are mounting a Kinect sensor on a performer's head to capture a corresponding visual interaction space that roughly matches the performer's visual field of view. Ultimately we want to augment the visual information of the performer with interactive possibilities, such as the placement of a virtual violin on which the performer can act. However head-mounted displays are only emerging, despite a long line of research into wearable displays [1]. Hence in our current work we emulate this property by a displaced display on a larger screen such as a monitor.

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

Body-centric interactions as we see them are interactions that operate in the space in front of a performer. This space can be populated by object but can also be simply an open space in which hands and gestures can conjure up sounds. The hands can act on imagined object such as an imagine violin, keeping the existing cultural association with that performance style. This grounds gestures while opening up new possibilities. In a sense body-centric performance allows the free use of the space in front of the performer, to be restricted for a particular performance purpose.



(a) (b)

Figure 1: (a) Barehanded body-centric interaction (b) The display of Air Violin. The four areas divided by the vertical lines represent String G, D, A, E from left to right. The left hand center is in the area of String D and no fingers are pressed, which corresponds to playing the open string D. The right hand simulates the bowing action by moving back and forth.

We designed Air Violin as an early example to demonstrate this interaction paradigm. We mount a Kinect sensor on a hat to be worn by the performer. The sensor is then connected to a laptop in order to detect hand motion and fingers in the camera's view, which in turn controls visualization and sound production. The system setup is illustrated in Figure 1. Air Violin is a violin-like musical instrument played with bare hands. The design looks to approach the style of playing known from performing a real violin. The user presses on the virtual strings with left hand fingers and moves the right hand like bowing, as illustrated in Figure 2. The left hand and the right hand each occupies half of the window of the visual feedback. For the left hand, the task is to recognize the hand posture and map the posture to a note. The four areas divided by lines on the left represent four strings of the violin, which are G, D, A and E. When the center of the left hand moves into an area, a note will be generated as if the corresponding string is pressed. We analyze the motion pattern of the right hand and map it to the bowing action. The synthesizer outputs sound in realtime according to the note and bowing action.

2. RELATED WORK

New ways of conceiving of violin performance as well as the use of gesture for musical performance have a long history, too extensive to review here in detail. For this we refer the read to [8, 6].

Hand tracking is a central problem in achieving virtual violin performance. A technique for solving this problem using gloves with sensors for gesture recognition has been proposed by Wang and Popovic [9]. Mistry and Maes [7] presented Sixth Sense, a wearable gesture interface, which uses a wearable camera to capture hand gestures. Hackenberg employs a time-of-flight camera to detect finger positions for real-time 3D gesture interactions [3]. The lower-cost Kinect with infrared camera can also enable interaction through finger, hand or body gestures.A number of authors address the problem of finger detection. In [4], Kinect sensor is used for segmentation and the k-nearest neighbors (KNN) algorithm is used for fingertip tracking. Digito is a fine-level gesturally controlled virtual musical instrument, which also uses Kinect sensor for hand and finger tracking [2].

There is great promise of emerging technologies that are immanently body-centric. A prime example of this are Google Glasses as they integrate camera, speaker and see-through display in a truly body-centric manner. Unfortunately, it is not yet commercial available. Once this technology becomes available it remains to be seen what level of interactions can be supported by the integrated camera. The must successful current implementations require the depth sensing of a Kinect, which may not appear in commodity wearable technology for a while.

3. IMPLEMENTATION

The hardware of the system consists of a Kinect and a computer. The system is developed using the OpenNI framework[1]. The sound module uses Synthesis ToolKit in C++ (STK)[2], which is a set of open source audio signal processing and algorithmic synthesis classes.

The construction of the system is illustrated as in The depth images in the video stream captured by the Kinect depth camera are processed using a set of OpenCV[3] methods. The advantage of using depth camera is that we can easily segment the hands using a threshold of the depth value. To determine the musical notes, the location of hand center and fingertips are tracked using a convex hull based implementation [5]. The center location is mapped to a string, which is either G, D, A or E. We analyze the finger locations and map them to the predefined hand postures which represent which finger is pressed using a template matching method. We precalculated the normalized average finger locations as the template for each posture. In the runtime, the euclidean distance between the finger coordinates and the template coordinates are computed. The posture corresponding to the closest distance is chosen. String selection is determined by the location of the hand center, and the hand posture determines the note on the string. The right hand motion simulates bowing. The system keeps a buffer of the past center locations of the right hand to determine whether to interpret the current state as the bowing motion or not. These parameters are then fed into the STK synthesis engine to drive a bowed string synthesis algorithm in real time. The performance of the finger tracking algorithm performs at a rate of about 0.2s. Other components of our system do not incur any further appreciable delay. While this is a not fully real-time yet, it allows a near real-time finger based performance to be demonstrated. Our method is not robust against jitter and noise leading to an occasional misregistration in the finger registration.

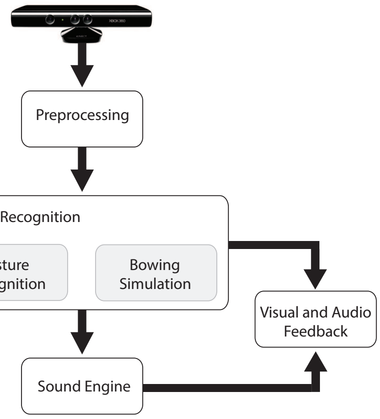


Figure 2: Modules of the system

4. DISCUSSION

We presented Air Violin, a violin-like musical instrument played with bare hands. It serves to illustrate the emerging paradigm of body-centric interactions for musical performance. Future work will include refinement of the finger tracking algorithm to improve frame rate and accuracy, hence enabling richer and more nuanced performance. We anticipate that emerging technologies such as the leap motion sensor and Google glasses will enable drastic evolution of this paradigm.

5. REFERENCES

[1] M. Billinghurst and T. Starner. Wearable devices: new ways to manage information. *Computer*, 32(1):57–64, 1999.

[2] N. Gillian and J. A. Paradiso. Digito: A fine-grain gesturally controlled virtual musical instrument. In *Proceedings of the 2012 International Conference on New Interfaces for Musical Expression (NIME12), Ann Arbor, USA*, 2012.

[3] G. Hackenberg, R. McCall, and W. Broll. Lightweight palm and finger tracking for real-time 3d gesture control. In *Virtual Reality Conference (VR), 2011 IEEE*, pages 19 –26, march 2011.

[4] T. Hongyong and Y. Youling. Finger tracking and gesture recognition with kinect. In *Computer and Information Technology (CIT), 2012 IEEE 12th International Conference on*, pages 214 –218, oct. 2012.

[5] H. Kato and T. Kato. A marker-less augmented reality based on fast fingertip detection for smart phones. In *Consumer Electronics (ICCE), 2011 IEEE International Conference on*, pages 127 –128, jan. 2011.

[6] E. Miranda and M. Wanderley. *New digital musical instruments: control and interaction beyond the keyboard*, volume 21. AR Editions, Inc., 2006.

[7] P. Mistry and P. Maes. Sixthsense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches*, page 11. ACM, 2009.

[8] J. A. Paradiso. Electronic music: new ways to play. *IEEE Spectr.*, 34(12):18–30, Dec. 1997.

[9] R. Wang and J. Popović. Real-time hand-tracking with a color glove. In *ACM Transactions on Graphics (TOG)*, volume 28, page 63. ACM, 2009.

[1] http://www.openni.org/

[2] https://ccrma.stanford.edu/software/stk/

[3] http://opencv.org/