The Augmented Djembe Drum – Sculpting Rhythms

Teemu Maki-Patola
M.Sc, researcher
Laboratory of Telecommunications Software and Multimedia, Helsinki University of Technology
+358 9 451 5849
tmakipat@tml.hut.fi

Perttu Hämäläinen
M.Sc, M.A., researcher
Laboratory of Telecommunications Software and Multimedia, Helsinki University of Technology
+358 50 596 7735
pjhamala@tml.hut.fi

Aki Kanerva
Research assistant
Laboratory of Telecommunications Software and Multimedia, Helsinki University of Technology
+358 50 544 8673
aki.kanerva@iki.fi

ABSTRACT
In this paper, we present an augmented djembe drum created by mounting a webcam inside a regular djembe. By moving your hands in the vicinity of the drum membrane, you can sculpt computer-generated rhythm patterns that imitate real-life djembe rhythms. You can also play the djembe for real, treating the patterns as interactive accompaniment. A computer vision system infers the 3d location of the performer's hands by tracking their shadows on the drum membrane. The individual drum hits of the patterns are automatic, but the player has direct control over their loudness, tempo and timbre. We explain the design and implementation of the instrument and share our design experiences. We also present qualitative results from testing the instrument with amateur musicians and experienced drummers.

Keywords
Augmented musical instrument, computer vision, drum, tactile feedback.

1. BACKGROUND
Gestural interfaces for sound control have been created since the Theremin in 1919 [21]. Their popularity has increased much more in the past two decades, thanks to quickly evolving technology.

Computer vision technology based on web cameras and inexpensive sensors (e.g. Phidgets [17]) greatly reduce the effort needed to use gestures as an input medium. Complementing input devices, several software tools, such as EyesWeb, Pure Data and Open Sound Control, [2], [14], [19], are available to help with creating interface prototypes rapidly. These tools handle functions from extracting control features from a webcam feed to routing the data to sound models, and also hosting the sound models themselves. However, moving from prototype to final product cannot often be done with these tools, as will be discussed later.

Thanks to increase in both processing power and research attention, computer vision is becoming more popular as an interaction method. One of its advantages is the lack of any wires or control devices, allowing the performer freedom of expression. Computer vision may be used for shape and feature recognition, and it offers higher precision in tasks such as evaluating color, object proportions, wavelengths or features that humans do not perceive.

On the topic of practical computer vision for user interfaces, there is a growing body of research [4], [5], [6], [11]. Practical systems often make simplifying assumptions, for example, that the user performs or at least tries to perform only motions that are relevant in a given application context [4], [5]. Markers, such as coloured gloves, can also allow the user to be tracked as 2d colour blobs instead of a fully articulated 3d skeleton [9], [13].

An article by Paradiso [15] and a book edited by Wanderley and Battier [23] offer a good introduction to existing gestural musical interfaces. Many prototypes have been created, and a few commercial gestural controllers exist as well [7], [22].

Computer vision or magnetic tracking is also used in some tangible interfaces to locate specific objects and buttons directly from a control surface [8], [16]. The instrument installations Music Table [20] and Augmented Groove [18] present controls layered on top of real objects, such as cards or LP records, visible on screen. Many traditional instruments have been expanded with additional sensors [15], [23]. However, the added control is often used to alter the existing sound through filters or effects. We have augmented a real musical instrument by adding a completely new way to play. The player can play the drum normally, use only the augmented playing method, or even play both versions at the same time.

The augmented djembe is different from our previous instruments based on computer vision [13]. The control features are not extracted directly from an image of an object or body part, but from lighting changes caused by body parts. The physical drum itself remains unaltered, because the technology is hidden.

One of our goals in building new instruments is to make the experience of playing available for a larger audience by reducing the amount of practice needed to play instruments. The augmented djembe follows this ideal as well, because the augmented playing method can be used even by beginners to create advanced results. The amount of control can then be gradually increased, such as hitting the drum every now and then, all the while staying in rhythm. This way, the augmented playing method supports the user's learning and makes practicing more rewarding.

2. AUGMENTED PLAYING METHOD
The player of an augmented djembe moves his hands above and on the surface of the physical drum, without striking the drum membrane. A rhythm pattern consisting of sampled djembe hits is automatically generated and output through a loudspeaker, and the player's hand motions control the timbre of the individual strikes in the pattern. This method was devised by

observing real-life djembe players. In this article, we refer to the non-augmented djembe as the 'real-life version'.

2.1 Real-life djembe

When playing the real-life djembe, the player strikes the drum alternately with left and right hands at a steady tempo and rhythm. Larger-scale dynamics are "sculpted" by modifying the loudness and timbre of the individual hits. Loudness is modified simply with strike strength, and hits can also be skipped, which we regard as changing the loudness to zero.

In addition to striking the drum membrane, the player controls the timbre of hits by striking at different locations. Striking near the centre produces a deep and dry sound, and a location near the edge results in a higher and more tonal and reverberant sound. The player uses this control to produce gradual changes in timbre over a fast pattern of strikes.

2.2 From real-life to augmented playing

Figure 1. Recording setup for analyzing hand motion during real-life djembe playing. To record hand motion, a magnetic motion sensor is attached to each hand.

We hypothesized that the djembe playing hand motions could be separated into components of different frequencies. The high frequency consists of the up-down motion of striking the drum, and the timbre and loudness modifications are seen as lower frequencies. If the high frequency was eliminated by automatically producing a pattern of strikes, the low frequencies could still be used to control the timbre and dynamics of the strikes in the pattern.

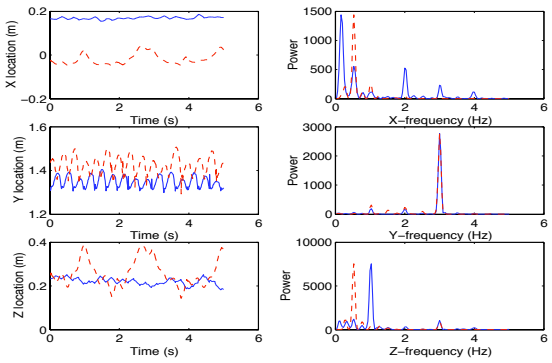

Figure 2. Motion of player hands along coordinate axes (left) and their respective frequency spectrums (right). The dashed line represents the right hand. The left side shows samples of movement data. Frequency spectra are evaluated from more comprehensive data. For example, the y-spectrum reveals that the drum is struck with alternating hands at a rate of 3 Hz or 180 bpm.

To prove this hypothesis, we recorded the motion of a djembe player's hands using a Polhemus magnetic tracker (see Figure 1). The location data was then analyzed in MatLab to extract separate movement trends.

By plotting the frequency spectrum of hand motion along coordinate axes (see Figure 2), we find out the different frequency components. Plotting vertical hand motion reveals that the membrane is struck at a rate of 3 Hz, visualized in the second row of Figure 2. The lower frequency components of dynamics in the recorded djembe pattern are 1 Hz for the left hand and 0.5 Hz for the right on the z axis, and the same 0.5 Hz for the right hand, and a 2 Hz pattern for the left on x axis, as illustrated in Figure 2. Combined, these patterns define the movement between centre and edge, which causes the most radical changes in the timbre of the sound.

3. IMPLEMENTATION

Figure 3. A web camera is mounted inside the physical djembe, facing towards the membrane. The software runs on a Windows laptop. Good speakers are also needed as the sound contains mostly low frequencies.

The augmented djembe consists of two parts: the physical djembe drum with a webcam mounted inside, and computer software. Both the sound model and the computer vision run inside the same software, a custom Windows application.

3.1 Hardware

The large viewing angle (70 degrees) of the Creative NX Ultra webcam allowed us to mount it at the narrow base of the drum, where it could be fixed securely, while still capturing the entire membrane in its field of view. The software runs on a regular Windows computer, and poses no special requirements. However, in minimizing audio latency, it is recommended to install either ASIO drivers supported by the soundcard, or the generic ASIO4All drivers [1]. Strong ambient lighting or a lamp positioned perpendicular to the drum membrane, pointing at the drum, is required.

3.2 Software

The initial prototype was constructed on the PureData [19] software, which was used for both input mapping and sound generation. Communication with the computer vision application was done through Open Sound Control [14] messages. Soon, the input mapping was moved to the custom computer vision application, leaving only the sound production for PureData. It then became apparent that a custom sampler would serve us better for further prototyping, and we abandoned the Pure Data platform.

3.3 Image analysis

Figure 4. Raw images from the web camera. The hands project blurred shadows on the surface of the drum membrane.

Figure 4 shows what the camera sees during playing, when a lamp causes hands to cast shadows on the drum membrane. The real-life djembe controls for timbre and loudness – strike position and strength – are mapped from hand location and height as seen by the camera. Based on the distance of a hand from the edge to the centre of the drum, the software selects a sample recorded from a similar position to play back. Conversely, the height of a hand is used to control the volume of the sample.

Despite this poor image data, it is possible to extract the required information in three steps.

3.3.1 Circle fitting

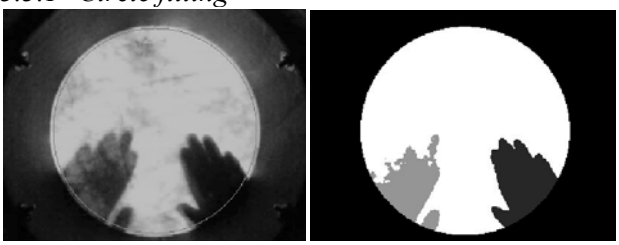

Figure 5. Pixel mask generation by fitting a circle constraining the lit area of the membrane. Note how the left hand's shadow is brighter due to the hand's distance from the surface.

First, a pixel mask is created to include only pixels inside the membrane area for further processing. This should be done every frame, because the camera is not absolutely fixed, and the djembe may move during playing. This has the added benefit of ignoring artifacts during an initialization phase, and being more tolerant of lighting changes during playing.

The translucent membrane receives more light than the inside of the drum body, so its pixels are brighter than those of the body. The membrane's pixel mask is created by fitting a circle to contain the bright pixels. We use genetic optimization to maximize a fitness function that has the radius and centre of the circle as parameters. This works well even when hand shadows are present in the image. The fitness function is:

$$f(r,x,y) = r\sum_{i=1}^{N}\left[I(x+r\cos\alpha_i, y+r\sin\alpha_i) - I(x+r'\cos\alpha_i, y+r'\sin\alpha_i)\right]$$

where r'=1.05r, I(x,y) is the intensity of the image pixel at coordinates x,y, and $\alpha_i = i2\pi/N$. Basically, fitness equals the intensity of the circle with radius r minus the intensity of a slightly larger circle with radius r'. This is maximized when the first circle is aligned at the rim of the drum. The sum of the intensities is multiplied by r to favour large circles, which speeds up the convergence of the optimization. According to our experience, it is sufficient to sample the circles at 16 points, that is, $N=16$.

3.3.2 Thresholding the membrane

The membrane of a typical djembe is not uniform in color. To remove the pixels that are not in shadow we simply threshold away the pixels that are brighter than the color variations of the membrane. As a result of the variations the final hands may become "broken", as seen in Figure 5.

3.3.3 Shadow area and average color

After thresholding, only hand shadows remain visible for further processing. The surface is divided into two halves, and for each half the amount of overlaying shadow surface and its average colour is evaluated.

Initially, we had used an angle histogram to determine whether zero, one or two hands were visible. The histogram calculated the amount of shadow mass as a function of angle from the center of the membrane. A valley in the histogram would indicate the space between hands. This approach was abandoned, because it was prone to errors when playing with spread fingers, which made it difficult to interpret clusters properly.

3.4 Sound model

The sounds of the augmented djembe were sampled from the same drum, and are played back using a software sampler. The sample matrix consists of strikes at varying distances from the membrane centre, with varying intensities. For each location and intensity pair, a number of different samples were recorded to avoid sounding too mechanical.

To further reduce monotony, each output sound consists of multiple samples recorded from nearby locations, with continuously changing weights. As a result, the sound model can produce a nearly infinite amount of different combinations even with a small selection of samples.

3.5 Control mapping

Strikes are generated for both hands at a steady tempo, which can be changed with the software's interface. Each hand strikes 8th notes alternately, resulting in a straight 16th-note pattern. To avoid the mechanical quality of strict tempo, the strike times vary randomly within 5% of the actual beat [3]. This simulates a human player, and sounds livelier.

The distance of a hand to the membrane's centre is mapped as weight in choosing samples to mix together from the sample matrix, effectively modifying the timbre.

The height of a hand (distance from membrane) is calculated from the average luminosity of the visible shadow on each membrane half. The higher the hand is, the brighter the shadow becomes, creating a penumbra. For this reason, strong ambient lighting or a lamp with dimensions, such as a fluorescent lamp, is required.

Thus, the locations of the player's hands control both the timbre and loudness in the rhythm pattern. For producing typical djembe rhythms, the user moves hands in smooth, rhythmic motions over the drum's surface. This motion resembles the low frequency motion of real-life djembe playing excluding strikes.

3.6 Visual feedback

Both the unprocessed camera image and the result of computer vision processing are shown on screen. This feedback helps the user understand how the sound is generated and sculpted.

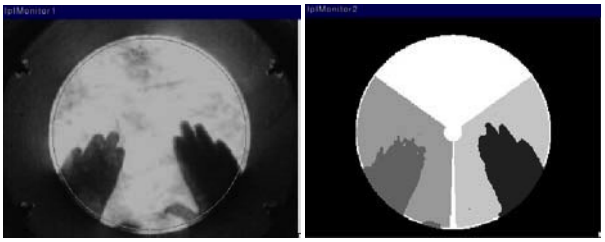

Figure 6. Visualization. Distance from centre is represented by darkening the according sector, and hand height visualized by darkening the hand itself.

4. USER TESTS

We conducted user tests to find out how people with different musical backgrounds experience the instrument, and to see what playing styles would be discovered. The informal test situation was a jam session where each user was allowed to experiment with the various features of the instrument. All comments given during test sessions were recorded.

4.1 Users

The test users consisted of six people. One was an experienced percussionist, three had a strong musical background and two were amateur musicians. Only the percussionist had previous experience of playing a djembe. One subject had experience with other hand drums, while the rest had none. Five of the subjects were male. The ages of the subjects were between 24 and 46 years.

4.2 Procedure

First, the real-life djembe was introduced to the user with a three-minute introduction video. The video taught how to hold the instrument properly, demonstrated playing and explained timbre and dynamics controls. After the video, the user was encouraged to experiment with the djembe.

Next, the augmented version was introduced by explaining its control logic with the test conductor showing simple demonstrations. The user was told that they could ask to change the tempo.

The user was then allowed to experiment with the augmented djembe for as long as they liked, which ranged with different users from half an hour to two hours. Users were encouraged to comment during the entire test.

Finally, the user was interviewed with informal questions:

1. How would you describe the playing experience?
2. What kind of playing styles did you find?
3. Were you able to achieve satisfying results?
4. Did you feel that you were in control of the instrument?
5. How did it feel to play together with the augmented djembe?
6. What would you change in the instrument?

4.3 Results

Most users were intrigued by the augmented instrument. They felt it was a good idea to have the augmented playing method attached to a real-life instrument: “Because the user interface is an actual drum, playing feels natural, like playing a real instrument”.

The users quickly learned to produce patterns with little or no dynamic variations, but maintaining a complex pattern for a longer period of time turned out to be more difficult than expected. According to some users, a partial reason for this was that they did not know much about typical real-life djembe patterns. We assume they also had difficulty in knowing how the hands should be moved to produce a pattern they had heard.

It took some time for each user to find a suitable playing style. Most subjects felt this was because the control approach was fundamentally different than their mental model of playing a drum, even if it was based on drum playing. One user commented that the instrument was easier to understand once you stopped thinking of it as a drum. Still, after some practice, most subjects found ways to produce rhythms they enjoyed.

4.3.1 Playing experience

Most users felt that the instrument felt natural and behaved logically. Many noticed some latency, but felt they could adapt to it. Some users commented even without being asked that the augmented instrument was easier to approach than the real-life djembe. These users enjoyed the ability to produce decent-sounding results even with little practice.

The user with the most musical experience commented that combining real-life djembe playing with the augmented method was the most fruitful in a musical sense. He quickly learned to use the augmented patterns as a dynamic background for his own djembe playing. He felt the steady patterns allowed him to practice real-life playing, and he was able to play more freely and use embellishments when the augmented patterns took care of staying in tempo. Nevertheless, the dynamics of the patterns could be controlled enough to be musically interesting instead of acting as a monotonic metronome.

The experienced percussionist felt that the augmented play mode was frustrating when he already knew how to play the instrument. It appears that a more fruitful approach for him would have been to use the computer vision to control the parameters of effects applied to the sound of the real djembe.

4.3.2 Playing styles

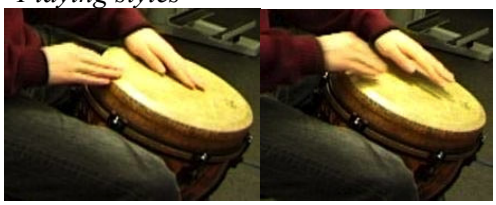

Figure 7. Small motions produce drastic changes in the sound. Lifting the hand only a few centimeters halves the amplitude.

For maintaining a particular rhythmic pattern, the most effective way to play was to keep one’s hands on the drum’s surface. It provided an anchor for repetitive movements, and made it easy to stay in tempo. Several users felt that without touching the drum, it was difficult to repeat motions accurately.

The most successful playing styles involved rocking the whole body with the rhythm. This way, the user was immersed in the rhythm of the pattern, which was additionally helped by closing the eyes and concentrating on the aural feedback of the drum’s sound. One user described the experience as being “submerged in a flow of mind.”

Figure 8. Touching the drum’s surface helps to repeat motions accurately. Here, the palm of the left hand stays in contact while its angle affects the size and intensity of the shadow. The right hand moves with a larger motion, controlling amplitude.

4.3.3 Problems

Users commented that the biggest problem was dropping out of rhythm without making a clear error. This seems to be the combined result of the nature of rhythm patterns and latency. The sound of a strike is based on what the camera sees at the time, which may be data that is over 50 ms old due to the

camera's 30 Hz refresh rate and the delay in starting the next sound. As a result, the player needs to adapt to performing controls slightly before the sound is heard.

However, the user has little feedback of the phase of his motion processing. For example, while the user keeps on emphasizing the first hit in a pattern, his control action may happen in a good time or be almost late. Both situations appear similar to him. If his action is almost late, his time precision may fluctuate just enough to cause the control action to suddenly happen too late missing the next hit. As a result the pattern is shifted, which confuses the user who did not perceive making any errors. Some users were able to adapt to this once they understood what caused the shifting: "you try to adapt your playing to the sudden error but until you understand why it happens, fixing it may be difficult."

4.3.4 Suggestions for changes

A simple way to reduce the shift error described above would be to use a camera with faster refresh rate. We considered adapting the tempo to the player dynamically, but concluded that this would lead to both user and software adapting to each other, resulting in even more confusion.

Many subjects suggested using presets for different rhythmic patterns, as well as different sounds to make the playing less monotonic. Presets could be selected with an attached switchboard, for example. Additionally, users wanted a way to modify the tempo quickly and easily.

5. DISCUSSION

The general impression based on user tests was that our approach has potential, even though there are a few technical issues, such as reducing the latency, to tackle. The augmentation's main benefit was in making the playing more rewarding already early on. Users were able to produce interesting results quickly, and gradually increase their interaction with the real-life djembe. This made the instrument easier to approach and supported the learning process by keeping the motivation high. The automated rhythm patterns also acted as examples of real-life djembe playing.

The conducted user test supports our earlier findings [12] of tangible objects and tactile feedback improving the feel of musical interfaces. Users enjoyed having a real, physical drum as the interface.

5.1 Software Tools

Using PureData allowed us to quickly see if our approach would lead to anything. However, as the instrument developed, the software soon became cumbersome for both the input mapping and sample playback. Once the concept was verified, we moved to a custom application.

Our earlier experiences with existing software tools such as PureData are similar – they are invaluable for rapid prototyping but become unwieldy when polishing final versions. Using highly general architectures is bound to make achieving specific goals more difficult than it is with software designed to achieve exactly those goals. Naturally, making everything by yourself requires a strong programming background but if you have it, it may be a faster approach than learning to expand existing software tools for your specific needs.

6. FUTURE WORK

As suggested by the user test subjects, the instrument would benefit from several additional features. First addition should be to add tempo control, realized as three to five presets in the up-most sector of the drum surface. Through using the instruments visual feedback, the upper sector could contain even pull-down menus for choosing different rhythm patterns, sounds and play modes.

7. CONCLUSIONS

We introduced an augmented djembe drum, which features a playing method augmented with computer vision and sample-based sound generation. The augmented method allows users to concentrate on sculpting rhythm pattern timbre and dynamics without concentrating on precise tempo.

In our user tests, we discovered that the augmentation allowed the users to produce musically interesting results with less practice making the augmented djembe easier to approach. Additionally, a novel playing style was discovered, where users played the real-life djembe accompanied by dynamic automated patterns.

The application of computer vision we developed for this instrument suggests that as long as the context is known well beforehand, accurate control data can be extracted even from seemingly poor image data.

8. ACKNOWLEDGMENTS

The authors would like to thank all the users who tested the instrument for their valuable feedback. The research was partly supported by Pythagoras Graduate School funded by the Finnish Ministry of Education and the Academy of Finland.

9. REFERENCES

[1] ASIO4All Universal ASIO Driver, http://www.asio4all.com/ (Link visited Jan 2006).

[2] EyesWeb software home page, http://www.eyesweb.org/ (visited 1.2006)

[3] Dahl, S. The Playing of an Accent – Preliminary Observations from Temporal and Kinematic Analyses of Percussionists. *Journal of New Music Research* No. 3: 225-233, 2000.

[4] Freeman, W.T., Anderson, D., Beardsley, P., Dodge, C., Kage, H., Kyuma, K., Miyake, Y., Roth, M., Tanaka, K., Weissman, C., Yerazunis, W., Computer Vision for Interactive Computer Graphics, *IEEE Computer Graphics and Applications*, May-June, 1998.

[5] Freeman, W.T., Beardsley, P.A., Kage, H., Tanake, K., Kyuma, K., Weissman, C.D., Computer Vision for Computer Interaction, *SIGGRAPH Computer Graphics Newsletter, Vol. 33, No. 4*, November 1999, ACM SIGGRAPH.

[6] Hämäläinen, P., Höysniemi, J., Ilmonen, T., Lindholm, M., Nykänen, A. Martial Arts in Artificial Reality, *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'2005)*, Portland, Oregon, 2-7 April 2005, ACM Press.

[7] I-Cube website. (Visited Jan 2006) http://infusionsystems.com/catalog/index.php

[8] Jordà, S., Kaltenbrunner, M., Geiger, G. Bengina, R. The Reactable. *Proceedings of the International Computer Music Conference (ICMC05).* Barcelona, Spain, 2005.

[9] Karjalainen, M., Maki-Patola, T., Kanerva, A., Huovilainen, A. and Janis, P. Virtual Air Guitar. *Proc. AES 117th Convention*, San Francisco, CA, October 28-31, 2004.

[10] Machover, T. Instruments, Interactivity, and Inevitability. *Proceedings of the NIME International Conference*, 2002.

[11] Moeslund, Thomas B.Computer vision-based human motion capture – a survey. Aalborg: Aalborg University, Laboratory of Computer Vision and Media Technology, 1999. (LIA Report; LIA 99-02.-ISSN 0906-6233)

[12] Mäki-Patola, T. User Interface Comparison for Virtual Drums. *Proc. Int. conference on New Interfaces for Musical Expression (NIME05)*, Vancouver, Canada, May 2005.

[13] Mäki-Patola, T., Kanerva, A., Laitinen, J., and Takala, T. Experiments with Virtual Reality Instruments. *Proc. Int. conference on New Interfaces for Musical Expression (NIME05)*, Vancouver, Canada, May 2005.

[14] Open Sound Control project home page, http://www.cnmat.berkeley.edu/OpenSoundControl/ (visited Jan 2006)

[15] Paradiso, J. Electronic Music Interfaces: New Ways to Play. *IEEE Spectrum*, 34(12), 18-30, 1997. Later expanded as an online article, 1998. (visited Jan 2006): http://web.media.mit.edu/~joep/SpectrumWeb/SpectrumX.html

[16] Patten, J., Recht, B., Ishii, H. Audiopad: A Tag-based Interface for Musical Performance. *Proc. Int. conference on New Interfaces for Musical Expression (NIME02)*, Dublin, Ireland, 2002.

[17] Phidgets sensor hardware, http://www.phidgets.com/ (visited Jan 2006)

[18] Poypyrev, I., Berry, R., Kurumisawa, J., Billinghurst, M., Airola, C., Kato, H. Augmented Groove: Collaborative Jamming in Augmented Reality. *ACM SIGGRAPH Conference Abstracts and Applications, 2000.*

[19] Puckette, M. PD – Pure Data, http://www.crca.ucsd.edu/~msp/Pd_documentation/ (visited Jan 2006)

[20] Rodney, B., Makino, M. Hikawa, N., Suzuki, M. The Augmented Composer Project: The Music Table. *In proceedings of the International Symposium on Mixed and Augmented Reality,* Tokyo, Japan, 2003.

[21] Theremin info pages. (Visited Jan 2006) http://www.theremin.info/

[22] The Yamaha Miburi System (visited Jan 2005) http://www.spectrum.ieee.org/select/1297/miburi.html

[23] Wanderley, M., Battier, M. Eds. *Trends in Gestural Control of Music*. Ircam - Centre Pompidou - 2000.

[24] Wanderley, M. *Performer-Instrument Interaction: Applications to Gestural Control of Music*. PhD Thesis. Paris, France: Univ. Pierre et Marie Curie - Paris VI, 2001.

[25] Wright, M., Freed, A., Momeni A. Open Sound Control: State of the Art 2003. *Proceedings of the 3rd Conference on New Instruments for Musical Expression (NIME 03)*, Montreal, Canada, 2003.