REAL-TIME AUDIO SYNTHESIS IN A WIRELESS INTERACTIVE SENSOR PLATFORM

Da-Lei Fang, Yi Qin, Qiang-Bin Chen
Shanghai Conservatory of Music
Dept. of Music Engineering
Shanghai, China

Jia-Liang Lu
Shanghai Jiao Tong University
Dept. of Computer Science & Engineering
Shanghai, China

ABSTRACT

In this paper, we introduce the real-time audio processing for live performances in Wireless Interative Sensor Platform (WIS platform) that is built by our team. WIS platform is designed based on on-body motion sensors and processing computers to support dynamic and interactive applications. It is composed of a capture system, a processing unit and an audio-visual display control unit. We embed several general real-time sound synthesis modules including additive synthesis, FM synthesis, scene sound synthesis and effect processing functions for audio synthesis to satisfy the distinct needs of the live performances. The wireless communication between multiple on-body sensor nodes and display control units has been investigated to enable smooth gesture recognitions and offer real-time audio and visual experiences to users. We also illustrate the audio synthesis through a recent interactive installation *Air-ergy*.

1. INTRODUCTION

Recently, audio synthesis through a real-time interactive system has become a popular research direction in computer music. Meanwhile, advances in wireless communication technologies and the reducing size of sensors have empowered the use of wearable sensors in interactive music performances. Many sensor systems [7, 10, 6, 11, 12] were developed for interactive music performances. Each of these system achieve a specific application. In [7], authors had an interesting discussion concerning the integration of the dancing body and the musical body in the Performance of Electronic Music. In [6], authors proposed a wireless system for gesture following for music pedagogy. In French ANR Interlude project [1], Mo interfaces were developed to explore novel gestural interfaces for musical expression. The MO handheld unit is designed to capture a wide range of gestures, and a real-time gesture and sound processing software was also developed.

Our idea is to develop a platform that fits the general requirements of the interactive music performances, so that artists can easily apply it in their works. WIS platform is developed in MAX/MSP with OSC message interfaces both real-time gesture capture system (on-body sensor nodes) and actuator system (sound synthesis and interactive video). Real-time sound synthesis unit, one of the core parts of WIS platform, acquires several general dynamic parameters via OSC. Additive synthesis, FM synthesis, scene sound synthesis and effect processing function modules are used according to these parameters. Moreover, the flexibility of WIS platform is enhanced so that users would feel free to embed the audio processing modules they need and will be able to remap the relationship between parameters and those audio modules.

The rest of the paper is organized as follows. Section 2 gives a brief description of WIS platform with three main parts. We provide the details of the control and audio synthesis unit in WIS platform in section 3 and section 4 . The installation Air-ergy is presented in section 5. This installation has been exhibited in the 3rd Shanghai International Electronic Music Week [3]. Section 6 concludes our work and gives future research directions.

2. WIS PLATFORM

WIS platform consists of three main parts: capture system, processing unit and display unit. However the primary purpose of WIS platform is to implement an open architecture to combine these three parts. One can easily adapt either capture or display part regarding to application's requirements and circumstances. The system architecture of WIS platform is given in figure 1.

2.1. Capture System

The capture system includes wireless on-body sensors and receivers. The sensing part is a SparkFun 9 Degrees of Freedom (9DoF) Razor IMU [2] that incorporates a triple-axis gyro, a triple-axis accelerometer and a triple-axis magnetometer. In addition, a modified XBee adapter with an XBee 2 RF module [5] is used as the wireless communication module whereby it sends and receives data via the TX/RX pins of 9DoF board. This portable device is powered by a 900Am.h lithium battery. All the components of the wireless on-body sensor are contained in a $6cm \times 4cm \times 1.5cm$ translucent plastic box as shown in figure 2.

The wireless on-body sensor sends out a packet including one-byte Node_ID, three two-byte X-, Y-, Z-axis acceleration data, four two-byte Quaternion data (from the 3×3 DCM Matrix (Direction Cosine Matrix) and Tait - Bryan angles data) and one one-byte Symbol Identifier *

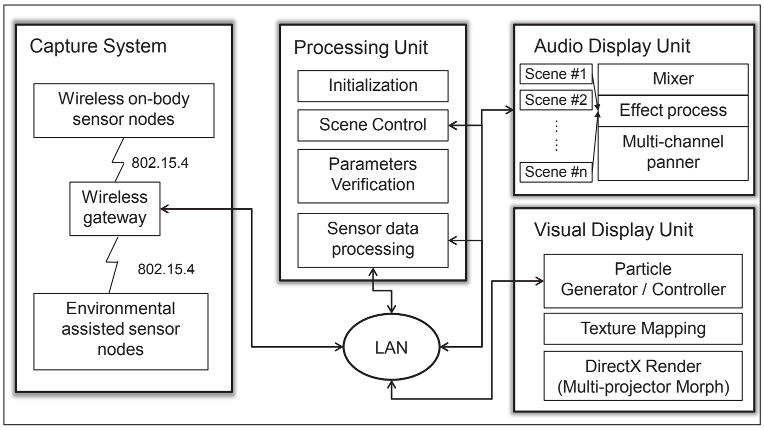

Figure 1. Architecture overview of WIS platform.

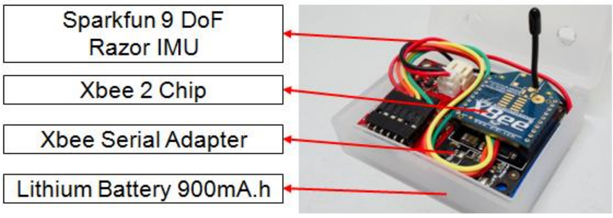

Figure 2. The wirelss on-body node.

for data synchronization. The data is refreshed every four milliseconds. The XBee transceivers work at 57600 bps.

After receiving and unpacking the data package, the Mega Board recovers the 3×3 DCM Matrix and Tait-Bryan angles data from Quaternion. Z_OSC library of Arduino is used in program to implement a standard OSC communication between Arduino-compatible hardware and computer or other devices. We divide all the data into two separate OSC data packages. One contains yaw, pith, roll float-type Tait-Bryan angles data and 3-axis acceleration int-type data; and the other one sends the nine DCM Matrix data of float type to the processing unit.

2.2. Processing Unit

The processing unit is composed of three modules. *Initialization* module is responsible for parameters and state initializations as well as for loading DSP. *Scene control* module loads pre-defined scenes with regards to the environment definition and the results of sensor data processing module. It sends OSC packets through LAN to control audio and visual display unit. *Parameters verification* module serves as a tool to visualize the parameters and to test them on-the-fly.

2.3. Audio-Visual Display Unit

The audio display unit consists of mixer, effect process and multi-channel panner modules. It interacts with processing unit to play different scenes as well as to generate sonic elements related to captured data. It is worth noting that processing unit and audio display unit are realized on the same computer to minimize the influence of delay on the audio system. Therefore, the control signal is directly sent through MAX/MSP program, but does not go through LAN. In section 4, the real-time audio synthesis in WIS platform will be more detailly presented. The visual display unit is implemented on the second computer for particle generation and control, as well as texture mapping and render. If multi-projector system is necessary for the application, then the morph should also be activated in this part.

3. REAL-TIME SYNTHESIS CONTROL

The OSC packets sent from capture system such as the yaw, pitch, roll and 3-D acceleration are processed, then the processing unit will send control packets to audio and visual units to interact.

3.1. Acceleration Data Processing

The acceleration data is processed and output by two ways: *continuous output* and *state output*.

Continuous output represents any movements, such as turn around and foot-lifting. Firstly, the processing unit gets the differences of acceleration in a small period. Then the absolute acceleration value will be computed. Finally, an envelope follower is used to smooth out the change. To make adjustment easier, all the parameters above can be stores into presets. So the same structure can be used in many sub-patches with different parameters presets. The continuous output is sent to many sub-patches, such as step recognition and wind synthesis.

State output can represent any movement states, such as fast movements or sudden stop. The continuous output passes through a series of comparators. Then four state triggers are output according the value of continuous output: slow, mid, high and fastest. Wind 1 and 2 use these data to change the timbre.

3.2. Step Recognization and Position

The step recognition is essential for in-door tracking and is widely use in Dead Reckoning solutions [8, 9]. In our realization, the step recognition is based on yaw, pitch and roll and 3-axe acceleration. Firstly, all the raw data go through the AIO with offset for adjustment. By setting time switch, the unit is able to synchronize data from different on-body sensors for processing. Secondly, differential and envelope operations are performed on the data series to output a continuous acceleration variation.

The 3D acceleration is enveloped to get direction and a series of values. A threshold value should be defined to trigger the step. A number of tests are executed to get this threshold. It is also possible to implement a self-learning algorithm to get this threshold. In order to eliminate repetitive trigger, a switch is used with time interval for the possible following trigger. During the exhibition, it turns out that the threshold range is applicable for 90% of utilization.

4. AUDIO SYNTHESIS IN WIS PLATFORM

4.1. Wind Scene Sound Synthesis

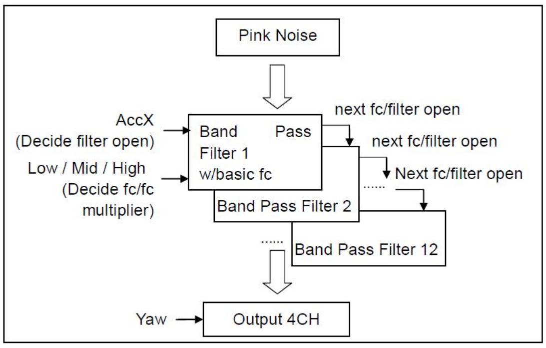

Figure 3. Wind scene diagram.

Wind Scene is a scene triggered by the rotation data from the sensor on the user's shoulder. This scene synthesis a breeze/wind sound effect. *Pink Noise* is used as the raw sound material for *Wind Scene*. In the first version, a low pass filter with resonance was used. Although this was enough to simulate a wind sound, we implement a second version, and try to bring more musical elements into the wind sound. 12 band pass resonators are used. The cutoff frequency of each next filter is a multiple of the cutoff frequency of the current filter. In this way a series of harmonics are obtained. The cutoff frequency is between 100 and 400Hz, and the multiples vary from 1.05 to 1.30. The variation of these parameters is triggered by the 3-level acceleration state output (low, mid and high) from sensing data processing part. The volume of each resonator is controlled by the continuous acceleration output. Thus, the sound of a fast movement straight after a slow movement will be different from a pure fast movement. The pan position of the wind among 4 output channels is controlled by Yaw data, so that users are easy to detect the relation between his/her movement and the sound position. The diagram of *Wind Scene* is given in figure 3, in which **fc** stands for filter cutoff frequency.

Wind 2 scene has a similar basic structure as *Wind Scene* above, but the volume of each filter is controlled by the height data from vvvv. The output of wind 2 is fixed to 4 channels.

4.2. Additive Synthesis

Additive synthesis consist 12 sine oscillators. The frequency (pitch) of 12 oscillators is the same as the 12 band pass resonators. The volume of each oscillator also has the similar control as 12 band pass filters. That means the additive synthesis has the same harmonics as wind 2, but the harmonics are more pure than wind 2. To make the sine sound lively, each oscillator has its own slightly random pitch and volume modulation. The output of additive synthesis crossfades with wind 2 by the Y position data from the AIO patch. The user can feel this crossing fade between the reality (wind 2) and the virtual reality (additive synthesis).

4.3. FM and Other Synthesis

FM synthesis is triggered by low acceleration state output, then a random note from C5 to B5 (60 to 71) is played. The structure of FM is 4 sets of operators with a frequency ratio of 7:1. Each carrier has a little detuned to create chorus effect. The output of the FM synthesis is fixed to 4 channels. FM synthesis is used as a pad sound. We also implement a sample record/playback system. The user can record his/her voice before the exprience, and these samples are played in a granular synthesizer during the exprience.

5. AIR-ERGY: AN INTERACTIVE AUDIO-VISUAL EXPERIENCE

In this part, we present *Air-ergy* as a work of interactive installation (videos are available on YouTube). It is exhibited in the 3rd Shanghai International Electronic Music Week [3] in October 2011. Figure 4 show a scene of *Air-ergy*. This work emphasizes the interactions between human body and its environment. By making use of WIS platform in a particular space, the demonstration of this work is triggered by the user's body action. The idea of this installation is to show the kinetic energy of a man in real world through interactive sound effects and videos in a particular space. This is a closed loop system that embodies a real physical world transiting to a virtual cyber world. Table 1 gives the audio synthesis of a set of gestures in *Air-ergy*.

Bubble up visual setting runs on two computers. One for audio (Max/MSP) and the other for video (vvvv [4]). In vvvv, we define a basic particle height with a LFO modulation to simulate a float state. When a step is detected in the Max/MSP, a trigger will be sent to vvvv through

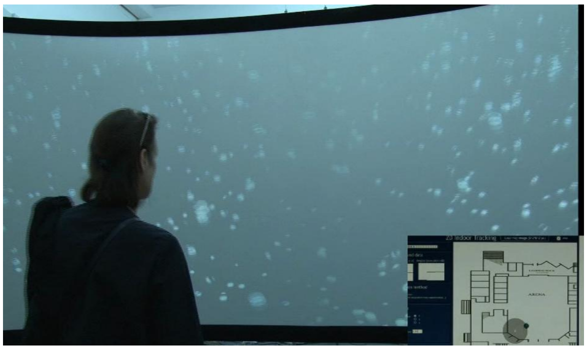

Figure 4. A scene of *Air-ergy*. On the right-bottom snapshot shows the user localization in the environment.

Table 1. Interactions implemented in *Air-ergy*

User gesture	**Audio-visual experience**
Slow body rotation	drifting breeze
Rapid body rotation	swift and wide oscillation of sound wave
Raise foot	particle floating up; variation on sound frequency
A step of user	virtual footstep sound
User's potions during a walking	position of sound in the audio field
Leave from the area	the particles settle down; deformed user voice looming

LAN using OSC. The vvvv will add a fast up ramp to the height of the particles. If no more steps are received, the height will fall slowly. If more steps are received, the height will be added cumulatively. The height is also sent back to Max/MSP through LAN using OSC. This parameter is used to determine the volume of each partials in wind 2 and additive synthesis.

6. CONCLUSIONS

WIS platform is an open platform designed for interactive music performance. We set the basic sound synthesis modules in real-time sound synthesis unit. Users are free to add other sound synthesis modules into WIS platform as well as redo the mapping relationship according to the specific requirements. Moreover, the wireless sensor systems or motion capture systems can be easily interfaced with WIS platform. Air-ergy is a good example that employ of WIS platform. We are currently working on the WIS platform to include more general audio synthesis modules and map the general parameters.

7. ACKNOWLEDGMENTS

We would like to thank Christophe Lebreton at GRAME and for a number of suggestions and discussions on this project as well as Quentin Pyanet at IRCAM for his contrubition to this project. Chao-Lun Ni at SJTU also has helped for the installation of *Air-ergy*.

This work is supported by NSF of China under grant No.61100210, the Key Science and Technology Program of STCSM under grant No. 10511500900, Shanghai MEC Innovation Project No. 12YS080 and Shanghai Key Laboratory for Art of Musical Acoustics. The authors from SHCM are supported by Construction Program for Innovative Research Team, Shanghai Institutions of Higher Education (2nd Phase) and Talents Training Mode Innovation Experimental Zone Construction Project by MOE and MOF of China.

8. REFERENCES

[1] "Interlude project, ANR: http://interlude.ircam.fr."

[2] "Sparkfun 9 dof razor imu: http://www.sparkfun.com."

[3] "The 3rd Shanghai International Electronic Music Week: http://www.emusicfestival.cn/."

[4] "vvvv - a multipurpose toolkit: http://vvvv.org."

[5] "Xbee 2 802.15.4 oem rf module: http://www.digi.com/."

[6] R. Aylward and J. A. Paradiso, "A compact, high-speed, wearable sensor network for biomotion capture and interactive media," in *Proc. of the Sixth Int'l Symposium on Information Processing in Sensor Networks, Massachusetts, USA*, Apr. 2007, pp. 380–389.

[7] C. Bahn, T. Hahn, and D. Trueman, "Physicality and feedback: a focus on the body in the performance of electronic music," *Proc. of the 2001 International Computer Music Conference, Havana, Cuba*, 2001.

[8] C. Cliff and H. Muller, "Personal position measurement using dead reckoning," in *Proc. of the 7th Int'l Symposium on Wearable Computers*, 2003, pp. 166–173.

[9] H. Leppakoski, J. Kappi, J. Syrjarinne, and J. Takala, "Error analysis of step length estimation in pedestrian dead reckoning," in *ION GPS*, 2002, pp. 1136 –1142.

[10] J. Ryan and C. Salter, " TGarden: wearable instruments and augmented physicality." *Proc. of the 2003 Conference on New interfaces For Musical Expression (NIME03), Singapore*, pp. 87–90, 2003.

[11] T. Todoroff, "Wireless digital/analog sensors for music and dance performances," in *Proc. of the 11th Int'l Conference on New Interfaces for Musical Expression (NIME), Olso, Norway*, 2011.

[12] G. Torre, M. Fernstrom, B. O'flynn, and P. Angove, "Celeritas: wearable wireless system," in *Proc. of the 7th Int'l Conference on New Interfaces for Musical Expression, New York, USA*, Jun. 2007.