Wireless Interactive Sensor Platform for Real-Time Audio-Visual Experience

Jia-Liang Lu
Dept. of CSE
Shanghai Jiao Tong University
800 Dongchuan Road
200240 Shanghai, P.R.China
jialiang.lu@sjtu.edu.cn

Da-Lei Fang
Dept. of Music Engineering
Shanghai Conservatory of Music
20 FenYang Road
200031 Shanghai, P.R.China
fangdalei@shcmusic.edu.cn

Yi Qin
Dept. of Music Engineering
Shanghai Conservatory of Music
20 FenYang Road
200031 Shanghai, P.R.China
qinyi@shcmusic.edu.cn

Jiu-Qiang Tang
School of Computer Science and School of Music
Carnegie Mellon University
5000 Forbes Avenue,
Pittsburgh PA 15213
jiuqiant@andrew.cmu.edu

ABSTRACT

WIS platform is a wireless interactive sensor platform designed to support dynamic and interactive applications. The platform consists of a capture system which includes multiple on-body Zigbee compatible motion sensors, a processing unit and an audio-visual display control unit. It has a complete open architecture and provides interfaces to interact with other user-designed applications. Therefore, WIS platform is highly extensible. Through gesture recognitions by on-body sensor nodes and data processing, WIS platform can offer real-time audio and visual experiences to the users. Based on this platform, we set up a multimedia installation that presents a new interaction model between the participants and the audio-visual environment. Furthermore, we are also trying to apply WIS platform to other installations and performances.

Keywords

Interactive, Audio-visual experience

1. INTRODUCTION

Recently, wireless communication technologies can be viewed as in an advanced stage. Meanwhile, motion sensors are becoming miniaturized. These two progresses of these two technologies empower the use of wearable sensors in interactive arts. A set of sensor systems [7, 11, 12] was applied in music and dance performances to enhance the interactions between performers and real-time scene effects. However, most sensor systems were developed to accomplish dedicated performances.

Rather than limited to one application, our system is designed as a basic platform on which artists could apply their arts preferences. In addition, WIS platform is an open platform that can be served with any wireless sensor nodes with corresponding receivers and send data to other program through standard OSC [2] packet. In this work, we decide to develop our own wireless on-body sensors as the capture devices of WIS platform.

With a growing interest in gesture recognition, several pioneer works have been proposed. In [7], a wearable sensor network has been improved for high-speed capture. In [12], Celeritas system consisting of with 6 DoF wearable sensors were used for dance performance. In [8], authors proposed a wireless system applying gesture following technology for music pedagogy. In French ANR Interlude project [1], MO interfaces were developed to explore novel gestural recognition approaches for musical expression. The MO handheld unit was designed to capture a wide range of gestures, and a real-time gesture and sound processing software was also developed in this work.

The main components of WIS platform, the real-time gesture capture system (on-body sensor nodes) and the actuator system (sound synthesis and interactive video), are implemented in Max/MSP with OSC support package. Thus, WIS platform is highly modulated and easy to be modified. The rest of the paper is organized as follows. Section 2 gives a detailed description of WIS platform with three main parts. We introduce an installation *Air-ergy* in section 3. The installation was exhibited in the 3rd Shanghai International Electronic Music Week [4]. It showed great stability of WIS platform. Section 4 concludes our work and gives future research directions.

2. WISP

WIS platform consists of three main parts: capture system, processing unit and display unit. The key idea of WIS platform is to implement an open architecture to organize these three parts. The artists can easily modify capture or display part of WIS platform according to working circumstances. For instance, we developed our own Zigbee compatible wireless on-body sensors and Arduino gateway broad for capture, but other commercial wireless sensor nodes such as MicaZ motes or IrisMotes can be perfectly compatible with WIS platform. The only criterion for evaluating the capture technique is how it adapts to the performing circumstances. This criterion should be also applied to the display part.

The system architecture of WIS platform is given in fig-

NIME'12, May 21 – 23, 2012, University of Michigan, Ann Arbor.

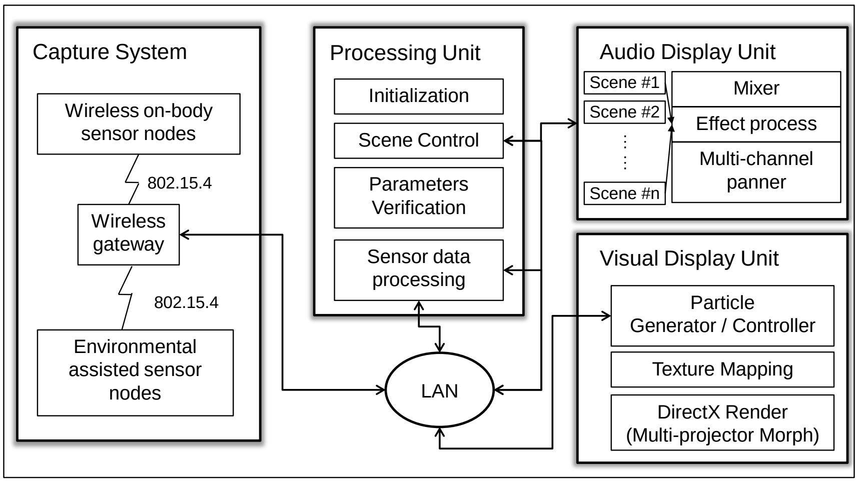

Figure 1: Architecture overview of WISP.

ure 1. The capture system is connected to a LAN via a dual-interface (IEEE802.15.4 and Ethernet) wireless gateway. The sensing data generated by wireless on-body sensor nodes are pre-processed and formatted to OSC messages by the wireless gateway, and received by processing unit eventually. In the processing unit, the sensor data are used to trigger the scene and generate control signals for audio and visual part. Moreover, user can create or modify toolbox module in processing unit to define the gesture patterns and build audio/visual effects matching his/her artistic contents in processing unit. The audio display unit consists of a mixer, effect processors and multi-channel panner modules. This unit interacts with processing unit to switch different scenes and also generates sonic elements according to the changes of sensor data. Normally, the processing unit and audio display unit would run on the same computer to minimize the influence of latency on the audio system. So that, instead of going through LAN, the control signal is directed sent trough two Max/MSP program locally. Meanwhile, visual display unit runs on the second computer for video particle generation and control, as well as texture mapping and rendering. While multi-projector system is used with this application, morph module should be activated in this part.

2.1 Capture System

The capture system includes several wireless on-body sensors and receivers. The receivers unpack the data package sent by sensors, process them and send them in OSC format to the processing unit via Ethernet.

The sensing task is secured by a SparkFun 9 Degrees of Freedom (9DoF) Razor IMU [3]. It is an attitude heading reference system 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 [6] is used as the wireless communication module whereby it sends and receives data via the TX/RX pins of 9DoF board. This portable device that includes the sensor board and the XBee module is supplied with a 900Am.h lithium battery. All the components of the wireless on-body sensor are settled in a $6cm \times 4cm \times 1.5cm$ translucent plastic box as shown in figure 2.

9DoF provides the acceleration data, angular rate and direction data in three dimensions, i.e. in X-, Y-, Z-axis. The data packet size needs to be reduced from 56 bytes to 23 bytes so that wireless communications junks are prevented. To do so, we compute the Quaternion from the 3×3 DCM

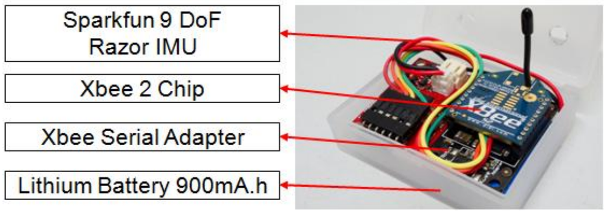

Figure 2: The wirelee on-body node is equipped with: a SparkFun 9 DoF Razor IMU; an XBee 2 802.15.4 RF module; an XBee serial adapter; and a 900Am.h lithium battery.

Matrix (Direction Cosine Matrix) and Tait - Bryan angles data. The package that the wireless on-body sensor sends out includes one one-byte Node_ID, three two-byte X-, Y-, Z-axis acceleration data, four two-byte Quaternion data and one one-byte Symbol Identifier ""*"" for data synchronization. The program refreshes data every four milliseconds on average, and therefore XBee modules could afford sending the packed messages with a speed of 57600 bps.

On the receiver part, the Arduino Mega Board, the microcontroller, works with the XBee Shield and the Ethernet Board to complete the procedure of getting data packages from portable device, processing data and sending data to computer via OSC. After Mega Board receives and unpacks the data package, it recovers the 3×3 DCM Matrix and Tait-Bryan angles data from Quaternion. To make the message more standard and efficient, 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 packets. 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, which has the type of float to the processing unit.

2.2 Processing Unit

The processing unit is composed of four modules:

1. *Initialization*: This module is responsible for parameter and state initialization as well as DSP loading. These operations are sequentially scheduled with fixed interval in order to avoid exceeding load on hardware. It also allocates dedicated memory space for the programs.

2. *Sensor data processing*: This module pulls the OSC packets from capture system through LAN and unpacks them. It processes the motion data captured by wireless on-body sensors and generates control data packets to audio and visual units. Typically, the yaw, pitch, roll and 3-D acceleration data are processed here. It is also in this unit that different pattern recognition models are implemented. In section 3.1, a step recognition module is further discussed.

3. *Scene control*: This module loads pre-defined scenes in terms of each particular performance environment and results of sensor data processing module. It sends OSC packets through LAN to control audio and visual display unit.

4. *Parameters verification*: This module serves as a tool to visualize the parameters and test them on-the-fly. Figure 3 gives one snapshot of this module, through which one could easily verify the motion data either in Tait-Bryan angles or DCM Matrix forms. We also develop a sonification sub-module to test the jitter of received messages. This module works with a background sound sample and jitter could be easily detected by these instable sounds.

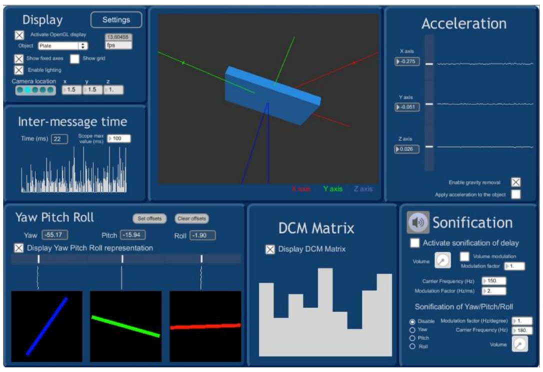

Figure 3: Parameter display and verification tool.

2.3 Audio-Visual Display Unit

As explained earlier, the audio display unit is implemented with the processing unit to maintain a low latency communication between each other, while the visual display unit is implemented in a separated computer.

A set of scenes is preset in audio-visual display unit, and each of them may consist of several modules, such as FM, preset mid/high, record play and additive. They are triggered by scene control module in the processing unit. The sound effect is related to the user motion, i.e. his movements. The multi-channel panner deals with the location of the user to help him feel the strong interaction between his movement and the environment. However, not all the effects are strictly associated with the sensor data; the artist could test and define the sound effects through the mixer module to realize his artistic design. In section 3.2, the technical details of an audio scene *Wind Scene* will be further discussed.

The visual display unit receives OSC packets from the processing unit in order to trigger video particle generation and control. In section 3.3, we introduce how this *bubble up* effect is generated after the step tracking function is triggered. The visual display unit is also responsible for texture mapping and rendering. Though the output is a single video source, in some circumstances multiple projectors are required for better experience. For instance, in our work *Air-ergy* two projectors are used. We developed a multi-projector morph module to handle it.

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

3. 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 was exhibited in the 3rd Shanghai International Electronic Music Week [4] in October 2011. This work emphasizes the interaction between the human body and the surrounding. To make use of WIS platform in a particular environment, we try to tie up the audio and video effects of this work with the body action of the user. The idea of this installation is to show how powerful the kinetic energy of people is through interactive sound effects and videos in a particular space. This closed loop system embodies a real physical world converted to a virtual world and then gets feedback from the latter. Table 1 lists the interactions implemented in *Air-ergy*.

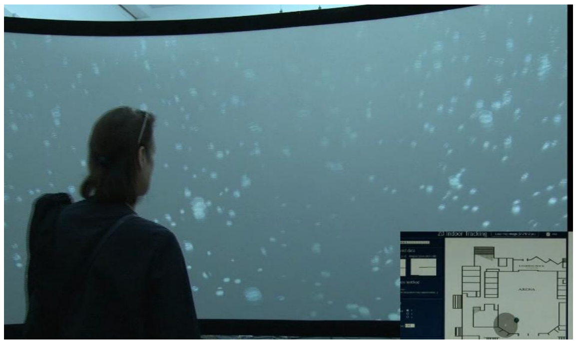

Figure 4: A scene of *Air-ergy*. The right-bottom snapshot shows the user localization.

Figure 4 show a scene of *Air-ergy*. Let us take three examples to introduce how to use WIS platform in a real artistic creation.

3.1 Step Recognition

The step recognition is essential for in-door tracking problem and widely use in Dead Reckoning solutions [9, 10]. In our realization, the step recognition is based on yaw, pitch and roll and 3-axe acceleration. First, all the raw data go through the AIO with offset for correction. By setting time switch, the unit is able to synchronize data from different on-body sensors for processing. Second, differential and envelope operations are applied on the data series to output a continuous and smooth acceleration variation.

The 3D acceleration is enveloped to get directions and a series of values. A threshold value should be defined to trigger the step. A number of tests are realized 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 next trigger. During the exhibition, it turns out that the threshold range is applicable for 90% of utilization.

3.2 Wind Scene Sound Synthesis

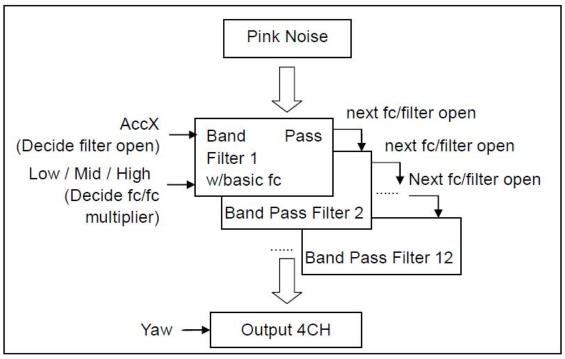

Figure 5: Wind scene diagram.

Wind Scene is a scene triggered by the rotation data of the sensor on the shoulder of the user. Since the designer intends to produce a breeze/wind sound effect. *Pink Noise* is used as the input signal for *Wind Scene*. In the first version, low pass filter with resonance is used. Though this is enough to simulate wind sound, we still implement a second version to involve more musical elements in the wind sound. 12 band pass resonators are used. The cutoff frequency of each new filter is the product of the cutoff frequency of the current filter. In this way the harmonic series would be obtained. The cutoff frequency is between 100 and 400 Hz, and the multiple factors vary from 1.05 to 1.30. The variation is triggered by a 3-level acceleration state output (low, mid and high) from sensing data processing part, and the volume of each resonator is controlled by the continuous acceleration output. Thus, the sound of a fast movement coming after a slow movement would be different from a pure fast movement. The position of the wind among 4 sound channels is controlled by Yaw data, so that users feel more comfortable to detect the relations between his/her movement and the sound position. The diagram of *Wind Scene* is given in figure 5, in which **fc** stands for filter cutoff frequency.

3.3 Bubble Up Visual Setting

Bubble up visual setting run on audio (Max/MSP) and video (vvvv [5]) computer. 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 would be sent to vvvv through LAN using OSC. The vvvv would add a fast up ramp to the height of the particles. If no more steps were received, the height would fall slowly. If more steps were received, the height would be added cumulatively. The height could also be sent back to Max/MSP through LAN using OSC. This parameter is used to determine the volume of each partial in additive synthesis. Each partial has the same frequency (pitch) corresponding to each band pass resonators. So the additive synthesis has the same harmonic series as *Wind Scene*, but with more pure tone. We use X position to cross fade between *Wind Scene* and additive. It means a cross fade of reality and virtual world.

4. CONCLUSIONS

The WIS platform is an open platform designed for interactive artistic creation. We present a capture system with our on-body sensors. Meanwhile, other wireless sensor systems or motion capture systems can be easily embedded to WIS platform. The display part can be tailored with artists' creation to help express in their own artistic language. *Air-ergy* is a good example built on WIS platform.

We are currently working on the WIS platform to include more environmental sensors and actuators to enrich the expressions of this platform. For instance, a more accurate in-door localization system is being developed on the WIS platform to improve the audio-visual experiences.

5. ACKNOWLEDGMENTS

We would like to thank Christophe Lebreton at GRAME and Prof. Qiang-bin Chen at SHCM for a number of suggestions and discussions on this project as well as Quentin Pyanet at IRCAM for his contribution. 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.

6. REFERENCES

[1] Interlude project, ANR: http://interlude.ircam.fr.
[2] Open Sound Control: http://opensoundcontrol.org/.
[3] Sparkfun 9 dof razor imu: http://www.sparkfun.com.
[4] The 3rd Shanghai International Electronic Music Week: http://www.emusicfestival.cn/.
[5] vvvv - a multipurpose toolkit: http://vvvv.org.
[6] Xbee 2 802.15.4 oem rf module: http://www.digi.com/.
[7] 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*, pages 380–389, Apr. 2007.
[8] F. Bevilacqua, F. Guedy, N. Schnell, E. Flety, and L. N. . I. N. 2007. Wireless sensor interface and gesture-follower for music pedagogy. In *Proc. of the 7th Int'l Conference on New Interfaces for Musical Expression, New York, USA*, 2007.
[9] C. Cliff and H. Muller. Personal position measurement using dead reckoning. In *Proc. of the 7th Int'l Symposium on Wearable Computers*, pages 166–173, 2003.
[10] H. Leppakoski, J. Kappi, J. Syrjarinne, and J. Takala. Error analysis of step length estimation in pedestrian dead reckoning. In *ION GPS*, pages 1136 –1142, 2002.
[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.