MusicGlove: A Wearable Musical Controller for Massive Media Library

Kouki Hayafuchi
Dept. of Intelligent Interaction Technologies
University of Tsukuba, Japan
kouki@ai.iit.tsukuba.ac.jp

Kenji Suzuki
Dept. of Intelligent Interaction Technologies
University of Tsukuba, Japan
kenji@ieee.org

ABSTRACT
This research aims to develop a wearable musical interface which enables to control audio and video signals by using hand gestures and human body motions. We have been developing an audio-visual manipulation system that realizes tracks control, time-based operations and searching for tracks from massive music library. It aims to build an emotional and affecting musical interaction, and will provide a better method of music listening to people. A sophisticated glove-like device with an acceleration sensor and several strain sensors has been developed. A realtime signal processing and musical control are executed as a result of gesture recognition. We also developed a stand-alone device that performs as a musical controller and player at the same time. In this paper, we describe the development of a compact and sophisticated sensor device, and demonstrate its performance of audio and video signals control.

Keywords
Embodied Sound Media, Music Controller, Gestures, Body Motion, Musical Interface

1. INTRODUCTION
The listening habits of people have been dramatically changing in recent years because people can bring massive libraries of digital music with small portable music players like *iPod*. In this situation, it is required to build a new system that people can find desired music from enormous numbers of digital media data. Many methods have been proposed to address this problem, for example, a method of graphical visualization to organize lucidity music libraries were suggested [1].

In order to allow users to provide more degrees of freedom, a variety of physical input devices such pen tablet, dial or glove shaped interface are commercialized and widely being used in various fields. In addition, intuitive input devices such as touch and haptics have been attracting social attention. To date there are a number of researches about the gesture interface for music [2, 3]. For example, systems for musical controller that are able to control electronic devices by using simple finger gestures were proposed in several studies such as FreeDigiter [4] and Ubi-Finger [5]. Recently, a system that controls the tracks like a DJ has also been proposed in [6].

In this study, we focused on a sophisticated interface that enables users to control the sound and music in intuitive and efficient manners. The glove-like input device is one of the conventional interfaces for human-computer interaction. The developed system, *MusicGlove*, has a role of interactive music player and explorer, which performs tracks control, time-stretching of audio and video signals, and information retrieval from massive music library as a result of hand gestures and body motion recognition. In particular, time-based multimedia interaction including audio and video signals becomes popular in these years. To date some rich time-based operations have also been proposed for different applications [7].

Gestural control allows the real-time control and high affinity for expressive performance. The target of *MusicGlove* project includes an all-in-one device that can control music and generate audio sound by itself. In this instance, the user can listen to music that is produced by the wearable device. Therefore, people can enjoy musical control by using *MusicGlove* at any time and place, even in transit or on the walking. The developed device and system can contribute to new listening style of music from massive music library.

2. SYSTEM OVERVIEW
In this study, the musical control is mainly divided into two functions; tracks control and audio/video time-based control. The tracks control is regarded as common manipulations of music player such as play, stop and skip to the next music. In addition, a function of searching tracks from music library is implemented, which is similar to the manipulation performed by a DJ. On the other hand, audio/video time-based control is regarded as signal processing which directly controls sound waveform such as change of tempo or addition of tonal effect, which is a resemblance function of audio signals.

2.1 Hardware Overview
The overview of the developed glove-like sensing device is shown in Figure 1. The device consists of one 3-axis acceleration sensor, 4 strain sensors, 1 microprocessor for signal processing and control, Bluetooth wireless module, a portable music player that is used in the *Wearable Music* play (all-in-one application), and a battery. The measurement range of the acceleration sensor is from $\pm$10g. As the sensor is fixed on external side of the wrist part, X, Y, Z-axis are also fixed at a given position. Four strain sensors are mounted at upside of index finger and mid finger, and also inner and exterior side of wrist as illustrated in Figure 2. The strain sensors provide analog values of bending of each position. The glove like device has light weight and

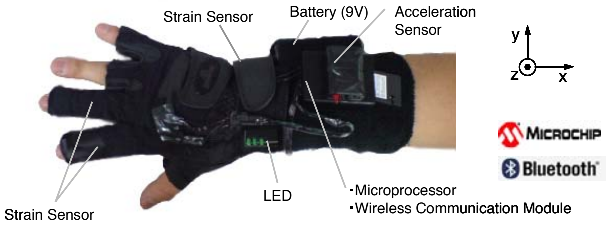


Figure 1: The overview of the MusicGlove input/output device

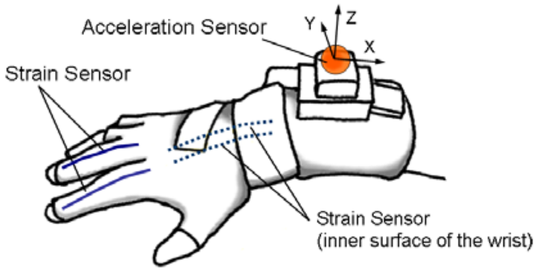


Figure 2: Arrangement of sensors

Table 1: Gestures and Functions

Mode	Function	Gesture
Tracks control 	Play Pause Volume up/down Next/Previous music	Bend index finger Make a fist Wrist up/down Pointing right/left
Sound control 	Fast forward Fast rewind Tempo up/down Scratch	Wrist rotation Wrist rotation Wrist up/down Scratch motion
Searching tracks 	Searching Shuffle	Hand Motion *Acceleration used to be forward to next music

is designed to satisfy the minimum requirement for musical control. The arrangement of sensors as illustrated in Figure 2 is determined as a result of preliminary experiments. The microprocessor is used to obtain sensor data, to perform gesture recognition, and then to transmit processed data to the wireless module. The device communicates with the host computer via Bluetooth. In addition, the music player can be connected to the microprocessor via dock connector port. In the all-in-one application, the microprocessor generates control signals for the music player.

3. GESTURES AND FUNCTIONS

Figure 3 shows the flow chart of data and signal processing. The acquired sensor data from the acceleration sensor and strain sensors can be transmitted to the host computer at 60Hz. Controls of sound effect, tracks and video are executed according to predetermined gestures in the host computer . The audio signals are presented by the loudspeakers or attached headphones.

The classification of performed gesture is described as follows. In particular, we focus on the hand posture based on index finger and wrist. All hand gestures are classified mainly into 3 categories according to the measured data by equipped sensors. Different category of gestures has a different role. While the user takes the particular posture, a predetermined style of musical controller is activated. While the user is holding the initial hand posture, the user is able to realize different control methods. When a control methods will be enabled, other control methods are disabled at the same time. This processing enables the system to avoid undesired behavior of the system, and it also helps to increase the number of usable gesture. Due to analysis of hand gesture for musical control, the user, therefore, do not need to have training and consciousness about the switching of the method.

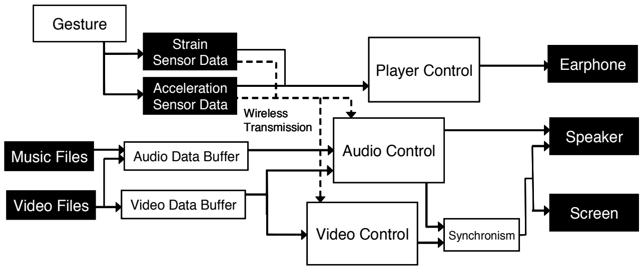


Figure 3: Flow diagram

3.1 Classification of the performed gestures

Max/MSP and EyesWeb[8] software architecture are used for controls of audio/video signals and data receiver. In this study, Eyesweb is solely used as a data receiver and transmitter to Max/MSP. However, the system, can be extended to incorporate with camera-based recognition and other performance installations. The media library is imported to a buffer in advance, and Max/MSP then receives sensor data from the glove device on a steady basis. A hand shape is recognized in a successive manner by means of some filtering techniques based on the sensor model.

The interaction is classified mainly into 3 styles: (1) Air disc jockey (DJ), (2) gestural conducting, and (3) wearable music. These include three common modes for musical control: i) tracks control mode, ii) sound control mode, iii) search for audio tracks mode, Each mode is continued until the user changes mode by changing his/her hand posture.

3.2 Three styles of interaction

(1) Music Player control - ***Air DJ*** : As shown in figure4, the system allows users to control musical features at the external computer, and audio sounds are produced from loudspeakers installed in the surrounding environment. The user wears a wireless sensor glove, and sensory data are transmitted to the computer. When the user does some hand gestures or body motion, the computer translates them into musical control and produce audio sounds.

(2) Gestural conducting - ***Air Conductor***: We explain about a gestural conducting which is regarded as a time-based interaction by using audio/video signal rendering. As illustrated in figure 3, the system allows users to control not only audio signals by user's gesture but also videos associated with music like a conductor. The performance video such as orchestra's playing or musical dance is controlled in accordance with the user's conducting behavior. The audio/visual performance provide high immersive environment.

A video sequence with polyphonic audio such as orchestra or quartet play is used in this study because the music quality is less tolerated compared to the video sequence with song or speech. By utilizing the acceleration sensor values, the control of audio volume and time-stretching of audio/video signals are carried out in real time. A modified phase vocoder algorithm with noise reduction technique is applied to achieve the time-stretched audio signals. Regarding the visual signal processing, a simple speed control of playback is used. The audio volume is also controlled by the accumulated value from the triaxial acceleration sensor. The tempo control is based on extracting beats as a inflection point of Z-axis acceleration value.

Figure 4: Air Disc Jockey (DJ)

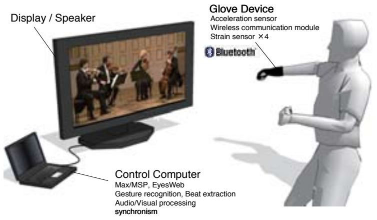


Figure 5: Air Conductor

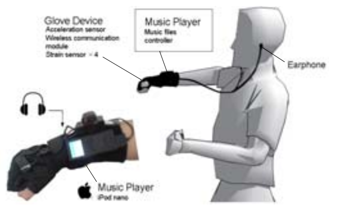


Figure 6: Wearable Music

When a user begins to swing his/her arm with a constant tempo and keeps three times within a certain tempo change, this conducting interaction is initiated.

(3) Wearable music: This is another application of the developed *MusicGlove*. A portable music player can be attached with the glove device, and the users are able to control music player by his/her gestures as illustrated in Figure 6. This enables the system to be stand-alone, and users can listen to music via headphone or earphones that are directly connected to the developed device without any other equipments. The embedded microprocessor produces a control signal to the player such as: play or stop tracks, skip to the next or previous music, fast forward and rewind, and volume control by means of acceleration and strain sensors. The predetermined gestures are the same as ones in the *track control* mode.

3.3 Common Control Mode

i) Tracks control mode: The control of tracks is done according to the hand posture, which includes the following functions: play, stop, skip to the next music, back to the previous music, and volume control. This mode is initiated when the user stretches index finger, and the hand is then shaped like pointing to the air. For example, playing music is done by bending down index finger a little. To skip to the next or previous music, the user should make pointing to the left or right by index finger. The volume control is done by using data from the acceleration sensor, in particular, the value of Z-axis. When the sensory value exceeds a predetermined threshold level, the amount of volume increase and decrease are done.
ii) Sound control mode: This mode is initiated by bending the wrist toward the palm. In this mode, the user is able to control audio sound features. The predetermined controls include: fast forward and rewind, changing tempo without time-stretching, and scratch. The acceleration sensors of X and Y axes caused by circular motion of hand are used for fast-forward and fast-rewind operations. A clockwise rotation corresponds to fast-forward, while fast rewind corresponds to the counterclockwise rotation. Regarding the tempo change, solely audio resampling by changing the speed results in the pitch-shifting effect. Scratch play like DJ is available as a result of filtering value from X-axis acceleration sensor. In order to prevent a false operation, the user must not generate acceleration signal in Y and Z axes when during scratch motion.
iii) Search for audio tracks mode: This mode is initiated by bending the wrist toward the back of hand. In this mode, the user is able to search audio tracks by grasping gesture at the air. In this mode, audio tracks presented to the user in a successive manner. Search for audio tracks is regarded as repetition of trial and error to choose a music (or album) from massive libraries. The system will present the first part of a track in a music library at each step of search successively according to the user's searching motion. The searching motion is regarded as a simple hand motion which is detected based on the accumulated value of the acceleration sensors in all-axes. The user listens to audio data, and acts a grasping motion of playing music when she/he finds a desired track or library. Grasping motion is detected based on the strain sensors. In addition, waving user's hand is regarded as "shuffle search." The user is able to choose a media library or tracks in a random manner. These manipulations provides users with intuitive search like grasping a music at the air.

4. PERFORMANCE DEMONSTRATIONS

In this section, we show some performance demonstrations with the developed device. We first describe time-series examples of sensors and gesture classification. Next, a example of waveform regarding the scratch motion will be shown with the spectrogram.

Behavior of Sensors: We carefully arranged the sensors' location and filtering algorithm in order to achieve the gesture classification. An example of time-series sensor data obtained from the strain sensors during a *tracks control* gesture are shown in Figure 7. The *tracks control* mode begins at the point A when the user took the predetermined posture. At the point C and D, the user makes gestures to be forward to next track, and back to previous track, respectively. At the point B and E, the user plays and stops the audio track. The Figure 8 represents gestures of *sound control*. This mode begins at the point A and the user made a gesture of scratch motion at B, C, and D. On the other hand, at two points E, scratch effect is not occurred. The acceleration value in Y axis exceeds the threshold because the user made different gestures from previous gestures. The system successfully distinguished the intended gestures for musical control. The Figure 9 represents a time-series sensor data during a gestural control of "search of audio tracks." This control is initiated when the user extends his wrist and the value of strain sensor exceeds the predetermined posture, indicated by A in the figure. The search of audio tracks is occurred according to the accumulated value of the summation of the triaxial acceleration as indicated by the points B, C and D. The user then makes *play* gesture at the point E, and the chosen track begins to play.

Waveform of audio output signals: We examined the quality of audio output signals modified by the user's gestural control. In this section, we particularly focus on the *Air DJ* style and present the waveform of audio output signals during scratch motions. The Figure 10(a) shows the spectrogram during a scratch motion measured by the developed device. The Figure 10(b) shows the spectrogram during a scratch motion which is performed by using a commercially available digital turntable system. The region indicated by dotted line represents the period of the scratch motions. There appears to be particular spectrum features because the scratch performs as a short fast-rewind and the particular frequency spectrum can be seen.

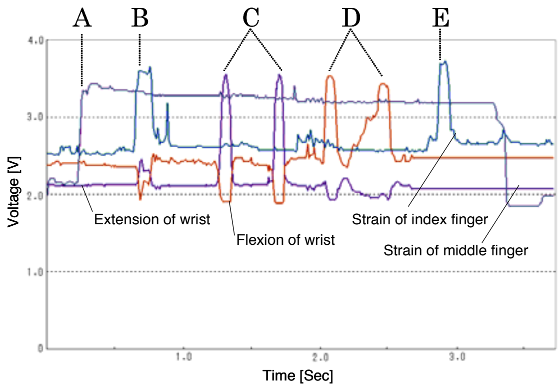


Figure 7: Tracks control mode

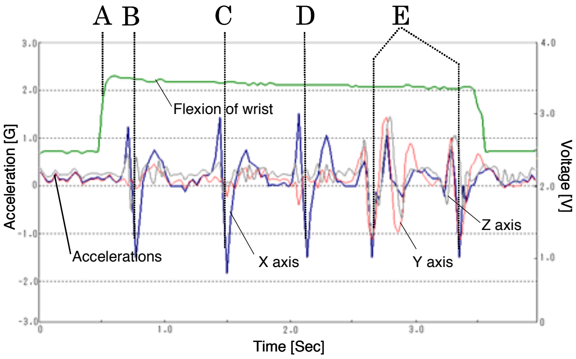


Figure 8: Sound control mode

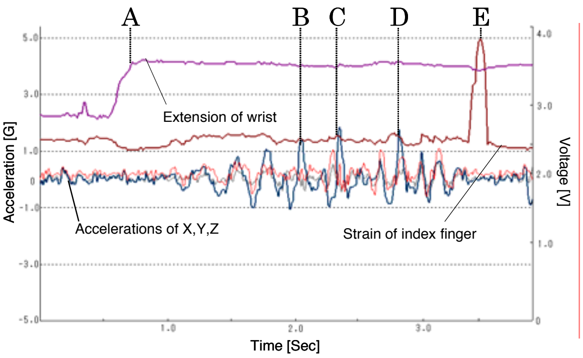
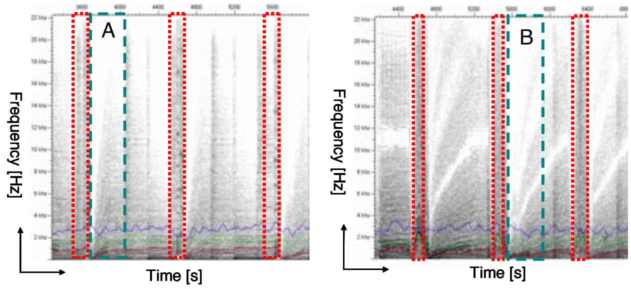


Figure 9: Search for audio tracks mode

In addition, a distinguishing spectrum feature can be seen in the region *A* of Figure 10(a). That wave pattern is similar to the region *B* which is seen in the Figure 10(b). The wave pattern is generated by a particular rotation of the turntable after the scratching action. That can be said that we have duplicated the behavior of digital turntable system. The waveform of audio output by the developed system has quite similar characteristics with one by the digital turntable system in terms of temporal transition. The response to gestural control is enough fast to realize natural sound effects by scratch.

5. DISCUSSION AND CONCLUSIONS

In this paper, we proposed a method of active music listening for massive media library. Different styles of musical interaction are realized by using the sophisticated glove-like input device. The developed system allows humans not only to control audio and video signals but also to search audio tracks from massive media library by hand gestures and body motion. In addition, it is possible to listen to music



(a) MusicGlove (b) Turntable

Figure 10: Spectrogram comparison

from the device itself. In recent years, we are able to access to enormous amounts of media data with a portable device. A better style of interaction than conventional control by means of buttons or cursor keys is definitely required. Some media players that accept haptic control on the LCD are already commercialized.

The challenges to associate human body motion with musical control have a quite long history and will be continued with the advancement of sensor and wireless technologies. We have been investigating the embodied musical interaction by means of expressive behavior and gestures [9], namely *Embodied Sound Media*. The next stage of this research includes the implementation of physiological sensors in order to extend the control capability by users. The effort to achieve a non-intrusive and transparent interface is continued, which allows humans to make more natural, exciting and artistic performance with machines.

Acknowledgement
A part of this work is supported by Japan Science and Technology Agency (JST), CREST "Generation and Control Technology of Human-entrained embodied Media."

6. REFERENCES

[1] M. Torrens, P. Hertzog and JL Arcos. Visualizing and Exploring Personal Music Libraries. *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR2004)* , 2004. Avaiable online.

[2] M. Wanderley, Gestural Control of Sound Synthesis. *Proc. of the IEEE*, 92(4), pp. 632-644, 2004.

[3] K. Ng, Music via Motion: transdomain mapping of motion and sound for interactive performances, *Proc. of the IEEE*, 92(4), pp. 645-655, 2004.

[4] C. Metzger, M. Anderson and T. Starner, "FreeDigiter: A Contact-free Device for Gesture Control," *Proc. of the Intl. Symp. on Wearable Comput.*, pp. 18-21, 2004.

[5] K. Tsukada, and M. Yasumura, "Ubi-Finger: a Simple Gesture Input Device for Mobile and Ubiquitous Environment," *Journal of Asian Information, Science and Life(AISL)*, Vol.2, No.2, pp. 111-120, 2004.

[6] K. F. Hansen and R. Bresin. DJ Scratching Performance Techniques: Analysis and synthesis, Proc. Stockholm Music Acoust. Conf., pp. 693-696, 2003.

[7] E. Lee, et al., Toward a Framework for Interactive Systems to Conduct Digital Audio and Video Streams, *Comput. Music J.*, 30(1), pp.21-36, 2006.

[8] Eyesweb, InfoMus Lab, [web] http://www.infomus.dist.unige.it/EywMain.html

[9] K. Suzuki et al., Robotic Interface for Embodied Interaction via Dance And Musical Performance, *Proc. of the IEEE*, 92(4), pp. 656-671, 2004.