KEER2010, PARIS | MARCH 2-4 2010

INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2010

GESTURAL INTERFACE TO EXPLORE AUDIO LIBRARIES AND ENHANCE MUSICAL EXPERIENCE

Kouki HAYAFUCHI * [a], **Kenji SUZUKI**[a]

[a] *Dept. of Intelligent Interaction Technologies, University of Tsukuba, Japan*

ABSTRACT

In recent years, the proliferation of online music distribution services for personal computers and portable music players has led to an increase in the amount of available music-related information. Accordingly, the computation and bandwidth loads for music retrieval and database manipulation by listeners has also increased considerably. In this paper, we propose a wearable interface that enables users to smoothly select music tracks and control the audio while listening. Unlike a precise music search, the developed interface allows the users to explore and control music by making gestures, i.e., with bodily movements, in real time. In particular, it aims to support the seamless selection of the user's favorite tracks at a specific time without specific data retrieval but by following the user's vague musings. The user can switch between tracks by performing a swishing motion in the air (similar to fumbling actions). We have developed a music exploration system that has a wearable interface. The system comprises a glove-like device for gesture recognition, an external computer for processing audio signals, and speakers. An acceleration sensor and several bend displacement sensors are installed in the glove-like device. The intuitive operation and the selection of music by fumbling, swinging, scratching, and grasping motions can enhance the musical experience. Furthermore, the developed system makes us enjoy these operations themselves without the necessity of holding any devices in our hands while listening to music.

Keywords: *Music technology, Gesture recognition, Wearable Interface*

1. INTRODUCTION

The recent proliferation of online music distribution services for portable music players and the associated network technology has led to a great increase in the vlume of available music-realated

* **Corresponding author:** kouki@ai.iit.tsukuba.ac.jp

information. Such music-related information retrieval systems play an important role in allowing users to listen to and selecting from a large collection of music. However, the computational and bandwidth loads for music retrieval and database manipulation have also increased concomitantly. In such a situation, without choosing specific tracks, users often listen to tracks without any regard to the order. This is a well-known style of listening, for instance, *shuffle play*. A music interface that lets the users explore for the desired music is now required [1].

Moreover, a desired interface that users can treat intuitively has been paid considerable social attention. In particular, a user interface that considers bodily movements, such as a touch pen and a wearable computer, and the concept of gestural control have attracted significant attention. An intuitive interface with which everyone can simply control the system and which everyone can understand easily is requested. For instance, when users want to perform an operation of ”throwing out,” it is easy for the operation unlike the command input with the button it. People simply shake their arm, and throw out to understand the input method. Therefore, it is easy for people to understand the system’s behavior intuitively when the requisite user action is similar to one that suits the related physics phenomenon.

Thus far, several types of interfaces using bodily movements have been reported. For instance, FreeDigiter [2] allows the users to control the audio-visual apparatus with a finger. Tsukada *et al*. proposed a system that can operate electronic equipment on the basis of hand and finger gestures [3]. Moreover, different types of interfaces to control electronic equipment and sound using bodily movements have been reported in the field of music technology. For instance, Kia C. Ng proposed a system that associates effect control and bodily movements to the sound of musical instruments [4]. However, the literature mainly focus on audio signal control in terms of the relation between gestures and the played music. Instead, this paper proposes that music be treated a controlled object, in addition to the control of sound and performance, as an integral part of the music listening experience. Thus far, systems that re-create a DJ whom K. F. Hansen presented for instance have been proposed for music control [5].

We have been exploring a novel media technology, called Embodied Sound Media; it generates and controls sound on the basis of a free bodily movement [6], [7]. In this paper, we propose a system that explores music by using bodily movements such as ”fumbling for your favorite music” in real time for the control of sound and music. Music control when roughly divided into two means sound control and music control. Sound control is related to the signal processing of the audio and controls the acoustic waveform directly. On the other hand, music control is a track-based control of playing and selecting tracks. In this paper, we investigate a method of exploring and changing music tracks like DJs do, in addition to the current sound control. This presents a new style of listening to music and a simple and intuitive operation using bodily movements at the same time.

Gestures are not always suitable to be associated with indexes like text information or commands but are quite suitable for the volume control of music, exploratory actions, and so on. Using bodily movements, we can enable a variety of operations to control audio tracks. Furthermore, compared to the traditional button operation, the proposed method provides a new style of listening to music.

2. SYSTEM OVERVIEW

2.1. Interaction design

This system allows users to explore music tracks from a massive audio library by using arm and hand gestures. The user performs exploratory actions to look for a music track using a glove-type wearable device, called MusicGlove, by moving his/her hands in the air. The movement makes the user appear as if he/she were fumbling for tracks. Music tracks are considered to be wafting in air, and we are made to evoke the movement that the user is gripping them. The user fumbles for music tracks with his/her hands, grips them, and explores for the music that he/she wants to listen to at that time. The device allows the user to conduct several types of exploratory actions, such as looking for a track, looking for a playlist, and shuffling. The operation includes some basic controls such as play, stop, and forward/backward. Moreover, audio controls and modifications, such as changing the tempo of the tune, scratching, and adding several sound effects of the tone to the tune, can be carried out. Because of the nature of wearable devices, this system can be used while walking, jogging, or even playing musical instruments. Moreover, applications with visual effects are also possible.

2.2. Overview of the system

Figure 1 shows the outline of the developed system. This system is composed of a glove-type device, a computer for music and audio processing, and loudspeakers. The acceleration sensor and the bend displacement sensors are installed in the glove and used for gesture acquisition. The device acquires the user's operation and transmits data to the computer via Bluetooth communication. Max/MSP is used for the analysis of the sensor data and the music control. The user first wears the glove and performs a regulated gesture. Max/MSP always processes sensory data from the glove. The threshold of each piece of sensory data is set to all functions in the system as the beginning trigger.

2.3. Hardware implementation

Figure 2 shows the overview of the glove-type device used with this system. The developed device consists of a 3-axial acceleration sensor, four bend displacement sensors, a microcomputer, a Bluetooth module, an LED that displays the current state of the device, and a battery. The

Figure 1: System Overview

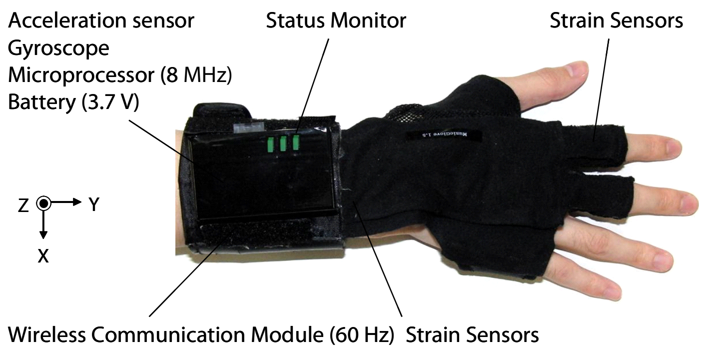


Figure 2: MusicGlove Overview

acceleration sensor with the range of ± 10 [G] is installed on the wrist. The direction for the backward/forward action is the X-axis, the right/left direction is the Y-axis, and the upper/lower direction is the Z-axis from the body to the arm. The bend displacement sensors acquire the bending or extension of the index finger, the middle finger, and the wrist. Each bend sensor detects only one direction of bending within a range of approximately 120 °. The arrangement of these sensors is determined on the basis of the result of a prior experiment. A microprocessor is used for processing all sensory data. The weight of the entire interface is approximately 138 [g], and the sampling data is fed to the external computer via wireless communications at approximately 60 [Hz].

3. GESTURAL CONTROL OF MUSIC

3.1. Flow of data

Figure 3 shows the flowchart of the audio/sound control. The sensory data obtained with each sensor in the MusicGlove are converted to 12-bit digital data and transmitted to the external computer. In the current implementation, 50 songs were selected from five different genres. The system gathers all 14 sensor values (3 for acceleration and 4 for bend displacement for each hand). Predetermined thresholds are used for the gesture recognition on the basis of the hand posture. The software executes the control and processing on the basis of a recognized gesture as a cue in real time and then generates the audio. The sound signal is presented to the user through the

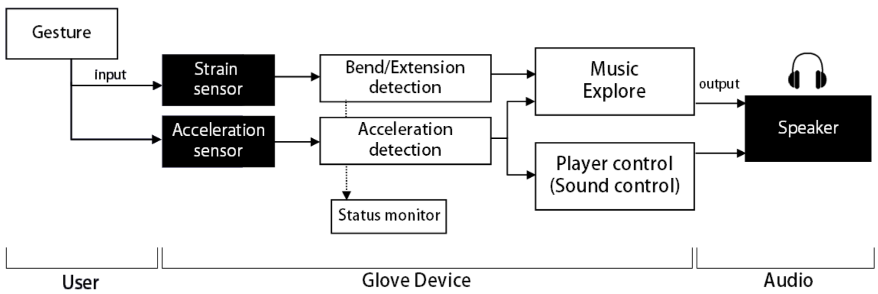


Figure 3: Flow Diagram

loudspeaker.

3.2. Gesture cues

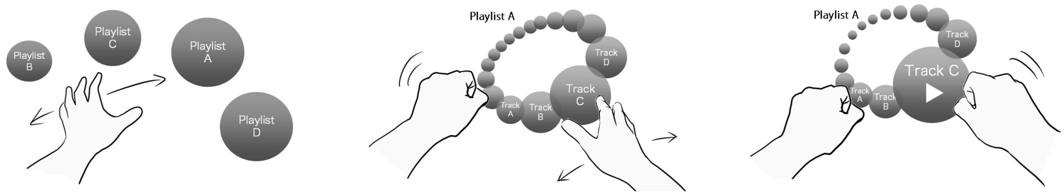
Table 1 shows the relation between the system's functions and their corresponding gestures, which are designed from the user study. Moreover, in order to prevent the misclassification of gestures during system operation, hand postures are divided into three categories. Particular postures are determined as the trigger of each category, and each posture is recognized on the basis of the installed bend displacement sensors. Each posture activates a different mode.

Table 1: Gestures and Functions

State	Function	Gesture
Music Explore	Switch playlist	Fumble one hand
(Extended wrist posture)	Switch track	Fumble one hand while grasping the other
	Switch continuously	Fumble one hand quickly while grasping the other
	Switch by shuffle	Scramble motion
Player Control	Play/Pause	Bend the index finger
(Pointing posture)	Volume up/down	Move wrist up/down
	Next/Previous tracks	Point right/left
Sound Control	Scratch	Scratch motion
(Flexed wrist posture)		

3.3. Music explore

There are four types of interactions designed as the switch for music tracks: (1) Switch each song, (2) switch continuously by an automatic operation, (3) switch playlist, and (4) switch by shuffling. The concept of exploratory action for exploring music is illustrated in Figure 4. The user can execute the switch of each playlist by moving one of his/her hands in the air; the music tracks in the playlist can also be switched by keeping the left hand steady in the air and moving the right hand simultaneously. At this time, music tracks in the playlist change one after another automatically according to the velocity of the arm gesture. Moreover, tracks can be shuffled when the user moves both her hands quickly. By these methods, the system allows the user to select a song as if he/she could find and grip the desired track.
The user first explores the playlist by swinging his/her left hand. Next, a playlist is gripped by the left hand and then the right hand is moved so that the user can explore the track in the selected playlist. Finally, the gripping gesture is carried out to select a track. The playlist has been switched by one hand, and the tracks in the playlist have been switched by using the other hand. This is inspired by an action to support an object by the left hand and to do a task by the right hand (for right-handed people) in day-to-day life. As a result, the user can explore for music naturally as if the music group were floating in the air and can fumble with the tracks. In the music search, when the gesture cue (extension of the wrist) is detected by the installed sensors, the sum of the absolute value from the 3D acceleration sensor is used for the calculation of the arm momentum. Moreover, the gripping operation is recognized by utilizing the bend sensors on the finger. The automatic flow of music tracks changes at shorter time intervals when the hand is moved faster. The switching time T of the music tracks is set to be proportional to the velocity V of the hand motion.



(a) Explore playlists (b) Explore tracks (c) Select a track

Figure 4: Concept of Music Exploring

$T = b(a - V)$
where a is a constant with dimensions [L][T^{-1}] and b is a constant with dimensions [L^{-1}][T^{2}]

In this operation, there are two types of presentments for the user while he/she explores music: the transition of music tracks and the sound effect as a feedback to a specific operation. The transition of tracks contributes to a smooth sound change by cross-fading the volume of each track. On the other hand, we have designed a ring sound that is played when a track is selected or when an automatic change is begun in order to provide feedback for the user operation.

3.4. Player control

In the Player Control mode, the system is used as a general music player, 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 his/her index finger and the hand is shaped as if the user were pointing to the sky. Particular gestures are prepared for the player-like control, for example, music is played by bending down the index finger a little. To skip to the next or previous music, the user has to point to the left or the right using his/her index finger. Volume is controlled by using data from the acceleration sensor, in particular, the value on the Z-axis. When the sensory value exceeds a predetermined threshold, volume is increased or decreased by a particular value.

3.5. Sound control

This mode is initiated by bending the wrist toward the palm. In this mode, the user can control the audio features, in particular, the scratch operation. The scratch is executed by quickly moving the wrist back and forth as if the user were rubbing a record. Scratch play is available as a result of the filtering value from the X-axis acceleration sensor. In order to prevent a false operation, the user must not generate an acceleration signal in the Y- and Z- axes during the scratch motion. Additionally, in order to generate the sound approximated to a more actual scratch, the playback speed will have been intermittently changed in a short time after the scratch.

4. EVALUATION EXPERIMENT

4.1. Arrangement of sensors

IIn order to investigate the arrangement of the sensors, we performed a user testing. Further, in order to confirm the usability in day-to-day life, the hand operation in actual life was recorded. The subject was asked to stay indoors with the glove equipped with the sensors. Only one subject was observed for this experiment. The graph shows of the posture of each incidence of the hand that appears during the experiment. Figure 5 shows the classified hand gesture based on the installed sensors. As shown in the figure, the occurrence rate of the gesture cues is lower than that of the other gestures. The gray bars show the gesture cues. Based on this result, we verified that the predetermined gesture cues were appropriate for executing each function.

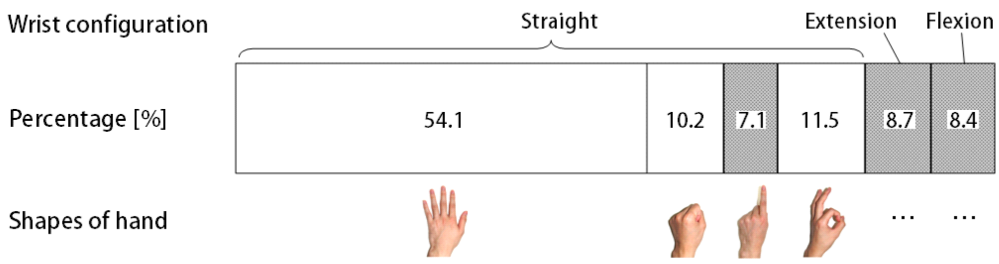


Figure 5: Percentage of Hand Shapes

4.2. Basic characteristics of performance

We conducted an evaluation experiment in order to clarify the characteristics of the performance. This experiment was aimed at clarifying the location as the interface of the system by comparing the amount of bodily movement with respect to the other interfaces. We compared the operation of Music Explorer to that of a commercially-available gestural interface, Nintendo Wii Remote. The arm momentum and the repetitions in the exploratory actions were recorded, and the operations carried out by the subjects were examined. The conditions of the experiment were as follows:

- Subjects: 5 (22-24-year-old males)
- Number of tracks: 50
- Number of playlists: 5
- Test duration: 10 min
- Used devices: MusicGlove, Wii Remote
- Measured values: magnitude of the arm motion and repetitions of exploratory actions

Table 2: Operations by Wii Remote

Function	Operation
Choose a playlist	Swing Wii Remote
Search for a track	Swing Wii Remote while pushing A button
Switch continuously	Swing Wii Remote fast while pushing A button
Select a track	Push B button

Figure 6 and Table 3 show the experimental result, which is averaged across five subjects. Figure 6 shows the difference with regard to the magnitude of the arm motion. It can be seen

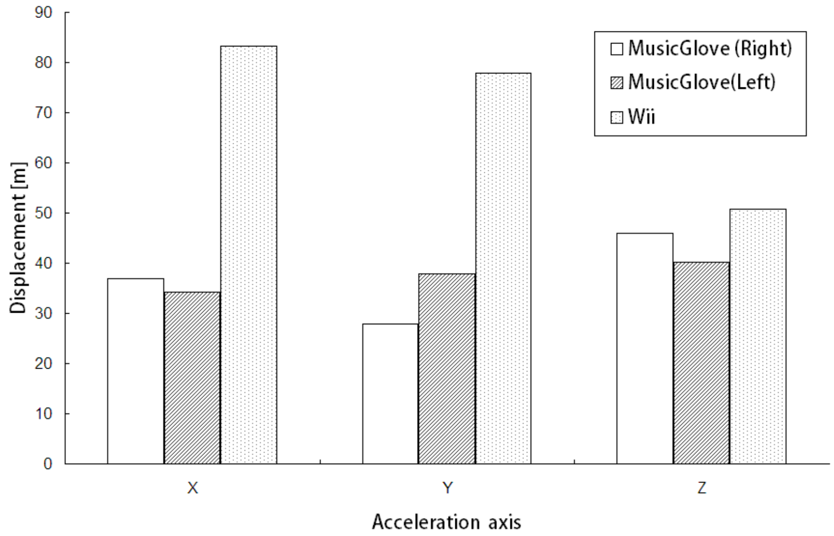


Figure 6: Mean number of Displacement Magnitude of Arm

that the magnitude of the motion in the case of Wii Remote is larger than that in the case of MusicGlove. On the other hand, the exploratory actions for the playlist and music tracks in the case of MusicGlove are not as often as in the case of Wii Remote. The amount of movement with MusicGlove is almost half of that with Wii Remote. We consider that this is because the subjects often either swung their hands slowly or fumbled. Moreover, in the repetition of the exploratory actions as in the case of MusicGlove, the playlist change is less and track change is more than that in the case of Wii Remote. The playlist change is caused by the left hand in the case of MusicGlove. On the other hand, the Wii Remote controller is held with the dominant hand, and the subjects use their left hand for assistance as in day-to-day life.

Table 3: Mean Number of Executed Functions

	Choose a playlist	Search for a track	Switch continuously	Total track explorings
MusicGlove	15.6	26.6	6.4	68.4
Wii Remote	25.4	19.2	3.8	81.4

4.3. NASA-TLX

In the above experiment, the subjects were asked to fill out the NASA-TLX [9] questionnaire after each session in order to examine the body and mental loads. The standards of the evaluation were as follows: mental demand, physical demand, temporal demand, performance, effort, and frustration. The result is given in Table 4. The mental load value in the case of MusicGlove was slightly higher than that in the case of Wii Remote. We consider that this shows a clear difference between Wii Remote that requires the use of only one hand and MusicGlove that requires the use of both hands. The mental workload is lower although the freedom of one hand is deprived because there are no other tasks and the subjects pay attention to listening to music. The MusicGlove has a great advantage to allow the user to use their hands freely without holding any devices. We consider the result in the case of MusicGlove without any buttons to be not very bad as compared

Table 4: Workload by NASA-TLX

	Demand			Performance	Effort	Frustration
	Mental	Physical	Temporal			
MusicGlove	2.0	2.4	2.0	2.0	2.2	2.0
Wii Remote	1.8	1.8	1.2	2.2	1.8	1.8

Values: 1-low workload (Good), 5-high workload (Bad)

to that in the case of the holding-type interface with buttons such as Wii Remote although some loads exist at the present stage. There are advantages of such an interface, such as a high degree of freedom of both hands (there is no necessity to hold any devices), and the intuitive operation is designed for making it easy to remember the required gesture. We obtained considerable positive feedback from subjects in terms of a real-time response and an appropriate gestural design. Some of this feedback is as follows: ”MusicGlove that includes gripped operations is easy to use for a corresponding function than Wii Remote with the button operation” and ”It is good to be able to execute the operation regardless of timing or place because there is no necessity of holding a part of the device.” These results and feedback verify the advantage of the wearable interface that makes the best use of bodily movements.

5. DISCUSSIONS

Not only audio but also visual displays are integrated with the MusicGlove system. We have developed a visual-audio installation where the CD jacket image of music is used for representing each track. Moreover, a CoverFlow-like control was achieved as an example of the application. This system enables the user to enhance the searched object to not only music but also the sound source from different instruments. Hamanaka *et al.* reported a headphone-type interface to seek musical instruments [8]. The individual sound source of different musical instruments is presented and switched by the user’s action using MusicGlove. Moreover, the volume and the pan of each sound can be controlled by hands. As a result, the user can listen to music by switching to his favorite musical instruments. The developed system is also designed to use while several scenes in a daily life. We have developed a prototype device embedded with a portable player. This device realizes a simple and gestural control of music and can be used while walking or jogging (Figure 7). In addition, the system that searches for the sound has also been developed as an application for searching for music. For instance, in the practice of guitar playing, the user can select different sound sources at the same time during the performance (Figure 8).

6. CONCLUSION

In this paper, a novel interface for exploring an audio library using several hand gestures was introduced. We have already developed several devices along the paradigms of Embodied Sound Media (ESM) technology. ESM is designed to formalize a musical sound-space on the basis of the conversion of free human movement into sounds. This technology includes the measurement of human motion, processing, acoustic conversion, and output. The first idea was to introduce direct and intuitive sound feedback within the context of not only embodied interaction between humans

Figure 7: Use while jogging **Figure 8**: Use while playing guitar

and devices but also social interaction among humans. In the present system, the data acquired with the glove device is transferred via wireless communication to and processed by the external computer.

Further development includes the integration of all systems from gesture recognition to audio output at the wearable device. In recent years, several types of portable music players are commercially available. The proposed device can contribute to a new style of wearable music to enhance the user's musical experience. Intuitive music manipulations are enabled by the proposed device, which allows the users to listen to music in an active manner.

REFERENCES

[1] M. Goto and T. Goto. Musicream: New Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, September 2005.

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

[3] 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.

[4] K. Ng. Music via Motion: Transdomain Mapping of Motion and Sound for Interactive Performances. In Proc. of the IEEE, 92(4), pp. 645-655, 2004.

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

[6] K. Hayafuchi and K. Suzuki. MusicGlove: AWearable Musical Controller for Massive Media Library. In *New Interfaces for Musical- Expression(NIME)*, 2008.

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

[8] M. Hamanaka and Seunghee Lee. Sound Scope Headphones: Controlling an Audio Mixer through Natural Movement. In *Proceedings of the 2006 International Computer Music Conference (ICMC2006)*, pp. 155-158, November 2006.

[9] NASA TLX: Task Load Index, [web] http://humansystems.arc.nasa.gov/groups/TLX/