Hindawi Publishing Corporation
International Journal of Computer Games Technology
Volume 2011, Article ID 612535, 10 pages
doi:10.1155/2011/612535

Research Article

Concert Viewing Headphones

Kazuya Atsuta,[1] Masatoshi Hamanaka,[2] and SeungHee Lee[3]

[1] *Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan*
[2] *Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan*
[3] *Faculty of Human Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan*

Correspondence should be addressed to Kazuya Atsuta, kazuya@music.iit.tsukuba.ac.jp

Received 1 September 2011; Revised 12 December 2011; Accepted 26 December 2011

Academic Editor: Suiping Zhou

Copyright © 2011 Kazuya Atsuta et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An audiovisual interface equipped with a projector, an inclina-tion sensor, and a distance sensor for zoom control has been developed that enables a user to selectively view and listen to specific performers in a video-taped group performance. Dubbed *Concert Viewing Headphones*, it has both image and sound processing functions. The image processing extracts the portion of the image indicated by the user and projects it free of distortion on the front and side walls. The sound processing creates imaginary microphones for those performers without one so that the user can hear the sound from any performer. Testing using images and sounds captured using a fisheye-lens camera and 37 lavalier microphones showed that sound locali-zation was fastest when an inverse square function was used for the sound mixing and that the zoom function was useful for locating the desired sound performance.

1. Introduction

In this paper, we describe an audiovisual interface dubbed *Concert Viewing Headphones* that enables someone viewing a video-taped musical concert to select particular performers or areas to view and listen to.

We define two requirements for the interface. First, the user can control it by simply performing the natural actions related to listening. Such actions are those of people in the audience at a concert hall; for example, they turn their heads in the direction of the viewing and/or listening target. Thus, with this interface, a user can better enjoy videos of concerts by selecting particular areas and/or performers on the stage by turning his or her head in their direction and cupping a hand to an ear. Second, the constructed device incorporating this interface is small enough for home use. Since projectors have been getting smaller and smaller, we were able to develop a headphone device equipped with a compact projector, an inclination sensor, and a distance sensor. This device detects the user's head direction, detects the distance between the user's cupped hand and ear, and outputs the corresponding image and sound. Moreover, it is small enough to be used in the home and many other environments.

Figure 1 shows the system flow of this interface. First, the user's head direction is detected by the inclination sensor. Next, the portion of the wide-angle image covering the whole stage corresponding to the head direction is extracted, and this portion is projected on the screen. At the same time, the recorded sounds are mixed so as to emphasize the sounds of the performers within the extracted portion (i.e., the projected image). If the user cups a hand to an ear to hear better, the projected image is enlarged to a degree corresponding to the distance between the user's hand and the distance sensor attached to the one of the headphones, enabling the user to better focus on a particular performer.

This interface has three features in particular.

(1) Use of Imaginary Microphones. Ideally, we would capture the sound for each performer through a microphone attached to the performer's music stand because the sounds are mixed so as to emphasize those within the selected ambit. However, this is difficult in terms of time and cost if there are many performers, as in an orchestra. There is

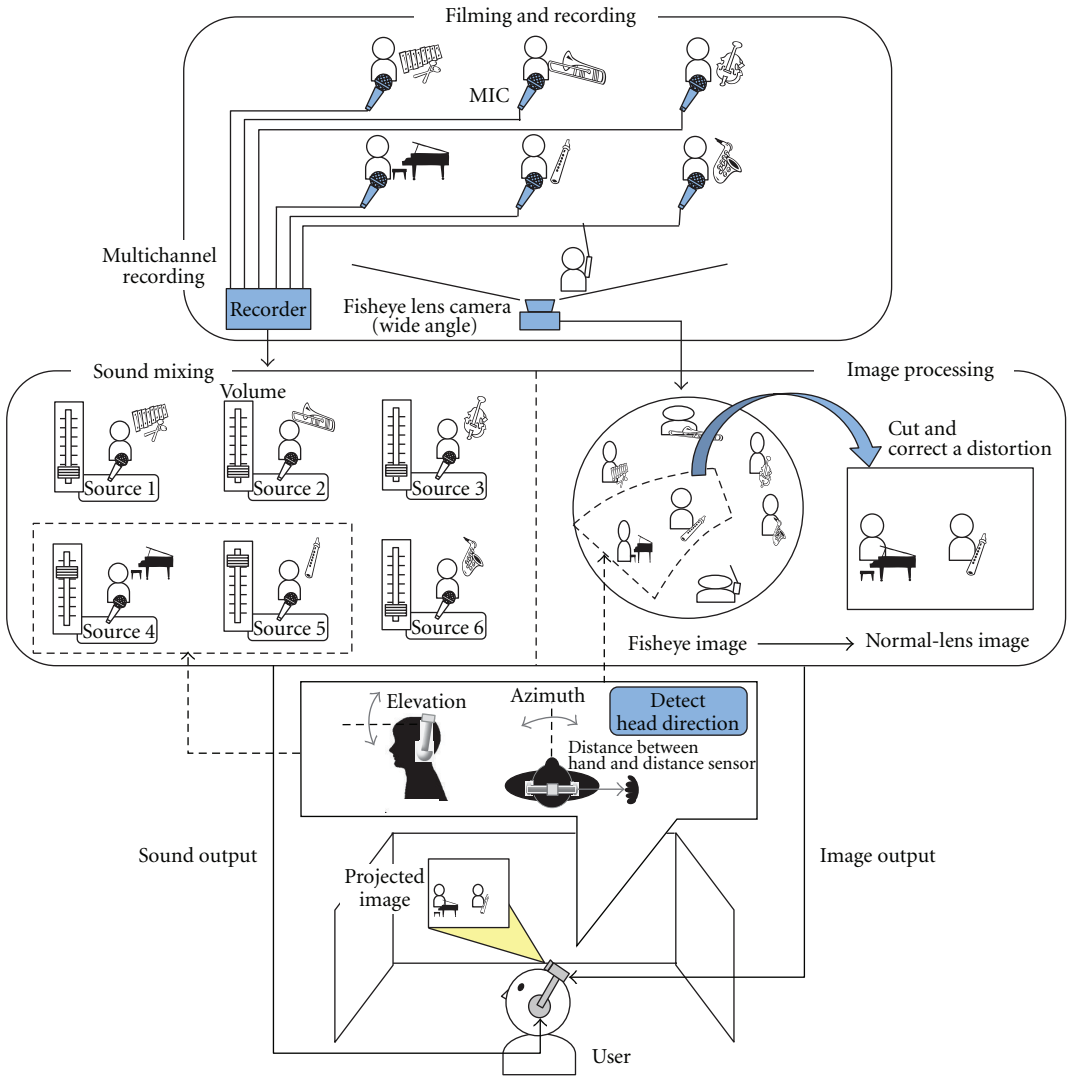

Figure 1: System flow of audiovisual interface.

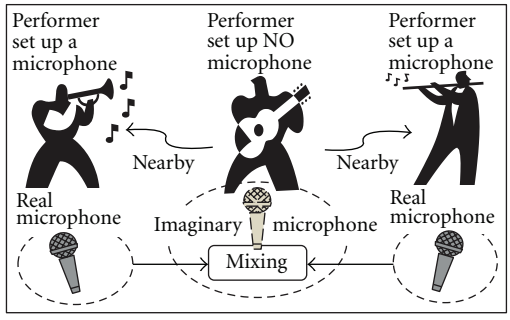

Figure 2: Creation of imaginary microphones.

thus a problem with mixing sounds if performers without a microphone are in the selected ambit. We solve it by creating imaginary microphones for those performers without a real microphone, making it possible to mix sounds as if each performer had a real microphone. The creation of these imaginary microphones is illustrated in Figure 2. The sounds for two performers with real microphones who were near a performer without one are mixed, and this mixed sound is used as the sound recorded by an imaginary microphone for that performer. The sounds are mixed in proportion to the distances between the performer without a real microphone and the two performers with real microphones.

(2) Use of Image Captured with Fisheye Lens. A wide-angle image covering the whole stage is captured using a camera with a circular fisheye lens. The user selects the target ambit from this image, enabling the user to selectively appreciate a particular portion of the image. The lens captures a 180° image that is characterized by low distortion at the center and large distortion around the edges. The distortion in the extracted area is corrected when the image corresponding to the gaze orientation is extracted. A nondistorted image is then projected. This is described in detail in Section 3.1.1.

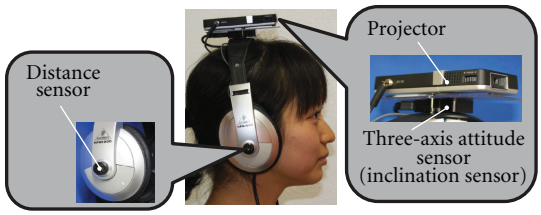

Figure 3: Concert Viewing Headphones.

(3) Projection of Image on Front Wall and Two Side Walls. The image is projected on the front wall and two side walls relative to the user. If the 180° image covering the whole stage was projected on only the front wall, it could be difficult to clearly see the performers at the ends of the stage. For this reason, *Concert Viewing Headphones* is premised on being used indoors and on the images being projected on not only the front wall but also the two side walls. Thus, the user can view the whole stage evenly because the performers at the ends of the stage appear on the corresponding side wall when the user's head turns in the direction of a side wall. However, when an image is projected on a side wall, it is projected at a tilt, so it is distorted and forms a trapezoid (*keystone distortion*). This distortion is compensated for by distorting the image counter to the *keystone distortion* before it is projected. This is described in detail in Section 3.1.2.

The *Concert Viewing Headphones* device developed in this study is shown in Figure 3. A projector and an inclination sensor are mounted on top of the headphones (i.e., above the user's head). This enables the image to be projected in the direction of the user's gaze.

We tested *Concert Viewing Headphones* by using a video of a musical concert recorded with a fisheye-lens camera and 37 microphones. Ten participants used our device to view the concert. They enjoyed listening to the music and watching the performance and were able to zero in on the sounds of particular performers.

2. Related Research

This is the first report of a device based on a pair of headphones with a projector that enable images and sound source to be viewed and listened to selectively. There has been some related research.

2.1. Immersive Displays. *Immersive displays* project images onto screens or on monitors that surround a user. We discuss two major systems below.

The first one is *TWISTER*, which presents stereoscopic images to a user [1, 2]. The user stands in a cylindrical booth, which displays live full-color 360-degree panoramic and stereoscopic images. There are many units arranged in a cylindrical pattern around the user. Each one of these units consists of two LED arrays (one array for the right eye, the other for the left) and a douser. When these units are spinning at high speed, a binocular parallax is caused by the douser, giving the images a stereoscopic effect.

The other one is *Ensphered Vision*, which projects images onto a full-surround spherical screen that surrounds the user, enabling him or her to experience virtual reality [3]. A single projector and a spherical convex mirror are used in order to display a seamless image. The spherical convex mirror diverges the light from the projector in the spherical screen. A planar mirror then bends this light so that the user can see the image from the center. These optical configurations enable the user to view a seamless wide-angle image.

These systems are not suitable for home use because they must be large enough to cover either one's whole body or one's head.

2.2. Multiview or Wide-Angle Image Systems. One can selectively appreciate the images on a DVD containing multiview images and contents captured with wide-angle images by switching among the images arbitrarily [4, 5]. In conventional DVDs containing live images, one cannot selectively switch among images because the images were captured with cameras having a narrow field of view. In contrast, with a DVD containing multiview images, one can select images for particular points on the stage. In contents using wide-angle images, one can arbitrarily select the portion of an image to be viewed from a panoramic image captured with omnidirectional cameras or with a camera equipped with a fisheye lens. Such systems are controlled by manipulating a remote control, or a mouse. Therefore, one cannot appreciate the image and sound by simply performing the natural actions related to listening. Moreover, in the course of using a mouse to select a portion of an image to be viewed from a panoramic image, one sometimes loses his or her place in the image. In contrast, if a user wearing the *Concert Viewing Headphones* device turns his or her head, the projected image is switched corresponding to this movement. Thus, the user can recognize in which direction he or she is looking at in the image.

2.3. Sound Scope Headphones. We previously developed an interface called "*Sound Scope Headphones*" that enables a user to appreciate sounds by selecting particular sound sources [6]. In conventional DVDs containing live images, one cannot select particular sound sources because the sounds are already mixed and cannot be remixed. In contrast, *Sound Scope Headphones* mixes the sounds on the basis of the user's head direction. The interface does not handle images. Furthermore, the interface does not use real sounds recorded at a concert but rather the music in the *RWC Music Database* [7]. In this study, we use real sounds recorded at a concert.

2.4. Head-Mounted Display. A system using a *head-mounted display* changes the images on the basis of the user's head direction [8]. The user wears a device, such as a pair of goggles or a helmet, and experiences virtual reality produced by images displayed on a monitor close to the user's eyes. However, only one person can use it at a time, and the user's eyes tend to become tired.

In a computer game using a *head-mounted display* (e.g., SONY HMZ-T1 [9]), a player experiences only the virtual

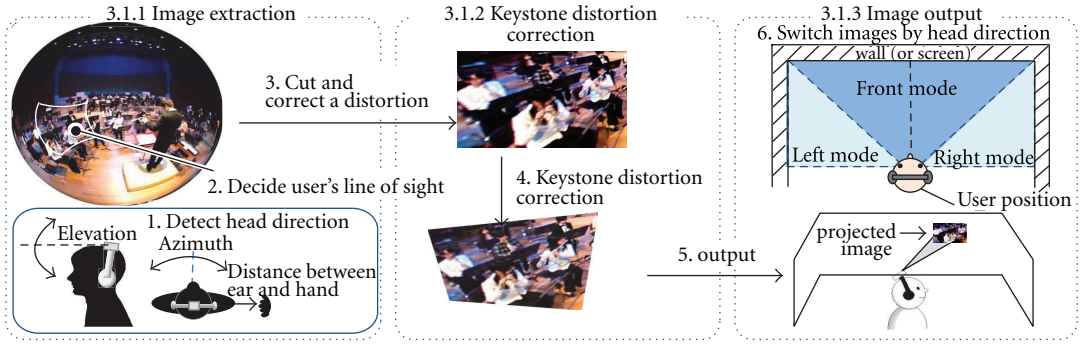

Figure 4: Image processing system.

world. In contrast, a computer game using an interface based on *Concert Viewing Headphones* can combine the real world with the virtual world by projecting the virtual images onto real-world objects. We discuss such a game more fully in Section 6.

2.5. Motion Captures. If the motions of a user turning his or her head and cupping a hand to an ear are captured with Microsoft's Kinect for the Xbox 360 [10], it is possible to produce images and sound corresponding to the motions. In this method, the image is projected on a monitor. However, it is unnatural visually to project 180° images that capture a whole stage on a flat surface, and, as mentioned in Section 1, it could be difficult to clearly see the performers at the ends of the stage. However, if the projection surface is extended by using monitors, it is difficult in terms of cost. In contrast, *Concert Viewing Headphones* projects a 180° image not only on the front wall but also on the two side walls relative to the user. That is, the performers at the ends of the image appear on the corresponding side walls. Thus, the user can view the whole stage evenly.

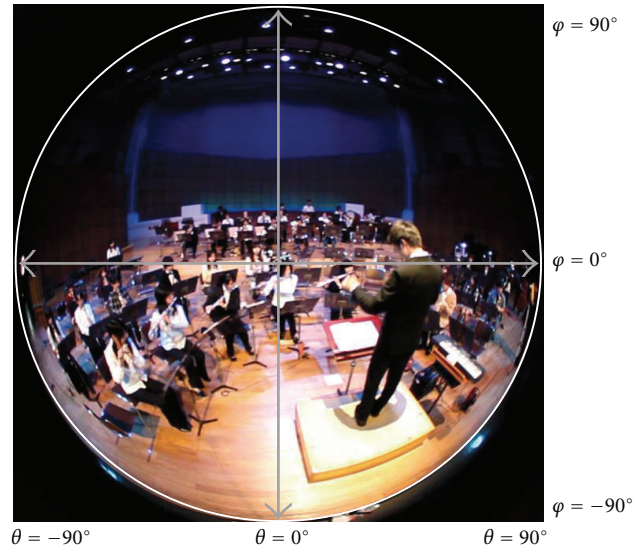

Figure 5: Coordinates on a stage.

3. Description

Concert Viewing Headphones, equipped with a projector, an inclination sensor, and a distance sensor, comprises two systems: image processing and audio processing. It first detects the angle of elevation and direction of the user's head. It then extracts the portion of the image corresponding to the elevation and direction and projects it onto the front wall or a side wall. At the same time, it mixes the sounds recorded by the microphones for that portion of the image and outputs those sounds.

3.1. Image Processing System. The image processing system comprises image extraction, keystone distortion correction, and image output, as illustrated in Figure 4.

3.1.1. Image Extraction. The first step in image extraction is to detect the user's head direction by using the elevation and azimuth of the user's head measured with a three-axis attitude sensor. Next, the gaze orientation, that is, the direction of the user's eyes, is determined on the basis of the detected head direction. Specifically, the sensors detect the user's head direction (elevation φ, azimuth θ), and the coordinates on the stage corresponding to the head direction are calculated. Thus, the image processing system determines which portion of the stage the user is looking at. As Figure 5 shows, the coordinates on the stage are calculated on the basis that the center of the fisheye image is $\varphi = 0°$, $\theta = 0°$ and of the ends of an image are $-90°$ and/or $90°$.

Furthermore, if the user cups a hand to an ear to hear better, the image is enlarged so as to enable looking at a particular area of the stage more closely. The degree of enlargement is proportional to the distance between the user's hand and the distance sensor. The corresponding portion of the image is then extracted.

In this study, we used only one camera in order to reduce cost and to enable a more convenient installation position. Therefore, we captured a whole stage with a circular fisheye-lens camera that can capture a wide-angle image at around 180°. We will now describe the principle for capturing with a fisheye lens and the method for distortion correction.

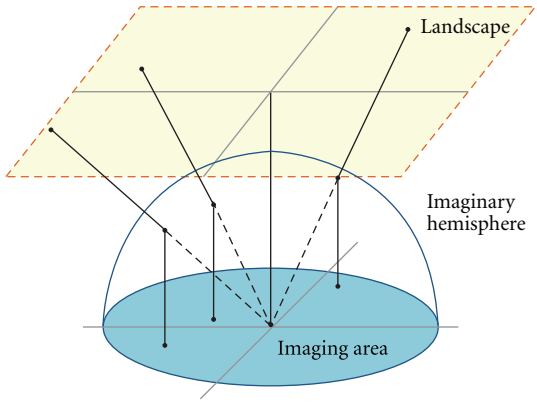

Figure 6: Principle for capturing with a circular fisheye-lens camera.

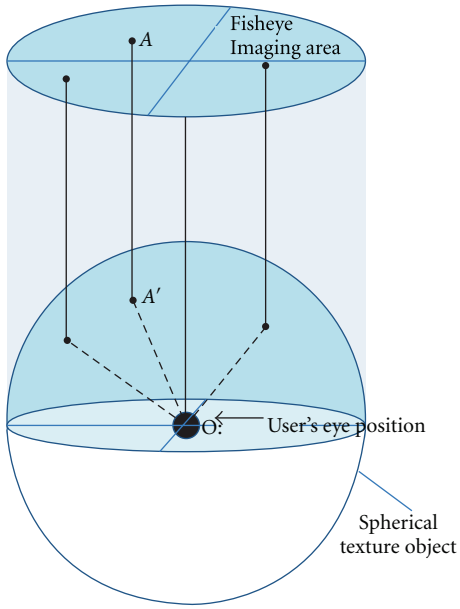

Figure 7: Texture mapping for distortion correction.

As Figure 6 shows, in a camera using the circular fisheye lens, the landscape is reflected in an imaginary hemisphere and projected onto the imaging area at right angles. Thus, a fisheye image is produced and output with low distortion at the center and large distortion around the edges.

The distortion is corrected for by centering the extracted image onto a point detected by the three-axis attitude sensor. Specifically, we correct the distortion by reversing the process based on the principle for capturing with a fisheye lens. In this study, we used OpenGL [11] for the process in this distortion correction. As Figure 7 shows, the fisheye image is projected onto the upper hemisphere of a spherical texture object in OpenGL. We define the center of this object as the user's eye position. Thus, when the user views the spherical surface from the center of the object, a nondistorted image is presented to the user. Figure 8 shows a distorted image and a corrected (non-distorted) image.

Figure 8: Distorted image and corrected image.

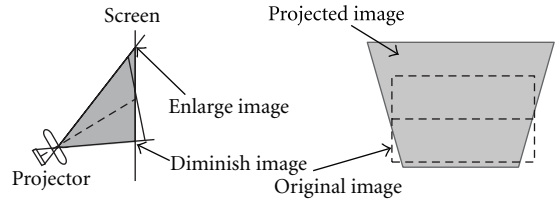

Figure 9: Image projected at a tilt on side wall.

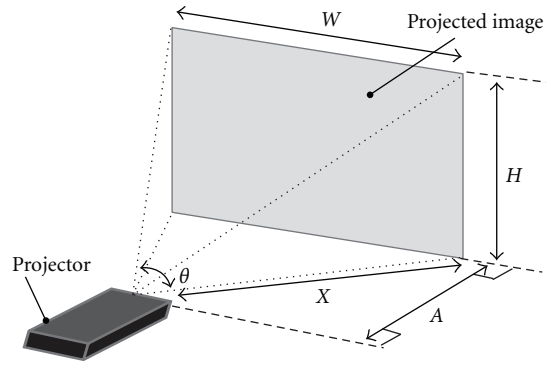

Figure 10: Configuration of projected image.

3.1.2. Keystone Distortion Correction. The image projected on a side wall is at a tilt, so it is either magnified or demagnified and forms a trapezoid, as illustrated in Figure 9. This *keystone distortion* must be corrected. The image is thus processed in accordance with the elevation and direction of the user's head so that it is not distorted when it is projected. This processing distorts the image counter to the keystone distortion. The correction is calculated by using the measurements from the direction and inclination sensors. The magnification or reduction percentage is determined by comparing the projection distance with the image projected on the front wall.

We use A as the minimum distance between the projector and screen, H as the height of the projected image, W as the width of the projected image, θ as the angle of view, and X as the distance between the projector and the edge of the projected image. Figure 10 shows the relationships among these elements.

In addition, when we tilt the projector, θ_1 is the angle between the center of the front wall and the major axis of the

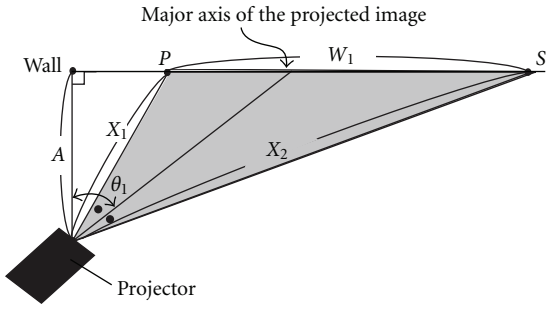

Figure 11: Configuration of image projected at a tilt.

projected image. X_1 and X_2 are the right and left edges of the projected image. Figure 11 shows the relationships among these elements.

The angle between the major axis of the projected image and the left end of the projected image, P, is $\theta/2$. Therefore,

$$X_1 = \frac{A}{\cos(\theta_1 - \theta/2)}. \quad (1)$$

The left side in the image is a maximum of X_1/X times enlarged compared to when it is projected in front of the wall.

Similarly,

$$X_2 = \frac{A}{\cos(\theta_1 + \theta/2)}. \quad (2)$$

The right side in the image is a maximum of X_2/X times enlarged compared to when it is projected in front of the wall. Therefore, the output image is corrected so that the left side in the image is a maximum of X_1/X times reduced, and the right side in the image is a maximum of X_2/X times reduced.

3.1.3. Image Output. Since it is visually unnatural to project 180° images on a flat surface, *Concert Viewing Headphones* is premised, as mentioned above, on being used indoors and on the images being projected on not only the front wall but also the two side walls. When an image is projected on a side wall, correction processing switches its shape appropriately for sidewall projection.

3.2. Audio Processing System. The audio processing system comprises *imaginary microphone creation, distance calculation,* and *mixing*. The system flow is shown in Figure 12.

3.2.1. Imaginary Microphone Creation. As mentioned in Section 1, ideally, there would be a sound source for every performer, but this is difficult in practice. Moreover, it would result in an enormous amount of audio data. Therefore, we record sound sources with microphones corresponding to each of the instrumental parts of the music. That is, a microphone is attached to the music stand of one performer for each group of performers playing the same instrument. For those performers without a microphone, an imaginary microphone is created for each one by mixing the sounds recorded with nearby real microphones on the basis of the distance between the performer and the position of the real microphones. It is possible that the sound of an imaginary microphone might include instrumental sounds that are distant from the imaginary microphone because, in general, microphones pick up sounds from all directions. However, sound pressure is decreased in inverse proportion to the square of the distance between a microphone and the position of a sound source. Moreover, in typical concerts, because the size of the stage is large enough, the performers are spaced out on the stage with enough distance between each instrumental part of the music. Therefore, in this study, most of the sounds picked up by microphones are from nearby the microphones and made up of one instrumental part. For this reason, we used a method that creates imaginary microphones on the basis of the distance between the performer and the position of the real microphones. We will discuss methods for creating a more realistic sound for those performers without a real microphone. These methods are discussed more fully in Section 6.

The sounds from all microphones are used to control the mixing rate for each performer's sound.

3.2.2. Distance Calculation. The mixing rate for each performer's sound is calculated on the basis of the distance between the performer and the center of the image extracted by the image processing system. The distance is calculated from the current projected image and the position of each performer in the coordinate system of the image captured with a circular fisheye lens.

3.2.3. Mixing. Whether there are any performers in the projected image is determined from the distance calculation information. The sound volume for those performers not in the projected image is set to zero. That is, their sound is muted. For performers who are in the projected image, their sound is emphasized so that they approach the center of the projected image. Prepared functions are used to determine the mixing rates needed to adjust the sound volumes so as to increase the volumes with a decrease in the calculated distance. Each performer's sound is multiplied by the corresponding mixing rate, the sounds are added together, and they are output. We prepared five mixing rate functions, as shown in Figure 13.

We normalized the height of the projected image from 0 to 1. When the distance from the center of the image is 0, the volume amplification rate of each function is normalized so that it has a value of 1. When the distance from the center of the image is 0.5, the volume amplification rate of each function is normalized so that it has a value of 0.1.

In Figure 13, line 1 is constant with the distance, and the volume is the same for all microphones. Line 2 is a negative gradient. If this function is applied, each performer's projected volume has an inverse relationship to the performer's distance from the center of the image. Line 3 is a normal distribution. If his function is applied, the user can hear the performers located at the center of the image and around the center. The volume for the performers at the edge of the

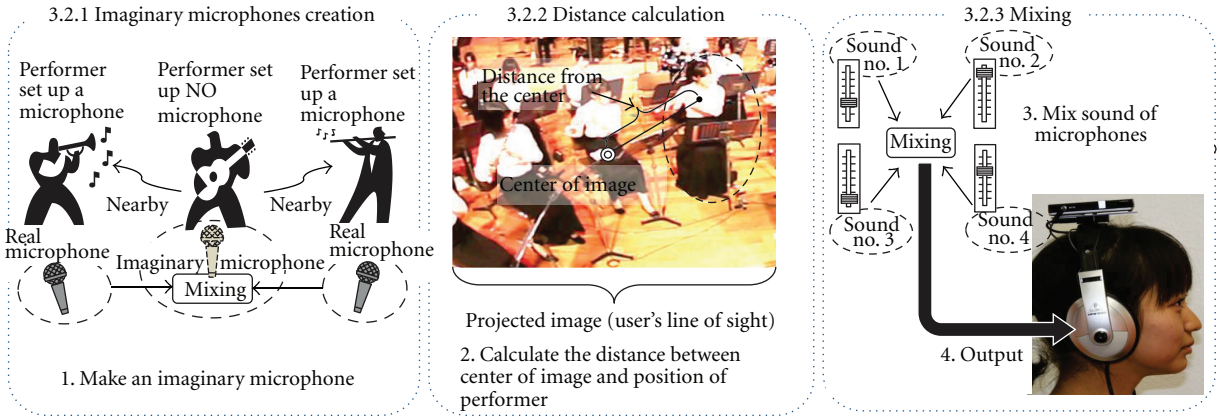

Figure 12: Audio processing system.

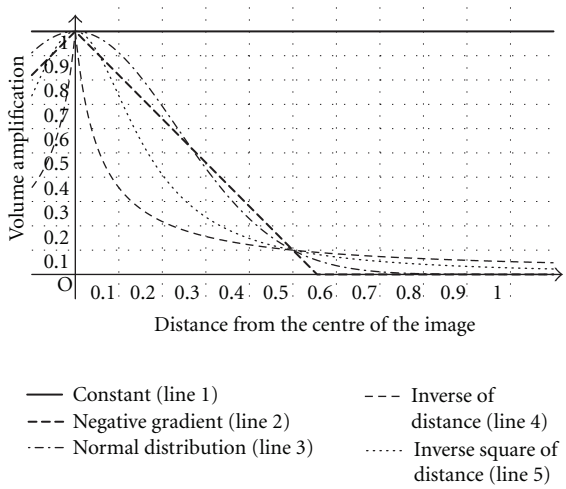

Figure 13: Mixing rate functions.

image is almost zero. Line 4 is the inverse of the distance from the center of the image. If this function is applied, each performer's projected volume has an inverse relationship to the performer's distance from the center of the projected image. Line 5 is the inverse square of the distance from the center of the projected image. If this function is applied, each performer's projected volume has an inverse relationship to the performer's distance from the center of the image. In Section 4, we describe our evaluation of these functions.

If a user controls the projected image so as to be able to view the whole stage, it is possible that the user does not want to emphasize a particular performer's sound but listen to all performers' sounds. For this reason, the user can select from two sound mixing modes. The first one is "constant mode," in which the volume is the same for the sounds of all performers who are in the projected image (line 1 in Figure 13). The other one is "emphasis mode," in which, as mentioned above, the sounds of the performers who are in the projected image are emphasized so that the performers approach the center of the projected image (e.g., line 5 in Figure 13).

Concert Viewing Headphones changes the image and sound mixing in accordance with the orientation of the user's head. When the user operates the zoom function by cupping an ear, performers at edge of the image are out of the enlarged image or are away from the center of the enlarged image. As a result, the sounds of the performers at the center of the projected image are emphasized.

3.3. Implementation. We used an attitude heading reference system (3DM-GX3-25, MicroStrain) as both a direction sensor and an inclination sensor. It outputs the Euler angles, rotation matrix, delta angle, delta velocity, acceleration angular rate, and magnetic field to a USB device small enough to be mounted on a pair of headphones.

We use a proximity sensor (Asakusa Giken Co., Ltd.) as the distance sensor. It detects the distance in the 0–6 cm range by infrared reflectance, making it well suited for measuring the distance between the hand and ear. It is mounted on the right headphone. We initially used an ultrasonic sensor, but it was unable to provide accurate measurements at close range.

Data is transmitted from the proximity sensor by using the *Arduino* which is an open-source electronics prototyping platform. Because the proximity sensor outputs data using serial communication, connecting it to a PC is problematic. We used a USB-serial conversion substrate (FT232RX) to enable the proximity sensor-to-USB connection.

A *mini USB projector* was mounted on the headphones. Because the 3DM-GX3 system is susceptible to magnetic force, we selected a projector with less magnetic force.

4. Evaluation

We evaluated our *Concert Viewing Headphones* by first creating an image and sound source. We video-taped and recorded a University of Tsukuba Symphonic Band concert at Nova Hall (a concert hall in Tsukuba city, Japan). The microphones and camera were configured as shown in Figure 14. The microphones were placed on the music stands. Because there were 37 instrumental parts in the performance, we used 37 lavalier microphones to record the

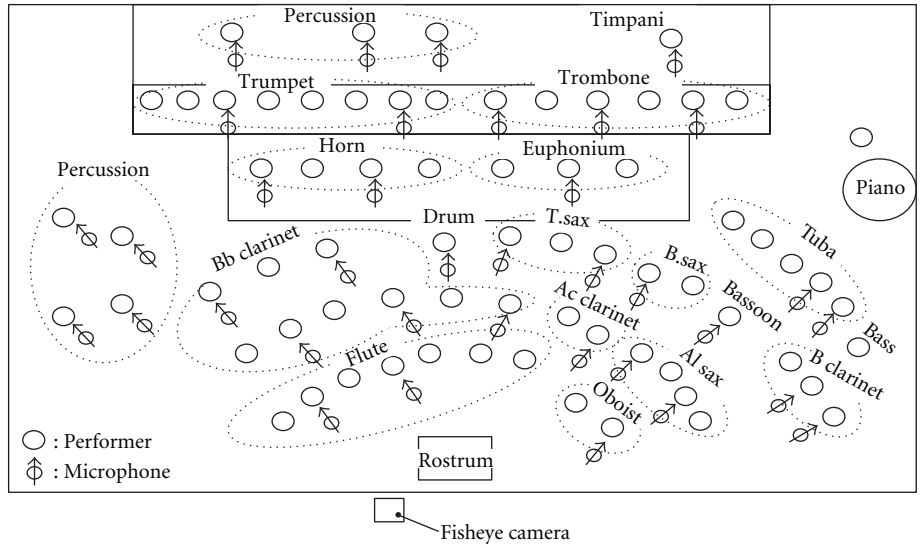

Figure 14: Microphone and camera positions.

Table 1: Average time to locate 440 Hz sine wave.

Function	Time without zoom system [s]	Time with zoom system [s]
Constant	112.5	56.2
Negative gradient	29.2	24.0
Normal distribution	32.6	17.8
Inverse of distance	30.9	20.8
Inverse square of distance	23.8	16.9

sounds. An image of the entire stage with all performers visible was captured with a 2.5 m high fisheye-lens camera. We experimented with the mixing functions described in Section 3.2.3 to identify the natural correspondence between changes in the image and the sound mixing.

4.1. Evaluation of Functions for Mixing. We located a point at which one could hear a 440 Hz sine wave at random positions in the image. The ten evaluation participants attempted to locate this point with only their ears so as to identify the natural correspondence between head movement and changes in the sound mixing. We recorded the time it took to do this. Each time a participant located the point, the sine wave shifted to another point at random. This was repeated five times for each function. We defined the function with which the participants could find the sine wave the fastest as the best one. The average times for each function are shown in Table 1. The time was the shortest for the "inverse square of distance" function with the zoom function.

All ten participants were adults in their 20's. In addition, three of them had been playing musical instruments on a daily basis, another three had been playing musical instruments for a period of time, and the other four had hardly played an instrument. We determined that there was not any difference that depended on the participants' musical

Figure 15: Participant using *Concert Viewing Headphones*.

experience in the result of each function. Figure 15 shows a participant using *Concert Viewing Headphones*.

4.2. Evaluation of Sound and Image. We demonstrated that the targeted ambit naturally corresponded to the changes in the sound mixing by using an eye-mark recorder. Specifically, the participants repeated the action of looking at a particular performer arbitrarily and transferring their gaze to another performer. In the meantime, we examined the relation between their viewpoint and the mixing rate of each performer's sound. As a result, we determined that the sound volume of a performer who the participant was looking at was highest in the mixing. When the participant was transferring his or her gaze, the mixing was switched in real time so that the sound volume of a performer who was closest to the participant's viewpoint was highest. However, when the participant turned his or her head quickly, we determined that the shifting of the image sometimes lagged behind that of the sound. If we reduced the file size of the image and/or sound, this lag would be improved, but

Table 2: Average score of each function.

Function	Average score rated by ten participants			Average
	Q1: Is the relation between the shifting of the image and that of the sound natural?	Q2: is the instrumental sound emphasized so that it approaches the center of the projected image?	Q3: when you turn your head quickly, is the relation between the shifting of the image and that of the sound natural	
Constant	2.6	2.7	3.1	2.8
Negative gradient	3.9	4.1	3.9	3.9
Normal distribution	3.4	3.6	3.6	3.5
Inverse of distance	3.6	3.3	3.1	3.3
Inverse square of distance	3.7	3.7	4.0	3.8

we think that the image and sound quality should not be degraded in the audiovisual interface. Therefore, we will improve the program of this interface so that the shifting of the image and sound is more natural.

4.3. Questionnaire. A questionnaire containing four questions was completed by the participants after the testing. Questions nos. 1 to 3 were rated on a 5-point scale with the following possible responses: completely disagree (1), somewhat disagree (2), uncertain (3), somewhat agree (4), and completely agree (5). Table 2 shows the average score of each function rated by the ten participants for each question. The score was highest for the "negative gradient" and "inverse square of distance" functions. For question no. 4, we asked "What is your feeling about the characteristic or difference among each of the five functions?" In the "inverse square of distance function," most participants had positive opinions, for example, "I could hear the difference among the sounds very clearly," and "I could notice many different instrumental sounds." Therefore, we could determine the effectiveness of using the inverse square of distance function.

5. Conclusion

Our *Concert Viewing Headphones*, equipped with a projector, an inclination sensor, and a distance sensor for zoom control, enables a user to selectively view and listen to specific performers in a video-taped group performance. It has both image and sound processing functions. The image processing extracts the portion of the image selected by the user and projects it free of distortion on the front and side walls. The sound processing creates imaginary microphones for those performers without one so that the user can hear the sound from any performer. Testing using images and sounds captured using a fisheye-lens camera and 37 lavalier microphones showed that sound localization was fastest when an inverse square function was used for the sound mixing. Moreover, the zoom function enabled the participants to indicate the desired sound performance.

6. Future Work

We will discuss the creation of the imaginary microphones in the audio processing system and design a method so as to be able to create a more realistic sound. For example, we will record the instrumental sounds within a more narrow area on the stage by using directional microphones and adjusting the position of the microphones. Besides this, we will estimate the acoustic transfer function in the position of the performers without a real microphone.

Furthermore, we will apply this interface to virtual reality games. As mentioned above, *Concert Viewing Headphones* can combine the real world with the virtual world because it can project virtual images onto real-world objects. For example, we plan to devise a game that uses an interface based on *Concert Viewing Headphones*. In a darkish room, there are real objects (e.g., clocks, shelves, boxes, etc.) located around a player wearing the interface. The player finds a particular imaginary musical instrument from among the imaginary instruments hidden in the objects similar to putting a spotlight on something. Specifically, each one of these objects corresponds to an instrumental sound and the image of a performer playing the instrument on the projected image. If the player's viewpoint is close to an object, an instrumental sound corresponding to the object is emphasized, and a performer playing the instrument is projected on the object. That is, the player can find the imaginary instruments from the change of sounds and the projected images by turning his or her head.

In order to improve on *Concert Viewing Headphones*, we plan to conduct further experiments on the image functions to determine whether the zooming and image changes are natural. Furthermore, we will conduct experiments with the participants of all ages.

Acknowledgments

The author would like to express our deepest gratitude to Sawako Miyashita who provided helpful comments and suggestions. They would also like to thank Atsushi Usami whose meticulous comments were an enormous help to them.

References

[1] Y. Kunita, N. Ogawa, A. Sakuma, M. Inami, T. Maeda, and S. Tachi, "Immersive autostereoscopic display for mutual telexistence: TWISTER I (Telexistence Wide-angle Immersive STEReoscope Model I)," in *Proceedings of the IEEE Virtual Reality Annual International Symposium*, pp. 31–36, Yokohama, Japan, 2001.

[2] S. Tachi, "TWISTER: immersive ominidirectional autostereoscopic 3D booth for mutual telexistence," in *Proceedings of the Asiagraph in Tokyo*, pp. 1–6, Tokyo, Japan, 2007.

[3] H. Iwata, "Full-surround image display technologies," *International Journal of Computer Vision*, vol. 58, no. 3, pp. 227–235, 2004.

[4] Google Maps with Street View, http://maps.google.com/intl/en/help/maps/streetview/.

[5] Immersive Media, http://www.immersivemedia.com/demos/index.php.

[6] M. Hamanaka and S. Lee, "Sound scope headphones: controlling an audio mixer through natu-ral movement," in *Proceedings of the International Computer Music Conference*, pp. 155–158, New Orleans, La, USA, 2006.

[7] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: music genre database and musical instrument sound database," in *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR '03)*, pp. 229–230, October 2003.

[8] Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya, "Telepresence by real-time view-dependent image generation from omnidirectional video streams," *Computer Vision and Image Understanding*, vol. 71, no. 2, pp. 154–165, 1998.

[9] Sony Introduces the Worlds First Personal 3D Viewer, http://news.sel.sony.com/en/press_room/consumer/television/release/60813.html.

[10] Xbox 360 + Kinect, http://www.xbox.com/en-US/Kinect.

[11] OpenGL: The Industry's Foundation for High Perfor-mance Graphics, http://www.opengl.org/.