A System for Visualizing Sound Source using Augmented Reality

1 Introduction

Since computers have become smaller and capable of higher levels of performance and with Internet technologies being widely used, computers have become common in our daily life. In particular, since the advance of high-performance devices such as smartphones, wearable computing, in which users fix devices to their body, has attracted a great deal of attention. Users can get information and services by these devices anywhere and at anytime (Kawai & Tomita, 2000). Environmental sounds are one of the most important information sources in our daily life, we can realize what is happening in our surroundings (Matthews, Fong & Mankoff, 2005). In recent years, there has been a lot of research on automatic sound source recognition by computers (Asano, Goto, Itou & Asoh, 2001). These pieces of research are focused on the algorithm of recognition and show the result to users in text or pictures (Ho-Ching, Mankoff & Landay, 2003). However, this is not an ideal interface for the hearing-impaired for showing them the direction of the sound source (Somervell, McCrickard, North & Shukla, 2002).

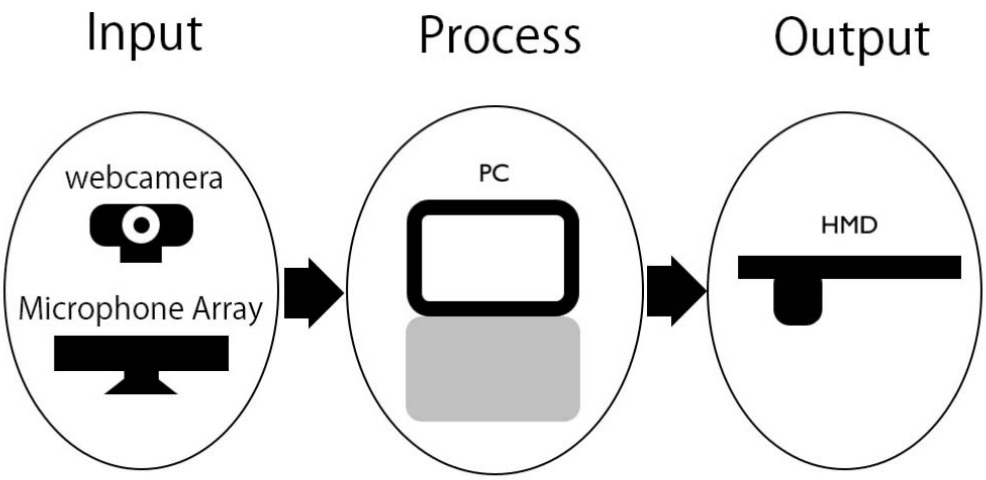


Figure 1: System Construction

We propose a new interface of sound source recognition for the hearing-impaired using augmented reality (AR) that can show the user information in the real world by using a web camera and head-mounted display (HMD) (Feiner, MacIntyre, Hollerer & Webster, 1997). Our system can recognize the environmental sound in realtime and inform the user of the type of sound by showing a virtual object in the user's sight. Furthermore the user can find the direction of the sound source by using a microphone array and locate the sound source through the AR marker attached to the object.

This article consists of six parts: reference research in Section 2, detail of the system design in Section 3, system implementation in Section 4, evaluation in Section 5, and finally, the conclusion in Section 6.

2 References

There is currently a lot of research on sound visualization. For instance, in Azar's research, to find the easiest way to understand how to visualize sound, several methods of sound visualization are discussed (Azar, Saleh & Al-Alaoui, 2007). They discuss methods of sound visualization for environmental sound and speech, and the results of visualization are displayed in one form of the system. They considered seven kinds of methods to visualize environmental sound and all of these results were displayed in graphs by using MATLAB. They also used text, icons, and American Sign Language (ASL) with text for three kinds of methods to visualize speech. The results of the examination, show that they figured out that ASL was the easiest way to understand how to visualize sound. However, ASL is only for those who have learnt it; moreover, it is hard to use only one ASL picture to explain all words.

In Kaper's research, they addressed digital sound synthesis in the context of Digital Instrument for Additive Sound Synthesis (DIASS) and sound visualization in a virtual reality environment by means of M4CAVE (Kaper, Tipei & Wiebel, 1999). In Kaper's project they visualized sound in a room-sized virtual reality environment. But when there were no speech sounds, it was still hard for humans to recognize when sound transferred into particles in a VR environment.

In Bertram's research, particles were used to visualize sound source positions in virtual reality (VR) (Bertram, Deines, Mohring, Jegorovs & Hagen, 2005),but showing the user the exact location of the source was challenging.

In Osaka's research they proposed a visualization environmental search system(Osaka, Saito, Ishitsuka & Yoshio, 2009). They used tones to describe environmental sound and showed them to users in 3D. Furthermore, similar tones are shown in the same area, so users can easily search the environmental sounds.

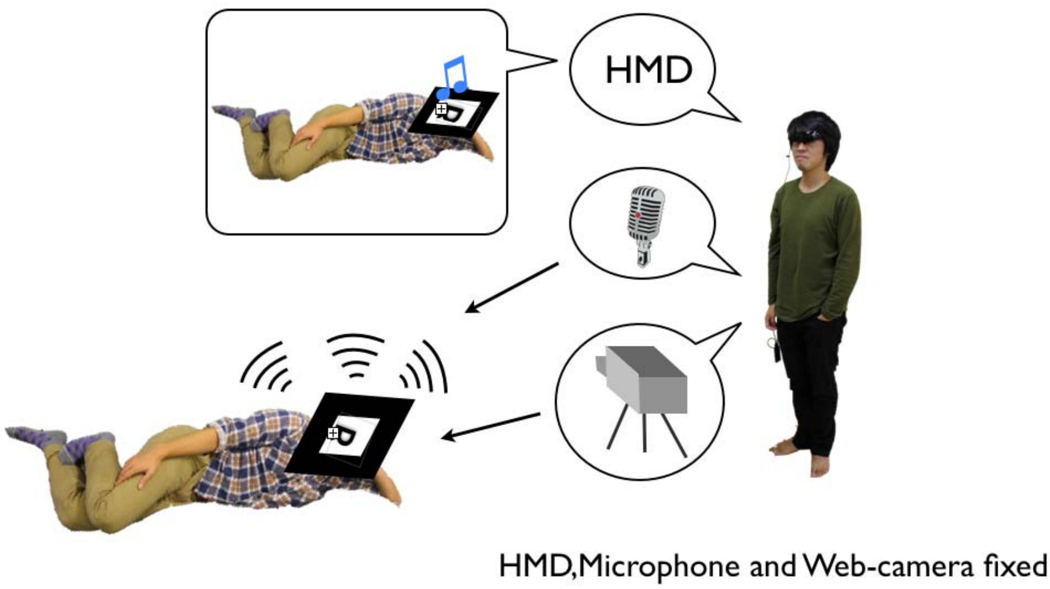


Figure 2: Proposed System Image

In Valin's research, they presented a robust sound source localization method in three-dimensional space that uses an array of eight microphones. The method is based on time delay of arrival estimation (Valin, Michaud, Rouat, & Letourneau, 2003). As result, they found that not all sound types are equally well detected, and the system functioned properly up to a distance between 3 and 5 meters. However, we proposed a system that cannot only recognize the direction of sound objects but also show a user exactly where the sound source is.

3 System Design

In this section we will first introduce the environmental assumptions of our system. Second we will then explain the detail of our system construction, and,third, we will explain the method. Finally, we will explain the method of visualization of a sound source by using AR and the method of matching the position of sound source by using a microphone array.

3.1 Environmental Assumptions

We assumed that the user is fixed with a PC, wearable camera, microphone, and a HMD.

3.2 System Construction

The construction of the proposed system is shown in Figure 1. Our system consists of input, process, and output. In the input part, the web camera captured images from the view of the user in realtime, and the microphone captured environmental sounds and detects the direction of sound source. In the process part, the program in the PC calculates the data, which are transfered from the input part, and transfers the result of the analysis to the device of the output part. In this part, the user can find the sound source on the HMD by showing the virtual object on the AR marker that is attached to the sound source by using AR.

The image of proposed system is shown in Figure 2: the microphone that is fixed to the user to capture the environmental sound in realtime, and the web camera captures the image that includes the AR markers. The result of sound source recognition and AR marker recognition are output to the user in HMD by showing the virtual object in the user's view. Thus, the user can see what is making the sound in the environment without hearing any sound. Furthermore, the user can exactly locate the sound source.

3.3 Sound Source Recognition Function

Figure 3: Sample of Matching

We used the mel-frequency cepstral coefficient (MFCC) to recognize the sound source. Human beings can easily recognize sounds that have a high frequency, but find it hard to recognize sounds that have low frequency. Mel-frequency used in MFCC has the same nonlinear character[10]. Therefore, using it is an ideal method to recognize environmental sounds in the smae way that human ears do. Furthermore, MFCC has a higher level of recognition accuracy than that of other methods(Zheng, Zhang & Song, 2001) (Clarkson, Sawhney & Pentland, 1998). We used MFCC to calculate features in the following four steps.

1. Use Fast Fourier transform (FFT) to calculate the spectrum of the sound in each window.
2. Map the spectrum onto the mel-scale by using triangular overlapping windows.
3. Take a log of the mapping result.
4. Take the discrete cosine transform after taking the log.

We thus calculate the features of the sound source to create the database of sound sources, and label each sound source. The system can be used to recognize the environmental sound in realtime through matching the features in the database with those that are calculated in realtime by using Euclidean distance. Finally, the nearest one is the recognition result.

3.4 Augmented Reality Function

We used AR to visualize sound source that is recognized for user. As AR is a kind of technology that is used to present the user with virtual objects in the real world, the user can check information when seeing the real world with a HMD (Hollerer, Feiner, Terauchi, Rashid & Hallaway, 1999). Augmented reality is an ideal interface in our daily life (Kato, Billinghurst, Poupyrev, Imamoto & Tachibana, 2004). We can make AR work for a low cost, and because the AR marker is used to show user the virtual object, the user can easily locate the sound source.

We made AR marker and pattern file for each sound source in the database and attached the marker onto the sound source object. Therefore, the system was used to visualize a sound source, the web camera captures the marker, and the system shows the result of recognition of the marker in the HMD worn by the user.

3.5 Sound Source Position Recognition Function

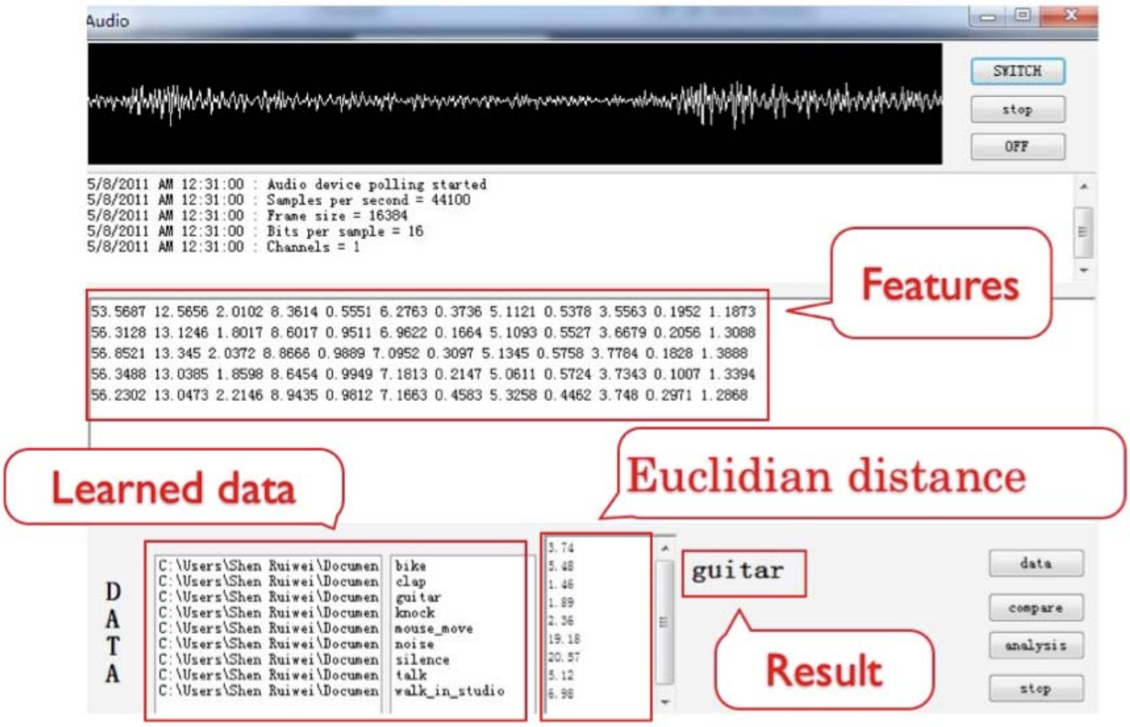


Figure 4: interface of sound source recognition

The same kind of sound may have several different sources in an environment at the same time, such as different people talking, different cellphones or copy machines. Even if the sound source is recognized by the system, the user can find identifying which is making sound difficult. Therefore, our system uses a microphone array to detect the direction of the sound source. Humans can find the direction of a sound source through the difference in intensity between their two ears (Rakerd, & Hartmann, 2010). For machines the direction of a sound source can be recognized through interaural time difference (ITD) or inter- aural level difference (ILD) (Nakadai, Matsuura, Okuno & Kitano, 2003). In our system, Microsoft's Kinect sensor is used to detect the direction of the sound source. The Kinect sensor's microphone array is made up of four microphones that are on the bottom of the Kinect sensor, and using APIs, which are released by Microsoft, the microphones can recognize the direction of the sound source in a range of 100 degrees in front of the Kinect sensor. In this way, the proposed system can definitely tell the user where the sound source is even if the same kind of sound source objects exist in the surrounding environment.

3.6 Position Matching of Sound Source Function

Our system shows the user virtual objects after matching the direction of sound calculated with the data from microphone array with the position of the marker of the sound source on the screen. First, we calculate the range of the web camera and the microphone array. We then transfer the direction of the sound source from degree to pixel onto the x coordinate of screen. Finally, if the direction of sound source was within 10 pixels of the center of the marker as Figure 3, we specify the sound source and show the user the object on the sound object that is making the sound.

4 System Implementation

We implemented our prototype system. In this section, we will explain the detail of the implementation of the input function, process function, and output function.

4.1 Input Function

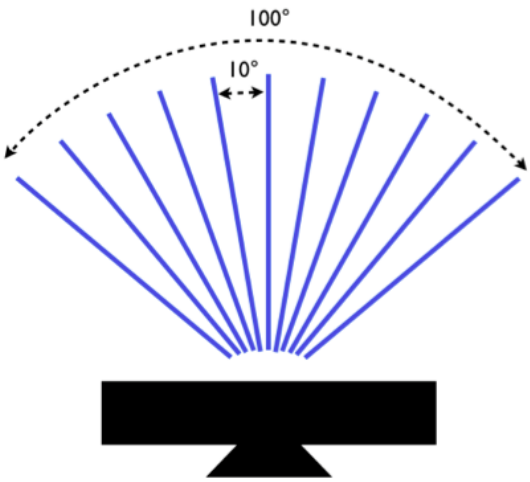


Figure 5: Beamforming of Kinect

Figure 6: Implementation of Prototype System

We used Digital Cowboy's Net Cowboy web-camera, Microsoft's Kinect for Xbox 360 as the microphone array, and the SDK, which is released by Microsoft for programming. Furthermore, we calculated the range of the web camera and microphone array. We found that microphone array ranged from -50 to 50 degree and that the web camera ranged from -30 to 30 degree.

4.2 Process Function

Our system has three parts: a sound source process, an AR process and a sound source direction. We will explain the details below.

4.2.1 Sound Source Process

To recognize environmental sound, we used a system to calculate features as learnt data before hand, and labelled each learned data of the sound source. Each piece of learned data of sound had 250 frames of data, and each frame had the features of one second. The interface of sound recognition is shown in Figure 4. Environmental sound was calculated into 12-dimension features by using MFCC, and we calculated the distance between the features in the database by using Euclidean distance. After three frames, the result of recognition was decided by majority, and then it was transmitted to the output part.

Figure 7: Sample of Visualization

4.2.2 Augmented Reality Process

Several toolkits for applications of AR have recently become available. These include the, ARToolkit in the C++ development environment, the FLARToolkit for ActionScript in Flash. In our system, we used NyArtoolkit for our C# development environment. We used the NyARToolkit to make pattern files of the same quantity with labeled learned data, attached these AR markers to the matching sound source objects. We used the NyARToolkit to detect the AR markers from the images, which were captured with a web camera mounted on the user, and images were processed into black and white by using a program and, AR markers were recognized through AR pattern files.

We used the toolkit to calculate the position of the AR marker from a 3D to a 2D position on the screen and transferred the 2D one to a sound source direction process.

4.2.3 Sound source direction process

The Kinect sensor cannot only track the motion of human being, but is also commonly used in sound recognition. There were four microphones at the bottom of the Kinect sensor, which uses them to detect the direction of a sound source as a microphone array. To make it work we used Kinect for Windows SDK (released by Microsoft) and SDK recognized the direction of sound the source by making use of beamforming. As shown in Figure 5, there were eleven beams in the range from -50 to 50 degrees to detect the direction. In our system, we used the property called MicArrayBeamAngle in the KinectAudioSource object to estimate the direction of sound source for which beam; furthermore, we used SoundSourcePosition to calculate the direction exactly. In our prototype system, we transferred the direction that was calculated by using the Kinect sensor from degree to pixel onto the x coordinate of the screen and matched it with the position of the marker on the screen. As shown in Figure 6 we fixed the web camera to the Kinect.

4.3 Output Function

To provide information on the surrounding environment, the user wore an HMD. An HMD is an ideal device for providing the position and other information of a sound source to a user because the HMD can display virtual objects in the real world environment in the user's field of view. When the AR markers were detected, the virtual objects showed on the AR markers, which were matched with the labels of the detected sound source.

5 Evaluation

We evaluated our system. First, we evaluated the accuracy of sound source recognition. We then tested the system in the supposition environment i.e, that markers were attached to the object. We tested the system to see whether the user could find the position of the sound source. We explain the evaluation our system by sound source recognition function, augmented reality function, sound source direction recognition function, system confirmation, and comprehensive evaluation.

5.1 Sound Source Recognition Function

We took six kinds of sound sources; each type has 4 samples. We made the system calculate the features for each sound source as learned data using MFCC, each piece of data has 250 frame samples. We took the sound of the circuit breaker, a car, a train, the ring of a cellphone, an alarm, and no sounds as sample sound. For evaluation we took 3 one-minute pieces of sound, and we set the system to recognize them. Because the system shows the user the result of recognition by major, the system may be unable to decide what the sound source is when each frame is different, we set the system to show the result for when the major result did not work. The recognition results of sample sound A, B, and C are shown in Table 1, 2, and 3.

The recognition accuracy of each piece of sound was 0.83, 0.81, and 0.92, and we found the follow points of sound recognition in our system.

- Our system recognized a sound source in one second, after three frames, and the recognition result was decided by majority. This causes a delay, of about 2 seconds. For instance, when a new sound source is presented, to recognize the new sound source, our system takes 2 seconds.
- In a situation where there were several kinds different kinds of objects making sound in the surrounding environment, the accuracy of recognition decreased.
- In a situation where there were unknown sound sources in the surrounding environment, the accuracy of recognition changed. For instance, in sample B, there were sounds of people talking and walking that were not learned data; therefore, the accuracy of sample B was lower than that of sample C.

5.2 Augmented Reality Function

We evaluated the AR part of the proposed system. Four samples of sound source visualization using our system are shown in Figure 7: the ringing of a cellphone, the sound of a guitar, the sound of a shredder crushing paper, and the sound of a copy machine printing.

There were several parameters in NyARToolkit's library that are used in the AR part of our system, such as the size of the AR markers and the threshold of the recognition rate, and we set the side of each AR marker at 80 millimeters and the threshold of AR the markers's recognition rate to be 0.6. If several different kinds of AR markers were present in one picture, the recognition accuracy of AR markers could be about 80%. Virtual objects, which were prepared beforehand, could be displayed on the AR markers, but the light and the distance between web camera and AR marker, could lead to misrecognition.

- If AR markers were under intense light, the binarization of AR markers could be problematic. Because of the reflection of the light, our system could not recognize the AR markers.
- If the marker captured was small, the system could hardly recognize it.
- If part of a marker was covered by an other object, the system could not recognize it.

Table 1: Recognition Result of Sample A for Each Second

A	Sound	Recognition1	Recognition2	A	Sound	Recognition1	Recognition2
1	circuit breaker	no sound	car	37	circuit breaker	circuit breaker	ring
2	circuit breaker	--	no sound	38	circuit breaker	circuit breaker	car
3	circuit breaker	circuit breaker	car	39	circuit breaker	circuit breaker	ring
4	circuit breaker	circuit breaker	car	40	circuit breaker	circuit breaker	car
5	circuit breaker＋car	car	ring	41	circuit breaker	circuit breaker	car
6	car＋circuit breaker	car	ring	42	circuit breaker	circuit breaker	car
7	car＋circuit breaker	car	train	43	circuit breaker	circuit breaker	car
8	car	car	alarm	44	circuit breaker	circuit breaker	car
9	car	car	alarm	45	circuit breaker	circuit breaker	alarm
10	car	car	alarm	46	circuit breaker＋train	circuit breaker	alarm
11	car	train	train	47	train	train	alarm
12	car	train	train	48	train	train	alarm
13	car	car	train	49	train	train	alarm
14	car＋circuit breaker	car	train	50	train	train	alarm
15	car＋circuit breaker	car	no sound	51	train	train	alarm
16	car＋circuit breaker	car	train	52	train	train	train
17	car＋circuit breaker	car	ring	53	train	car	alarm
18	circuit breaker＋car	car	car	54	train	train	alarm
19	circuit breaker＋car	circuit breaker	car	55	train	train	train
20	circuit breaker＋car	circuit breaker	car	56	train	train	train
21	circuit breaker＋car	circuit breaker	car	57	train	train	train
22	circuit breaker＋car	circuit breaker	ring	58	train	train	train
23	circuit breaker＋car	car	ring	59	train	train	train

24	circuit breaker＋car	car	car	60	train	train	alarm
25	circuit breaker＋car	circuit breaker	car	61	train	train	train
26	circuit breaker＋car	car	ring	62	train	train	train
27	circuit breaker＋car	circuit breaker	car	63	car	car	train
28	circuit breaker＋car	circuit breaker	car	64	car	car	train
29	car＋circuit breaker	circuit breaker	ring	65	car	car	train
30	car	circuit breaker	no sound	66	car	car	train
31	car	car	train	67	car	car	train
32	car	car	no sound	68	car	car	ring
33	car	car	no sound	69	car	car	ring
34	circuit breaker＋car	circuit breaker	no sound	70	car	car	ring
35	circuit breaker	circuit breaker	no sound	71	car	car	car
36	circuit breaker	circuit breaker	no sound	72	car	car	car

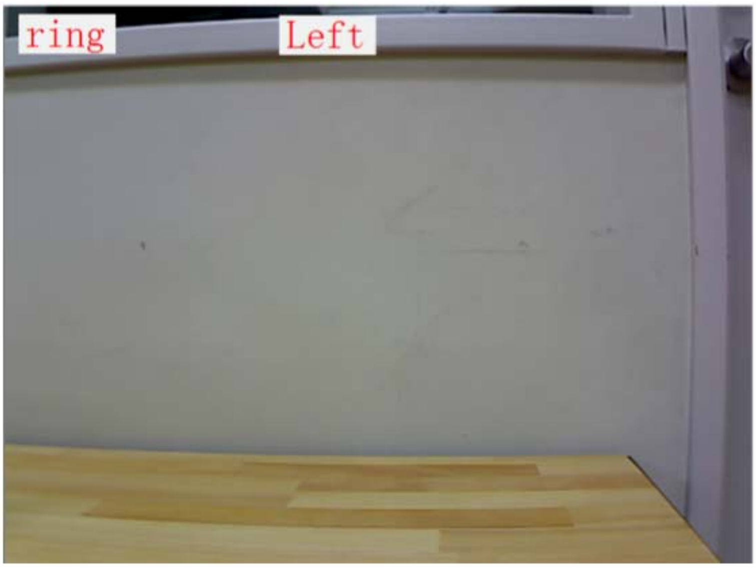
Table 2: Recognition Result of Sample B for Each Second

B	Sound	Recognition1	Recognition2	B	Sound	Recognition1	Recognition2
1	circuit breaker	no sound	no sound	27	car＋circuit breaker	car	car
2	circuit breaker	circuit breaker	ring	28	car＋circuit breaker	circuit breaker	car
3	circuit breaker	circuit breaker	ring	29	car	--	car
4	circuit breaker	circuit breaker	ring	30	car	--	train
5	circuit breaker	circuit breaker	ring	31	car	--	train
6	circuit breaker	circuit breaker	car	32	car	car	train
7	circuit breaker	circuit breaker	ring	33	car＋circuit breaker	--	car
8	circuit breaker	circuit breaker	ring	34	circuit breaker＋car	circuit breaker	car
9	circuit breaker	circuit breaker	car	35	circuit breaker＋car	circuit breaker	train
10	circuit breaker	circuit breaker	car	36	circuit breaker＋car	circuit breaker	car
11	circuit breaker	circuit breaker	car	37	circuit breaker＋car	circuit breaker	car
12	circuit breaker	circuit breaker	car	38	circuit breaker＋car	circuit breaker	car
13	circuit breaker	circuit breaker	ring	39	circuit breaker＋car	circuit breaker	car
14	circuit breaker	circuit breaker	car	40	circuit breaker	circuit breaker	car
15	circuit breaker	circuit breaker	car	41	circuit breaker	circuit breaker	car
16	circuit breaker	circuit breaker	car	42	circuit breaker＋train	circuit breaker	car

17	circuit breaker	circuit breaker	car	43	train	circuit breaker	train
18	circuit breaker	circuit breaker	car	44	train	train	train
19	circuit breaker	circuit breaker	car	45	train	train	train
20	circuit breaker+car	circuit breaker	car	46	train	train	train
21	circuit breaker+car	circuit breaker	car	47	train	train	train
22	circuit breaker+car	circuit breaker	car	48	train	train	train
23	circuit breaker+car	circuit breaker	car	49	circuit breaker ＋train	train	train
24	circuit breaker+car	circuit breaker	train	50	circuit breaker	train	train
25	car＋circuit breaker	circuit breaker	ring	51	circuit breaker	train	car
26	car＋circuit breaker	car	ring	52	car	circuit breaker	car

Table 3: Recognition Result of Sample C for Each Second

C	Sound	Recognition1	Recognition2	C	Sound	Recognition1	Recognition2
1	circuit breaker	no sound	ring	28	train	train	train
2	circuit breaker	no sound	car	29	train	train	train
3	circuit breaker	circuit breaker	car	30	train＋circuit breaker	train	train
4	circuit breaker	circuit breaker	car	31	train＋circuit breaker	train	alarm
5	circuit breaker	circuit breaker	car	32	car	train	alarm
6	circuit breaker	circuit breaker	train	33	car	car	train
7	circuit breaker	circuit breaker	car	34	car	car	train
8	circuit breaker	circuit breaker	car	35	car	car	train
9	circuit breaker	circuit breaker	ring	36	car	car	train
10	circuit breaker	circuit breaker	car	37	car	car	train
11	circuit breaker	circuit breaker	car	38	car	car	train
12	circuit breaker	circuit breaker	car	39	car	car	train
13	circuit breaker	circuit breaker	car	40	car	car	train
14	circuit breaker	circuit breaker	car	41	car	car	train
15	circuit breaker	circuit breaker	car	42	car	car	car
16	circuit breaker	circuit breaker	car	43	car	--	car
17	circuit breaker	circuit breaker	car	44	car	--	car



18	circuit breaker	circuit breaker	car	45	car	circuit breaker	car
19	circuit breaker	circuit breaker	train	46	car	circuit breaker	ring
20	circuit breaker	circuit breaker	car	47	car	car	ring
21	circuit breaker	circuit breaker	car	48	car	car	no sound
22	train＋circuit breaker	train	alarm	49	car	car	train
23	train	train	alarm	50	car	car	train
24	train	train	train	51	car	car	train
25	train	train	train	52	car	car	train
26	train	train	train	53	car	car	train
27	train	train	train				

Figure 8: View of Subject

5.3 Sound Source Direction Recognition Function

We evaluated the sound source direction recognition part. We used the microphone array, which is in the Kinect to detect the direction of the sound source. A sample of the microphone array used to detect the direction of a sound source is shown in Figure 5. This figure shows two of the same AR markers in different positions, and through the detection of the sound source direction, we found whether the object making sound is the right one in the picture.

To detect the direction of a sound source, we fixed the web camera on the Kinect so that it could capture video in a range of 40 degrees. We used Kinect to detect the direction of a sound source in this range in front of it at the same time and matched the direction with the result of the AR marker recognition to specify the position of the sound source. We found that the position of the sound source, which was calculated by the Kinect, was in a range of 20 pixels on the x vector of the AR marker on the screen. We considered the reasons for this are as follows.

- To match the result calculated by the Kinect with the position of AR marker, we needed to convert degrees to pixels; this conversion could have caused the deviation.
- We used Microsoft's official SDK for Kinect to detect the direction of a sound source. Therefore, the Kinect may have had a deviation that interfered with detecting the direction.
- Furthermore, the Kinect could only detect the direction in a range of 100 degrees and 3 meters in front of it. Therefore, the user could hardly find sounds that the sound sources were beside or far from him or her.

5.4 System Confirmation

We tested our prototype system experimentally. We prepared a cellphone, shredder, guitar, and alarm as four kinds of sound source. We attached different AR markers on each object, and put them around the subject. The subjects wore a web camera and Kinect. When they were hearing the sound by themselves, the subjects wore earphones. The subject's view is shown in Figure 8. When the sound source object was within the range of web camera, the subject found an 'X' on the top of the sight. Whether the sound source was out of range, there would be a 'Left' if the sound source was on the left of the subject, or a 'Right' if the source was on the right of the subject.

We received the following comments from the subjects.

- The sound source was very easy to understand.
- Because of AR, the position of the sound source was very easy to locate.

We found the following problems with the prototype.

- There was a delay in showing the user the position of the sound source.
- When the sound source is behind the user, the microphone array told them the direction, but the accuracy was lower than one in front of the user.

5.5 Comprehensive Evaluation

We confirmed that the source and direction of sound could be effectively recognized, and that augmented reality was implemented, and thus that the user could use our system to recognize and visualize environmental sounds. When there was only a single sound source in the surrounding environment such as at home or when doing some simple work, and especially when a source was near a user, our system provided information on the sound source and visualized the sound source to satisfy the user's need. However, in outdoor environments, it was difficult to recognize the sound source. The recognition accuracy falls because of the complexity of the sound source environment. Furthermore, if the AR marker was taken far away from the web camera mounted on the user, the recognition accuracy of sound source position could fall.

6 Conclusion

We propose a sound source visualization system that uses augmented reality so that the hearing impaired and those who are working indoors can recognize environmental sounds, and, in particular, sounds indicating danger. The prototype system, which was implemented, recognized a sound source and the position of the sound source and visualized the sound source in the user's field of vision by making use of AR.

As future work, we will also discuss how to use the method to recognize several sound sources at the same time. We need to raise the recognition accuracy in complex sound environments because the user needs to grasp the details of the surrounding environment. We also need to evaluate the new method of recognition and to test our system in the outdoor environment, evaluate the effectiveness and efficacy of its visualization.

7 Acknowledge

This research was supported in part by a Grant in aid for Precursory Research for Embryonic Science and Technology (PRESTO) from the Japan Science and Technology Agency and by a Grant-in-Aid for Scientific Research(A)(20240009) from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

8 References

Kawai, Y. and Tomita, F. (2000), "A Visual Support System for Visually Impaired Persons Using Acoustic Interface", in IAPR Workshop on Machine Vision Applications 2000 proceedings of the international conference in Tokyo, Japan, 2000, pp. 379-382.

Matthews, T., Fong, J. and Mankoff, J. (2005), "Visualizing Non-Speech Sounds for the Deaf", in ACM International Conference On Computers and Accessibility 2005 proceedings of the international conference in Maryland, USA, 2005, pp. 52–59.

Asano, F., Goto, M., Itou, K. and Asoh, H. (2001), "Real-time Source Localization and Separation System and Its Application to Automatic Speech Recognition", in European Conference on Speech Communication and Technology 2001 proceedings of the international conference in Aalborg, Denmark, 2001, pp. 1013–1016.

Ho-Ching, F. W., Mankoff, J. and Landay, J. A. (2003), "Can You See What I Hear? The Design and Evaluation of a Peripheral Sound Display for the Deaf", in International Conference on Human Factors in Computing Systems 2003 proceedings of the international conference in Fort lauderdale, USA, 2003, pp. 161–168.

Somervell, J., McCrickard, D. S., North, C. and Shukla, M. (2002), "An Evaluation of Information Visualization in Attention-Limited Environments", in Joint Eurographics - IEEE TCVG Symposium on Visualization 2002 proceedings of the international conference in Barcelona, Spain, 2002, pp. 211– 216.

Feiner, S., MacIntyre, B., Hollerer, T. and Webster, A. (1997), "A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment", in International Symposium on Wearable Computers 1997 proceedings of the international conference in Massachusetts, USA, 2997, pp. 208– 217.

Azar, J., Saleh, H. A. and Al-Alaoui, M. A. (2007)," Sound Visualization for the Hearing Impaired", International Journal of Emerging Technologies in Learning, Vol. 2 , No. 1.

Bertram, M., Deines, E., Mohring, J., Jegorovs, J. and Hagen, H. (2005), "Phonon Tracing for Auralization and Visualization of Sound", in IEEE International Symposium on Visualization 2005 proceedings of the international conference in Minneapolis, USA, 2005, pp. 151–158.

Osaka, N., Saito, Y., Ishitsuka, S. and Yoshio, Y. (2009), "An Electronic Timbre Dictionary and 3D Timbre Display", in International Computer Music Conference 2009 proceedings of the international conference in Montreal, Canada, 2009, pp. 9–12.

Zheng, F., Zhang, G. and Song, Z. (2001), ""Comparison of Different Implementations of MFCC", Journal of Computer Science and Technology, pp. 582–589.

Clarkson, B., Sawhney, N. and Pentland, A. (1998), "Auditory Context Awareness via Wearable Computing", in International Workshop on Perceptual User Interfaces 1998 proceedings of the international conference in San Francisco, USA, 1998, pp. 4–6.

Gook, K. H. and Sikora, T. (2004), "Audio Spectrum Projection Based on Several Basis Decomposition Algorithms Applied to General Sound Recognition and Audio Segmentation", in European Signal Processing Conference 2004 proceedings of the international conference in Vienna, Austria,2004, pp. 1047–1050.

Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K. and Tachibana, K. (2000), "Virtual Object Manipulation on a Table-Top AR Environment", in International Symposium on Augmented Reality 2000 proceedings of the international conference in Munich, Germany,2000, pp. 111–119.

Rakerd, B. and Hartmann, W. M. (2010), "Localization of Sound in Rooms. V. Binaural Coherence and Human Sensitivity to Interaural Time Differences in Noise", Journal of the Acoustical Society of America, pp. 3052–3063.

Nakadai, K., Matsuura, D., Okuno, H. G. and Kitano, H. (2003), "Applying Scattering Theory to Robot Audition System: Robust Sound Source Localization and Extraction", in International Conference on Intelligent Robots and Systems 2003 proceedings of the international conference in Las Vegas, USA, 2003, pp. 1147–1152.

Kaper, H. G., Tipei, S. and Wiebel, E. (1999), "Data sonification and sound visualization", Journal of the Computing in Science and Engineering, Vol. 4 No.1, pp. 48–58.

Valin, J. M., Michaud, F., Rouat, J. and Letourneau, D. (2003), "Robust Sound Source Localization Using a Microphone Array on a Mobile Robot", In International Conference on Intelligent Robots and Systems 2003 proceedings of the international conference in Las Vegas, USA, 2003, pp. 1228–1233.

Hollerer, T., Feiner, S., Terauchi, T., Rashid, G. and Hallaway, D. (1999), "Exploring MARS: Developing Indoor and Outdoor User Interfaces to a Mobile Augmented Reality System", International Journal of Computers and Graphics, Vol. 23 No. 6, pp. 779–785.