KIKIMIMI
Voice Separation System for Automating Post Evaluation of Learning Support System

Takahiro Nakadai[1], Tomoki Taguchi[1], Ryohei Egusa[2], Miki Namatame[3], Masanori Sugimoto[4], Fusako Kusunoki[5], Etsuji Yamaguchi[2], Shigenori Inagaki[2], Yoshiaki Takeda[2] and Hiroshi Mizoguchi[1]

[1]*Department of Mechanical Engineering, Tokyo University of Science, 2641 Yamazaki, Noda-shi, Chiba-ken, Japan*
[2]*Graduate School of Human Development and Environment, Kobe University, Hyogo, Japan*
[3]*Tsukuba University of Technology, Ibaraki, Japan*
[4]*Hokkaido University, Hokkaidou, Japan*
[5]*Department of Computing, Tama Art University, Tokyo, Japan*

Keywords: Kinect Sensor, Microphone Array, Signal Processing, Learning Support System.

Abstract: A learning support system with a body experience has a favorable influence on learning in children because they obtain a sense of immersion. It is important to evaluate the learning effect of a learning support system. However, the learning effect of the learning support system was almost evaluated manually in previous research. The authors propose an evaluation system called "KIKIMIMI" for automating the post-evaluation of a learning support system by reactions from the learner's voice. In this paper, we report the validity of KIKIMIMI as a system for automating a post-evaluation.

1 INTRODUCTION

A learning support system with a body experience has a favorable influence on learning in children because they obtain a sense of immersion. It is also very good for children to interact with computers. The extinct animals learning system is the major example (Figure 1-2) (Tomohiro Nakayama, 2014). The extinct animals learning can only be learned with a textbook because it is not possible to go back in time. Therefore, a learning support system with a body experience is effective.

However, the learning effect of the learning support system was almost evaluated manually in previous research using, for example, a questionnaire, an interview, a post-evaluation by a recording, and so on. The authors focused on a post-evaluation using a learner's voice recording.

The technology for automating the post-evaluation of a learning support system by reactions from the learner's voice has been thoroughly researched (Tomoki Taguchi, 2014). The technology researched until now is shown in Figure 1. Automation of the post-evaluation by a static learner's voice recording is shown in Figure 1. However, research on the

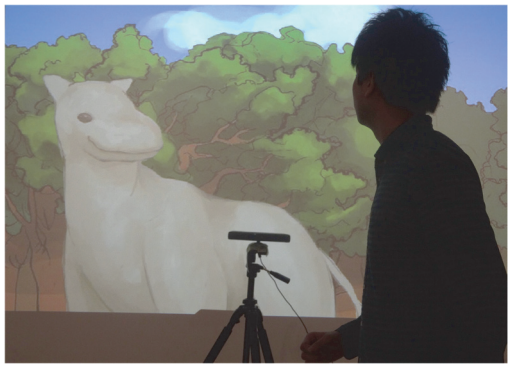

Figure 1: The extinct animals learning system.

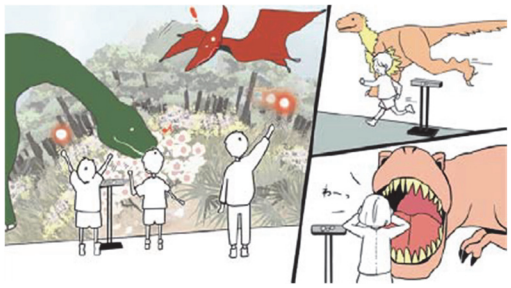

Figure 2: Learning support system.

Nakadai T., Taguchi T., Egusa R., Namatame M., Sugimoto M., Kusunoki F., Yamaguchi E., Inagaki S., Takeda Y. and Mizoguchi H..
KIKIMIMI - Voice Separation System for Automating Post Evaluation of Learning Support System.
DOI: 10.5220/0005476303010306
In *Proceedings of the 7th International Conference on Computer Supported Education* (CSEDU-2015), pages 301-306
ISBN: 978-989-758-108-3

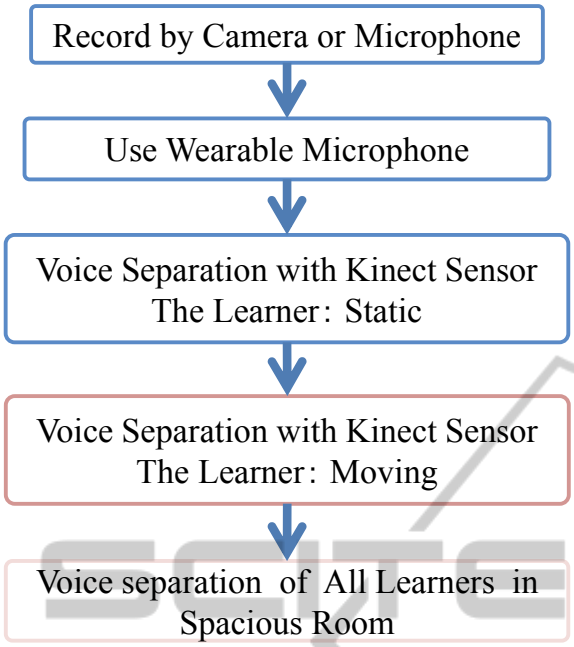

Figure 3: Flowchart of the developed technology.

evaluation by a moving learner's voice recording is being developed. If this research advances, all of the learners' voices in a spacious room can be separated selectively. This research is very important (Figure 3).

In this research, we propose an evaluation system called "KIKIMIMI" for automating the post-evaluation of a learning support system by reactions from the learner's voice. In this paper, we describe the details of KIKIMIMI.

2 AUTOMATING THE POST-EVALUATION SYSTEM

2.1 Problems Associated with Wireless Microphones

Here, we describe the methods used to record a learner's voice from previous research. It is important for the field of education to perform an evaluation by recording a voice. As a resort, learners were recorded by a wearable microphone, e.g., a noise-cancelling microphone and wireless microphone (Masafumi Goseki, 2012). A wearable microphone has some problems. It is a burden for children to wear a heavy wireless microphone. In addition, a distraction, psychological and physical burdens, and an increase in tension of children are also problems associated with the use of a wireless microphone. A distraction is caused when children recognize the wearable microphone. As a result, children's interest decreases. Psychological and physical burdens are caused when the wearable microphone is too heavy for children. As a result, the children cannot act as a natural learner. An increase in tension is caused when the learner is in an unnatural state. Children almost never have a wearable microphone in a general environment. Children may be uncomfortable because they are unnatural with the wearable microphone. Therefore, a noncontact recording technique is needed in the field of education. We propose KIKIMIMI to solve these problems.

2.2 Proposed Techniques

We now describe the details of the proposed techniques. There are many necessary techniques for noncontact recordings, e.g., a hands-free function, noise-canceling, and a voice-separation technique (Takahiro Nakadai, 2014). These techniques do not need a medium to record using a wearable microphone. The voice-separation technique can selectively separate to remove noise. This technique is noncontact recording.

In this research, we utilized the voice-separation technique because it is a noncontact recording technique and could capture sound at an objective angle locally. This technique is combined with a microphone array with signal processing that can separate children's voices from background noise in general living environments.

We proposed the KIKIMIMI system for automating the post-evaluation of a learning support system by reactions from the learner's voice in this research.

3 VOICE-SEPARATION SYSTEM

3.1 Separation Technique

Here, we describe the voice-separation technique applied to KIKIMIMI. We selected the method called delay-and-sum beamforming (DSBF) because it is robust in real environments (Ngoc-Vinh Vu, 2010). DSBF is a method in which an objective voice can be amplified. The method uses the time lag of an input signal during signal processing. The outline of this method is shown in Figure 4. The delay times are calculated by the distance from an objective talker to each microphone. An objective

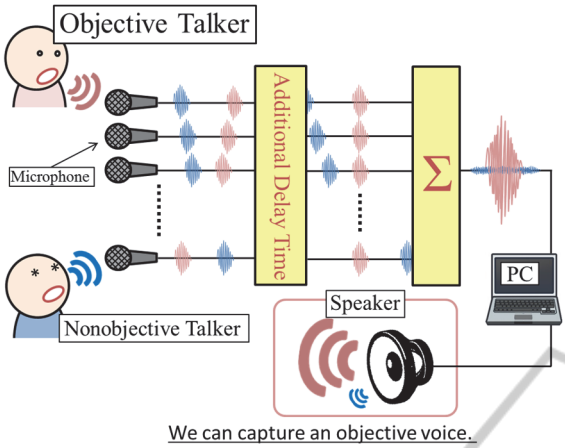

Figure 4: Schematic of the DSBF method.

voice can be amplified with an additional delay time. This position of an objective voice is called "focus," which is applied for voice separation of the objective talker.

However, the moving learner's position is uncertain when a moving learner is playing with the learning support system. In order to determine their focus, we need a way to recognize the moving learner's position. Therefore, we selected the Kinect sensor by Microsoft (Riyad A. El-laithy, 2012). We can capture a learner's head-position information with the recording-time information from the depth sensor. We can continue updating to determine the focus of the moving learner's head position. As a result, the voice-separation technique is able to be applied to a moving learner. We can capture the objective learner's voice from the background noise.

3.2 Details of KIKIMIMI

A photograph of the Kinect sensor is shown in Figure 5. The details of the Kinect sensor are shown in Figure 6. The microphones in this Kinect sensor are installed as four elements. Four microphone elements can synchronize a recording. The depth sensor can obtain the head-position information of a detected person by recording the time information. Four recording files and the head-position information can synchronize a recording.

The details of KIKIMIMI are shown in Figure 7. Four recording files are obtained from the four microphone elements in the Kinect sensor. We can capture the learner's head-position information with the recording-time information from the depth sensor. We can continue updating to determine the focus of the moving learner's head position combined with the head-position and recording-time information. As a result, the voice-separation technique is able to be applied to a moving learner. We can capture the objective learner's voice from the background noise.

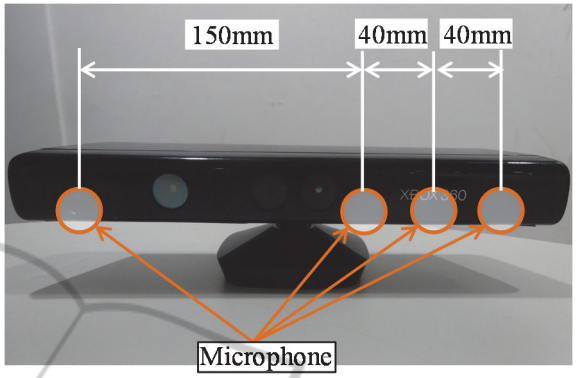

Figure 5: Photograph of the Kinect sensor.

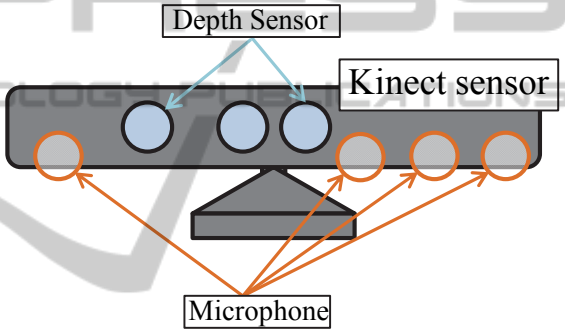

Figure 6: Kinect sensor.

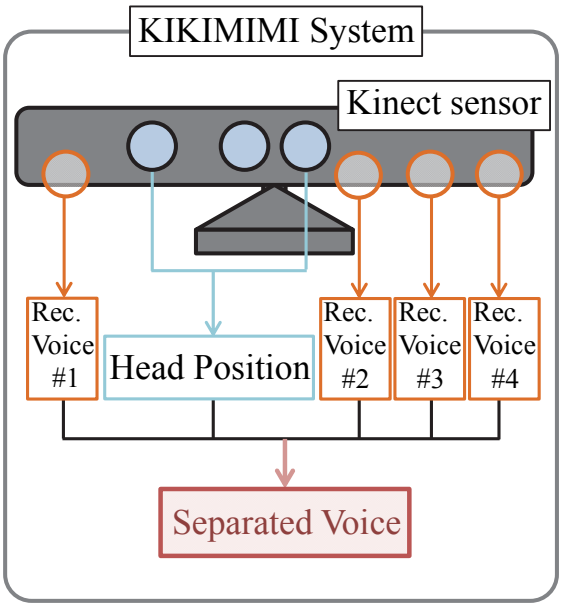

Figure 7: Voice-separation system.

4 OPERATING KIKIMIMI

4.1 Operating Environment

Here, we describe the operating environment when KIKIMIMI is operating, as shown in Figure 8–10. First, KIKIMIMI was installed. Second, the learner moved linearly in front of KIKIMIMI (2.5–3.0 m). Finally, two people talking were placed near the learner.

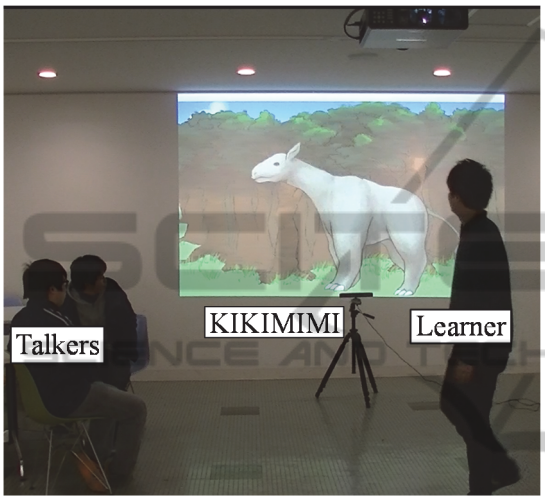

Figure 8: Operating state of KIKIMIMI.

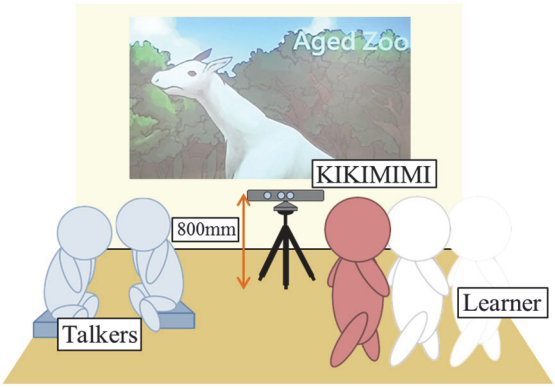

Figure 9: Operating environment #1.

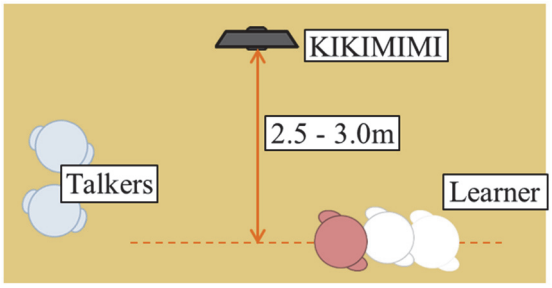

Figure 10: Operating environment #2.

Two people were talking without minding anything. We confirmed that KIKIMIMI operated in an environment with confusing background noise such as environmental noise and another talking voice.

KIKIMIMI operated using the Kinect sensor. We can obtain four recording files. We can also obtain the learner's head-position information with recording-time information from the depth sensor. Some information obtained by KIKIMIMI was utilized for the sound-separation technique.

4.2 Operating Results

Some results obtained by KIKIMIMI are now presented. We compared the results before voice-separation signal processing with those after this signal processing. The results are shown in Figure 11–12. Figure 11 shows a waveform before voice-separation signal processing. Figure 12 shows a waveform after voice-separation signal processing.

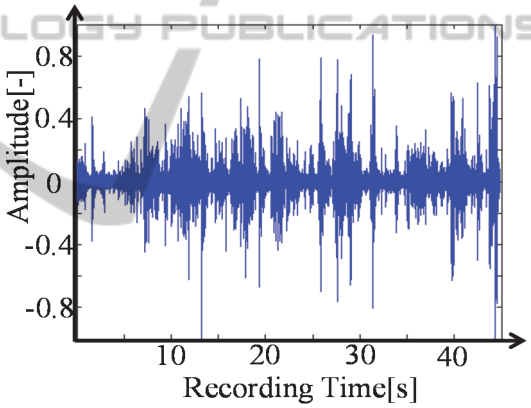

Figure 11: Before signal processing.

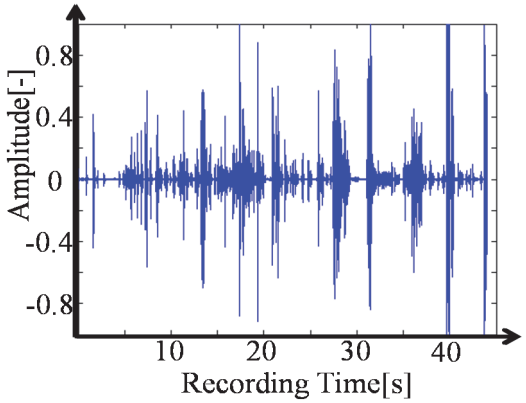

Figure 12: After signal processing.

The vertical and horizontal axes show the voice amplitude and recording time, respectively. Figure 11 shows are very noisy waveform. The voices of

two people talking and environmental noise are present in Figure 11. Consequently, the learner's voice was not able to be heard because it was obscured by noise. Figure 12 is clearer than Figure 11. We can recognize some of the recorded voices, e.g., "It's very big," "Have a good coat of hair," "They live in such a forest," "It's cute," and so on. We can also recognize a nonverbal voice in some of the recorded voices, e.g., "Oh," "Awesome," "What," and so on. As a result, we can also recognize nonverbal voices or reactions.

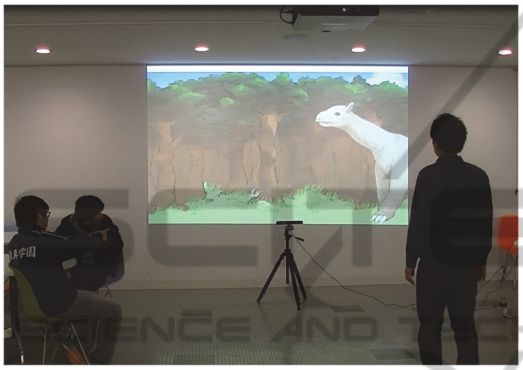

Figure 13: Operating KIKIMIMI after 10 s.

Figure 14: Operating KIKIMIMI after 20 s #1.

Figure 15: Operating KIKIMIMI after 20 s #2.

Figure 16: Operating KIKIMIMI after 30 s.

Figure 17: Perspective from Kinect sensor #1.

Figure 18: Perspective from Kinect sensor #2.

The operating states of KIKIMIMI are shown in Figure 13–16. Figure 13 shows the operation of KIKIMIMI after 10 s, where the learner gets used to the learning support system. In the same way, this is the period of time during which the two people begin talking. In Figure 14–15, the operation of KIKIMIMI is shown after 20 s, where the learner moves linearly. Figure 16 shows the operation of KIKIMIMI after 30 s, where the learner was finding

a characteristic of the learning support system. The learner stopped and vocalized nonverbal voices.

As a reference, the perspectives from the Kinect sensor are shown in Figure 17–18. From the above, we can obtain a noncontact recording of the moving learner's voice when they are playing with the learning support system. Consequently, it is concluded that KIKIMIMI is a suitable system for automating a post-evaluation.

5 CONCLUSIONS

In this paper, we proposed an evaluation system called "KIKIMIMI" for automating the post-evaluation of a learning support system by reactions from a learner's voice. The focus of this study was on the post-evaluation of a learner's voice recording. We selected a voice-separation technique because it could capture a noncontact recording and sound at an objective angle locally. We could capture a learner's head-position information by recording the time information from the depth sensor. We aimed to obtain a noncontact recording of the moving learner's voice when they are playing with the learning support system.

In this research, we confirmed that KIKIMIMI operated in environment with confusing background noise for automating a post-evaluation. As a result, it is suggested that KIKIMIMI can possibly capture an objective voice.

In the future, KIKIMIMI will be used as a system for automating a post-evaluation.

ACKNOWLEDGEMENTS

This work was supported in part by Grants-in-Aid for Scientific Research (B). I am particularly grateful for the illustration (Figure 2) produced by Ms. Midori Aoki.

REFERENCES

Tomohiro Nakayama, Ryuichi Yoshida, Takahiro Nakadai, Takeki Ogitsu, Hiroshi Mizoguchi, Kaori Izuishi, Fusako Kusunoki, Keita Muratsu, and Shigenori Inagaki, "Immersive Observation Support System toward Realization of 'Interactive Museum' -Observing 'Live' Extinct Animals while Walking in a Virtual Paleontological Environment-," Proceedings of the 11th International Conference on Advances in Computer Entertainment Technology (ACE2014), Poster_149(1)-(4), November 2014.

Tomoki Taguchi, Ryohei Egusa, Masanori Sugimoto, Fusako Kusunoki, Etsuji Yamaguchi, Shigenori Inagaki, Yoshiaki Takeda, and Hiroshi Mizoguchi, "Developing Voice Separation System for Support Education Research: Determining Learner Reaction without Contact," Journal of Convergence Information Technology (JCIT), Vol. 9, No. 3, pp.12-17, May 2014.

Takahiro Nakadai, Tomohiro Nakayama, Tomoki Taguchi, Ryohei Egusa, Miki Namatame, Masanori Sugimoto, Fusako Kusunoki, Etsuji Yamaguchi, Shigenori Inagaki, Yoshiaki Takeda, and Hiroshi Mizoguchi, "Sound-Separation System using Spherical Microphone Array with Three-Dimensional Directivity-KIKIWAKE 3D: Language Game for Children," International Journal on Smart Sensing and Intelligent Systems (S2IS), Vol. 7, No. 4, pp.1908-1921, December 2014.

Masafumi Goseki, Ryohei Egusa, Takayuki Adachi, Hiroshi Mizoguchi, Miki Namatame, Fusako Kusunoki, Shigenori Inagaki, "Puppet Show for Entertaining Hearing-Impaired, Together with Normal-Hearing People—A Novel Application of Human Sensing Technology to Inclusive Education," International Conference on Innovative Engineering Systems (ICIES2012), pp.121-124, December 2012.

Nogoc-Vinh Vu, Hua Ye, J. Whittington, J. Devlin, and M. Mason, "Small Footprint Implementation of Dual-Microphone Delay-and-Sum Beamforming for In-Car Speech Enhancement," IEEE International Conference on Acoustics, Speech, and Signal Processing, pp.1482-1485, March 2010.

Riyad A. El-laithy, Jidong Huang, and Michael Yeh, "Study on the Use of Microsoft Kinect for Robotics Applications," IEEE/ION International Conference on Position Location and Navigation Symposium (PLANS), pp.1280-1288, April 2012.