Development of an Automatic Location-determining Function for Balloon-type Dialogue in a Puppet Show System for the Hearing Impaired

Ryohei Egusa[1,2], Shuya Kawaguchi[3], Tsugunosuke Sakai[3], Fusako Kusunoki[4], Hiroshi Mizoguchi[3], Miki Namatame[5] and Shigenori Inagaki[2]

[1]*IPSJ Research Fellow, Tokyo, Japan*
[2]*Kobe University, Hyogo, Japan*
[3]*Tokyo University of Science, Chiba, Japan*
[4]*Tama Art University, Tokyo, Japan*
[5]*Tsukuba University of Technology, Ibaraki, Japan*

Keywords: Puppet Show, People with Hearing Impairment, Dialogue Presentation, Balloons.

Abstract: People with hearing impairments have a tendency to experience difficulties in obtaining audio information. They have difficult to watch puppet shows. In this study, we have undertaken the development of a dialogue presentation function in a puppet show system for the hearing impaired. The dialogue presentation function that was developed is an automatic location-determination system for balloon-type dialogue. The balloon-type dialogue turns the lines of dialogue into text and displays it as balloons in the background animation of the puppet show. This function automatically places the balloon-type dialogue in the vicinity of the locations of the puppets. In the evaluation test, 24 college students with a hearing impairment were used as participants, and a comparison was made between a system that implemented the automatic location-determination function for balloon-type dialogue and a system that did not implement said function. These results indicate the effectiveness of the location tracking function for balloon-type dialogue as a means for ensuring access to audio information in puppet shows.

1 INTRODUCTION

In this study, we have undertaken the development of a dialogue presentation function in a puppet show system for the hearing impaired. The dialogue presentation function that was developed is an automatic location-determination system for balloon-type dialogue. This function automatically places the balloon-type dialogue in the vicinity of the locations of the puppets.

Puppet shows are an important cultural form; they are deeply rooted in the cultures of Europe, Asia, Africa, South America, and Oceania (Los Angeles County Museum. and Zweers, 1959; Blumenthal, 2005).

People with hearing impairments have a tendency to experience difficulties in obtaining audio information, and the percentage who rely primarily on visual information is higher among such people than among people whose hearing is unimpaired (Marschark and Hauser, 2012). Owing to this, it is difficult for the hearing impaired to watch puppet shows. One reason is that audio information such as the lines of dialogue and background music are important in understanding the story in puppet shows. Another reason is that the facial expressions and movements of puppets are limited compared to those of humans, and less information can be obtained from them. Despite this, studies of systems that would ensure information access for the hearing impaired are scarce in the field of puppetry.

Dialogue presentation is one technique for ensuring that the hearing impaired has access to information (Rander and Looms, 2010; Stinson et al., 2014). Egusa et al. (2016) present authors have undertaken the development of a puppet show system that implements a dialogue presentation function that facilitates the viewing of such shows by the hearing impaired. In the puppet show system in Egusa et al. (2016), a balloon-type dialogue presentation function was implemented as a way to

Egusa, R., Kawaguchi, S., Sakai, T., Kusunoki, F., Mizoguchi, H., Namatame, M. and Inagaki, S.
Development of an Automatic Location-determining Function for Balloon-type Dialogue in a Puppet Show System for the Hearing Impaired.
DOI: 10.5220/0006377203400344
In *Proceedings of the 9th International Conference on Computer Supported Education (CSEDU 2017) - Volume 2*, pages 340-344
ISBN: 978-989-758-240-0

ensure information access. The balloon-type dialogue presentation function turns the lines of dialogue into text and displays it as balloons in the background animation of the puppet show. The balloons have been embedded beforehand in the background animation and can be displayed using a PC operation. The balloons are an effective dialogue presentation method on the following two scores. First of all, the lines that have been turned into text are displayed in the vicinity of the puppets. According to Fukamauchi et al. (2007), it is known empirically that persons with a hearing impairment actively utilize their peripheral vision during communication. Owing to the fact that the balloons are displayed in the vicinity of the puppets, it can be expected that the movements of the puppets and the lines can be viewed and heard simultaneously. Second, the shape of the balloons connects the lines with the puppet that is currently speaking. In venues such as a seminar, where multiple speakers may be talking in succession, the effectiveness of providing dialogue in a form that indicates the respective speakers has been confirmed (Wald, 2008; Hisaki, Nanjo and Yoshimi, 2010). With regard to the matching of the uttered content to a speaker, a method by which the audio information that has been turned into text is displayed in the vicinity of the speaker as balloon-type dialogue is superior as it can easily convey the connection between the speaker and the utterance in a mere glance (Hong et al., 2010; Hu et al., 2015). In puppet shows as well, when several characters are speaking, it is important that the character who is speaking is clearly indicated visually.

In Egusa et al. (2016), it is clear that balloon-type dialogue is effective as a method for ensuring access to the information in a puppet show to children with a hearing impairment. However, there was still a need to improve the method whereby such information was displayed. The reason for this is that the location of the display of the dialogue has been determined prior to the performance. The puppeteer must change the location of the puppets according to the location of the dialogue that is displayed, and inevitably this results in a time lag in positioning the puppets properly relative to the dialogue. In addition, the locations of the dialogue and the puppets appear different from the viewpoint of the puppeteer and that of the spectators. Owing to these reasons, there is a chance that the following three kinds of information may be lost.

First, it is possible that information about the movements of the puppets may be lost. When the relative locations of the dialogue and puppets are far apart, a spectator cannot keep both the puppets and the dialogue display within his or her field of vision. Owing to this, the movements of the puppets may be missed while the spectator reads the dialogue text.

Second, it is possible that it may become unclear to which character the lines of dialogue belong owing to the fact that the relative locations of the dialogue text of the speaker and the puppets are not fixed.

Third, there is a possibility that the dialogue and the characters may overlap, and it may not be possible to read the lines from the spectator's position.

In order to reduce the possibility of the information loss indicated above, it is necessary that the following three conditions be maintained at all times during the performance of a puppet show.

1) The location of the dialogue is displayed in the vicinity of the puppets.

2) The display is in a location where the puppets and the dialogue do not interfere with one another.

3) The dialogue is displayed in real time with the speaking of the character.

In order to achieve these conditions, we adopted the method of measuring the locations of the puppets with a depth image sensor and automatically determining the locations of the dialogue. By this means, it is anticipated that the location of the dialogue can be determined appropriately and in real time, and the loss of information can be reduced.

In this article, we report about the development of the automatic location-determination function for balloon-type dialogue, the details of the system, an overview of the evaluation test, and the preliminary analysis results. The purpose of the evaluation test was to assess the effectiveness of the automatic location-determination function for balloon-type dialogue in ensuring access to audio information. In the evaluation test, 24 college students with a hearing impairment were used as participants, and a comparison was made between a system that implemented the automatic location-determination function for balloon-type dialogue and a system that did not implement said function.

2 SYSTEM DESIGN

In order to achieve the three conditions described in Section 1, we undertook the development of a balloon-type dialogue presentation system for puppet shows that employs a range image sensor. First, we give an explanation of an overview of a puppet show. Figure 1 shows the configuration of a stage for a puppet show. The stage is composed of a screen, projector, PC, and range image sensor. One puppeteer operates one character. The animation operator is in charge of the operation of the background animation and the control of the depth image sensor by means of a PC. The background

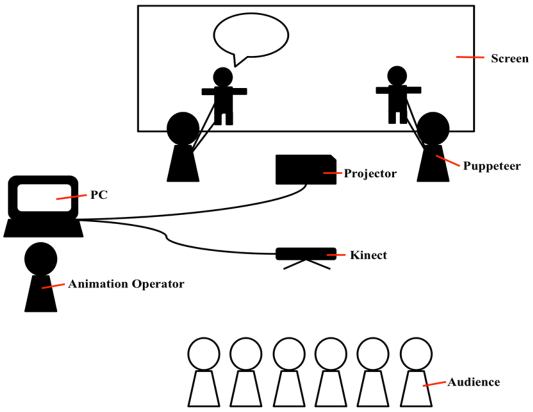

Figure 1: The configuration of a stage for a puppet show.

animation is projected on the screen by the projector.

Next, we recount the details of the configuration of the balloon-type dialogue system. Figure 2 shows the settings of the automatic location-determination function for balloon-type dialogue. Kinect V2 (made by Microsoft) (hereinafter “Kinect sensor”) is employed as the depth image sensor for the control of the automatic location-determination function. The region recognized by the Kinect sensor is 4,000 mm vertically and 3,300 mm horizontally from a distance of 3 meters. The recognition region can be divided into a user-specifiable number of subregions. Numerical values can be selected for the width and for the maximum number of subregions. For this study, it was set as four regions with a minimum width of 500 mm in order to minimize unforeseen changes of the dialogue display location resulting from shaking of the puppeteer’s hands.

Lastly, we provide a description of the mechanism of the automatic tracking function. First, the Kinect sensor tracks the location of the hands of the puppeteer within the recognition region. Next, the Kinect sensor determines the location of the puppeteer’s hands being captured in subregions 1–4. Finally, on the background animation, the balloon-type dialogue is displayed in the balloon display region corresponding to the subregion where the hands are. By this means, the three conditions are satisfied at all times, and the location of the dialogue is determined appropriately and in real time.

Figure 3 is the display of a dialogue in a stage performance. In the venue shown in Figure 3 (top), the dialogue is displayed in the part that is directly above the location of the puppet in the center of the figure. In the venue shown in Figure 3 (bottom), the place of displaying dialogue is moved into the left end of the screen in accordance with the puppet moving.

3 EVALUATION

3.1 Method

Participants: 24 college students with a hearing impairment (group 1: 10 students; group 2: 14 students).

Tasks: In the evaluation test, a new system with an automatic location-determination function for the dialogue and a conventional system with a line projection function wherein the location of the dialogue was determined beforehand were prepared, and the authors had the participants observe these. For the subjective evaluation, subjects evaluated each of the systems individually for the puppet show

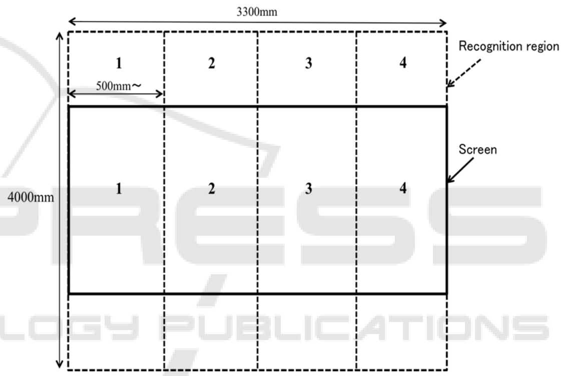

Figure 2: Specifications of the Automatic Tracking Function.

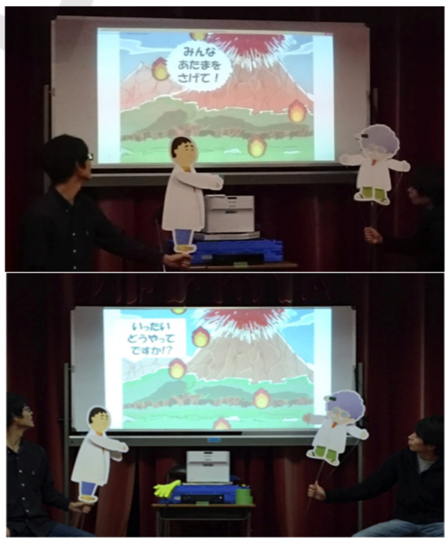

Figure 3: The display of a dialogue in a stage performance.

that was performed, and a comparative evaluation was performed wherein subjects were asked at the end of the show which of the two systems was superior. In this article, we report the results of a preliminary analysis of the subjects' comparative evaluation of the systems.

The data for the comparative system evaluation were the subects' answers to the question of which was superior, the new system or the conventional system, with regard to the following three aspects. Aspect 1 is the understanding of the content of the balloons, for which there were two items: "I could understand the contents of the balloons better" and "The balloons were easy to read." Aspect 2 is the ease of simultaneous recognition of the balloons and of the behavior of the puppets, for which there were two items: "The location of the balloons was just right" and "I could see both the balloons and the movements of the puppets." Aspect 3 is the ease of identifying the speaking character, for which there was one item: "I could tell which character was speaking."

Procedure: We prepared five short scenes of a puppet show. The scenes consisted of the elements of two characters on stage, balloon-type dialogue of 5–6 phrases, background animation, and action corresponding to the situation.

The participants viewed the same scenes of the show performed with the new system and with the conventional system. Each time they finished watching a scene, the participants completed the subjective evaluations by system for that scene. The viewing order differed for each group: Group 1 viewed the scenes first by the conventional system and then by the new system; Group 2 viewed the scenes first by the new system and then by the conventional system. This was done in order to achieve a counterbalance. After the viewing of all of the scenes was finished, the participants completed the comparative evaluation of the systems.

3.2 Result

Table 1 summarizes the results of the comparative evaluation. For all items, the number of affirmative responses for the new system surpassed the number of affirmative responses for the conventional system. For the items related to the understanding of the content of the balloons, "I could understand the contents of the balloons better" and "The balloons were easy to read," the new system was positively evaluated by 70% of the participants. This indicates that acquisition of the lines of dialogue was easier with the new system than with the conventional system.

For the items related to the ease of simultaneous recognition of the balloons and of the behavior of the puppets, "The location of the balloons was just right" and "I could see both the balloons and the movements of the puppets," more than 60% of the participants selected the new system. This may be the result of the fact that the loss of information related to the behavior of the puppets was reduced since the location of the balloons relative to the puppets was set appropriately.

Similarly, for the item related to the ease of identification of the character speaking, "I could tell which character was speaking," 70% of the participants selected the new system. This may be a result of the fact that the loss of information related to the identification of the speaking character was reduced since the location of the balloons relative to the puppets was set appropriately.

The participants' responses were classified into 'positive responses for new system' (selected "new system") and 'neutral and negative responses' (selected "conventional system" or "neither"). Then, for each item, the difference between the positive responses and the neutral/negative responses was analyzed using Fisher's exact test. It was found that the number of positive responses exceeded that of the neutral/negative responses for all item, and the difference was statistically significant.

Table 1: Results of Subjective Evaluation by Hearing-Impaired College Students.

Item	New system	Conventional system	Neither
I could understand the contents of the balloons better**	17	2	5
The balloons were easy to read**	16	5	3
The location of the balloons was just right**	15	2	7
I could see both the balloons and the movements of the puppets**	15	2	7
I could tell which character was speaking**	17	2	5

New system: System incorporating an automatic tracking function.

Conventional system: System in which the location of the balloons is pre-set.

$N = 24$. $^{**}p < .01$

4 CONCLUSIONS AND FUTURE WORKS

In this study, the development of an automatic tracking function for balloon-type dialogue, whose aim is the improvement of the usability of the dialogue projection function, was undertaken. In the analysis of the results of the evaluation test, it was seen that the location tracking function for balloon-type dialogue was positively evaluated with regard to the understanding of the dialogue content, its recognition simultaneous with that of the behavior of the puppets, and identification of the character speaking.

These results indicate the effectiveness of the location tracking function for balloon-type dialogue as a means for ensuring access to audio information in puppet shows.

However, in the case of the aspect related to the simultaneous recognition of the dialogue and of the puppets' behavior, about 60% of the participants chose the new system, and no major difference against the conventional system was observed. A detailed analysis of the subjective system evaluation for each scene can be noted here as one issue to be addressed in the future. By this means, it is anticipated that it will be possible to more clearly identify the kind of situations for which the automatic location-determination function for balloon-type dialogue is particularly effective.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number 26282061, 16K12382, and 15J00608.

REFERENCES

Blumenthal, E. (2005). *Puppetry: A World history 1st Edition*. Abrams, New York, NY.

Egusa, R., Sakai, T., Tamaki, H., Kusunoki, F., Namatame, M., Mizoguchi, H. and Inagaki, S. (2016). Designing a Collaborative Interaction Experience for a Puppet Show System for Hearing-Impaired Children. In *Computers Helping People with Special Needs, ICCHP2016, Part II, LNCS 9759*, pp.424-432.

Fukamauchi F., Nishioka, T., Matsuda, T., Matsushima, E. and Namatame, M. (2007). Exploratory Eye Movements in Hearing Impaired Students − Utilizing Horizontal S-shaped Figures −. In *National University Corporation Tsukuba University of Technology Techno Report, 14*, pp.177-181.

Hisaki, I., Nanjo, H. and Yoshimi, T. (2010). Evaluation of Speech Balloon Captions for Auditory Information Support in Small Meetings. In *the 20th International Congress on Acoustics*. Sydney, Australia: Australian Acoustical Society, pp.1-5.

Hong, R., Wang, M., Xu, M., Yan, S. and Chua, T-S. (2010). Dynamic captioning: video accessibility enhancement for hearing impairment. In *the 18th ACM international conference on Multimedia*. Firenze, Italy: ACM, pp.421-430.

Hu, Y., Kautz, J., Yu, Y. and Wang, W. (2015). Speaker-Following Video Subtitles. In *ACM Transactions on Multimedia Computing, Communications, Applications,* 11(2), Article 32, 17 pages.

Los Angeles County Museum. And Zweers, J. U., 1959. *History of Puppetry. Los Angeles County Museum*. L.A., CA.

Marschark, M. and Hauser, P. C. (2012). *How Deaf Children Learn.* Oxford University Press, Inc., Oxford, UK.

Rander, A. and Looms, P. O. (2010). The Accessibility of Television News with Live Subtitling on Digital Television. In *the 8th International Interactive Conference on Interactive TV & Video*. Tampere, Finland: ACM, pp.155-160.

Stinson, M. S., Francis, P., Elliot, L. B. and Easton, D. (2014). Real-time Caption Challenge: C-print. In *the 16th international ACM SIGACCESS Conference on Computers & Accessibility*. Rochester, NY, USA: ACM, pp.317-318.

Wald, M. (2008). Captioning Multiple Speakers using Speech Recognition to Assist Disabled People. In *Computers Helping People with Special Needs, ICCHP 2008, LNCS5105,* pp.617-623.