Wearable MC System
a System for Supporting MC Performances
using Wearable Computing Technologies

Tomonari Okada
Grad. School of Engineering, Kobe Univercity
1-1, Rokkodaicho, Nada, Kobe, Hyogo
657-8501, Japan
tomonari.okada919@gmail.com

Tsutomu Terada
Grad. School of Engineering, Kobe Univercity
and PRESTO, Japan Science and Technology
Agency
1-1, Rokkodaicho, Nada, Kobe, Hyogo
657-8501, Japan
tsutomu@eedept.kobe-u.ac.jp

Tetsuya Yamamoto
Grad. School of Engineering, Kobe Univercity
1-1, Rokkodaicho, Nada, Kobe, Hyogo
657-8501, Japan
tetsuya.y@gmail.com

Masahiko Tsukamoto
Grad. School of Engineering, Kobe Univercity
Institute for Clarity in Documentation
1-1, Rokkodaicho, Nada, Kobe, Hyogo
657-8501, Japan
tuka@kobe-u.ac.jp

ABSTRACT
A master of ceremonies (MC) plays an important role to ensure all events progress smoothly because unexpected interruption make them unsuccessful. MCs must have various abilities such as being able to memorize the content of given scenarios and manage problems that occur unexpectedly. Moreover, since unskilled MCs cannot intuit the atmosphere in the audiences during an event, they cannot control this smoothly. Therefore, we propose a wearable system that solves these problems for MCs achieved through wearable computing technologies. Our system has functions to support MCs in carrying out their duties smoothly, such as a robust voice-tracking function for them to read scripts, a user interface that does not interrupt other tasks, and a function that enables MCs intuit grasp the atmosphere of the audience. We implemented a prototype of the wearable MC system and actually used it at several events. The results we obtained from actually using it confirmed that it worked well and helped MCs to carry out their official duties smoothly.

1. INTRODUCTION
MCs (masters of ceremonies) play an important role in various kinds of events such as academic meetings, performances on stage, and wedding ceremonies. They manage events according to time constraints and given elaborate scenarios that are rehearsed in advance. In addition to this, MCs need to have many other skills to manage events smoothly. For example, it is difficult for unskilled MCs to handle sudden requests, and to remember prescribed scenarios on stage. It is also difficult for them to manage various unexpected problems and to respond to requests by the audience. This means that unskilled MCs take too long to prepare for events and cannot manage them while with taking all factors into account.

Therefore, we propose a *wearable MC system* that solves these problems on stage by MCs using wearable computing technologies. This system consists of a wearable system that is used by an MC and a server that is used by an operator. The wearable system is in constant communication with the server to acquire necessary information. By using a Head Mounted Display(HMD) and an input interface in the wearable system, the MC can obtain information from the system without the audience noticing. In addition to this, the proposed system automatically tracks the MC's speech and indicates his current position in the script. Moreover, it has a function to intuit the atmosphere by analyses the sound captured by microphones placed on stage. Utilizing these functions, a skilled/unskilled MC can effortlessly manage the stage. We implemented a prototype of the wearable MC system and actually used it at several events. The results obtained from actual use confirmed that our system worked well and helped the MC to ensure the event ran smoothly.

This paper is organized as follows. We related work in Section 2 and our design and implementation of the new system is introduced in Section 3. The experiments we used to evaluate the system are described in Section 4 and Section 5 describes the actual use of at system in several events. Finally, we present our conclusions and explain work we plan to do in Section 6.

2. RELATED WORK
Various systems that support presentations have been proposed [1][2]. For example, *Presentation Sensei* [3] is a system for training people to make presentations, which offers instant feedback on the speed of speaking, intonation, and eye contact with the audience by analyzing presentations captured with a Web camera and microphone. It has a function to alert presenters when some of these indices exceed predefined thresholds. After the presentation has finished, it generates summaries of analysis. This system is also aimed at being used for practice, and not in actual presentations. In contrast, our system can be used at actual events. Moreover, the content to

support MCs is different from that supporting presenters. The purpose of presenters is to pass on information to the audience clearly while the purpose of MCs is to control the atmosphere on stage.

People speaking to audiences need to control their computers while communicating when they use various support systems. This control should not be noticed by the audience. Travis et al. developed a mobile device Hambone[4] for this purpose. Hambone uses two small piezoelectric sensors placed on either wrist or ankle. When a user moves his hands or feet, the sounds of his bone is generated by movement are transmitted to Hambone via bone conduction. Hambone then sends a packet to a mobile device or a computer. The received data are analyzed using hidden Markov models (HMMs) and are mapped to a set of commands to control applications. Because the motion from these gestures is sufficiency small, Hambone can be available for our system. However, it is too hard to control correctly without any stress on the presenter.

A prompter and a wireless headset are typical devices for supporting MCs on stage. Because both devices can display instructions from the director, they make it easy to change scenarios. However, it is expensive to install are these devices and places where they can be used are restricted. Wireless earphone microphones are hands-free and enable brief communications. However, only limited amounts of information can be given each time and it is inconvenient to repeatedly obtain information. Furthermore, existing devices to support MCs do not assume various settings such as outside stages. The system we propose uses wearable computing technologies to solve these problems. First, an HMD is used as a portable display because no large displays need to be set up on stage and MCs can obtain information in any place from any direction.

3. SYSTEM DESIGN

MCs must remember given scenarios is detail and sequentially present events on stage on time according to the scenario invoked. Some MCs may forget scenarios and therefore cannot manage well on stage. Moreover, MCs need to deal with various kinds of problems flexibly. If they are confused, audiences will soon become aware of their anxiety and then smooth transitions in events are not likely to be achieved. Considering these problems, we established three main requirements for the MC support system.

- **System should display necessary information automatically.**

  An MC needs various information to present events on stage, such as the scenario (acting script), the content of the next presentation, the time left, and the situation in the convening in the convening hall. An MC has to be able to browse such information at all times without any disruptions.

- **MC can communicate operator.**

  The operator's role is to send helpful information to the MC so that he/she can shift smoothly from event to event and deal with problems. An MC using the system needs to communicate through the operator with a user interface that does not interfere with other tasks.

- **MC must control atmosphere on stage.**

  One of the most important things for MCs is to manage the stage smoothly and control the atmosphere of the audience on stage[6]. Reading the atmosphere includes *when to take*
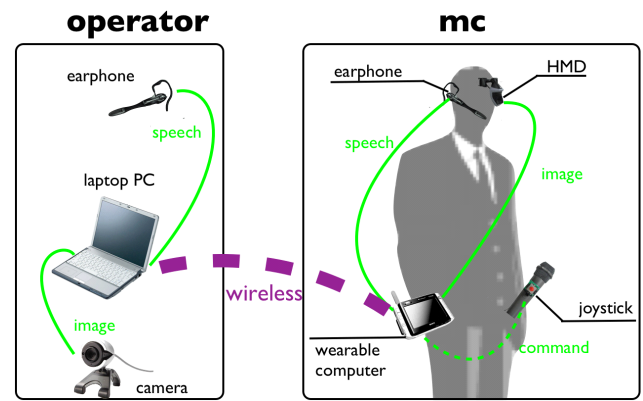

Figure 1: System structure.

pauses while talking and *when to stop applause*. Appropriate momentary pauses represent an important technique where MCs can empathize with the audience. For example, when the stage is buzzing with excitement, MCs often wait for five seconds before speaking[6]. Moreover, applause is typically stopped using three patterns. *Promotion* involves the timing when an MC gives the audience the chance to applaud. *Waiting* involves timing when the MC moves on to the next event when the applause has finished. *Cutting* involves the time when the MC has begun to speak when the volume of applause has decreased and the MC is going to completely stop it. The system needs to give instruction with appropriate timing.

3.1 System structure

Figure 1 shows the structure of our proposed system which consists of a wearable system for the MC and a server for the operator. The MC wears a wearable PC, an earphone, a microphone, an input interface, and an HMD. The operator uses a laptop PC and the server with a microphone and a camera to capture the atmosphere. The wearable system automatically tracks and presents the current position on the script from where the MC is speaking by using voice-recognition techniques we developed. The operator sends information by text or by voice to the MC. Two computers are connected via a wireless LAN and a small wireless joystick is fixed behind the microphone of the MC to enable him/her communicate with the operator without this being noticed by the audience. The microphone connected to the operator's PC captures the sound generated by the audience, and the system recognizes the context of the audience.

The operator gives MCs the instructions to solve problems and to more on to the next event on stage by operating his/her PC. Two computers can exchange data wirelessly.

The microphone on the operator's PC acquires sound on the stage of in front of audiences, and the difference between audiences's sound and applause is recognized by a fast Fourier transform (FFT). However, the recognition rate% on an actual stage is greatly changed due to several factors, including noise such background music (BGM). Then, to improve recognition, the system takes different pictured of the audience and the takes elapsed time into account in assessment. As a result, it meets the third requirement for the MC to control the atmosphere on stage.
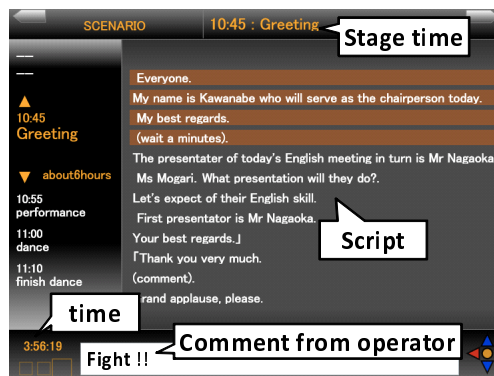
Figure 2: A scenario screen
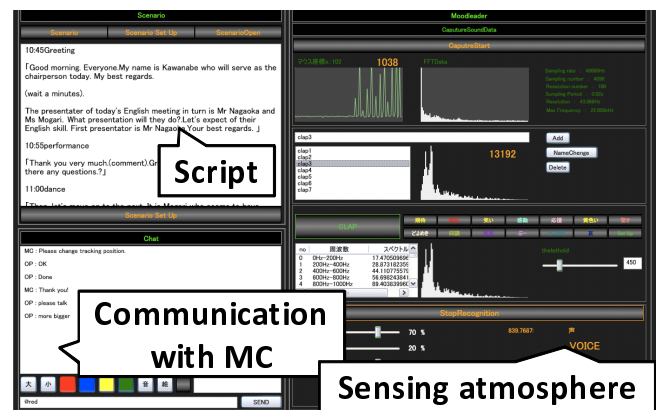

Figure 3: A communication screen

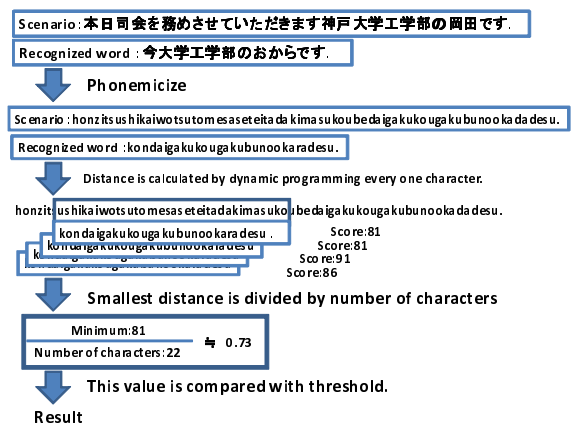**Figure 4: An operator screen**

Figure 5: Our tracking method

3.2 System function

The application is divided into two,i.e., an MC module and an operator module. Figures2 and 3 have examples of screen shots on the MC's HMD. Figure 2 shows a scenario screen of the script to be spoken at that time, the remaining time, comments by the operator and the content given as ad-lib. The background color of texts is changed by the system. Figure 3 shows a communication screen that is used for communicating with the operator. It shows the help buttons have registered the commands in advance. The MC can ask the operator to assist by using the help buttons. Figure 4 shows an operator screen. The operator mainly sends instructions on this screen. He/she can change the size and color of characters by commands which start from @. For example, *@big @red* means the size of text is enlarged and the color of the text is changed to red. *@beep* makes a sound on the MC's computer. Further, he/she can change the scenario, the tracking position of words, and the image on this screen. He/she can provide a good suggestions since he/she knows the atmosphere of the audience.

3.2.1 Tracking of scenario

As previously discussed, the MC sometimes loses the place from where he/she is to read on the HMD screen. This often occurs when he/she takes his/her eyes off the HMD. As we often used an HMD with a resolution of 800 × 600 pixels because commercial products with this resolution were available space was more limited than that with a display for desktop PCs. Therefore, the operator had to frequently scroll the lines for the scenario. Then, the back-

ground color of the lines to be read were changed and scrolled with voice recognition. It was difficult to recognize a voice spoken on stage with a high degree of accuracy. Voice recognition is often used to operate robots, where a user can utter commands again if recognition fails. However, an MC on stage cannot repeat himself and therefore higher recognition is required. In addition, MCs do not need to speak according to the script and sometimes ad-lib. The system must recognize which lines the MC is speaking and whether he/she is conforming to the scenario. The recognition technique is shown in Figure 5.

First, recognized words are phonemicized, and they are compared with the sentences in the script using Dynamic Programing. Japanese is composed the Chinese and hiragana characters. Hiragana is a Japanese syllabary, in which each character represents one more, and it has a total of 50 syllables as shown in Figure 6. Both sentences in the script and recognized words are translated into hiragana, and they are compared by using Dynamic Programing,whichis a technique that compares two character strings, and measures the distance between them. In our system, a hiragana sentence with recognized words and a sentence in the script are calculated while doing the slide by one syllable, and the smallest distance in the result is extracted. We set the window size to the number of syllables in the recognized words. The recognized words are determined

consonant / vowel	a	k	s	t	n	h	m	y	r	w
	あ	か	さ	た	な	は	ま	や	ら	わ
a	a	ka	sa	ta	na	ha	ma	ya	ra	wa
	い	き	し	ち	に	ひ	み		り	
i	i	ki	si	ti	ni	ha	mi		ri	
	う	く	す	つ	ぬ	ふ	む	ゆ	る	を
u	u	ku	su	tsu	nu	hu	mu	yu	ru	wo
	え	け	せ	て	ね	へ	め		れ	
e	e	ke	se	te	ne	he	me		re	
	お	こ	そ	と	の	ほ	も	よ	ろ	ん
o	o	ko	so	to	no	ho	mo	yo	ro	nn

Figure 6: Japanese syllabary

Figure 7: A set up screen

Figure 8: MC's interface

Table 1: Commands operated with joystick

operation	command
lean up	scroll up
lean down	scroll down
lean left	switch to the communication screen
lean right	switch to the script screen
click	send enter key
double click	send emergency to the operator
Press and hold	switch to mouse mode

3.2.2 MC's interface

The MC operates the system with a small joystick behind the microphone as shown in Figure 8. Therefore, audiences usually do not notice the joystick being operated. We can click and input directions with this joystick that is connected to wearable PC via wireless communication. As precise pointing like that with a mouse is difficult for the MC, several commands are assigned to directions this joystick can manipulate as shown in Table 1. The operator can customize the content of commands and the number of commands.

3.2.3 Recognizing atmosphere in audience

The system analyzes what the applause means and suggests intervals for appropriate pauses to the MC according to the flow in Figure 9. First, the microphone on the operator's PC selects sound on stage. The concrete procedure for this recognition is given in Figure 10. The frequency spectrum is obtained from the FFT function and applause and voice are distinguished according to the spectrum distribution. Next, the Euclidean distance between the stored data and the sound on stage is calculated. Since the number of people and the size of the stage changes from event to event, the data from the applause are captured by the operator at each event. Since there is often BGM on stage, it is difficult to recognize this only from the frequency spectrum distribution. We therefore additionally use image and scenario information at higher levels in the recognition process. If the differences in images taken within a short periods of time by openCV[8] are larger than a predetermined threshold, the system determines that it is applause, and not a regular voice. Further, the threshold for applause is set for each location in the scenario. When applause does not finish within a certain time interval, the MC has to stop it with appropriate timing. Applause is typically stopped with three patterns. The system gives instructions according to appropriate timing. It also recognizes the positive or negative reactions by the audience. The results are obtained by calculating the distance between the sample data for positive and negative feedback at high and low frequencies. When the audience does not quiet on down after the MC has spoken the five-second rule, i.e., waiting for five seconds, is useful for getting there attention. This timing is also suggested by the system. Furthermore, the instructions are de-
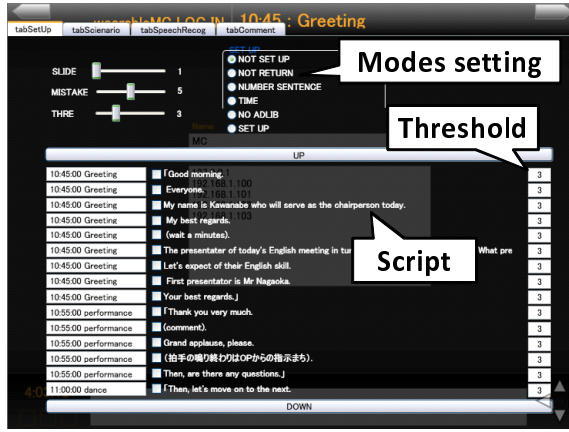
to be part of the scripts if the extracted distance is lower than the threshold. Otherwise, it is judged as ad-lib and the system stores the recognized texts with time stamps. Stored ad-lib can be seen later to evaluate the meeting and improve of the scripts. The operator can change the tracking position manually when recognition fails and the tracking lines are moved to a very different place.

Moreover, the scenario on stage includes a structured salutation and many ad-lib lines when guests are introduced. Therefore, we set the threshold for ad-lib lines in all places in the scripts to improve recognition accuracy. Figure 7 shows a screen of settings. Six elements can be set.

- The dialog does not return part of the script.
- Limitations with sentences
- Limitations with time
- Only the script is read.
- The set up for the threshold level
- Nothing is done.

The recognition rate is improved by appropriately setting these parameters.

Figure 9: Flow for recognizing audience atmosphere

$$d = \sqrt{\left(x_{g1} - x_{r1}\right)^2 + \left(x_{g2} - x_{r2}\right)^2 + \ldots + \left(x_{g100} - x_{r100}\right)^2}$$

d : Distance
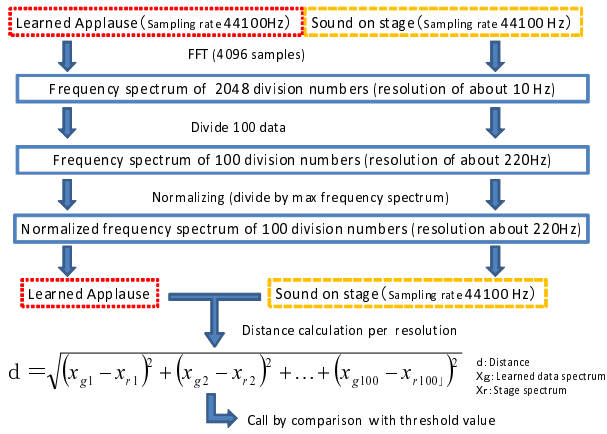Xg : Learned data spectrum
Xr : Stage spectrum

Figure 10: Technique for recognizing atmosphere for sensing

termined by the frequency spectrum, volume, image information, time, and scenario information from the applause and voices.

4. EVALUATION

4.1 Evaluation of scenario tracking

We evaluated the recognition rate for tracking a voice that read a given scenario. This evaluation was done in a performance on stage that we held at Luminarie, where a chairman used our proposed system. A scenario and some ad libs were used in this experiment just like on a real stage. The threshold was determined from a preliminary experiment. Table 2 summarizes the results obtained from the evaluation. Sixty-seven sentences were recognized on a real stage and 22 sentences were recognized as ad-libs. Note that the ad-libs could be recognized with 100% accuracy by using this tracking technique. No incorrect tracking was done even if the chairman spoke about a different scenario. The recognition results for the scenario were 84%. Ten words were recognized as ad-libs and one item was recognized as wrong lines. Tracking was not done once every five times. However, this is not a problem in practical use because tracking can return to the correct position by using recognized sentences. However, the operator should change

Table 2: Judgment result

Kind	Correct answer	Failure	Recognition Percent
Scenario	56	11	84%
Ad-lib	22	0	100%

Table 3: Recognition Percent where a loud BGM exists

Average vol	Maximum vol	Recognition Percent
-54.51dB	-25.08dB	75%
-43.43dB	-25.19dB	90%
-32.12dB	-15.39dB	62.5%
-26.86dB	-6.27dB	0%

to manual operation when the results of recognition differ markedly from the right lines.

We also assumed that the recognition rate would fall when BGM and audience noise were mixed on an actual stage. We carried out experiments in four kinds of sound environments. Table 3 lists the recognition rates. In the fourth result, which was the loudest environment, no one could recognize voice. However, this was not a problem because it was difficult for the MC to confirm his own voice, which we assumed was an unusual situation for the MC. Since there were few failures in other environments, this is not a serious problem in the practical sense. In addition, when the MC began to speak on stage, the BGM volume was adjusted to a low level and sound from the audience was also low. Therefore, the amount of marginal sound did not badly affect voice-recognition.

Table 4 summarizes the recognition rate(%) when we set up various conditions. The scenario we used in this evaluation was for Our Laboratory Quinquennial Convention. The results indicate that appropriate set tings improve the recognition rate.

4.2 Evaluation of sensing atmosphere in audience

We evaluated the recognition rate for applause. The number of samples for FFT, its resolution, and the threshold to recognize as to applause were determined from a preliminary experiment. The recognition rate(%) whether the recognized results were applause was evaluated by using the data from the voice file of an actual event on stage(sampling rate of 44100 Hz for about 1 hour). Table 5 lists the results from the evaluation. The recognition results for applause were 76% and 98% for voice. Also, the second step in Table 5 indicates the recognition rate (%) where a loud BGM exists(the date length was about 5 minutes). The recognition result for the applause were 66% and 89% for voice. The recognition rate(%) for applause was low because it was difficult to detect the start and end of applause. However, the recognition rate(%) for applause is sufficient if the system is only used among certain volumes, the instructions of waiting and cutting after it has recognized applause. These results indicate that if applause on stage is detected once, how to applause and voice can almost be recognized, and the system can instruct the MC appropriately treat applause.

Moreover, the third step in Table 5 shows the recognition rate(%) when we set the threshold so that the event on stage was scheduled to end according to the time in the scenario information. The threshold for applause was set to 450. Therefore, the threshold for the last 30 seconds on stage was lowered. This indicates that an

Table 4: Recognition Percent when we set up some modes

Situation	Recognition Percent
not set up	76.47%
not return	76.47%
limit of stage	85.29%
time	85.29%
no ad-lib	85.29%
set up threshold level	91.17%

Table 5: Recognition Percent as to whether result is applause

Result is applause	data number	Correct	Percent
all	48714	47289	97%
applause	231	176	76%
voice	48483	47113	98%
applause in a loud BGM	data number	Correct	Percent
all	3830	3383	88%
applause	118	79	66%
voice	3712	3304	89%
set up by scenario information	data number	Correct	Percent
all	3830	3709	97%
applause	118	87	73%
voice	3712	3622	98%

elaborate set up can improve the recognition rate.

5. ACTUAL USE IN EVENTS
5.1 A Luminarie stage

The system was used on stage at an event at the Kobe Luminarie that was held on 13th and 14th December, 2008 to test and verify how effective our proposed system was. The Kobe Luminarie has been held annually since December 1995 to commemorate the victims of the Hanshin-Awaji earthquake and has been a symbol of reconstruction. Figure 11 shows the appearance in which the system was used. The chairman introduced the proceedings and introduced some performers who were in front of the audience. The system that was used on stage was a prototype of the one we propose . As there was no voice tracking system, the chairman had to scroll on the lines manually. The scenario and the instructions from the operator were displayed on his HMD. The chairman prepared the whole system including putting on the HMD and booting PC in five minutes. He could introduce the events smoothly on stage because he saw the scenario at all times wherever he went despite his nervousness on stage. However, he lost his position in the lines when he took eyes off the HMD and he did not see the operator's new instructions. Therefore, we added voice-tracking and alert systems to assist the chairman.

There was another problem in that a ten key which was the main input for the system, was not able to be operated by touching a button on the small computer that was installed around his waist. As such problems generate extra tension, we have to consider the use of microcomputers without OSs on PCs and need to prepare a paper version of the scenario.

5.2 Repetitive use of system

The first author chaired a seminar at our laboratory to test and improve the system. All speakers at this seminar introduced research project in front of an audience, and after they had made their the presentations, the participants discussed the project. There were six presentations and spoke each time four speakers. The audience

Figure 11: A snapshot of an actual use of the system at Luminarie stage

consisted of about 20 people in a conference room at Kobe university. The chairman made six main comments.

- Even when the scenario was changed, I did not need to change anything. I could easily prepare myself with this system.
- When there were few comments from the audience, I could consult with the operator. He was made several good suggestions.
- It was possible to immediately deal with problems like sudden changes of in the scenario.
- I never lost the sight of the place I was at in the lines.
- I could find the scenario even when I looked at PowerPoint slides the presenter projected and listen to him.
- I could easily check the time.

However, there were three main problems.

- HMD was not useful because there was no need for me to move in the conference room. A normal PC would have been good enough.
- The operator could not speak in the quiet room.
- The input of text when taking notes was impossible on the present interface.

We tested and it was confirmed that the MC support system using wearable computing technologies was effective in a laboratory environment. However, when the Chairman's position was fixed, the HMD offered little. It will be necessary to introduce the same system to other staff to provide more dynamic support.

5.3 Use by beginners of system

At a conference on the 5th anniversary of our laboratory on 30th October, 2009, an inexperienced MC used this system so that it could be tested. Figure12 shows the appearance of the system in

Figure 12: A snapshot of an actual use by beginners of system

use. The MC was not well trained in the use of this system. The first scenario was a five-minute greeting and the MC read out text displayed on his HMD. He did not memorize any lines. He used the system for four hours at this ceremony. As a result, he was able to progress smoothly from one event to the next even though he was not well prepared. He made four observations;

- I did not need to remember formal greetings.

- The setup time was short and the system could be used while I was moving freely.

- The timing for scrolling was difficult and it was necessary to get used to it.

- I was able act calmly as an MC from event to event.

We also asked the audience for their opinions. Almost of all were of the opinion that the MC had smoothly shifted from event to event. We confirmed that the system can support MCs without any need for special skills and the wearable computing technologies used in it were effective. More improvements are needed to remedy the timing for scrolling and the inconvenience of information retrieval.

6. CONCLUSIONS

We proposed a wearable MC system that MCs with little experience could use to carry out their responsibilities on stage smoothly. It makes it possible for MCs to communicate with an operator through, a hidden interface. They can confirm many kinds of information on the screen of an HMD. It is easy for MCs to locate the reading position for their lines easily even if they take their eyes off the display due to the use of a voice tracking system. The system informs MCs of the timing of momentary pause and suggests appropriate timing to them. It will be necessary in future work to develop a flexible input interface and a dependable system to deal with sudden problems with PCs. There is also a problem in ensuring a stable power supply for extended periods. It is possible to not only apply the system to support MCs but also to supporting presentations, to supporting scenarios on stage, to supporting displays of singer lyrics and to supporting various performances on stage.

7. REFERENCES

[1] S.Elrod, R.Bruce and R.Gold: Liveboard: a large interactive display supporting group meetings, presentations, and remote collaboration, Proc. of the SIGCHI conference on Human factors in computing systems (SIGCHI1992), pp.599-677, (1992).

[2] H.U.Hoppe, W.Luther, M.Muhlenbrock, W.Otten and F.Tewissen: Interactive Presentation Support for an Electronic Lecture Hall, Proc. of the Advanced Research in Computers and Communications in Education, pp923-930, (1999).

[3] K.Kurihara, M.Goto, J. Ogata, Y.Matsusaka, and T.Igarashi: Presentation Sensei: A Presentation Training System using Speech and Image Processing, Proc. of the 9th International Conference on Multimodal Interfaces (ICMI2007), pp.358-365, (2007).

[4] T.Travis, S.Palinko, E.S.Poole, and T.Starner: Hambone: A Bio-Acoustic Gesture Interface, Proc. of the 11th IEEE International Symposium on Wearable Computers (ISWC2007), pp.1-8 (2007).

[5] J.Ikeda, Y.Takegawa, Y.Terada, and M.Tsukamoto: Evaluation on Performer Support Methods for Interactive Performances Using Projector, Proc. of the 6th International Conference on Advances in Mobile Computing and Multimedia (MoMM2009) pp.205–211 (2009).

[6] Raita Goto: Shikai Kanji Dandori No Shikata, Takahashi Syoten, 2008.(in Japanese)

[7] Julius, http://julius.sourceforge.jp/.

[8] openCV, http://opencv.jp/.

[9] Kobe Luminarie, http:/www.kobe-luminarie.jp/.