A Context-aware Audio Presentation Method in Wearable Computing

Shinichi Yataka
Grad. School of Engineering, Kobe Univercity
1-1, Rokkodaicho, Nada, Kobe, Hyogo
657-8501, Japan
s.yataka@stu.kobe-u.ac.jp

Kohei Tanaka
Advanced Technology R&D Center, Mitsubishi Electric Corporation
8-1-1, Tsukaguchi Honmachi, Amagasaki, Hyogo 661-8661, Japan
Tanaka.Kohei@ah.MitsubishiElectric.co.jp

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@teralab.info

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

ABSTRACT

Audio usage is one of the more widely-applicable methods of information presentation in wearable computing environments since it can be used hands-free, requires only small devices like earphones, and does not interfere with most tasks compared with other methods including visual displays. However, since presented sound is often drowned out by ambient noise or conversational voice, a user is forced to turn the volume up to catch the audio information. Therefore, we propose an audio information presentation method that takes into account the user contexts. In our proposed method, a system estimates the user contexts based on the data from wearable sensors and a microphone, and then controls and presents the audio information so that it can be surely audible by changing the volume and the timing for presentation. The evaluation results confirmed the effectiveness of our proposed method.

1. INTRODUCTION

Wearable computing has recently attracted a great deal of attention due to the technological advancement of computer miniaturization. Since users in wearable computing environments can browse information in various situations by using different information presenting devices, the existing methods for information presentation for desktop computing environments are not always sufficient enough for users to fully grasp the information. For example, when a user wearing a head mounted display (HMD) is paying attention to a real-world task, he/she sometimes misses the information displayed on the HMD. Moreover, it is dangerous for a user to continuously look at visual information on a HMD when he/she is walking or driving.

Here, we focus on the audio presentation that is used as a widely applicable interface in current wearable computing environments because it does not require anything but an inexpensive device and allows users a passive reception of information. Users can acquire information using audio presentation without needing to completely concentrate on the information. However, a problem still remains in that the recognition accuracy of the audio information depends on the user contexts. For example, a user in a loud environment cannot correctly catch the information presented at a low volume. Furthermore, a user who is concentrating on another task may miss unexpectedly presented audio information in at a low volume or at an inappropriate time. On the contrary, audio presented at a high volume may annoy the user because it may interfere with a conversation or his/her concentration. To solve this problem, we propose an audio presentation method that takes into account the user contexts in order for the user to be able to correctly and comfortably catch the presented information. In this research, we investigated the relationship between the recognition ratio of the presented audio information and the user contexts, and propose a mechanism for controlling the volume and timing of the audio presentation.

The rest of this paper is organized as follows. In Section 2 we describe the related work and Section 3 describes our environmental assumptions. We describe our proposed method in Section 4 and Section 5 presents the results of our evaluation experiments. Finally, we summarize the paper in Section 6.

SAC'11 March 21-25, 2011, TaiChung, Taiwan.

2. RELATED WORK

A lot of information presentation methods have been proposed in order to assist a user in catching the information in wearable computing environments. Tanaka proposed an information layout method for an optical see-through HMD that took into consideration the viewability based on the sight behind the display [11]. The layout for displaying information is dynamically changed to ensure better recognition of the information. Although this method resolves the problem of the difficulty in seeing information on a HMD, there is still the problem of missing the information because the user does not always look at the HMD.

Another research area related to our scope is the researches on information notification methods. There are several researches on how to notify mobile users about mail reception or schedule alerts. Kern and Shiele proposed a method that maps the user contexts obtained from body-worn sensors to the interruptability that is explained on two axes of the personal and social interruptability in order to decide whether to notify the user [3]. Nomadic Radio changed the amount of information in a notification, from simple alerts to full text reading, by using the priority of the information, the history of the interaction between the user and the system, and whether the user is talking [10]. In these studies, the amount of presented information is controlled according to the user contexts. Our method presents the information without reducing the amount of information and the user comfortably receives the information. Additionally, our method can improve these methods by controlling the volume and the timing of the presented information.

Just as in our research, there are many systems and applications using audio information in wearable computing environments. SoundWeb [5] is a system for browsing voice information structurized by using a hypertext-style link in wearable computing environments. It plays a signal sound along with the part with a link to another set of voice information and enables a user to know of the existence of the link and to catch the structured information by voice. SWAN [14] presents various kinds of information, such as the direction of travel, the existence of nearby bus stops, and convenience stores, by using non-language sound. Aritsuka et al. proposed a system that outputs spatial sound according to the GUI layout for visually handicapped people [1]. NewsComm [9] automatically retrieves news contents on the network and delivers them according to the user preferences. However, since the user contexts are not taken into consideration in these audio presentation applications, the user may not be able to correctly and comfortably acquire the information. Our proposed method aims to solve this problem, and it can be integrated into these researches to make them more effective.

3. ENVIRONMENTAL ASSUMPTIONS

We put some assumptions on wearable computing environments in this research. In general, wearable computing has several features, i.e., it can be used hands-free, and it is always on. By using such features, there are several promising application areas such as in health-care [8] and navigation systems [2]. In such applications, various information including any abnormalities in the health conditions and the instructions from a remote manager are often given by audio because a user can passively catch the information passively. Therefore, we focus on audio information presentation in this paper.

We use a bone-conduction earphone as our audio information presentation device. This is because standard earphones and headphones that obscure the user's ears are dangerous in wearable computing environments since it makes it difficult for him/her to hear ambient sound. Bone-conduction earphones are installed without obscuring the ear, and the ambient sound can be caught at the same time as the presented information.

In addition, we assume that a user wears several sensors and a microphone to acquire the user contexts. These sensors are commonly used in wearable computing environments because they are small, do not obstruct his/her daily life, and are assumed to be used in various applications such as navigation and health-care services. Figure 1 shows a example of a typical user we assumed in this paper.

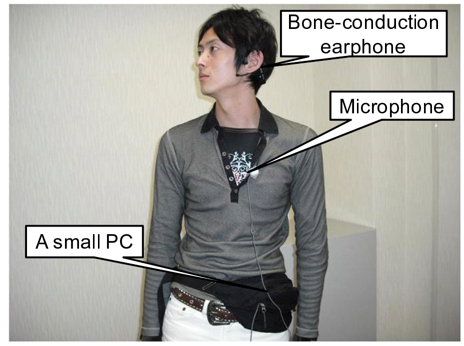

Figure 1: A snapshot of a typical user assumed in this paper

4. PROPOSED METHOD

Our method presents audio information taking into account the user contexts in wearable computing environments, which is aiming at enabling the user to correctly and comfortably catch the information. In particular, when the system outputs audio information such as a news flash or a voice guide in a navigation service, our method firstly recognizes the user context, and then it adjusts the volume and the timing of the audio presentation for an easy and comfortable acquisition of the information. In addition, the way of presentation is controlled according to the required accuracy for recognizing the information. For instance, when an instruction manual is read out in a work support system, a mishearing of information may lead to dangerous situations while the system should control the audio not to interfere with the user's work.

In this section, we firstly describe our preliminary evaluation to clarify the relationship between the accuracy in recognizing the presented audio information and typical user contexts. Next, we propose a method for controlling the way of presentation based on the above evaluation. Lastly, we show the system design of our prototype.

4.1 Preliminary Evaluation

When the volume of the presented audio information is too low, a user may not be able to catch the information. On the other hand, a user may feel annoyed if the presentation volume is needlessly large (The influence on this factor is evaluated in Section 5). Therefore, to find an appropriate volume level with respect to a given context, we conducted a preliminary evaluation concerning the relationship between the recognition rate of the presented audio information and the typical user contexts.

4.1.1 Hypothesis

We put hypotheses on the factors in the user contexts that affect the recognition accuracy of audio information in this research. We classified the user contexts into two categories: One of them is an *activity context*, which is based on his/her own context, such as *walking* and *having a conversation*, and the other is an *environmental context*, which is the context concerning the user's surroundings, such as *a loud* or *crowded environment*.

Here, we make the hypothesis that a physical factor such as *the amount of body movement* or a mental factor such as *the concentration on other tasks* affects the recognition rate of the audio information in the *activity contexts*. In addition, the activity context *conversation* should be separately considered because it is generally difficult to hear a story from two or more people at the same time. As an important factor of for the *environmental context*, we chose *the volume of ambient sound*.

4.1.2 Evaluation Method

The test subjects were five college students aged from 22 to 23. Each test subject wears a small PC around their waist, a microphone on their neck, and a bone-conduction set of earphones. The system presents them with audio information where a man sequentially reads out four alphabets (400∼450 [Hz]) with an interval based on the Poisson arrival of a one minute average 3 ∼ 5 times in each situation. The test subjects mark the heard characters. The BGM (with Japanese lyrics) is played as the ambient sound from the speaker. Based on the hypotheses described in Section 4.1.1, we investigated five activity contexts that a lot of people perform every day; *no task*, *walking*, *jogging*, *using a desktop PC*, and *having a conversation*, as shown in Figure 2. We chose *walking* and *jogging* for investigation of how the amount of the body movement affects the recognition rate, *using a desktop PC* for the concentration on another task, and *having a conversation* for the effect of conversation. Note that *using a desktop PC* means a task that requires a small amount of concentration, such as creation of a document or web browsing, and *having a conversation* means a situation where three or four college students discuss a certain theme and a person controls the timing for presenting the audio information to the test subjects while he/she is having a conversation.

In this evaluation, we used a Sony VAIO UX90S (OS: Windows XP, CPU: 1.20 GHz, Memory: 1 GB) to control the audio presentation, TEMCO HG40SAN-TBT bone-conduction earphones, and Panasonic stereo microphone RP-VC201. The presentation volume was -5 dB, -15 dB, -25 dB, and -35 dB compared with the maximum volume.

4.1.3 Evaluation Result

Figure 3 shows the accuracy in changing the volume of the presented audio in the *no task* context. In the figure, the average accuracy was 97% at a -5 dB presentation volume. In this situation, most of the errors happened when trying to distinguish between two alphabets in which their vowels are similar such as b[bi:] and d[di:], which is difficult for Japanese to distinguish. We can see that the recognition accuracy decreased according to the increase in volume of the ambient sound, except for in the case of a -5 dB presentation

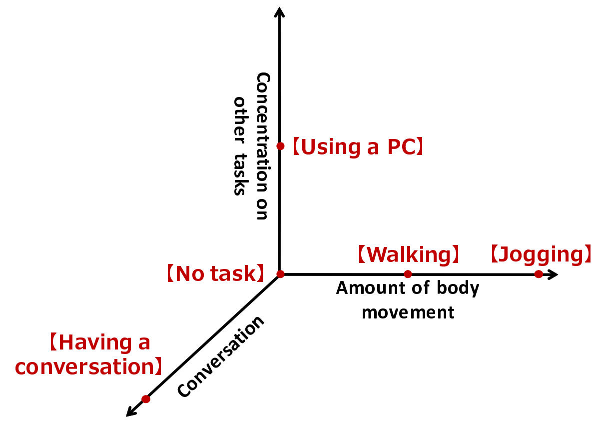

Figure 2: User contexts for the evaluation

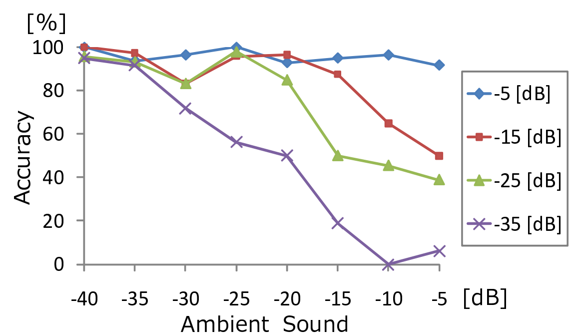

Figure 3: Accuracy of *no task*

volume.

Next, Figure 4 shows the differences in accuracy by the activity contexts. In all the activity contexts except for *having a conversation*, there is a tendency for the accuracy to fall as the volume of ambient sound increase or the presentation volume is reduced while there is little difference among the contexts. Therefore, we can see that the differences in *the amount of the body movement* and *the concentration on other tasks* do not have much influence on the recognition of the audio information. On the other hand, the accuracy in *having a conversation* moves up and down regardless of the volume of the ambient sound. We believe that this is because the test subjects concentrated on speaking and listening to a story told by others while the degree of concentration frequently changes during the conversation.

4.1.4 Decision of Presentation Volume

We clarify the relationship between the volume of the ambient sound and the presentation volume necessary to recognize the audio information in each activity context using the results of a preliminary evaluation. Figures 5 and 6 show the points where the accuracy exceeds the threshold T for the contexts of *no task* and *having a conversation*. In the figures, the presentation volume necessary for recognition increases as the volume of the ambient sound increases in the *no task* context. In this case, we can draw a approximated

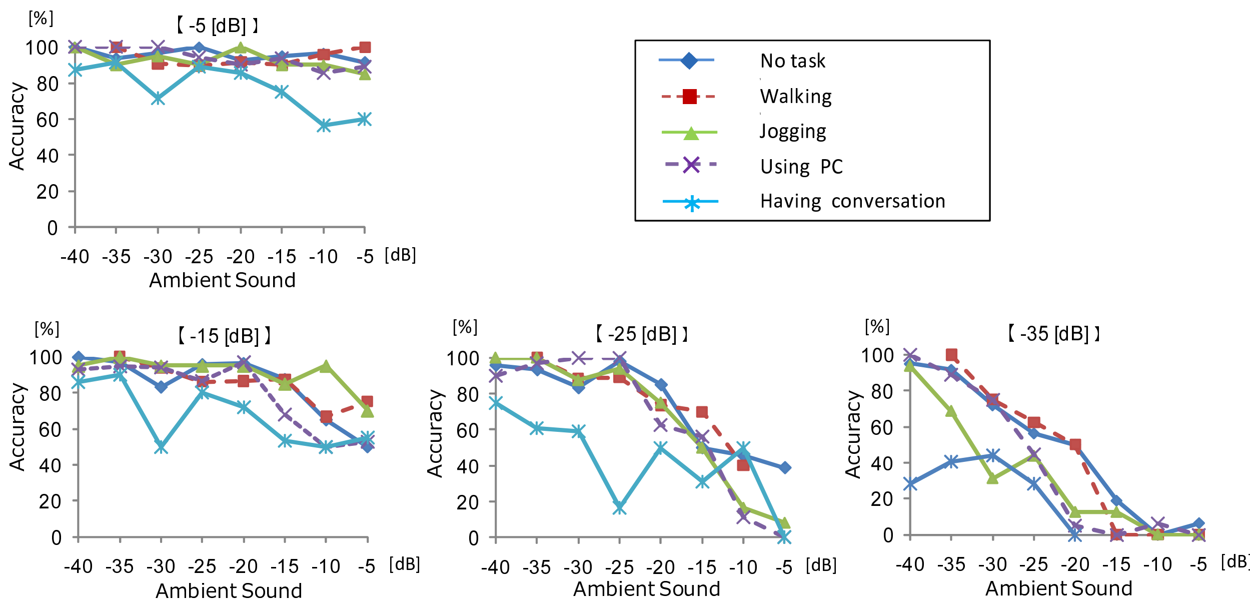

Figure 4: Accuracy of each presentation volume

line that represents the relationship on the decibel scale between the volume of the ambient sound and the minimum requirements on the presentation volume. In the other activity contexts except for *having a conversation*, a similar tendency is seen. On the other hand, there is no correlation in the context of *having a conversation*.

Here, $T = 94\%$ implies that the test subjects surely catches the presented information, which is based on the average accuracy (97%) with the confidence interval (3%) when the test subjects in *no task* heard the information at -5 dB under the conditions that the volume of the ambient sound was -30 dB or less. $T = 85\%$ implies that he/she catches the rough meaning, which is decided as the accuracy where the test subjects make a mistake at a 10% probability.

However, when the volume of the ambient sound is too large, the required presentation volume becomes impossibly high. Moreover, a proportional relationship is not seen for the *conversation* context and it is assumed that an insufficient amount of accuracy is obtained even if the presentation volume is changed. In these cases, it is necessary to use other methods such as presenting an alert before presentation to get the user's attention or pausing the presentation until it becomes possible for the user to comfortably hear the presented contents.

4.2 Audio Information Presentation Method

Our proposed method achieves an audio information presentation that takes the user contexts into account using the following steps:

1. Detection of the user contexts.
2. Decision on the best way of presentation.
3. Presentation of audio information.

The system executes the above-mentioned steps n seconds before presenting the information ($n = 2$ in our prototype). Figure 7 shows the operational chart of our proposed

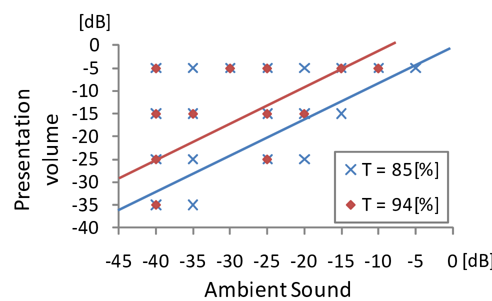

Figure 5: Point diagram in *no task*

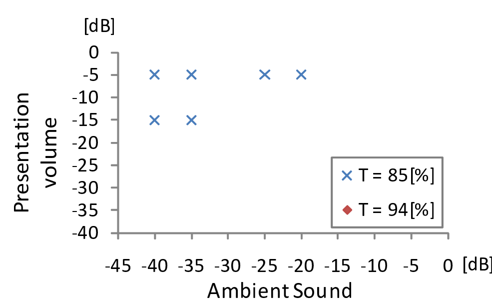

Figure 6: Point diagram in *having a conversation*

method. Here, it is assumed that the necessary recognition accuracy is decided by a user beforehand.

Detection of User Contexts

The system measures the volume of the ambient sound using a microphone and acquires the user's activity using a microphone and 3-axis acceleration sensors. The methods

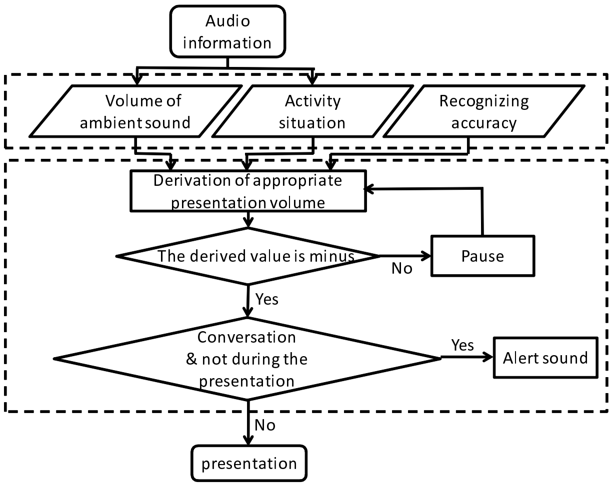

Figure 7: A flow of proposing method

of recognizing the activity contexts are widely proposed and our proposed method assumes that *no task*, *walking*, *jogging*, and *using a desktop PC* are recognized using the method in [4][6] and *having a conversation* is recognized using the method in [10].

Deciding the Best Way of Presentation

The way of presentation is decided based on the results from the preliminary evaluation. In particular, the presentation volume is obtained using the approximation line obtained in Section 4.1.4. If the derived volume is too large and the presentation is impossible, the system presents it after it becomes possible. If the user has a conversation, it sounds an alert before the contents are presented to get the user's attention, and presents it at a volume derived using the proportional relationship for a *no task*.

4.3 Implementation

We implemented a prototype system that presents audio information in consideration of the user contexts. We used Microsoft Visual C# 2005 on Windows XP to develop the prototype and DirectSound to control the audio information presentation. Figure 8 shows the system structure of our prototype. The system recognizes the activity contexts using a microphone and 3-axis acceleration sensors. In particular, it recognizes the *no task*, *walking*, *jogging*, and *using a desktop PC* using three 3-axis acceleration sensors installed on the right arm, waist, and right leg by SVM (Support Vector Machine) [13] using the mean and variance of the sensing data for 0.5 seconds as the feature values to learn and recognize. *Having a conversation* is recognized using a microphone installed on the neck. The system checks whether the captured sound includes language using the open-source large vocabulary CSR (Continuous Speech Recognition) engine Julius [7] to detect *having a conversation*.

In addition, we implemented two applications that use the proposed prototype, one which reads out the updated contents of the timeline in Twitter [12] and one which delivers gourmet information, such as nearby restaurants. They also decide the way of presenting using our method.

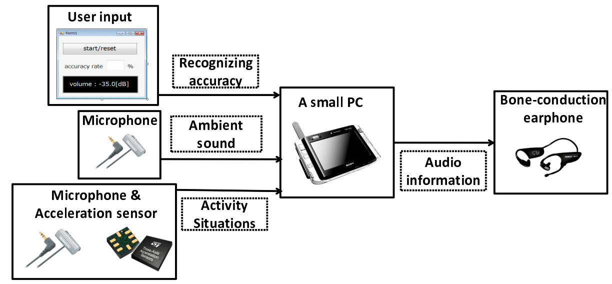

Figure 8: System structure

5. EVALUATION

We conducted four experiments to evaluate our method. We investigated the accuracy in recognition of the presented contents and the subjective evaluation, such as the appropriateness and annoyance, in order to show the effectiveness of three methods: the dynamic change of volume, alert before presentation, and pausing on improper timing. In addition, we experimented using the Twitter voice client implemented to show the usefulness of presenting audio information, such as a news flash and navigation messages, using our method.

5.1 Evaluation on Dynamic Change in Volume

We set up an experiment using a prototype system to evaluate the effectiveness of the dynamic change in the presentation volume. The experiment was done in the same manner as that for the preliminary evaluation. Furthermore, they were requested to subjectively evaluate the appropriateness of the presentation volume by using a scale from 1 to 5 (1: too small – 3: appropriate – 5: too large). We compared our method with the methods that present the contents at constant volumes of -5 dB and -35 dB for *no task*, *walking*, and *operating a PC*. The threshold T was 85%.

Figure 9 shows the results of the accuracy and the subjective evaluation for each activity context. The accuracy in our method exceeded the threshold T (85%) regardless of the activity context or the volume of the ambient sound. On the other hand, when comparing the proposed method with the case where the presenting volume was fixed to -5 dB, there was no significant difference for any condition. This indicates that a high enough accuracy can be obtained by using the proposed method that does not usually need a large presenting volume.

As for the subjective evaluation, the score in the case of -35 dB was close to 3 under quiet conditions and approached 1 as the volume of the ambient sound increased. In the case of -5 dB, we can see that the evaluation score greatly exceeds 3 under quiet conditions. This indicates that the test subjects felt uncomfortable with the needlessly large volume under quiet conditions while the score of our method was close to 3 regardless of the volume of the ambient sound. Therefore, we can say that people receive the presented information accurately and comfortably when using our method.

In addition, we conducted two questionary investigations with the test subjects after the experiment; "What do you think of the audio information presentation in wearable computing environments?" and "Did you feel uncomfortable with the change in volume?" The answers to the first question were: "Audio is useful as a secondary use in situations

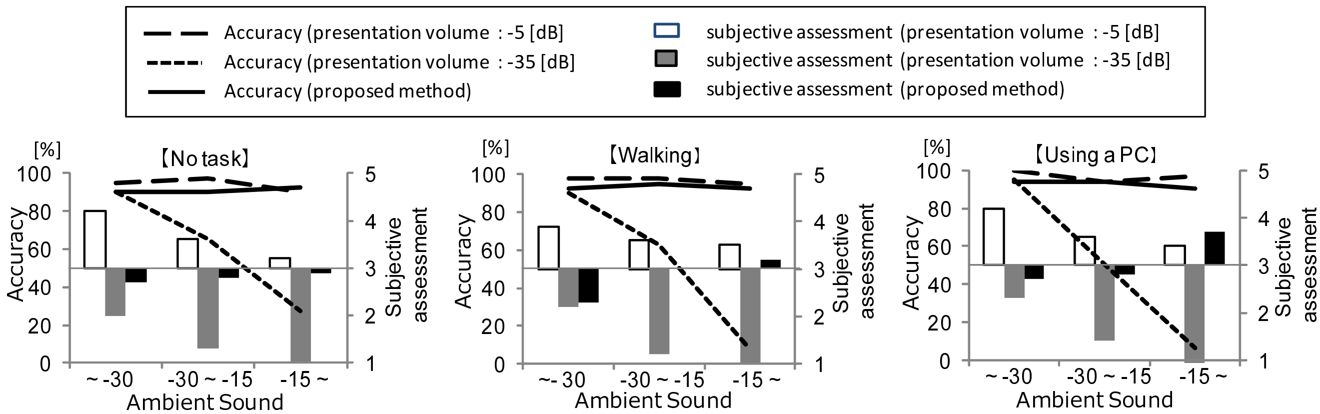

Figure 9: Evaluation results of dynamic change in presentation volume

where I cannot use display devices," "It was annoyed that the contents were read out whenever email is received," and "I want to use the proposed system as an email checker. I think it is convenient if the sender and subject fields are read." In general, presenting a large amount of information seems to enlarge the user's stress. However, an effective information presentation can be achieved using the audio presentation together with other information presentation methods, such as visual presentation. The answers to the second question were: "I was pleased with the change because the ambient sound affects my hearing," "It would be useful if we could choose it to be automatic or manual," and "I felt uncomfortable when suddenly presented with information at a larger volume than expected." They indicate that it is necessary to appropriately change the volume.

5.2 Evaluation of Alert Sound

We evaluate the effectiveness of an alert sound before the contents are presented. We conducted the experiment in the same way as our preliminary evaluation and compared it with the presentation without it. The test subjects subjectively evaluated the annoyance of the presentation method by using five stages (1: good - 5: bad). Here, the presentation volume was changed dynamically using our proposed method. The activity contexts experimented on were *no task*, *walking*, *jogging*, *using a desktop PC*, and *having a conversation*. The alert sound is a signal sound at -5 dB.

Figure 10 shows the results of the accuracy and the subjective evaluation for each activity context. The accuracy reaches a high value close to the threshold T (85%) regardless of the alert presence in all contexts except for *having a conversation*. There was no significant difference between the accuracy with an alert and without it by a Fisher's exact test (significance level was 5%) so that we can say that sounding an alert does not affect the recognition of the audio information in all the activity contexts except for *having a conversation*. On the other hand, in *having a conversation*, there was a significant difference when the ambient sound is -15 dB or less and the accuracy with an alert was higher. This indicates that an alert sound can turn the attention of a user who was concentrated on speaking and listening to a person's story to the contents.

There was little difference in the subjective evaluation of the annoyance level between the presentation with an alert and that without it. A Wilcoxon-signed rank test (significance level is 5%) indicated no significant difference in all the activity contexts or the volume of the ambient sound. Accordingly, we can say that an alert does not affect the annoyance of the presentation. In summary, if a user is having a conversation, the audio information presentation with an alert is useful because it raises the recognition rate without affecting the annoyance level of the presentation. Moreover, we can see the tendency where the annoyance level grows regardless of the alert presence as the volume of the ambient sound increases. On the other hand, because the presentation volume is changed according to the volume of the ambient sound, the ambient sound does not affect the accuracy very much. Briefly speaking, when the ambient sound is large, a user can receive the audio information accurately but uncomfortably.

5.3 Evaluation of Pausing

We evaluate the effectiveness of pausing when it is improper to present a sound. We conducted an experiment in the same way as used for the preliminary evaluation and compared the case of pausing with the case of presenting at a maximum volume. The test subjects subjectively evaluated the annoyance level similarly to the test described in Section 5.2. In the experimental environment, a large BGM volume was played from a speaker. The activity contexts experimented were *walking*, *jogging*, and *using a desktop PC*. The reason why we did not experiment in *having a conversation* is that it is difficult to have a conversation in very noisy surroundings. The pausing interval was set to 10 seconds.

Figure 11 shows the results of the accuracy and the subjective evaluation for each activity context. The accuracy for pausing was generally higher than in the other cases. There was a significant difference in *jogging* and *using a desktop PC* by a Fisher's exact test (significance level was 5%). Accordingly, it can be said that pausing at improper timing enables the user to more correctly catch the presented information when the volume is impossible to present.

The subjective evaluation of the annoyance level was generally low, and there was a significant difference in *jogging* and *using a desktop PC* by a Wilcoxon-signed rank test (significance level was 5%). This seems because the test subjects

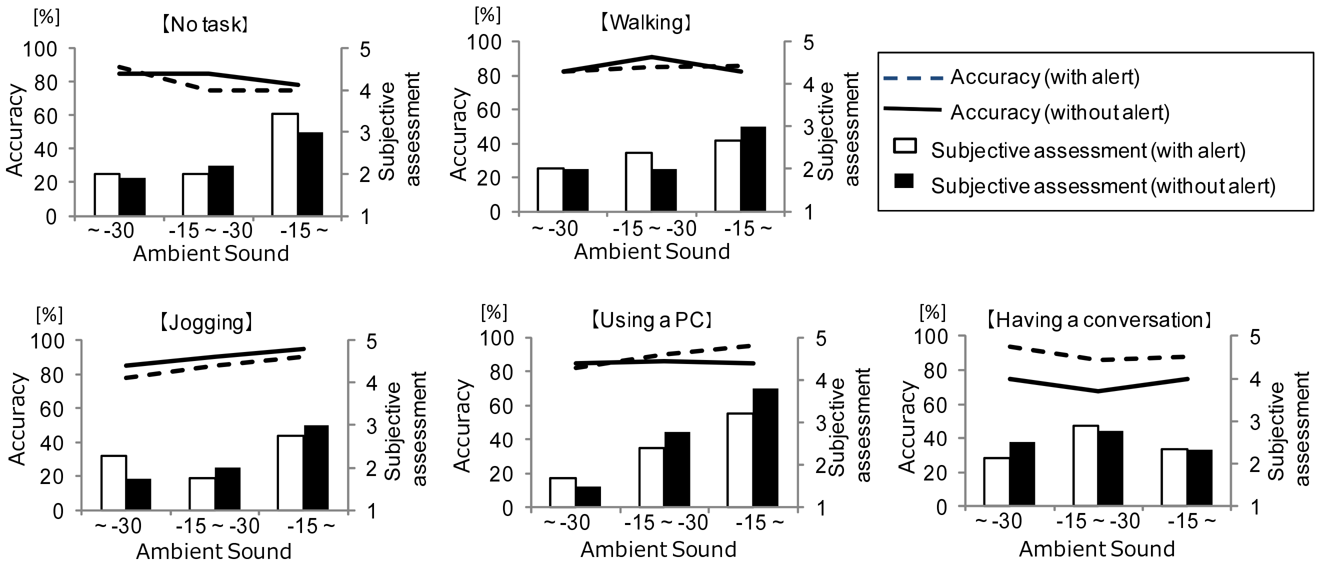

Figure 10: Evaluation results on the effect of alert

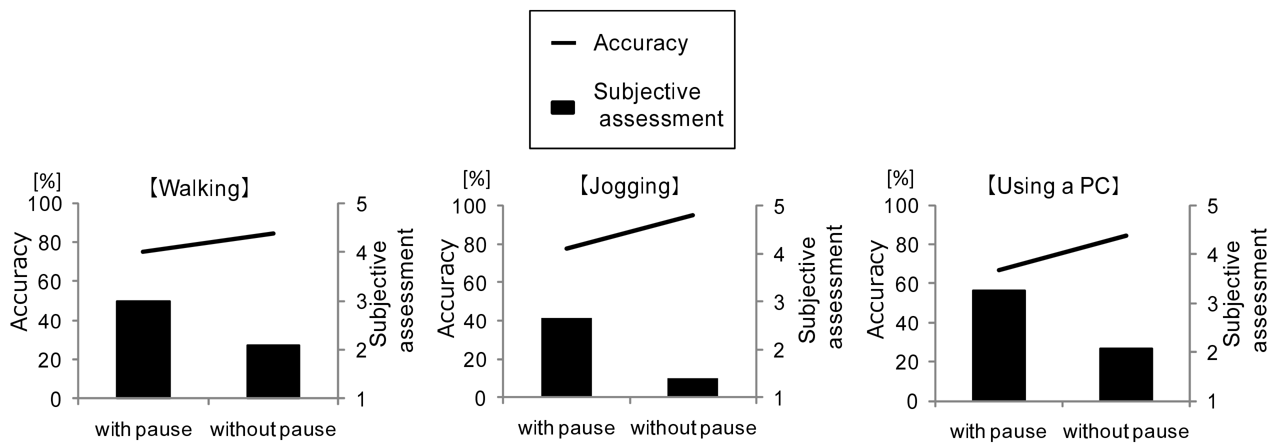

Figure 11: Evaluation results of pausing

felt uncomfortable with a larger volume. When the ambient sound is too large, pausing is effective because it raises the recognition rate and lowers the annoyance level.

5.4 Evaluation on Application

To show the usefulness of presenting audio information such as news flash and navigation messages using our method, five college students who use Twitter usually used the implemented Twitter voice client in a real environment and we discussed their feedback. We compared our proposed method with that presenting at constant presentation volumes of -5 dB and -35 dB. An example of the feedback concerning the presentation methods is shown below.

- The presentation using the proposed method was comfortable and easy to hear.
- The presentation at -5 dB was too large a volume, and thus uncomfortable.
- There did not seem to be so many differences between the proposed method and at a smaller volume (-35 dB) and both were comfortable.

From the feedback, we believe that our method correctly and comfortably enables users to catch audio information. It is thought that the reason for the last opinion is that the experimental environment was quiet, and the presentation volume derived by our proposed method was near -35 dB. Next, examples of other feedback (ex. about the audio presentation of general information such as Twitter updates or voice synthesis performances) are shown below.

- It was convenient because I might not purposely open a web browser.
- I was pleased that updated information is presented.
- All the presentation method was a little annoying when concentrating on another work.

- The user-name is not easily caught.
- Even if the URL is presented, I do not understand it.

We believe from the above-described feedback that it is useful to present general information by using audio delivery. This is because audio information allows users a passive reception of the information and is suitable for wearable computing environments. However, negative feedback where the subjects were annoyed by the presentation while working were also received. Although the updated information on Twitter was presented this time, emergent or real-time information such as earthquake news flashes and schedule information need to be presented even if it interrupts work. Therefore, we need to evaluate the relation between the character of the information and the presentation method and add the necessary system processing, such as not presenting information of low importance while working. Moreover, we believe that the reason for the feedback concerning not easily catching the alphabets such as user-names and URLs is that each alphabet was read out (ex., the word "yes" was read "wai, i, εs") because Japanese voice synthesis was used. These problems correspond with the dictionary data associated with the alphabet and pronunciation. However, because it is impossible to understand even if a URL is presented by a voice, it is believed that the audio information presentation is improper according to the presented content.

6. CONCLUSION

We proposed an audio information presentation method that takes into consideration the user contexts for wearable computing in this paper. We investigated the relationship between the recognition of the audio information and the user's activity and the ambient sound. Based on the preliminary evaluation, we designed and implemented a system that changes the way of presentation of the audio information based on the user's activity, the volume of the ambient sound, and the recognition accuracy that the user desires. We confirmed from our evaluation that our proposed method enables users to correctly catch the audio information in the way that they do not feel pressured.

In the future, we will generalize the factors in the user contexts to further affect the audio information recognition and to consider how to recognize it. Moreover, we need to investigate the recognition rate of the activity contexts and evaluate how this affects the accuracy of the audio information and, the subjective evaluation. In addition, we predict that the change in presentation volume while the contents are being read out annoys the users, while the presented information might not be caught to until the last moment because of the change in the user context if the volume is unchanged. Our proposal of a comfortable presentation method that solves this problem is a future task and we plan to consider the change in presentation volume in the delimitation of the word's or sentences, and to derive a presentation volume that takes the present presentation volume into consideration.

7. ACKNOWLEDGEMENTS

This research was supported in part by PRESTO, Japan Science and Technology Agency and a Grant-in-Aid for Scientific Research(A)(20240009) and Priority Areas(21013034) from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

8. REFERENCES

[1] Aritsuka, T., and Hataoka, N. Gui representation system using spatial sound for visually disabled. In *Int'l. Symposium on Simulation, Visualization and Auralization for Acoustic Research and Education (ASVA97)*, pages 415–420, 1997.

[2] Cheverst, K., Davices, N., Mitchell, K., Friday, A., and Efstratiou, C. Developing a context-aware electronic tourist guide: Some issues and experiences. In *the SIGCHI conference on Human factors in computing systems (CHI 2000)*, pages 17–24, 2000.

[3] Kern, N. and Schiele, B.,. Context-aware notification for wearable computing. In *Int'l. Symposium on Wearable Computing (ISWC 2003)*, pages 263–274, 2003.

[4] Murao, K., Terada, T. and Nishio, S. *Toward Construction of Wearable Sensing Environments, Book Chapter*. Number 278. Wireless Sensor Network Technologies for Information Explosion Era (Book Series: Studies in Computational Intelligence), Springer-Verlag, 2010.

[5] Nakamura, S., Shoji, T., Tsukamoto, M. and Nishio, S. Soundweb: Hyperlinked voice data for wearable computing environment. In *Int'l. Symposium on Wearable Computing (ISWC 2005)*, pages 14–17, 2005.

[6] Naya, F., Ohmura, R., Takayanagi, F., Noma, H. and Kogure, K. Workers' routine activity recognition using body movements and location information. In *Int'l. Symposium on Wearable Computing (ISWC 2006)*, pages 105–108, 2006.

[7] Open-Source Large Vocabulary CSR Engine Julius. http://julius.sourceforge.jp/.

[8] Ouchi, K., Suzuki, T., and Doi, M. Lifeminder: A wearable healthcare assistant. In *Int'l. Workshop on Smart Appliances and Wearable Computing (IWSAWC 2002)*, pages 791–792, 2002.

[9] Roy, D, K., and Schmandt, C. Newscomm: A hand-held interface for interactive access to structured audio. In *the SIGCHI conference on Human factors in computing systems (CHI 1996)*, pages 173–180, 1996.

[10] Sawhney, N., and Schmandt, C. Nomadic radio: Speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human Interaction*, 7(3):353–383, 2000.

[11] Tanaka, K., Kishino, Y., Miyamae, M., Terada, T., and Nishio, S. An information layout method for an optical see-through head mounted display focusing on the viewability. In *Int'l. Symposium on Mixed and Augmented Reality (ISMAR 2008)*, pages 139–142, 2008.

[12] twitter. http://twitter.com/.

[13] V. Vapnik. *The nature of statistical learning theory,*. Springer, 1995.

[14] Wilson, J., Walker, N, B., Lindsay, N., Cambas, C., and Dellaert, F.,. Swan: System for wearable audio navigati on wearable computers. In *Int'l. Symposium on Wearable Computing (ISWC 2007)*, pages 91–98, 2007.