Context-aware Multimodal Interfaces Enhancing Ubiquitous Learning

Mengmeng Li
The University of Tokushima, Japan
Lemonrain99@gmail.com

Hiroaki Ogata
The University of Tokushima, Japan
ogaoga@mb2.tcn.ne.jp

Bin Hou
The University of Tokushima, Japan
myvstar@gmail.com

Noriko Uosaki
The University of Tokushima, Japan
orchard.place@gmail.com

ABSTRACT
This paper proposes multimodal interfaces for a mobile learning system called SCROLL (System for Capturing and Reminding of Learning Log). The system is designed to help learners to record what they have learned (called as learning log) and recall their learning logs by providing them with quizzes. With the help of multimodal interfaces, learners can interact with the system by speech and body movements. It not only increases the entertainment of learning but also facilitates learners to use the system in the environment, in which it is not comfortable for learners to watch the display and input with keyboard. Another feature of the interface is that it adapts to the learners' context. It means that it can suggest learners to choose appropriate way to interact with the device in different environment.

Author Keywords
Multimodal interface, context-awareness, mobile learning, ubiquitous learning log

INTRODUCTION
Mobile learning is known as learning anytime and anywhere. And with the evolution of the mobile technology, mobile devices company the learners 24 hours. But it is still not possible for learners to study anytime and anywhere, because for most mobile learning systems users have to input using keyboard and receive the output by watching the display. But from morning till night, there are lots of situations in which learners' hands and eyes are not available, for example when we are walking, standing with heavy baggage, or wearing gloves. In such situations, they cannot input with keyboard using their hands or watching on the screen, but they can speak with their voice, move their body and hear with their ears. In these situations, a hands and eyes free interface are desired. Therefore, this paper proposes a multimodal interface which can increase the availability of the mobile learning systems.

The interface is built on a learning log system called SCROLL (System for Capturing and Reminding of Learning Log). The main functions of SCROLL consist of helping learners to record and share their learning experiences and recall what they have learned in quizzes. Usually learners can interact with the system by touching and viewing the display of the devices. However, even though mobile learning is thought to help learners with anytime and anywhere learning (Motiwalla, 2007), the normal interface is not available for them to use in some situations, such as when a learner is walking or the learner is wearing gloves. Therefore, to enhance the usability of mobile learning in such situations, a multimodal interface consisting of speech and body movements is proposed in this paper. What' more, the system can also be aware of the learner's environments such as whether it is too noisy for the devices to hear the learners' voice or whether the learner is moving at a high speed (like running) so that it is impossible for him to use the device and so on. Based on these judgments, the system can also suggest which interaction modality the learner should use.

The rest of the paper is constructed as follows. Next section introduces the related work. Then, both the "learning log" term and SCROLL are introduced in detail. In the fourth section, we describe the design of the multimodal interface. At last, conclusions and the future work are presented.

RELATED WORK
Before our study, many researchers have tried to enhance mobile learning using multimodal interface in the literature in technology enhanced learning field. For example, Wolff & Eichner integrated speech processing technology with a web-based e-learning environment in 2004 (Werner, Wolff, Eichner, & Hoffmann, 2004). Bischoff developed a text-to-speech interface on a context-aware mobile learning environment in 2007 (Bischoff, 2007). Motiwalla also did a case study on enhancing mobile learning using speech recognition technologies in 2007 (Motiwalla & Qin, 2007). In 2009, Kondratova discussed the multimodal interface in mobile learning and proved that it was technically possible to implement speech-based and multimodal interaction with a mobile device and to achieve significant level of user acceptance and satisfaction with technology (Kondratova & others, 2009). However, he also pointed out that there were some important

contextual constrains that limit applications with speech-only interfaces in mobile learning, including social and environmental factors, as well as technology limitations (Kondratova & others, 2009). Thereby, in this study we manage to propose several possible kinds of multimodal interface for learners to break through those contextual constraints, such as body movement, which is also an effective interaction way studied by researchers (Jaimes & Sebe, 2007). The next section we firstly introduce the mobile learning system called SCROLL, on which our proposed interfaces are based.

WHAT IS SCROLL?

SCROLL stands for System for Capturing and Reminding of Learning Log. The term of learning log is defined as a recorded form of knowledge or learning experience acquired in our daily lives and it serves as memory storage for notable or important knowledge to review, to remind and to reflect (Li, Ogata, Hou, Uosaki, & Yano, 2012). For example, the notes taken in the foreign language learning can be called a learning log (shown in Figure 1(1)). But besides the text information, a learning log also contains photo, audio, video and some meta-data such as the location, time and so on.

An important goal of SCROLL system is to help learners recall what they have learned after they archived their learning logs. When a learner captures his learning log, besides the location based property mentioned above, a number of things are designed for learners to encode as retrieval cues. For instance, according to the picture superiority effect, the learning logs with pictures are much more likely to be remembered rather than those without pictures (Nelson, Reed, & Walling, 1976). In addition, according to the basic research on human learning and memory, practicing retrieval of information (by testing the information) has powerful effects on learning and long-term retention. And compared with repeated reading, repeated testing enhances learning more (Karpicke, Butler, Roediger, & others, 2009). For these two reasons, the quiz function taking advantages of the pictures, locations and so on is proposed. Three types of quizzes can be generated automatically by the system, which are image multiple-choice quiz (shown in Figure 1(2)), text multiple-choice quiz (shown in Figure 1(3)) and yes/no quiz (shown in Figure 1(4)). Because such three quizzes stands for the basic interaction way in mobile learning systems, how to support learners to do these quizzes in different situations are mainly discussed in this paper.

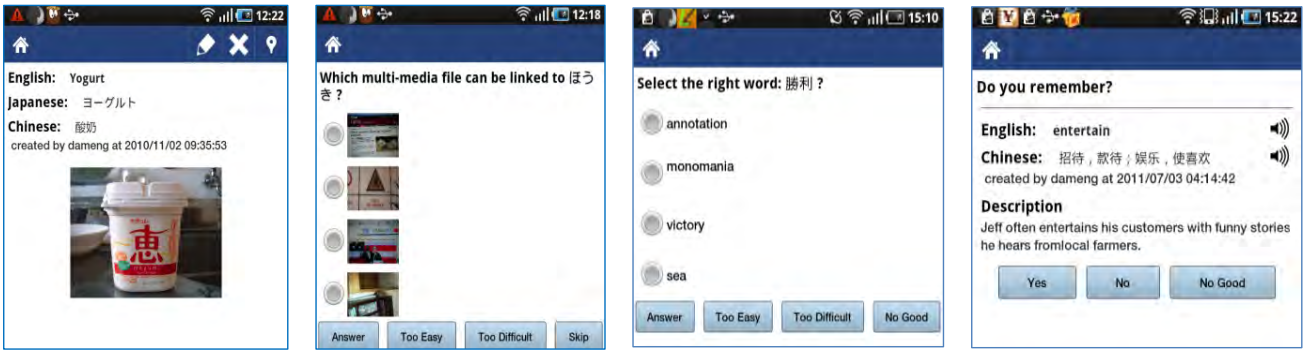

Figure 1 **(1) A learning log** **(2) Image-based Quiz** **(3) Text-based quiz** **(4) yes-no quiz**

DESIGN OF CONTEXT-AWARE MULTIMODAL INTERFACE

As far as we concerned, among all the human organs People's hands, mouth and head are possible for learners to input and eyes and ears can be used to catch the output. Table I and II show available human organs for input and output, the corresponding situations and the appropriate interfaces. The following sections describe two multimodal interfaces and a context-aware feature.

Available human organs for input	Situations	Interfaces
Hands	Normal way	Keyboard based
Mouth	Walking, wearing gloves, etc.	Speaking
Head	Studying in the library, traveling on the train	Body movement

Table 1. human-organs for input, used situations and interfaces

Available human organs for output	Situations	Interfaces
Eyes	Normal way	Display based way
Ears	Walking, driving.	Voice

Table 2. human-organs for output, used situations and interfaces

Speaking to the device

Voice modality is the first considered way when we cannot use our eyes and hands. And it has already been studied widely and proved to be possible. Among the three kinds of SCROLL, the text-based multiple choice quizzes can be assisted with this interface. For example, when learners speak the nature language, such as "Please give me some quizzes of Japanese" or "I need some quizzes of Japanese", the system can interpret them into "get quizzes of Japanese". Then, the system reads out the quizzes, learners can also provide their answers in voice. After recognizing the answers, the system will check the answers and provide more explanations in voices. Parsing Expression Grammar (PEG) technology can be used to analyze the construction of the speech to get both learners' commands and their objects. Similarly, the system can also provide more different commands to support other system functions besides quizzes.

Moving your head

Besides speech, body movement way is also a possible way. This is because the front camera of smartphones can be used to track learners' body movement. We propose to catch learners' head movement. For instance, if the learner nods his head, it means "yes" while if the learner shakes his head, it means "no". This kind of interface can be used in the social context that is not available to speak loudly, such as in the train or subway. This method can be used in the "yes-no" quizzes.

Shaking devices

Because the recent smartphones are equipped with a range of sensors, such as accelerometer, ambient light sensor, GPS, microphone, and so on, it is also possible to use these sensors to interact with device. For example, we propose to use accelerometer sensor on the device to catch the device movement in front and back direction or left and right direction. The front and back direction can be interpreted into "yes" while the left and right direction stands for "no". This method can also be utilized in "yes-no" quizzes.

Situated multimodal interface

Because the interfaces of the system include three kinds: text-based, speech and body movements, it is necessary to recommend the learner to choose the appropriate method in different situation. Consequently, we can recommend learners the appropriate interface for learners by detecting the learners' current environment via these sensors. Two aspects of the environment detected by the system are listed in the follows:

1. Whether the surround environment is noisy and whether learner is wearing earphone: This is used to judge whether the environment is suitable for the system to read out for learners and whether it is suitable for learners to respond the system using voice. The speaker of the device is used to detect the sound of the environment to judge whether it is noisy. The system can also check whether the earphone is working.
2. Whether learner is moving in a high speed. This is used to judge whether it is appropriate to provide a learner speech based interface, because it is impossible for learner to see the text information if the learner is walking or jogging. We distinguish learners' movement way based on the Table III by using the speed data obtained from GPS.

Speed	***Movement Way***
0~5km/h	Walking (Knoblauch, Pietrucha, & Nitzburg, 1996)
5~10km/h	Running, Jogging, Riding on bicycle
10~50km/h	Bus/Car
>50km/h	Train/Subway (Toshiaki, Ryota, Hirokazu, & Tadashi, 2005)

Table 3 speed and movement way

CONCLUSIONS

This paper proposes a multimodal interface for mobile learning system. The interface is supposed to provide three benefits learners. Firstly, it will unbind the learners' hands and provide a more comfortable way especially in quite a lot of situations, such as when the learners are walking. Secondly, it will provide a diversified way for learners to interact with the system and increase learners' interests. Thirdly, the system can be aware learners' environment and recommend appropriate interface option for learners based on the judgment of the environment. The multimodal interface contains the speech, shaking the device and tracking head movements. As for the future work, we will implement and evaluate the interfaces.

ACKNOWLEDGMENTS

This research work was supported by JST PRESTO, and the Grant-in-Aid for Scientific Research No. 21650225 from the Ministry of Education, Science, Sports, and Culture in Japan.

REFERENCES

Bischoff, A. (2007). The Pediaphon-Speech Interface to the free Wikipedia Encyclopedia for Mobile Phones, PDA's and MP3-Players. *Database and Expert Systems Applications, 2007. DEXA'07. 18th International Workshop on* (pp. 575–579).

Jaimes, A., & Sebe, N. (2007). Multimodal human-computer interaction: A survey. *Computer Vision and Image Understanding*, *108*(1-2), 116–134.

Knoblauch, R. L., Pietrucha, M. T., & Nitzburg, M. (1996). Field studies of pedestrian walking speed and start-up time. *Transportation Research Record: Journal of the Transportation Research Board*, *1538*(-1), 27–38.

Kondratova, I., & others. (2009). Multimodal interaction for mobile learning. *Proceedings of the 13th International Conference, HCI International 2009*.

Li, M., Ogata, H., Hou, B., Uosaki, N., & Yano, Y. (2012). Personalization in Context-aware Ubiquitous Learning-Log System. *2012 IEEE Seventh International Conference on Wireless, Mobile and Ubiquitous Technology in Education* (pp. 41–48).

Motiwalla, L. F. (2007). Mobile learning: A framework and evaluation. *Computers & Education*, *49*(3), 581–596.

Motiwalla, L. F., & Qin, J. (2007). Enhancing Mobile Learning Using Speech Recognition Technologies: A Case Study. *Management of eBusiness, 2007. WCMeB 2007. Eighth World Congress on the* (pp. 18–18).

Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 523.

Toshiaki, Z., Ryota, H., Hirokazu, A., & Tadashi, K. (2005). Development of a method for estimating transportation modes with handy GPS equipment. *The 4th ITS Symposium*, 1–6.

Werner, S., Wolff, M., Eichner, M., & Hoffmann, R. (2004). Integrating speech enabled services in a web-based e-learning environment. *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on* (Vol. 2, pp. 303–307).