A Context-aware Multimodal Interface for Mobile Learning

Mengmeng Li, Ogata Hiroaki, Bin Hou, Noriko Uosaki
The University of Tokushima, lemonrain99@gmail.com

Abstract

This paper proposes multimodal interfaces for a mobile learning system called SCROLL. 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, which is not suitable for reading and keyboard-based inputting. 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.

1. Introduction

Recent years, many researchers have tried to enhance mobile learning using multimodal interface in the literature in technology enhanced learning field. For example, 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. 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 [1]. Thereby, in this paper we propose several possible kinds of multimodal interface for learners to break through such contextual constraints.

The interfaces are built on a 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 (shown in Figure 1). [2] has introduced SCROLL in details. Usually learners interact with the system only by touching and viewing the display of the devices. Therefore, how to support learners to do interact with the system in more different situations are mainly discussed in this paper. The rest of the paper is constructed as follows. In next section, we describe the design of the multimodal interface. Finally, conclusions and the future work are presented.

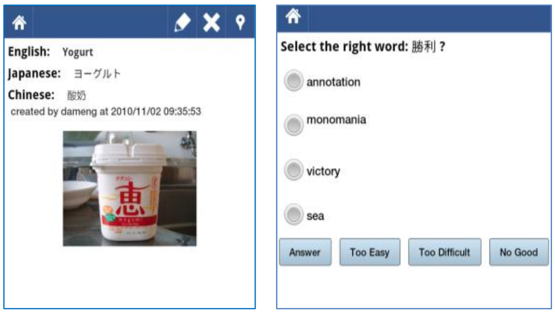

Figure 1 Interface of SCROLL

2. 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

2.1 Speaking to the device

Voice modality is the first considered way when we

CollabTech 2012 , August 27-29, 2012, Hokkaido, Japan.

cannot use our eyes and hands. And it has already been studied widely and proved to be possible. In SCROLL, a function called text-based multiple choice quizzes can be assisted with this interface. For instance, when learners speak to the device, such as "Please give me 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.

2.2 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.

2.3 Shaking your device

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.

2.4 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 3 by using the speed data obtained from GPS.

Speed	Movement Way
0~5km/h	Walking [3]
5~10km/h	Running, Jogging, Riding on bicycle
10~50km/h	Bus/Car
>50km/h	Train/Subway [4]

Table 3 speed and movement way

3. Conclusions

This paper proposes a multimodal interface for mobile learning system. There are at least three benefits providing for learners. Firstly, it can 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 provides a diversified way for learners to interact with the system and it increases 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 firstly.

References

[1] I. Kondratova and others, "Multimodal interaction for mobile learning," in Proceedings of the 13th International Conference, HCI International 2009, 2009.

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

[3] R. L. Knoblauch, M. T. Pietrucha, and M. Nitzburg, "Field studies of pedestrian walking speed and start-up time," Transportation Research Record: Journal of the Transportation Research Board, vol. 1538, no. -1, pp. 27–38, 1996.

[4] Z. Toshiaki, H. Ryota, A. Hirokazu, and K. Tadashi, "Development of a method for estimating transportation modes with handy GPS equipment," The 4th ITS Symposium, pp. 1–6, 2005.