A Smart Speaker Application to
Assist Japanese Onomatopoeia Learning: A Prototype

Takahiro Tagawa
Research Institute for Information Technology, Kyushu University, Japan
tagawat@cc.kyushu-u.ac.jp

Meng Jin
Graduate School of Integrated Frontier Sciences, Kyushu University, Japan
2FS18206N@s.kyushu-u.ac.jp

Hitoshi Inoue
College of Healthcare Management, Japan
jin-inoue@healthcare-m.ac.jp

Abstract: This paper describes the characteristics of smart speaker and how and why this device is expected to play a new role in language learning. Next we describe the characteristics of words called "onomatopoeia", and the difficulties in learning them. We build a prototype of the application for smart speakers to assist learning Japanese onomatopoeia, implemented as a simple quiz application that plays the audio file as the question and requests the learner to answer what onomatopoeia to be used to express the sound. Then we discuss the findings from the prototype, and the future vision of the application of smart speakers to be combined with other learning assistant/management systems.

Introduction

Recent years the human interface that can understand the speech of human and answer or correspond it became real, by the combination of voice recognition technology, voice synthesis technology, and natural language processing technology. By such interface a human can operate or order the computer systems by his/her voice, or receive various assistance based on requests by voice. Such systems and programs are called "Virtual Assistant", and now they are used in combination with the smartphones and speakers ("Smart Speaker").

Such systems can establish "a communication natural to humans" between humans and the machines(computers), so it is natural to expect that such computer systems/services will be of great help for learning or education that deals with(or includes) the communication. In particular, application of such technologies in the field of language learning is expected, because the language learning requires listening, speech, and iteration.

On the other hand, in the field of foreign language learning, it is pointed out that learning words called "onomatopoeia" to express animal voices, natural sounds, body movements, impressions and emotions, is difficult. The onomatopoeia is often used by native speakers because they are the natural expression of their feeling, so learning onomatopoeias is important for non-native speakers such as foreign students to establish the foundation of life and social relationship with local people.

In this paper, we propose a use of smart speakers for learning onomatopoeias, build a prototype of application of smart speaker to assist Japanese onomatopoeia learning for non-native speakers, and report findings acquired from the prototype.

Japanese Onomatopoeias and Mimetic Words

"Onomatopoeia" is a word to express sounds of something or voices of animals or humans. Mimetic words (or mimicry words) are used for imitating body movements, situations, sense, and feeling. These are often collectively referred as "onomatopoeia". Japanese language includes many onomatopoeias, however many non-native Japanese speakers feel learning Japanese onomatopoeias difficult. Even the foreign students who passed the exam of Japanese Language Proficiency Test (https://www.jlpt.jp/) tend to avoid using Japanese onomatopoeia. The following factors can be understood as the reasons for it:

(1) Usages of almost all of Japanese onomatopoeia are based on Japanese traditions and cultures to express the feeling about the subject, and other languages do not necessarily have words with the exact same meaning.

(2) Onomatopoeias are often used for informal expression of subjective feeling, impression. The use of onomatopoeias is often regarded as infantile (childish) and tends to be avoided in public, so the necessities to learn Japanese onomatopoeia are given a low priority in education of Japanese as the second language.

However, the understanding and the use of the onomatopoeia are important for the non-native Japanese speaker such as international students, to establish intimate and close communication with the society in Japanese people.

As the teaching material, there are many textbooks of Japanese language that include onomatopoeias, however the usages of onomatopoeia are closely related to the situations and feeling, so the textbook is not enough because of the lack of the reality.

There is a teaching material specially developed to assist learning Japanese onomatopoeia for foreign students (http://nihongo.hum.tmu.ac.jp/mic-j/KIRAKIRA-material/index.html). This learning content consists of seven video clips that deal with the students' life and the subtitles are inserted to emphasize the occurrence of onomatopoeias, so the learner can recognize the relationship between the situation shown in the video and the word (onomatopoeia) to express the situation.

On the other hand, the use of the word or speech itself also plays the essential part of the language learning. The learner of the language learns not only from reading and/or listening, but from the using (writing and speaking) words and phrases.

Here we focused attention on the smart speakers as the teaching material or media for onomatopoeia, because of their speech recognition function and audio playback functions. The smart speaker can play a sound file of voices and sounds, and also can conduct a test/quiz to ask what word(onomatopoeia) to be used to express the sound, and the learner need to answer the question with his/her own voice. We can combine a hearing (the understanding of the sound) experience and a speech. This combination is often said to be an essential factor of the language learning.

Smart Speakers

A smart speaker is a wireless speaker with voice command interface that offers interactive actions and hands-free activation of some applications and services. Users of smart speaker can operate some computer (or cloud) system with their own voices, instead of (or as an assistance for) using the keyboard or touch panel or mouse. The speaker has a connection with the server of virtual assistant operated by the vendor (such as google, amazon, etc.), and the server uses the result of natural language processing of voice input to answer the command or input by the user. This type of interface, called Voice User Interface (VUI for short) has advantages compared to text input/output in (1) being intuitive (2) quick input (3) hands-free operation (4) availability of various audio sources. These factors contribute to the growth of the use of the smart speakers at home.

Main vendors of smart speakers and virtual assistants offer the developing environment of the applications(apps) to the users and the user can develop the applications for their own purpose, not only using pre-produced applications developed by vendors and/or other users.

Characteristics of the VUI described above can be expected to open a new possibility of the applications of technology for education, especially for the fields that include speech, linguistic interaction or listening behavior, like language education.

There are researches on the use of smart speakers for second language learning (Moussalli 2016) (Dosen2017), but they are generally the overview of the possibilities of application and not necessarily on the application of smart speaker itself or development of the systems (apps) for specific learning issue. Here we propose the use of smart speaker for learning Japanese onomatopoeia, and show a prototype of the system (app).
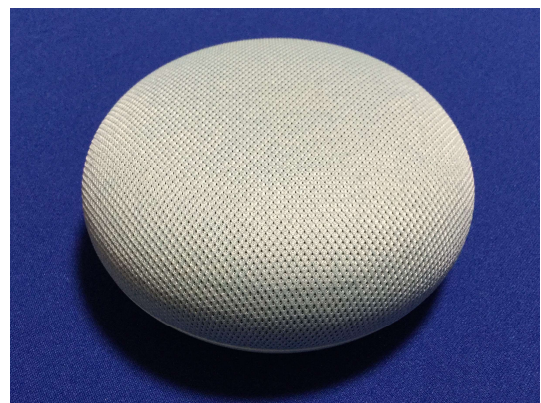

Figure 1.   Smart Speaker
(Google Home mini)

Smart Speaker Application to Assist Japanese Onomatopoeia Learning: A Prototype

We made a prototype of a smart speaker application for learning Japanese onomatopoeia. We adopted "Google Home mini" smart speaker (Figure 1) and "google assistant" (application that offers VUI) as the basis of our prototype.

Through using google assistant the user can use various applications (functions) registered in the platform called "Actions on Google". Also, the user can design and develop their own application that uses the VUI, with the tool called "Dialogflow". We developed our prototype using this Dialogflow.

This application is designed as the quiz system. The speaker plays the sound (natural sound, or human/animal voice) and asks the user (learner) to answer the correct onomatopoeia word. The speaker tells whether the answer was correct or incorrect, and proceeds to the next question. This application works as shown in the diagram in Figure 2. The sound used as the question is stored as mp3 file in the web server. The sound file is referred using SSML (Speech Synthesis Markup Language) from the index.js in Dialogflow, instead of the script line of synthesized speech. For example, the description, '<speak><audio src="https://XXXX/YYYY/cat-cry1.mp3" clipBegin="0s" clipEnd="5s"></audio></speak>' is used to play the mp3 file of crying cat voice with five second length.
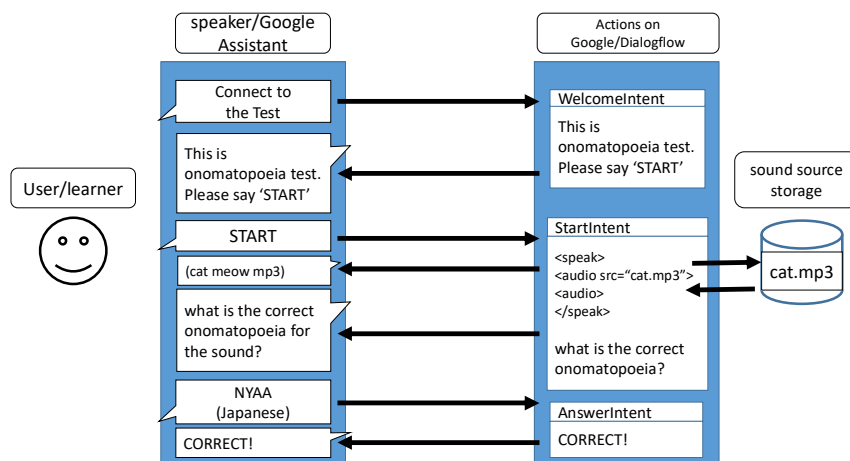

Figure 2. Diagram to Show the Process of Quiz in Our Prototype Application

Findings from Prototype

The smart speaker application described above is just a prototype and we have not conducted the quantitative evaluation of the effectiveness yet. However, through building the prototype we acquired following findings.

1. The number of onomatopoeias that have original sound/voice is not large. Many of onomatopoeias are used to express the emotion, atmosphere (impression), movement, and they do not necessarily occur with sound. In the field of entertainment (such as TV show or movies) these emotion, impression and movements are often expressed with some sound effect for dramatic impact, however understanding of such sound effect is context-dependent, and the sound itself does not have one-to-one correspondence with the onomatopoeia.
2. Instead of the use of sound file, verbal descriptions of impressions or movements can be used as the question of the quiz. This usage is just the same thing as "the riddle". However sometimes the pronunciation of words artificially synthesized as the answer has some intonation that is strange to the native Japanese speaker. This needs attention for the use of smart speaker for language learning purpose.
3. In the prototype above we included the quiz as a part of program. This implementation has merit that the structure of the system can be simple and easy to develop. On the other hand, this implementation lacks the scalability, in the sense that adding questions is often troublesome. Dividing the design of the system (application) and the management of the learning content is necessary.

Conclusion

In this paper we proposed the application of smart speaker for learning Japanese onomatopoeia, showed the prototype design of the smart speaker application of it, and discussed its merits and the limitations. We implemented the application software as the simple multiple choice quiz to answer the correct Japanese onomatopoeia word after the application plays the sound of some voice of human or animal, or natural sound. This prototype was only built for

verification/experimental proof of the applicability of smart speakers for second language learning or speaking, and the type of the questions or type of events the learner experience may be better with the different design of interaction, for example, with the "fill-in-the-blank" questions, or role-play in some situation.

Also, it is already tested that the smart speaker applications can work in cooperation with the LMSs (Kita 2018). This means that the learning contents (teaching materials) can be set up as a part of course on LMS, and the records of the learning can be stored on the LMS and can be reviewed by the teachers, or students in their selves. This also implies that the teachers (professors) can set up complex algorithm such as changing the learning items adaptively along with the learning score of a student, not by creating program of smart speakers, but by using the standard function of LMSs.

References

Kita, T., et al. (2018) Learning Activities on an LMS via Smart Speakers, *IPSJ-SIG TechnicalReports,* 2018-CLE-26(16), pp. 1-5 (in Japanese) .

Dizon, G. (2017) Using Intelligent Personal Assistants for Second Language Learning: A Case Study of Alexa. *TESOL Journal*. 8. 811-830.

Moussalli, S., et al. (2016) Are Commercial "Personal Robots" Ready for Language Learning? Focus on Second Language Speech, *Proceedings of EUROCALL 2016 Conference*, pp.325-329

Acknowledgement