Study on the Use of Speech Recognition Function to Practice Speaking English Using the Voice Translator "Pocketalk"

Harumi Kashiwagi, Kobe University, Japan
Min Kang, Kobe University, Japan
Kazuhiro Ohtsuki, Kobe University, Japan

The Asian Conference on Education & International Development 2021
Official Conference Proceedings

Abstract

Although some speech recognition software is highly developed, few studies have focused on how this technology should be adapted for foreign language learners with various proficiency levels, including Japanese students. Thus, this study explores the use of speech recognition to support the practice of English speaking by using the voice translator "Pocketalk." English sentences spoken by 95 Japanese university students were identified by Pocketalk's speech recognition function. Afterward, a five-point Likert scale was used to measure the usefulness of the activity with Pocketalk and the affective factors related to speaking English. The results indicated that students tended not to distinctly pronounce the difference between the /n/ sound and the /m/ sound. In addition, when the end of the words such as "terribly" and "stooped" were not pronounced distinctly, they tended to be incorrectly recognized as "terrible" and "stupid." Questionnaire results showed over 70% of the students expressed a positive attitude toward their interaction with Pocketalk, and over 90% of them paid more attention to their pronunciation. Using its recognition function, we could identify how the spoken sentences were actually recognized, which provided clues for correcting their pronunciation. Regarding the affective factors, no significant relationship was found between students' responses to the usefulness of their interaction with Pocketalk and their nervousness in speaking English or their negative feelings toward pronunciation. These results suggest a positive potential for Pocketalk's speech recognition function regardless of their affective factors.

Keywords: Speech Recognition, Voice Translator, Speaking Practice, Language Use

Introduction

According to the English proficiency promotion plan for students by the Japanese Ministry of Education, Culture, Sports, Science and Technology (2015), comprehensive development of the four core language skills: listening, speaking, reading, and writing is required more than ever in English education, and increasing emphasis is being placed on strengthening the ability to output in English. However, it has been reported that many students feel that they are not proficient at speaking English (Kashiwagi, Kang, & Ohtsuki, 2018). When examining the English learning environment, students do not have much opportunity to use English outside of English language classes; they need a practice environment where they can become familiar with speaking English. Using speech recognition technology could be one such way to facilitate their English-speaking practice, especially when natural opportunities to practice are scarce.

Some speech recognition software or applications are highly developed, and they are very accurate for native speakers; they often use their voice dictation functions to type documents or emails. However, few studies have focused on how this technology could be adapted to foreign language learners of various proficiency levels, including Japanese students. Research regarding to what extent foreign language learners' speaking is accurately recognized is needed. Further, it is also important to determine how foreign language learners perceive speech recognition, as some learners with low English-speaking abilities may demonstrate negative attitudes toward this technology.

Thus, this study explores the use of speech recognition to support the practice of English speaking through the voice translator "Pocketalk" (Pocketalk Home Page, n.d.). Pocketalk is a two-way translation device that provides consistently accurate translations across 82 languages. To investigate the following research questions, we conducted an experiment in which English sentences spoken by Japanese university students were identified by the speech recognition function of this translation device. Afterward, a feedback questionnaire was administered measuring the usefulness of the activity with Pocketalk and the affective factors related to speaking English.

1. To what extent are the students' spoken sentences recognized accurately, and which words or parts of speech are not spoken accurately by the students?
2. How do the students feel about the speech recognition function of Pocketalk to practice speaking English?
3. Is there any significant relationship between students' responses to the usefulness of their interaction with Pocketalk and their nervousness in speaking English or their negative feelings toward pronunciation?

The rest of this paper describes the existing literature and our experiment's methodology. It then discusses the results, along with our conclusions, the study's limitations, and recommendations for additional research.

Literature Review

In recent years, speech recognition technology has evolved to enable native speakers to apply its voice dictation function to type documents or emails. However, although speech recognition technology has become highly developed, it is not yet sophisticated enough to recognize English speech by all levels of English language learners (Chapelle & Voss, 2016).

According to Blake (2016), oral dictation activities in language learning can take advantage of speech recognition software, for example, through word recognition or short sentence repetition. Further, according to McCrocklin (2016), the introduction of this technology helps students become more autonomous in their pronunciation practice. Yoon and Zechner (2017) proposed to combine human and automated scoring for the assessment of non-native speech, and with advances in automatic speech recognition, more accurate feedback can be expected (O'Brien et al., 2018).

Building on the previous research reviewed above, we consider how we should adapt this technology to foreign language learners with various levels of proficiency, with a specific focus on Japanese students.

Experiment

Participants

Participants were 95 first-year students who were learning English at a university in Japan. They took a review quiz in their English language classes using Pocketalk, then they answered a post-practice questionnaire.

Procedures

We first provided students with 20 Japanese sentences and their English translations in advance of the review quiz. The same exact sentences were used in the quiz. We instructed them to practice speaking the English sentences without looking at the textual information. The sentences were expressions related to "poor physical condition" and "illness and injury." Next, we administered the review quiz individually to each student. There were three versions of the review quiz, and five sentences were provided in each version of the quiz. Fifteen sentences in total among 20 sentences were used in the quiz. The sentences used in each quiz and the number of participants taking each quiz are shown in Table 1. Students were asked to translate the Japanese sentences into English. They orally answered each question twice. The answer sentences spoken by the students were recognized by the speech recognition function of Pocketalk. The device translated the student's spoken English sentence into English text through the voice-to-text translation function. Two doctoral students who were international students checked whether the students' recognized sentence matched the correct given sentence. These doctoral students were not native speakers of English, however, they were highly advanced ESL (English as Second Language) speakers who had sufficient English proficiency level to deal with international academic settings. We calculated the percentage of correct answers based on whether the recognized sentence matched the correct given sentence. For example, if a student missed one word in a sentence, that was not considered correct. After the quiz, a questionnaire was conducted to gather students' responses regarding the usefulness of the activity with Pocketalk, their nervousness in speaking English, and their negative feelings toward pronunciation, shown in Table 2. The values were scored using a five-point Likert scale (1= agree, 2= moderately agree, 3= neutral, 4= moderately disagree, and 5= disagree).

Quiz version	English sentences	Number of participant s	Percentage of spoken sentences recognized as correct answers (%)
Version 1	He might be depressed.	31	45
	I have terribly stiff shoulders.		19
	You grind your teeth so loudly.		23
	My eyes are a bit irritated.		84
	Your snoring disturbed my sleep.		23
Version 2	I strained my back. It hurts so much.	31	68
	Can you prescribe a Chinese herbal medicine?		35
	It's a throbbing pain.		71
	Do you have medicine for hay fever?		81
	She dresses neatly.		6
Version 3	My eyes are itchy.	33	82
	I want to fix my stooped shoulders.		6
	I toss and turn a lot in my sleep.		52
	Don't push yourself too hard.		91
	Please give me a compress for my sprain.		30

Table 1: English Sentences of The Review Quiz

Results and Discussion

RQ1: To what extent are the students' spoken sentences recognized accurately, and which words or parts of speech are not spoken accurately by the students?

First, we analyzed to what extent Pocketalk identified that the recognized spoken sentences were the correct given sentences by calculating the percentage of students' spoken sentences that matched the correct given sentences. Table 1 shows the results of the 15 sentences that were provided in the three versions of the quiz. The percentages of the sentences recognized as correct vary widely, from 6% to 91%. While four sentences among 15 were spoken accurately by more than 80% of the students, the three sentences highlighted in yellow in Table 1 were spoken accurately by fewer than 20% of them.

Next, based on the observation of the doctoral students, we analyzed the recognized sentences that fewer than 20% of the students answered correctly in more detail. The sentence "I have terribly stiff shoulders" tended to be recognized as "I have terrible stiff shoulders." The sentence "I want to fix my stooped shoulders" tended to be recognized as "I want to fix my stupid shoulders." From these results, it appears that the end of the words "terribly" and "stooped" were not pronounced distinctly. Regarding the sentence "She dresses neatly," the students' spoken sentences were often recognized as sentences that were more dissimilar than those for the previous three sentences, such as "She dresses me today," "She dresses in Italy," or "She dresses nearly." The results suggest that students tend not to distinctly pronounce the difference between the /n/ sound and the /m/ sound.

Using the recognition function of Pocketalk, we could identify how the spoken sentences were actually recognized, which provides clues for correcting students' pronunciation.

RQ2: How do the students feel about the speech recognition function of Pocketalk to practice speaking English?

Here, we investigated how the students felt about the speech recognition function of Pocketalk to practice speaking English by using a post-practice questionnaire. Q1, Q2, and Q3 in Table 2 concern students' interaction with Pocketalk, Q4 is regarding their nervousness in using spoken English, and Q5 concerns their negative attitudes toward English pronunciation.

Questionnaire items	Agree	Moderately Agree	Neutral	Moderately Disagree	Disagree
Q1: Pocketalk is useful when I practice speaking English for self-study.	35 (36.8%)	36 (37.9%)	11 (11.6%)	11 (11.6%)	2 (2.1%)
Q2: Pocketalk helped me pay more attention to my pronunciation.	63 (66.3%)	26 (27.4%)	4 (4.2%)	2 (2.1%)	0 (0%)
Q3: I noticed the words and phrases that I have trouble pronouncing by using Pocketalk.	34 (35.8%)	39 (41.1%)	14 (14.7%)	6 (6.3%)	2 (2.1%)
Q4: I feel nervous when I use English in spoken communication.	46 (48.4%)	33 (34.7%)	7 (7.4%)	6 (6.3%)	3 (3.2%)
Q5: I am not good at English pronunciation.	35 (36.8%)	36 (37.9%)	11 (11.6%)	12 (12.6%)	1 (1.1%)

Table 2: Questionnaire Results. (*n* = 95)

According to the results of Q1 in Table 2, a total of 71 students (74.7%) showed agreement with Q1 (i.e., "Pocketalk is useful when I practice speaking English for self-study"), which indicates that almost three-quarters of the students found this device useful. Further, 89 students (93.7%) showed agreement with Q2: Pocketalk helped almost all the students pay attention to their pronunciation. Regarding Q3, 73 students (76.9%) showed agreement, indicating that many students noticed their weak points in English pronunciation thanks to the recognized results of the device.

It is important to note that the correlation coefficient between Q1 and Q3 (r_{Q1Q3}=0.40) in Table 3 (addressed in the next section as part of RQ3) shows a weak relationship between students' responses regarding Pocketalk's usefulness and their awareness of their English pronunciation weaknesses with this device. In the review quiz, the students orally answered each question only twice, which was not enough to conclude whether there is any relationship between them; however, noticing the weak points in pronunciation through interaction with Pocketalk can be regarded as one of this device's useful features.

The analysis of students' responses supports the idea that students have a positive attitude toward their interaction with Pocketalk. Particularly, its speech recognition function can show students not only whether the spoken sentence is correct but also where the error is in the spoken sentence. This device has the potential to give students an idea of how their pronunciation is actually perceived, which helps to make them aware of where their deficiencies are. This result is also consistent with the results of RQ1.

Meanwhile, from the results of Q1, 13 students (13.7%) showed a negative attitude toward the usefulness of Pocketalk. Although further investigation is needed concerning their reasons and factors contributing to the negative attitudes, other approaches for English speaking practice need to be considered for these students.

RQ3: Is there any significant relationship between students' responses to the usefulness of their interaction with Pocketalk and their nervousness in speaking English or their negative feelings toward pronunciation?

For RQ3, we investigated the relationship between students' responses about Pocketalk's usefulness and their affective factors in speaking English. To analyze the relationships between the variables, Spearman's rank-order correlation coefficients on the questionnaire data were determined; results are shown in Table 3.

	Q1	Q2	Q3	Q4	Q5
Q1	—				
Q2	0.16	—			
Q3	0.40*	0.12	—		
Q4	0.13	0.06	0.13	—	
Q5	0.06	-0.13	0.17	0.45*	—

Table 3: Correlations among Questionnaire Items *p<.05

From the correlation coefficient between Q1 and Q4 (r_{Q1Q4}=0.13) in Table 3, no statistical relationship exists between students' responses to the usefulness of Pocketalk and their nervousness in speaking English. Additionally, the correlation coefficient between Q1 and Q5 (r_{Q1Q5}=0.06) shows no statistical relationship between students' responses to the usefulness of Pocketalk and their negative feelings toward pronunciation.

These results suggest that even if students get nervous when speaking English or have negative feelings toward English pronunciation, it does not mean they also have a negative attitude toward using Pocketalk. In addition, because of the weak relationship between Q4 and Q5 (r_{Q4Q5}=0.45), it is suggested that students who become familiar with English pronunciation by using Pocketalk can reduce nervousness regarding speaking English.

Limitations and Recommendations

Certain limitations of the current study should be mentioned. First, in this study, we could observe how the sentences spoken by the students were actually recognized; however, further investigation is needed to determine more details related to common pronunciation errors by preparing more and varied sentences. Pocketalk saves the logs of the recognized results, which can be downloaded as a CSV file. Using these logs, we could identify in more detail how the spoken sentences were actually recognized and provide feedback for correcting students' pronunciation. Second, we found a weak relationship between students' responses on the usefulness of Pocketalk and their awareness of their weak points in English pronunciation through this device; however, the students orally answered each question only twice, which did not provide sufficient data for analysis. Further studies are needed to investigate this relationship. Third, 13 students showed a negative attitude toward the usefulness of Pocketalk. Although further investigation is needed to determine the reasons and factors behind this response, other approaches for English speaking practice should be considered for these students. Lastly, incorporating speech recognition into speaking practice usually focuses on pronunciation, but it is also important to consider how we can incorporate this function in other ways. We hope to consider some task-based activities with the speech recognition function in the future.

Conclusion

This study explored the use of speech recognition to support the practice of English speaking by using the voice translator Pocketalk. We conducted a study in which English sentences spoken by 95 Japanese university students were identified by the speech recognition function of this device. Afterward, a five-point Likert scale was used to measure the usefulness of the activity with Pocketalk and the affective factors related to speaking English.

The results indicated that students tended not to distinctly pronounce the difference between the /n/ sound and the /m/ sound. In addition, when the end of the words "terribly" and "stooped" were not pronounced distinctly, they tended to be recognized as "terrible" and "stupid." Results from the questionnaire showed that more than 70% of the students expressed a positive attitude toward their interaction with Pocketalk, and more than 90% of them paid more attention to their pronunciation from using the device. With its recognition function, we could identify how the spoken sentences were actually recognized, which provides clues for correcting their pronunciation. Regarding the affective factors, no significant relationship was found between students' responses to the usefulness of their interaction with Pocketalk and their nervousness in speaking English or their negative feelings toward pronunciation. These results suggest the potential to use its speech recognition function in English classes regardless of students' affective factors.

As a continuation of this study, we hope to prepare more sentences to identify how the spoken sentences are actually recognized, using the aforementioned data logs. We also will investigate further how we should incorporate the speech recognition function into speaking practice by considering some task-based activities.

Acknowledgement

This work was supported by JSPS Grant-in-Aid for Scientific Research Number JP18K02822.

References

Blake, R. (2016). Technology and the four skills. *Language Learning & Technology*, *20*(2), 129–142.

Chapelle, C. A., & Voss, E. (2016). 20 years of technology and language assessment in Language Learning & Technology. *Language Learning & Technology*, *20*(2), 116–128.

Japan Ministry of Education, Culture, Sports, Science and Technology. (2015). *English proficiency promotion plan for students*. Retrieved from https://www.mext.go.jp/a_menu/kokusai/gaikokugo/__icsFiles/afieldfile/2015/07/21/1358906_01_1.pdf [in Japanese].

Kashiwagi, H., Kang, M., & Ohtsuki, K. (2018). A basic study on the conformity of Japanese university students in language communication activities. *Official Conference Proceedings of The Asian Conference on Education & International Development 2018*. Kobe, Japan (pp.299–309).

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, *57*, 25–42.

O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. L., … Levis, G. M. (2018). Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation*, *4*(2), 182–207.

Pocketalk home page. (n.d). Retrieved January 11, 2021, from https://www.pocketalk.com/

Yoon, S.-Y., & Zechner, K. (2017). Combining human and automated scores for the improved assessment of non-native speech. *Speech Communication*, *93*, 43–52.

Contact email: kasiwagi@kobe-u.ac.jp