様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　４ 年　 ６ 月　 ８ 日現在

機関番号：１３３０１
研究種目：基盤研究(C)（一般）
研究期間：2018～2021
課題番号：１８Ｋ００８２５

研究課題名（和文）A cross-institutional study into speech recognition for language learners analyzing usage patterns

研究課題名（英文）A cross-institutional study into speech recognition for language learners, analyzing usage patterns

研究代表者
Ｇａｒｙ　Ｒｏｓｓ（Ross, Gary）

金沢大学・薬学系・准教授

研究者番号：１０７０８１４２

交付決定額（研究期間全体）：（直接経費）　　3,500,000円

研究成果の概要（和文）：全体として、音声認識の使用は、言語学習者の学習成果を向上させるメリットがあることが示唆されました。結果は有意差のないものでしたが、音声認識の練習を行った学生は、従来の方法よりもテストの点数が高い傾向にありました。さらに、学生は音声認識の使用について、話し言葉を練習するためのシームレスな方法として肯定的な見方をしていました。音声認識はまだ新しい技術であり、モバイルを含むさまざまなタイプのアクティビティを開発し、それが学習成果の向上につながるかどうかを判断するためには、さらなる長期的な研究が必要です。

研究成果の学術的意義や社会的意義
The lack of speech practice in Japanese L2 classrooms is a serious obstacle to English language education. Lack of spoken and presentation skills affects Japan scientifically and politically. This study shows that Speech Recognition Technology can lead to language and speaking improvement in Japan.

研究成果の概要（英文）：Overall findings suggest the use of Speech Recognition holds merit for improving learning outcomes in language learners. Although the result fell just outside of significance, students tended to score better in tests using Speech Recognition practice than using traditional methods. Furthermore, students were positive about the use of Speech Recognition as a seamless way to practice spoken language. Speech Recognition is a new and improving technology and further long-term work is needed to develop different types of activities (including mobile) and to determine whether these lead to improved learning outcomes.

研究分野：Computer Assisted Language Learning

キーワード：speech recognition　speech synthesis　autonomy　feedback

１．研究開始当初の背景

The lack of speech practice in Japanese L2 classrooms has been a serious obstacle to English language education in this country. Despite structural attempts to remedy this situation, cultural issues, student confidence, and institutional issues have stood in the way of progress (King, 2013). Researchers have attempted to use mobile and online learning to improve autonomy, but spoken language has not been the focus of these attempts due to technological limitations (Daniels & Bateson, 2012; Ross, 2013). Speech Recognition has shown potential to evaluate speakers and improve autonomy (Ross, 2016) but there is still a lack of information regarding the most effective activities & feedback methods, the levels of improvement, and student attitudes to such systems.

２．研究の目的

The purpose of the study was to assess the effectiveness of Speech Recognition and Synthesis on learning outcomes and engagement in Japanese L2 learners, and to develop new types of activities using this developing technology.

３．研究の方法

The research was divided into two stages:
In the first stage activities and refinements were developed for the use of speech synthesis and recognition with feedback. I
In the second stage, the effectiveness of speech recognition and synthesis on learning outcomes, autonomy and engagement was assessed:
The main research was conducted on two English language classes (73 students) composed of 2nd-year Pharmacy and Pharmaceutical Science students at the University of Kanazawa. Classes would be randomly assigned learning outcomes using (a) traditional methods such as written exercises versus (b) a speech recognition activity to learn a particular construct or notional pattern. Later performance and attitudes to the two activities would be compared. In order to ensure that each student were subject to the same learning goals per the curriculum, students were given the opportunity to use the other learning method as a follow-up. i.e. those students who used traditional methods were given access to the speech-recognition system, and vice-versa. The results of these follow-up practices were not tracked. The initial plan was to include mobile devices in these studies. However, iOS (Apple iPhone) did not support the HTML5 Speech Recognition API until April 2021 (caniuse.com/speech-recognition), and due to

a significant portion of students being on iOS combined with technical obstacles, finding a workaround proved unfeasible.

４．研究成果

1) The first stage of the research involved ensuring that the speech recognition system functioned effectively and had enough cross-platform support to enable all students proper access. The system could give prompts to which the student would respond in (C) a set pattern (Computer: *Did you go yesterday?* Student: No, I went today), (SO) semi-open questions (C: *What would you buy if you won 100,000 yen?*), and (O) totally open questions (C: *What did you do last weekend?*). Students were asked to rate each type in terms of (i) perceived benefit, and (ii) technical ease of use. The purpose of these early questionnaires was to ensure the system worked on a technical level, and thus how well the students did were not part of this part of the study. Both (C) and (SO) received statistically similar scores [4.0 / 4.1 and 4.1 / 3.9 respectively]. However, the open ended question (O) received a lower score. [3.2/2.5]. This was not altogether surprising. Speech recognition struggles with breaks between sentences, with long pauses interpreted as end of speech and/or a break not being interpreted as a new sentence. This is generally a non-issue with short utterances but causes major interface and interaction issues for learners with longer utterances. Furthermore, sentence errors are only viewed in a fleeting manner in (C) and (SO) as any practice makes way for a new one. On the other hand, open ended style practices involve a long string of sentences appearing on the screen complete with errors. Feedback indicated that students found it distracting, and a less effective use of the technology. Based on these results it was decided that until the technology improves, open ended practices would only be used occasionally, be ungraded, and to inject variety into activities.

Since transcribing speech to written text is the basic function of speech recognition it allows a student's utterance to be compared with a target sentence, and to be graded accordingly. For example, 9 out of 10 target words being correct would give a score of 90%. While such a system doesn't account for the complexity of possible speech errors, the feedback score broadly corresponds to speaking accuracy (Ross, 2017). Another measurement is the 'confidence score'. This is the score the Recognition Engine gives to how confident it has transcribed what the speaker actually said. (This is analogous to a human being's certainty on hearing something correctly: a noisy environment or unfamiliar accent lessening the score). In general, in open ended activities only the

confidence score can be used as there is no target to act as a comparison. At the same time the system displays to the student what the computer 'heard' (i.e. transcribed). A small study was carried out on closed activities to compare student preferences for the two different scoring systems versus no scoring. I.e. (a) accuracy score, (b) confidence score, (c) no score (just the transcript). A Likert-Scale based on the usefulness for each methodology yielded a definite preference for scoring (Accuracy: 4.2, Confidence: 4.0) over no-score (3.1). Discussions with the students however revealed they didn't fully understand what the confidence score was measuring and assumed it was accuracy. Thus it was decided to only use the accuracy score where possible and only to store the confidence score for future analysis and reference.

2) The main body of the study compared how students performed on short-quizzes based on two different kinds of activity: semi-open ended practices vs. closed (drill-style) practices and this was also compared to how students performed when doing more traditional approaches (Control Group). Careful consideration was given to making sure each study approach took a similar time (approx 30 - 50 minutes each week). The scores across activities were compiled into an overall score to avoid statistical multiple comparison problems. It needs to be noted too that the bulk of this stage occurred during Covid-19. It's possible that in an environment where online work became required, students were less willing to do online work than they would otherwise have been.

Each student over the course of the study performed approximately 7,500 utterances with a total of 553,927 utterances (N=73). Students alternated between doing traditional study methods such as online quizzes or grammar based activities and speech recognition based activities.

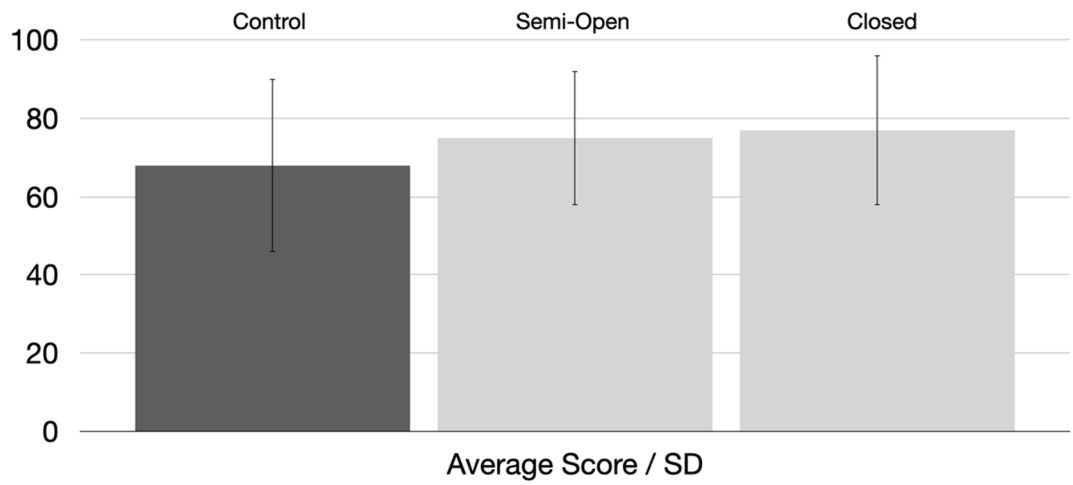

Figure 1: Overall, average score of 73 students on standardized quizzes having done preparatory work using (a) traditional methods (control), (b) semi-open speech-recognition activities, (c) closed SR activities

An ANOVA test revealed $p = 0.059$ for differences between the groups with scores of 68%, 75%, 77% (Fig. 1) for the Control, Semi-Open Speech Recognition, and Closed Speech Recognition Group respectively. This was marginally outside significance. However, the Speech Recognition groups did tend to perform better than the control group indicating that more research is needed and different activities may yield higher scores as the technology progresses. It should be noted too these tests were of the written/listening/grammar variety; spoken tests where one would expect the best improvement to occur were not part of the assessment. This was due to reliable tests of speaking ability not being available for such a large group and the added difficulty of carrying out such tests during Covid. It was notable too that feedback regarding self-study indicated that students were much more willing to use Speech Recognition Systems (3.2 vs 4.1). Students commented that little preparation was needed for such study: ‘one click and the activity runs itself’, and being able to see a score after each utterance and practice was motivating. The ability to study with a mobile device was a common request as it would enable one to study “sitting outside”. The scope of the study didn’t allow for such comparisons but informal discussions indicated that mobile-use was an important consideration in light of Covid restrictions. Of course, other types of study allow mobile use so while it is not a particular advantage of Speech Recognition Learning it is worth noting for further design of such activities, such as making spoken observations on the move.

Overall findings suggest the use of Speech Recognition holds merit for improving learning outcomes in L2 learners. The students tended to score better in tests having completed Speech Recognition practice even though the tests themselves were not speech oriented. Furthermore, students were positive about the use of Speech Recognition as an easy way to practice. As an improving technology further long-term work is needed to develop different kinds of activities, especially more open-ended and on mobile devices, and to determine whether these lead to improved learning outcomes.

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　1件）

１．著者名 Gary Ross, Stephen Henneberry, Glen Norris	４．巻 1
２．論文標題 Speaking with Your Computer: A New Way to Practice and Analyze Conversation	５．発行年 2019年
３．雑誌名 AI and Machine Learning in Language Education	６．最初と最後の頁 152-167
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計6件（うち招待講演　1件／うち国際学会　0件）

１．発表者名 Gary Ross, Stephen Henneberry
２．発表標題 Online speech: utilizing speech recognition in the classroom today
３．学会等名 JALTCALL 2020
４．発表年 2020年

１．発表者名 Gary Ross
２．発表標題 [Keynote] Technology: Bringing us together in a chaotic world
３．学会等名 PanSIG 2020（招待講演）
４．発表年 2020年

１．発表者名 Gary Ross, Glen Norris, Stephen Henneberry
２．発表標題 An analysis of online speech in the classroom
３．学会等名 JALTCALL 2019
４．発表年 2019年～2020年

１．発表者名
Gary Ross

２．発表標題
AI and the Online Speaking Revolution Today

３．学会等名
JALT 2019

４．発表年
2019年～2020年

１．発表者名
Gary Ross

２．発表標題
Online speech evaluating the effectiveness of speech recognition for language learners

３．学会等名
EUROCALL 2019

４．発表年
2019年～2020年

１．発表者名
Gary Ross, Glen Norris, Stephen Henneberry

２．発表標題
Speaking with your computer: a new way to practice conversation

３．学会等名
JALTCALL 2018

４．発表年
2018年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関