The Impact of ASR-based Interactive Video Activities on Speaking Skills: Japanese EFL Learners' Perceptions

Sachiko Nakamura
Tohoku University
<sachiko.nakamura.b6@tohoku.ac.jp>

Ryan Spring
Tohoku University
<spring.ryan.edward.c4@tohoku.ac.jp>

Shizuka Sakurai
Tohoku University
<shizuka.sakurai.c4@tohoku.ac.jp>

Abstract

This study looked at how practically ASR-based interactive video assignments can be integrated into EFL classrooms for additional out-of-class speaking practice, and what effects it will have on students. We created an ASR-based interactive video assignment using Google Scripts and gave it to students as a homework assignment between lessons in which students were learning to give opinions and respond to interrogatives. We used pre- and post-treatment surveys to examine shifts in students' attitudes and gather their responses pertaining to the interactive video assignment. In general, students thought the assignment was good practice, and the majority of students showed positive shifts in confidence and feelings of liking English ($p < .001$ for both, $rs = .28$ and $rs = .48$, respectively). However, we found that students over-focused on pronunciation, especially when they were more proficient learners ($p = .035$, $r = .015$). Therefore, we believe that while the activity was somewhat successful, the feedback likely needs to be changed for ASR-based systems when the focus should not be pronunciation. Future possibilities are also discussed.

Properly participating in speaking activities in foreign language classrooms is imperative for students to improve their L2 speaking skills. This is partly true because they often suffer from

a "poverty of opportunity" (Kato, 2020), in that they often have fewer chances to speak in their target language outside of the classroom than second language learners. Therefore, providing English as a Foreign Language (EFL) learners with enough opportunities to practice speaking and making the most of classroom speaking activities such as discussions is essential. However, many EFL learners have negative feelings toward speaking English in class and feel hesitant to actively participate in discussions, and this is especially well documented in Japan (e.g., Ayedoun et al., 2015; Maftoon, & Ziafar, 2013; Watanabe, 2013; Yashima et al., 2004; Yashima et al., 2018; Yasuda & Nabei, 2018). This high level of anxiety towards speaking English often hinders students from actively participating in speaking activities in class, creating a downward spiral of non-practice that prevents them from acquiring speaking skills. In fact, this phenomenon has led to an entire sub-field of study, known as "willingness to communicate" (WTC), which describes how amenable learners are to enter discourse in their target language (e.g., Al-Murtadha, 2019; Pawlak & Mystkowska-Wiertelak, 2015; Yashima et al, 2004).

Despite the aforementioned circumstances we believe that recent technology has made it possible to provide some monitored speaking practice outside of the classroom, and if the out-of-class practice is similar enough to in-class activities, students might feel less hesitation towards speaking in class, which may increase their speaking time in class and in total. Specifically, we tested if interactive video materials combined with automatic speech recognition (ASR) technology could provide meaningful speaking practice to students outside of class, and if that would in turn raise their willingness to speak in class and redouble the amount of speaking practice that they receive in general. While there are some studies about ASR-based practice, they generally focus on pronunciation, report discrepancy in their results regarding changing student perspectives, and do not include an interactive video element. Therefore, this paper seeks to fill in the gaps in previous studies by investigating how such interactive video activities affect EFL learners' perceptions toward speaking English.

Previous Studies

Active Participation in Speaking Activities

Although many EFL learners show hesitation towards speaking English in their EFL classrooms, Japanese students are often cited as having particularly low WTC (e.g., Cutrone, 2009; Ito, 2022; Osterman, 2014; Watanabe, 2013; Yasuda & Nabei, 2018). This is potentially due to a number of reasons such as the lack of communication confidence in English (Fushino, 2010; Matsuoka & Rahimi, 2010) and the culture in which face-threatening acts and competition tend to be avoided and harmony, indirect speech, and silence are valued (Harumi, 2011; Maftoon & Ziafar, 2013). However, more recent studies have also pointed to the fact that WTC is extremely complex and can be influenced by a number of personal differences such as personality traits such as general trust and L2 proficiency (e.g., Ito, 2022; Yashima et al., 2018). The general lack of WTC, coupled with the fact that using English is not necessarily required in Japanese society, causes learners in Japan to lack exposure to English and have high anxiety towards speaking it. Accordingly, they tend to miss chances to nurture self-confidence in speaking English, culminating in most learners refusing to or being extremely resistant to speaking in EFL settings (e.g., Watanabe, 2013).

Regardless of the exact reasons that WTC is low in Japan, it has been well documented to the point that a number of techniques and teaching methodologies have been developed to attempt

to overcome it. For example, communicative approaches based on social constructivism such as Project Based Language Learning (PBLL), Problem Based Learning (PBL), and Collaborative Online International Learning (COIL) have been suggested (e.g., Beckett, 2006; Kato et al., 2023; Loi & Hang, 2021; Ng, 2009; Nishio et al., 2020; O'Dowd, 2018; Spring, 2020a; Stoller, 2006; etc.). Though such methodologies have shown success in boosting L2 speaking skills and WTC, there are inevitably personal differences in learners that affect uptake, and most of these methodologies require special attention that cannot always be integrated into EFL classes. For example, though PBLL and COIL programs have shown to improve learners' L2 speaking ability overall, studies have suggested that learners do not show much or any improvement through such methodologies if they do not complete assigned study outside of class or have a proper attitude towards learning (e.g., Spring, 2020a). Furthermore, though PBLL, COIL, and other communication-centric methodologies can help develop speaking skills, they often require diverting a large percentage of class time (e.g., Beckett, 2006; Kato et al., 2023; Ng, 2009; Stoller, 2006) or must be given outside of the classroom, and can also require resources that are not available to every L2 teacher (i.e., COIL programs generally require a partner institution with a similar number of participating students) (e.g., Kato et al., 2023; Nishio et al., 2020; O'Dowd, 2018). Therefore, while some of these methodologies can surely help with WTC, they might not be available or appropriate for all EFL contexts.

Several other studies have also suggested interventions to assist with WTC in the classroom that could be used in a wider range of situations more easily than those mentioned in the previous paragraph. For example, Cutrone and Beh (2018) found that task-based instruction had a greater positive impact on WTC than other teaching styles. Similarly, Matsuoka et al. (2014) found that having students give small presentations as part of their class increased WTC. Furthermore, studies such as Kitaoka (2023) have also suggested that enjoyable activities, such as those that integrate music, can increase students' WTC. However, most of these studies focus on what can be done in the classroom, and therefore, there is still a need for more research into how extra-curricular, localized speaking practice can be incorporated into EFL classes in small chunks to potentially increase WTC. We believe one such area with great potential is through newly emerging technologies.

Automatic Speech Recognition (ASR)

ASR technology has increasingly been applied to language learning likely due to the fact that it can be implemented at no cost and the instantaneity of feedback (e.g., Guskaroska, 2020; Spring & Tabuchi, 2022). While ASR technology has developed over the years, the focus of such applications thus far has almost entirely been on improving pronunciation. For example, studies have explored how ASR helps to improve suprasegmental features of pronunciation (Bozorgian & Shamsi, 2020) and the pronunciation of vowel sounds (Guskaroska, 2020). Other research focuses on using ASR as a diagnostic tool to measure the intelligibility of learners' speaking (Mroz, 2018) and how ASR-assisted practice (Spring & Tabuchi, 2021) affects overall intelligibility. Furthermore, a meta-analysis by Ngo et al. (2023) seems to suggest that ASR practice is generally beneficial for EFL pronunciation learning, although their study seems to conflate some non-ASR studies into the meta-analysis, and they found that the effectiveness is influenced by feedback type as well as whether or not peers are included in the practice. Though most of the aforementioned studies signal that ASR-based practice can help improve EFL learners' pronunciation, the same technology has not been applied very much to general

speaking practice, nor are there many studies that focus on how ASR-based practice might influence speaking more holistically.

However, there have also been concerns about using ASR in teaching L2 pronunciation. For starters, some studies have found that students sometimes doubt the ability of ASR to judge pronunciation properly (e.g., Guskaroska, 2020; Inceoglu et al., 2020), and there may be some truth to this claim, as ASR models are generally based on input from native speakers. Furthermore, ASR generally uses not only pronunciation, but also the probability of n-grams (continuous sequences of words) with surrounding words to guess what speakers are saying, and thus careful selection of sentences and phrases is required in order to ensure that pronunciation is being tested and that non-nativelike turns of phrase are not affecting the outcome (e.g., Ngo et al., 2023; Spring & Tabuchi, 2021). However, despite students' perceptions, Spring (2020b) has shown a strong correlation between the intelligibility of learners' pronunciation as judged by five human native-speakers and the percentage of words correctly guessed by Google's ASR technology. Therefore, students becoming frustrated might be more likely due to the ASR task simply being too difficult for them, or due to unrealistic expectations (i.e., they would have to exhibit native-like pronunciation in order to obtain nearly perfect ASR output, which is often what students are expecting).

Importantly, though many studies claim that ASR-based practice helps students improve their L2 pronunciation (e.g., Bozorgian & Shamsi, 2020; Guskaroska, 2020; Ngo et al., 2023; Spring & Tabuchi, 2021; 2022; Wang & Crosthwaite, 2021), there is some discrepancy regarding how ASR-based training affects students' attitudes. Several studies report that learners generally have positive attitudes towards using ASR. For example, Bueno-Alastuey (2011) argues that ASR helped students become more motivated in language learning, which in turn leads to their eagerness to speak and confidence in speaking, and Ahn and Lee (2016) reported that students liked using ASR and found it informative. On the other hand, other studies suggest students can become frustrated by the ASR because the ASR is unable to interpret their speech perfectly (e.g., Guskaroska, 2020; Inceoglu et al., 2020). For these reasons, other studies such as Ngo et al. (2023), Wu et al. (2022) and Evers and Chen (2022) suggest using ASR in groups so that students can get feedback from their peers as well as the ASR system. However, it is still unclear how ASR-based activities affect learner attitudes and perspectives, as this is less studied than improvement in pronunciation (e.g. Ngo et al., 2023) and it is even less clear what the impact will be when using ASR-based activities for general speaking practice instead of pronunciation practice.

Using Video Media in EFL Education

Video media has been increasingly used in EFL education, and many studies suggest it has a number of positive effects on learning. For example, studies suggest that students generally have a positive attitude toward learning English using videos (e.g.,; Fujita, 2019; Nakamura & Spring, 2020) and that they increase student motivation (e.g., Kadoyama, 2008; Kondo, 2018). Many studies also suggest that using videos is effective for learners to acquire language, including vocabulary (Kadoyama, 2008), phrasal expressions in a meaningful way (Yamaguchi, 2018), listening comprehension skills (Kang, 2019), and even reading comprehension and speed when combined with subtitles (Nakamura & Spring, 2020). Thus, using videos in combination with ASR-based practice might help the activities to be better received by students and increase uptake. However, there are no specific studies that explore how videos are

combined with ASR-based activities to create interactive practice, and thus it is still unclear how well these two different modes will blend together to impact learning or learner perceptions.

Factors influencing EFL speaking

A number of studies have noted many factors that can influence L2 speaking, but it is not yet known exactly how these factors comprehensively contribute to improvement and willingness to communicate. For example, learner enjoyment has been invoked as a contributor to both (e.g., Brantmeier, 2005; Dewaele & Alfawzan, 2018; Kitaoka, 2023), and studies such as Saito et al. (2018) argue that using L2 more frequently with positive emotions results in acquisition. Furthermore, Schultz (2017) suggests that the correlation between self-perception, self-assessment, enjoyment and achievement is complex and differs depending on the learner's level of proficiency. Thus, while there is some disagreement amongst the previous studies, it is likely that improving L2 speaking requires some degree of consideration to the students' perceptions and feelings towards the activities.

Other works that provide hints as to what might help L2 learners improve both speaking ability in general and active participation in class often point to the amount of practice and learner anxiety. For example, Nazara (2011) suggests that students should have enough practice time in speaking English, and Liu (2006) notes that students feel less anxious about speaking English with more exposure to oral English. Furthermore, studies such as Kitaoka (2023) and Cutrone (2009) recommend that teachers improve their learners' WTC by starting with activities in which learners would have lower levels of anxiety and gradually moving on to those causing higher levels of anxiety. There are also numerous studies that report a correlation between the time students spend engaging in activities and the improvement of their language (e.g., Krause & Coates, 2008; Nauffal, 2012). From those studies, giving enough time to speaking practice could lower anxiety and contribute to improvement of general English ability. However, it is not yet known if out-of-class speaking practice, such as ASR-based interactive video practice, will provide sufficient practice to raise student engagement and help alleviate anxiety.

Research Questions

Based on the studies above, ASR-based practice with interactive videos might be able to afford L1 Japanese EFL learners with valuable speaking practice outside of the classroom, even when curricula are set such that not much classroom time can be afforded to speaking activities. Such practice may reduce their anxiety towards speaking, in turn making them more willing to speak in class. However, this has yet to be tested, and there is no way to know how such practice will alter their perceptions of speaking English in general, if at all. Therefore, we seek to shine light onto these topics by answering the following research questions:

1. What beliefs do Japanese EFL learners hold about the effects of ASR-based interactive video activities on their speaking skills?

2. What changes are observed in the learners' perceptions of the effects of ASR-based interactive video activities on their speaking skills?

Methods

Participants

The study was conducted with 387 first-year university EFL students in various majors at a Japanese university. Their TOEFL ITP® scores indicated that they are mostly of the Common

European Framework of Reference (CEFR) B1 level, but the range was from A2 to C1: (range: 390-643, M = 503, SD = 36.64). Students had taken six years of formal English instruction before this study and had not spent any significant amount of time abroad (no more than three months). At the time of the study, they were taking two English classes a week: a speaking and listening class (wherein the study was conducted) and a reading and vocabulary class. The curriculum at the university outlines particular skills, phrases, and vocabulary that students are required to acquire in each class but allows instructors to use their own methods to teach them and in whichever order they choose. One of the skills that needed to be addressed in the speaking and listening class was the ability to respond appropriately to interrogatives and give their opinions with supporting details, and this is the skill which we attempt to address in this study. Prior to the study, the authors gained the approval from the research ethics review board in their affiliation. We also provided the participants with informed consent forms, outlining the purpose of the research, and used only the data of those who consented to them.

Procedure

Students learned how to state opinions in class and did some basic worksheet-based practice. Specifically, they were first presented with phrases used to state opinions, such as *as far as I'm concerned*. Then they practiced using them in pairs, asking each other for opinions with various questions (e.g., "Is it better to have more time or more money?"). Students were also encouraged to state supporting details when giving their opinions. They were then given the ASR-based video practice assignment as a homework assignment. Finally, they were asked to conduct role-plays and interactive speaking practice with peers after the video practice assignment.

In the ASR-based interactive video activity, students watched multiple related video clips in which two English speakers talked about an issue (e.g., what topic to talk about in a presentation). At the end of each video clip, one of the speakers looked at the camera and asked a question, encouraging the student watcher to respond or give their opinions. The students then pressed a "speak" button which activated their device's local ASR and spoke their answers. When they finished speaking, they pressed a "stop" button. The recognized text appeared on the screen and students could check if their utterances were recognized as they intended to speak them. If they were not satisfied that the text accurately represented what they wanted to say, they could press a "retry" button which erased the text and allowed them to rerun the ASR. However, they were unable to type their responses or inserted them in any way other than via the ASR, so the instructors could be relatively sure that they actually practiced speaking outside of the classroom. When students finished all of their responses, they pressed the "send" button to send their responses to their instructor.

The activity described above was coded in HTML and Google Script, and the code is available at https://github.com/springuistics/Video-ASRforEFL. We selected these programming languages for ease of adaptability across devices and also because it is relatively easy for others to change colors, texts and videos in the HTML code, and this allows for other interested parties to use the code themselves. Additionally, Google Script was selected because it allows for the HTML form to accept student answers and write the data into a predetermined Google Sheet, which makes it easy for instructors to monitor students' progress and answers. Furthermore, we added a feature to the code used in this study to track and record the number of times that

students pressed the "retry" button. This allowed us to obtain further insights into how many times students attempted the speaking practice.

We collected survey data regarding students' feelings toward the assignment and towards speaking English before and after the assignment. Three of the questions were recorded in Likert-scale style, and one was a multiple-selection question. Furthermore, we asked one open-ended question in the post-treatment survey regarding students' feelings towards the use of videos and the assignment in general. The exact questions can be found in the Appendix.

Data Analysis

We applied a mixed-methods approach to the study, employing both statistical analysis and qualitative analysis of the open-ended questions.

Due to the ordinal nature of the data, we used Wilcoxon signed-rank tests to check for the differences of the results between the Likert-scale questions on the pre- and post-surveys with effect sizes interpreted according to Plonsky and Oswald (2014). Furthermore, we examined the descriptive statistics of the questions regarding which areas of speaking the participants said they wanted to improve on before the activity, and which areas they felt they improved on via the activity. To examine the impact of general L2 proficiency and the number of retries on changes in the participants' perceptions and feelings of having improved, we employed generalized linear models (GLMs). Specifically, we used ordinal models for Likert-scale questions that exhibited statistically significant change, using a transformed delta value (i.e., the posttreatment score minus the pretreatment score) as the dependent variable, and pretreatment question scores, TOEFL ITP® scores, and the number of retries to verify whether or not the latter two independent variables had any impact on changes in perception after correcting for initial pretreatment scores. In order to determine the impact of L2 proficiency or retries on feelings of having improved, we first examined which areas of speaking many of the participants claimed to have improved. For areas where at least one-third of the participants claimed improvement, we employed Bernoulli Generalized Linear Models with having claimed improvement on this aspect of speech or not as the dependent variable, and TOEFL ITP® scores, the number of retries, and the binomial answer on the pretreatment survey regarding whether or not the participants specifically felt that they wanted to improve that aspect of speaking as independent variables.

Open-ended questions were analyzed via a conceptually-clustered matrix to find common themes and help organize them in an interpretable way. Students who did not leave any comment or left the open-ended question blank were excluded from this analysis. Finally, we compare the results of the open-ended questions to the statistical analyses to see which areas overlapped and if any opinions from the open-ended questions could help explain any of the quantitative results we found.

Results

Quantitative Analysis

The descriptive statistics of the Likert-scale questions and item-selection survey questions regarding which aspects of speaking the participants said they wanted to improve in (pretreatment) and felt they improved upon (post-treatment) are presented in Table 1. Furthermore, the change in the average scores of the Likert-scale questions is depicted in Figure

1, and statistical testing is provided in Table 1. According to these results, some detectable differences were found from pre- to post-treatment with regards to questions 1 and 2, but not in question 3. Specifically, there was a small positive change in confidence (question 1) and a medium-sized positive change in the participants' feeling of liking speaking English (question 2), however, we cannot be sure if there was really a decrease in their anxiety towards speaking English (question 3). Furthermore, the aspects of speaking that students felt they most improved upon were pronunciation, contents, and fluency (question 4, in rank order).
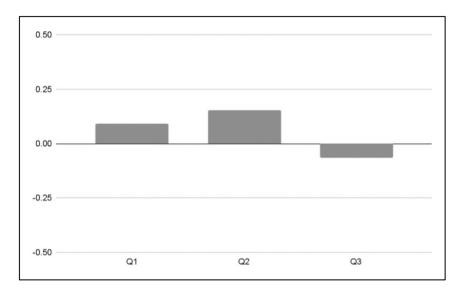

Figure 1. Change in Likert-Scale Questions

Table 1. Descriptive Statistics (include all post-test questions and significance testing when the same questions were on both pre- and posttests. Also include TOEFL and retries data.)

	Pre		Post		Z	p	rs
	M	*SD*	*M*	*SD*			
Q1: Confidence	2.11	0.88	2.21	0.83	2.92*	<0.01	0.28
Q2: Liking	3.19	0.85	3.35	0.83	4.82*	<0.01	0.48
Q3: Hesitation	3.40	1.10	3.34	1.13	1.55	0.12	0.14

Q4: Improvement	The aspects they wish to improve (%)	The aspects they feel they have improved (%)
Vocabulary	80.74	19.55
Grammar	57.51	14.73
Fluency	65.72	34.56
Pronunciation	58.92	54.67
Confidence	44.48	21.81
Contents	41.36	42.49

Retry data (range: 0-223, *M* = 19, *SD* = 26.94) was collected from 179 students. Table 2 provides both the straight correlations for both TOEFL ITP® Scores and retries against the change in Likert-scale questions 1 and 2, and the binomial data regarding whether or not students reported having felt that they improved on either pronunciation, contents, or fluency.

The results suggest some moderate impact of pretreatment score on change in Likert-scale questions, meaning that students who already liked English or felt anxiety were less likely to show change, likely due to some amount of ceiling effect (i.e., if one starts at a maximum score, it is impossible to exhibit upward change). Though other statistically significant correlations were found, the effect sizes were all negligible and likely had little true impact, as indicated by the GLMs.

Table 2. Correlation between Survey Questions, TOEFL ITP® Scores, and Retries

Measure	TOEFL ITP® Scores	Retries	Pretreatment Question
Change in Q1	$rs = -.03$	$rs = .02$	**$rs = -.42$
Change in Q2	$rs = -.08$	$rs = .01$	**$rs = -.36$
Improved Pronunciation	*$r = -.10$	*$r = .16$	$r = .02$
Improved Contents	$r = .06$	$r = .04$	$r = .01$
Improved Fluency	$r = .04$	$r = -.02$	*$r = .10$

$*p < .05, **p < .01$

Table 3 provides the results of the GLM models for each. The GLMs did not reveal any statistically significant impact of TOEFL ITP® scores or retries on change in Likert-scale questions. However, higher TOEFL ITP® scores did seem to be negatively correlated with feelings of having improved pronunciation after correcting for answers in the pretreatment survey.

Table 3. Results of GLM

Overall Model Results			Relationship to Coefficients			
Model	X^2	p	TOEFL Scores	Retries	Pre-Survey Question	Intercept
Q1	2.833	.418	.589	.758	.095	.007
Q2	2.946	.400	.829	.735	.098	.001
Pronunciation	8.612	0.035*	.015*	.228	.611	.019
Contents	2.805	0.423	0.282	0.946	0.266	0.205
Fluency	2.452	0.484	0.644	0.300	0.327	0.854

$*p < .05, **p < .01$

Qualitative Analysis

Table 4 shows the results of the conceptually clustered matrix analysis of the open-ended survey questions. The results suggest that students felt the assignment was good in multiple points. About 9.9% of the students answered that it was good pronunciation practice. Students' efforts to make ASR recognize their language correctly seems to have made them realize the importance of using proper pronunciation. Other students answered that the assignment was:

generally good (7.9%), good speaking practice (5.0%), practical practice (3.3%), a good opportunity to speak English (2.0%), and good listening practice (1.3%).

The awareness of the importance of pronunciation is also reflected in negative comments. Thirty-two of the students (10.6%) answered that pronunciation was difficult, while only 2 (0.7%) answered that stating opinions was difficult. This shows, although this assignment was intended to provide practice stating opinions, students paid more attention to the pronunciation rather than stating opinions. Furthermore, 14 (4.6%) expressed a preference for voice recording rather than using ASR, which is potentially due to their frustration with getting the ASR recognize their speech. This also implies that the students might have faced difficulties in pronouncing English words accurately. The comments about the functional aspects (e.g., *"It didn't work well." "Retry function should be improved so that students can use it sentence by sentence."*) are not addressed here as they are not directly related to the students' perspectives toward linguistic aspects. It will be discussed in the next section.

Table 4. Conceptually Clustered Matrix for Open-ended Survey Question (*n* = 303) (percentages in parentheses).

Titles	Number of students (*n* = 303)
Positive opinions: 89 (58.9%)	1. It was a good pronunciation practice. 30 (9.9%) 2. It was generally good. 24 (7.9%) 3. It was a good speaking practice. 15 (5.0%) 4. It was a practical practice. 10 (3.3%) 5. It was a good opportunity to speak English. 6 (2.0%) 6. It was a good listening practice. 4 (1.3%)
Negative opinions: 62 (41.1%):	1. Pronunciation was difficult. 32 (10.6%) 2. Recording voice is better. 14 (4.6%) 3. It was generally difficult. 7 (2.3%) 4. I wanted feedback. 3 (1.0%) 5. In-class activity is better. 2 (0.7%) 6. Listening was difficult. 2 (0.7%) 7. Stating opinions was difficult. 2 (0.7%)
Tally	
Total: 151 Positive Opinions 89 (58.9%) Negative Opinions 62 (41.1%)	

Discussion

This study investigated the amount of ASR-based interactive video speaking practice engaged in by students and the change of their perceptions toward English. The results suggest that the activity can have a small impact on students' perceptions overall, regardless of how often they retry it or their ability level (as measured by TOEFL ITP®). Furthermore, the results suggest that the activity can make students feel they have improved their pronunciation, but this is mostly true for students with lower L2 proficiency. Students said they mainly needed to improve vocabulary, fluency, and pronunciation (in that order), but they felt they improved their pronunciation, contents, and fluency (in that order) so there was some mismatch between what they wanted and what the activity provided. The results of the qualitative analysis

demonstrate that the top opinions for both positive and negative are related to pronunciation, which implies that simply displaying the transcription provided by the ASR probably made students focus on pronunciation perhaps more than they should have; i.e., they were meant to focus on using the target phrases and their ability to provide a number of details, not necessarily on pronunciation at all.

One important question that is left somewhat unanswered by this study is what is 'too much' focus on pronunciation. Although pronunciation is important for effective communication, it only needs to be intelligible, and not necessarily perfect (e.g., Munro, 2010). Furthermore, according to the trade-off hypothesis (Skehan & Foster, 1997), L2 learners can only focus on one or two of three aspects (i.e., fluency, complexity, accuracy) at a time. Essentially, the task that we created for this study asks students to focus on complexity; i.e., using novel phrases and providing as many details as possible requires focus on lexical complexity at the very least, and likely also syntactic complexity. Therefore, providing feedback that gives clues as to their pronunciation accuracy runs the risk of making the task too cognitively demanding and might cause students to focus on pronunciation to the point that they no longer focus on giving details and using the target phrases. Alternatively, it might be appropriate to provide feedback on the pronunciation accuracy of just target phrases, but not for non-target phrases, which would allow for students to focus on two of three aspects of speaking (i.e., complexity and accuracy), but would lessen cognitive load by limiting their focus to the accuracy of only predetermined target phrases. However, the fact that many students in the survey reported thinking about their pronunciation and fluency suggests that the task could be enhanced by limiting feedback to certain areas to improve focus and lessen cognitive load.

Another important consideration that should be made based on students' comments regarding pronunciation is that of motivation. As indicated by previous studies such as Guskaroska (2020) and Inceoglu et al. (2020), students can sometimes feel demotivated by the results of ASR transcripts. Based on student comments, we found that this occurred to some degree amongst the participants of this study as well. However, we also found positive comments regarding the ASR being good pronunciation training, so we might need to consider a way to provide feedback on pronunciation without it being demotivating. Since many of the negative comments had to do with how the ASR transcribed many of the reasons that students were giving for their opinions, a good balance could be struck by making the ASR simply give affirmative, positive feedback when target words or phrases are pronounced correctly, rather than transcribe everything that the learners are saying. This would reduce the cognitive load, as suggested above, while hopefully still providing some degree of positive feedback and reducing demotivating negative feedback. Another potential alternative is to conduct the activity in class with peers so that they can receive the advice of peers regarding their pronunciation, as suggested by works such as Ngo et al. (2023), Wu et al. (2022), and Evers and Chen (2022).

Based on the results above, we feel that we need to make improvements to the activity in the future to meet students' needs to improve particular target vocabulary, avoid too much focus on pronunciation aspects, and give feedback to the students in a more positive way. We could perhaps deal with those issues by changing the feedback from a visualization of what the students wrote (i.e., ASR writing) to a simpler version, i.e., if the students answer a question correctly using the target vocabulary, they move on to the next video, and if they do not, they have to answer again or are sent to a different video. Theoretically, under such conditions students would not care too much about how to pronounce the words and would concentrate

more on vocabulary use because the words they said are not spelled out, which in turn will lead to more positive attitudes toward using ASR in the speaking practice. There were also many comments about functional aspects, showing their frustration of using ASR and requesting improvement, such as students who mentioned wanting to use the retry function sentence by sentence (the current version only allows retrying the full answer). We will have to alter functions in order to ease their frustration, which we hope will eventually lead to more positive perceptions toward speaking English.

Finally, we do think that increasing opportunities to practice speaking while lowering their anxiety towards speaking English is still needed and that this type of activity has promise as one potential type of opportunity. Specifically, we feel that using technology such as that discussed in this paper will be the key for English learners to practice speaking out of class. However, in the future, we need to examine what influences learners' perceptions in speaking interactively, especially negative feelings such as hesitation and anxiety, and adjust the difficulty levels, topics, and the amount of video activities. Finally, in this study, we used devices' built-in ASR, but as technology develops we will have to constantly consider how the latest developments can be implemented to give students chances to speak English outside the classroom.

Conclusion

Based on the results of this study, we conclude that an ASR-based interactive video speaking activity can be practically given to EFL students as a homework assignment, even in cases where pre-set curricula limit teacher freedom. Furthermore, we found that although students were generally favorable towards the practice and it did give them more confidence in speaking, they tended to over-focus on pronunciation, despite the fact that the purpose of the activity was to focus on giving opinions and responding to particular interrogatives. The most likely reason for this was the type of feedback (i.e., a full textual representation of the ASR transcription), and therefore, future studies should consider providing different feedback when the purpose is not to improve pronunciation, specifically.

About the Authors

Sachiko Nakamura is a senior assistant professor at the Institute for Excellence in Higher Education at Tohoku University. Her primary interests revolve around TESOL with a focus on speaking practice, automatization, and the use of multimedia in EFL learning. Currently, she is particularly enthusiastic about the creation of speaking practice activities and methods. ORCID ID: 0009-0004-1743-9084

Ryan Spring is an associate professor in the Institute for Excellence in Higher Education at Tohoku University. His research interests include applications of cognitive linguistics to second language acquisition, objectively measuring speaking and writing, and the use of multimedia in EFL teaching. He currently serves as the vice-president of the Association for Teaching English through Multimedia. ORCID ID: 0000-0003-4810-9825

Shizuka Sakurai is an associate professor at the Institute for Excellence in Higher Education at Tohoku University in Japan. Her research centers on EFL academic listening and the creation of listening materials utilizing AI technology. Presently, her focus lies in exploring pedagogical methods to improve note-taking skills during lecture listening. ORCID ID: 0009-0001-9061-1942

To Cite this Article

References

Ahn, T. Y., & Lee, S. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, *47*(4), 778–786. https://doi.org/10.1111/bjet.12354

Al-Murtadha, M. (2019). Enhancing EFL learners' willingness to communicate with visualization and goal-setting activities. *TESOL Quarterly, 53*(1), 133-157. https://doi.org/10.1002/tesq.474

Ayedoun, E., Hayashi, Y., & Seta, K. (2015). A conversational agent to encourage willingness to communicate in the context of English as a foreign language. *Procedia Computer Science*, 60, 1433–1442. https://doi.org/10.1016/j.procs.2015.08.219

Beckett, G.H. (2006). Project-based second and foreign language education: theory, research and practice. In G.H. Beckett & P.C. Miller (Eds.), *Project-Based Second and Foreign Language Education: Past, Present and Future* (pp.3–18). Information Age.

Bozorgian, H., & Shamsi, E. (2020). Computer-assisted pronunciation training on Iranian EFL learners' use of suprasegmental features: A case study. *Computer-Assisted Language Learning Electronic Journal*, *21*(1), 93–113. http://callej.org/journal/21-2/Bozorgian-Shamsi2020.pdf

Brantmeier, C. (2005). Nonlinguistic variables in advanced second language reading: Learners' self-assessment and enjoyment. *Foreign Language Annals*, *38*(4), 494–504. https://doi.org/10.1111/j.1944-9720.2005.tb02516.x

Bueno-Alastuey, M. C. (2011). Perceived benefits and drawbacks of synchronous voice-based computer-mediated communication in the foreign language classroom. *Computer Assisted Language Learning*, *24*(5), 419–432. https://doi.org/10.1080/09588221.2011.574639

Cutrone, P. (2009). Overcoming Japanese EFL learners' fear of speaking. *Language studies working papers*, *1*(1), 55–63.

Cutrone, P., & Beh, S. (2018). Investigating the effects of task-based language teaching on Japanese EFL learners' willingness to communicate. *The Journal of Asia TEFL, 15*(3), 566–589. http://dx.doi.org/10.18823/asiatefl.2018.15.3.566

Dewaele, J. M., & Alfawzan, M. (2018). Does the effect of enjoyment outweigh that of anxiety in foreign language performance? *Studies in second language learning and teaching*, *8*(1), 21–45. https://doi.org/10.14746/ssllt.2018.8.1.2

Evers, K., & Chen, S. (2022). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning, 35*(8), 1869–1889. https://doi.org/10.1080/09588221.2020.1839504

Fujita, R. (2019). How do TV-drama-based materials affect the listening abilities of EFL learners with different proficiency levels? *ATEM Journal: Teaching English through Multimedia, 24*, 17–30.

Fushino, K. (2010). Causal relationships between communication confidence, beliefs about group work, and willingness to communicate in foreign language group work. *TESOL quarterly*, 44(4), 700-724. https://doi.org/10.5054/tq.2010.235993

Guskaroska, A. (2020). ASR-dictation on smartphones for vowel pronunciation practice. *Journal of Contemporary Philology*, *3*(2), 45–61. https://doi.org/10.37834/JCP2020045g

Harumi, S. (2011). Classroom silence: Voices from Japanese EFL learners. *ELT journal*, 65(3), 260–269. https://doi.org/10.1093/elt/ccq046

Inceoglu, S., Lim, H., & Chen, W. H. (2020). ASR for EFL pronunciation practice: segmental development and learners' beliefs. *Journal of Asia TEFL*, *17*(3), 824–840. http://dx.doi.org/10.18823/asiatefl.2020.17.3.5.824

Ito, T. (2022). Effects of general trust as a personality trait on willingness to communicate in a second language. *Personality and Individual Differences, 185*, https://doi.org/10.1016/j.paid.2021.111286

Kadoyama, T. (2008). Teaching communication through the use of films. *ARELE: Annual Review of English Language Education in Japan*, *19*, 243–252. https://doi.org/10.20581/arele.19.0_243

Kang, E. (2019). Enhancing Korean EFL learners' vocabulary learning and listening comprehension through video captions. *STEM Journal*, *20*(2), 91–108. http://dx.doi.org/10.16875/stem.2019.20.2.91

Kato, F. (2020). *Strategies for growing and enhancing university-level Japanese programs.* Routledge.

Kato, F., Spring, R., & Mori, C. (2023). Incorporating project-based language learning into distance learning: Creating a homepage during computer-mediated learning sessions. *Language Teaching Research, 27*(3), 621–641. https://doi.org/10.1177/1362168820954454

Kitaoka, K. (2023). An empirical study of the effect of using music on EFL students' motivation, willingness to communicate, and shyness. *The Society for Teaching English through Media, 24*(1), 55–70. https://doi.org/10.16875/stem.2023.24.1.55

Kondo, A. (2018). Eiga wo shiyou shita shidou ni yoru nihonjin daigakusei eno eigogakushu ni kakawaru doukizuke eno eikyou [The effect on the motivation related to English learning of Japanese university students instructed with fils]. *ATEM Journal: Teaching English through Multimedia*, *23*, 17–30.

Krause, K. L., & Coates, H. (2008). Students' engagement in first-year university. *Assessment & Evaluation in Higher Education*, *33*(5), 493–505. https://doi.org/10.1080/02602930701698892

Liu, M. (2006). Anxiety in Chinese EFL students at different proficiency levels. *System*, *34*(3), 301–316. https://doi.org/10.1016/j.system.2006.04.004

Loi, N. V. & Hang, C. T. T. (2021). Integrating project work into English proficiency courses for pre-service teachers. *Teaching English as a Second Language Electronic Journal (TESL-EJ), 25*(3). https://tesl-ej.org/pdf/ej99/a12.pdf

Maftoon, P., & Ziafar, M. (2013). Effective factors in interactions within Japanese EFL classrooms. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *86*(2), 74–79. https://doi.org/10.1080/00098655.2012.748641

Matsuoka, R., Matsumoto, K., Poole, G., & Matsuoka, M. (2014). Japanese university students' willingness to communicate in English: The serendipitous effect of oral presentations. *Journal of Pan-Pacific Association of Applied Linguistics, 18*(1), 193–218.

Matsuoka, R., & Rahimi, A. (2010). The positive effect of conference participation on reducing L2 communication apprehension. *Procedia-Social and Behavioral Sciences, 9*, 1845–1854. https://doi.org/10.1016/j.sbspro.2010.12.412

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, *51*(3), 617–637. https://doi.org/10.1111/flan.12348

Munro, M. J. (2010). Intelligibility: Buzzword or buzzworthy? In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd pronunciation in second language learning and teaching conference*, (pp. 7–16). Iowa State University. https://www.academia.edu/12031228/Proccedings_of_the_2nd_Pronunciation_in_Second_Language_Learning 184 _and_Teaching_Conference

Nakamura, S., & Spring, R. (2020). How watching subtitled YouTube videos can affect EFL listening and reading abilities. *ATEM Journal: Teaching English through Multimedia, 25*, 3-16.

Nauffal, D. I. (2012). Assessment of student engagement: An analysis of trends. *Tertiary Education and Management*, *18*(2), 171–191. https://doi.org/10.1080/13583883.2012.656696

Nazara, S. (2011). Students' perception on EFL speaking skill development. *Journal of English Teaching, 1*(1), 28–43. https://doi.org/10.33541/jet.v1i1.50

Ng, P. C. L. (2009). The Power of Problem-based Learning (PBL) in the EFL classroom. *Polyglossia, 16*, 41–48.

Ngo, T. T., Chen, H. H., & Lai, K. K. (2023). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL,* 1-18. https://doi.org/10.1017/S0958344023000113

Nishio, T., Fujikake, C., & Osawa, M. (2020). Language learning motivation in collaborative online international learning: an activity theory analysis. *Journal of Virtual Exchange*, *3*, 27-47. https://doi.org/10.21827/jve.3.35780

O'Dowd, R. (2018). From telecollaboration to virtual exchange: State-of-the-art and the role of UNICollaboration in moving forward. *Journal of Virtual Exchange*, *1*, 1–23. https://doi.org/10.14705/rpnet.2018.jve.1

Osterman, G. L. (2014). Experiences of Japanese university students' willingness to speak English in class: A multiple case study. *SAGE Open, 4*(3), 2158244014543779. https://doi.org/10.1177/2158244014543779

Pawlak, M., & Mystkowska-Wiertelak, A. (2015). Investigating the dynamic nature of L2 willingness to communicate. *System, 50*, 1-9. https://doi.org/10.1016/j.system.2015.02.001

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079

Saito, K., Dewaele, J. M., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, *68*(3), 709–743.

Schultz, L. M. (2017). Affect with Chinese Learners of English: Enjoyment, Self-Perception, Self-Assessment, and Abilities across Levels of Language Learning 1. *Quarterly Journal of Chinese Studies*, *5*(2), 65–81.

Skehan, P. & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*(3), 85–211.

Spring, R. (2020a). Maximizing the benefits of video-creation PBLL in the EFL classroom: A preliminary analysis of factors associated with improvement in oral proficiency. *STEM Journal, 21*(4), 107–126. https://doi.org/10.16875/stem.2020.21.4.107.

Spring, R. (2020b). Using multimedia tools to objectively rate the pronunciation of L1 Japanese EFL learners. *ATEM Journal: Teaching English through Multimedia, 25*, 113-124.

Spring, R., & Tabuchi, R. (2021). Assessing the practicality of using an automatic speech recognition tool to teach English pronunciation online. *STEM Journal, 22*(2), 93–104. https://doi.org/10.16875/stem.2021.22.2.93

Spring, R., & Tabuchi, R. (2022). The role of ASR training in EFL pronunciation improvement: An in-depth look at the impact of treatment length and guided practice on specific pronunciation points. *Computer Assisted Language Learning Electronic Journal*, *23*(3), 163–185. http://callej.org/journal/23-3/Spring-Tabuchi2022.pdf

Stoller, F. (2006). Establishing a theoretical foundation for project-based learning in second and foreign language contexts. In G. Beckett & P. C. Miller (Eds.), *Project-based second and foreign language education: Past, present, and future* (pp. 19–40). Information Age Publishing.

Wang, H., & Crosthwaite, P. (2021). The affordances of WeChat voice messaging for Chinese EFL learners during private tutoring. *Computer Assisted Language Learning Electronic Journal, 22*(1), 230–253. http://callej.org/journal/22-1/WangCrosthwaite2021.pdf

Watanabe, M. (2013). Willingness to communicate and Japanese high school English learners. *JALT Journal, 35*(2), 153–172.

Wu, X., Liu, X., & Chen, L. (2022). Reducing EFL learners' error of sound deletion with ASR-based peer feedback. In W. Jia, Y. Tang, R. S. T. Lee, M. Herzog, H. Zhang, T. Hao, & T. Wang (Eds.) *Emerging technologies for education. SETE 2021* (pp. 178–189). Springer. https://doi.org/10.1007/978-3-030-92836-0_16

Yamaguchi, A. (2018). Toward successful multimedia learning: Cases of self-directed EFL learners. *Teaching English through movies: ATEM journal*, (23), 31–42.

Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language learning*, *54*(1), 119–152. https://doi.org/10.1111/j.1467-9922.2004.00250.x

Yashima, T., MacIntyre, P. D., & Ikeda, M. (2018). Situated willingness to communicate in an L2: Interplay of individual characteristics and context. *Language Teaching Research, 22*(1), 115–137. https://doi.org/10.1177/1362168816657851

Yasuda, T., & Nabei, L. (2018). Effects of coping strategies on language anxiety of Japanese EFL learners: Investigating willingness to communicate. *Journal of Language Teaching & Research*, 9(5), 905-915. http://dx.doi.org/10.17507/jltr.0905.03

Appendix

Survey Questions (translated from Japanese)

Pre-survey

1. Do you have confidence in your English speaking ability?

    (1) I don't have any confidence (2) I don't have much confidence (3) I can't say either (4) I have confidence (5) I have a lot of confidence

2. Do you like speaking English?

    (1) I hate speaking English very much (2) I hate speaking English (3) I can't say either (4) I like speaking English (5) I like speaking English very much

3. Do you feel hesitant (anxious, nervous, and shy) about speaking English in front of people?

    (1) I don't feel hesitant at all (2) I don't feel very hesitant (3) I can't say either (4) I feel hesitant (5) I feel very hesitant

4. What aspects of speaking English do you want to improve? (vocabulary, grammar accuracy, fluency & speed, pronunciation, confidence, contents, others)

Post-survey

1. Do you have confidence in your English speaking ability?

    (1) I don't have any confidence (2) I don't have much confidence (3) I can't say either (4) I have confidence (5) I have a lot of confidence

2. Do you like speaking English?

    (1) I hate speaking English very much (2) I hate speaking English (3) I can't say either (4) I like speaking English (5) I like speaking English very much

3. Do you feel hesitant (anxious, nervous, and shy) about speaking English in front of people?

    (1) I don't feel hesitant at all (2) I don't feel very hesitant (3) I can't say either (4) I feel hesitant (5) I feel very hesitant

4. What aspects of speaking English do you think improved by the assignment? (vocabulary, grammar accuracy, fluency & speed, pronunciation, confidence, contents)

5. Please write your comments about the speaking assignment if any.