ReCALL (2023), 35: 1, 74–88
doi:10.1017/S095834402200012X

RESEARCH ARTICLE

Review of research on applications of speech recognition technology to assist language learning

Rustam Shadiev
Nanjing Normal University, China (rustamsh@gmail.com)

Jiawen Liu
Nanjing Normal University, China (liujw9797@163.com)

Abstract

Speech recognition technology (SRT) is now widely used in education because of its potential to aid learning, particularly language learning. Nevertheless, SRT has received only limited attention in earlier review studies. The present research aimed to address this gap in the field. To this end, 26 articles published in SSCI journals between 2014 and 2020 were selected and reviewed with respect to domain and skills, technology and their application, participants and duration, measures, reported results, and advantages and disadvantages of SRT. The results showed that English received much more attention than any other language, and scholars mostly focused on facilitating pronunciation skills. Dragon Naturally Speaking and Google speech recognition were the most popular technologies, and their most frequent application was providing feedback. According to the results, college students were involved in research more than any other group, most studies were carried out for less than one month, and most scholars administered a questionnaire or pre-/posttest to collect the data. Positive results related to gains in proficiency and student perceptions of SRT were identified. The study revealed that improved affective factors and enhanced language skills were advantages, whereas a low accuracy rate and insufficiency (i.e. lack of some useful features to support learning efficiently) of SRT were disadvantages. Based on the results, the study puts forward several implications and suggestions for educators and researchers in the field.

Keywords: review; speech recognition technology; language learning

1. Introduction

Communication plays an important role in language learning. It incites reflection and helps learners become included in a conversation. Therefore, there is a demand for language learners to communicate and to practice communication skills, such as speaking and pronunciation. However, some constraints exist for learners to communicate and practice their communication skills in language learning classrooms, hampering the improvement of communicative performance. For example, language learning class time is limited, so students cannot spend much time on communication and drills using the target language. In addition, teachers who need to assess the speaking abilities of learners and offer feedback are usually not able to provide feedback to each individual learner given high instructor–student ratios (Ehsani & Knodt, 1998; Oh & Song, 2021).

Scholars suggest that this issue can be addressed by using speech recognition technology (SRT) (Ehsani & Knodt, 1998; Oh & Song, 2021; Shadiev, Hwang, Chen & Huang, 2014; Shadiev, Hwang, Huang & Liu, 2016). In the process of speech recognition, SRT receives the speaker's verbal input,

Cite this article: Shadiev, R. & Liu, J. (2023). Review of research on applications of speech recognition technology to assist language learning. *ReCALL* 35(1): 74–88. https://doi.org/10.1017/S095834402200012X

analyzes it, and then generates an output, for example, in the form of a text (McKechnie *et al.*, 2018; Radha & Vimala, 2012). In the past two decades, SRT has made remarkable progress (Oh & Song, 2021; Shadiev, Wang, Wu & Huang, 2021; Xiao & Park, 2021). Major advances made during this period include the maturity and continuous improvement of the hidden Markov modeling (HMM) approach, which became the mainstream SRT; more attention has also been given to research on knowledge-based SRT and to artificial neural networks in SRT. These advances have contributed to making SRT more sophisticated; for example, it has the ability to adapt to different speakers, meaning new users do not need to train on SRT (i.e. a user reads text into the system and it analyzes the verbal input to fine-tune the speech recognition) to recognize continuous speech and to constantly improve the recognition rate in use.

SRT has been employed in language learning research to facilitate skills such as pronunciation (Ahn & Lee, 2016), listening (Mirzaei, Meshgi, Akita & Kawahara, 2017), writing (Arcon, Klein & Dombroski, 2017), grammar (Bodnar, Cucchiarini, de Vries, Strik & van Hout, 2017), and vocabulary (Cavus & Ibrahim, 2017). The results of related studies were mostly positive (e.g. skills were facilitated), and learners showed great interest and developed a positive attitude toward using SRT (Shadiev, Sun & Huang, 2019; Shadiev, Wu, Sun & Huang, 2018). Scholars have reported that SRT offers more opportunities to communicate and practice language skills, and learners receive timely corrective feedback for further improvement of their communicative abilities. For example, language learners in Ahn and Lee (2016) used SRT for self-regulated speaking practice. Learners spoke in the target language, and SRT provided immediate feedback to spoken utterances. Based on the feedback, the learners became aware of their errors and modified their speech. Shadiev, Huang and Hwang (2017) applied SRT to lectures in English as a foreign language. The system received speech input from the instructor and generated texts that were synchronously shown to nonnative-English-speaking students. Students were able to listen to the instructor and simultaneously read generated texts. This method enhanced the comprehension of the lecture content of students, especially those with low language abilities. Texts were useful for students to follow the instructor's speech and to understand lecture content (Shadiev *et al.*, 2014).

Several review studies on SRT have been carried out (see Appendix A in the supplementary material). Scholars have suggested that SRT facilitates language learning. Some of these studies focus on applications of the technology to education and others on technology development. For example, Ehsani and Knodt (1998) introduced the principles and components of SRT. In addition, the performance and implementation of technology were explained, and a number of applications were evaluated. Ehsani and Knodt (1998) proposed current and future trends in computer-assisted language learning (CALL) related to SRT. McKechnie *et al.* (2018) reviewed studies published between 2007 and 2016 and focused on a specific group of learners (e.g. children with speech sound disorders). Scholars reviewed the current state of the field and evaluated the quality of current SRT. McKechnie *et al.* (2018) identified 18 tools, whose accuracy rate was less than 80% (i.e. percent agreement between SRT and human judgment when used for evaluating words containing mispronunciations). They argued that SRTs have been used effectively to analyze foreign language pronunciation (e.g. phoneme and prosodic) in children. Radha and Vimala (2012) provided an overview of SRT and its development between 2003 and 2010. In addition, they explained a variety of speech feature extraction techniques and speech recognition approaches. Finally, the performance evaluation measures available for SRTs were reviewed. According to Radha and Vimala (2012), the Mel frequency cepstral coefficient is the most frequently used feature extraction technique, and the HMM is the most popular recognition technique among scholars. Shadiev *et al.* (2014) reviewed articles on SRT published between 1999 and 2014 to understand how it has been used to support learning. In addition, scholars have aimed to analyze all research evidence to understand how SRT can enhance learning. According to Shadiev *et al.* (2014), Dragon Naturally Speaking, Windows Speech Recognition, and IBM ViaVoice were the most popular recognition tools. They were applied to aid the learning of

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

different types of learners (e.g. learners with cognitive and physical disabilities or foreign students) in different ways (e.g. to aid comprehension of lecture content in a foreign language or support fluent communication among learners in a cyber-classroom).

An analysis of the previous review studies (Appendix A) showed that most of them are outdated. One study reviewed articles published more than 20 years ago (Ehsani & Knodt, 1998), and the other covered articles published more than 10 years ago (Radha & Vimala, 2012). Technology is developing very fast, and there could be many changes in the field in just a few years. There is no doubt that the accuracy rate of current SRT is much better than that of SRT 10 years ago. Therefore, recent technologies have much more potential and a greater variety of applications to offer in the field of education. Furthermore, the above-mentioned review studies have different focuses. Some studies focused on applications of SRT to support education in general but not language learning specifically (Shadiev *et al.*, 2014). Ehsani and Knodt (1998) and Radha and Vimala (2012) focused on technical aspects of SRT, its approaches or performance evaluation techniques, which are different from its educational aspects. McKechnie *et al.* (2018) targeted a very specific group of learners (e.g. children with speech sound disorders), and thus their findings may be of limited interest to those who target different groups of learners.

Therefore, a gap in the literature exists. In particular, researchers and practitioners need updated views on modern SRTs and their applications to language learning, their domains, and the skills analyzed in recently published studies. Furthermore, the methods (e.g. participants, interventions, and measures) of previous studies and the reported results, including advantages and disadvantages of technology applications, remain unclear. The present review aims to fill this gap. To this end, the following research questions were addressed:

1. What were the domains and skills in the reviewed articles?
2. What technology did scholars use and what were their applications?
3. Who were the research participants and what was the duration of the interventions in the reviewed articles?
4. What measures did scholars use in the reviewed articles?
5. What results were reported of the reviewed articles?
6. What were the reported advantages and disadvantages of the reviewed studies?

2. Methodology

The methodology of this review study consists of three major steps: article search, article selection, and content analysis. Articles were searched based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Figure 1). PRISMA is an evidence-based set of items intended to assist authors in preparing and reporting a wide array of systematic reviews and meta-analyses. Articles were searched using the Web of Science database. Scholars suggest that this database "is one of the most extensive, popular and relevant research databases for the academic community" (Caseiro & Santos, 2018: 8). In this study, we applied the following keywords in combination to search research articles: speech, voice, recognition, learning, instruction, and education. We selected these keywords because they were frequently used in related review studies (McKechnie *et al.*, 2018; Shadiev *et al.*, 2014).

Researchers screened titles and abstracts of articles and selected those that matched the following criteria: (1) articles published from 2014 to 2020, (2) articles published in journals related to education and educational research and indexed by the Social Sciences Citation Index (SSCI), (3) articles reporting research on applications of SRT to assist language learning, and (4) articles published as full texts and in English. According to Shadiev *et al.* (2021: 5), "the SSCI is an important channel with high authority for journal retrieval and paper references in the field of social sciences." In other words, research articles published by SSCI journals have higher impacts in the field and are rigorously reviewed (Duman, Orhon & Gedik, 2015).

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

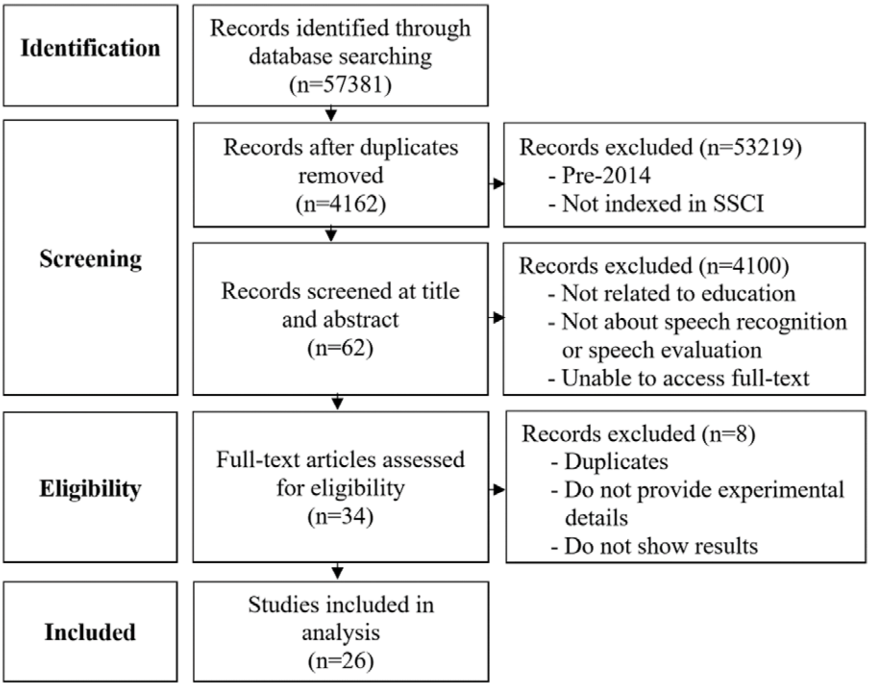
Identification
Records identified through database searching (n=57381)
Screening
Records after duplicates removed (n=4162)
Records excluded (n=53219)
- Pre-2014
- Not indexed in SSCI
Records screened at title and abstract (n=62)
Records excluded (n=4100)
- Not related to education
- Not about speech recognition or speech evaluation
- Unable to access full-text
Eligibility
Full-text articles assessed for eligibility (n=34)
Records excluded (n=8)
- Duplicates
- Do not provide experimental details
- Do not show results
Included
Studies included in analysis (n=26)

Figure 1. Systematic search flowchart

Eligibility assessment resulted in the selection of 26 articles (Appendix B). Content analysis was carried out using open coding (Creswell, 2014). The following coding schemes were derived from the content analysis: (1) domain and skills – language learning domain and skills that were assisted by SRT, (2) technology and its application – types of SRTs and their application to support language learning, (3) participants and duration – participants and duration of language learning activities, (4) measures – types of data collection instruments, (5) reported results – results that were reported in reviewed studies, and (6) advantages and disadvantages – advantages and disadvantages of SRT applications to support language learning. Two researchers conducted the article search, article selection, and content analysis. Each worked independently first to search articles, select appropriate articles, and analyze their content, and they then compared the obtained results. All discrepancies were resolved by discussion to reach a consensus.

3. Results and discussion

Information about publications in specific journals is included in Appendix C. Most studies on SRT were published in *Computer Assisted Language Learning* (CALL). This result is not surprising because CALL publishes articles on the use of computers to assist language learning, instruction, and assessment.

3.1 Domain and skills

The results related to domain and language skills are presented in Appendix D and Appendix E, respectively. The domain included language learning ($n = 23$), cross-cultural learning ($n = 2$), and distance learning ($n = 1$). According to the results, in the language learning domain, SRT was applied to English ($n = 17$), Dutch ($n = 3$), French ($n = 2$), and Spanish ($n = 1$) learning. In the cross-cultural learning domain, scholars used SRT to promote multilingual communication

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

and cross-cultural understanding of the participants. In the distance learning domain, SRT was used to support educational technology courses taken by students from China and Japan. The results showed that SRT was used to assist different language skills: pronunciation ($n = 15$), listening ($n = 5$), writing ($n = 3$), communication ($n = 3$), grammar ($n = 2$), word recognition ($n = 1$), and vocabulary ($n = 1$). Some studies focused on a single skill and others on more than one. For example, Tsai (2019) used SRT to improve student pronunciation only, whereas Cavus and Ibrahim (2017) focused on facilitating pronunciation as well as listening.

Some of our results echo those obtained in earlier studies. For example, Ehsani and Knodt (1998) and McKechnie *et al.* (2018) mentioned that SRT was applied to the domain of language learning. Studies by Shadiev *et al.* (2014) and Radha and Vimala (2012) are not comparable because they reported applications of SRT to education in general. In contrast to previous studies, the findings of the present study revealed new domains, such as cross-cultural learning and distance learning. Perhaps due to advancements in SRT, educators and researchers have considered its applications to such new domains. For example, the students in Shadiev *et al.* (2018) represented 13 nationalities, and SRT combined with machine translation enabled them to communicate with each other in their native languages. A student from one nationality spoke in his or her native language, and a student from another nationality could understand speech content because the technology received speech input and translated it. In this way, the students exchanged culture-related information. Another difference between previous studies and the present research is that none of the earlier studies considered any skills that could be supported by SRT.

Based on our results, we suggest that SRT can be applied not only to assist language learning but also in other learning contexts, such as cross-cultural learning and educational technology learning. In the language learning domain, successful research projects were implemented with respect to English, Dutch, French, and Spanish. In addition, SRT was employed to assist cross-cultural learning and educational technology learning of distant students. Educators and researchers who are planning to use SRT in these domains may find useful information from the reviewed studies to guide their teaching and research.

The results also showed certain domains and languages that received more attention from researchers. That is, language learning was the main domain, and English was the dominant language in the reviewed studies. English was the dominant language in the reviewed articles, as it is considered one of the most commonly spoken languages (Cavus & Ibrahim, 2017). This finding is in line with those obtained in other review studies (McKechnie *et al.*, 2018; Shadiev & Yang, 2020). For example, McKechnie *et al.* (2018) found that SRTs were used to support the learning of 13 different languages, most commonly English.

Our results demonstrated that SRT could be applied to facilitate skills such as pronunciation, listening, writing, grammar, and word recognition. Therefore, educators and researchers who apply SRT may focus on these skills. The results suggest that the most commonly supported skill was pronunciation and that the remaining skills were targeted least frequently. The reason is that SRT is capable of receiving voice input, analyzing it, and providing corrective feedback so that language learners can improve their pronunciation (Cavus & Ibrahim, 2017; Tsai, 2019). In terms of improving listening, SRT can generate texts from the speech input of the instructor, and language learners are able to listen to the instructor and read transcriptions simultaneously to confirm unfamiliar vocabulary or what they misheard from the speech (Mirzaei *et al.*, 2017; Shadiev *et al.*, 2017). Furthermore, Shadiev *et al.* (2019), Shadiev *et al.* (2018), and Yueh, Lin, Liu, Shoji and Minoh (2014) proposed extending the speech-to-text recognition process with computer-aided translation to assist communication among learners without a common language. Learners spoke to the system, generated translated texts from speech input, and translated texts were shown to language partners. For writing, SRT can help learners demonstrate compositional ability independent of transcription; for example, technology mitigates challenges with spelling and allows students to compose written content using their rich oral vocabulary

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

(Arcon *et al.*, 2017; Baker, 2017). For vocabulary, word recognition and grammar learning, learners had to speak vocabulary words and their order in a sentence to SRT so that it could analyze learners' speech and provide immediate corrective feedback, such as the correct pronunciation, their recognition, and order (Bodnar *et al.*, 2017; Cavus & Ibrahim, 2017; Matthews & O'Toole, 2015).

3.2 Technology and its application

Seven types of SRTs (Appendix F) and three different applications (Appendix G) were identified. Scholars used Dragon Naturally Speaking ($n = 4$), Google speech recognition ($n = 4$), Windows Speech Recognition ($n = 2$), automatic speech recognition (ASR)-based CALL system ($n = 1$), partial and synchronized captioning (PSC) ($n = 1$), and Julius ($n = 1$). The results show that SRTs were not specified in 13 studies. According to the data, SRT was used to provide feedback ($n = 19$), showing texts generated by technology ($n = 9$) and giving commands to the system ($n = 3$).

The results showed that the SRTs Dragon Naturally Speaking, Google speech recognition, Windows Speech Recognition, ASR-based CALL system, PSC, and Julius have been used to assist language learning. Most recognition technologies feature continuous speech recognition (i.e. when SRT transcribes naturally dictated speech) and thus were used in reviewed studies in writing practice, listening skills development, and pronunciation training. Therefore, they might be considered by educators and researchers in their future studies. Two recognition technologies, Dragon Naturally Speaking and Google speech recognition, were the most frequently used. One reason is because both Dragon Naturally Speaking and Google speech recognition are designed to recognize speech independent of the speaker (i.e. they are designed to respond to a word or phrase regardless of who speaks) and can be installed on mobile phones, tablet PCs, or desktop and laptop computers with either Microsoft Windows or Mac operating systems. Although Dragon Naturally Speaking is not free, it is mature and undergoes constant improvement. For this reason, Dragon Naturally Speaking is used mostly by professionals (e.g. doctors, lawyers, court stenographers). On the other hand, Google speech recognition is a cloud-based service that is available free of charge and features a high rate of accuracy. Therefore, Google speech recognition is mostly used by casual users. Microsoft Windows Speech Recognition is the third most popular technology. Windows Speech Recognition comes pre-built into the Windows operating system, and thus is free to anyone with a Windows PC but unavailable to those who use different computer operating systems.

In comparison to Dragon Naturally Speaking, Google speech recognition, and Windows Speech Recognition, SRTs such as the ASR-based CALL system, PSC, and Julius were developed by researchers for academic purposes only and therefore are not publicly available. These tools were used as an alternative to existing speech recognition methods. For the ASR-based CALL system, researchers judged learners' pronunciation by comparing HMM with built-in corpus approaches (Wang & Young, 2015). In PSC, the system transcribed speech input in the form of a set of words or phrases, which appeared on the screen in a one-by-one sequence synchronized with the utterance (Mirzaei *et al.*, 2017). Julius was implemented to control a camera and to access demographic information and learning statistics of students using voice commands in a distant learning environment (Yueh *et al.*, 2014). It should also be noted that in contrast to Dragon Naturally Speaking and Google speech recognition, Windows Speech Recognition, ASR-based CALL system, PSC, and Julius are speaker-dependent recognition systems; that is, they are trained by the individual who will be using the system. Furthermore, they are only available through desktop and laptop computers.

Some of our results are similar to those that were reported in other review studies. For example, Shadiev *et al.* (2014) also found that Dragon Naturally Speaking was among the most frequently used SRT. Shadiev *et al.* (2014) also mentioned two other popular systems, Windows Speech

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

Recognition and IBM ViaVoice. Although IBM ViaVoice was popular a decade ago, it has since been discontinued and thus is no longer used. On the other hand, Google speech recognition was not mentioned in Shadiev *et al.* (2014) because Google's first effort at speech recognition came only recently. It was not widespread among educators and researchers before, and therefore previous review studies did not mention it.

The results showed that SRTs were not specified in 13 studies. That is, half of the reviewed studies simply did not specify their technology, which is problematic. We suggest that scholars should specify their technology in order to disclose what SRTs were used, their applications to learning, and their impacts (i.e. either positive or negative) on learning outcomes. According to the results, providing feedback was the most frequent application. A learner spoke a word or read a sentence, and the technology provided feedback so that a learner could see where and what kind of mistakes she made. For example, students in Ahn and Lee (2016) interacted with the system to identify errors in their speech, and students in de Vries, Cucchiarini, Bodnar, Strik and van Hout (2015) received different types of corrective feedback (e.g. implicitness and explicitness, reformulations or prompts to improve pronunciation). In showing texts generated by the technology applications, a lecturer or student spoke to the system, which then generated texts from the voice input and showed them a whiteboard or computer screen (Shadiev *et al.*, 2017). Such applications are popular in lectures for enhancing students' comprehension of lecture content, especially when lectures are delivered in a foreign language as a medium of instruction (e.g. English as a medium of instruction) (Mirzaei *et al.*, 2017). To give commands to the system applications, a lecturer or student spoke to the system to execute a command. For example, students in Arcon *et al.* (2017) composed and modified their essays by executing specific word-processing editing commands through a speech recognition interface. Such an approach was useful in improving speaking skills. Previous related studies reported similar applications of SRT. For example, the most popular were providing feedback (Ehsani & Knodt, 1998; McKechnie *et al.*, 2018) or dictating (Shadiev *et al.*, 2014). Other related review studies focused on the technical aspect of technology instead of the educational aspects (Ehsani & Knodt, 1998; Radha & Vimala, 2012).

3.3 Participants and intervention duration

Identifying the educational level and size of a population is essential for researchers and language teachers in their design of future research. Studies usually cover diverse contexts, target learners at different educational levels, and involve different numbers of learners. Approaches that were found to be successful for learners of one education level may not be as effective for learners of another education level. For example, Arcon *et al.* (2017) carried out their research with primary school students in the context of persuasive writing. Arcon *et al.* (2017) warned that their results could not be generalized to other populations and contexts; for example, written compositions made by adult learners may produce different patterns of results. It will also be inappropriate for young learners (e.g. preschool or primary school students) to participate in language learning activities designed for adult learners (e.g. university students), as the activities might be more complex and difficult. The number of participants can give confidence in the obtained findings as well as recommendations to researchers and educators about how many participants should be involved in specific learning activities. Therefore, without knowing what kind of learners or how many of them were involved in a particular study, it is difficult to develop appropriate and effective learning activities supported by technology.

The educational level and the number of participants are shown in Appendix H and Appendix I, respectively. College ($n = 20$), elementary school ($n = 4$), junior high school ($n = 2$), and preschool ($n = 1$) students were involved in the reviewed studies. In two studies, the educational level of participants was not specified. The number of participants in the reviewed studies ranged from 10 to 341. There were 12 studies in which the number of participants was less than

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

30, 10 studies with a number between 30 and 60, and four studies in which more than 60 participants were involved.

Intervention duration is shown in Appendix J. The results demonstrate that there is a range of intervention durations: less than one hour ($n = 5$), less than one day ($n = 2$), less than one week ($n = 1$), less than one month ($n = 9$), more than one month ($n = 6$), or not specified at all ($n = 3$). Some examples of short-term and long-term interventions are provided. For example, in Arcon *et al.* (2017), students trained on SRT and composed persuasive texts on assigned topics. They did so in three sessions, each lasting for approximately 20 minutes. Students in Dalim, Sunar, Dey and Billinghurst (2020) interacted with the system to improve their language skills, and the entire session lasted between 30 and 35 minutes. Hsu (2016) arranged a three-month self-regulated speech-recognition-based course, and Wang and Young (2015) asked participants to practice their language skills twice a week and to complete their learning content in eight weeks.

According to the results, college students made up the majority of the participants. This is because college students are a large group of language learners, and it is convenient for researchers to conduct experiments with such learners. Primary and secondary school students were involved in research to a lesser extent. This is because this group of learners has less experience in both language learning and technology usage. Our results are different from those obtained by McKechnie *et al.* (2018). In their review study, scholars mostly involved participants up to 16 years (i.e. secondary or high school students). Only one study had participants up to 21 years (i.e. college students). This is perhaps because McKechnie *et al.* (2018) focused on a specific group of learners (e.g. children with speech sound disorders).

In the face of such research information, we believe that younger groups will be more involved in research and that with the continuous development of technology, more suitable SRT for young children will be developed. Thus, it is necessary for researchers to take other groups of learners into consideration, provide them with necessary and easy-to-use software, and teach them learning strategies to apply technology to language learning more efficiently (e.g. Shadiev *et al.*, 2016).

The number of studies with small and large samples (i.e. small and large numbers of participants) was almost the same. Although small samples can reduce the workload of the experiment, we found that studies using small sample populations acknowledged it as a limitation and noted that their results could not be generalized beyond that particular group involved. For example, 21 students participated in the study of Arcon *et al.* (2017), and 22 participants were involved in Baker (2017). Some scholars argued that they aimed for a large group of participants to obtain a sample that is likely to be representative of population values. To make their conclusions robust and valid, Hsu (2016) had 341 students participate in the experiment, and it is worth mentioning that in Ahn and Lee (2016), there were more than 1,000 participants.

Regarding duration, eight studies were relatively short (less than one week). It is possible that such a short duration may result in frustration with the learning experience involving technology and negatively affect the learning outcomes. For example, students need to train on SRT and learn how to use it to learn more efficiently. Scholars have suggested that at least two weeks are needed for this process (Shadiev *et al.*, 2019; Shadiev *et al.*, 2018). Therefore, educators and researchers need to consider having their intervention for a longer period of time, with at least a few weeks dedicated to technology training.

3.4 Measure

The primary goal of including this dimension was to explain the overall data collection preference and the tendency of the research. The measures (i.e. data collections instruments) used are shown in Appendix K. They are a questionnaire ($n = 24$); pre-/posttest ($n = 14$); interviews ($n = 11$); content of reflective notes, created texts, and think-aloud protocols ($n = 3$); learning logs ($n = 3$); EEG recordings ($n = 2$); fieldwork method ($n = 1$); eye tracking ($n = 1$); task analysis

($n = 1$); language learning logs ($n = 1$); and usability review ($n = 1$). Scholars used questionnaires to investigate participants' perceptions ($n = 17$) of SRT usefulness or student cognitive load ($n = 3$). For example, Bodnar *et al.* (2017) measured student perceptions of corrective feedback based on SRT. Pre-/posttest was administered in reviewed studies to measure student proficiency before and after the intervention ($n = 9$). For example, Mirzaei *et al.* (2017) employed comprehension tests to measure the effects of the intervention. In the interviews, scholars mostly explored student learning experiences ($n = 8$). For example, teachers were interviewed in Baker (2017) to gather information on how students were learning with the support of technology.

During the application of SRT, various measurements provide not only scientific evidence of the effectiveness of the intervention but also useful feedback on the technology so that it can be improved. We can also see that researchers pay attention to both learning outcomes and learning experiences in the technology use process. Therefore, educators and researchers may consider these aspects in future planned studies.

Research methods can be divided into qualitative and quantitative research. Scholars used pre-/posttest design to investigate the effectiveness of applying SRT to assist learning by comparing learning outcomes of the control and experimental groups. For example, in quantitative research, Cavus and Ibrahim (2017) administered pre- and posttests to explore the improvement of language skills. Scholars compared the results of the tests between two groups to demonstrate the effects of their intervention. Questionnaires were mostly used to measure participants' perceptions of the intervention. For example, in qualitative research, Ahn and Lee (2016) aimed to measure the perceptions and attitudes of their students toward the application of speech-to-tech recognition in terms of its design, convenience, and efficacy. Interviews are mostly administered by researchers to obtain insights into participants' learning experiences with the technology. For example, in qualitative research, Liakin, Cardoso and Liakina (2017) interviewed the participants to explore their learning experience with SRT during language learning. Some scholars even collected the data using mixed methods (e.g. Arcon *et al.*, 2017; de Vries *et al.*, 2015; Liakin *et al.*, 2017). This approach is common in research because it enables the triangulation of data from different sources to make results and conclusions reliable and robust (Shadiev *et al.*, 2017).

Log data were collected by scholars to objectively reflect on the state of students' use of technology. When such data are combined with an interview or questionnaire data, the results become more convincing because researchers are able to triangulate data collected from various sources. Interestingly, among these measures, very few studies have employed physiological measurement tools to obtain physiological data. Such data objectively reflect the physiological state of learners and enable us to know whether SRT is effective in improving physiological learning outcomes (e.g. learning or visual attention) and to make appropriate changes in the intervention design.

3.5 Results in reviewed studies

The results in the reviewed studies are shown in Appendix L. The results can be divided into five areas: (1) gains in proficiency – there were changes in certain language skills after the intervention; (2) perceptions – students' perceptions of the intervention; (3) questions, suggestions, or approaches – some questions were raised after the intervention and suggestions or approaches were proposed; (4) system design – results concerned the system design; and (5) learning logs – records of system usage for language learning. Proficiency gains in terms of writing (Baker, 2017; Haug & Klein, 2018), pronunciation (McCrocklin, 2016; Tsai, 2019), and grammar (de Vries *et al.*, 2015) were reported in reviewed studies. For example, Haug and Klein (2018) argued that SRT played an important role in learning writing strategies. Haug and Klein (2018: 1) found that using the technology resulted in "large gains in text quality, word count, and variety of argument moves."

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

The effects of SRT applications are different depending on students' abilities. For example, the technology significantly improved the writing outcomes of students who struggled with writing (MacArthur & Cavalier, 2004) but it did not help students with better writing abilities (Arcon *et al.*, 2017). In another study by Shadiev *et al.* (2017), texts generated by SRT from the instructor's speech during lectures in a foreign language were very useful for low language ability students. They listened to the instructor and read texts to better comprehend lecture content. However, the same texts were not useful and were even distracting for high language ability students because listening to the instructor was enough for them (Shadiev *et al.*, 2017). This issue should be considered in future studies – for example, how technology can help students with lower language abilities without hurting the learning of students with greater abilities.

In pronunciation learning, SRT helped improve pronunciation significantly, not only in terms of the accuracy of pronunciation but also in awareness of pronunciation errors and linguistic features (Tsai, 2019). SRT is a potential tool for practicing oral grammar, as it detects errors in the speech of learners, and then learners receive immediate corrective feedback (de Vries *et al.*, 2015). SRT feedback can be of different forms, such as texts generated by technology, graphs with sound waves, or echoed speech. Of course, more feedback and details on the mistakes made or specific instructions on how to improve them will be beneficial for improving learning outcomes. However, it is still not clear how to combine multiple feedback forms, what combination of feedback forms, and in what quantity feedback needs to be provided. Future studies may consider exploring this aspect.

Regarding perceptions, participants generally had positive perceptions of their experiences using SRT. SRT helped students practice speaking in private spaces (Ahn & Lee, 2016), enabled them to use the target language in situated contexts and assisted their understanding of language appropriateness (Wang & Young, 2014). In addition, students became aware of their errors and were able to modify their utterances (Liakin *et al.*, 2017). Most of this evidence is subjective in nature and prone to bias. Therefore, future studies may consider exploring learning experiences (cognitive load, learning attention, learning emotions, etc.) using objective approaches based on psychological data (EEG or eye tracking). For example, it is possible to know objectively and precisely what part of the learning content is more difficult or when learners are overloaded when learning a new language.

In the reviewed papers, some scholars posed questions and proposed suggestions or approaches. For example, Liakin *et al.* (2017) claimed that language learners do not have enough time and opportunities for pronunciation practice in language classrooms. To address this issue, they introduced SRT implemented on mobile devices so that learners can practice this skill for as long as they need, anytime and anywhere. Tsai (2019) reported that although SRT used for pronunciation learning was useful to provide feedback related to such linguistic features (e.g. pitch variation), some learners doubted the validity of its scoring system and complained about its feedback. That is, learners followed the model utterance regarding pausing, and their scores were lower than when they read without any pauses. Other learners complained that feedback from the system was not as clear as the feedback from the instructor or that no feedback on how to correct their production was provided. Therefore, researchers have suggested that instructors may integrate peer interaction activities into the learning process; such an approach may enable assistance from peers that cannot be provided by the software (e.g. some support or strategies to obtain better learning outcomes).

Results related to SRT design were positive. The SRT "interface appeared good and learners could find the areas of their interests easily" (Yu *et al.*, 2016: 997). Teachers in van Doremalen, Boves, Colpaert, Cucchiarini and Strik (2016) thought that SRT was easy to use and that students were satisfied with the tool. Scholars used learning logs to know how often and how long learners used SRT (Wang & Young, 2014). In addition, we found that scholars do not get the most out of their systems. For example, it is possible to add some features that can record the learning behavior of learners (e.g. what functions they used and how frequently).

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

Learning behavior analysis did not receive much attention in the reviewed studies, but it can help reveal behavior patterns, behavior habits, and behavior rules to promote our understanding and optimization of language learning through SRTs. Therefore, scholars may consider exploring learning behavior in the future. We could not find learning analytics and big data approaches in the reviewed studies. It is possible to employ such approaches to further optimize language learning experiences; they can help instructors better support struggling language learners by personalizing the learning process and adapting their teaching when necessary.

Finally, scholars may consider combining SRT with other tools according to different learning purposes. For example, in the process of assisting word recognition and pronunciation learning, dictionaries, voice recording and playback, as well as other functions that may be frequently used by learners, can be integrated. In writing training, SRT can be combined with a proofreading function so that feedback on grammar, spelling, and punctuation can also be provided.

Most previous review studies also reported similar results, and most of them were positive. For example, Ehsani and Knodt (1998), McKechnie *et al.* (2018), and Shadiev *et al.* (2014) found an improvement in students' language learning. However, the results of the reviewed studies did cover other dimensions – for example, cross-cultural learning and learning system design. Other review studies, such as that by Radha and Vimala (2012), did not consider such results at all.

3.6 Advantages and disadvantages of using speech recognition technology

The reported advantages and disadvantages are shown in Appendix M. The advantages can be divided into 10 different aspects: improving affective factors ($n = 18$), enhancing language skills ($n = 14$), promoting interaction ($n = 5$), creating a self-paced learning environment and improving autonomy ($n = 5$), increasing learning involvement ($n = 3$), self-monitoring errors ($n = 2$), enhancing intercultural sensitivity ($n = 2$), supporting learner differences ($n = 1$), reducing task completion time ($n = 1$), and developing awareness of intelligibility ($n = 1$). Scholars have reported that using SRT is beneficial for improving affective factors during language learning by, for example, decreasing frustration and anxiety or increasing enjoyment, confidence, and motivation. Dalim *et al.* (2020) received many positive comments regarding the system – for example, using the system was enjoyable and exciting. Mroz (2018) reported affective gains, such as an increase in learning motivation and self-confidence, as well as learners' willingness to communicate, particularly for apprehensive learners. Language skills development – for example, pronunciation (Cavus & Ibrahim, 2017), listening (Mirzaei *et al.*, 2017), or writing (Arcon *et al.*, 2017) – was also reported. Cavus and Ibrahim (2017) claimed that student skills (i.e. listening, vocabulary, or pronunciation) showed significant improvements as a result of using SRT. In terms of promoting interaction, SRT enabled a novel type of interaction between language learners and technology. Students were able to speak to the system and be provided with immediate feedback. Students in Ahn and Lee (2016) liked the system, as they found speaking practice with SRT to be more motivating, enjoyable, and interactive; students felt like conversing with a real person.

Some disadvantages were also reported: the low accuracy rate of the system ($n = 9$), its insufficiency (i.e. SRT lacked some useful features to support learning efficiently) ($n = 8$), the system placing a burden on some students ($n = 3$), and being time-consuming to use ($n = 1$). Disadvantages were not specified in nine studies. Scholars have reported that the system generates texts from speech input with errors (Arcon *et al.*, 2017). Usually, the accuracy rate decreases in a noisy environment when several people speak at the same time (Dalim *et al.*, 2020) or speech input is lengthy (Mroz, 2018). To combat these issues, Shadiev *et al.* (2017) and Yueh *et al.* (2014) provided strategies, such as training on the technology beforehand or using it in a quiet environment. More useful strategies to use SRT for learning and achieving higher accuracy rates were proposed by Shadiev *et al.* (2016). In terms of insufficiency, some tools lacked features that could efficiently support learning. For example, Wang and Young (2014) found that implicit feedback showing learners' pronunciation scores and audio waveforms was insufficient to

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

recognize their pronunciation errors. Therefore, it was suggested that feedback be provided in an "explicit format of immediate audio replay (recast) with a textual description" (Wang & Young, 2014: 230). Scholars also warned that recognition tools might place a burden on students. For example, students felt frustrated by the low accuracy rate of SRT as a result of their accent when they practiced writing skills (Arcon *et al.*, 2017) or pronunciation (McCrocklin, 2016), and they had to spend much time completing their tasks using SRT. To avoid this problem, students need to be trained on SRT beforehand, the instructor needs to provide students with necessary practice strategies, and certain SRTs should be used that are able to recognize speech in English with certain types of accent. Another example is that texts generated by SRT were distracting to students with high language abilities because they were able to understand the content without any support (Shadiev & Huang, 2020). Therefore, a more flexible approach to using SRT was proposed; learners may switch on displays with texts when they need them and switch them off when support is no longer required. It is possible that the quick conversion from speech to text may, to a certain extent, indulge some students' laziness. To avoid this, educators and researchers need to design learning activities in such a way that students have manageable assignments, understand the affordances of SRT for their learning, know what to do and how to do it, have a positive learning experience, are motivated, and can obtain assistance and guidance whenever needed.

Some results of the present study on advantages and disadvantages of using SRT are in line with the results from related review studies; for example, SRT is beneficial for enhancing language skills, or problems associated with accuracy rates of SRT have been reported previously (Ehsani & Knodt, 1998; McKechnie *et al.*, 2018; Shadiev *et al.*, 2014). However, reported advantages and disadvantages of SRT for language learning varied in the present review studies compared to earlier review studies. For example, we found that SRT was beneficial for improving affective factors or enhancing intercultural sensitivity. Reviewed studies also mentioned the insufficiency of SRT or that it placed a burden on language learners. Such variety of reported advantages and disadvantages can be accounted for the fact that earlier review studies focused either on SRT applications to support education in general but not on language learning specifically (Shadiev *et al.*, 2014) or on technical aspects of SRT instead of educational (Ehsani & Knodt, 1998; Radha & Vimala, 2012). For these reasons, our findings may be of greater interest to those who target language learning process supported by SRT.

4. Conclusion

The review study's results demonstrate that SRT has gained considerable attention in CALL research. The reported results were mostly positive and demonstrated that SRT has great potential to assist language learning. Based on our results, several suggestions for educators and researchers can be made. First, it is suggested that educators and researchers consider applying SRT to the field of language learning. It can assist learners not only in improving their pronunciation but also in writing, listening, and grammar learning.

It is also suggested that educators and researchers provide appropriate assistance to students when SRT is applied. The reason is that few students have experience with SRT and its applications. In addition, providing various support mechanisms, such as useful strategies, timely feedback, and guidance to learners when they use technology, is important. The greatest value of this technology applied to education is realized only when technological and educational strategies are combined. Learners need to be trained on SRT in advance. This will help them get better acquainted with SRT, know its strength and limitations, and then use it for language learning more effectively.

Educators and researchers may use various recognition technologies. Dragon Naturally Speaking and Google speech recognition were recognized as the most frequently used and mature technologies. Recognition technology can be used for different purposes, such as to provide

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

feedback on voice input so that language learners recognize their mistakes and correct them, to show texts generated by the technology to students during lectures in a foreign language so that they comprehend learning content better, or give commands to the system that were found to be useful for interaction practice.

If SRT is used by young learners, a friendlier and easy-to-use interface should be developed. Scholars may also consider different sample sizes and durations of studies based on their research purposes. However, a longer duration of the intervention is preferred, as longer exposure of students to technology enables them to first understand it and its applications better and utilize technology more efficiently. Various data collection instruments can be used. Scholars may consider a questionnaire, pre-/posttests, and interviews, as they were frequently used in reviewed studies. Scholars may also consider using several measures at the same time to increase the rigor and validity of the research. In addition, educators and researchers can use measures based on physiological data, as they are proven to be useful in research because of their objectivity.

Some disadvantages were acknowledged in the reviewed studies, with the most frequent being the recognition accuracy rate. Disadvantages along with their solutions need to be considered. For example, the usage of useful learning strategies (e.g. using technology in a quiet environment, adding vocabulary that is frequently mistranslated into the technology's database, or speaking clearly and loudly) (Ehsani & Knodt, 1998; McKechnie *et al*., 2018; Shadiev *et al*., 2014) can help address some of these issues.

There are also some future trends that can be considered by educators and researchers in the field. Apart from the language learning domain and pronunciation skills, we found that scholars applied SRT to other domains and focused on other skills, such as cross-cultural learning courses and communication skills. However, these domains and skills have received little or no attention from researchers. Therefore, future studies may consider them to help increase the little existing knowledge in the field. We found that all reviewed studies focused on individual learning only – that is, when a learner learns alone. Educators and researchers may consider designing communicative learning activities supported by SRT (e.g. conversations, discussions, or meetings) in which learners are able to practice their communication skills by interacting with other learners. SRT can support such activities because it is speaker independent, smart, and features high accuracy rates (Shadiev *et al*., 2014).

The results demonstrated that various supporting mechanisms were used when SRT was applied, such as providing learners with effective strategies to use SRT, timely feedback, and guidance by the instructor. Future studies may consider exploring the role of such support and its effects on learning outcomes. Different applications of SRT or their extension should also be considered in the future to improve the benefits of SRT. For example, texts generated by SRT can be translated and shown to learners so that their language skills can be improved by using both transcripts and translations simultaneously. As most studies involved college students, another promising research direction is to involve learners from other educational levels. Furthermore, some of the reviewed studies did not explicitly state important information; for example, SRTs were not specified in 13 studies, and this issue should be considered by scholars in the future.

For SRT development, scholars need to consider how to make it smarter – that is, how to make the system able to understand the context in which the learning experience is taking place. When the system experiences ambiguous words, it should select the most appropriate one and the one that relates to the context. Developers also need to consider making SRT easier to use. This is especially important for young learners. If this group of learners will participate in SRT-related studies in the future, developers need to provide them with appropriate technology based on their educational level and experience in language learning and technology usage. Finally, the system should become more accurate. SRT has existed for a few decades, and scholars consistently claim that its accuracy rate is improving dramatically. However, even today, some authors still report issues associated with the recognition accuracy rate and that misrecognition negatively affects

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

learning. This is the biggest issue in the field of SRT-supported language learning. Therefore, researchers should focus on improving it so that it does not interfere with the learning process.

Supplementary material. To view supplementary material referred to in this article, please visit https://doi.org/10.1017/S095834402200012X

Author contributions. Jiawen Liu: Methodology, investigation, writing – original draft preparation; Rustam Shadiev: Conceptualization, writing – reviewing and editing, supervision.

Ethical statement and competing interests.

1. The dataset will be provided on request upon completion of this project.
2. The authors declare no competing interests.
3. No ethical issues exist in this systematic review study.

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

Arcon, N., Klein, P. D. & Dombroski, J. D. (2017) Effects of dictation, speech to text, and handwriting on the written composition of elementary school English language learners. *Reading & Writing Quarterly*, 33(6): 533–548. https://doi.org/10.1080/10573569.2016.1253513

Baker, E. A. (2017) Apps, iPads, and literacy: Examining the feasibility of speech recognition in a first-grade classroom. *Reading Research Quarterly*, 52(3): 291–310. https://doi.org/10.1002/rrq.170

Bodnar, S., Cucchiarini, C., de Vries, B. P., Strik, H. & van Hout, R. (2017) Learner affect in computerised L2 oral grammar practice with corrective feedback. *Computer Assisted Language Learning*, 30(3–4): 223–246. https://doi.org/10.1080/09588221.2017.1302964

Caseiro, N., & Santos, D. (Eds.). (2018). *Smart specialization strategies and the role of entrepreneurial universities*. Hershey, PA: IGI Global. Available from: https://www.igi-global.com/book/smart-specialization-strategies-role-entrepreneurial/197442

Cavus, N. & Ibrahim, D. (2017) Learning English using children's stories in mobile devices. *British Journal of Educational Technology*, 48(2): 625–641. https://doi.org/10.1111/bjet.12427

Creswell, J. W. (2014). *Educational research: Planning, conducting, and evaluating quantitative*. Boston, MA: Pearson Education.

Dalim, C. S. C., Sunar, M. S., Dey, A. & Billinghurst, M. (2020) Using augmented reality with speech input for non-native children's language learning. *International Journal of Human-Computer Studies*, 134: 44–64. https://doi.org/10.1016/j.ijhcs.2019.10.002

de Vries, B. P., Cucchiarini, C., Bodnar, S., Strik, H. & van Hout, R. (2015) Spoken grammar practice and feedback in an ASR-based call system. *Computer Assisted Language Learning*, 28(6): 550–576. https://doi.org/10.1080/09588221.2014.889713

Duman, G., Orhon, G. & Gedik, N. (2015) Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL*, 27(2): 197–216. https://doi.org/10.1017/S0958344014000287

Ehsani, F. & Knodt, E. (1998) Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. *Language Learning & Technology*, 2(1): 54–73.

Haug, K. N. & Klein, P. D. (2018) The effect of speech-to-text technology on learning a writing strategy. *Reading & Writing Quarterly*, 34(1): 47–62. https://doi.org/10.1080/10573569.2017.1326014

Hsu, L. (2016) An empirical examination of EFL learners' perceptual learning styles and acceptance of ASR-based computer-assisted pronunciation training. *Computer Assisted Language Learning*, 29(5): 881–900. https://doi.org/10.1080/09588221.2015.1069747

Liakin, D., Cardoso, W. & Liakina, N. (2017) Mobilizing instruction in a second-language context: Learners' perceptions of two speech technologies. *Languages*, 2(3): 1–21. https://doi.org/10.3390/languages2030011

MacArthur, C. A., & Cavalier, A. R. (2004). Dictation and speech recognition technology as test accommodations. *Exceptional Children*, 71(1), 43–58.

Matthews, J. & O'Toole, J. M. (2015) Investigating an innovative computer application to improve L2 word recognition from speech. *Computer Assisted Language Learning*, 28(4): 364–382. https://doi.org/10.1080/09588221.2013.864315

McCrocklin, S. M. (2016) Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, 57: 25–42. https://doi.org/10.1016/j.system.2015.12.013

McKechnie, J., Ahmed, B., Gutierrez-Osuna, R., Monroe, P., McCabe, P. & Ballard, K. J. (2018) Automated speech analysis tools for children's speech production: A systematic literature review. *International Journal of Speech-Language Pathology*, 20(6): 583–598. https://doi.org/10.1080/17549507.2018.1477991

Mirzaei, M. S., Meshgi, K., Akita, Y. & Kawahara, T. (2017) Partial and synchronized captioning: A new tool to assist learners in developing second language listening skill. *ReCALL*, 29(2): 178–199. https://doi.org/10.1017/S0958344017000039

https://doi.org/10.1017/S095834402200012X Published online by Cambridge University Press

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

Oh, E. Y. & Song, D. (2021) Developmental research on an interactive application for language speaking practice using speech recognition technology. *Educational Technology Research and Development*, 69(2): 861–884. https://doi.org/10.1007/s11423-020-09910-1

Radha, V. & Vimala, C. (2012) A review on speech recognition challenges and approaches. *World of Computer Science and Information Technology Journal*, 2(1): 1–7.

Shadiev, R. & Huang, Y.-M. (2020) Investigating student attention, meditation, cognitive load, and satisfaction during lectures in a foreign language supported by speech-enabled language translation. *Computer Assisted Language Learning*, 33(3): 301–326. https://doi.org/10.1080/09588221.2018.1559863

Shadiev, R., Huang, Y.-M. & Hwang, J.-P. (2017) Investigating the effectiveness of speech-to-text recognition applications on learning performance, attention, and meditation. *Educational Technology Research and Development*, 65(5): 1239–1261. https://doi.org/10.1007/s11423-017-9516-3

Shadiev, R., Hwang, W.-Y., Chen, N.-S. & Huang, Y.-M. (2014) Review of speech-to-text recognition technology for enhancing learning. *Journal of Educational Technology & Society*, 17(4): 65–84.

Shadiev, R., Hwang, W.-Y., Huang, Y.-M. & Liu, C.-J. (2016) Investigating applications of speech-to-text recognition technology for a face-to-face seminar to assist learning of non-native English-speaking participants. *Technology, Pedagogy and Education*, 25(1): 119–134. https://doi.org/10.1080/1475939X.2014.988744

Shadiev, R., Sun, A. & Huang, Y.-M. (2019) A study of the facilitation of cross-cultural understanding and intercultural sensitivity using speech-enabled language translation technology. *British Journal of Educational Technology*, 50(3): 1415–1433. https://doi.org/10.1111/bjet.12648

Shadiev, R., Wang, X., Wu, T.-T. & Huang, Y.-M. (2021) Review of research on technology-supported cross-cultural learning. *Sustainability*, 13(3): 1–23. https://doi.org/10.3390/su13031402

Shadiev, R., Wu, T.-T., Sun, A. & Huang, Y.-M. (2018) Applications of speech-to-text recognition and computer-aided translation for facilitating cross-cultural learning through a learning activity: Issues and their solutions. *Educational Technology Research and Development*, 66(1): 191–214. https://doi.org/10.1007/s11423-017-9556-8

Shadiev, R. & Yang, M. (2020) Review of studies on technology-enhanced language learning and teaching. *Sustainability*, 12(2): 1–22. https://doi.org/10.3390/su12020524

Tsai, P. (2019) Beyond self-directed computer-assisted pronunciation learning: A qualitative investigation of a collaborative approach. *Computer Assisted Language Learning*, 32(7): 713–744. https://doi.org/10.1080/09588221.2019.1614069

van Doremalen, J., Boves, L., Colpaert, J., Cucchiarini, C. & Strik, H. (2016) Evaluating automatic speech recognition-based language learning systems: A case study. *Computer Assisted Language Learning*, 29(4): 833–851. https://doi.org/10.1080/09588221.2016.1167090

Wang, Y.-H. & Young, S. S.-C. (2014) A study of the design and implementation of the ASR-based iCASL system with corrective feedback to facilitate English learning. *Journal of Educational Technology & Society*, 17(2): 219–233.

Wang, Y.-H. & Young, S. S.-C. (2015) Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system. *Journal of Computer Assisted Learning*, 31(6): 493–504. https://doi.org/10.1111/jcal.12079

Xiao, W. & Park, M. (2021) Using automatic speech recognition to facilitate English pronunciation assessment and learning in an EFL context: Pronunciation error diagnosis and pedagogical implications. *International Journal of Computer-Assisted Language Learning and Teaching*, 11(3): 74–91. https://doi.org/10.4018/IJCALLT.2021070105

Yu, P., Pan, Y., Li, C., Zhang, Z., Shi, Q., Chu, W., Liu, M. & Zhu, Z. (2016) User-centred design for Chinese-oriented spoken English learning system. *Computer Assisted Language Learning*, 29(5): 984–1000. https://doi.org/10.1080/09588221.2015.1121877

Yueh, H.-P., Lin, W., Liu, Y.-L., Shoji, T. & Minoh, M. (2014) The development of an interaction support system for international distance education. *IEEE Transactions on Learning Technologies*, 7(2): 191–196. https://doi.org/10.1109/TLT.2014.2308952

About the authors

Rustam Shadiev is a professor at Nanjing Normal University and distinguished professor of Jiangsu province. He has been appointed as fellow of the British Computer Society and is a senior member of IEEE. He was selected as the Most Cited Chinese Researchers in the field of education by Elsevier in 2020 and 2021. His research interest includes technology-assisted language learning and cross-cultural education.

Liu Jiawen is a graduate student at the School of Education Science, Nanjing Normal University. She is interested in topics related to technology-enhanced language learning.

Author ORCiD. Rustam Shadiev, https://orcid.org/0000-0001-5571-1158
Author ORCiD. Jiawen Liu, https://orcid.org/0000-0002-3843-845X