



# Using texts generated by STR and CAT to facilitate student comprehension of lecture content in a foreign language

Rustam Shadiev<sup>1</sup> · Ai Sun<sup>2</sup>

Published online: 23 November 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

In this study, we applied a combination of speech-to-text recognition (STR) and computer-aided translation (CAT) technologies during lectures in English as a foreign language to facilitate student comprehension of the lecture content. The instructor lectured in English, the STR system generated texts from the voice input, and the CAT system then simultaneously translated the STR texts into the students' native language. We aimed to test the feasibility of applying combined STR and CAT technologies to facilitate student comprehension of lecture content in a foreign language. To this end, we designed an experiment. Three groups with twenty students each were formed. All students attended the same lectures: (a) students in the control group attended lectures without any support, (b) students in experimental group 1 attended lectures with STR support (i.e., they were presented with texts in English generated from the instructor's speech by STR), and (c) students in experimental group 2 attended lectures with STR and CAT support (i.e., they were presented with texts in their native language that were translated from English by STR and CAT). We compared the posttest results of the students in the three groups. We also explored the effects of our approach with respect to different levels of foreign language ability. Finally, we surveyed the perceptions of students in experimental group 2 about the usefulness of the translated texts for their learning. Our results showed that applying STR and CAT technologies together was a useful approach: the translated texts helped significantly improve student learning performance compared to that of the students in the control condition. Translated texts were beneficial for students, as the students were able (a) to confirm some words that were not clearly spoken by the instructor or to find the meaning of some words with which the students were not familiar and (b) to complement spoken lecture content with translated content to help information processing and enhance comprehension. When comparing students with different language abilities, we found that students with low language abilities benefited from the translated texts the most. The students' language ability was significantly lower than that of the high-ability students before the experiment; however, the low-ability students' learning performance showed no significant difference from the high-ability students after the experiment. Finally, most students

---

Extended author information available on the last page of the article

perceived translated texts as useful for their learning, and they intended to use the texts in the future for learning purposes.

**Keywords** Comprehension · Lecture · Foreign language · STR · CAT

## Introduction

Our educational society has become more globalized in recent years due to rapid technological developments. As a result, it is common practice in many countries to organize international academic events with English as the main communication language. Such events take place in either a physical or virtual space. Nevertheless, it has been suggested that not all participants in such academic events benefit from these events, especially those participants for whom English is not their native or first language. Related studies document instances when, due to limited language ability, participants must exert extra effort to achieve comprehension, while some participants are still unable to comprehend (Camiciottoli 2005; Debusse et al. 2009; Miller 2007; Parmar et al. 2015; Pearce and Scutter 2010). It has been suggested that the inability to comprehend lecture content presented in a foreign language is associated with increased cognitive load (Bloomfield et al. 2010). The reason for such a difficulty is that working memory limits learners' cognitive capacity to accommodate demands imposed by listening to content delivered in a foreign language (Paas and Sweller 2014). According to Kurz (2009), attending to the auditory channel involves a greater cognitive load compared to attending to the visual channel, because speech is continuous and transitory in nature. Thus, a learner must receive verbal input, retain it in working memory and process it. After that, a learner must integrate the processed information with what follows, all the while continually adjusting one's understanding to prior knowledge (Chen and Chang 2009). A learner must keep up with the flow of the verbal input, constantly organizing and anticipating on the basis of whatever has been spoken by the speaker (Diao et al. 2007). More importantly, a learner does not have control over elements of the verbal input, such as the amount or delivery rate. Furthermore, it is impossible to pause the speech if it is too fast or to rewind it to the parts that were not understood (Graham 2011). For these reasons, verbal input processing is very complex and imposes a heavy cognitive load on working memory (Keysar et al. 2012).

Several approaches have been proposed in the literature to address the issue of comprehending spoken content delivered in a foreign language. For example, Nisbet et al. (2005) have suggested applying speech-to-text recognition (STR) technology, which synchronously generates text streams from speech input that are then shown to students on their computer screens or on a projector screen. Texts generated by STR can help students attain a better understanding of a lecture and improve their note-taking during lectures (Wald and Bain 2008). Furthermore, STR texts are useful for confirming what is being said (Ryba et al. 2006). Since

the speaker speaks in English and the texts generated by STR are also in English, we assume that the content of academic events can still be difficult to comprehend for participants who are less fluent in English. Therefore, this issue needs to be considered.

The latest advancements in information technology have enabled us to access a wide range of powerful computing tools. One of them is computer-aided translation (CAT) technology, which translates texts from one language into another (Godwin-Jones 2011). According to Hwang et al. (2012), applications using STR technology can be advanced with CAT technology to conduct lectures in different languages. For example, when STR technology generates texts from speech input in English, CAT technology can simultaneously translate them into participants' native languages, thereby making the content of an academic event understandable. Following this notion, we aimed to apply combined STR and CAT during lectures presented in English to translate the content for nonnative English speaking participants into their mother tongue. We tried to explore whether our approach will be helpful in enhancing participants' comprehension of lecture content. Therefore, we tested the feasibility of our approach in this present study.

## Theoretical background

Scholars have used different theories and hypotheses to explain how learners comprehend learning content delivered to them in a foreign language. According to the input hypothesis, learners are able to comprehend learning content in a foreign language when they receive information that they can understand (Krashen 2014). Good enough theory suggests that the depth of information processing can vary for a number of reasons, e.g., proficiency in a foreign language (Ferreira et al. 2002). According to this theory, learners do not always fully process learning content, and their information processing system has a tendency to develop shallow and superficial representations when content is difficult to comprehend (Ferreira et al. 2009). As a result, formed representations are often shallow and incomplete (Ferreira et al. 2002). Following the notion of input hypothesis and the good enough theory, the affective filter hypothesis was proposed (Krashen 1985). This hypothesis states that learners experiencing negative emotions such as fear or anxiety while learning content in a foreign language may fail to comprehend the content, because their comprehension ability will be constrained (Krashen 2014). Usually, learners have negative emotions when they need to study learning content that is difficult and when their language proficiency is low (Cheng 2000).

How a learner processes content presented in different forms (e.g., lecture content delivered as a speech, PowerPoint slides, or transcripts) can be explained by the cognitive theory of multimedia learning (Clark and Mayer 2016; Mayer 2009). This theory states that visual and verbal content is processed in different parts of the brain. A learner receives visual content through the eyes and processes it via the visual channel, whereas verbal content is received through the ears and processed via the verbal channel. When listening to speech, a learner pays attention to the verbal message, and then parses and segments the message into words that are

retained in the verbal working memory. After that, a learner transforms the words into verbal mental representations, and connections are mentally constructed to organize words into cause-and-effect chains. When learning from visual content, a learner pays attention to the content, selects images, and holds them in the visual working memory. After that, a learner mentally builds connections that organize images into cause-and-effect chains. Finally, the verbal mental model, the visual mental model, and prior knowledge are merged by constructing referential connections among them.

In contrast, according to cognitive load theory (Sweller 1994), the same information presented in auditory and written forms makes the information redundant and gives rise to a split-attention effect that leads to increased cognitive load (Clark and Mayer 2016; Mayer and Moreno 2003). The redundancy effect (Sweller et al. 2011a, b) is likely to take place during lectures where lecture transcripts are presented simultaneously, since the same information is also presented to the learners in verbal (i.e., lecture content delivered as a speech) and visual (lecture transcriptions) forms. As a result, the redundancy effect may hinder learning, so less information needs to be presented instead of more information in a multimedia format (Sweller et al. 2011a, b). However, Clark and Mayer (2016) have argued that in some particular situations, for example, when the lecture content is delivered in a foreign language and when it is difficult to understand, multimedia content is useful and even necessary. In such situations, visual and verbal information can be useful for complementing each other in the information-processing process.

Another important principle of cognitive load theory is the expertise reversal effect (Kalyuga et al. 2003), which concerns the effectiveness of learning material for learners with differing levels of prior knowledge. That is, the expertise reversal principle states that instructional techniques that are highly effective for novice learners may not be effective when used by more knowledgeable learners (Kalyuga 2014). The reason is that our working memory has a limited capacity to process learning information and different degrees of information element interactivity, i.e., the number of elements that must be attended to in order to understand the information (Kalyuga et al. 2003). If information has low-element interactivity, then each element of the information can be learned individually. Thus, such information does not impose a heavy cognitive load and can be learned easily (Sweller et al. 2011a, b). For example, in the chemistry periodic table, each chemical symbol stands for one element that should be processed in working memory, so that students can learn each symbol individually with no reference to other symbols. On the other hand, when information has high-element interactivity, individual elements interact and should be learned simultaneously rather than as individual elements (Kalyuga 2014). This may cause a heavy cognitive load and make the information difficult to learn. For example, when students learn the various ways in which symbols are manipulated in a chemical equation, the symbols cannot be learned in isolation but the entire equation (i.e., including all its elements) should be considered (Sweller et al. 2011a, b).

The expertise reversal principle has been applied to different domains, e.g., science or mathematics, in which various concepts were learned and which were related to various content expertise (Kalyuga et al. 2003). This principle has been successfully used in foreign or second-language learning studies (Chen et al. 2012; Lee and Kalyuga 2011; Sweller 2017; Yeung 1999). In the field of language learning, expertise level has been defined as a language proficiency and has been measured by various language proficiency tests (Chen et al. 2012; Lee and Kalyuga 2011). It has been suggested that individual vocabulary may represent low-element interactivity information, whereas high-element interactivity information can be represented by a combination of vocabulary, word order and sentence structure (Lee and Kalyuga 2011; Sweller et al. 2011a, b). Therefore, when learners have no related prior knowledge, for example, they do not know the vocabulary used in the learning content, and it will be difficult for them to process the learning information. In addition, learning is more difficult during lectures delivered in a foreign language: scholars have explained that during information processing of such lecture content, learners must receive and retain information in their working memories and then integrate the information with what follows, all the while continually adjusting their understanding to their prior knowledge (Chen and Chang 2009).

Sweller (2017) and Yeung (1999) have also clearly pointed out that expertise level can be related to the language comprehension level. For example, they have suggested that translations are essential for novice learners and should be integrated with learning content; however, translation should be eliminated entirely from the learning content of students with higher levels of expertise. In other words, “an instructional design that is suitable for novices gradually loses its effectiveness with increasing expertise and may become dysfunctional for more expert learners” (Sweller 2017). Scholars have explored the expertise reversal effect in different language learning situations in multimedia learning environments. For example, Chen et al. (2012) investigated the effects of written text on the comprehension of spoken English as a foreign language when it was simultaneously displayed to learners. Yeung (1999) examined the effects of cognitive load when readers of varying levels of expertise were provided with vocabulary definitions during reading activities to facilitate their reading comprehension and vocabulary learning. Lee and Kalyuga (2011) developed effective techniques for reducing learner cognitive overload while using pinyin (a phonetic system) to learn the Chinese language.

Informed by related theories, we provided students with lecture content in multiple modalities. We assumed that such an approach can be useful for students attending lectures in a foreign language to facilitate their understanding of lecture content. In addition, we attempted to investigate how our approach is beneficial for students with different language abilities (i.e., low vs. high). For example, lecture content presented in multiple modalities can be beneficial for facilitating the understanding of language learners with a low level of proficiency. In contrast, presenting content in multiple modalities will be redundant and counterproductive for language learners at high levels of proficiency, as processing multimodal content requires additional cognitive resources (Kalyuga et al. 2003).

## The use of speech-to-text recognition and computer-aided translation technologies for education

According to earlier related research, STR technology can be a useful tool to assist student learning during lectures (Hwang et al. 2012; Kheir and Way 2006; Kuo et al. 2012; Ranchal et al. 2013; Ryba et al. 2006). STR technology synchronously transcribes text streams from a lecturer's speech input, which are then shown to students on their computer screens or a projector screen (Nisbet et al. 2005). In the study by Hwang et al. (2012), the students who attended lectures in online learning environments often experienced problems associated with network connections and could not hear the lecturer, so they read STR texts to better follow the lecturer. Kheir and Way (2006) adopted STR technology during lectures to assist the learning of hearing-impaired students: using STR was the only way for these students to attend and comprehend the lecture. Ranchal et al. (2013) also used STR for education. Their students received lecture transcriptions during and after the lectures. It was found that when transcripts were available during lectures, the students paid more attention to the instructor instead of focusing on the note-taking process. After lectures with the lecture transcripts, the students reviewed the lecture material, took additional notes, made comments, and searched for key terms using keywords and time periods. In the study by Ryba et al. (2006), the students were nonnative speakers of English and attended lectures in English. When students encountered unfamiliar vocabulary, misheard their lecturer or could not understand portions of the lecture, they read STR texts to facilitate their comprehension of the lecture content.

Related studies have suggested that CAT can aid learning, particularly second or foreign language learning. CAT translates texts from one language into another (Godwin-Jones 2011). Hermet and Desilets (2009) used CAT for writing activities such as composing essays or correcting grammatical and lexical errors in essays. ElShiekh (2012) applied CAT to a research writing course and explored the translation process from English into Arabic and vice versa. He found that one can translate and search for appropriate words to express opinions and ideas using CAT. Omar et al. (2012) introduced CAT to students to support their online discussion in a foreign language. Omar et al. (2012) suggested that CAT is capable of checking grammar and spelling and helps overcome problems when constructing sentences. Shadiev et al. (2018) adopted CAT during a cross-cultural learning project. Representatives of thirteen different cultures communicated and exchanged culture-related information with each other in their native languages, and CAT helped translate the communication content. According to Shadiev et al. (2018), cross-cultural learning took place, and CAT played an important role in this because CAT enabled multilingual interaction among the participants.

### Research questions

Related theory and literature informed the design of this study. We applied STR and CAT technologies during lectures in a foreign language, following the general recommendations of the cognitive theory of multimedia learning. We aimed to explore whether our approach can be useful for students attending lectures in a foreign language

and whether the approach can facilitate their understanding of lecture content. In addition, we attempted to investigate how our approach is beneficial for students with different language abilities (i.e., low vs. high). Finally, we investigated participants' perceptions of our approach. The following research questions were addressed:

1. Do students who use texts generated by STR and CAT technologies perform differently from those who do not use them?
2. How can the differences be accounted for regarding low or high English as a foreign language (EFL) ability?
3. What are students' perceptions of our approach?

## Method

### Participants

We invited potential participants by distributing and displaying a poster with information about our study. Sixty students from one state university who majored in social sciences were recruited. Most of the participants were between 18 and 22 years old, and all of them were native speakers of Russian. Before our study began, we adequately explained the study to the participants and obtained informed consent.

### Experimental procedure

In the beginning, we collected participants' demographic information and administered an EFL ability pretest. After the test, all students attended two lectures on general topics given in English. The lectures were on general topics: the first lecture, "Photography", was about two friends taking pictures and how to take good pictures, and the second lecture, "From matchmakers to dating services", was about marriage traditions and customs around the world. The reason for selecting lectures on general topics was the capacity of the STR and CAT technologies. Shadiev et al. (2018) urges that, at present, "STR and CAT should not be considered a well-rounded professional translation mechanism from voice input". That is, STR and CAT technologies are not yet sufficiently mature to be applied to lectures on academic topics. Such topics contain specific and complex terminologies and concepts that these technologies have difficulty translating accurately. In contrast, topics containing simple, everyday vocabularies and sentences are preferred when using STR and CAT technologies, as such content can be translated with a high accuracy of more than 90% (Shadiev and Huang 2016). Therefore, we selected general topics for our lectures, with simple content related to daily life.

We applied STR system during the lectures. The system received speech input from the instructor and simultaneously generated texts from that input. STR texts were displayed on computer screens for the students to read during both lectures. CAT system was also employed, and in addition to the STR function of receiving speech input from the instructor, the system also translated the STR output from English into Russian. Translated texts were then displayed on the computer screens



during lectures for the students. We randomly divided the students into three groups, with twenty students in each:

1. Control group—students attended a lecture without any support;
2. Experimental group 1—students attended a lecture in which the STR technology generated texts;
3. Experimental group 2—students attended a lecture in which STR texts were translated into Russian by CAT.

We carried out a posttest after each lecture and compared the outcomes of the students in the three groups. After that, we administered a questionnaire survey to the students in experimental group 2 to survey their perceptions regarding the usefulness of the translated texts for learning. Finally, we conducted one-on-one semi-structured interviews with twenty students randomly selected from experimental group 1 and experimental group 2, that is, with ten students from each group, to explore the possible reasons for our main findings related to the research questions.

### **Application of speech-to-text recognition and computer-aided translation**

The Windows Speech Recognition system was applied as STR technology for this study. The system received speech input from the instructor and simultaneously generated texts from that speech input, which was then displayed for the students on their computer screens. Earlier studies have claimed that the Windows Speech Recognition system is an accurate and valuable tool for supporting students' learning during lectures in a foreign language (Nisbet et al. 2005; Ryba et al. 2006; Wald and Bain 2008). The Google Translate system was employed as CAT technology to translate STR texts from English into Russian. The Shadiev et al. (2016) has argued that the accuracy rates of CAT can be as high as 88% for Russian and 89% for Chinese during bilingual cross-cultural communication. An even higher accuracy rate can be attained when using the technology to translate shorter and less complicated sentences; CAT generates more errors when translating longer and more complicated sentences, because it considers a highly limited linguistic context (Mellebeek et al. 2005). Another way to improve the accuracy rate is to train these technologies as well as to add unfamiliar domain-specific terminology to their databases (Hwang et al. 2012; Shadiev et al. 2014). Wald and Bain (2008) suggest that the accuracy rate can reach more than 90% after such preliminaries are addressed. We also followed these useful guidelines to improve the accuracy rate; thus, in our study, the accuracy rate of all translated texts was higher than 95%. Wald and Bain (2008) claim that texts with accuracy rates higher than 75% are reasonably accurate, acceptable and useful for students and can enable teaching and learning. In this study, the instructor employed the STR system during lectures in English; the system generated texts from the voice input, and the STR texts were displayed on computer screens for the students to read during both lectures. The accuracy rate of STR-generated texts was 100%. An extract from a STR-text is included in "Appendix 2". The instructor also employed the CAT system during lectures; the system translated STR



texts from English into Russian, and CAT texts were displayed on computer screens for the students. An extract from a CAT-text is included in “[Appendix 2](#)”.

Some may confuse our approach of employing STR and CAT technologies with the grammar–translation method. The grammar–translation method is a method of teaching foreign languages by which students learn grammatical rules and then apply those rules by translating sentences between the target language and the native language. Please note that English was not the main subject for students to learn in this study. Instead, students learned about general topics, but the information was presented in English. Therefore, we employed STR and CAT technologies that translated lecture content from English into Russian, and the translated texts were shown to students to help them better comprehend the lecture.

## Data collection

Data were collected from evaluations via tests, a questionnaire survey and interviews. We triangulated different data sources to ground the findings related to the effectiveness of texts generated by STR and CAT for learning. In other words, the results of the learning performance outcomes and their comparison between the control and experimental groups were supported by the questionnaire survey and interview results.

We carried out one pretest and two posttests. To measure the effectiveness of our treatments, i.e., an application of (a) STR and (b) STR and CAT, for students’ learning performance, we compared test scores of students in three groups. The pretest was carried out before the experiment to measure the EFL ability (i.e., the ability of an individual to understand while reading and listening to information in English) of students. The pretest included nine multiple-choice items (i.e., we measured reading ability with five items and listening skills with four items). We scored the pretest results on a 9-point scale (with “9” as the highest score); each correct answer to an item was scored as “1”, while each incorrect answer received a “0”. The posttests were carried out after lectures, i.e., one posttest after one lecture, to measure learning performance. Each posttest included the following: (a) five multiple-choice items to measure information recognition (each correct answer to one item was scored as “1”, while each incorrect answer received a “0”); (b) two open-ended questions to measure information recall (each complete and correct answer was scored as “2”, each partly and correctly answered item was scored as “1”, and incorrect answers received a “0”); and (c) one summary writing task to measure understanding of the learning content (“5” was the highest score). We scored the posttest results on a 14-point scale (with “14” as the highest score). All tests were designed by experienced EFL teachers. The pretest was developed based on the General English Proficiency Test (GEPT) exam and contained items related to listening and reading skills. The posttests were created based on the learning content of the lectures. Student answers to the open-ended questions and summary writing tasks were first coded using the sentence as a coding unit and were then scored by three raters. Major differences in assessment were resolved through discussion. All three raters were experienced in EFL teaching. The interrater reliability of the content was evaluated using Cohen’s kappa. The analysis results exceeded 0.90, indicating high reliability.

Although students wrote their answers for the open-ended questions, we did not assess their writing skills. In addition, we did not assess students' prior knowledge. The reason for these choices is that the content of the two lectures was general and unique. That is, two characters, Daniel and Winnie, were introduced in the Photography lecture, and it was mentioned that Daniel has a small digital camera and that Winnie has a large professional camera. We could not assess students' prior knowledge, because they did not know such information prior to the lecture. However, the students' comprehension was measured using the following test item:

Which statement is true?

- A. Only Daniel has a small camera; Winnie has a large camera.
- B. Winnie has a small camera. Daniel has a large camera.
- C. Daniel and Winnie have the same digital camera.
- D. Winnie and Daniel use film in their cameras.

We evaluated students' perceptions of our approach using a questionnaire survey (see "[Appendix 1](#)"). We adopted two dimensions of the technology acceptance model (TAM) (Venkatesh and Davis 2000) for the questionnaire. Venkatesh and Davis (2000) demonstrated that TAM is valid and reliably measures users' acceptance of technology. In addition, TAM has been successfully used in a wide array of educational technology research areas (Shadiev et al. 2016; Hwang et al. 2012; Ryba et al. 2006). The two adopted dimensions were (items 1–6) the usefulness of the treatment (i.e., "STR" for experimental group 1 and "STR and CAT" for experimental group 2) for learning—the degree to which a student believes that using the treatment for learning would enhance his or her learning performance—and (items 7–9) behavioral intentions to use the treatment for learning in the future—a major determinant of whether a student would actually use the treatment for learning. Responses to the questionnaire items were scored using a five-point Likert scale, anchored by the end-points "strongly disagree" (1) and "strongly agree" (5). Twenty valid answer sheets to the questionnaire were obtained from twenty students from experimental group 2. We employed Cronbach's  $\alpha$  to assess the internal consistency of the survey, and the values exceeded 0.90.

We conducted one-on-one semistructured interviews with students to explore their experiences using texts generated by STR and CAT as well as their perceptions of our approach. Students were asked the following questions: (1) Did you use the texts generated by STR and CAT during the lectures, and for what reasons? (2) Were these texts useful for learning during lectures and why? Each interview took approximately 30 min. We used open coding to analyze the interview data. First, all interviews were audio-recorded, with the permission of the interviewees, and they were then fully transcribed for analysis. Then, the text segments that met the criteria for providing the best research information were highlighted and coded. Next, the codes were sorted into categories, i.e., codes with similar meanings were aggregated. Established categories produced a framework to illustrate findings relevant to the research questions. Two coders were involved in the coding process. Differences in coding and categorization were resolved through discussion until a consensus was reached. The

interrater reliability of the interview data was evaluated using Cohen's kappa, and the result exceeded 0.90, which indicated that the interrater reliability was high.

## Data analysis

We adopted the following methods of statistical analysis: (1) one-way multivariate analysis of variance (one-way MANOVA) to evaluate the difference between three groups on three tests and (2) an analysis of covariance to measure the difference between high- and low-ability students in experimental group 2, controlling for their EFL ability. We set a prior alpha-level at 0.05, since an alpha level of less than 0.05 is accepted in most educational research as statistically significant.

All ethical guidelines were met when this research was carried out, and approval from the relevant institutional ethics committee was obtained, under whose supervision the data were collected and reported.

## Results and discussion

Our results are presented in the following order. First, we report results related to the assessment of students in the three groups on the pretest and the two posttests, and we compare these results among the groups. Second, we present the results on the EFL ability and performance assessment of students in the three groups with respect to different language abilities, and we also compare the results among the groups. Third, we report our results related to differences in learning performance between low-ability and high-ability students in experimental group 2 and explore the relationship between language ability and the benefits of translated texts for learning. Finally, we present the results of the questionnaire survey on the perceptions of students in experimental group 2 regarding our approach to applying STR and CAT during lectures in English.

### Learning outcomes assessment across different experimental conditions

The results of the assessment with respect to students in the three groups are presented in Table 1. According to the results of the statistical analysis (Table 2), there was no significant difference in EFL ability among students in the control group ( $M=4.60$ ,  $SD=2.23$ ), experimental group 1 ( $M=3.95$ ,  $SD=2.11$ ), and experimental group 2 ( $M=4.65$ ,  $SD=2.45$ ),  $F=.591$ ,  $p=.557$ . After lecture 1, our assessment

**Table 1** Descriptive statistics for assessment of learning performance in three groups: means and standard deviations

Test	Control group		Experimental group 1		Experimental group 2	
	M	SD	M	SD	M	SD
Pre-test	4.60	2.23	3.95	2.11	4.65	2.45
Post-test (lecture 1)	3.65	1.75	4.35	2.13	5.80	2.48
Post-test (lecture 2)	4.10	2.42	5.05	2.99	7.15	3.25

**Table 2** One-way MANOVA by condition: main effects obtained for all measures across different treatment conditions

Source	Dependent variable	df	Mean square	F-value	<i>p</i> value	Post hoc
Treatment	Pre-test	2, 57	3.050	0.591	.557	
	Post-test (lecture 1)	2, 57	24.050	5.226	.008	CG < EG2 ( <i>p</i> = .007)
	Post-test (lecture 2)	2, 57	48.717	5.746	.005	CG < EG2 ( <i>p</i> = .005)

CG control group, EG2 experimental group 2

results showed that a significant difference exists in the posttest scores of students ( $F=5.226$ ,  $p=.008$ ). Post hoc analysis results demonstrated that students in the control group ( $M=3.65$ ,  $SD=1.75$ ) had significantly lower performance compared to the students in experimental group 2 ( $M=5.80$ ,  $SD=2.48$ ),  $p=.007$ . However, the results showed no significant difference in posttest scores between students in the control group and experimental group 1 and between students in experimental group 1 and experimental group 2.

Similar results were obtained for lecture 2—a significant difference existed in the posttest scores of students ( $F=5.746$ ,  $p=.005$ ). Post hoc analysis results revealed that students in the control group ( $M=4.10$ ,  $SD=2.42$ ) had significantly lower performance compared to students in the experimental group 2 ( $M=7.15$ ,  $SD=3.25$ ),  $p=.005$ . In addition, there was no significant difference in performance on the post-test between students in the control group and experimental group 1 and between students in experimental group 1 and experimental group 2. These findings may suggest that at the beginning of the experiment, all students had similar EFL ability. However, after the first and second lectures, students who were provided with a translation of the lecture content outperformed students in the control condition. This finding is in line with the cognitive theory of multimedia learning (Clark and Mayer 2016; Mayer 2009). The theory suggests that providing multimedia content can be beneficial for learning, as different media complement one another during the information-processing process. No significant difference was found between the learning performance of students in the control group and experimental group 1 and between students in experimental group 1 and experimental group 2. This result may suggest that during lectures in English, transcriptions generated by STR technology in English were not as useful for learning as the translated texts from English into Russian. Previous studies have shown that STR texts are beneficial for learning during lectures in a foreign language (Nisbet et al. 2005; Ryba et al. 2006; Wald and Bain 2008), and their results contradict ours. The following reason may explain this contradiction. Students in previous studies had very different language learning backgrounds. That is, they were foreign students in an English-speaking country (Nisbet et al. 2005; Ryba et al. 2006; Wald and Bain 2008) for whom English was the main language of instruction. In our study, students learned English as a foreign language, and their main language of instruction was Russian, whereas students in Shadiev et al. (2017) were from Taiwan, for whom English was also a foreign language. However, in contrast to our students, the Taiwan students were exposed to English texts and subtitles much more frequently. Chen (2004) argued

that subtitles/transcriptions are common practice in Taiwan and are widely used for learning and entertainment. For example, almost all educational and entertainment TV channels (both in Mandarin and English) feature subtitles/transcriptions that are presented along with audio output. For this reason, students in Taiwan are more familiar with using subtitles/transcriptions, are more experienced with using them and are more skilled with better strategies to use them during learning compared to the students in our study. It is also possible that students in previous studies and this research had different EFL abilities; however, we did not test this, and our pretest is not comparable to the pretests used in the previous studies. However, we assumed that students in previous studies had higher EFL levels. For example, as mentioned earlier, students in Nisbet et al. (2005), Ryba et al. (2006) and Wald and Bain (2008) were foreign students and to enter a university in an English-speaking country, they needed to have very high scores on EFL exams. Moreover, Taiwan is ranked as a top country based on the average level of skill in English as a foreign language (First 2013). Our students were not measured on English as a foreign language, and thus, we assume that their EFL abilities were lower compared to those of the Taiwanese students. These reasons may explain the degree of usefulness of transcriptions and translations during lectures. Our students said in the interviews that translated texts were more useful for learning and understanding the lecture content because the translated texts were in their native language. The following are two extracts from the interviews with the students.

Actually, my English ability is not very good, and so I have difficulties in understanding speech or text in English. When the lecture content was translated into my native language and the translated texts were shown to me during lectures, I found them very useful in aiding my learning and comprehension of the lecture content (Student 1).

Translated texts were in my native language, and this is why I was able to perfectly understand the lecture content. Otherwise, I always feel frustrated when I am in lectures done in English, because I do not understand their content even with transcriptions, as my language ability is low (Student 2).

This finding is in line with the expertise reversal effect (Kalyuga et al. 2003). Kalyuga (2014) argued that learning material and techniques that are highly effective for students with lower language ability may not be effective when used by students with a higher language ability.

### **Assessment of learning outcomes across different experimental conditions and language abilities**

Next, we divided students in each group into low- and high-ability students based on their pretest scores. High-ability students were the top ten students in a group, and low-ability students were the bottom ten students in a group. With this grouping approach, we aimed to make the size of the high- and low-ability groups as large as possible to draw conclusions with a high degree of confidence. The results of

**Table 3** Descriptive statistics for assessment of learning performance in three groups (high ability and low ability): means and standard deviations

Ability/test	Control group		Experimental group 1		Experimental group 2	
	M	SD	M	SD	M	SD
Low ability						
Pre-test	2.80	1.39	2.20	.79	2.50	1.08
Post-test (lecture 1)	2.90	1.10	3.20	1.13	4.10	1.37
Post-test (lecture 2)	2.20	1.39	3.90	2.81	4.80	1.99
High ability						
Pre-test	6.40	1.17	5.70	1.42	6.80	1.13
Post-test (lecture 1)	4.40	2.01	5.50	2.32	7.50	2.17
Post-test (lecture 2)	6.00	1.56	6.20	2.86	9.50	2.46

**Table 4** One-way MANOVA by condition and ability (high ability and low ability): main effects obtained for all measures across different treatment conditions

Source	Ability/dependent variable	df	Mean square	F-value	p value	Post hoc
Treatment	Low ability					
	Pre-test	2, 27	.900	.721	.495	
	Post-test (lecture 1)	2, 27	3.900	2.673	.087	
	Post-test (lecture 2)	2, 27	17.433	3.793	.035	CG < EG2 ( $p = .034$ )
	High ability					
	Pre-test	2, 27	3.100	1.988	.157	
Post-test (lecture 1)	2, 27	24.700	5.235	.012	CG < EG2 ( $p = .011$ )	
Post-test (lecture 2)	2, 27	38.633	6.949	.004	CG < EG2 ( $p = .008$ ) EG1 < EG2 ( $p = .013$ )	

CG control group, EG1 experimental group 1, EG2 experimental group 2

the pretest and posttests are presented in Table 3, and we include the results of the comparison among these scores in Table 4. According to the results, there was no significant difference in pretest scores among the three groups across different EFL abilities. This finding is consistent with the one we reported earlier (see Table 1). We also found no differences among the three low EFL ability groups on posttest scores after lecture 1. However, there was a significant difference between the scores of low EFL ability students in the control group and low EFL ability students in experimental group 2 on the posttest after lecture 2. That is, the learning performance of the control group ( $M = 2.20$ ,  $SD = 1.39$ ) was lower than that of experimental group 2 ( $M = 4.80$ ,  $SD = 1.99$ ),  $p = .034$ . The reason why there was no significant difference between groups after lecture 1 was revealed through the interview results. Low-ability students said that they were unable to immediately discern the strengths of our approach during the first lecture (Code 3.2 in Table 5), because this was their first lecture of this type (Code 3.1). However, later, during lecture 2, they became familiar with our approach and learned how it could be beneficial for learning (Code

**Table 5** Interviews data coding

Category	Code	Definition	Examples
1. Strategy	1. Confirmation	To confirm meaning of unfamiliar vocabulary	A student read transcripts to confirm and understand meaning of unfamiliar vocabulary
	2. Supplementation	To supplement verbal content with transcripts	A student supplemented verbal content with transcripts to enhance comprehension
	3. Verification	To verify verbal content with transcriptions	A student verified spoken content of the lecturer with transcription to enhance comprehension
	4. Translate	To translate new vocabulary using transcripts	A student translated new vocabulary using transcripts to understand lecture content better
2. Advantage	1. Experience	Experience to use transcriptions during lectures was beneficial for learning	A student performed better after using transcriptions for some time compared to the beginning of this study
3. Other	1. Novelty	The treatment was novel to a student	It required some time to get used to the treatment because it was new to a student
	2. Lack of understanding of benefits from the treatment	A student did not understand what are benefits of the treatment	In the beginning of this study, a student did not understand what benefits of the treatment were
	3. Earlier understanding of benefits from the treatment	A student understood what are benefits of the treatment from the start	A student understood what benefits of the treatment were right from the beginning
	4. Late understanding of benefits from the treatment	A student understood what are benefits of the treatment with some lag	A student understood what benefits of the treatment were after using it for some time
	5. Not popular approach	The treatment is not popular among instructors	Not many instructors adopt the treatment during lectures so students do not have future perspective to use it



3.4). Students then implemented some useful strategies (Category 1 in Table 5) to take advantage of our approach, and as a result, their performance was much better compared to that of those in the control group. Useful strategies for using translated texts during lectures were (a) to confirm and understand the meaning of unfamiliar vocabularies (Code 1.1) and (b) to supplement spoken lecture content with translated textual content to enhance their comprehension (Code 1.2). In addition, we found that no significant difference existed between the control group and experimental group 1 and between experimental group 1 and experimental group 2 on posttest scores after lecture 2. This result echoes the one reported earlier and suggests that during lectures in English, transcriptions generated by STR technology in English were not as useful for learning as translated texts from English into Russian.

With respect to the high EFL ability students in the three groups, their EFL ability did not differ significantly before the experiment. However, students in experimental group 2 ( $M=7.50$ ,  $SD=2.17$ ) outperformed those in the control group ( $M=4.40$ ,  $SD=2.01$ ) after lecture 1,  $p=.011$ . Furthermore, after lecture 2, experimental group 2 ( $M=9.50$ ,  $SD=2.46$ ) performed significantly better than the control group ( $M=6.00$ ,  $SD=1.56$ ),  $p=.008$ , and the experimental group 1 ( $M=6.20$ ,  $SD=2.86$ ),  $p=.013$ . High-ability students know a wider variety of learning strategies and utilize them better than low-ability students. Therefore, high-ability students discerned the strengths and limitations of our approach faster and utilized associated learning strategies earlier (Code 3.3), so their performance was better after the first lecture. With more experience during lecture 2, students in experimental group 2 performed even better (Code 2.1) and, as a result, were able to outperform both groups on the posttest for lecture 2.

### **Differences in learning outcomes of experimental group 2 students across different language ability levels**

When we compared test scores for students with different EFL abilities in experimental group 2, i.e., low ability versus high ability, we found an interesting phenomenon. The comparison of pretest scores between low- and high-ability students in experimental group 2 showed a significant difference; that is, scores of the former ( $M=2.50$ ,  $SD=1.08$ ) were much lower compared to those of the latter ( $M=6.80$ ,  $SD=1.13$ ),  $t=-8.677$ ,  $p=.000$ . However, when we compared the scores on the posttests of students with low EFL ability and students with high EFL ability in experimental group 2, our results demonstrated that there was no significant difference after lecture 1 ( $F=.714$ ;  $p=.410$ ) and lecture 2 ( $F=.937$ ;  $p=.347$ ). This finding may suggest that our approach was more beneficial for low EFL ability students than for those with high EFL ability.

Results of the questionnaire showed that our approach was perceived by most students as useful for learning during lectures in English ( $M=4.17$ ,  $SD=0.65$ ). In addition, most students demonstrated their high behavioral intentions to use our approach in the future for learning ( $M=3.65$ ,  $SD=1.03$ ). Students mentioned that this approach was useful for understanding lecture content (Category 1), especially when the lecture contained some unfamiliar terminology (e.g., name of a city or

dish). Students were able to find translations of these words and understand their meaning (Code 1.4). In addition, this approach was useful for confirming some new words that the students were not familiar with (Code 1.1); students could read transcripts to confirm and understand the meaning of unfamiliar vocabulary. Furthermore, this approach helped students compare what they heard with what they read to enhance comprehension (Code 1.3). However, some students mentioned that not many instructors use such an approach during lectures, so the students felt that they will have fewer opportunities to use this approach in the future (Code 3.5). Therefore, the mean value for students' behavioral intentions is lower compared to the approach's perceived usefulness.

## Limitations

Three limitations of this study must be noted. First, the sample size was relatively small, which may limit the generalization of the research results to the wider population. Second, the lectures were on general topics; therefore, the research results have limited applicability to specific academic topics. Third, we did not use any comparable tests but made an assumption that students in previous studies and this present research had different EFL abilities to interpret the degree of usefulness of transcriptions and translations during lectures. Thus, it is suggested that researchers and educators in the field need to address these issues in the future by, for example, involving larger sample sizes, applying STR and CAT to lectures in a foreign language on specific academic topics, and using comparable tests.

## Conclusions

Our statistical results showed that applying STR and CAT together during lectures in English was beneficial for the learning of nonnative English-speaking students. Students learned better when STR and CAT were used together than students who learned without any support. Translated lecture content was particularly useful for low EFL ability students. Furthermore, the questionnaire results demonstrated that most students perceived our approach as useful for their learning during lectures in English and that they intended to use it for learning in the future.

Based on our results, we suggest applying STR and CAT technologies together to support the learning of students during lectures in a foreign language under conditions similar to those of our study, i.e., the general topics of the lectures and the specific demographics of the participants. To make better use of this novel approach, students need more time to become acquainted with the two technologies. This familiarity will help the students identify the strengths and limitations of translated texts during lectures and determine what learning strategy to use and how to obtain texts with higher accuracy rates. We also suggest that researchers and educators need to encourage low-ability students to use translated texts during lectures in a foreign language more frequently, because they benefit from these tools the most.

**Funding** There is no funding to report for this study.

## **Compliance with ethical standards**

**Conflict of interest** The authors declare that they have no conflict of interest.

## **Appendix 1**

1. The treatment improves my understanding of a lecture.
2. The treatment increases my productivity during a lecture.
3. The treatment enhances my learning effectiveness during a lecture.
4. The treatment improves my learning performance during a lecture.
5. The treatment helps me accomplish a learning task more quickly.
6. Overall, I found the treatment to be useful during a lecture.
7. I intend to continue using the treatment for learning in the future.
8. I plan to use the treatment for learning often.
9. I will strongly recommend others to use the treatment for learning.

## **Appendix 2**

### **STR-text**

#### **Photography**

Hello everyone, today I am going to talk about photography. Do you have a camera? Do you enjoy taking pictures? Daniel and Winnie are taking pictures today, so we are learning about photography.

Anyone can be a photographer. You just need a camera. Daniel has a small camera. Winnie has a big camera.

### **CAT-text**

#### **ФОТОГРАФИЯ**

Привет всем, сегодня я собираюсь поговорить о фотографии. У тебя есть камера? Вам нравится фотографировать? Дэниел и Винни сегодня фотографируются, поэтому мы учимся фотографии.

Любой может быть фотографом. Вам просто нужна камера. У Даниэля маленькая камера. У Винни большая камера.

## References

- Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult? Factors affecting second language listening comprehension*. College Park: University of Maryland.
- Camicciottoli, B. C. (2005). Adjusting a business lecture for an international audience: A case study. *English for Specific Purposes*, 24(2), 183–199.
- Chen, S. J. (2004). Linguistic dimensions of subtitling. Perspectives from Taiwan. *Meta: Journal des traducteurs/Meta: Translators' Journal*, 49(1), 115–124.
- Chen, I. J., & Chang, C. C. (2009). Cognitive load theory: An Empirical study of anxiety and task performance in language learning. *Electronic Journal of Research in Educational Psychology*, 7(2), 729–746.
- Chen, I. J., Chang, C. C., & Yen, J. C. (2012). Effects of presentation mode on mobile language learning: A performance efficiency perspective. *Australasian Journal of Educational Technology*, 28(1), 122–137.
- Cheng, X. (2000). Asian students' reticence revisited. *System*, 28(3), 435–446.
- Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (4th ed.). San Francisco, CA: Wiley.
- Debusse, J. C., Hede, A., & Lawley, M. (2009). Learning efficacy of simultaneous audio and on-screen text in online lectures. *Australasian Journal of Educational Technology*, 25(5), 748–762.
- Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology*, 120(2), 237–261.
- ElShiekh, A. A. A. (2012). Google translate service: transfer of meaning, distortion or simply a new creation? An investigation into the translation process & problems at google. *English Language and Literature Studies*, 2(1), 56–68.
- Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15.
- Ferreira, F., Engelhardt, P. E., & Jones, M. W. (2009). Good enough language processing: A satifying approach. In *Proceedings of the 31st annual conference of the cognitive science society*. Austin: Cognitive Science Society.
- First, E. (2013). *EF English proficiency index*. Retrieved on March 17, 2017.
- Godwin-Jones, R. (2011). Emerging technologies: Mobile apps for language learning. *Language Learning & Technology*, 15(2), 2–11.
- Graham, S. (2011). Self-efficacy and academic listening. *Journal of English for Academic Purposes*, 10(2), 113–117.
- Hermet, M., & Desilets, A. (2009). Using first and second language models to correct preposition errors in second language authoring. In *Proceedings of the fourth workshop on innovative use of NLP for building educational applications* (pp. 64–72). Association for Computational Linguistics.
- Hwang, W. Y., Shadiev, R., Kuo, T. C. T., & Chen, N. S. (2012). Effects of speech-to-text recognition application on learning performance in synchronous cyber classrooms. *Educational Technology & Society*, 15(1), 367–380.
- Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 576–597). New York: Cambridge University Press.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31.
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, 23(6), 661–668.
- Kheir, R., & Way, T. (2006, June). Improving speech recognition to assist real-time classroom note taking. In *Rehabilitation Engineering Society of North America (RESNA)*. Conference, Atlanta.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Reading: Addison-Wesley Longman Ltd.
- Krashen, S. (2014). Case histories and the comprehension hypothesis. *TESOL Journal*, 9(1), 70–91.
- Kuo, T. C., Shadiev, R., Hwang, W. Y., & Chen, N. S. (2012). Effects of applying STR for group learning activities on learning performance in a synchronous cyber classroom. *Computers & Education*, 58(1), 600–608.

- Kurz, I. (2009). The impact of non-native English on students' interpreting performance. In G. Hansen, A. Chesterman, & H. Gerzymisch-Arbogast (Eds.), *Efforts and models in interpreting and translation research* (pp. 179–192). Amsterdam: John Benjamins.
- Lee, C. H., & Kalyuga, S. (2011). Effectiveness of on-screen pinyin in learning Chinese: An expertise reversal for multimedia redundancy effect. *Computers in Human Behavior*, 27(1), 11–15.
- Mayer, R. E. (2009). *Multimedia learning*. New York: Cambridge University Press.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43–52.
- Mellebeek, B., Khasin, A., Owczarzak, K., Van Genabith, J., & Way, A. (2005). Improving online machine translation systems. In *Proceedings of the X MT summit* (pp. 290–297). Phuket: Thailand.
- Miller, L. (2007). Issues in lecturing in a second language: Lecturer's behaviour and students' perceptions. *Studies in Higher Education*, 32(6), 747–760.
- Nisbet, P., Wilson, A., & Aitken, S. (2005). Speech Recognition for Students with Disabilities. In *Proceedings of the inclusive and supportive education congress, ISEC 2005 conference*. Delph: Inclusive Technology.
- Omar, H., Embi, M. A., & Yunus, M. M. (2012). ESL learners' interaction in an online discussion via Facebook. *Asian Social Science*, 8(11), 67–74.
- Paas, F., & Sweller, J. (2014). Implication of cognitive load theory for multimedia. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 27–42). New York: Cambridge University Press.
- Parmar, J. R., Tejada, F. R., Lang, L. A., Purnell, M., Acedera, L., & Ngonga, F. (2015). Assessment of communications-related admissions criteria in a three-year pharmacy program. *American Journal of Pharmaceutical Education*. <https://doi.org/10.5688/ajpe79686>.
- Pearce, K., & Scutter, S. (2010). Podcasting of health sciences lectures: Benefits for students from a non-English speaking background. *Australasian Journal of Educational Technology*, 26(7), 1028–1041.
- Ranchal, R., Taber-Doughty, T., Guo, Y., Bain, K., Martin, H., Robinson, J. P., et al. (2013). Using speech recognition for real-time captioning and lecture transcription in the classroom. *IEEE Transactions on Learning Technologies*, 6(4), 299–311.
- Ryba, K., McIvor, T., Shakir, M., & Paez, D. (2006). Liberated Learning: Analysis of University students' perceptions and experiences with continuous automated speech recognition. *Journal of Instructional Science and Technology*, 9(1), 1–19.
- Shadiev, R., & Huang, Y. M. (2016). Facilitating cross-cultural understanding with learning activities supported by speech-to-text recognition and computer-aided translation. *Computers & Education*, 98, 130–141.
- Shadiev, R., Hwang, W. Y., Chen, N. S., & Huang, Y. M. (2014). Review of Speech-to-Text Recognition technology for enhancing learning. *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 for face to face seminar to assist learning of non-native English participants. *Technology, Pedagogy and Education*, 25(1), 119–134.
- Shadiev, R., Wu, T. T., & Huang, Y. M. (2017). Enhancing learning performance, attention, and meditation using a speech-to-text recognition application: Evidence from multiple data sources. *Interactive Learning Environments*, 25(2), 249–261.
- Shadiev, R., Wu, T. T., Sun, A., & Huang, Y. M. (2018). Applications of speech-to-text recognition and computer-aided translation for enhancing cross-cultural learning: Issues and their solutions. *Educational Technology Research and Development*, 66(1), 191–214.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312.
- Sweller, J. (2017). Cognitive load theory and teaching English as a second language to adult learners. *Contact Magazine*, 43(1), 10–14.
- Sweller, J., Ayres, P., & Kalyuga, S. (Eds.) (2011a). The redundancy effect. In *Cognitive load theory* (pp. 141–154). Springer New York.
- Sweller, J., Ayres, P., & Kalyuga, S. (Eds.) (2011b). *Cognitive load theory*. New York: Springer.
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186–204.

- Wald, M., & Bain, K. (2008). Universal access to communication and learning: The role of automatic speech recognition. *International Journal Universal Access in the Information Society*, 6(4), 435–447.
- Yeung, A. S. (1999). Cognitive load and learner expertise: Split-attention and redundancy effects in reading comprehension tasks with vocabulary definitions. *The Journal of Experimental Education*, 67(3), 197–217.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Dr. Rustam Shadiev** is a professor in the School of Education Science, Nanjing Normal University, China. His research interests include HCI for collaboration, speech to text recognition and computer assisted translation technologies, and cross-cultural learning.

**Dr. Ai Sun** is an assistant professor at the School of Humanities and Social Sciences, Harbin Institute of Technology in Shenzhen, China. Her research interests include e-learning and computer assisted language learning.

## Affiliations

Rustam Shadiev<sup>1</sup> · Ai Sun<sup>2</sup>

✉ Ai Sun  
sun\_ai@hotmail.com

<sup>1</sup> School of Education Science, Nanjing Normal University, No. 122, Ninghai Road, Nanjing 210097, Jiangsu Province, China

<sup>2</sup> School of Humanities and Social Sciences, Harbin Institute of Technology in Shenzhen, The University Town of Shenzhen, Nanshan District, Shenzhen 518055, China