Compilation and Translation Review
Vol. 15, No. 2 (September 2022), 199-242
DOI: 10.29912/CTR.202209_15(2).0006

Exploring Automatic Speech Recognition Technology for Undergraduate Sight Translation Training

Gracie Peng

Sight translation (ST), a form of oral translation of written texts, has been widely adopted in both professional and training settings. Translators may also benefit from learning ST skills. This study explored the use of automatic speech recognition (ASR) technology in ST training at the undergraduate (UG) level. We recruited 21 UG students for an elective course of ST to examine the accessible options of ASR for ST pedagogy. We describe the process of incorporating ASR in UG classes and report on students' experiences of using ASR for ST learning. Data were collected through class observations, discussions, and a semi-structured questionnaire. Students were motivated, confident, and competent when applying ASR technology on mobile phones and computers. The questionnaires were used to evaluate participants' insights on the advantages and disadvantages of incorporating ASR in ST exercises and of the ASR platforms they used. This study provided a better understanding of the applicability and limitations of ASR in ST training at the UG level and identified areas for future research.

Keywords: Sight Translation, automatic speech recognition technology, interpreting pedagogical innovation, undergraduate interpreting training

Received: June 15, 2021
Revised: November 1, 2021
Accepted: November 23, 2021

Gracie Peng, Assistant Professor, Department of Foreign Languages and Literature, Tunghai University, E-mail: graciepeng@thu.edu.tw

編譯論叢
第十五卷　第二期（2022年9月），199—242

大學視譯教學結合語音科技之初探

彭貴絹

視譯，是由口譯員將書面資訊口譯出來，常見於口譯服務及口譯教學現場，其訓練也常作為同步口譯的入門。另有研究指出，視譯訓練也有助筆譯員在工作時更著重文本意涵不受用字約束。本研究召集選修大學視譯課程 21 位同學，旨在探討大學視譯教學結合語音技術的議題。為明確研究課題，我們探索了如何選用可行的語音科技，描述如何在合適的階段中融入語音科技，報告參與本研究學生操作後的心得。透過課堂觀察和討論，以及一份半結構式的問卷，收集質性的研究數據。結果發現，學生對於手機及電腦上的語音科技不僅躍躍欲試且十分自信，顯示出此種結合頗能促發學習動機。問卷結果也記錄了參與者使用語音科技練習視譯的經驗，探討其優劣之處，及針對所用的技術平臺提供意見與反思。希望透過本研究，我們可以更加理解語音技術應用在大學視譯教學上的可行性及相關限制，和探討未來研究的可行方向。

關鍵詞：視譯、語音科技、口譯教學創新、大學口譯教學

收件：2021 年 6 月 15 日
修改：2021 年 11 月 1 日
接受：2021 年 11 月 23 日

彭貴絹，東海大學外國語文學系助理教授，E-mail: graciepeng@thu.edu.tw。

Introduction

Sight translation (ST), a hybrid activity involving a written source text being orally translated into another language in real time, has been widely adopted in various professional settings. Some suggest that due to its real-time nature of cognitive process, ST is more appropriately defined as mode of interpreting rather than translation (e.g., Herbert, 1952), and Čeňková (2015) proposes that "sight interpreting" as a better term to encapsulate the specific process of this mode of interpreting. Nevertheless, given the popular adoption of the term of ST professionally and pedagogically, we will continue to use the term for the consistency and clarity of discussion in the current study. ST is often practiced in bilateral meetings where written documents are prepared to be orally translated onsite by an interpreter recruited to offer consecutive interpreting to facilitate communication. In conferences and press conferences, for instance, statements, reports, or relevant documents of interest to the meeting participants are often prepared and sight translated on site. In community settings, such as police stations, courtrooms, doctors' surgeries, similarly, interpreters are often asked to back translate some written records of interviews, witness statements, or medical reports of patients (Čeňková, 2015).

In interpreter training, ST has been used in a number of stages for different purposes. Some schools, for instance, include ST as part of their aptitude tests to evaluate a candidate's ability to grasp and deliver the main message of a text within a limited time. It is also frequently offered as a preparatory exercise for simultaneous interpreting training to encourage fast reaction and comprehension through visual input of the text, as well as flexibility of the oral output of interpreting learners. In addition, ST training plays an important role to prepare students for simultaneous interpreting with text, that they would be encouraged to demonstrate skills of effective

and efficient scanning of a text but not necessarily to identify key information linearly for real-time processing.

ST skills have also been suggested to be useful for translators, that they would learn to appreciate the importance of prioritizing meaning over words. Dragsted and Hansen (2009), for instance, observe the behavioral differences between translators and interpreters. The numbers of word produced per minute by interpreters doing ST (142), translators performing ST (74) and translators working with written text (17) indicate a huge disparity of productivity among the three groups. They also found that the quality of work of the written translation by translators did not significantly surpass that produced by the other two groups of translators and interpreters. Gorszczyńska's (2010) research also shows a number of advantages of combining ST and written translation. In addition to facilitating the efficiency and effectiveness of translation work, ST helps translators to process information faster and therefore reduce translation costs. In a nutshell, the incorporation of oral modality for translators appears to have enhanced work flow without compromising the quality of work, and consequently ST training for translators would be relevant and beneficial.

Sight Translation in Undergraduate Interpreting Classes

ST is useful in developing oral skills and language transfer skills through syntactically restructuring and paraphrasing of the source text (Ilg & Lambert, 1996). Despite its importance in the field and in interpreter training, ST has not attracted much scholarly attention in the existing literature on interpreting and interpreter training. It has neither attracted as much interest than other topics of interpreting have (Čeňková, 2015). A handful of studies have focused on ST training, with an emphasis on European language pairs (e.g., Agrifoglio, 2004; Ilg

& Lambert, 1996; Moser-Mercer, 1995; Weber, 1990). Having said that, ST has also been offered widely in many undergraduate (UG) interpreting classes, either as a standalone course, or part of translation or interpreting classes to encourage students to acquire basic yet essential interpreting related skill components.

Ersözlü (2005), for example, proposes that UG students should be developing the following ST skills, namely reading and comprehension, domain knowledge, detailed reading, dealing with unknown words, chunking skills, and meaning retention. These skills are believed to be vital for one to achieve accurate, coherent and fluent ST performances. In practical terms, Chen (2015) summarizes that in the UG ST class, it is very important to guide students to focus on meaning, not words during comprehension, to be able to identify keywords and segment meaning units during preparation, and to develop the skill of reading ahead. Lee (2012) stresses that, in addition to comprehending a text, reading exercises for students at this level shall be used to reinforce students' capacity of text analysis as one of the major steps in ST skills development. As Li (2015) elaborates further, analytical reading and its sub-skills such as "text analysis, speed reading, identification of main ideas and their links, and concentration" (p. 173) plus the commands of non-verbal skills such as appropriate pauses and intonation, are key elements of successful ST performances (p. 173). He also concurs with Chen (2015) on the significance of acquiring the vital yet challenging skills of segmenting a source text into meaning units and how to link those units into a sensible ST text for UG student interpreters (Li, 2015). Similarly, Krapivkina (2018) reinstates that with chunking skills in place, one can prioritize the hierarchical significance of meaning units within a text, which also serves as one of the keys to successful ST.

The researcher's experience of teaching ST to UG students also corresponds to the previous discussion, that it is of pedagogical advantage to break down the rather complicated process of ST into manageable steps to start with. In addition to

comprehending and analyzing written message of the source language accurately, one shall be able to produce a spoken text of the same message in the target language coherently and fluently. Some skills such as text comprehending, keywords identification, and segmenting a longer text can be taught via paper-based exercises with teacher-led demonstrations in class for students. The acquisition of other relevant skills such as reading ahead and aloud (A&A), oral paraphrasing and message summary, as well as sight translating of various types of texts, for instance, would be facilitated and better supported with other technical resources such as audio-recording and the potential use of voice technology, especially for self-reflection or peer feedback afterwards. Being able to reflect on one's performances effectively and efficiently also helps one to raise awareness of one's strength and weakness, develop a better sense of ownership of one's development and hopefully achieve better learning outcomes as a result.

Automatic Speech Recognition Technology in Sight Translation

Applications of Technology in Translation and Interpreting

The rise of technological advances has opened a number of possibilities for teaching and learning not just in traditional classes where conceptual knowledge is passed on to students, but especially in language classes where skill acquisition is of paramount importance. The same applies to translation and interpreting classes. In both translation and interpreting training at the postgraduate level, for instance, professional preparation has been one of the priorities. Situated learning is therefore critical to actively engage students, so that they are exposed to scenarios as closely as possible to realistic professional environments and requirements. In translation industry nowadays, translators, being freelance or in-house, are

expected to work competently with a number of computer assisted translation (CAT) tools to ensure and enhance their work quality and efficiency. Consequently, to equip students with knowledge and skills of using major translation-related technology, ranging from translation memory, terminology managers and databases, to project management software and many others, is core to many translation schools. In other words, in addition to linguistic skills for translation, developing students' technology literacy, i.e., the ability of working with major CAT tools, for example, often takes a prominent position in the curriculum designs.

In interpreting, the application of technology and information to training and professional practice is heading to the similar direction, too. Drechsel (2019) points out that despite the seemingly uneasy relationship with technology in general, the interpreting industry is catching up, too. The comparatively recent development of computer assisted interpreter training (CAIT) indicates the trend of closer relationship between interpreter training and information technology (e.g., Sandrelli & Jerez, 2007). Professionally, interpreters today have to work in settings where technology plays an essential part, such as the interpreting booth for simultaneous interpreting in a conference as a start, and the rising need for remote interpreting services to cope with the COVID-19 lockdowns in many parts of the world. Terminology management tools, for instance, are also being explored and tested by many interpreters to help with meeting preparation (e.g., Fantinuoli, 2017). The use of digital pens in Consec-simul (Orlando, 2014), a hybrid of consecutive and simultaneous interpreting, has become a possibility for interpreters under certain circumstances, too.

Voice Technology and Automatic Speech Recognition in Sight Translation

The applications of voice technology are also on the rise in our daily life. For one, voice assistants are widely seen on various platforms, such as Siri on iPhone, the Google assistant on Google, and Alexa with Amazon. With voice assistants, one can

search the internet, make phone calls, or even access online systems via giving voice commands. Voice technologies are also applied extensively in business. Companies nowadays have adopted voice technologies to empower their telephone customer services. For example, when customers phone in a bank with queries, they would be asked to answer few questions such as their names, dates of birth or the names of their pets to verify their identities. Some companies have gone further by adopting voice biometrics in this regard, a technology of using callers' voiceprint to confirm their identities.

Voice technology proposed here to facilitate ST learning, on the other hand, refers specifically to automatic speech recognition (ASR), which enables voice-to-text processing to produce a written text. A variety of computer-assisted interpreting (CAI) tools also incorporate the function of ASR. Some interpreters under certain circumstances would use speech recognition (turning voice to text) to reduce the number of notes to be taken for consecutive interpreting, or to enhance the accuracy of simultaneous interpreting with a transcript produced from voice recognition (Gillies, 2019). It is also observed that many translators, when facing a longer translation project with a tight deadline, would produce their first drafts of work by ST with voice recognition software in conjunction with a word processor for a more efficient workflow (Chen, 2015). Relevant studies also suggest that translators with training in ST and voice recognition software may be more effective regarding grasping the meaning of a text as well as producing translation than those who are without such experiences. In other words, such translators are better prepared for the faster workflow in today's market (Chen, 2015). To sum up, it is therefore evident that the combination of ST and voice technology has a positive impact on translators' work efficiency and could be a useful setup to motivate students to be more engaged with similar activities in ST classes.

Automatic Speech Recognition Technology for Sight Translation Training

As discussed previously, as a skill based on a balanced combination of sub-skills and competencies, ST shall be taught with clear scaffolding to support UG students to ensure positive learning experiences and results. One common yet useful pedagogical approach to raise students' awareness of their progress in ST, for instance, is to allow them to listen to the recordings of their performances after completing the assigned tasks. When reviewing oral performances by listening to the recordings, either for self-critiquing or peer-feedback, it is often observed that students' remarks tend to be impressionistic and scattered due to the transient nature of audio input. To allow them to capture more details from the audio recordings to substantiate their remarks, the researcher used to suggest students to transcribe some of the audio recordings of their ST performances into texts.

Even though transcribing an audio recording certainly helps to reveal various aspects of ST learning for students, such as word choices, backtracking, coherence of the target text, it can be very time-consuming and less welcome by some students. During the search for a more efficient yet still effective alternative to conventional transcription, the use of voice technology, i.e., ASR had caught the researcher's attention as a possible solution. By adopting ASR technology, the transcription exercise would be further facilitated to allow students to observe and deliberate their performances more concretely. In the meanwhile, the researcher had been inspired to seek various ways to incorporate such technology more dynamically and effectively to support the teaching and learning of ST. This has become the drive for the study to come to place. Questions on the availabilities and accessibilities of various ASR options, approaches to incorporate ASR technology in ST training to maximize its pedagogical benefits, as well as how the ST students react to such introductions, therefore, serve as the backbone of the current study.

Research Questions

To address the issues concerning applying ASR technology in ST training for UG students in this study, we have set up three research questions as our guidelines. First, we hope to explore accessible and available options of voice technology for ST learning at the UG level. The accessibility and availability of any technology determines the feasibility of its application in various situations. What might work perfectly for one scenario might not be affordable for the users we have in mind. Likewise, what is available and affordable for users might not operate as successfully as expected in scenarios the technology is called for. Therefore, to identify the possible options becomes one of the first questions we hope to answer.

Secondly, like other forms of interpreting, to successfully perform ST of a text requires good coordination of a number of steps and skills. To explore the application of voice technology in ST training, it is vital to consider when and how the technology is to be incorporated to enhance teaching and learning of certain ST steps and skills.

Last but not least, it is of utmost importance to explore users' feedback on using such technology to handle their tasks and whether it facilitates or impedes their workflow and learning. It is also useful to enquire if there are other possible approaches that they would hope to apply the technology to support their learning during the process.

Methods

As an exploratory study, we aim to look at issues around the research questions from various perspectives to cover as much as ground as possible within the context of the current study. Data are gathered from class observations,

interactions with participants both in and outside the class, textual evidence from their class performances and homework, as well as a semi-structured questionnaire.

It is worth noting that as the study is driven and developed to explore more efficient and effective approaches to support and enhance students' ST learning, which corresponds well with the nature of action research, "a research method to enable and support educators in pursuing effective pedagogical practices by transforming the quality of teaching decisions and actions, to subsequently enhance student engagement and learning" (Clark et al., 2020, p. 8). It is vital to recognize that findings of such type of research tend to be less conclusive, as relevant knowledge emerges through collaboration from all participants. Researchers are often the educators themselves, and the role of "educator-researcher" allows them to take advantage of own teaching experiences and subjective knowledge from classroom situations to take part more effectively in their unique pedagogical context (Clark et al., 2020).

In other words, such methodology allows us to gather data from multiple sources to address the questions we set forth for our investigations from various perspectives, which in return allows us to follow the cycle of action, evaluation and reflection.

Subjects

To answer the proposed research questions regarding the pedagogical applications of ASR in the teaching and learning of ST for UG students, we involved 21 students in an elective course of ST as participants. Thirteen of them were seniors and eight were juniors, and all of them have given consent to join the project to explore the use of ASR technology to support their learning of ST.

The semester-long course was offered as a part of the Translation and Interpreting curriculum program at the university level. The objectives of this UG elective course

(Introduction to ST) are to enhance students' basic understanding of ST principles and applications, and their appreciation of different text types and genres, as well as to offer them an opportunity to explore and acquire basic ST skills, i.e., reading ahead, clear articulation, rephrasing, comprehending and analyzing a source text, segmentation, extracting message quickly, expressing ideas logically and fluently without obvious signs of hesitation in the target language. It is also hoped to offer students an opportunity to monitor and reflect on their progress and work closely to form better understanding of their strength and weakness of language skills.

The majority of takers (18) were from a department of foreign languages and literature, while only few (3) from non-language departments. Most students (17) have had taken few basic interpreting classes, while a few (4) were completely new to interpreting classes.

Initial Contact With Automatic Speech Recognition and Preliminary Choices

To explore ASR options for ST pedagogy at the UG level, it is vital to identify available and accessible apps or online resources that students are able to make use of. It is also relevant to consider how such applications can be used in different learning scenarios such as in classroom learning, after class practice, or for their ST assignments.

In a regular language learning lab, there are headsets and microphones for students to listen to audio feeds and to record their performances, which are centrally controlled by the teacher at the console. Students rarely have access to laptops or desktop computers, not to mention other voice recognition tools on site. The only available devices are their smart phones, allowing them to access online information and applications such as voice input apps, or inbuilt function of voice recognition or dictation at the fingertips.

From social interaction with students in classroom, we learned that quite a few (9/21) students have used voice recognition in dealing with some of their daily tasks such as composing short text replies, taking simple notes, giving voice commands to activate AI mobile assistants (e.g., Siri on iPhone or Google assistant on Android phones) to perform internet searches, making phone calls, playing music, setting alarms, and many others. Other students (11), however, are less familiar with the technology and unsure if such functions would work on their phones. Few (4) are concerned that that their phones might not be "advanced" enough to perform such tasks via ASR technology.

Considering the various platforms and setups required by different ASR technology, the participating students were encouraged to bring in devices of their choices (e.g., smartphones, tablets or laptops) to explore ASR technology. The classroom was wired with university Wi-Fi, so that the internet connection was available and reliable. To initiate the process, students were invited to explore what the voice technology on their phones could do in turning speeches to texts. The two chosen options by students are Apple dictation and Google Voice typing, as both are free and common applications on their phones, and both support ASR input in both Chinese and English.

Settings and Questionnaires

The class met once a week for two hours for 18 weeks in a semester. During the first half of semester (week one to nine), students were guided to grasp the basic ideas, skills and principles of ST through class lectures, exercises and classroom discussion. Such arrangement allowed students time to reflect on different challenges facing themselves during the process. For instance, during their weekly assignments, students were asked to practice various elements of ST (such as reading A&A) or other exercises with written texts, such as segmenting longer

texts, identifying keywords and message, same-language oral paraphrasing and summary, to prepare them for the whole process of ST later on.

As action research allows the researcher (also the educator) the liberty to take action to address challenges in the educational context, some of the exercises in the ST classes were modified to be pedagogically suitable for the participants. Take reading A&A for instance, it was introduced as one of the most basic exercises for the UG students at the start of learning ST skills. Different from what Weber (1990) suggests as an activity of reading ahead a source text for an oral production of the target text, the exercise of reading A&A in this study was monolingual (i.e., Chinese to Chinese, and English to English) to demonstrate how one shall allocate one's cognitive resources properly to tend to few different yet relevant tasks almost simultaneously. (i.e., reading, understanding and sounding out the text with appropriate delivery skills continuously). Reading aloud without reading ahead is likely to result in stilted delivery as well as random pauses and hesitation when textual complexity increases. On the other hand, reading ahead without sounding the text out might give students the false impression that it was no different from reading for information, rather than as a preparatory step for ST to happen afterwards. In other words, when reading A&A is done successfully, it does not only enable students to appreciate the significance of fluent delivery for audience, but also facilitate their textual comprehension at the same time.

To encourage students to focus on acquiring those skills, many of the exercises were paper-based with audio recordings when necessary. To review performances at different stages more concretely with textual evidence, students were asked to transcribe their performances verbatim (i.e., word-for-word transcriptions). All the deficiencies such as hesitations, redundancies, back-tracking, and even noises of "ums" and "ahs" shall also be retained, which would help to form a more realistic record of their performances for self-critiques or peer

feedback. Such exercises would demonstrate how intrusive and distracting of a performance with poor delivery and unwanted noises.

ASR was introduced after the Midterm exams from week ten in two sets of exercises (Set 1 and 2) to test the capacities of ASR technology with regard to supporting ST learning. In week ten, the first set of exercises (Set 1) was to incorporate ASR with reading A&A and segmenting (in both Chinese and English). The second set of exercise (Set 2) in week 12 was designed to engage students in using ASR to record their ST in Chinese (Table 1).

Table 1

Setup for Incorporating ASR in ST Exercises

Exercises	Set 1	Set 2
Timing	Week 10	Week 12
Steps	1. Read ahead & aloud 2. Segment the source text	1. Warm up - read ahead & aloud 2. Segment the source text 3. ST with ASR (English to Chinese)
Objectives	1. Explore and assess ASR options 2. Read ahead/aloud with ASR 3. ASR to raise awareness of one's delivery	1. ASR to record one's ST in text 2. Explore pros and cons of such practices 3. Explore ways of treating ASR results
Materials	Appendix A	Appendix B

The arrangements aimed to offer students opportunities to taste and test ASR technology using both their Chinese (A language) and English (B language). As working from Chinese to English has proven to be considerably more challenging for this group of UG students, they only worked from English into Chinese, their mother tongue, for the ST exercises with ASR, instead the other way round.

In addition, both sets of exercises involved working with the same texts at least twice if not more times from different approaches. It was to ensure that

students focused on practicing with the target ST skills, rather than on dealing with new textual challenges, being it cognitively, linguistically or phonetically.

In summary, Set 1 exercises were designed to apply ASR technology to support students to acquire the sub-skills of ST; Set 2 aimed to offer students the opportunity of incorporating ASR into the full set of ST, involving reading A&A a written text and orally translating of the text almost simultaneously, and in the meanwhile, recording one's performances via ASR.

Regarding the exact amount of time and effort invested in ASR exploration, it was difficult to quantify due to individual differences among participants such as motivations, prior technical proficiency, as well as learning styles. In addition, to maximize the coverage on using ASR for ST learning and other language exercises, students were encouraged to explore freely, so that they could share their observations in class discussion. With such disparities, it is inevitably a challenge to measure the correlations between the introduction of ASR and students' learning outcomes. Yet their explorations, observations and experiences have become the major source of data which also enrich our discussion and investigation as a result.

To better record students' views and experiences on using ASR technology in the two sets of exercises over the weeks, we devised a very simple semi-structured questionnaire to collect their experiences and record their feedback (Appendix C). On the questionnaire, they were asked to reflect on three aspects of ASR application in ST exercises: (a) the advantages of such application, (b) the disadvantages or challenges, and (c) comparison of different platforms of using ASR applications as well as reasons for their personal preferences.

Results and Discussion

Automatic Speech Recognition Options Explored

Before adopting ASR in ST exercises, we chose two available and affordable options (ideally free) for the participants to explore. There are paid products, proven very powerful and effective such as "Dragon Naturally Speaking," which is highly recommended by language professionals such as journalists, writers and even translators for their full range of functions with commendably high accuracy rates. Participants in the study, however, would find such options unaffordable and inaccessible. The two chosen options for students at the initial stage were Apple dictation and Google Voice typing, as all participating students had access to either system. Students were encouraged to explore other options on available platforms, i.e., on a mobile phone or a desktop/laptop computer. The comparison of the two ASR systems is summarized in Table 2.

Table 2

Comparison of Apple Dictation and Google Voice Typing

ASR systems	Apple dictation	Google Voice typing
Languages	Support both English & Chinese	
Accuracy	Mostly acceptable, with errors due to reasons in both languages ☑ Better with Chinese recognition than that of Google	Mostly acceptable, with errors due to reasons in both languages
Platforms	☑ Smart phone: iOS ☑ Computer for Mac/MacBook & iPads	☑ Smart phone: Android ☑ Computer – on Google Docs
Online/Offline	☑ Online Offline available on mobile phones for English	☑ Online Offline available on mobile phones for both languages
Punctuation	☑ Voice commands for both languages	☑ Voice commands for English only

Apple Dictation vs. Google Voice Typing and Platforms

To begin with, it is noted that both Apple dictation and Google Voice typing are known and used by participating students. Nine of them had started exploring the in-built ASR technology on their mobile phones, be it on iOS (for iPhones) or Android (for the majority of non-iPhones) system before taking part in the course of ST. It is observed that users all were quite satisfied with the accuracy rate of recognition, and efficiency of ASR text production of both systems.

This does not mean, however, that the ASR is free of errors or the ASR texts flawlessly prepared as a result. One major reason for recognition errors in both Chinese and English was users' mispronunciations. When working with ASR in English, for instance, students reported that many of the mistakes were made due to wrong stress of certain vocabulary, or mumbled annunciation of challenging vowels and consonants for EFL users like themselves. Figure 1, for example, shows how a student followed the instructions given in Set 2 exercises to check her ASR output closely and highlighted the disparities with different color codes to indicate corresponding reasons: dark gray for technical problems while light gray for human errors.

Figure 1

Example of Student's Reading A&A With ASR and Error Analysis

So far it seems that the graduates have kept their advantage. A changing jobs market has generated more opportunities for graduates and people who went to university are likely to be earning more and are less likely to be unemployed.	So far it is seems that the bridges have kept their advantage a changing jobs market have generated more opportunity is for a bridges and people who went to university are likely to be earning more in our life to be an employee.

Other issues such as wrong parsing of words and phrases due to pauses, hesitations, backtracking, or stilted delivery were observed in both languages, too. In terms of text production, students reported that they were impressed by the almost real-time production of texts they orally produced. For instance, one shared in class that, "it was exciting to see what we just said in actual words straightaway," but "it's also shocking to see all the nonsense I produced" (Student 5).

Few (3) reported that iPhone dictation would only function well with strong and steady internet connection. A shared complaint about both systems is the lack of automatic punctuation with ASR texts. To produce a punctuated text, one must either use verbal command to insert punctuation marks or to create new lines during their oral production of the text.[1] Users can also add the punctuations back to the ASR text afterwards if they hope to use the text for other purposes.

Regarding platforms for incorporating ASR to ST exercises, the most noticeable constraint with ASR on mobile phones (both Apple dictation and Google Voice typing) is the one minute limitation, that the ASR function switches off automatically after around one minute and needs to be restarted if one wishes to continue. It could be very frustrating for students when they had to pause to organize their thoughts and construct their sentences in ST, or when students are half way through their reading A&A exercises. ASR on a desktop or a laptop computer, especially with Google Docs, in contrast, did not have such limitation and students could therefore work without the pressure of being cut short during the process.

Languages

The chosen ASR options need to be able to support both Chinese and English inputs, two working languages for the group of participants in the study. It is

[1] One shall speak out the verbal commands such as "comma" for "," within a text. To produce "Hello John, the check is in the mail!", one shall say "Hello John comma the check is in the mail exclamation mark." Retrieved from https://support.apple.com/en-us/HT208343

understood that in general, English ASR input is fully supported by both Apple dictation and Google Voice on various platforms, both online and offline, a welcoming feature for many when reliable internet connection was an issue.

Google Voice input is widely accessible in Android system on mobile phones with both Chinese and English, online and offline. Supporters of Google Voice also suggested that the system could deal with mix-code input, that when they had to use both Chinese and English in one sentence, the system could mostly detect such mixture and offer sensible outputs as a result. Chinese input via ASR, however, appeared to be less ideal. Apple dictation on iPhones, on the other hand, only supports offline input in English, but not in Chinese, so that it could be very frustrating for users when internet connection is not reliable.

Punctuation

To ensure a text makes sense, punctuation plays an important role. In other words, without proper punctuation in places, it would be hard for human minds to comprehend a text efficiently and correctly. It is the same with ASR systems, that a large part of intelligent recognition relies on how a text is punctuated. Through the initial exploration, it was realized that neither of the suggested ASR systems offered automatic punctuation. Google Voice typing, for example, did not support Chinese punctuations by voice typing, which could only be added via touch typing afterwards. Apple dictation, on the other hand, allowed verbal input of punctuation in both Chinese and English. Nevertheless, using verbal commands within oral input of speech, or ST performances, inevitably, creates unnatural interruptions to the flow of speech and thoughts.

According to some (7) participants, however, the need for post-editing offered them an opportunity to review their performances more closely, which therefore became a useful step for self-reflections. With a unpunctuated text containing errors

of recognitions in places, they needed stay actively engaged with not only the text, but also the process they just had gone through with their ST exercises. In addition, few (3) suggested that, in place of the full range of punctuation marks, they simply use "new line" to segment a text into manageable meaning units as the first step to prepare the text for ST. It is especially useful during the stage of reading A&A.

Students' Feedback on Automatic Speech Recognition Technology in Sight Translation Learning

From the questionnaires and classroom discussion, while there appeared to be clear advantages of such application, there are a number of concerning aspects, too. Students also shared their views on the choices of different platforms and systems of ASR application.

Advantages and Positive Discoveries

The majority of students were very positive about the benefits of ASR in enhancing their ST learning in several areas: increasing work efficiency, raising awareness of aspects for improvement as well as encouraging learning motivations.

A great majority of students (18/21) indicated that ASR helped them to save a lot of time in transcribing their oral performances into text, and with the acceptable rate of accurate recognition, the main message could be largely recorded. Compared with their previous experience of transcribing audio recordings of their ST by lengthy repetition of listening and typing, ASR certainly helped to save a lot of time and the result was immediately viewable. The text could also be edited and saved as files straightaway. Few (5) of them observed that the ASR technology appeared to be smart enough to ignore some of their hesitations and intrusive noises such as "ums" and "ahs" during their performances, so that the majority of ASR text contained complete and sensible sentences. Such texts allowed them to

examine the coherence and concision of their performances more effectively, rather than following some fleeting impressions without any solid evidence. Although ASR results were not without flaws, the majority of students (16) were pleased with the overall accuracy. Few (2) even suggested that texts were 80% correct, which could serve as a useful draft should one hope to use the ASR text to understand the main message of the source text.

In addition, many (12/21) suggested that the use of ASR technology in ST exercises helped to raise their awareness of various aspects of their performances, such as their delivery skills, their use of language in terms of grammatical correctness and concision, as well as the overall coherence of the text being produced. Due to the nature of ASR, that it takes time for the application to recognize the audio input before turning into text, students reported that they deliberately aimed to pace themselves properly to enhance the clarity of their pronunciation and intonation in order to maintain or increase the accuracy rate of ASR.

For example, one participant shared in classroom discussion that, "it's useful to have ASR as a tool to check my pronunciation and intonation, and I felt great when it could recognize what I said correctly" (Student 3). Another shared that, "I learned that I had some mispronunciations for so long without myself knowing, that I finally got them right to 'pass' the check of ASR" (Student 11).

They also stated that the visual feedback of ASR text served as a useful reminder for them to pay more attention to the text structures, word choices and expressions, as well as grammatical agreements when practicing ST. "The ASR text honestly reflected my careless and ungrammatical expressions. I then became more careful during my ST exercises and tried hard to produce sensible sentences, or at least not to create so much nonsense" (Student 8), said one participant.

Few (3) students summarized that slowing down for better accuracy rate of ASR has actually served as a useful breathing space, that students could use to

organize their ideas and output better without being too nervous and careless when recording their ST.

Quite a few students (7) also claimed that incorporating ASR to their ST exercises has been an interesting and motivating experience. They were excited to know that the technology available could be applied to their learning, and very keen to explore the pros and cons of the ASR technology and to test its limits, too. The welcoming attitude has been a useful driving force for students to be more engaged with their learning in general.

Disadvantages and Challenges

The application of ASR in ST exercises has not been impeccable, either. The major complaints are mismatches of recognition, a number of technical problems as well as distraction and stress as a result.

The biggest frustration of using ASR in ST exercises was the problem of mismatches, or recognition errors, that the majority of students (15/21) complained about. Some causes were traceable, that students realized that their pronunciation of certain words or expressions was not accurate or clear enough for the tool to achieve accurate recognition in both languages. Some (8) have also observed that proper names (e.g., names of organizations) or fixed expressions such as the four-word idioms in Chinese often failed to be recognized accurately. When the input was not "recognizable" for the ASR system, it would be automatically changed into something totally irrelevant to the context, which could surprise the users afterwards. Others (4) pointed out that the ASR system they used did not seem to be able to distinguish a pause in the middle of a sentence from a stop of an utterance, especially when they were hesitating or their delivery was monotonous.

Other noticeable technical challenges include the one minute limitation, abrupt switch-off of the ASR tool, the lack of automatic punctuation and others.

Five students were specifically frustrated by the in-built time limit of ASR on their mobile phones, that the function of ASR switched off automatically after around one minute. Eight reported that the ASR system shut down if they paused for too long in the middle of their delivery, or when there was something "beyond recognition" by the system. For example, one student suggested that:

> I used my iPhone to do my last (home)work and found out that it had some flaws. For instance, the process of recognition was so short that I had to pause a lot and clicked on the bottom to continue. But this time, with Google Voice, there's no such a problem. It can go on and on until I stop. (Student 17)

In addition, seven found the lack of automatic punctuation or the need of inserting punctuations by speaking out the commands rather inconvenient and unnatural. For example, one stated that "speaking out all the punctuations feels very unnatural, as nobody would speak like that" (Student 10). Yet it was not totally without benefits, as claimed by another, that:

> Having to go back to the whole unpunctuated text after recording and to add back the punctuations is quite tiring and confusing sometimes, but I suppose it's not a bad idea to see what I had actually produced and check how I could have done better. (Student 3)

A few students (4) pointed out that to ensure better results of ASR, one shall work in a quiet place, use a dedicated headphone/microphone and stay online. It was problematic for students in shared accommodation where quiet corners were few and the internet connection was not always strong and reliable. Others (4) also discovered that some ASR (e.g., iPhone dictation) was not able to deal with mix-codes, that when they inserted few English words in a Chinese text, the system would guess the English sounds it received and turned them into random Chinese expressions.

When working with the ASR text from their ST exercises, students reported that the need for repair (recognition errors, omission of recognition, and sentence parsing) and editing (punctuations) could be daunting at first when they were new to the application. Few (3) have also observed that compared to audio recording, the ASR text failed to record the prosodic features of their ST, such as clarity of utterance, fluency, intonation, pace, and so on. As a result, some texts might appear fragmental and less connected without apparent textual connectives, which is an expected limitation of written representation of an oral text.

The technical difficulties as discussed inevitably posed some challenges to students while they incorporated ASR to ST learning and practice. A few (5) claimed that it was distracting and some (2) found it stressful to have to deal with ASR while having to be busy with the task of ST already. Some (3) suggested that they were "worried and nervous" that ASR would suddenly stopped, especially when they were struggling to sight translate, or their utterances might be longer than the one minute limitation. Some (4) also found it rather disappointing to realize that the ASR failed to function in the middle of their work, that only half or part of their performances was recorded.

In summary, by exposing students to the application of ASR in a variety of ST exercises, we are able to conclude that the pros and cons are equally vital in keeping us informed of the strengths and limitations of such technology in the teaching and learning of ST.

Comparing Platforms

When asked to compare the two platforms of ASR application, a mobile phone or a laptop/desktop computer, the results have been quite interesting, too (Table 3).

Table 3

Comparisons of ASR Platforms

Platform	Supporter (number)	Pros	Cons
Mobiles	iPhone (11)	1. Familiarity 2. Better recognition 3. Portability 4. Better connection (Wi-Fi) 5. Fast ASR results	1. One-min cut-off 2. Sudden cut-off when connection's weak 3. No offline input for Chinese 4. No mix-code input
	Android (3)	1. Support mix-code input 2. Support offline ASR for both languages with downloaded packs	1. Sudden cut-off 2. No punctuation in Chinese 3. Slow with errors due to older system/phone
Laptop/ Desktop	Google Voice in Google Doc (5)	1. Bigger screen to see context 2. Easy to edit (errors/ punctuation/save) 3. No sudden cut-off 4. Support mix-code input	1. Headphone/Mic for better result 2. Less accurate for Chinese recognition 3. Slower but fuller ASR text 4. Confusing interface
	Apple Memo voice (2)	1. Support offline ASR 2. Good accuracy for both languages	None mentioned

The majority of students (14/21) indicated that they preferred to use ASR on their mobile phones over computers. Interestingly, almost all of them (11/14) were iPhone users. They all suggested that due to their familiarity with their phone and frequency of using them, the ASR system seemed to be able to recognize their input faster and more accurately than Google Voice typing on a computer. They appreciated the portability of the device, which facilitated the use of ASR. Some claimed that the internet connection via the mobile phone appeared to be more reliable than that of their computers in shared accommodation. Few (4) phone users, however, were disappointed by the poor performances of ASR on their mobile phones.

In the meanwhile, a few (4) students favored using Google Voice typing on the computer. For one thing, they preferred the bigger screen for a better view of the whole text and its ability of saving the ASR document as a file straightaway. Another advantage of Google Voice typing, according to its supporters, was that it would not stop functioning or cut itself off like other ASR on mobile phones. It was also easier to insert punctuation or correct errors afterwards when necessary. The Google Voice typing supporters were all pleased with its accuracy rate of recognition in English, but less so with Chinese. One student suspected that the Google system might not respond to Taiwanese Chinese as well by stating that, "Google finally functioned better in Chinese after I deliberately put on the Beijing accent" (Student 6).

Three also shared their observations of the comparably slow appearance of text via Google Voice typing, that:

> It took longer for Google Voice typing to present the text of my work after I finished talking. At first, I was worried that it was not working, but it turned out that it just needed more time to process longer input. The results were actually surprisingly satisfactory and complete. (Student 6)

Two shared their positive experiences of using Google Voice typing on their Android phones, that Google Voice seemed to be able survive well offline the needed language packs downloaded. They also discovered that Google Voice typing could deal with mix-code input, that when they had to use both Chinese and English in one sentence, the system could mostly sense such mixture and offer sensible outputs. One (Student 9) reported that he had been using Apple's Memo voice on his iPad, which allowed him to use ASR offline and the accuracy for Chinese recognition was satisfactory. The other two used Apple laptops (Mac) and reported their satisfaction with its inbuilt ASR.

Best Use of Automatic Speech Recognition Technology

Having explored the two suggested options of ASR applications as well as students' experiences and feedback of applying the technology to their ST exercises, it is therefore vital to discuss how to incorporate the ASR technology to better support different stages of learning ST, its respective strengths and space for improvement. As explained in Settings, we devised two sets of ST exercises (Set 1 and 2) for students to incorporate ASR technology to explore and test the technology in various scenarios.

Set 1 Reading Ahead and Aloud

One of the very first steps of ST is to read through the source text for comprehension. Reading, in general, is often done in silence for information retrieval, while reading in ST is about comprehending the intended message of the source text before communicating the textual message to the target audience orally. To enhance better comprehension of a text, reading A&A as well as segmenting the text into meaning units, also known as chunking, are among the indispensable exercises to be introduced in ST classes.

Reading A&A can be taken either as one activity, or two distinctive yet highly related monolingual skills, as one of the warm-up activities for ST training. To read A&A in ST, one shall orally deliver a written text accurately, clearly and fluently without obvious signs of hesitation, random pauses or backtracks. To achieve this, one needs to be able to coordinate several skills more or less at the same time. Being able to recognize and pronounce the words and phrases in a text, for instance, is only the start. In addition to understanding how various parts of texts are strung together to form a text, one shall also be able to reproduce the textual meaning via oral delivery of those words, phrases and sentences with appropriate prosodic features such as pauses, stress and intonation. Last but not the least, by

combining all the steps together, one strives to comprehend the constant flow of textual input while reading out the new textual segments almost at the same time. In other words, reading A&A even monolingually is a version of simultaneous activity, involving a synchronization of eye-movement to scan through the text, effective cognitive effort for textual comprehension, as well as fluent verbal output of the textual segments virtually simultaneously.

Reading A&A can be used to engage students to work with both of their working languages: Chinese and English in this study. To make the task meaningful but not overly challenging for the participants, we chose few texts in both Chinese and English on subjects of a general nature at between 200-400 words in length each as both classroom practices and homework exercises (e.g., Appendices A & B). When in class, the task of reading A&A was firstly demonstrated by the trainer without any technical support. Students then were given opportunities to practice on their own and with their peers. Their performances were recorded for them to listen to afterwards. It was reported that listening to their own voices through recording was an unfamiliar but interesting experience. Due to the transient nature of audio recording, however, it was difficult to reflect on the performances meaningfully. As a result, the self-evaluation and peer feedback tend to be impressionistic due to the lack of clear trace of evidence to support their observations.

Reading A&A With Automatic Speech Recognition to Enhance Delivery and Text Segmentation

With ASR, students were able to visualize their oral output and identify space for improvement. Despite occasional technical challenges, most participants were motivated to see their verbal performances in written forms. They also reported an increased level of awareness of the prosodic features of their reading aloud. For example, to ensure a more desirable ASR output, i.e., a higher rate of accurate

recognition, they started to pay attention to their pronunciation, intonation, pauses, stress, and paces of delivery in both languages. Some reported that it was easier in Chinese, their mother tongue, than in English. Others had come to realize that their reading A&A in English was in urgent need of improvement, and had taken ASR as a useful reminding tool for self-improvement. In addition to enhancing their delivery skills, it was noted that students would slow down their delivery to facilitate ASR results, which also in return allowed themselves space to read ahead efficiently. When one can read ahead efficiently, one would be able to proceed to the next vital step for ST: segmenting the ongoing text into manageable meaning units. In other words, the application of ASR at the initial stage of ST training also helps to enhance students' skill acquisition of text segmenting.

To segment a text into meaning units, one needs to comprehend the message well enough to draw clear boundaries between meaning units. Punctuation marks, for instance, are useful indicators under normal circumstances to guide a reader to comprehend a text. Texts sight translators need to work with could sometimes be very challenging with long and complicated sentence constructions, needing more than punctuation marks for one to dissect and digest the message accurately and efficiently. It would be less of an issue for native speakers of the working language, or people with strong language competencies and relevant knowledge of the subject matters. It could be, however, very demanding for non-native speaker of the language with less solid understandings of certain subject matters, i.e., some participants in this study.

When one needs to read out such texts, one would deploy other means, such as intonations and appropriate pauses, to mark the progress of a message accordingly. If one fails to apply those skills successfully, the flow of reading-aloud would be disrupted with apparent pauses at wrong places or backtrackings. From the two sets of exercises of the study, it was interesting to observe that when students were able to read A&A well with appropriate intonations, stress, and

pauses to segment the message correctly and efficiently, their comprehension of the text was more likely to be in place.

In short, one vital benefit of using ASR was to encourage better reading A&A, consequently enhancing better delivery, segmenting and comprehension skills for students.

Set 2 Sight Translation With Automatic Speech Recognition (English to Chinese)

In addition to reading A&A, we have also incorporated English-Chinese ST exercises to explore how ASR could facilitate and support students' ST learning. During the two phases of exercises, i.e., first on reading the source text A&A in English, and the other on ST of the text from English into Chinese with ASR, students were asked to record their voice performances at the same time with another device. The audio recordings serve as useful references for the less satisfactory ASR results due to technical issues as described earlier.

From class discussion as well as their reflection notes, the majority of participants were excited and positive about the experiences of working with ASR on their ST exercises despite the known technical challenges. They were more impressed than disappointed by the general accuracy rate of their chosen ASR tools as well as the usefulness of having textual recordings of their verbal performances. To ensure better recognition, many have also reported that they paid extra attention to their delivery skills such as articulation, pronunciation, intonation, and suitable pauses when producing their ST in Chinese. Some suggested that they would make extra efforts to organize their sentences better to avoid verbal fillers such as "ums" and "uhs" and unwanted backtrackings in their performances. In other words, ASR seemed to have helped to raise self-awareness of their own performances in both language as well as delivery skills. It is, however, by no means to suggest that ASR helped to improve the quality of students' ST performances like magic. A few

participants with space for improvement on basic skills such as comprehending the source text found working with ASR in ST exercises distracting and somehow frustrating. For instance, few participants reported that using ASR inevitably impact on their limited cognitive and linguistic resources for the ST tasks at hand.

To summarize, the application of ASR in the ST exercises has useful pedagogical implications such as raising students' awareness of their linguistic outputs as well as their delivery skills, but it could also cause confusion and be counter-productive when students were less prepared or familiar with the technical setups and details.

Automatic Speech Recognition to Support Reviewing Sight Translation Performances

After completing Set 2 exercises, we looked into ways to exploit the ASR texts of ST performances. One was to them to support students to review their ST performances in addition to the usual audio recordings. When evaluating ST performances by listening to audio recordings of their own performances or their peers', students tended to offer over generalized feedback such as "I think I covered most of the ideas, but I am not very sure to be honest" (Student 2), or "he sounded ok but something was not quite right" (Student 17).

It is also observed that during ST, when being cognitively occupied by various tasks almost simultaneously, many students would find it challenging to monitor their own performances. The addition of the ASR-generated texts of ST provided students with useful visual evidence for closer and more reliable critiques. Although ASR texts at the current state still contain technical faults as detailed earlier and errors from students' performances such as mispronunciations and backtrackings, they could, on the other hand, record and reflect some major ST infelicities such as message distortions, clumsy choices of terms and phrases, or incoherent utterances. Such evidence is greatly welcome by students for the purposes of self-critique and peer-feedback. Some (4) were delighted to report that

once an ASR text was cleaned up and rectified, it could be a useful translation draft of the corresponding source text. Few (2) suggested that they would try to adopt the combinations of ST and ASR technology for their translation exercises, which corresponds to what Chen (2015) suggested, that the combination of ST skills with ASR technology might enhance translators' work efficiency.

Yet it is sensible to recognize the realistic challenges of such arrangements. For instance, collecting one's ST performances in both spoken and written forms requires rather complicated technical coordination. The current technology of either mobile phones or computers would not allow people to use ASR and audio recording concurrently on one device. One needs a device for audio-recording, and another with ASR functions; in the meanwhile, one needs to read and comprehend the source text closely for ST. Few (3) students found such arrangements confusing, distracting and stressful, especially when they also had to extra attention to the ASR device to prevent it from shutting down or other malfunctions.

To address the conflicts, some (2) motivated students suggested that when they worked on the ST assignment, they would focus on the source text only without paying attention to ASR text to avoid distraction, and to ensure better work flows. The audio recording of their work would later be fed into some online ASR tool to turn their work into text, should they need a written record of their work for reflections and reviews.

Other Advantages and Possible Applications of Automatic Speech Recognition

It is realistic to recognize that the ASR-generated texts wound need considerable amount of post-editing to function for general purposes. For example, ASR texts would not be able to capture extra-linguistic features such as intonation and stress like audio-recordings do, yet neither could most written texts. ASR texts,

however, could complement the audio recording of the same piece of ST performance for students to review their work from different perspectives to inform a more holistic assessment. The addition of ASR texts, therefore, is believed to be pedagogically beneficial for the teaching and learning of ST.

In addition, few (3) enthusiastic students suggested that ASR on their mobile phones be of great use for other language exercises such as oral summaries in either mono-language such as in English or Chinese, or cross-language settings, e.g., from English to Chinese, or vice versa. When they incorporated ASR in their exercises, the appearance of the ASR text would remind them of the significance of concision, proper word choices as well as logical connections between ideas and sentences. For instance, one participant remarked that, "I feel motivated and a sense of purpose when seeing what I said appearing as actual words on the screen with voice recognition" (Student 19). Another reported that, "I feel that I have to be responsible for what I say, so I try very hard to not to produce too much nonsense or hesitate too much" (Student 21).

It is also noted that the introduction of ASR to ST classes has also driven up students' learning motivations. Students nowadays in general are mostly confident and competent of exploring new technology and applications. Many (16/21) expressed strong interests in incorporating new technology at their hand with classroom learning and welcomed the opportunity to test the possibilities.

Conclusions

Summary

The project set out to explore possible applications of available ASR technology for UG students to facilitate their learning of ST. It is clear that the

technology under discussion has progressed remarkably over the years and there are accessible and affordable options like Google Voice and Apple dictation on different platforms for users. We explored the options from various aspects such as platforms, language-specific issues as well as their current limitations. From students' feedback, it is useful to know that the performances of ASR from both Google Voice as well as Apple dictation are generally acceptable and satisfactory, and it has been of help to examine their respective pros and cons and how they compared in different tasks. As to when and how to incorporate the technology to ST learning, we have investigated few major settings and discovered some advantages of introducing ASR to students' learning in general. It is noted that such introduction has raised students' awareness of their performances in delivery skills such as their articulation and intonation from tasks such as reading A&A, and how those skills would impact on other vital skills in ST such as message segmentation for better comprehension. It also has offered them the opportunity to examine and reflect on their ST output more closely with textual evidence. In general, despite few inherent technical deficiencies to be resolved hopefully soon with technological advancement, the participating students have expressed strong interests in and motivations for the combination of learning and technology they have at hand. In a nutshell, this study has addressed its research questions and fulfilled its objectives by identifying workable protocols for the participants to explore, experiment and experience the various approaches of working with available ASR technology to support students' ST learning and practices.

Contributions

The paper portraits an honest picture of a ST elective class in university where there were various constraints of resources and pedagogical challenges facing a group of less homogenous students. Yet more importantly, we explored how we

could incorporate available and accessible ASR technology to facilitate the learning and teaching of ST under the circumstances. Through the steps designed by the study, the students were offered the opportunity to explore the possibility of combining the ASR technology available with their ST learning. During those phases and tasks, they were encouraged to investigate the pros and cons of such arrangements and other potential applications. For instance, ASR has facilitated the efficiency of transcribing voice recordings for self-reflection, and helped to raise students' awareness of their delivery skills and how those skills could enhance textual comprehension for both the sight translators and the end-users of the ST. It also inspired some participants to explore other platforms and pedagogical applications of such technology.

Their feedback and observations also served as useful basis and reminders for themselves when they hope to adopt ASR further. The experiences allowed the participants to develop better judgements of when and how to adopt ASR for certain tasks and when concessions would be needed to address the limitations of current technology. In other words, the process has consequently encouraged learning autonomy by raising students' awareness of not just their own performances but also of resources available and how to tailor the arrangements to maximize the advantages of such incorporation.

Limitations

As with the majority of studies, the design of the current study is subject to limitations. As a qualitative study, data generated from the participating students' views and feedback might appear to be of subjective nature and the conclusions provisional. It is, however, important to note that as an exploratory study following the basic principles of action research, the results gathered have successfully addressed the research questions set out at the start of our work. Despite the limited

size of subject population, the multifaceted data collected have served useful clues to inform the key areas of investigation. In addition, the heterogeneous backgrounds and competencies of participants posed challenges for the researcher in terms of validating the effects of ASR application in ST exercises. For instance, some are more technologically proficient than the others; while their linguistic or ST sub-skills might not be as ready as their counterparts, or vice versa. With regard to when and how to adopt ASR technology in ST learning, we have experimented with a few relevant setups, but have not exhausted all the possibilities due to the limited scale of the project. For instance, should there be sufficient fund to acquire purposely-built ASR tools such as Dragon products, we would be able to streamline the workflow better without the limitation observed in current study. In short, with more resources available in the future, it would be of great advantage to investigate more systematically on how learners could maximize the benefits of ASR technology in learning ST.

Future Work

Undoubtedly, results of the study would serve as a valuable basis to inform further investigations and pedagogical developments. To deepen the understanding of ASR on facilitating ST learning for UG students, for example, it might be useful to eliminate as many technological limitations of ASR as possible by choosing one reliable tool for participants to work with. Instead of generating rich, yet rather diverse feedback and perspectives, we would be able to collect empirical data for more systematic and effective analyses to establish better understanding of this field of discussion. In addition, it would also be of great use to work with participants of more advanced levels of readiness for language profession of translation and interpreting, in order to explore the combination of ASR with their professional life and its impacts on their work flow and efficiency. All in all, it is

hoped that with the current study, we have brought technology and UG interpreting learning a step closer to address not only the pedagogical needs, but also to initiate a positive beginning to understand and explore further how technology and language professions interact today and tomorrow.

References

Agrifoglio, M. (2004). Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting*, *6*(1), 43-67. https://www.doi.org/10.1075/intp.6.1.05agr

Čeňková, I. (2015). Sight interpreting/translation. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies* [VitalSource Bookshelf version] (pp. 374-375). Routledge.

Chen, W. (2015). Sight translation. In H. Mikkelson & R. Jourdenais (Eds.), *The Routledge handbook of interpreting* (pp. 144-153). Routledge.

Clark, J. S., Porath, S., Thiele, J., & Jobe, M. (2020). *Action research*. New Prairie Press. https://newprairiepress.org/ebooks/34

Dragsted, B., & Hansen, I. G. (2009). Exploring translation and interpreting hybrids. The case of sight translation. *Meta: Translators' Journal*, *54*(3), 588-604. https://doi.org/10.7202/038317ar

Drechsel, A. (2019). Technology literacy for the interpreter. In D. B. Sawyer, F. Austermühl, & V. E. Raído (Eds.), *The evolving curriculum in interpreter and translator education: Stakeholder perspectives and voices* (pp. 259-268). American Translators Association.

Ersözlü, E. (2005). Training of interpreters: Some suggestions on sight translation teaching. *Translation Journal*, *9*(4). https://www.translationdirectory.com/article755.htm

Fantinuoli, C. (2017). Computer-assisted preparation in conference interpreting. *The International Journal for Translation & Interpreting Research*, *9*(2), 25-38. https://www.doi.org/10.12807/ti.109202.2017.a02

Gillies, A. (2019). *Consecutive interpreting: A short course*. Routledge.

Gorszczyńska, P. (2010). The potential of sight translation to optimize written

translation: The example of the English-Polish language pair. In O. Azadibougar (Ed.), *Translation effects. Selected papers of the CETRA research seminar in translation studies 2009* (pp. 1-12). Katholieke Universiteit Leuven, Centre for Translation Studies. https://www.arts.kuleuven.be/cetra/papers/files/paula-gorszczynska-the-potential-of-sight.pdf

Herbert, J. (1952). *The interpreter's handbook: How to become a conference interpreter*. Librairie de l'Université.

Ilg, G., & Lambert, S. (1996). Teaching consecutive interpreting. *Interpreting*, *1*(1), 69-99. https://www.doi.org/10.1075/intp.1.1.05ilg

Krapivkina, O. A. (2018). Sight translation and its status in the training of interpreters and translators. *Indonesian Journal of Applied Linguistics*, *7*(3), 695-704. https://www.doi.org/10.17509/ijal.v7i3.9820

Lee, J. (2012). What skills do student interpreters need to learn in sight translation training? *Meta: Translators' Journal*, *57*(3), 694-714. https://www.doi.org/10.7202/1017087ar

Li, X. D. (2015). Designing a sight translation course for undergraduate T&I students: From context definition to course organization. *Revista Española de Lingüística Aplicada*, *28*(1), 169-198. https://www.doi.org/10.1075/resla.28.1.08li

Moser-Mercer, B. (1995). Sight translation and human information processing. In A. Neubert, G. M. Shreve, & K. Gommlich (Eds.), *Basic issues in translation studies: Proceedings of the fifth international conference* (pp. 159-166). Kent State University.

Orlando, M. (2014). A study on the amenability of digital pen technology in a hybrid mode of interpreting: Consec-simul with notes. *Translation and Interpreting*, *6*(2), 39-54. http://www.trans-int.org/index.php/transint/article/download/301/165

Sandrelli, A., & Jerez, J. D. M. (2007). The impact of information and communication

technology on interpreter training: State of the art and future prospects. *The Interpreter and Translator Trainer*, *1*(2), 269-303. https://www.doi.org/10.1080/1750399X.2007.10798761

Weber, W. K. (1990). The importance of sight translation in an interpreter training program. In D. Bowen & M. Bowen (Eds.), *Interpreting: Yesterday, today, and tomorrow* (pp. 44-52). John Benjamins.

Appendix A

Texts for Reading A&A

Automatic Speech Recognition Technology for Sight Translation (1)

Dear all,
Please use both texts (Chinese first then English) below to practice **Reading A/A** with your ASR APP (e.g. iPhone and Google Voice). Please keep the ASR text the way it comes out without further editing.

Reflection questions:

1. Please compare and contrast the ASR results with the original texts. Are you happy with them and why?
2. Do you find it harder or easier to work with either language and possible reasons for that?
3. How has the ASR technology helped or did not help with your work?

Chinese Text for Reading ahead and aloud

(from https://kknews.cc/zh-tw/health/6anzlyq.html)

糖可說是現代最常見的食品添加物，從各種糖果、餅乾、蛋糕、麵包等甜食，或煙燻滷味、醬燒等鹹食，裡面都加了不少糖。炎炎夏日大家更是人手一杯大杯糖飲。然而，愈來愈多的動物實驗和人體實驗結果，讓我們看到糖帶來的壞處。攝取很多果糖時，無論是人抑或老鼠的實驗裡，都會看到發炎指數升高。當餵食小白鼠高糖飲食，小白鼠會產生嚴重的發炎反應，長出乳癌並轉移到肺部。目前認為，吃一堆糖與肥胖、胰島素阻抗性、糖尿病、脂肪肝、癌症、慢性腎病變都有關係，可說是健康大敵，而且這種壞處並無法用「另一種好東西」來反制。什麼意思呢？有個實驗讓小白鼠吃「容易發炎」的高糖飲食和「抗發炎」的魚油，而結果發現，這時魚油無法帶來抗發炎的好處！所以別想說平常愛吃甜食沒關係，後來多吃個保健食品就好了…其實，平時吃高糖飲食，還會讓抗發炎的魚油失效呢。所以請記得，平時如果你想吃個甜甜，吃水果就好了，只要不沾糖粉，僅攝取真實存在水果裡的天然糖份並不會大幅影響健康，但盡量請不要吃額外添加的糖，否則這一點都不療癒，反而很傷身。

(438 words)

English Text for Reading ahead and aloud

(from https://theconversation.com/depression-men-far-more-at-risk-than-women-in-deprived-areas-127264)

Depression is a major cause of disability around the world, and if left untreated, can lead to substance abuse, anxiety and suicide. Major depressive disorder is a particular form of the condition which affects many people, potentially causing loss of pleasure in activities that once used to bring joy. It can also lead to feelings of worthlessness, imbalances such as oversleeping or insomnia, and trigger thoughts of suicide. This is the condition we examined during our new study, which showed that living in a deprived area can lead to major depressive disorder in men, but not in women.

Before explaining these findings, it is important to provide some further background on this condition. There are certain factors which can place you at increased risk for major depression. Being diagnosed with a serious chronic ailment, such as diabetes or cancer, now or in the past, can increase your risk for it. As can experiences of trauma, such as physical or sexual abuse, or being raised in a dysfunctional family in which there was a high degree of marital discord.

These, however, are all individual factors – or personal circumstances – which can negatively affect your mental health. And most of the research on depression has indeed focused on such personal factors. But there are characteristics beyond the level of the individual – such as attributes of the communities in which we live – that can also have a profound effect on our mental well-being. (239 words)

Appendix B

Text for Basic ST Exercises

Voice Technology for Sight Translation (2)

Dear all,
Step 1. Please use the Text 1 as the warm-up to practice Reading A/A with your ASR APP (e.g. iPhone and Google Voice). While you read A/A, please chunk the text into meaning units.
Step 2. When you're ready for the ST, please incorporate your voice technology to the process and also record your audio performances at the same time.
Step 3. Repeat Step 2 with Text 2 (Earning power)
Step 4. Reflections – What are the advantages/disadvantages of using ASR for these exercises? Any problems? Your possible solutions?

Text 1 Does it really matter which university you study at?
(Read A&A → Segment → ST)
This is the question that's been chosen by the BBC News audience - and it is a very immediate concern for hundreds of thousands of families wrestling with university application forms. Of course, on the idealistic side of things, what really matters is that someone is following a course that they really like and in a place
that suits their needs. But there are thornier worries about the cost of university and how much degrees are worth after graduation. The evidence suggests that going to university remains a good investment.
Organisations such as the Organisation for Economic Co-operation and Development (OECD) have tracked whether the rising number of students will erode the benefits in the jobs market. So far it seems that the graduates have kept their advantage. A changing jobs market has generated more opportunities for graduates and people who went to university are likely to be earning more and are less likely to be
unemployed. (166 words)

Text 2 Earning power (Segment → ST)
The Institute for Fiscal Studies, Cambridge University, Harvard University and the Institute of Education, UCL published research on graduate earnings in England earlier this year. It analysed the incomes of 260,000 graduates and showed a very wide spectrum of likely earnings. At the top was a cluster of universities,
headed by the London School of Economics, Oxford and Cambridge.
In these three institutions, 10% of their male graduates had earnings above £100,000 a decade after leaving university. The LSE was the only place where 10% of female graduates were also in this top earning bracket. There is an earnings pecking order - with about another 30 or so universities, not identified by name, where 10% of graduates are earning above £60,000. And at the bottom, there are some more awkward figures.
There are 23 universities where male graduates are likely to end up earning less than non-graduates - and there are nine universities where that is also the case for women. But there is another important factor
cutting across this - the differences between subjects. (175 words)
(from https://www.bbc.com/news/education-38015829)

Appendix C

Questionnaire

Turn Your Voice Into Words
Voice Technology for Sight Translation

Dear all,
Having worked with your chosen ASR tools and platforms over the past few weeks, it's time to reflect and record our observations and thoughts. Please tick the boxes which apply to your situation and use the following questions to record your experiences.

Table C

APPs I used	☐ iPhone dictation ☐ Others (please specify)	☐ Google voice type
Platforms I used		Activities
☐ Mobile phone	☐ iPhone ☐ Android	☐ Reading ahead/aloud ☐ Sight Translation ☐ Chunking ☐ Others (explain)
☐ Laptop		
☐ Computer		
☐ Others (please specify)		

1. What are the most noticeable advantages of using ASR over the past few weeks in your ST exercises? (name 3)

2. What are the disadvantages or noticeable problems with the tools you used? (name 3)

3. How do you compare the two systems/platforms that you used?

4. What is your preferred option and why?