ARTIFICIAL INTELLIGENCE IN SIMULTANEOUS INTERPRETING TRAINING: AN EXPERIMENTAL STUDY OF SPEECH-TO-TEXT TECHNOLOGY

Dr. Safa'a Ahmed
Associate Professor of Interpreting and Translation
Faculty of Languages, MSA University, Egypt

ABSTRACT

As the world is rushing relentlessly to incorporate technology and artificial intelligence into various walks of life under allegations like 'development', researchers should investigate the potential impacts of such a movement from different perspectives. Simultaneous interpreting (SI) is no exception per se. The present study aimed to explore how useful the technological advances achieved in artificial intelligence can be in SI training (process and performance) through an experimental study of a speech-to-text technology. It adopted both qualitative and quantitative methodological approaches using analysis, comparison, assessment, questionnaire and experiment as research tools. In this human-machine interaction, sample original English speeches (in Language B) were interpreted simultaneously into Arabic (Language A) by participants/trainees representing fourth year university students, with the help of a speech-to-text model. The significance of the study lies mainly in its potential implications for the industry, training and education, and research. It found out that STT in its current form is a failure and that the suggested model proved some success although the results were quite modest.

Keywords: Simultaneous Interpreting Training; Artificial Intelligence; Speech-To-Text Technology; Speech Recognition; Performance Assessment

الذكاء الاصطناعي في التدريب على الترجمة الشفهية: دراسة تجريبية على تقنية تحويل الكلام إلى نص مكتوب

أ.م.د. صفاء أحمد

أستاذ مشارك الترجمة الشفهية والتحريرية

كلية اللغات ، جامعة MSA ، مصر

ملخص

نظرًا لأن العالم يندفع بلا هوادة لدمج التكنولوجيا والذكاء الاصطناعي في مختلف مناحي الحياة تحت ظل ادعاءات وشعارات مثل "التطور" وغيرها ، يجب على الباحثين التحقيق في الآثار المحتملة لمثل هذا الدمج من وجهات نظر وعلوم مختلفة. والترجمة الفورية ليست مستثناه من هكذا أمور. تهدف الدراسة الحالية إلى استكشاف مدى فائدة استخدام التقدم التكنولوجي الذي يحدث في الذكاء الاصطناعي في التدريب على الترجمة الفورية (عمليةً وأداءً) أو عدم فائدتها، وذلك من خلال دراسة تجريبية على تقنية تحويل الكلام إلى نص مكتوب (Speech-to-Text) STT. واعتمدت على المنهجين النوعي والكمي من أجل ذلك، واستخدمت التحليل والمقارنة والتقييم والاستبيان والتجربة كأدوات بحث. في هذا التفاعل بين الإنسان والآلة ، تمت ترجمة عينة من خطابات باللغة الإنجليزية (من اللغة ب) ترجمة فورية إلى اللغة العربية (اللغة أ) من قبل مشاركين/ متدربين يمثلون طلاب السنة الرابعة بالجامعة، بمساعدة إحدى تطبيقات تحويل الكلام إلى نص مكتوب. تكمن أهمية الدراسة بشكل أساسي في آثارها المحتملة على هذه الصناعة والتدريب عليها وتعلها وإجراء الأبحاث عليها. فتوصلت إلى أن تقنية STT في شكلها الحالي غير مفيدة وأن النموذج المقترح أثبت بعض النجاح على الرغم من أن النتائج جاءت متواضعة للغاية.

الكلمات المفتاحية: التدريب على الترجمة الفورية؛ الذكاء الاصطناعي؛ تقنية تحويل الكلام إلى نص مكتوب؛ التعرف على الكلام؛ تقييم الاداء

ARTIFICIAL INTELLIGENCE IN SIMULTANEOUS INTERPRETING TRAINING: AN EXPERIMENTAL STUDY OF SPEECH-TO-TEXT TECHNOLOGY

Dr. Safa'a Ahmed
Associate Professor of Interpreting and Translation
Faculty of Languages, MSA University, Egypt

INTRODUCTION

> Robust research methods are required to cover the potential impact of emerging technologies on interpreting.
>
> -Amato, Spinolo and Gonzalez Rodriguez (2018:43)

Simultaneous interpreting is such a complex, heavily-loaded cognitive activity for humans that research and practice have been trying to develop ideas and tools to help interpreters or facilitate their work along the history of the profession. Technological developments, particularly the progress achieved in artificial intelligence (AI) and speech-to-text (STT) or speech recognition (SR) systems/ applications/ or tools, are precipitous and daunting. SR is "the ability of machine/program to identify words and phrases in spoken language and convert them into machine-readable format", while STT refers to "the process of converting spoken words into written texts" (Trivedi et.al.2018:37). SR task is "to convert speech into a sequence of words by a computer program" (Huang and Ding 2010:339). The two terms, SR and STT, are used interchangeably in this study, though the latter often refers to a wider process of speech understanding.

Developments have become visible in many applications in various fields, like health care, car systems, military aircrafts, people with disabilities, education, to name but a few. The profession of simultaneous interpreting (SI) is no exception. These technological advances may range from rooms fully equipped with high quality devices (e.g. audio and video equipment, booth consoles, laptops, etc.), to Computer Assisted Translation (CAT) and Computer Assisted Interpreting (CAI) tools (e.g. software memories, dictionaries and programmes as seen in remote interpreting (RI) (as opposed to in situ- or face-to-face interpreting) settings (e.g. video-conferencing and teleconferencing). Remote interpreting "is a broad concept which is commonly used to refer to forms of interpreter-mediated communication delivered by means of information and communication technology" where not all the event participants are present at one place (Fantinuoli 2019:8). Therefore, the use of machine-assisting tools in the field signifies a fertile area for both research and the industry.

The present study aims to explore how useful the technological advances achieved in artificial intelligence can be in SI training (process and performance) through an experimental study of a speech-to-text technology (STT), using the Otter.ai model. It adopts both qualitative and quantitative methodological approaches using analysis, comparison, assessment, questionnaire and experiment as research tools. In this human-machine interaction, sample original English speeches (in Language B) were interpreted simultaneously into Arabic (Language A) by participants representing fourth year university students (called trainees here), with the help of the speech-to-text model. The contribution of the study lies mainly in the attempt to explore the functionality or otherwise the disfunctionality of using an AI tool, like STT, in human SI training; hence come its potential implications for the industry, training and education, and research. Also on the methodological level, it combines between both qualitative and quantitative methods to give results more validity. Moreover, its quantitative approach depended not only on two questionnaires to measure the trainees' opinions, but also on two experiments to measure and compare their performances.

In addition to this introduction, which sets the scene with some preliminary concepts, and the conclusion, the paper is divided into three sections: a literature review, theoretical and methodological frameworks and a results and discussion section.

1. LITERATURE REVIEW

1.1 Speech-To-Text Developments

The very early attempts at speech recognition can be said to start in the 1950s and 60s and they reached a machine performance that could understand 16 English words and distinguish 9 consonants and 4 vowels. At the beginning of the 1970s, the U.S. Department of Defense funded a Speech Understanding Research programme (SUR) creating a system called "Harpy" "which was shown to be able to recognize speech using a vocabulary of 1,011 words, and with reasonable accuracy" (Juang and Rabiner 2004:9). Other systems like Hearsay II and Hear What I Mean (HWIM) met SUR's performance goals. Also Bell Laboratories developed a device that managed to understand more than one person's voice. During that period, "the computational power available was only adequate to perform speech recognition on highly constrained tasks with low branching factors (perplexity)" (Huang, Baker and Reddy 2014:102).

An important development occurred in speech recognition systems in the 1980s. The inclusion of the Hidden Markov Model helped use statistics, rather than the old straight forward, fixed models, to decide the word probability of an unknown sound in an unlimited number of words (Juang and Rabiner 2004:12). It did not rely on speech patterns or fixed templates. Industries and businesses benefited from the new system applications as in the Julie doll which responded to children's spoken words; the only problem was that the speaker had to pause after each word. The decade ended in using neural networks in speech recognition models where they managed to recognize a few phonemes and words.

The 1990s witnessed the release of the first speech recognition system, "Dragon Dictate" for consumers. Microsoft created applications for speech recognition on Windows 95. Huang, Baker and Reddy summarizes the achievements of Hearsay-I as "one of the first systems capable of continuous speech recognition", the Dragon as "one of the first systems to model speech as a hidden stochastic process" and Harpy as the first to introduce "Beam Search, which for decades has been the most widely used technique for efficient searching and matching" (2014:96). Then in 1987, Sphinx-I managed to recognize speaker-independent speech and in 1992 Sphinx-II made use of some tied parameters to balance between trainability and efficiency, a matter which achieved the best recognition accuracy in DARPA evaluation in 1992 (ibid.).

The advances taking place at the level of speech data and computing power also affected significantly STT models and enabled them to deal with larger tasks. Juang and Rabiner indicate that "While we are still far from having a machine that converses with humans on any topic like another human, many important scientific and technological advances have taken place, bringing us closer to the “Holy Grail” of machines that recognize and understand fluently spoken speech" (2004:3). As Apple, Microsoft and Google started the adoption of SR applications in their products, devices managed to deal with almost unrestricted multimodal dialogues despite some "remaining challenges", Huang, Baker and Reddy argue, adding that "the speech community is en route to pass the Turing Test in the next 40 years with the ultimate goal to match and exceed a humans speech recognition capability for everyday scenarios" (2014:95).

1.2 Technology in SI

The world's feverish propaganda for the inevitability of technological change together with the epidemic panic from machine IQ to reach 10,000 (while Einstein's was 150) have given impetus for this research

> to look into SI as integrated into the emerging technology. (Ahmed 2022:4)

Generally speaking, technology has been used more intensively in Machine Translation (MT) and MI than in human simultaneous interpreting. In the latter, "technological support is scarce, except for electronic devices used for terminology support in the booth (Desmet et.al.2018:13-14).

Desmet et.al. made an experimental pilot study to investigate the potential impact of using technology which helps in the interpretation of numbers in the booth. They suggest that visual assistance of the speaker's text is "likely to boost performance" of interpreting numbers and names if displayed on a screen while they are pronounced (2018:16). They found out that using automated speech recognition (ASR) system improved the overall accuracy of interpreting numbers from 56.5% to 86.5 % and reduced errors two thirds. Without ASR systems, participants omitted or approximated numbers in many cases; with ASR, omissions dropped by nearly 90%. They add that now "limited applications exist in conference rooms with voting systems, where the results of votes are displayed on a screen in the booth, but the targeted use of natural language processing applications could make it possible in the near future to extract numerical information from online speeches" (p.17). Support in this regard comes mostly through providing terminological memories during SI process.

In an important experiment carried out by Lamberger-Felber (2001; cited in Desmet et.al.2018), 53% to 68% fewer errors of interpreting names and numbers were reported after interpreters had been provided with the source text in the booth, compared to their performance without the text. In line with this idea, Mead investigated how beneficial writing down numbers by a booth collaborator, providing a visual text by the speakers before the event, or projecting presentation slides can be (2015; as cited in Desmet et.al 2018).

Fantinuoli calls the current state of using technology in interpreting "the technological turn" (2018b:2). He refers to three areas in this respect: CAI, RI and MI. He says that technology can:

> improve the interpreters' work experience, by relieving them of the burden of some of the most time-consuming tasks (such as the creation and organization of terminology) and by supporting them in carrying out numerous activities, from the retrieval of preparatory

documents to their analysis in a way appropriate to their profession. (p.4)

This can happen before and during SI. He distinguishes two categories of the technologies that can be used in SI: *setting-oriented*, used to improve the setting, and *process-oriented*, to facilitate the SI process itself. Today, the CAI tools used mostly in interpreting focus on terminological systems (Fantinuoli 2018a).

The crucial point in using any CAI tool during interpreting is the simultaneity of information availability with the speaker, i.e. the interpreter gets support within the ear-voice span. Desmet et.al. (2018:17) explain that ASR "has the potential of speeding up the look-up process and solving the cognitive effort and latency of manual querying". However, they assert that "Given the current state of the art in asr and its foreseeable progress, it seems to be a matter of time before this technology is used in cai tools to support interpreters with terminology look-up, and/or with information-dense content" (ibid.).

STT models are used as a part of CAT to transfer a speech into a text readable for the machine which in turn translates the text and produces the translation in a written form. The same applies to CAI but the product comes in an oral form "in which a human interpreter makes use of computer tools designed to support and facilitate some aspects of the interpreting task – mainly subject preparation and information access – with the goal to increase quality and – to a minor extent – productivity" (Fantinuoli 2019:7). In other words, CAI can help the interpreter in the preparation phase and rarely during the event as a glossary to look up terms. In fact, MI is scarcely used in limited, more often informal settings, like travelling, restaurants and some simple services on a small scale. Because interpreting is a complex human activity, MI is still facing many challenges. This is why "there is reason to believe that the development of an MI system that could systematically compete with human interpreters will require a lot of effort, and probably a lot of time" (ibid.8). Therefore, complete MI seems to be totally an invalid idea at least for the present time and we should look for tools to help human interpreters through human-machine interaction rather than replacing them.

1.3 WER and Latency Assessment of STT

Drastic attempts have been undergoing with a view to mimic human performance '*accuracy*' in transforming a speech to text by machines and to get this with low '*latency*' (time gap between the speech and its computational text) in order to reach a synchronous speed of the delivered speech. Juang and Rabiner argue that "the challenge of designing a machine that truly functions like an intelligent human is still a major one going forward.. it will take many years before a machine can pass the Turing test, namely achieving performance that rivals that of a human" (2004:21).

One of the most common ways to assess the accuracy of STT systems is word error rate (WER), referring to the total number of incorrectly transcribed words divide the total word count of the original speech then multiply by 100. Errattahi et.al. (2018; as cited in Filippidou and Moussiades 2020:77) mentions three errors in STT systems: *substitution* when a word is replaced by another, *deletion* when it is omitted and *insertion* when a new word is added. WER has advantages like its simplicity and easy use and disadvantages as it does not evaluate how good a system is in its own, but it can compare systems; it also gives more weight to insertions than deletions (ibid.). Most assessments of WER systems talk of an average error 5%. For instance "Microsoft claims to have a word error rate of 5.1%. Google boasts a WER of 4.9%.. For comparison, human transcriptionists average a word error rate of 4%" (Does WER Matter 2021). According to Tech Radar, Dragon's offerings rank at the top of SR models for 2021. Nuance says the software dictates at speed of 160 words per minute with an accuracy 99% (Speech Recognition Software 2021). Additionally according to Sonix, while IBM claimed 5.5% WER, Google 4.9% and Microsoft 5.1% in 2017 (What is Word Error Rate 2022). This means that "the accuracy of speech recognition systems remains challenging task in the research field" (Dhanraj 2020:521).

Latency has been improved a lot but still constitutes a problem for STT models particularly if we talk about a tool to help in SI. Indeed, neither complete accuracy nor zero latency has been attained yet, despite the progress achieved in deep neural networks a decade ago. The future of ASR may lie in, as Li argues, "a significant trend of moving from deep neural network based hybrid modeling to end-to-end (E2E) modeling.. While E2E models achieve the state-of-the-art results in most benchmarks in terms of ASR accuracy, hybrid models are still used in a large proportion of commercial ASR systems at the current time" (2020:1).

However, if we keep in mind that humans can speak 150 words/minute and can only type 40 words/minute on average, the job of the machine cannot be undermined or underestimated, hence comes the purpose of the present paper.

From this review, a noticeable gap exists in our understanding of and research on how the advances achieved in AI can influence the SI field. As Desmet et.al. argue, "While the development and adoption of cai (computer-assisted interpreting) tools has been limited, scientific research on the impact of their use has been even scarcer" (2018:17). Fantinuoli also stresses that "more empirical studies are still needed to understand if CAI tools will be able to meet interpreters' real-life requirements" (2019:10).

2. THEORETICAL AND METHODOLOGICAL FRAMEWORKS

2.1 Theoretical Framework

For three decades (1970s-1990s), Seleskovitch was developing the Interpretive Theory of Translation which has been used as a foundation for understanding how interpretation occurs. The EU asked her and Lederer to write a book on teaching interpretation 'Pédagogie raisonnée de l'interprétation' (1989), translated into English in 1995. Her Theory is "a coherent construct with high explanatory power, based on practical experience of both interpreting and translation" (Lederer 2010:173). The main tenet of the theory is that interpreters translate sense rather than words (Seleskovitch 1999); an idea criticized by Saussurean and literary translation scholars.

Seleskovitch divides the interpreting process into three phases. First, the interpreter *comprehends* the sense of the original message keeping in mind the extra-linguistic features necessary to understand it. Second, words disappear in a *deverbalistion* phase (mental representation of intended meaning; Chesterman and Wagner simply state that deverbalisation is used "to get away from the surface structure of the source text, to arrive at the intended meaning" (2002:9-10; Lederer 2010:176). Third, he *reformulates* or expresses what he comprehended in a way that does not seem strange to the target hearer. Each of these phases needs attention or 'efforst' as Gile (1995) call it. The interpreter's cognitive load may be affected when, for instance, reading the original text, reading it with some sort of latency between the speaker and the transcription, or reading with about 5% WER in the transcription. Here lies the core of the present study.

2.2 Methodological Framework

There is an unprecedented feverish rush for integrating AI into various disciplines and professions, including SI. From this problem statement, the researcher formulated the study aim and research questions. It aims to explore how functional or dysfunctional a speech-to-text system (STT), called Otter.ai, can be in SI training for university undergraduate students/trainees. It raises the following research questions:

RQ1-What is the opinion of SI participants about the use of STT technology in their training before the experiments?

RQ2-How functional or dysfunctional is using STT technology in STT training after the Otter.ai experiment?

RQ3-If improvements are suggested to SST model, how functional or dysfunctional can they be in SI training?

RQ4-What is the opinion of participants about the use of STT technology in their training after the experiments?

RQ5- In the light of all the results, how useful or otherwise useless can STT technology be as a potential tool to help interpreters?

To answer these questions, the objectives are to:

1-Prepare the source speeches for the experiments;

2-Make a questionnaire before the experiments to measure the trainees' opinions;

3-Measure the trainees' SI performances before the experiments;

4-Make an experiment (1) to assess the difference between the trainees' performance before using Otter.ai and after using it;

5-Make another experiment (2) to assess the difference between the trainees' performance before using the suggested improvements to the STT model and after using it;

6-Make a questionnaire after the experiments to measure the trainees' opinions;

7-Analyse and discuss results to explore the usefulness vs. uselessness of STT technology in both its present and suggested improved form.

To this end, the researcher has adopted both a qualitative and quantitative method of research to give results more validity and reliability particularly as they are taken from various perspectives, from both the trainer's and the trainees'. In addition to questionnaire and experiment, research tools include analysis, assessment and comparison.

Participants. Participants were selected randomly from a total number of 140 fourth year undergraduate students (called here trainees) of the Faculty of Languages, MSA University, Fall 2021-2022 semester. They registered in this simultaneous interpreting course after they passed some written translation units and a consecutive interpreting one. The course is a core unit for all the Faculty students, i.e. trainees come from four different minors (where students choose 4 units in either translation, TEFL, comparative literature or executive skills minor). The 140 trainees were divided into 8 groups to be manageable. Test subjects are 60 random trainees who were willing and able to complete the questionnaires before and after the experiments, out of them only 30 trainees managed to make a default performance then take the two experiments. The test subjects belong to 3 groups only (who completed the whole task).

Questionnaires. Two questionnaire were designed to be addressed to the 60 participants before and after the experiments. Instructions were explained and some information about the topic was delivered before filling in the questionnaires. In the first one, participant were asked about their knowledge and opinions about using an STT model to help them during the interpreting process. In the second, they were asked about their opinions having gone through the actual experiments and what opportunities or challenges could be raised in this regard.

Experiments. As explained above, 30 participants/trainees managed to complete the experiments. They were provided with instructions and some background about AI and STT technology before the questionnaires and the experiments. Then each was given a source speech to prepare and interpret; this performance will be a '*default*' rendition against which the performances in the two experiments are compared. The same speech was given in an '*otter.ai*' experiment in which he had an access to the speech transcribed live by the speech-to-text model, Otter.ai. Otter.ai is a technology company which employs AI and machine learning to develop STT transcription and translation applications. Its free availability, simplicity and easy application encouraged me to use in the experiment.

In the second experiment, some improvements were introduced to the model: almost no errors and no latency. To the best knowledge of the researcher, no such model is currently available in the market. That is why the improved model was presented through subtitling hoping that a future model can be as acceptable, to a great extent, as subtitling. The trainee performed the '*suggested*' experiment immediately after experiment 1. The reason why there is no time gap between performances (1 and 2) is to avoid any probability of improved performances that may result from repetition rather than exposure to the tested tools, AI/otter.ai and suggested improvements. In other words, trainees were not allowed to have time to check what they have done in the previous performance(s). That was done to avoid assigning another speech to the same trainee that could include variables which may cause the next performance to improve or deteriorate due to those variables rather than the tested tools; the researcher thus held source speech difficulty variables constant through exposing each trainee to the same one text.

The experiments aim to explore the differences in SI trainees' performances when the interpreter is helped by an STT model and a suggested improved one. The independent variables are the Otter.ai and the subtitling tools and the performances before and after the experiment represent the dependent variables. Each trainee has three performances recorded, analysed, assessed and compared.

Tests and Rubric. In this human-machine interaction, source English speeches (in Language B) were carefully selected on the basis of similar speech difficulty, speed rate, accent, etc. to be interpreted simultaneously into Arabic (Language A). All the speeches represent a similar topic, extracted from the UN Climate Change Conference COP26 held at Glasgow, Scotland, 2021 and G20 Summit 'The Future of Humanity is at Stake', 2021. Speakers are Prince Charles, Belgian Premier and Barbados Premier and each speech lasts about 7 minutes. The possibility of a difficulty bias in the speeches is erased since each student is exposed to one source speech, rather than two or three speeches. The tests/speeches are also exposed to a panel of jury to assess their validity, reliability and consistency.

The researcher used a rubric she developed based on previous studies on quality assessment of simultaneous interpreting (cf. Wu 2010; Ahmed 2020, 2018, 2016 and 2015). It consists of three criteria: Fidelity and Completeness, Presentation and Delivery, and Audience Point of View. By Fidelity and Completeness we mean content completeness and accuracy, faithfulness to the speaker's message, and contextual consistency; it weighs 50% of the total assessment or mark given to the trainees. Presentation and Delivery criterion includes SI skills and strategies, the interpreter's language abilities and knowledge; it weighs 30%. The third criterion, Audience Point of View, refers to the audience trust in the interpreter's message, personality and aptitude; it weighs 20%.

The researcher is aware of the study limitations, such as time and space limitations. Also it could give more reliable results with a larger size of participants/trainees. Furthermore, it is limited to Otter.ai as an STT tool; more experiments can be applied to other STT applications.

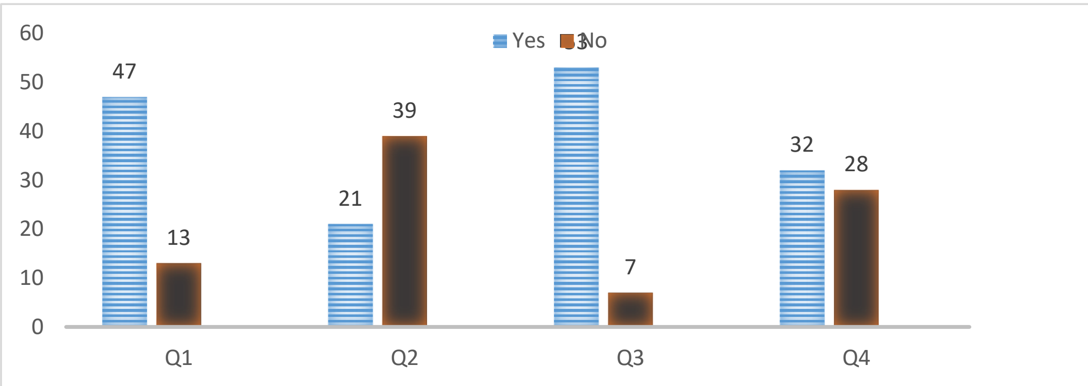
Thus based on the aim, objectives and research questions (RQs), the results and discussion section is divided into three parts: Opinions before the Experiments; Opinions after the Experiments and Performance Comparisons reflecting results from both experiments.

3- RESULTS AND DISCUSSION

3.1 Opinions before the Experiments

In the first questionnaire, the 60 participants are asked 10 questions (Qs). In Table 1, four questions are raised. In Q1, "*Have you ever heard about AI?*", their basic knowledge about AI was measured: 47 responded Yes and 13 No, which means that the majority 78.3% know about AI, while 21.7% do not. In Q2, "*Have you ever heard about using AI technology in SIM?*", 21 said Yes and 39 No, i.e. 35% only know about this while 65% do not; this may be attributed to the fact that the profession has been known in its classical form for a long time and it may take some time to incorporate technology and more specifically AI in its processes. What is interesting is their answers to Q3, "*Have you ever heard about speech-to-text (STT) technology*?", where 53 (88.3%) responded with Yes and 7 (11.7%) with No. This implies that they are not able to make connections between SIM and recent technological trends. However, 32 (53.3%) have used STT and 28 (46.7%) have not according to Q4 "*Have you ever used it?*", close percentages which indicates that some of the majority, who responded to Q3 with Yes, have superficial knowledge in this regard.

Table 1: Basic Knowledge



Then, participants who have used STT are asked in Q5 "*If yes, please specify which software*!", see Figure 1.

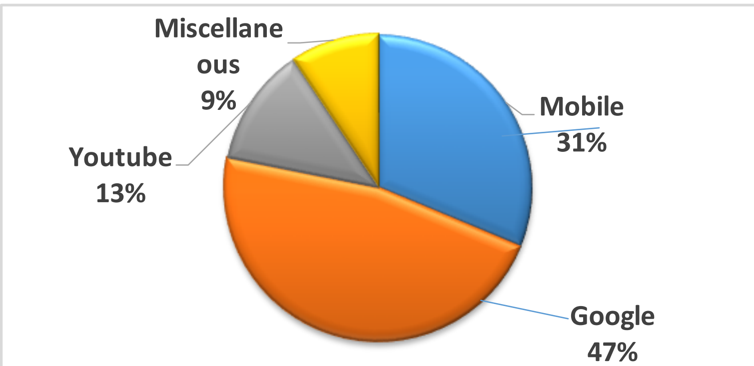


Figure 1: Which Software Participants Have Used

15 (47%) participants use Google, 10 (31%) participants use the application on mobile, 4 (13%) use YouTube and 3 (9%) use miscellaneous apps. This indicates the popularity of Google in this connection. The context of such usage is addressed in Q6, "*If yes, please write down in what context!*" where answers varied from educational context (21 participants i.e. 65.5% use it in education) and social context (11 participants i.e.34.5% use it for social purposes), see Figure 2.

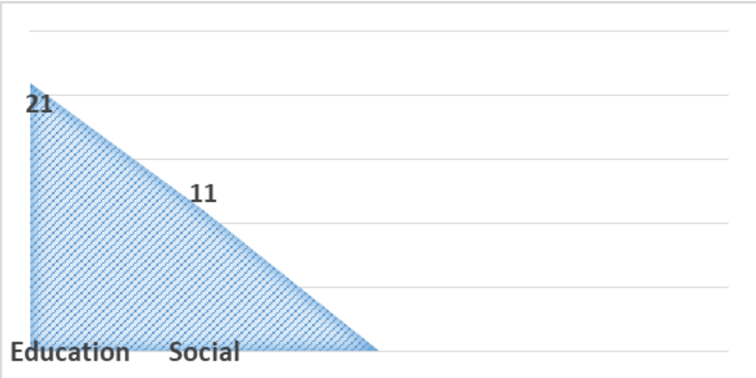


Figure 2: STT Context

Then Q7 measures their opinions before the experiments regarding "*How useful can the use of STT technology be and help improve performance?*", see Figure 3. Indeed the majority expected it to be useful and the minority was pessimistic, while some participants remained neutral. 13 (21.6%) said it can be very useful, 31(51.6%) useful, 10 (16.6%) neutral, 6 (10%) useless; none chose very useless. These numbers will be compared to the numbers after the experiment.

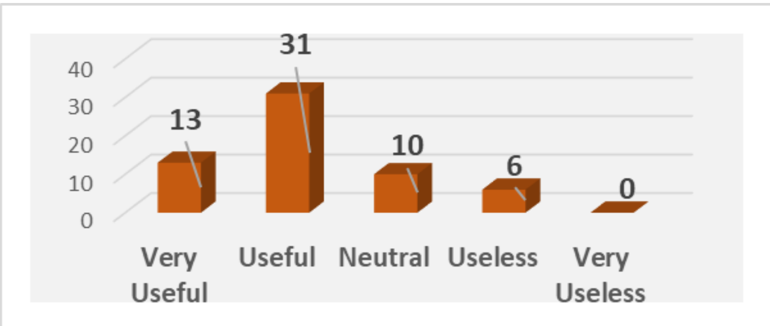


Figure 3: How Useful can STT be?

A technical SI question is raised to see their expectations in Q8, "*Do you think that the use of speech-to text technology can distract the simultaneous interpreter during the performance?*". They were asked to rate from 5-1 where: 5= Yes surely, 4=Yes to a great extent, 3= Yes to some extent, 2= To a little extent, and 1= Not at all; see Figure 4.

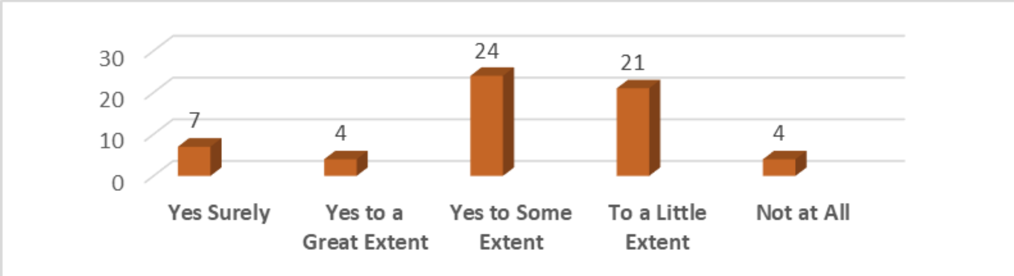


Figure 4: Rate STT Possible Distraction Level

7 (11.6%) participants expected that the use of an STT tool during SI can surely distract the interpreter, 4 (6.6%) said it can to a great extent, 24 (40%) can to some extent, 21(35%) can to a little extent and 4 (6.6%) cannot distract him at all. That's to say the majority were not confident whether STT will help or rather hinder.

However, the majority of participants seemed to be enthusiastic about using the tool in their trainings and exams with a general impression that technology will always help. Q9 asked them to rate their answers to the following question from Totally Agree to Totally Disagree: "*Do you recommend the use of STT technology in SI training?*" as in Figure 5. 16 (26.5%) participants Totally Agreed to recommend it, 27 (45%) Agreed, 10 (16.6%) Did Not Know, 6 (10%) Disagreed and 1 (1.6%) Totally Disagreed.

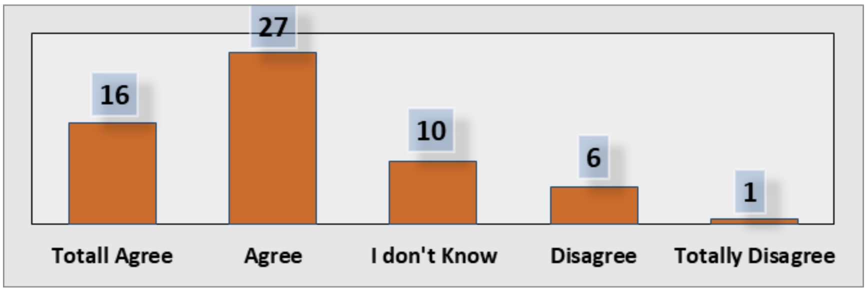


Figure 5: Recommending STT in SI Training

Q10 is an open-ended question asking "*Why?*" did they choose their answers in Q9. That's to say they list the opportunities and/or challenges they expected and this is why the total number of answers exceed 60, see Figure 6.

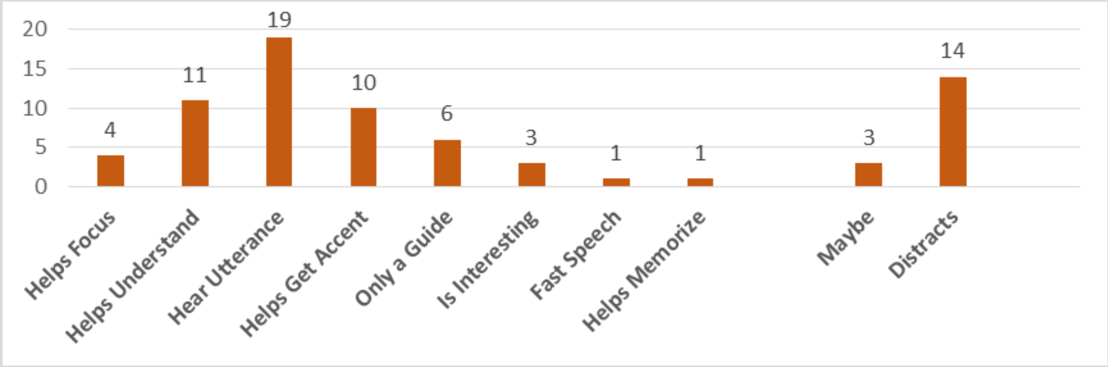


Figure 6: Opportunities and Challenges

On the one hand, opportunities include: 4 opinions think that STT tool can help the interpreter focus; 11 think it helps understand the message; 19 think it helps them hear the utterance and unclear or missed words; 10 see it helps understand the speaker's accent; 6 view it as a possible guide; 3 consider it an interesting tool to use; 1 opinion regards it as an asset to help in fast source speeches; and 1 believes it helps memorize the messages. On the other, challenges include: 3 opinions are not sure whether it is a helping tool or not; 14 indicate that it can distract the interpreter. The larger number of opinions seems to be worried that the interpreter may miss some original words during the complex SI process due to the multi-tasks, or as Seleskovitch calls it 'phases' of SI, required to be performed simultaneously by the interpreter.

3.2 Opinions after the Experiments

After the experiments, another questionnaire was designed to address 6 questions. Q1 asks: "*Do you think that the use of speech-to text technology helps the Simultaneous Interpreter improve his rendition during the performance?*". 43 (71.6%) responded with Yes, while 17 (28.4%) with No. This corresponds to Q7 in Figure 3, where 44 (73.3%) expected the tool to be useful, 6 (10%) responded negatively and 10 (16.6) preferred to be neutral. By comparison, the positive answers are more or less the same, but those who were neutral before the experiments have become confident after the experiment that it is not useful.

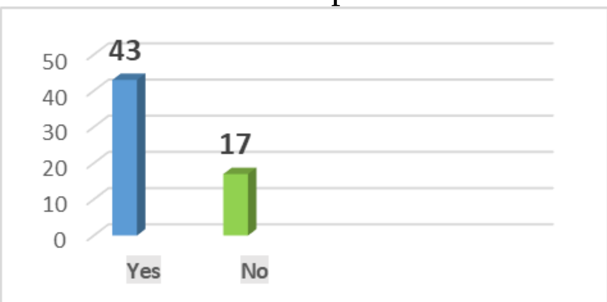


Figure 7: Useful vs. Useless Tool

Q2 asks participants "*How would you rate the use of speech-to text technology during your performance?*" on a scale from Very Easy to Very Difficult, see Figure 8.

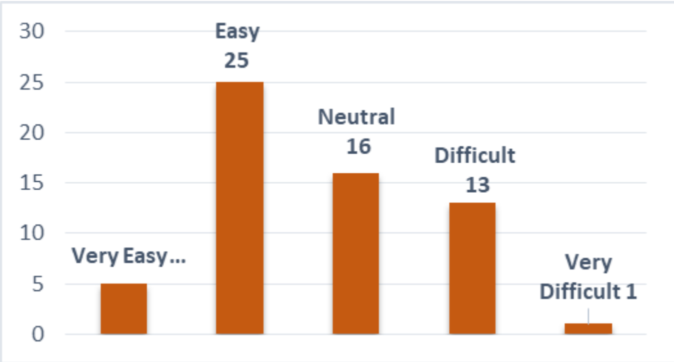


Figure 8: How Easy or Difficult is STT Tool

5 (8.3%) said it is Very Easy, 25 (46.6%) Easy, 16 (26.6%) Neural, 13 (21.6%) Difficult and 1 (1.6%) Very Difficult. Absolute majority found the model easy while about 23% found it difficult; the rest remained in between. Still closely related to the previous question, Q3 explores "*How did you find the use of speech-to text technology during performance?*" through a scale from Very Useful to Very Useless, see Figure 9.

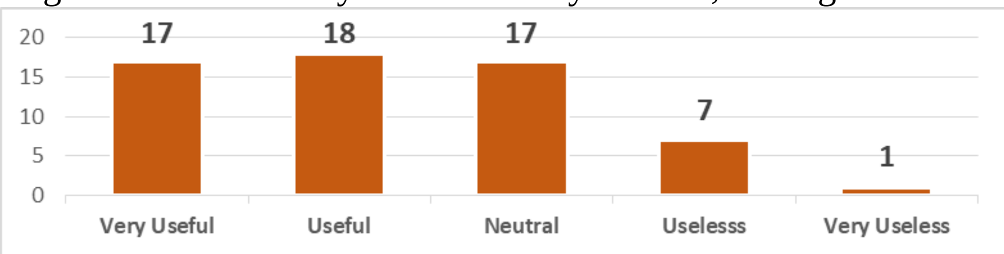


Figure 9: How Useful or Useless is ST Tool

17 (28.3%) said it is Very Useful, 18 (30%) Useful, 17 (28.3%) Neutral, 7 (11.6%) Useless and 1 (1.6%) Very Useless. Absolute majority had a positive feedback about the usefulness of the tool while minority felt negative, and 28% remained neutral. The difference between Q2 and Q3 is that of easiness vs. usefulness and vice versa, i.e. not every easy model is necessarily useful.

Now moving to the possibility of getting distracted by the tool, participants were asked to rate their answer to "*Do you think that the use of speech-to text technology distracts the Simultaneous Interpreter during the performance?*" in Q4 on a scale from 5-1, where 5= Yes surely, 4=Yes to a Great Extent, 3= Yes to Some Extent, 2= Yes to a Little Extent, and 1= Not at all. In Figure 10, 10 (16.6%) participants responded Yes Surely it distracts the interpreter, 11(18.3%) Yes to a Great Extent, 16 (26.6%) Yes to Some Extent, 15 (25%) Yes to a Little Extent and 8 (13.3%) Not at all. The majority found it distracting in a way or another during SI.

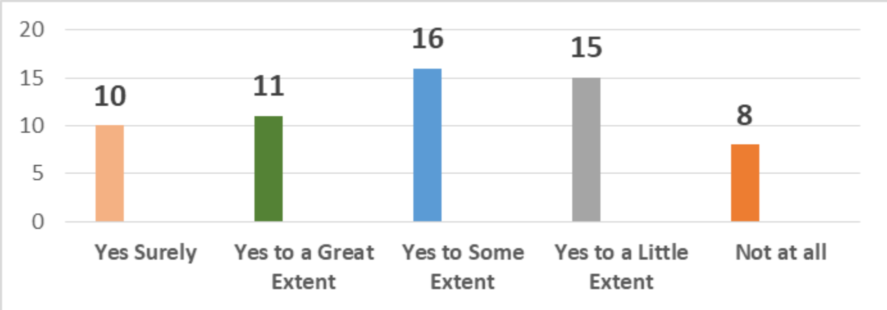


Figure 10: Can STT Distract the Interpreter

Q5 asks them to rate their answers to the question "Do you recommend the use of speech-to text technology in SIM training?" on a scale from Totally Agree to Totally Disagree, see Figure 11.

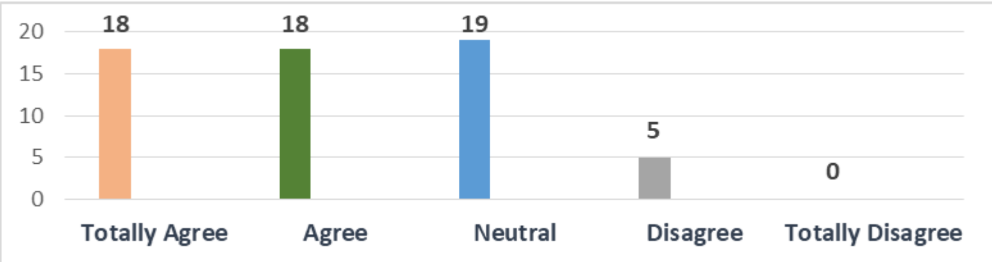


Figure 11: Recommending STT in SI Training

18 (30%) Totally Agreed to recommend using STT in SI training, 18 (30%) Agreed, 19 (31.6%) Neutral, 5 (8.3%) Disagreed and none (0%) Totally Disagreed. In comparison to the same question asked before the experiments (see Q9, Figure 5) the results correspond to 26.6%, 45%, 16.6%, 10% and 1.6% respectively. Accordingly, the majority still had a positive feedback though decreased from a total of 72% to 60%. The minority were still negative and feedback decreased from 16.6% to 11.6%. Meanwhile the Neutral or I Don't Know category increased from 16.6% to 31.6% but I think it may not be attributed to an increase in indecisiveness, rather than being convinced that STT tool is an opportunity and a challenge in the same time.

Finally Q6 investigated participants' opinions regarding "*During your performance, what areas did the speech-to text technology help you improve or vice versa, in your opinion?*". Strength points include the following: STT application helped 18 to interpret generally and interpret numbers and names especially, helped 3 to understand the original message, 7 to understand the speaker's accent, 17 to get unclear or missed source words, 4 to get well with fast speech, and 1 to memorize, see Figure 12. However, 5 others referred to the few mistakes made by the STT tool, 2 mentioned that using the tool needs practice and 1 said 'I Don't Know'.

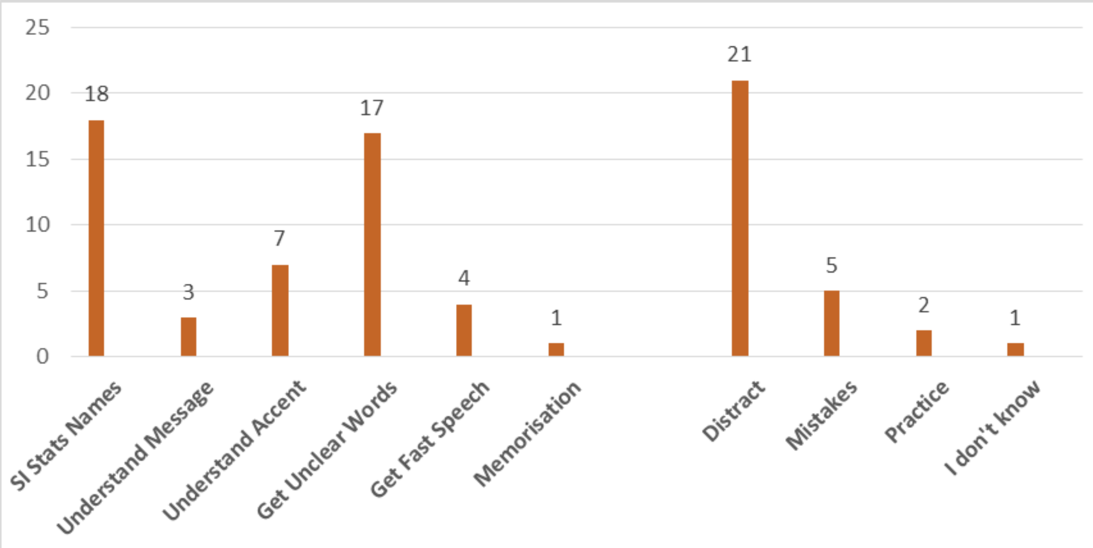


Figure 12: STT Strengths and Weaknesses

In comparison to the same question about participants' expectations before the experiments, 18 opinions noticed that STT helped them interpret numbers and names. If we add the first category to the second to refer to understanding the message, then opinions increased from 11 to 21. Understanding the accent increased from 7 to 9. 'Get Unclear or Missed Words' or 'Hear the Utterance' decreased from 19 to 17. 'Get Fast Speech' increased from 1 to 4 which indicates how useful it can be in case the speaker talks too fast that the trainee may need help through a written transcription of the source speech. 'Memorization' category has not changed, which is natural for three reasons: short memory only is applicable in SI, the interpreter does not really need a tool for memorization, and words vanish during the deverbalisation phase according to Seleskovitch. On the other hand, it is quite interesting to notice that their opinions that STT distracts the interpreter increased from 14 to 21. They also noticed the mistakes the tool can possibly make and the need for practice for it may affect performance. Moreover, the total number of opinions who regarded the tool as only a guide (6) and those who said Maybe (3) before the experiments decreased from 9 to 1 after;

this is understandable after going through the actual performance and being more decisive.

Until this stage, the positive results are mostly modest and challenges increased.

3.3 Performance Comparisons

In the next sub-section, the 30 trainees, who managed to complete the experiments, were

asked to make a default performance against which the results of the experiments are compared. The first experiment used Otter.ai (STT tool) to record the second performance, referred to as *Otter.ai*. The second experiment suggests an improvement in both word error (approximate percentage of WER 5%) and latency (voice span between the utterance and its transcription) through using Google subtitling transcription on YouTube; trainees recorded the third performance referred to as *Suggested* in the Table. The three performances, then, are assessed according to the rubric explained in the methodology. The three descriptors are: Message Fidelity and Accuracy (50% of the total mark) referred to as *Ms.*, SI Skills and Language (30%) referred to as *SI& Lang.*, and Audience Effect (20%) referred to as *Aud.* in Table 2 (cf. Ahmed 2020). The two experiments results are compared to the default performance and the difference is referred to as *Diff.* in the three descriptors plus the total in Table 2. Remarks are written to indicate the trainee's Level according to his default performance where Level A ≥ 80%, Level B = 60-79% and Level C < 60%. Furthermore, Remarks show whether the performance *improved* or *deteriorated* when the descriptors of the recorded Otter.ai and Suggested performances are compared to those of the default.

Table 2: Comparison of Performances

Code	Type of SI Performance	Dif. in Ms. 50%	Dif. in SI& Lang. 30%	Dif. in Aud. 20%	Dif. in Total 100%	Remarks
A1	Default					Level B
	Otter.ai	0	0	-2	-2 %	Performance deteriorated
	Suggested	-4	-1	-3	-8 %	Performance deteriorated
A2	Default					Level B
	Otter.ai/Test	+5	+4	+2	+11%	Performance Improved
	Suggested	+12	+5	+4	+10%	Performance Improved
A3	Default					Level C
	Otter.ai/Test	+2	+2	+2	+6	Performance Improved
	Suggested	+5	+5	+5	+15	Performance Improved
A4	Default					Level B
	Otter.ai/Test	-5	-2	-1	-8%	Performance deteriorated
	Suggested	+3	+2	+2	+7%	Performance Improved
A5	Default					Level A
	Otter.ai/Test	-2	-3	-2	-7%	Performance deteriorated
	Suggested	+2	+1	0	+3%	Performance Improved

Code	Type of SI Performance	Dif. in Ms. 50%	Dif. in SI& Lang. 30%	Dif. in Aud. 20%	Dif. in Total 100%	Remarks
A6	Default					Level B
	Otter.ai/Test	-4	-3	-4	-11%	Performance deteriorated
	Suggested	-3	-1	-1	-5%	Performance deteriorated
A7	Default					Level A
	Otter.ai/Test	+1	+1	0	+2%	Performance Improved
	Suggested	+2	+1	0	+3%	Performance Improved
A8	Default					Level B
	Otter.ai/Test	-3	-4	-4	-11%	Performance deteriorated
	Suggested	0	-1	-2	-3%	Performance deteriorated
A9	Default					Level A
	Otter.ai/Test	-3	-2	-1	-6%	Performance deteriorated
	Suggested	0	+1	+1	+2%	Performance Improved
A10	Default					Level A
	Otter.ai/Test	-2	-3	-4	-9%	Performance deteriorated
	Suggested	+1	+1	+1	+3%	Performance Improved
A11	Default					Level B
	Otter.ai/Test	-5	-3	-3	-11%	Performance deteriorated
	Suggested	-2	-2	-1	-5%	Performance deteriorated
A12	Default					Level B
	Otter.ai/Test	-2	-2	-1	-5%	Performance deteriorated
	Suggested	+2	+1	+1	+3%	Performance Improved
A13	Default					Level A
	Otter.ai/Test	-3	-3	-2	-8%	Performance deteriorated
	Suggested	-2	-3	-1	-6%	Performance deteriorated
A14	Default					Level B
	Otter.ai/Test	0	0	-1	-1%	Performance deteriorated
	Suggested	+2	0	+1	+3%	Performance Improved
A15	Default					Level B
	Otter.ai/Test	-4	-2	-3	-9%	Performance deteriorated
	Suggested	-2	0	-3	-5%	Performance deteriorated
A16	Default					Level C
	Otter.ai/Test	-5	-4	-3	-12%	Performance deteriorated
	Suggested	+5	+3	+4	+12%	Performance Improved
A17	Default					Level B
	Otter.ai/Test	-5	-3	-4	-12%	Performance deteriorated
	Suggested	+3	+1	+1	+5%	Performance Improved
A18	Default					Level A
	Otter.ai/Test	-2	-2	-1	-5%	Performance deteriorated
	Suggested	-1	0	-1	-2%	Performance deteriorated
A19	Default					Level A
	Otter.ai/Test	-3	-2	-2	-7%	Performance deteriorated
	Suggested	-2	-1	-1	-4%	Performance deteriorated
A20	Default					Level C
	Otter.ai/Test	-2	-2	-2	-6%	Performance deteriorated
	Suggested	+3	+3	+3	+9%	Performance Improved
A21	Default					Level C
	Otter.ai/Test	-2	0	-1	-3%	Performance deteriorated
	Suggested	+3	+4	+5	+12%	Performance Improved
A22	Default					Level C
	Otter.ai/Test	-6	-3	-3	-12%	Performance deteriorated
	Suggested	+4	+5	+3	+12%	Performance Improved
A23	Default					Level A
	Otter.ai/Test	-3	-3	-3	-9 %	Performance deteriorated
	Suggested	+1	+1	-1	+1 %	Performance Improved
A24	Default					Level C
	Otter.ai/Test	-1	-1	0	-2%	Performance deteriorated
	Suggested	+2	+3	+3	+8%	Performance Improved
A25	Default					Level C

Code	Type of SI Performance	Dif. in Ms. 50%	Dif. in SI& Lang. 30%	Dif. in Aud. 20%	Dif. in Total 100%	Remarks
	Otter.ai/Test	-2	-2	-1	-5%	Performance deteriorated
	Suggested	+8	+10	+4	+22%	Performance Improved
A26	Default					Level B
	Otter.ai/Test	+1	+2	-1	+2%	Performance Improved
	Suggested	+3	+3	+2	+8%	Performance Improved
A27	Default					Level B
	Otter.ai/Test	-3	-1	-2	-6%	Performance deteriorated
	Suggested	-1	-1	-1	-3%	Performance deteriorated
A28	Default					Level A
	Otter.ai/Test	+1	+1	+1	+3%	Performance Improved
	Suggested	+4	+3	+3	+10%	Performance Improved
A29	Default					Level B
	Otter.ai/Test	-5	0	-3	-8%	Performance deteriorated
	Suggested	-5	-2	-3	-10%	Performance deteriorated
A30	Default					Level B
	Otter.ai/Test	-1	+1	0	0	No Change in Performance
	Suggested	+5	+2	+1	+8%	Performance Improved

Dif.= Differences; Ms.= Message; SI & Lang.= SI Skills and Language; Aud.= Audience. Level A: ≥ 80%; Level B = 60-79%; Level C< 60%.

From this Table, the following remarks can be made. According to the Default performance, 9 (30%) trainees are Level A, 13 (43.3%) Level B and 8 (26.7%) Level C. Noticeably, 5 (16.6%) trainees improved both their Otter.ai and Suggested performances. Both performances deteriorated in 10 (33.3%) cases. One of them improved in 14 cases (46.6%). 1 case had no change in Otter.ai but improved the Suggested performance, see Figure 13. The significant number of trainees (33%) whose Otter.ai and Suggested performance deteriorated far exceeds those (17%) whose two performances improved, a matter which may affect negatively our impression and consequently our decision about the use of STT applications in SI. Yet, the 47% of those who had one of the two performances improved can emerge as a possible positive result. The final conclusion will be reached when all the statistics are completed.

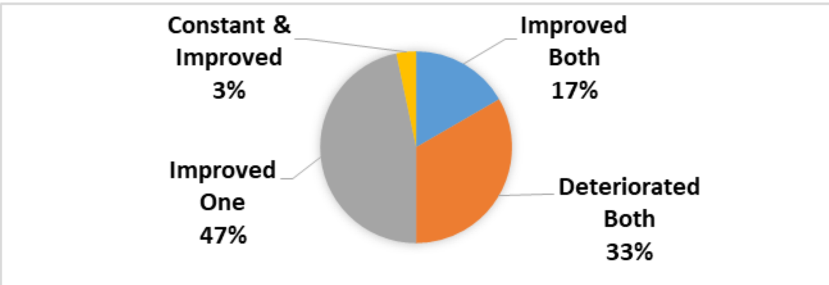


Figure 13: Change in Performances

The Final implications of Table 2 are presented in Table 3:

Table 3: Average Change in Both Experiments

Category	Change in Otter.ai/STT Model	Change in Suggested Model	Difference Remarks
Average change in relation to improved performances	+4.8% (For 5 Trainees)	+7.7% (For 21Trainees)	Improve
Average change in relation to deteriorated performances	-6.6% (For 25 Trainees)	-5% (For 9 Trainees)	Improve
Descriptors: Ms.	-1.9%	+1.9%	Improve
Descriptors: SI& Lang.	-0.16%	+1.3%	Improve
Descriptors: Aud.	-1.6%	+0.3%	Improve
Total Performance	-4.3%	+3.4%	Improve

STT/Otter.ai. The performances of 5 trainees improved using the STT tool (Otter.ai) with an average change of 4.8%, while the performances of the other 25 trainees deteriorated with an average of -6.6%. That's to say the STT tool improved 5 and deteriorated 25 performances, a failure! Also the percentages are not major. Additionally by comparing the two percentages, we will find that deterioration percentage is bigger than improvement by -1.8%, again a failure of the tool.

Suggested Model. The performances of 21 trainees improved using the suggested improved model with an average of 7.7%, while the performances of 9 trainees deteriorated by a percentage of -5%. That's to say the suggested model improved 21 and deteriorated 9 performances, a success despite the modest percentage of improvement. Still deterioration here is better than Otter.ai's in regard to the number of trainees and the percentage. The difference between the two percentages is 2.7%, a modest success.

Message. It is noticeable that the Message Fidelity and Accuracy are negatively affected by Otter.ai with a percentage of -1.9%. This Percentage improved in the Suggested model to 1.9%, a matter which indicates the failure of STT tool and the modest success of the Suggested model. Failure may be attributed to the very complex nature of SI which needs full attention during hearing and understanding the original message, deverbalisation and reformulation of the rendition. It seems that the STT transcription distracted the trainees causing an average negative deterioration in all the investigated categories. WER (about 5%) and latency are likewise added to the additional cognitive load required by the trainee in reading the transcription resulting in such a failure.

SI Skill & Language. Again the STT model proves its failure concerning this descriptor with a negative percentage of -0.16%. Meanwhile the Suggested model showed a modest success of 1.3%. SI skill is negatively affected by the cognitive load, WER and latency, although some participants expected that the STT tool would help in understanding difficult, unclear or missed words. Even the improvement (1.3%) achieved in the light of the Suggested model is minimal in relation to their expectations.

Audience. The same holds good to the Audience descriptor. Otter.ai failed with a percentage of -1.6% because the trainee seemed hesitant and sometimes inconsistent in his attempt to compromise simultaneously between what he hears and what he reads with some errors and latency. The Suggested model, which overcomes errors and latency, modestly succeeded with 0.3% for the previous reason.

Total Performance. Hence, the average total performance using Otter.ai failed with a percentage of -4.3%, whereas the average total performance using the Suggested model succeeded modestly with 3.4%.

By and large the Suggested model is better than the STT model regarding the average improvement vs. deterioration, the average of the three descriptors and the average total performance with a range of percentages from a maximum of 7.7% to a minimum of 0.16%.

Levels. It seems, according to the study results (see Table 3 above), that Level C trainees have benefited most from the experiments: from a total of 8, 1 trainee improved both Otter.ai and Suggested, 7 improved one of them and none deteriorated any of the two. However, improvements are generally modest as explained before. Among 9 Level A trainees, 2 improved both, 4 improved one of them and 3 deteriorated both. Similarly, out of 13 Level B trainees, 2 improved both, 5 improved one of them and 6 deteriorated both. This can indicate that the first experiment (Otter.ai/STT) is dysfunctional for all Levels and that the second experiment (with the Suggested improved model) is functional, though with modest results, for Level C trainees rather than Levels A and B. A transcription may represent an opportunity to lift up Level C trainees' weak performance and perhaps language.

CONCLUSION

From the beginning of this study, the aim was made clear, namely to explore the functionality or otherwise the disfunctionality of using an AI tool, like STT, in human SI training given to undergraduate university students. It raised 5 research questions and attempted to answer them through both a qualitative and quantitative methodology using analysis, comparison, assessment, questionnaire and experiment as research tools with an SI theoretical framework derived principally from Seleskovitch.

The study investigated the opinions of SI participants/trainees about the use of STT technology (like Otter.ai application) in their training before the experiments. It found out that they have some knowledge about AI and little or no information about using AI in SI. They use STT applications for social and educational purposes mainly and 73.2% expected that STT to be useful in interpreting. But they seemed not aware of possible distraction when such a tool is used during SI: 6.6% expected distraction, 35% did not and 40% remained in between. Expected opportunities included focus, comprehension, hearing better, getting accent, helping in fast speeches and memorization, while the challenge was mainly possible distraction.

After the experiments they were asked to give their opinions about the use of STT technology in another questionnaire. 71.6% found the STT tool useful, whereas 28.4% found it useless; 50% found it easy and 23.2% difficult, the rest were in between. When asked to rate the usefulness vs. the uselessness of the tool on a scale, 28.3% chose Very Useful, 30% Useful, 28.3% Neutral, 11.6% Useless and 1.6% Very Useless. This is why 60% would recommend it in SI training while 8.3% would not. Opportunities and challenges were more or less the same as in the first questionnaire but their realization of the cognitive load of possible distraction of the tool increased after the experiment. Until that stage, the positive results were mostly modest and challenges increased.

The results of the two experiments, the STT model and the suggested improved model, were compared to the trainee's default performance and the study have reached the following conclusions. STT improved very few performances and the average improvement was modest. The suggested model proved to be better than STT as the performances of 21 trainees improved with an average of 7.7%, while the performances of 9 deteriorated by a percentage of -5%. This means a success despite the modest percentage of improvement. Also deterioration in performance using the latter model is better than STT's in regard to the number of trainees and the percentage. The difference between the two percentages is 2.7%, a modest success.

In regard to the three descriptors of the rubric used in the assessment of the performances, we can notice first that Message Fidelity and Accuracy were negatively affected by STT with a percentage of -1.9%. This Percentage improved in the Suggested model to 1.9%, a matter which indicates the failure of STT tool and the modest success of the

Suggested model. Failure may be attributed to the complex nature of SI which necessitates full attention during comprehension, deverbalisation and reformulation stages. In fact, STT transcription seemed to distract the trainees causing an average negative deterioration in all the investigated categories. WER and latency are added to the trainee's cognitive load resulting in such a failure. Second, the STT model also proves its failure concerning SI Skill and Language descriptor with a negative percentage of -0.16%, meanwhile the other model showed a modest success of 1.3%. SI Skill is negatively affected by the cognitive load, WER and latency, although some participants expected that the STT tool would help in understanding difficult, unclear or missed words. Furthermore, the improvement (1.3%) achieved through the Suggested model is minimal in relation to their expectations. Third, the same applies to the Audience descriptor, where STT failed with a percentage of -1.6% because the trainee seemed hesitant and sometimes inconsistent. The Suggested model, which overcomes errors and latency, modestly succeeded with 0.3%.

Generally, the average total performance using Otter.ai failed with a percentage of -4.3%, whereas the average total performance using the Suggested model succeeded modestly with 3.4%. The study also found out that Level C trainees have benefited most from the experiments. To sum up, the STT experiment has proved dysfunctional for all Levels and the suggested model experiment proved functional, despite the modest results, for Level C trainees particularly which implies that transcription may represent an opportunity to lift up Level C trainees' weak performance and perhaps language.

This topic needs further research and experimentation because of the potential significant implications of such studies on SI profession, field, discipline and training.

Acknowledgements

I would like to thank **Mehiar Abdel-Aziz Al-Mansour** and **Yarah Abdel-Aziz Al-Mansour**, senior students at the Faculty of Engineering, MSA University, for their help in getting this tremendous work done. Their professional advice and technical support were invaluable before and during the experiments.

Appendix of Abbreviations

AI Artificial Intelligence

ASR Automated Speech Recognition

CAI Computer Assisted Interpreting
CAT Computer Assisted Translation
MI Machine Interpreting
MT Machine Translation
RI Remote interpreting
SI Simultaneous Interpreting
SR Speech Recognition
STT Speech-To-Text
WER Word Error Rate

REFERENCES

Ahmed, Safa'a Ahmed. (2022). Technology and Artificial Intelligence in Simultaneous Interpreting: A Multidisciplinary. *Occasional Papers in the Development of English Education*, *CDELT*, 78(1) April, 325-353.

Ahmed, Safa'a Ahmed. (2020). Quality Assessment of Simultaneous Interpreting: Teaching and Learning Perspective to English and Arabic Renditions. *Manchester Journal of Artificial Intelligence & Applied Sciences*, Vol 1(1), 55-61. Also published as conference proceedings, *Applied Research International Conference on Education, Literature & Language* (ARICELL) 2020, 24-25 February 2020, Oxford, United Kingdom, 54-69.

Ahmed, Safa'a Ahmed. (2018). Omission in Simultaneous Interpreting: A Multidisciplinary Perspective to U.S. Presidential Debates. *Occasional Papers in the Development of English Education*, *CDELT*, 65(1), July, 469-511.

Ahmed, Safa'a Ahmed. (2016).Translation and Shaping the Arab Identity in a Post-colonial Globalized World: A Multi-disciplinary Approach. *English Language and Literature Studies, ELLS*, X(1) (December), 403-456.

Ahmed, Safa'a Ahmed. (2015). Simultaneous Conference Interpreting and Mediation: An Application of a Cognitive-linguistic Approach to Professional Interpreting from English into Arabic. *Philology* (Faculty of Al-Alsun, Ain Shams University, Cairo), 64 (June), 25-52.

Amato, A., Spinolo, N., & Gonzalez Rodriguez, M.J (eds.). (2018). *Handbook of Remote Interpreting: Shift in Orality Erasmus+ Project, Shaping the Interpreters of the Future and of Today*. Bologna: University of Bologna, Department of Interpretation and Translation.

Chesterman, Andrew & Wager, Emma. (2002). *Can Theory Help Translators? A Dialogue between the Ivory Tower and the Word face*. Manchester: St. Jerome.

Desmet, Bart, Vandierendonck, Mieke, & Defrancq, Bart. (2018). Simultaneous Interpretation of Numbers and the Impact of Technological Support. In Claudio Fantinuoli (ed.), *Interpreting and Technology*. Berlin: Language Science Press. pp.13–27.

Dhanraj, Joshuva Arockia et.al. (2020). A Review on Recent Trends and Development in Speech Recognition System. *Journal of Advanced Research in Dynamical & Control Systems*, Vol. 12, 01-Special Issue, 521-528.

Does WER Matter. (20 Jan. 2021). Smartaction. Uploaded 20 January. Retrieved from https://www.smartaction.ai/blog/does-word-error-rate-matter/

Errattahi, R., El Hannani, A., & Ouahmane, H. (2018). Automatic Speech Recognition Errors Detection and Correction: A Review. *Proc. Comput. Sci*. 128, 32-37.

Fantinuoli, Claudio. (2019). The Technological Turn in Interpreting: The Challenges that Lie Ahead. *BDÜ Conference Translating and Interpreting 4.0*, Bonn 2019. pp.1-18. Retrieved from https://www.researchgate.net/publication/336231448.

Fantinuoli, Claudio. (2018a). Computer-assisted Interpreting: Challenges and Future Perspectives. In Gloria Corpas Pastor & Isabel Durán-Muñoz (eds.), *Trends in E-tools and Resources for Translators and Interpreters*. Leiden: Brill. pp. 153-174.

Fantinuoli, Claudio. (2018b). Interpreting and technology: The upcoming technological turn. In Claudio Fantinuoli, *Interpreting and Technology*. Berlin: Language Science Press. pp.1-12.

Filippidou, Foteini & Moussiades, Lefteris. (2020). A Benchmarking of IBM, Google and Wit Automatic Speech Recognition Systems. *Artificial intelligence and Applications and Innovations*, Vol. 583, 73-82.

Gile, Daniel. (1995). *Basic Concepts and Models for Interpreter and Translator Training.* Amsterdam/Philadelphia: John Benjamins.

Huang, Xuedong and Deng, Li. (2010). An Overview of Modern Speech Recognition. In Nitin Indurkhya & Fred J. Damerau (eds.), *Handbook of Natural Language Processing*. Chapman and Hall/CRC. pp.339-366.

Huang, Xuedong, Baker, James Baker & Reddy, Raj. (2014). A Historical Perspective of Speech Recognition. *Communications of the ACM*, Vol. 57 (1), January, 94-103.

Juang, B.H. & Rabiner, Lawrence R. (2004). Automatic Speech Recognition – A Brief History of the Technology Development. 1-24. Retrieved from https://web.ece. ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-.pdf.

Lederer, Marianne. (2010). Interpretive Approach. In Yves Gambier & Luc van Doorslaer (eds). *Handbook of Translation Studies*, Vol.1. John Benjamins, pp.173-181.

Lamberger-Felber, Heike. (2001). Text-oriented Research into Interpreting: Examples from a Case-study. *Hermes*, Vol. 26. 39–63.

Li, Jinyu. (2020). Recent Advances in End-to-End Automatic Speech Recognition. *Journal of Advanced Research in Dynamical and Control Systems*, 12 (1)- Special Issue. 521-528. ResearchGate, 1-27. Retrieved from https://www.researchgate.net/publication/339245470_A_Review_on_Recent_Trends_and_Development_in_Speech_Recognition_System. Retrieved on 24/5/2022.

Mead, Peter. (2015). Numbers. In Franz Pöchhacker (ed.), *Routledge Encyclopedia of Interpreting Studies*. Abingdon/New York: Routledge. pp.286–288.

Seleskovitch, Danica. (1999). *Simultaneous Interpretation: A Cognitive-pragmatic Analysis*. Amesterdam/Philadelphia: John Benjamins.

Seleskovitch, Danica, Lederer, Marianne. (1989). *Pédagogie raisonnée de l'interprétation.* Luxembourg: OPOCE and Paris: Didier Erudition, (2[0] édition augmentée 2002). Trans. Jacolyn Harmer: *A Systematic Approach to Teaching Interpretation*. (1995). Silver Spring: RID.

Speech Recognition Software: Past, Present, and Future. (2021). Summa Linguae. Updated 18 June 2021. Retrieved from https://summalinguae.com/language-technology/speech-recognition-software-history-future/

Trivedi, Ayushi; Pant, Navya; Shah, Pinal; Sonik, Simran; and Agrawal, Supriya. (2018). Speech to text and text to speech recognition systems-A review. *IOSR Journal of Computer Engineering (IOSR-JCE)*, Vol. 20(2-1) (Mar.- Apr.), 36-43.

What is Word Error Rate and Who is Winning. (2022). Sonix. Retrieved May 2022 from https://sonix.ai/articles/word-error-rate)

SOURCES OF DATA

-Belgium Premier Speech. (22 November 2021). Statement at the UN Climate Change Conference #COP26 in Glasgow (to be interpreted from minute 6:11-12:40). Retrieved from https://www.youtube.com/watch?v=Bo6aKazcHMc&t=375s

-Barbados Premier Speech. (1 November 2021). President Mia Mottley Speech at World Leaders Summit of COP26 at Glasgow. (6:44 minutes). Retrieved from http://www.youtube.com/watch?v=gBZBKEg6bIU

-Prince Charles Speech. (31 October 2021). G20 Summit 'The Future of Humanity is at Stake'. (6:55 minutes). Retrieved from http://www.youtube.com/watch?v =FaY0M3bDjT8