universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Can Computer-Assisted Interpreting (CAI) Tools Be Reliable Artificial Boothmates?

Testing the Usability of *InterpretBank*'s Automatic Speech Recognition Feature in a Remote Interpreting Setting

verfasst von / submitted by

Violetta Bereznaya, B.A.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 070 342 331
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Translation Englisch Deutsch
Betreut von / Supervisor:	Univ.-Prof. Dragos Ciobanu, PhD

Table of contents

List of tables

List of figures

List of abbreviations

AI	Artificial intelligence
AIIC	International Association of Conference Interpreters
ASR	Automatic speech recognition
C	Coordination effort
CAI	Computer-assisted (or computer-aided) interpreting
CAIT	Computer-assisted interpreter training
CAT	Computer-assisted (or computer-aided) translation
CIRIN	Conference Interpreting Research Information Network
COVID-19	Coronavirus disease 2019
D	Deletions
ELIS	European Language Industry Survey
EU	European Union
GALA	Globalization and Localization Association
I	Insertions
ICT	Information and communications technology
ISO	International Organization for Standardization
L	Listening and analysis effort/listening effort
LSP	Language service provider
LT	Language technologies
M	Short-term memory effort
MT	Machine translation
P	Speech production effort
P1	Participant 1
P2	Participant 2
P3	Participant 3
P4	Participant 4
PPT	PowerPoint
R	Reading effort
RI	Remote interpreting
RSI	Remote simultaneous interpreting
S	Substitutions
SI	Simultaneous interpreting
SL	Source language
SS	Source speech
ST	Source text
TL	Target language
TS	Target speech
TT	Target text
WER	Word error rate

1. Introduction

1.1 Relevance and contribution

Digital transformation has been underway for decades. Innovative technologies have increasingly been incorporated into operational processes in various businesses, with the ultimate goal of optimizing productivity and reducing costs. In the last few years, however, the speed of digital transformation has accelerated exponentially, especially against the backdrop of the recent pandemic. A new wave of disruptive technologies has made its way into our everyday lives, and the way people interact with technologies to communicate with each other, as well as access and exchange information is no longer considered a prospect for the future but has become a necessity for survival in the current environment. (Kudyba 2020:286)

Digitalization and the continuous development of information and communication technologies (ICT) have had a major impact on society and on a wide variety of facets of life in general, ranging from economics to business, to politics, to culture, and other spheres (Fantinuoli 2019:2). The field of interpreting as well as many other language-related professions did not remain unaffected by this digital shift either. Fantinuoli (2018; 2018b; 2019) refers to this shift as a 'technological turn,' adding that the interpreting sector is currently on the verge of a period of "fundamental and irreversible change" (2019:2), the main driver of which will be technology. Corpas Pastor and Fern (2016:36) also assert that the future of interpreting lies in technology tools. This thesis will build on this latest research and focus specifically on the use of technology tools in the interpreting field.

Fantinuoli (2019:2) explains that the driving wheels of this technological turn are connected to three main factors. First, he describes the 'anthropological drive,' which is based on the human need to accelerate certain processes and the tendency to automate and, thus, facilitate steps in work processes. The second factor he mentions is the 'economic drive,' which ultimately means that in order to optimize and maximize economic performance, processes need to be as effective, as productive, and as cheap as possible. He ascribes the third 'psychological drive' specifically to the interpreting sector. Although there has been a general fear of change and reluctance towards the use of technology in interpreting amongst interpreters themselves which will be discussed later in detail, they must now face the fact that technology has entered the discipline and accept that technological advances and developments and their influence cannot be avoided. The attitude towards such developments has recently slowly

begun to change and interpreting-specific tools are no longer being considered a threat to the profession, but rather as an opportunity to modernize it. Berber-Irabien (2010:208) adds that in order to continue to thrive professionally, interpreters need to adapt to these new realities of a rapidly changing world and learn how to use technology tools to their own advantage. As Pym also states, "technology, for better or for worse, is here to stay" (Pym 2011:5).

Tripepi Winteringham (2010:96) spoke on this more than a decade ago, claiming that the interpreting sector will undergo major changes and be reshaped by the ongoing spread of technologies. She adds that this will possibly lead to a broader use of the new forms of interpreting in preference to the more traditional ones. Given the growing emphasis on cost-efficiency and the ever-growing use of English as the *lingua franca* of communication, she even proposed that remote interpreting (RI), a mode of interpreting which will be explained later in detail, would become the preferred form of mediation in the foreseeable future. Poechhacker (2016:217) agrees that recent developments in telecommunications and digital data processing technology stand to markedly transform the way interpreting is practiced in the twenty-first century. The language services sector should therefore capitalize on these positive developments in order to ultimately boost their potential benefits and thus reinforce agility, cooperation, and efficiency. This would also help identify potential limitations and mitigate risks associated with the enhanced use of language technologies and artificial intelligence in the field of interpreting (European Union Institutions 2019:2).

The interplay of interpreting and technology is a topic that is still underrepresented in the field of interpreting studies. Many researchers (Corpas Pastor 2018:157, Fantinuoli 2018b:153; 2019:10, Ortiz and Cavallo 2018:26, Prandi 2017:76; 2020:7, Tripepi Winteringham 2010:96) agree that there is not enough literature dedicated to the topic of computer-assisted interpreting (CAI) tools, as only a few authors have addressed the topic. This will be discussed in more detail in Chapter 2.3, examining diverse aspects of the interpretation workflow. Moreover, even fewer scholars have addressed the use of technology in the interpreting booth, as well as the degree to which interpreters have embraced technology. This can be attributed to the fact that CAI tools are newer than and not as widespread as computer-assisted translation (CAT) tools, for instance. This has resulted in a lack of empirical tests that are crucial for finding the golden standard in terms of functionalities desired by the target group to be able to optimize their performances (Fantinuoli 2018b:164). Scholars therefore propose that further investigation is necessary to gain more insight on the influence of technology on the interpreting sector and the interpreting process, as well as to prepare future generations of

interpreters to adapt to and keep up with a constantly changing market. Continuing to assess these tools and their impact will not only help to gradually increase the quality of CAI tools, but also to better understand and meet interpreters' real needs. In addition, Moser-Mercer (2003:15) acknowledged the importance of conducting further research to investigate human factors in remote interpreting in order to be able to establish and guarantee the right and appropriate working conditions for interpreters who work remotely. The ultimate goal would therefore be to ensure that the high level of performance interpreters are asked to provide can be maintained without causing additional psychological stress.

1.2 Problem statement and research question

Over the last few years, the first bachelor's and master's theses have been written on the topic of technology in interpreting, which shows an interest for this field not only among trainers and practitioners, but also among trainees (Prandi 2015:49). Building on this previous research, this thesis aims to investigate further by testing the usability and applicability of the computer-assisted interpreting tool *InterpretBank* with an Automatic Speech Recognition (ASR) feature within a remote simultaneous interpreting (RSI) setting, as well as its perception by professional interpreters.

The focus of this research is on ASR in CAI, because it is a feature that is said to have the potential to revolutionize the interpreting sector (Fantinuoli 2019:14). For this reason, it is even more important to properly examine and test these types of CAI tools. Testing the ASR tool within an RSI setting could lead to further results as well, especially with regard to the cognitive load of the interpreter, since simultaneous interpreting (SI) is a mode in which interpreters usually work at the limit of their processing capacity, as Gile (2009:182) suggests. Besides, RSI is increasingly becoming the most frequently used setting at present. Furthermore, since most studies (Desmet et al. 2018, Defranqc and Fantinuoli 2020, Prandi 2015, Pisani and Fantinuoli 2021, Wang and Wang 2019) were conducted with students, conducting an experiment with active professional interpreters may lead to different results. As Daniel Gile (2021:2) points out, while the work of Fantinuoli and his team on ASR and *InterpretBank* has great potential and has delivered impressive results in terms of technical achievements so far, the exclusive use of students as probands in the reported experiments is self-limiting and perhaps even misleading. In this respect, he adds that since professional interpreters have had the time to significantly hone relevant cognitive skills over the years, their interpreting performance with *InterpretBank* would perhaps be considerably better than what experiments

with students have reported so far. He emphasizes that this would not only apply to accuracy "as regards whatever is ASR-ized, but hopefully to overall performance" (Gile 2021:2), which does not seem to have been assessed yet.

In particular, this study is designed, first, to test this ASR tool's impact on both the interpreting quality with regard to accuracy and fluency of dealing with numbers, as there is broad consensus that these represent a major stress factor during SI (Braun and Clarici 1996:87, Desmet et al. 2018:14, Gile 2009:171). Secondly, the professional interpreters' experience when using the tool will be documented. Both aims will be investigated by means of an experiment and pre- and post-experiment interviews. The overarching goal is to provide an answer to the research question of whether *InterpretBank*'s ASR feature provides helpful input and can be considered a reliable 'electronic boothmate' in a remote setting or proves itself to be an additional hurdle for the interpreter.

1.3 Structure of the thesis

Chapter 2 discusses the concept of remote interpreting and computer-assisted interpreting, explaining the characteristics of RI and introducing some software solutions used for CAI in order to provide some insight into recent technological developments in the field of interpreting. Next, Chapter 3 presents a theoretical framework that explains how cognitive resources are allocated during simultaneous interpreting – the widely applied Effort Model by Daniel Gile – and introduces coping strategies used to dealing with difficulties during simultaneous interpreting (SI). Chapter 4 explains the method of investigation, the participants involved, and the software used in this study, as well as the assessment criteria. Chapters 5 and 6 include a thorough analysis of the data obtained in the empirical study and assess this data in relation to the research questions. Lastly, in order to provide deeper insights into the effect of using a CAI tool in a remote interpreting situation, after the acknowledgement of this study's limitations, the conclusion drawn from the results of the experiment will be presented in Chapter 7.

2. Technological advances in interpreting

The language services sector as a whole has been influenced by technological advances that have been embedded into the profession, both indirectly, such as through the invention of the Internet, and directly, such as the implementation of computer-assisted translation solutions for written translation. Some of these technological developments have contributed to major changes in the interpreting sector. Thanks to the introduction of wired systems for the transmission of sound, for instance, which was the first major technical solution for interpreting, simultaneous interpreting has now become the standard form of interpreting in international organizations and multilingual conferences (Fantinuoli 2019:4). Without the use of technology, simultaneous interpreting would not be possible as we know it today. This mode of interpreting also introduced a new mode of processing information and required the professionals to acquire special skills and specific training (Berber-Irabien 2010:25).

However, as Berber-Irabien (2010:26) states, while the field of translation has clearly benefitted from technological innovations and improvements throughout the ages, the progress has not been as evident for interpreting. The only major breakthrough for conference interpreting has been the emergence of the technology used for simultaneous interpreting in the 20th century. It was not until the 2010s that the next truly radical step for the profession started to become evident: the practical and general introduction of the use of computers and ICTs in conference interpreting, such as remote interpreting or the use of CAI tools. Ziegler and Gigliobianco (2018:120) add that throughout the past few years, the general conditions for conference interpreting have continuously been changing, not only thanks to globalization and diversified market needs, but also due to digitalization and remarkably fast developing information and communication technologies.

Fantinuoli (2018:4, 2019:6) identifies at least three areas in interpreting-related technology advances that will be mainly affected by the so-called technological turn: remote interpreting, computer-assisted interpreting, and artificial intelligence (AI) present in machine interpreting. He also suggests that although these three areas are different and independent, they will soon intertwine and become subcomponents of a changing interpreting meshwork and ultimately have a major impact on how interpreting will be performed, perceived, and delivered in the near future. The focus in this research paper is on remote interpreting and computer-assisted interpreting.

2.1 Remote interpreting

Progress in technology has made it possible for conference participants to "meet" without having to be physically present at the given location, creating new opportunities for real-time interaction. Such technologies also allow interpreters to provide their services from a distance, i.e., in remote mode (Tripepi Winteringham 2010:94). The most notable advantage of this type of communication is the extended reach, which is enhanced even further through interpretation (Böhm et al. 2017:20). The earliest form of remote interpreting (RI; sometimes also referred to as 'distance interpreting') is over the phone, the most recent is web-based videoconferencing. There are several ways of setting up such communication events, whether through a three-way call, or a dialog between two parties, or a mutual conference with several participants, all of which can be mediated by a distant interpreter. Poechhacker (2016:184), however, makes a distinction between remote interpreting and videoconference interpreting, with the latter being a scenario where the interpreter is present at the same location as one of the parties.

Mouzourakis (2006:46) makes a more precise distinction between RI and videoconferencing, ascribing RI to situations in which interpreters are not present in the meeting room, but work from a screen and use headphones without necessarily having a direct view of the meeting room or the speaker. On the other hand, he ascribes videoconferencing to a situation where interpreters are still physically present in a booth within the room where the meeting is taking place. Most participants are gathered within this setting and other participants join remotely through a video connection. Corpas Pastor (2018:154) defines video remote interpreting where the videoconferencing is also online as 'cloud interpreting.' In this paper, the term that will be used and adhered to is 'remote interpreting,' which Ziegler and Gigliobianco (2018:122) define as "interpreting of a speaker in a different location from that of the interpreter, enabled by information and communications technology (ICT)."

While until recently only large companies or wealthy organizations could have access to long-distance real-time audio and video communications, advances in ICT have made RI more affordable and widespread. However, this mode of interpreting has its specificities, and several aspects that have to be taken into account by both the interpreters and the participants in the conference or communications event. These include, for instance, its effect on the interpreting process and performance; the importance of the allocation and number of participants involved in the communication; the difficulties in the management and coordination of communication; the setup of the interpreter in terms of ergonomics; or the issue

of having only a partial view of the speaker, the screen with PPT slides, and of the other participants (Fantinuoli and Prandi 2018c:173). Given the increased technical complexity and possible disadvantages due to the remoteness, a remote simultaneous interpreting situation requires strict technical and organizational coordination and more intensive preparation than in-presence interpreting (Böhm et al. 2017:26). Thus, in order to make sure that interpreters can perform adequately and deliver a service of quality, it is necessary to fulfill several technical requirements.

One of the first attempts to lay down the minimum technical requirements to assure adequate performance quality in remote interpreting was made in 2000. The International Association of Conference Interpreters (AIIC) together with other national interpreters' associations, the European Parliament, the European Court of Justice, and the joint conference and interpreting service of the European Union and the World Customs Organization published the Code for the use of new technologies in conference interpreting (Ziegler and Gigliobianco 2018:128). In January 2019, the AIIC Taskforce on Distance Interpreting published their *Guidelines for Distance Interpreting*, which were revised again later in March 2020. These guidelines include best practice recommendations and standards applicable to working conditions for remote conference interpreting. Apart from basic standards that are also valid for simultaneous interpreting situations other than in remote mode, such as provision of access to relevant documents or confidentiality and data protection, these guidelines include special requirements regarding correct equipment and some technical aspects, such as prevention of acoustic feedback and acoustic shock, image quality, lip sync, and latency. In addition, AIIC also put together a list of tips for speakers to follow to make it easier for the interpreters to perform their task in understanding the speakers and getting their message across. These include such aspects as using an external high-quality plugged-in microphone, raising the computer or camera to eye level, and scheduling a rehearsal meeting with the interpreters for them to familiarize themselves with the speaking style of the speaker. (AIIC 2020)

Furthermore, there are several internationally agreed upon and commonly used standards published by the International Organization for Standardization (ISO) available. These norms cover the technical requirements for the equipment used for simultaneous interpreting, which has to be set up and monitored by simultaneous technology experts at big events, as well as for the quality and transmission of image and sound to the interpreters in presence-based interpreting situations (ISO 2603, 4043, 20108, and 20109). Apart from some specific requirements for the modality of Distance Interpreting in ISO 20108, which are

officially referred to as *Simultaneous interpreting — Quality and transmission of sound and image input — Requirements*, the other ISO standards mentioned here have not been specifically defined for distance interpreting. They do, however, contain a number of relevant parameters for simultaneous interpreting that also apply to this usage case (AIIC 2020). Provided that the sound and image transmission requirements can be met in accordance with the ISO 20108 standards, remote simultaneous interpretation can even take place for live events. It is already being practiced in the form of so-called 'interpretation hubs,' i.e., dedicated sites where interpreters can find a professional environment with the necessary equipment such as booths, professional consoles, technicians, etc. (Fantinuoli and Prandi 2018c:165), and with the help of technologies from the broadcasting sector (Böhm et al. 2017:26).

2.1.1 Experiments with remote interpreting

According to Ziegler and Gigliobianco (2018:119) and Fantinuoli and Prandi (2018c:165), experimenting with remote interpretation in conference settings has shown many sides to this mode of interpreting. First of all, from a financial point of view, RI is a much more affordable option due to the great reduction of travel costs for interpreters, speakers, and participants. Furthermore, RI can provide a solution to problems that may arise with regard to availability or other issues. These include, for example, situations where no local interpreters are available for a specific language combination, or other organizational matters, such as if an interpreter cannot be hired at short notice or the room design does not allow for interpreting booths. Due to these reasons, both research teams agree that the interest in this interpreting mode has been increasing in the past few years, which leads to believe that this was even the case prior to the recent pandemic.

As Tripepi Winteringham (2010:94) states, international organizations, primarily the United Nations and the European Union, have already been experimenting with RI for a few years for these reasons. As far as interpreters are concerned, they can benefit from this mode of interpreting as well, as it allows them to work from home or from anywhere else without having to travel long distances and possibly even take on more assignments. Furthermore, from an ecological point of view, distance interpreting can lead to a considerable reduction in the carbon footprint of meetings in the broadest sense (European Union Institutions 2019:8).

This mode of interpreting, however, does not come without some disadvantages. As Tripepi Winteringham (2010:95) argues, RI is considered to be one of the most difficult forms

of interpreting due to the number of drawbacks it presents. Some empirical tests and experiments with remote conference interpreting that have been conducted since the 1970s by the United Nations and more recently by the European Union have shown, among others, flaws such as problems with audio and video quality, the partial loss of relevant, contextual information due to the remoteness, and psychological factors, such as fatigue, headaches, increased stress levels, and loss of motivation and concentration (Moser-Mercer 2003:15).

In addition, several studies showed that interpreters would report a feeling of alienation associated with not being physically present at the event, not having the possibility to get the right feel for the situation, and not being able to interact directly with the other participants of the event. All these problems, the physical and psychological distance, and a lack of control, can be attributed to the condition of remoteness, the limited view of the speaker, and the audience. (Tripepi Winteringham 2010:95, Ziegler and Gigliobianco 2018:120)

Mouzourakis (2006:52) states that initially, conference interpreters consistently rejected RI, mostly due to the poor image and sound quality that was observed in the early videoconferencing and RI tests. However, the more experiments were conducted with RI, the more apparent it became that working in RI was not as popular not only due to the poor technical conditions, but also due to the number of physiological and psychological complaints. A closer look is taken on the effects RI has on the cognitive load in Chapter 3.4.

From the above-mentioned drawbacks, the other two main factors hindering a broader implementation of remote conference interpreting thus far have been technological limitations. These limitations come about due to the lack of sufficient bandwidth for the synchronized transmission of high-quality sound and image, and the more or less general reluctance to using so-called ‘new technologies’ among conference interpreters. In recent years, however, due to globalization, digitalization, and the rapidly growing information and communications technologies sector, the general conditions for RI have been constantly improving. New and better technology tools, hardware and software, as well as network infrastructures have made it possible to simultaneously transmit high-definition video and high-quality sound, offering more possibilities to conduct high-quality online meetings, which, in turn, might help overcome existing technological, physiological, and psychological problems related to RI (Ziegler and Gigliobianco 2018:120).

Fantinuoli (2019:7), however, argues that even though RI, for the most part, was initially perceived negatively by interpreters, due to the improvement of technology over the

last two decades, the results of more recent empirical tests do not seem to provide any clear empirical evidence of negative effects on quality or on the interpreter's general well-being brought about using such technologies. He proposes that although RI will most likely not completely replace in-situ interpreting, there is still a reason to believe that this mode will become 'mainstream' due to all of its advantages. These include enhancing service availability, facilitating its provision, and reducing costs, because in this case, economic forces will prevail over the general feeling of mistrust among practitioners or over interpreters' concerns, such as fatigue, stress, and alienation. He also compared these concerns with those that interpreters felt when simultaneous interpreting was first introduced. There was a fear of a loss of quality in their performance and the belief that the relegation into interpreting booths and not being on stage with speakers or diplomats they were interpreting for would lead to a worsening of the profession's prestige and, ultimately, of their social status. However, as a result of geopolitical changes in the second half of the 20th century, the large-scale introduction of simultaneous interpreting together with the increasing demand for interpreting services in actuality led to a professionalization of the whole sector and, in turn, to a general improvement of the occupational status of interpreters. He therefore proposed that RI might influence the professional market in a similar way, leading to a productivity effect and ultimately generating a new wave of professionalization (Fantinuoli 2019:11).

2.2 Computer-assisted interpreting

Fantinuoli (2018b:155) divides the technologies in interpreting into two categories: setting-oriented technologies, which determine the external conditions in which the interpreting takes place, such as remote and telephone interpreting, and process-oriented technologies, which are designed to support the interpreter during different sub-processes and the various phases of accomplishing the interpreting task, which are described below. The latter aim to directly influence the interpreting process, its associated cognitive load and ultimately the quality of the outcome. One technological development that is considered to be process-oriented is computer-assisted (or computer-aided) interpreting, or CAI.

CAI tools emerged in the world of interpreting in the 2010s and were developed to help interpreters prepare for their job assignments and to provide them with assistance during the individual phases of their workflow: from preparation to the actual interpreting process and to follow-up work after the assignment. Thus, the main purpose of CAI tools is to make

interpreters' work more efficient in terms of preparation and command of terminology and ultimately improve the quality of their performance (Prandi 2018:29).

As Tripepi Winteringham (2010:89) states, the topic of CAI tools was somewhat first addressed in publications back around the turn of the millennium. She explained that only a limited number of studies were dedicated to the practical use of ICT in interpreting, whereas more research was done on telephone interpreting, on Computer-Assisted Interpreter Training (CAIT) and on the development of tools and software programs particularly created for trainees and interpreting students for practicing and studying purposes. Only in the past few years has the topic of CAI software begun to gain more attention, after the first programs designed specifically for interpreters had entered the market (Fantinuoli 2018b:157).

In the past, interpreters relied mainly on their experience, world knowledge, language skills, and talent. They also had hard copies of a variety of resources, such as dictionaries and glossaries. Since conference interpreters mostly work in pairs or in teams, they could also build networks and thus rely on each other for help. With the start of the Internet era, a new access to information was presented to interpreters that they now extensively use to prepare for their assignments (Fantinuoli 2012:1). According to Corpas Pastor (2018:142), there have been some attempts to meet interpreters' needs in preparing for their assignments and assisting them in various interpreting contexts and modes.

When they were first introduced, CAI tools, or 'interpreter-specific ICTs' as referred to by Berber-Irabien (2010:50), were mainly, but not exclusively, used for pre- and post-interpreting research. Most of the CAI tools available today offer features such as collecting corpora, creating, compiling, storing, and managing bilingual or multilingual glossaries and terminology databases. They are either used during the preparation process or during the actual interpreting task to look up specialist terminology when in the booth.

According to Corpas Pastor (2017:7), aside from terminology management CAI tools, software programs designed particularly for interpreters also include note-taking CAI tools (which assist consecutive interpreters with their notetaking), speech-to-text converters (which automatically transcribe speech into text), CAIT tools, and other assisted applications. These other computer-assisted applications that can be considered a type of CAI software include such tools as text-to-speech converters (which allow users to listen to words, texts, e-mails, etc. in several languages and formats, and practice and check pronunciation), and unit converters (which convert units from one system to another).

In her descriptive study from 2010, Berber-Irabien came to the conclusion that terminology databases in the form of linguistic tools that allow the search from various source languages (SL) to various target languages (TL) according to the domain were the preferred method for accessing terminology in the booth. Similar results were obtained by Corpas Pastor and Fern (2016:24), as they identified that bilingual dictionaries and personal glossaries were the preferred tool used by interpreters to search for terminology while interpreting. This can be attributed to the fact that specific technical terms are one of the obstacles that can influence interpreting quality, and such tools could potentially help interpreters use these terms more accurately and consistently, as well as memorize them for future assignments (Desmet et al. 2018:17). The number of available CAI tools, however, is limited, and their functionalities do not always cover all phases of the interpreting process (Prandi 2018).

In order to better understand what an interpreter's workflow consists of, researchers (Gile 2009:175, Kalina 2007:113, Will 2015:182) agree that the interpreting process can be divided at least into three parts: before, during, and after the interpreting task. Kalina (2007:113) defines four dimensions that describe the main components and conditions of an interpreter's workflow. First, the 'pre-process' phase includes trained interpreting skills and expertise, preparatory activities, such as knowledge acquisition, compiling of resources, managing of relevant documents and creation of glossaries – i.e., basic terminology work – and coordination. All these activities take place before the actual event. This phase of the interpreting process is crucial, as interpreters have to do preparatory work mostly in advance and without knowing exactly which problems (i.e., problems with comprehension, terminology, etc.) may arise while interpreting.

The preparatory phase is followed by the 'peri-process' dimension which outlines the external factors and working conditions (i.e., participants, languages, team, documents available, etc.). The 'in-process' phase covers the actual production by speakers and interpreters and any conceivable data related to these. This part of the interpreting process is then followed by 'post-process' work, which mainly focuses on quality assurance. Within each dimension, individual factors can be isolated and measured or assessed separately, as each phase has an impact on the overall interpreting quality (Kalina 2007:113).

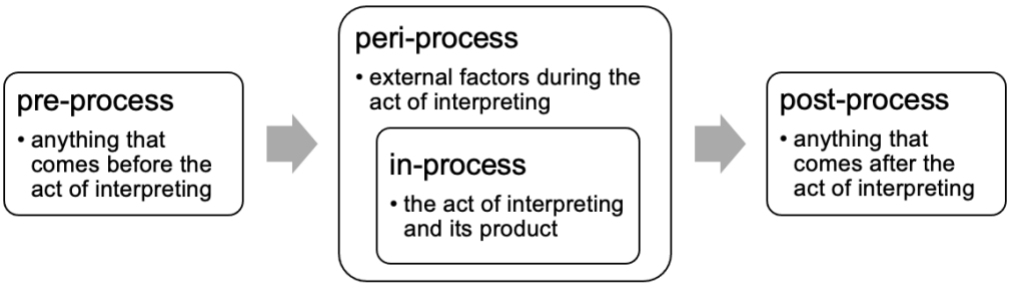

Figure 1: Dimensions of the interpreting process. Own elaboration adapted from Kalina 2007:113

Will (2015:183) describes the last phase of the workflow as the phase of follow-up work, including supplementary terminology work that takes place after the end of the conference.

Fantinuoli (2016:43) adds that aside from being small in number, the software applications developed for computer-assisted interpreting thus far are quite diverse in terms of their design, scope, and functionalities. They also mostly just cater to the needs of the actual developer, who although is usually an interpreter themselves, does not represent all needs and wishes of the interpreting community as a whole (see Chapter 2.2.1). Ortiz and Cavallo (2018:14) provide a detailed review of available CAI tools and classify their use into three categories: training, preparation, and the actual interpreting phase. The authors also make a distinction according to whether or not the tools were created specifically for the community of interpreters, either designed within the academic environment or described by their developers as such.

Furthermore, Fantinuoli (2016, 2018b) distinguishes between first-generation and second-generation CAI tools. The first are rather basic and are designed to help interpreters manage and organize terminology lists and multilingual glossaries in a more systematic way, but do not include any other advanced features. The list of first-generation software tools includes *Interplex, Terminus, Interpreters' Help, LookUp*, and *DolTerm*. Second-generation tools aim to extend first-generation ones by providing more advanced features and functionalities that go beyond basic terminology management. These include, for instance, automatic terminology extraction from preparation documents, support for the memorization of specialist terminology or advanced search algorithms for better acquisition of information from corpora and other resources or help in the organization of textual material. These can all be used in the different phases of the interpreting process—before and after the interpreting

task, as well as in real time, i.e., during interpreting itself. According to his paper of 2018, the list of second-generation software consists of *Intragloss* and *InterpretBank* (Fantinuoli 2018b:164) As Tripepi Winterhingham (2010:88) states, interpreting consists of several exhausting tasks in terms of the cognitive process, such as encoding, memorizing, and decoding messages. These tasks demand a high level of accuracy, which is why one would expect a larger range of technological tools would be available to support those stages. The number of these tools, however, so far remains relatively low.

Several authors list a series of available tools in their papers: Tripepi Winteringham (2010:92) presents a theoretical overview of the usefulness of ICTs in interpreting practice and mentions *Voxtec Phraselator*, a speech recognition software program. Costa, Corpas Pastor and Durán (2014:28) mainly list terminology management tools which were mentioned above, with the addition of the *Interpreter's Wizard*, as well as note-taking programs for consecutive interpreting *(Evernote, Inkeness, Penultimate, LectureNotes, PenSupremacy*, and *My BIC Notes*), but also include unit conversion tools (*ConvertUnits, OnlineConversion, Convert, Converto, Convert Units for Free, Unitsand Unit Converter*) and training programs for interpreters (*Black Box, InterpretaWeb, Linkterpreting*).

The second generation of CAI tools are supplemented and enhanced by the help of AI, which automates many operations that still need to be performed semi-manually (Fantinuoli 2019:8). As Tripepi Winteringham (2010:92) adds, some companies such as IBM have developed so-called speech-to-speech translation systems that are activated by impressive leaps in speech recognition and machine translation, and speech synthesis technology that enable free-flowing communication between any two speakers of the source and target languages. These systems represent a kind of second-generation technology that emerged as an evolution of the first CAI programs, which were developed to enable one-way oral translation through previously produced authoritative translations. Recently, Fantinuoli (2017) developed a prototype for the integration of Automatic Speech Recognition (ASR) into CAI tools. One software solution that features an ASR function is *InterpretBank,* which will be analyzed in Chapter 4.4.

In general, ASR aims at converting speech signals into a sequence of words, either for text-based communication and dictation purposes or for device controlling (Errattahi et al. 2018:32). ASR has been around for over three decades and has been used in many fields, but in the past few years it has been receiving even more attention, with major global companies

such as Microsoft, Amazon and Apple investing considerable funding and research for enhancing their commercial products Cortana, Alexa, and Siri (Fantinuoli 2017:28). The standard metric of assessing an ASR system for its usefulness and performance is the Word Error Rate (WER), which is defined as the proportion of word errors to words processed (ibid.).

Generally, ASR systems are most effective when the conditions are well controlled. As Ortiz and Cavallo (2018:24) suggest, for an ASR piece of software to be used efficiently in different interpreting scenarios it must be speaker-independent. This means that it should not need to be trained to correctly transcribe the speech of each speaker, because that would limit its functional application. On the other hand, designing a speaker-independent ASR, which would be able to recognize any type of speech, is faced with the challenge of understanding pronunciation variations, such as accents or dialects, thereby reducing its accuracy. Thus, the largest challenges for an ASR system are interruptions or distortions to the speech, such as poor articulation, a constantly changing speaking pace, background noise and natural sound, accents, sloppy pronunciation, hesitation, or repetition. Unlike a written text that is grammatically and lexically well-structured, natural spoken extemporaneous language is generally filled with hesitations, unfinished sentences, corrections, etc., which are much harder for a computer to deal with than for a human brain (Desmet et al. 2018:14). Then again, it is important to note at this point that everything in machine translation or ASR mainly depends on the training data. If the training data was ungrammatical, it would still be possible to continue to speak in an ungrammatical manner and still receive whatever quality the target was in the training data. Nevertheless, as Gile (2009:161) suggests, machines still do not have the same amount of lexical and syntactic information and world knowledge as a human interpreter. They will not be able to recognize natural continuous speech with 100 percent accuracy and reliability, even if they may have completed a 'learning process' with the voice of individual speakers. This puts a lot of pressure on an ASR software and underlines the fact that such systems must go through various processes to be able to reach satisfying results.

Besides, interpreters summarize information, while ASR systems transcribe speech input verbatim. It is up to the interpreter to decide what to do with the information delivered by the ASR. Therefore, the utilization of an ASR system in an interpreting situation also strongly depends on the way it presents its output – e.g., whether it is a transcription of a whole speech or solely key units like numbers or specialized terms. In the latter case, using ASR could be compared to having a virtual boothmate who writes down such information for the interpreter, provided that the information is recognized accurately by it.

Furthermore, in order to efficiently use ASR in the booth, the system needs to have a fast response time, which was another prerequisite listed by Fantinuoli for *InterpretBank*, without waiting for a long speech excerpt to solve such an ambiguity. In addition, in continuous speech, it is not easy for a machine to detect the correct separation of words, and ASR focuses on the recognition of words pronounced separately. Apart from the previously mentioned low WER, other prerequisites for *InterpretBank*'s ASR that were pointed out by its developer include the ability to recognize a wide vocabulary range and customize that vocabulary to recognize technical terms (Ortiz and Cavallo 2018:25).

In recent years, however, ASR has made rapid progress, primarily thanks to advances in deep learning and powerful computing platforms. The quality of ASR has also improved drastically, leading to its adoption in many different applications (Ali and Renals 2018:20). Such improvements are expected to continue in the years to come.

Fantinuoli (2019:9) speaks about the various phases of technologization in the interpreting field. Presently, the technologies that are being adopted have the potential to positively influence specific aspects of human interpreters' workflow (CAI) and to make the workflow more productive (RI). The better support CAI tools offer, the higher the probability of them becoming an indispensable element in the interpreters' workflow, just like CAT tools are for specialized translators, which have become increasingly more complex over time, incorporating more and more features. Fantinuoli (ibid.) continues by stating that once some operations start to become automated with the help of Artificial Intelligence (AI) and prediction algorithms, such as the collection of preparatory documents or multilingual terminology, that would eventually lead to a further step. These tools would be used (with human involvement) during the actual interpreting process, which will be discussed later in detail, with interpreters either using them in the booth for simultaneous interpreting or as a reference during consecutive interpreting. He (ibid.) suggests that the final step of technological development in the human-machine interaction would thus be represented by full automation of interpreting. Whether or not it is realistic that the profession will ever enter this third stage is debatable: Tripepi Winterhingham (2010:93) argues that the potential of the human mind and its ability to comprehend meanings and nuances beyond words and clarify any kind of misunderstandings can never be replaced by machines, albeit they may provide fast-paced help for interpreters. She adds that the human presence in interpreting practice is inevitably essential and interpreting, which is considered to be the oldest form of intercultural communication, could never become an alienating profession that will solely be performed by CAI tools, but rather a

profession that will be increasingly supported by new technologies (2010:96). According to Ortiz and Cavallo (2018:26), it seems that although presently, automation technology is close to being able to provide interpreters with truly useful tools, it remains far from being able to achieve the level of maturity required to replace humans. They also add (2018:11) that while people oppose the possibility of technology ever replacing human translators and interpreters, they still continue to seek out software and computer programs that can facilitate their work.

It should also be noted that interpretation is performed under very strict timelines that leave no additional time for proofreading and quality control prior to delivery. Currently, there is no digital or AI model that could combine an accurate low-latency bilingual speech recognition model with efficient machine translation (considering the technical specificities of each domain) and a text-to-speech tool that sounds reasonably natural (European Union Institutions 2019:11). As Böhm et al. (2017:27) state, recent research projects where computers are taught to interpret, such as the Lecture Translator of the Karlsruhe Institute for Technology which offers an English interpretation of a lecture to help foreign students understand what the lecture is about, still do not provide satisfactory results. The system has significant weaknesses as far as "human components" of speech are concerned, such as accents, sentences that are not well-formed, wit or irony, or non-verbal elements such as body language, etc. Although one can argue that these may be of little importance to students in this particular case, according to Böhm et al. (ibid.), they are still considered important aspects that need to be taken into consideration to enable satisfactory results. Whether or not these problems could be overcome, for example, by the use of neural networks, is not foreseeable at present.

2.2.1 CAI tools: perception and use

According to Costa et al. (2014:27), unlike translators, for whom a variety of computer-assisted tools are available, interpreters have not benefited from the same level of automation or innovation. While language technologies have already had a transformative effect on the core tasks of translators, who have already become dependent on specific technologies and beneficially use them in their everyday work (Fantinuoli 2019:5, Wang and Wang 2019:137, European Union Institutions 2019:5), the same cannot be said about interpreters. Computer-assisted translation has now become standard practice and the quality of machine translation (MT) is increasing, which leads to believe that within the next years, translators will most likely spend most of their time post-editing (Desmet et al. 2018:13). It is important to note, however, that there is a difference between syntactic and pragmatic translation and, while MT can handle

the former better, it does not handle the latter. Interpreters, on the other hand, still mostly rely on traditional or manual methods, and as Corpas Pastor (2017:7) states, the technological advances in interpreting have been slow. It can be argued, however, that due to the complexity of the interpreting process, the challenges that interpreters are usually faced with during their work represent a different kind of problem that cannot necessarily be feasibly solved with technology in the same amount of time and with the same resources. Costa et al. (2014:33) claim that even though the number of technologies designed particularly for interpreters is slowly increasing due to the growing interest towards interpreters' needs, they are still "insufficient and unable to fulfill all the necessary requirements" (Costa et al. 2014:33). Ortiz and Cavallo (2018:20) add that a state of constant professionalized development in which companies are interested in regularly developing and marketing solutions has not been reached yet, and that most solutions rather come about as a result of academic research.

The Host Paper presented by the institutions of the European Union[1] ("EU"), which examines the current state of language technologies (LTs) and AI and their use within the EU's translation, interpretation, and conference services, also mentions how in the field of conference interpretation, language technologies and AI have not yet matured into a tool that is able to provide the level of performance required to satisfy a reasonably demanding audience. Quality interpreting as delivered at the level of international organizations not only requires cognitive, cultural, intellectual, and emotional capacities, in particular the ability to use the semantic context in the way the human brain does, but also operational restrictiveness such as security or confidentiality, which software is not yet fully able to offer. Notwithstanding, it is important to mention that technology and software programs including AI could significantly improve and facilitate interpreters' preparation for an event, as well as assist them with useful tools while working and accordingly upgrade their performance in the booth. Furthermore, state-of-the art technology, automation processes, and AI have the potential to revolutionize and enhance supporting services of conference interpretation, such as managing demand for interpretation and planning interpreters' assignments (European Union Institutions 2019:5).

Moreover, interpreters cannot count on the help of or use technology tools in the same way as translators, as they do not have the same amount of time to look for an ideal translation of a certain word or statement in the target language when they are interpreting. They can only use technology without the time constraint during the preparation process, i.e., prior to the

[1] The European Parliament, the Council of the European Union, the European Commission, and the Court of Justice of the European Union

actual performance of the interpreting task, to gather relevant information (Corpas Pastor and Fern 2016:4).

It is worth mentioning that technologies in the field of interpreting are far less developed than in the field of translation. However, since automatic translation keeps delivering satisfactory results, automatic interpreting will therefore be dependent on the advances made in the translation sector and ultimately on the availability of accurate speech-to-text and software (Desmet et al 2018:14).

As many experts and researchers suggest (Berber-Irabien 2010, Corpas Pastor and Fern, 2016:34, European Union Institutions 2019:10, Fantinuoli 2018b:154, Prandi 2017:76, Tripepi Winteringham 2010:89, Wang and Wang 2019:137), the attitude of many practitioners and even academics towards interpreter-specific technologies and the use of CAI tools is, however, still rather negative. Until recently, the impact of technology tools on the profession had been rather minimal. Before this had changed due to recent developments in the interpreting field brought about by the recent pandemic, Fantinuoli (2018b:163) explains that there were several reasons for this.

First, there is an evident degree of reluctance to the use of technology during interpreting, because that could perhaps raise doubts about the pure intellectual activity of interpreting. Second, as Tripepi Winteringham (2010:90) states, the use of these tools has thus far been considered by some practitioners, at least in the booth, to some extent as "unnatural." This assumption can be attributed to the fact that using such tools is usually rather time-consuming and distracting in an activity that requires concentration and fast-paced decoding and delivery. Cognitive processes during simultaneous interpreting have not been completely ascertained (see Chapter 3). This makes it hard to create a software that will enable a smooth workflow and not interfere too much with the already complicated cognitive process.

Another reason Fantinuoli (2018b:163) mentions is that too little effort has been made in investigating the role of software tools in interpreting. Furthermore, from an economic point of view, since interpreting plays a "marginal role" in the language services industry, as Fantinuoli (2018b:163) claims, software developers have not invested as much time and money into the establishment of interpreter-specific software, as it is easier to focus on CAT or MT. This also makes it hard to measure the economic gain in using dedicated software. In this context, there is an evident gap between practitioners and software developers in the interpreting industry. Ortiz and Cavallo (2018:20) add that a state of constant professionalized

development, in which companies are interested in regularly developing and marketing solutions, has not been reached yet. Moreover, universities and other institutions providing conference interpreting training usually do not introduce interpreter trainees to CAI tools (Fantinuoli 2018b:163).

Whether interpreting actually plays a marginal role in the language services industry from an economic point of view, as stated by Fantinuoli (2018b:163), is debatable: according to recent European Language Industry Survey (ELIS) results and other Globalization and Localization Association (GALA) reports on the language services industry as well as Böhm et al. (2017:5), the market for language services, which includes interpreting, is considered one of the most dynamic markets in the world and especially in the EU, and interpreting is a big part of the biggest language service providers' work. Therefore, there is an urgent need to design more programs and technologies that automate the process, boost the productivity, and facilitate the intensive activities of an interpreter in all stages of the work process (Costa et al. 2014:33).

According to Corpas Pastor (2018:141), most technological advances in interpreting have so far been seen as "irrelevant" or even "useless," because they differ so much from what interpreters' work practice is actually like. This is also accompanied by the common concern about technology leading to the loss of quality and the dehumanization of interpreting. Fantinuoli (2018:8) agrees that in the case of RI, this negativity derives from arguments in defense of quality, and as for CAI, in defense of the intellectual dimension of the interpreting activity. Pym (2011:4) claims that professional conference interpreters resist using new technology due to the fear of technologically induced changes, and this type of resistance is usually a "defense of old accrued power, dressed in the guise of quality" (Pym 2011:4). The reason behind this attitude is the unspoken danger of falling behind, of losing market power and ultimately of becoming obsolete. However, as Pym points out, this attitude may change when the profession will "turn the new to suit its own strategic purposes" (Pym 2011:4). On the other hand, however, other conference interpreters state that the use of technology ensures competitiveness in this age of fast information and the conference interpreter who cannot use technology in their work is at a disadvantage (Berber 2009:83, cited by Tripepi Winterhingam 2010:96).

There are still many debates around the impact of CAI tools on both the process and the quality of interpretation. As Prandi (2018:40) argues, most of them are not based on empirical

data which is very limited, but rather on personal beliefs and assumptions that have not been proven empirically. As for any other tool, much depends on the user's personal preferences.

In 2016, Corpas Pastor and Fern conducted a survey among various types of interpreters, including consecutive, simultaneous, whispered, sight, liaison, sign and over-the-phone interpreters, asking them about their needs and requirements related to language technology. The results of their survey showed that the majority of interpreters seemed to either not be aware of the existence of technology resources or to agree that the ones that they do know about do not completely satisfy their needs (Corpas Pastor and Fern 2016:37).

To better understand what the exact needs of those interpreters are, Corpas Pastor and Fern (2016:25) asked the participants about what other resources they use for their work, resulting in the following: consecutive interpreters agree that due to the lack of time, it is not convenient to use technology tools while interpreting in consecutive mode, and most of them only used technology resources during the preparation process, with only 20% or less of the tools being computer-aided translation tools, translation memory systems, machine translation systems, term extractors, termbanks, and corpora. The rest consists of monolingual and bilingual dictionaries, glossaries, thesauri, parallel texts, and other printed materials. Corpas Pastor and Fern (ibid.) claim that this comes to prove that interpreters believe that tools would be of utility. Most consecutive interpreters who work in conferences also mentioned that the original speaker is also a key element, and nowadays more and more speeches are memorized prior to the event, or even read out during the event, which makes the interpreter's job much more challenging. Because of this, various conference interpreters stated that their job would be facilitated much more if the client offered source material beforehand to prepare for the assignment (such as subject matter, information on the speaker, conference organizer, keywords, etc.). Since most consecutive interpreters mentioned that they use pen and paper to take notes, and there are some note-taking applications, such as *Evernote*, that have been designed and are also commendable for the interpreter, in the survey, a majority of the interpreters alluded to the lack of time during the task to consult any kind of technology tools.

Simultaneous interpreters agree that in the booth, there is a high level of pressure and limited time for technology, and most of them stated that during the performance of the interpreting task, they do not rely on support from technology resources. The main resource that they use prior to the interpreting task and during the preparation process is bilingual dictionaries, their own glossaries, thesauri, databases, termbanks, web-based resources, and

audio input material, such as videos and recordings. During the actual interpretation, and depending on the specific circumstances, most interpreters responded that they use their own glossaries, dictionaries, and web-based resources. However, all of them agreed that having a boothmate who acts as technical support by quickly looking up an unknown term on his or her laptop or tablet or writes down numbers and proper names is essential. Whispered interpreters gave results that were similar to those of simultaneous interpreters, because most of the simultaneous interpreters participating in their study also work in whispered interpreting. They also added that glossary management tools in the form of portable software should be available for interpreters. This would give them the opportunity to create their own glossaries and access them should they encounter an unknown term. According to them, the issue does not lie in the lack of tools, but the bigger problem is rather the fact that interpreters do not receive enough information about the assignment beforehand in order to adequately prepare for it. As far as sight interpreters are concerned, they mentioned the need for a quick access tool, if possible, a software application, which will enable them to take a quick glance at a term. Interpreters in this domain also added that they would like to have more background information at their disposal, such as parallel texts, PPTs or speeches. The lack of previous information they receive from the client was also mentioned by over-the-phone interpreters and liaison interpreters as a difficulty in preparing for an upcoming task. (ibid.)

One major conclusion that can therefore be reached form these survey results is that no matter what mode they are working in, all interpreters need as much information about the assignment as possible in order to prepare for it, and during the actual assignment, the majority of interpreters agree that there is simply not enough time to consult technology. Those who do use technology mostly use bilingual dictionaries and glossaries. The interpreters also expressed the need for better specialized dictionaries to cater for different fields. Most of them, however, did mention that they believe that easy-to-use, quick, and effective technology tools would be useful for interpreters and could potentially affect the quality of their performance, agreeing that they would all like to be able to access such tools online, as portable software or as application software (Corpas Pastor and Fern 2016:25).

Böhm et al. (2018:25) also mention the importance of being able to prepare for an event in the context of RI: the more the interpreters are familiar with a communication situation and the interaction to be expected at an event in advance or can familiarize themselves with it, for instance, by rehearsing a procedure beforehand or having a speech manuscript at hand, the better quality can be achieved for a remote interpreting situation.

Scholars such as Fantinuoli (2018) and Corpas Pastor and Fern (2016) agree that the fact that such a low number of interpreters use technology tools can also be attributed to the general lack of awareness of the existence of tools designed for interpreters. According to Fantinuoli and Prandi (2018c:170), interpreters are usually not very well-informed about CAI tools, and that lack of interest could potentially become problematic and counterproductive once technology starts improving even more and entering certain market segments, which will very likely happen much sooner than most practitioners think. This was also proven in a survey conducted by Prandi (2020:3) among interpreter-training institutions on whether they have integrated new terminology management solutions for interpreters into their curricula. The results of the study showed that while all the respondents included technologies applicable to interpreting, only few of them included CAI tools. The reasons listed were, for instance, a reported lack of lecturers able to teach them, a lack of financial resources, and technical and budgeting issues. Prandi (ibid.:7) suggests that there must be a certain degree of confusion around what a CAI tool actually is, because the respondents seem to have interpreted 'CAI' as a synonym for 'ICT,' particularly related to setting-oriented technologies such as remote interpreting and videoconferencing. Although the researchers in this case should have defined what they meant by this term and given examples to the survey respondents in order to obtain more accurate results, they claim that this still may confirm the theory that there is a general lack of information among conference interpreting trainers, which can likely be attributed to the novel nature of such tools. The lack of interest could be attributed to the fact that most conference interpreters already have established their own systems of managing terminology or preparatory documents and are therefore not interested in finding out more and trying out new tools or systems.

As Prandi (2017:76) states, even though CAI tools have not yet received the same amount of attention as computer-assisted translation, the interest for new solutions for providing interpreters with help and assist them in their workflow seems to be growing. The results of the survey conducted by Corpas Pastor and Fern in 2016 (:34) showed that while most of the interpreters believe that they would be useful, some still lack faith in technology tools, whether it's due to the lack of time or because they believe technology tools are more of a distraction rather than a source of help. In their paper from 2018, however, Fantinuoli and Prandi (2018c:162) assert that recently there has, in fact, been a growing interest in technology-based interpreting, presumably by both interpreters and their clients. RI, for instance, which became even more relevant in the past couple of years, can be an effective solution to problems

surrounding the delivery of interpreting services. In the case of CAI tools, they can potentially improve interpreters' performance in highly specialized settings and hereby enhance quality and professionalism, and more interpreters need to be made aware of that fact. Even though at the moment there is so much debate around CAI tools' maturity for use in the interpreting phase, Wang and Wang (2019:137) assure that sooner or later, CAI tools will end up having more advantages than disadvantages for interpreters to embrace and benefit from.

2.3 Current state of research on remote interpreting and CAI

Although several studies have been conducted on the use of technology in interpreting as described below, there are still some questions yet to be answered and research topics yet to be examined thoroughly.

Fantinuoli and Prandi (2018c:173) suggest that more experiments should be conducted with RI by simulating conferences and real-life situations, as it would improve the familiarization with the anomalies of this mode of interpreting. It would shed light on the various implications and risks that derive from distance interpreting and that must be taken into consideration. Ziegler and Gigliobianco (2018:123) agree that while many interesting aspects regarding distance interpreting have been addressed based on other interpreting specializations, such as legal interpreting, there is still a need to conduct more similar large-scale projects directly or indirectly related to distance simultaneous conference interpreting in other fields as well. Poechhacker (2016:184) proposes that another topic of research that requires more coverage is the effect of RI on the quality of the delivery and on the working conditions of interpreters, especially with regard to such aspects as stress, visual access and psycho-social factors. As this area is evolving more and more, he proposes that another focus of research should be on the impact of technology and digital tools for terminology management and speech recognition on the interpreting (ibid.:218).

Cammoun et al. (2019:125) studied the attitudes of interpreters regarding simultaneous interpretation with text. In their experiment, they observed the interpreters' strategies when dealing with text in simultaneous interpretation, paying attention to both the benefits of simultaneous interpretation with text (SI+T) as well as the pressure applied on the cognitive load. The focus of their study was to determine whether there were any differences in the interpreters' attitudes depending on the moment they received the text: long before, shortly before, just before or right after the speaker began his or her speech. The results of their study

showed that the majority of the participants used the text during interpreting most of the time and found it to be useful, especially when dealing with difficulties, such as numbers, heavy accents, or fast speakers. That said, whether or not they actually enjoyed using the text is another matter. As far as the cognitive constraints of dual input is concerned, the interpreters' perception was that cognitive constraint is difficult to manage only in scenarios where they were given the text just before and just after the speaker has begun talking. Thus, their study led to the conclusion that even though the written text of a speech can be somewhat distracting and cause certain cognitive constraints in some cases, it is possible to propose and apply certain strategies that will help integrate the written text into the interpreting process. Then, the text may ultimately be considered as a useful and helpful tool for enhancing the interpreters' performance (ibid.:129). However, since their survey was carried out on SI with printed additional material and included manuscripts of the whole speech and not just numbers, these survey results do not necessarily apply to SI with text that appears in real time with the help of an ASR system. It is therefore necessary to test ASR system and its usability in another way to reach further conclusions.

In their 2016 survey, Corpas Pastor and Fern aimed to learn about interpreters' relationship with CAI tools, asking them to indicate which tools they use as well as their opinion as to whether the available CAI tools satisfy the needs of the profession. As mentioned before, the results showed that interpreters were not satisfied with the available technology tools or claimed that they did not use them, mostly due to their belief that there is not enough time to do so. It is however worth mentioning that this survey was conducted in 2016, which means that there may have been some changes to this by now. The authors (2016:36) assert that technology tools will have a major impact on interpreting in the near future and call for more research to be carried out in order to raise awareness for future interpreters, so that they can also benefit from these tools.

Due to the fact that the ASR feature in *InterpretBank* is relatively new and has not yet been tested as often as the other features, it would be useful to conduct more tests to achieve more results. Furthermore, Fantinuoli (2017:33) proposed that future research needs to be carried out to determine whether working with an ASR tool will cause the interpreter to experience a visual and cognitive overload or whether the quality of the interpretation of specialized text will increase with the help of the tool.

Although Wang and Wang (2019:135) conducted an experiment as a direct response to this research question, they explored the effects of the ASR tool on the performance, in terms of accuracy and fluency, of consecutive interpreters as well as their cognitive process during interpreting, and not that of simultaneous interpreters. The results of their study showed that the tool improved interpreters' interpreting accuracy, but as far as fluency is concerned, the impact was mixed. Regarding fluency, three participants showed better performance when interpreting with the reference, three showed worse performance and three did not have any change in their performance at all. This result is therefore not conclusive regarding the consequence of using the CAI tool and its effect on interpreting fluency. Another conclusion that could be reached is that the tool particularly benefits those who have better language proficiency and have more interpreting experience, while those who are relatively new to the task and have a low language proficiency were negatively affected. This was shown by the fact that the use of the tool has increased interpreters' cognitive load. Therefore, the present study aims to confirm whether professional interpreters do in fact benefit from an ASR-enhanced CAI tool. In addition, since their study concentrated on the effects of the tool in consecutive interpreting, it is evident that conducting a similar study in simultaneous interpreting, adding to that, in a remote setting, could help reach further conclusions.

Fantinuoli also mentions (2020:74) how in-booth CAI tools have been empirically tested in a number of studies (Prandi 2015, Prandi 2017, Gacek 2015), but none have made use of ASR. Neither were any of them conducted in an RI setting. Furthermore, it must be noted that testing the tool by someone other than the developer and/or without his help is important for obtaining unbiased results. Prandi (2017; 2020) and Fantinuoli (2018b) suggest that there are still not enough process-oriented studies exploring the effects of CAI tools on an interpreter's cognitive load, as most of the studies conducted thus far (e.g., Gacek 2015, Desmet et al. 2018) focus on the product or the quality of the interpretation. Cammoun et. al. (2009:134) have also called for the need for further research, observation, and experimentation with text in SI so that all professional conference interpreters can make better and more efficient use of it and develop strategies for overcoming the cognitive constraints and taking advantage of the cognitive benefits.

One study conducted by Fantinouli and Defrancq (2020:99) assessed the ASR tool's system performance in the context of numbers, the results of which showed that the ASR provision tool helped most participants improve the overall performance and the accuracy with which they interpret numbers, for almost all number types. Since the participants in this study

were students, it was proposed by the authors to find out whether professional interpreters would benefit from the ASR tool to the same extent.

The same question was raised by Desmet et al. (2018) who conducted a similar experiment, albeit with a mock-up system, but also with students, and provided some evidence that the technological support system positively influenced the students' interpreting performance. The results showed that when interpreters have access to technological support with numbers, it improves the overall accuracy on numbers from 56.5 to 86.5 percent, reducing the amount of errors by two thirds, especially errors on complex numbers and decimals, the two categories that are most often interpreted incorrectly. Furthermore, since the design of the experiment carried out by Desmet et al. (2018:25) is not based on automatic recognition of numbers, but rather with the help of Microsoft PowerPoint presentations including the numbers in the speech that were prepared beforehand, they claim that ASR might not achieve the same results, i.e. perfect recognition and minimal latency. Fantinuoli's (2017) experiment on the ability of *InterpretBank's* ASR feature to indicate relevant information, however, showed that the transcription of numerals was completed without errors – i.e., all types of numbers correctly appeared on the screen. Notwithstanding, it is worth mentioning that his experiment was not conducted in a remote setting, which, in turn, may lead to different results.

Results from the study conducted by Desmet et al. (2018:23) with the presentation of numbers show that all participants perceived the technological support as helpful. A few of them, however, reported that they were sometimes distracted by the system, as it caused them to focus their attention on numbers at the expense of other information. It was therefore suggested that future research should be conducted to ascertain whether distraction can be reduced by providing number support in different ways.

Pisani and Fantinuoli (2021:193) conducted an experimental study, the objective of which was to analyze the impact of using ASR on the simultaneous interpretation of numbers. This study compared the performance of two groups of students interpreting a speech dense in numbers, one with the help of a real-life ASR-enhanced CAI tool, the other without any form of technological support. The experiment showed that ASR turned out to be effective in providing support during the interpretation of a speech dense in numbers, which was proven by a significant reduction of the error rate. The error rate amounted to 39.8% among interpreters without technological support and to 14.8% among interpreters with the support of ASR. The participants who used the ASR tool, however, stressed that they could have made better use of

the software if they had had the opportunity to practice beforehand. The researchers also compared their experiment to that of Desmet et al. (2018) who used a simulated transcription and manual synchronization between suggestions and original speech, where the average error rate without technological support was 43.5%, compared to 39.8%, and the average error rate with technological support was 13.5%, compared to 14.8% in Pisani and Fantinuoli's experiment, who claim these results suggest that the quality of today's technological advances, at least in a comparable setting and in the context of numbers, is already mature enough to be used in real-life applications.

As years went by, from a technical perspective, the quality of ASR output has indeed dramatically improved, and as Fantinuoli (2019:10) claims, it has reached a level that seems suitable to be used in the interpreting setting. It remains to be seen whether the same level of quality and the same results can be achieved in an RI setting.

Prandi (2015:29) conducted a pilot study on students' use of CAI tools to look up terminology in the booth while interpreting terminology-dense texts. The results showed that some students excessively relied on the software program, while others perceived it to be more of a distraction. The aim of this study was to determine whether the tool should be incorporated into the curriculum of trainee interpreters, which is why the subject of this study were students working together with boothmates. The results of this study are therefore not applicable to experienced interpreters, which gives another reason to have professional interpreters as probands.

Gacek (2015:79) investigated the use of *InterpretBank* in the booth, focusing on its usability during simultaneous interpreting in comparison to paper glossaries. The analysis of the obtained experimental data showed that the use of interpreter-specific terminology software is more efficient in improving the terminology rendition in terms of correctness and completeness than paper glossaries. The feature that was used in his study was *ConferenceMode* (now changed to *BoothMode*) which enables interpreters to look up terminology while in the booth. It would therefore be necessary to test another, newer feature of *InterpretBank* to reach further conclusions.

Biagini (2016, cited by Fantinuoli 2016:50) conducted a similar study, the aim of which was to determine whether the use of a digital glossary in the booth could be seen as a distraction or as additional support in the interpreting process. In his experiment, he compared the performance of two groups of interpreters of a terminology-dense text, the first group used

InterpretBank and the second a traditional glossary on paper. The results showed that all interpreters had a better performance when using the software, which he suggests could be attributed to the fact that in this case, CAI tools help decrease the efforts needed to search for specific terminology, or simply the time it takes to do so, as opposed to using other traditional methods, thus leaving more capacity for other tasks.

Costa, Corpas Pastor and Durán (2014:2) suggest that in on-site interpreting settings where the clients are in the same place as the interpreter, the possibility to use technological support or tools that assist interpretation is limited. In this respect, one could assume that it would be more convenient to use CAI tools in an RI setting, and this study will attempt to provide insight into this as well.

3. Interpreting and cognition

Simultaneous interpreting is a highly demanding cognitive task and one of the most complex activities the human brain can perform (Prandi 2017:77; Seeber 2011:176; Tripepi Winteringham 2010:88). Several attempts have been made by various scholars and scientists (e.g., Gile 2009, Seeber 2011) to establish a theoretical model for such a highly complex cognitive activity. Due to the purpose and scope of this thesis, the model that was chosen to briefly describe how cognitive resources are allocated during simultaneous interpreting is the widely applied Effort Model by Daniel Gile, which is well known in the interpreting community. The concept and cognitive load model are kept simple and discussed briefly to give a general overview of the cognitive processes during SI. The aim is to use this model in the present study to gain a better idea of what kind of impact a technology tool may have on certain efforts used in the interpreting process.

3.1 Daniel Gile's Effort Model

Gile (2009:159) distinguishes between 'non-automatic mental operations' that require attentional resources and 'automatic mental operations' that do not. While automatic operations happen very fast, non-automatic operations take time; these operations take processing capacity from a limited available supply, and when the supply available for a particular task is insufficient, it causes the performance to deteriorate. He does, however, emphasize that this distinction is a simplification of reality, but it makes it easier to understand the operations making up interpreting, which clearly include non-automatic elements.

The mental operation that takes place during simultaneous interpreting consists of the following components that Gile refers to as 'Efforts,' due to their effortful nature: a listening and analysis effort, a speech production effort, and a short-term memory effort. The listening and analysis effort (L) consists of all comprehension-oriented operations, from the subconscious analysis of the sound waves of the source-language speech through the identification of words to the final comprehension of the meaning of what was heard. The speech production (P) effort comprises a set of operations, ranging from the mental visualization of the message to be delivered to planning the speech output and ultimately performing the speech plan, including self-monitoring and self-correction when necessary. Short-term memory operations (which other scholars sometimes refer to as 'working memory,' due to them being an important determinant of cognitive operations) that last up to a few

seconds are performed during SI without interruption and include the storage of information for later use (Gile 2009:160).

Using these terms and definitions, Gile (2009:167) describes simultaneous interpreting as a process that consists of the three core Efforts: Listening and Analysis Effort L, Short-term memory Effort M and Speech production Effort P, plus a Coordination Effort C which corresponds to resources required to coordinate the three other Efforts. Therefore, the formula can be presented as follows:

$$SI = L + P + M + C$$

The Effort model that was first proposed in 1995 and revised in 2009 was designed for explanatory purposes, and as mentioned before, Gile himself (2009:158) points out that this model is a somewhat simplified demonstration of a complex cognitive process that takes place during simultaneous interpreting. However, due to the complexity of the operations involved, including the sharing of the interpreter's working memory capacity by speech comprehension and speech production, it is hard to evaluate the added cognitive load to each Effort at each task. Therefore, as Gile (2009:169) argues, the Effort Model assumes that cognitive load is added only in the sense that when two Efforts are operated simultaneously, it results in heavier load than if only one effort was being used, and the simultaneous use of three Efforts leads to higher cognitive pressure than the simultaneous operation of two, and so on. It is also important to note that the Effort Model still assumes that the interpreter is not affected by emotion and that, as the task progresses, these emotions do not change, amplify, or impact the performance.

3.2 Problem triggers

For interpreting to proceed smoothly, the total processing capacity requirements may not exceed the interpreter's total available processing capacity. The objective of the Effort Model is to help understand how difficult simultaneous interpreting is and to highlight the well-known recurrent problems that are often mentioned by the interpreting community and in the literature. These so-called 'problem triggers' include proper names, numbers, lists, fast speeches, heavy accents or strong regional dialects, poor speech logic, poor sound, and so on. The Effort Model framework associates problem triggers with increased processing capacity requirements which may exceed available capacity or cause attention management issues. If there is not enough processing capacity for an Effort, it may result in a whole chain of events which can last several seconds and cause a deterioration of interpreting quality.

In that sense, problem triggers can be reduced by preparing for, and acquiring information about, the event where the interpreting is to take place, especially with regard to relevant facts, names, ideas, terms, and expressions that might come up during the conference. This is where CAI tools come in handy. That way, interpreters can also enhance their ability to anticipate and decrease processing capacity requirements, in this case, for the L and P Efforts, thus leaving more capacity for other tasks and reducing risks of saturation (Gile 2009:175). However, during the event itself, numbers pose a major problem trigger and an obstacle to their interpretation because of their low predictability and lack of conceptual representation (Braun and Clarici 1996:87). Because no anticipation is possible and the number can only be understood once it is uttered by the speaker, the L and M Efforts are increased. Against this backdrop, the question arises whether having additional references and faster access to relevant information during the actual interpreting process, e.g., directly in the booth, rather than beforehand, could also help reduce processing capacity requirements and, thus, leave more room for other efforts.

Gile (2009:181) also mentions the efforts that take place during SI with text, which he refers to when speakers read a text which has also been given to interpreters. In this case, a new Reading Effort is added, and SI with text can be modelled as follows:

> Simultaneous Interpreting with text = Reading Effort + Listening Effort + Memory Effort + Production Effort + Coordination Effort.

He does point out, however that while having a written text of the speech may help the interpreter and reduce memory problems and the effect of acoustic difficulties and unusual accents, it may result in increased cognitive load, or perhaps even saturation, arising from the need to simultaneously follow both the vocal speech and the written text. Again, the question arising here is whether having solely the problem triggers (i.e., numbers or proper names) displayed as a written text with the help of ASR as opposed to having the whole speech transcribed as a text would, on the contrary, lead to decreased cognitive load, as it would presumably help reduce the Listening and Memory Efforts. This may confirm the assumption that ASR may potentially integrate better into the interpreting workstation than a traditional CAI tool, i.e., one with a look-up mechanism (Pisani and Fantinuoli 2021:186).

To eliminate a possible narrative that available processing capacity of the brain is enough to cover all the needs interpreting requires and that interpreting failures or errors happen because of insufficient linguistic or extralinguistic knowledge rather than because of

cognitive tension between processing capacity supply and demand, Gile proposes the so-called 'tightrope hypothesis' (2009:182). This hypothesis suggests that most of the time, interpreters work close to saturation. Empirical studies conducted based on this hypothesis showed that when interpreters are asked to interpret the same speech twice in a row, the performance of interpreting still deteriorates in the second round, even when no difficulty can be identified in the source speech. According to Moser-Mercer (2003:12), this can only be explained by the processing-capacity limitations and, thus, higher levels of fatigue. This confirms the idea that interpreter's performance degrades over time, in particular regarding the completeness of information transfer (ibid.).

However, it can be argued that this may also suggest that interpreters have better short-term memory and worse long-term memory. While such an assumption cannot be generalized, according to Poechhacker (2016:108), despite being of obvious relevance to the receptive and productive processes in interpreting, there is little evidence to date of specific long-term memory skills in interpreters. Short-term memory (or 'working memory') resources have become a central concern in research on interpreters' cognitive processing, whereas the role of long-term memory in interpreting has attracted little specific research interest (ibid.). Moreover, the construct of working memory has proved useful for addressing issues associated with the degradation of performance when task demands exceed the individual's ability to perform it with complete success (Liu at el. 2004:20).

3.2.1 Tactics and strategies

Dealing with such factors and problem triggers that induce stress and high processing loads, especially under the temporal and cognitive constraints of SI, requires coping strategies, or 'tactics,' as referred to by Gile (2009:201). Different coping strategies have been dissected by several researchers and practitioners (e.g., Seeber, Chernov) throughout the past years. Shlesinger (1999, cited by Poechhacker 2016:127), suggests distinguishing between 'process-oriented' strategies for coping with high-load-inducing speech input and 'product-oriented' strategies for communicating effectively with the target-language audience.

Gile (2009:201) makes another distinction between 'comprehension tactics' that are used to deal with comprehension difficulties, 'preventive tactics' that are used for instances when time or processing capacity pressure is such that the interpreter believes a problem is about to occur, and 'reformulation tactics.' When a comprehension problem arises, interpreters

may stall, or 'wait,' as referred to by Seeber (2011:190) and delay their response up to a few seconds in order to win some time for thought whilst they receive more information from the source-language speech. If an interpreter didn't properly hear or understand a technical term, name, or number, another comprehension tactic is to try to reconstruct the said word or speech segment in their mind with the help of the context, using their knowledge of the language, the subject, and the situation, i.e., their extra-linguistic knowledge. Another strategy is asking for help from their boothmate or consulting resources or relevant documents in the booth. This is where ASR can potentially come in handy.

Preventive tactics include taking notes when names or figures come up (also a potential use of ASR) or changing the time lag between hearing a speech segment and reformulating it in the target language (TL), thus controlling to a certain extent processing capacity requirements for individual Efforts. If interpreters feel that their memory may potentially be overloaded, they may choose to reformulate speech segments earlier than they normally would, sometimes even prior to having a full picture of what the speaker wants to say, thus saving short-term memory capacity requirements by providing earlier relief to short-term memory. If the source-language speech includes an enumeration, another preventative tactic consists of changing the order of elements in the enumeration by reformulating the last elements first and thus freeing the short-term memory from the information, and then moving on to the other elements. The idea of preventative tactics is to restrain the risks of saturation or individual deficits. (ibid.)

The first three reformulation tactics are the same as those for comprehension issues (i.e., stalling, using the boothmate and consulting other resources in the booth), complemented by several other tactics. If a term or speech segment was not heard or understood by the interpreter, it can be generalized, or it can be explained or paraphrased. If they missed a specific piece or information in the SL-speech because they did not have enough processing capacity available for the Listening and Analysis Effort, interpreters could also either inform the audience or delegates of the information loss, e.g., "the interpreter is sorry, they missed the last number," or refer them to another information source, such as the screen with PPT slides or conference handouts. They may also decide to deliberately not render the information in the TL and omit it completely. Such triage strategy is used either because some information has disappeared from short-term memory or because it appears irrelevant and other information with more value requiring the interpreter's attention may be lost if the unimportant part is rendered as the interpreter detects a high risk of saturation. (Gile 2009:201)

As Poechhacker (2016:128) states, the most widely discussed strategy of SI, however, is the previously mentioned 'anticipation,' which is when a simultaneous interpreter produces a sentence constituent before the corresponding constituent has appeared in the source-language input. There are two types of anticipation: 'linguistic anticipation,' which is the prediction of words based on familiar lexical-grammatical patterns, and 'extralinguistic anticipation,' which describes the interpreter's knowledge of the probability of the speaker reacting or speaking in a particular way according to the context or situation at hand (Gile 2009:173). In general, anticipation serves to reduce uncertainty, thus contributing to more cognitive relief. This is where extralinguistic knowledge and, as previously mentioned, conference preparation become important. While an ASR system will not directly help with anticipating information, it may still provide cognitive relief by displaying vocabulary or numbers on the screen and leaving more capacity for anticipation and, thus, producing.

3.3 CAI tools and cognitive processes

Due to the large number of uncontrolled variables involved in experimental settings related to testing CAI software and human-machine interaction and the complexity of the simultaneous interpretation process, researchers are still working on developing rigorous theoretical frameworks for the design of empirical studies around ASR-based CAI tools (Pisani and Fantinuoli 2021:187). In her paper, Prandi (2018:36), for example, expands the cognitive load models proposed by Seeber (2011) and Gile (2009), arguing that their models represent the allocation of cognitive resources during 'standard' simultaneous interpreting, i.e., without CAI tools or any specific conditions under which the interpreting activity is performed. This means that it does not necessarily apply to the scenario where CAI tools are used, such as if an interpreter queries a terminological database during SI.

In such a case, looking up a certain term will recruit manual-spatial resources. Then, once a list of terminological pairs appears, the interpreter will need to visually identify the term needed on the screen, which will require visual-spatial resources. After identifying the necessary term, the interpreter will read it out loud and thus use visual-verbal resources in the same stage of the process. She therefore proposes that using a digital glossary during simultaneous interpreting results in more cognitive resources being recruited than during standard SI, as "cognitive load is not static, but rather varies constantly during the interpreting process, as a function of the cognitive resources recruited" (Prandi 2018:36). This leads her to believe that the integration of ASR in a CAI tool would require no manual-spatial resources,

unlike the traditional manual lookup, which would, in turn, lower the total interference score (ibid.). However, it must be noted that for this to be the case, both ASR and CAI must fulfill a series of specific requirements to be successfully integrated into the interpreting process, as they cannot automatically know which particular aspects the interpreter has problems with and display them on the interpreter's screen. This includes, for instance, creating a glossary and implementing it into the system beforehand, so that the ASR would be able to recognize and present the necessary extracted data.

When CAI tools are integrated into the interpreting process, they are directly linked to and might therefore influence the cognitive processes that take place during the central tasks of interpreting (Fantinuoli 2018:4). Since CAI tools are supposed to provide help and assist the interpreter during the task, they should not act as an additional distraction or hurdle for the interpreter to be effective.

Prandi (2017:80) states that using an electronic glossary in simultaneous interpreting has something in common with sight translation, because both tasks require visual-verbal resources at the perception and cognition stage and involve a manual-spatial response and visual-spatial resources. Even though this assumption is debatable, as searching for equivalents in a glossary is different from site-translating a source language text, this leads her to believe that performing an interpreting task with the use of a CAI tool or a digital glossary is more difficult than doing so without, since more resources are used and shared.

When using terminology tools in the booth, interpreters need to have at hand the right term at the right moment without affecting the balance of the other various processes that happen during the interpreting task in order to avoid cognitive overload. Fantinuoli (2016:46) describes at least three conditions that should be met to enable the use of a terminology tool during interpreting. First, interpreters need to be able to release resources that can be used for other tasks while in the booth by preempting part of the cognitive load from the interpretation to the preparation process. Second, the looking up activity should be selective and focused; and third, the tool should be designed to minimize the cognitive load added to the interpreting process. As Tripepi Winteringham (2010:91) suggests, the process of looking for the right term results in distraction and loss of concentration for the interpreter, as humans cannot fully focus with attention on two different tasks at the same time. The searching mechanism should therefore be designed in a way that the time constraints and the complex cognitive processes

involved in simultaneous interpretation are taken into account. The use of automated systems may therefore be more convenient in such a scenario.

This is confirmed by Desmet et al. as well (2018:17), where they state that for a knowledge management tool to be as efficient in the booth as possible, reference material must be accessed quickly and not cause any additional pressure on the cognitive load of the interpreter. To accomplish this, they assert that the tool must fulfill several requirements in terms of its design and functionalities, such as good search algorithms, clear presentation and ergonomic operation, and fast retrieval of information. To this effect, they propose that ASR could potentially speed up the look-up process and thus solve the cognitive effort and latency of manual retrieval.

Defrancq and Fantinuoli (2020:76) assert that an automatic support system should only present very accurate information, because otherwise it could not only have a negative impact on both the interpreting quality and process, but also cause the interpreters to reduce their level of trust in the system and perhaps even refuse to use support system altogether. Moreover, the visual input should be presented in an ergonomically suitable and legible format. The latency should be minimal as well – since there is already a time pressure in the booth, interpreters should not have to interrupt their delivery in order to wait for the numbers to be displayed on the screen.

As far as ASR is concerned, Defrancq and Fantinuoli (2020:97) state that ASR causes extra cognitive load, since ASR output is an extra source of information requiring attentional resources. The higher the cognitive load of interpreters, the less likely they are to direct their attentional resources to external sources, and it seems to be amplified by how the ASR output is displayed. In their experiment, the entire transcript of a speech was displayed, with highlighted units of interest. Numbers were displayed as transcribed by the ASR system without performing any kind of normalization (e.g., numbers were either transcribed as digits or as words), and the participants expressed that the running transcript distracted them during their interpretation. This led the researchers to believe that in their experiment, interpreters used ASR the least when they needed it the most (i.e., for very complex numbers). In this regard, they suggest that the experiment could be repeated by changing the visualization function to display only the numbers without the running script, since, according to them, it is reasonable to assume that minimizing the visual input as much as possible could increase the usability of the tool.

In light of this, Pisani and Fantinuoli (2021:194) conducted another experiment which was similar to the one by Defrancq and Fantinuoli, except for the way suggestions were presented to the users: instead of being embedded in the complete transcription, suggestions were shown isolated with no embedded context. The participants who used the ASR tool expressed that the additional task of reading the transcribed numbers did not raise any particular difficulty. However, some probands stressed that it would be useful to get to grips with the software before using it in a real-life situation. Nevertheless, as mentioned before, the experiment showed that ASR turned out to be helpful, which was proven by a significant reduction of the error rate. The error rate dropped from 39.8% without technological support to 14.8% with the support of ASR. In particular, ASR helped the participants avoid phonetic perception errors and improved their performance for complex or very long numbers. The researchers propose that this phenomenon could be partly explained by the larger amount of elements that need to be processed in complex numbers: without technological support, this numerical class could entail more significant memorization and elaboration difficulties. (Pisani and Fantinuoli 2021:195)

Another interesting point was that the sole availability of ASR already made the participants in Defrancq and Fantinuoli's experiment feel secure and enabled them to improve their interpreting accuracy, regardless of whether it was used. This led the authors to the hypothesis that the mere availability of assistance might reassure interpreters and thereby reduce stress, which will ultimately improve their performance. They emphasize, however, that to confirm this statement, more research is needed, and if their hypothesis is confirmed, it would mean that the effect of ASR to cognitive load is even more complex than expected (Defrancq and Fantinuoli 2020:98).

3.4 RI and cognitive processes

Studies have shown that cognition is closely connected to physiological processes that take place in the human body and that alternating the interpreters' acoustic and visual input influences vital functional systems, such as the respiratory system and metabolism, which ultimately leads to an increased level of stress, early onset of fatigue and other phenomena (Moser-Mercer 2003:11). As previously mentioned, considering the technological, communicative, cognitive, physiological, and psychological aspects, the element of remoteness has a significant impact on the technological setup and the requirements regarding components and transmission channels that need to be met in order to allow communication to take place

between the different parties involved. It is also important to take other important aspects into consideration, such as verbal and non-verbal communication elements that need to be acknowledged, captured, and transmitted between the locations. These include such elements as gestures, facial expressions, postures, etc. (Ziegler and Gigliobianco 2018:122)

Moser-Mercer (2003:14) conducted a study on the technical feasibility of remote interpreting and its impact on human factors such as psychological and physiological stress and fatigue as well as the interpreters' performance by comparing remote interpretations with live interpretations. In order to study the effects of fatigue under live and remote conditions, the interpreters' performance was evaluated. When comparing the effects of fatigue on performance for the same interpreters working either live or in a remote condition, there were significant differences: the same interpreter will be less tired and provide a higher-quality interpretation in live conference conditions as opposed to remote conditions, and the error rate increase occurred past the normal end of turn time, i.e., past the 30-minute mark on average. The onset of fatigue under remote conditions, as evidenced by a decline in performance, appears to occur sooner, i.e., somewhere between 15 and 18 minutes into a 30-minute turn.

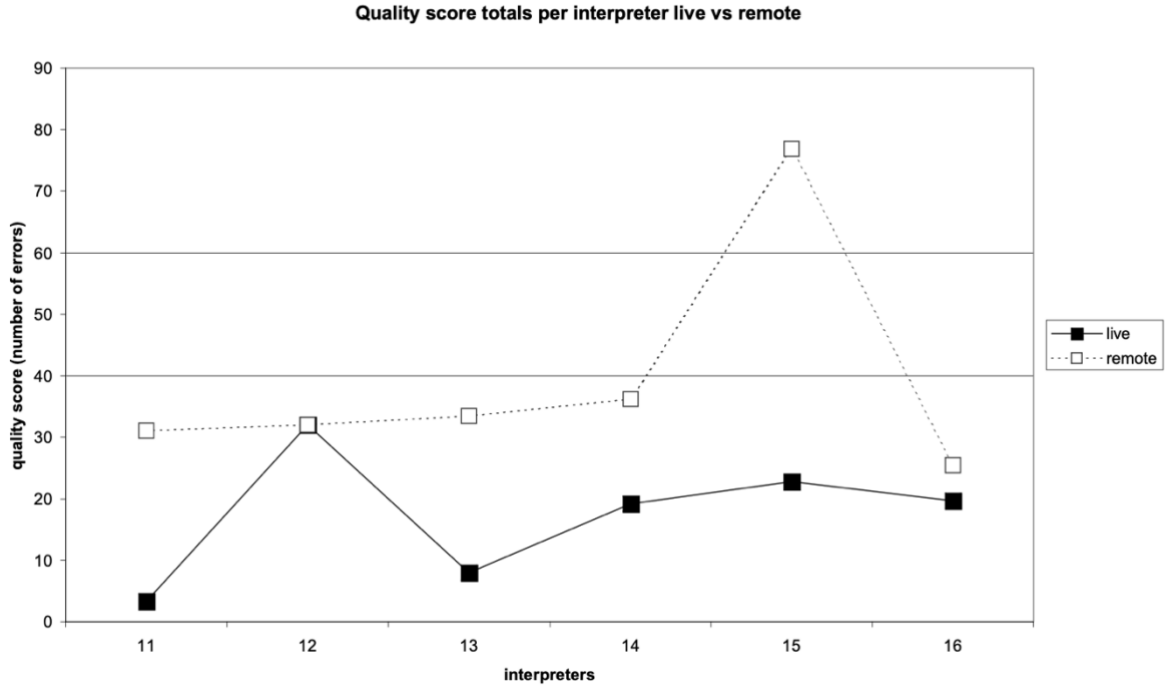

Figure 2: Quality score totals per interpreter, live versus remote interpreting. Moser-Mercer 2003:13

These results led to the conclusion that for the same group of interpreters, working live in a conference room is psychologically less stressful, according to their self-reports, and less tiring, according to performance indicators, and, thus, leads to better performance overall, whereas

interpreting remotely increases an interpreter's mental workload and leads to fatigue faster. These findings were attained by controlling individual differences in performance, by adopting a within-subject design and comparatively assessing the performance of the same interpreters working under two conditions – live and remote – at the same conference, i.e., with the same technical material and often even for the same speakers. The interpreters were working according to a prescribed rotation schedule both in the conference room and at the remote site. A within-subject design was used to eliminate excessive variation in human factors measurements and in output quality. For this reason, Moser-Mercer (2003:14) claims that any difference in performance must be attributed to the interpreters' working conditions and their impact on their performance. She concluded that interpreters seem to be under higher psychological stress when working away from the conference room, mostly because they experience a lack of control of the situation (ibid.).

Nevertheless, the remote interpreting scenario seems not only to present a new environment for interpreters, in which they must use more strenuous problem-solving strategies, but also to induce more than the usual physiological and psychological strain. This can be attributed to the fact that the coordination of image and sound, the piecing together of a distant reality, and the accompanying sense of lack of control all engross mental resources that are already overstretched in this highly complex skill (Moser-Mercer 2003:15). A subject of further research would therefore be to assess if a CAI tool with an ASR feature would help decrease the level of fatigue or the stress level as perceived by the interpreters by providing them with support, which this thesis aims to investigate.

4. ASR in CAI: an empirical study

Given the fact that relevant research calls for more empirical and experimental studies, as mentioned before, the empirical part of this paper is a mixed-method study that aims to test the impact of a CAI tool with an ASR feature used in a remote setting on both interpreting quality with regard to fluency and accuracy of numbers as well as the interpreters' experience when using the tool. The aim is to investigate whether *InterpretBank*'s ASR tool can ultimately be considered as a suitable and usable solution in an RSI setting. The following chapter presents the concrete method of investigation that was used in order to collect and analyze the necessary data and to be able to answer the research question.

4.1 Organization and method of investigation

The experiment consisted of testing the software with the help of four participants who were individually asked to interpret a short number-dense English speech into Russian via online meeting. The entire meeting was recorded.

Prior to the actual experiment, a preliminary test was carried out with a speech and without interpretation to check how the system would display the numbers. During this test, it became clear that the speech may not exceed five minutes, because the online experimental version of *InterpretBank*'s ASR only displays the numbers for five minutes without having to be restarted. It was therefore decided to shorten the speech and limit it to approximately 635 words and to limit the delivery rate so that the system would not be overloaded. The participants received an e-mail one week before the experiment was conducted with information about the experiment and the topic of the speech.

Since a 'normal' RSI setting usually consists of the interpreter hearing the source speech in their headphones and interpreting into the microphone that feeds another channel, whereas some platforms even offer a special interpreting feature, the conference situation simulated in this experiment is somewhat simplified in terms of ergonomics. During the speech, the interpreters were given access to the ASR output by means of the screensharing feature of the teleconference platform *Zoom*. The interpreters were asked to record their interpretation on a voice-recording app of their smartphone and subsequently submit the audio file for review.

Apart from the ASR output as well as the interpreting product that were taken into consideration in the examination, another focus of this research is on how professional

interpreters felt about using such a technology tool in interpreting. Due to the nature of this research, the method of participatory inquiry and primary means of data collection that was used in this qualitative research were interviews. The strategy of the interview that was adopted was 'semi-structured,' because it allowed the researcher to have a set of prompt questions to guide the discussion and the participants to express their thoughts and ideas, thus allowing the researcher to create a balance between simultaneously having a certain level of control and flexibility (Hale and Napier 2013:98). The interview consisted of two parts: several questions were asked before the actual interpreting procedure and several right after.

In order to be able to evaluate the interviews and summarize the insights gained from them, it is essential to transcribe the recorded audio material (Mayring 2016:89; Mayer 2013:47). There are several different approaches to transcription, but since the present study is primarily concerned with the insights and experiences conveyed and the content thematic level is in the foreground (Mayring 2016:91), the interviews were transcribed in standard language (see Appendix). Paralinguistic features and non-verbal utterances, such as stuttering, coughing, unintelligible expressions, sounds of hesitation, etc., were not included in the transcription of the interviews (except for laughter and head nods), as they are not relevant for the present research interest, and in such cases, an elaborate notation system, according to Meuser and Nagel (1991: 445, cited by Mayer 2013:47), is superfluous. Filler sounds, word and sentence breaks as well as stuttering were smoothed out or omitted, word doublings were only recorded if they were used as a stylistic device for emphasis. If a word or sentence was expressed in another language, in this case Russian, it was translated to English in the transcription and written in italics, with an indication in parentheses that it was originally uttered in Russian. In order to make the transcription as clear and as understandable as possible, every fifth line was numbered in the margin, the interviewer's questions and comments were indicated in bold. To keep the interviews anonymous, participants were abbreviated by the letter P and coded in the transcripts as P1-P4 and the interviewer by the letter I. The source speech that was read out to the participants by the interviewer was not included as a whole text in the transcription of the interviews, but only as an indication in asterisks that it took place, as it was the same in every interview and is individually presented in the Appendix.

However, since the interpretations were analyzed in terms of fluency, they were transcribed verbatim, according to the guidelines for conversation transcription by Froschauer and Lueger (2003:223). Every fifth line was numbered in the margin, pauses in speech were marked with dots (...) (one dot per second), unintelligible words were indicated by dots in

parentheses, where one dot was also used for each second. Non-verbal utterances and hesitations were marked as normal text, with the Cyrillic letter 'э' representing the sound. Half-sentences, false starts, and interruptions were marked with a dash (–), followed by repairs and repetitions, if applicable. Stretched speech, such as vowel and consonant lengthening, was marked with spaces between letters. Increased speech rate was indicated by underlined words. For better legibility, all these features were indicated in bold. Grammatical errors such as the incorrect usage of tenses or declension forms did not constitute a subject of investigation, therefore no special attention was paid to them.

4.2 Participants

The participants in this study were sought out according to their professional experience as a conference interpreter. The interpreters were personally contacted by the researcher via *WhatsApp* and asked if they were interested in participating in the present study. After agreeing to take part in the experiment, they were informed that they would be contacted again as soon as more details on how and when the experiment would be conducted were available.

All participants have between 15 and 35 years of interpreting experience and have all graduated from a postgraduate interpreting program. All the participants were native speakers of Russian. Two participants were male, two female.

Before the interviews began, the participants were once again asked for their explicit consent to record the interview, with assurances that the data would be treated confidentially and anonymized. The interviews were recorded using the recording feature in *Zoom*. It was up to the participants to decide whether they wanted their webcam to be switched on or off, as it would not have provided any additional relevant data for the present study. Nevertheless, all participants had their webcams turned on. Before the interpreting session, the participants were asked if they understood the assignment and if their connection was stable enough and if they could hear the speaker and see their screen so as to eliminate any technical difficulties that may have arisen.

Shortly before the speech was read aloud to the participants, they were informed that the speaker would be sharing their screen and the ASR software would be running in the background. However, to ensure validity of the experiment, they were asked to interpret as they usually would under real circumstances, in a normal interpreting scenario. It was therefore up to them to decide whether or not they would consult the software for help. They were also

assured that their interpretations would be treated confidentially and solely assessed in terms of what was relevant for this study.

4.3 Source speech

The source text was a speech in English by a CEO of a multinational conglomerate corporation presenting its performance at an annual general meeting. The speech was selected taking into account the objective of assessing the impact of the source-language difficulty with regard to numbers. In order to achieve the set goal of number density, the speech was shortened and adjusted by including more sentences containing figures and leaving out more specific details about the company's products and the individual financial reports of its subsidies. The speech was then read out loud to the participants in each experiment by the author of the present study. The speaker managed to keep the speech rate within a range of 115 to 132 words per minute for the experiment, which is close to the ideal speed for the interpretation of improvised speeches, as the input rate recommended for SI by advocates of different schools of thought ranges between 95 and 120 words per minute (Seeber 2011:186). The speech included an introduction of approximately one minute with only one number, followed by a descriptive part that progressively included more numbers and ended with a conclusion. The speech contained a total of 36 numerals. Among the number types were positive integers, ordinal numbers, decimal numbers, and dates. The numbers in the source text were compared with the participants' interpretations and assigned to an error category. The entire speech is presented in the Appendix.

4.4 InterpretBank

The software solution that features an ASR function chosen in the present study is *InterpretBank*. Fantinuoli, its founder and developer, defines it as "a software developed as part of a doctoral research project at the University of Mainz/Germersheim. The overall aim of the tool is to create an interpreter workstation which allows conference interpreters to optimize their workflow before, during and after the event they were called upon to interpret [at]" (Fantinuoli 2016:45).

InterpretBank is a modular tool that comprises the following parts: an *Edit* view to create, manage, and modify specialized multilingual glossaries, a corpus-based *Document* view with automatic text collection and terminological extraction to create glossaries using preparatory documents, *Memorization* view to learn conference-related terminology, and a

dedicated conference *Booth* module to look up terms in the booth. All these modules can be accessed in the upper right corner.

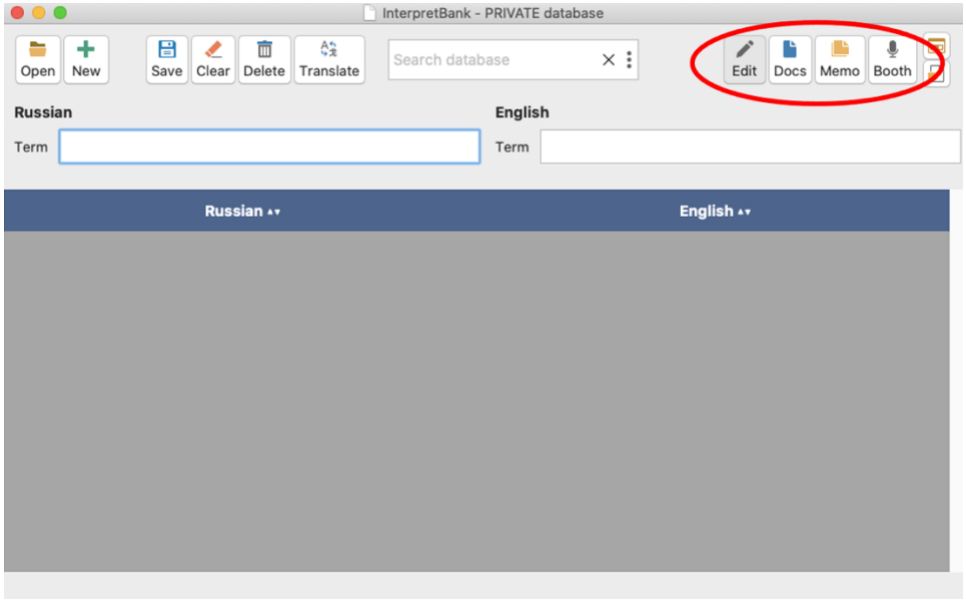

Figure 3: Screenshot of *InterpretBank* interface and its components. InterpretBank 2021

The modules are individual tools designed to complete a particular task of the interpreting workflow, and they are all incorporated into one software and seamlessly interact with each other (Fantinuoli 2012:73; Fantinuoli 2016:46; InterpretBank 2021).

The reason for choosing the trial version of this particular software for this experiment is because on the basis of the survey conducted by Prandi (2020:4) among 25 interpreter-training institutions, *InterpretBank* is by far the most widely used CAI tool in interpreter training. It is important to note, however, that her survey focused only on process-oriented CAI tools as a subset of technologies applied to interpreting and only on the educational setting and was conducted between October and December 2017. According to the results, *InterpretBank* is the tool students are most often introduced to, followed by *Interplex* and *Interpreter's Help*. Other tools that institutions present to their students include, in descending order of popularity, *Terminus*, *Flashterm.eu*, *Glossary Assistant*, *Intragloss*, *LookUp* and *Glossarmanager* (ibid.).

The ASR feature in *InterpretBank* a prototype of a web-based ASR-enhanced CAI tool that is designed to assist interpreters in the booth by transcribing a speech delivered by a speaker in real-time and automatically prompting the interpreter with numerals and their units of measurement as well as translation options for specialized terminology.

> The tool's workflow is straightforward: firstly, the acoustic signal that the interpreter receives in the headset is sent to the sound card of the computer

> equipped with the ASR-CAI tool. The audio signal is then sent to the server with the speech recognition engine which returns the real-time transcript of the speech. (Pisani and Fantinuoli 2021:183)

The experimental cloud-based ASR feature in *InterpretBank* has been designed with two main models of data visualization: the first model presents the entire speech transcript, detecting terminological units and numbers and highlighting these units of interest; the second model (shown below) suppresses the text stream and visualizes only the extracted units of interest, i.e., numbers and units of measurement, in a vertical prompt. That way, the interpreter is presented with less information load, albeit with a loss of contextual information, which aims to potentially reduce the distraction factor and the risk of cognitive overload. (Pisani and Fantinuoli 2021:183)

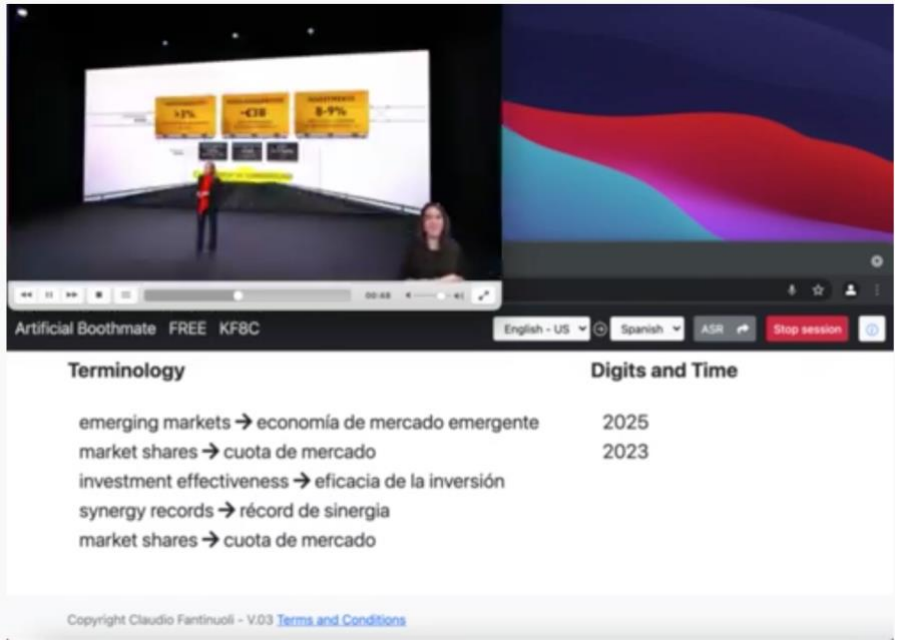

Figure 4: Screenshot of "Video of InterpretBank ASR for numbers and terminology." InterpretBank 2021

In the present study, the online experimental version of the tool[2] "*Augmented interpreter - CAI v 1.2*" was used, which displays only numbers without full transcription of the whole speech. Before starting the program, the source language must be selected in the lower left part. ASR works for the source languages English, German, Chinese, Spanish, French, Italian, Polish, Portuguese, Korean, and Japanese. For all languages except German, Polish, Korean, Japanese, and French, it is possible to choose between the different language varieties: for English, for instance, one can choose between Australia, Canada, New Zealand, South Africa, United Kingdom, or USA. The ASR engine can be activated by clicking on the microphone symbol.

[2] Available online at https://interpretbank.com/app/asr-num/index

The ASR output, i.e., the number and measurement unit, is then displayed in a vertical prompt, with the newest information highlighted in red and showing up on top. The rest of the numbers automatically move down and become smaller in size. The maximum amount of numbers displayed is six, whereas the number at the very bottom will at some point be covered by the microphone symbol.

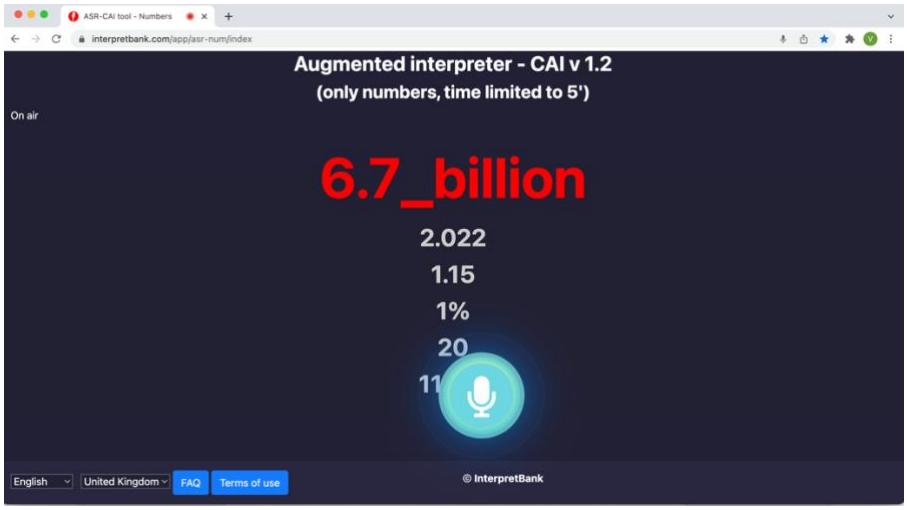

Figure 5: Screenshot of ASR for numbers. Augmented interpreter - CAI v 1.2 2022

Because the ASR feature of *InterpretBank* is still in the prototype stage, it is important perform case studies and test it in order to gain some insight into its effects on interpreting performance and to highlight possible deficiencies or present suggestions for improvement.

4.5 Interview

As previously mentioned, to be able to substantiate findings obtained from the literature with well-founded views from professional practice, the empirical method chosen in this study were interviews. In order to be able to conduct a topic-focused interview that also has a certain structure, it is recommended to create an interview guideline with openly formulated questions (Mayer 2013:28). Despite the use of a guideline, however, interviewers can always respond to other points of conversation, and it is possible to deviate from the fixed sequence of questions at any time (ibid.: 36).

When creating the interview guide, which should serve as a kind of basic framework and helpful support, it is generally possible to include both open and closed questions (Froschauer and Lueger 2003:33). The questions for the interviews were elaborated taking into account the research question as well as the respondents. Although there were occasional deviations from the order of the questions, where some questions were asked earlier or others

were skipped for the time being in order to keep the dynamics of the conversation going, as Mayer (2013:37) describes, and the questions were not always worded identically but adapted to the course of the interview. There were a few open-ended questions so that there was enough room for the respondents' own comments on the issues. Following Hale and Napier, the purpose of the in-depth interview is to "gain an understanding" into the interpreting experience as well as "the meaning they make of that experience" (Hale and Napier 2013:96), so as to gain an insight into the interpreters' opinions on using references while interpreting. The interview guideline with the questions that were asked in the interview is stated in the Appendix.

4.5.1 First part of interview

As mentioned before, there has been a general reluctance to using CAI tools in the booth, especially among professional interpreters. First, the experimental group was asked to specify how long they have been working as a professional interpreter. Then, in order to gain insight into the interpreters' views, familiarity and general experience with RSI and CAI tools, the participants were asked questions before interpreting such as whether they have ever interpreted in remote mode, and if so, how often, what their experience has been like and what their preparation process is usually like, as well as whether they have ever used any CAI tools before and which tool(s) in particular.

4.5.2 Second part of interview

In order to analyze how the interpreters perceived the tool and its applicability as well as their experience with it, after the interpreting session, the participants were asked, following the questions 3, 5, and 7 asked by Wang and Wang (2019:127) in their experiment, whether they felt that seeing the written text helped them or was more of a distraction, how reliable they found the transcription, and how often, if at all, they used it, and to justify their answers.

A common feature of the previous studies conducted on CAI tools is that they solely focus on the product of simultaneous interpreting and the terminological quality of the rendition. In this study, an attempt was made to gain some insight into the effects of CAI tools on the cognitive processes. For this purpose, the interpreters were asked to specify what they were mostly looking at during the speech – the ASR output, the speaker, or anything else. In addition, to get a sense of how they allocated their efforts during SI with text, after a brief explanation of what is meant by 'efforts,' the participants were asked to rank their efforts used (L, M, P, and R), i.e., what they were most concentrated on during interpreting, in order of

importance, with #1 being the effort most concentrated on to #4 being the effort least concentrated on. Furthermore, they were asked to rate how satisfied they were with their performance on a scale of 1-5, as well as their strategies for overcoming any challenges or difficulties that may have arisen during the interpreting session.

In order to gain insight into how a professional interpreter envisions their ideal ergonomic setup when interpreting remotely with such a CAI tool, they were asked to specify what their workplace in this case would look like, e.g., how many screens or any other devices they would use in such a scenario.

4.6 Assessment criteria for ASR output, level of fatigue and interpretation quality

As mentioned before, some empirical tests and experiments with remote conference interpreting showed that some flaws, such as problems with audio and video quality, and the partial loss of relevant, contextual information due to the remoteness, have led to fatigue and increased stress levels (Moser-Mercer 2003:15). Therefore, in addition to the interview questions, as an attempt to comparatively assess their perceived level of fatigue before and after interpreting with the ASR tool, the probands were also asked to perform a task, once before and once after the interpreting session. They were asked to say the alphabet backwards while being timed, the focus hereby being on the times they stumbled, made mistakes, or restarted. This reversed recitation approach was chosen as it is one of the methods that have been widely used in the past as part of standard mental status examinations, with the primary measures of interest being a person's ability to complete such a task and the number of errors made. Such methods have been included in neuropsychological tests as measures of complex attention and speed of information processing. (Williams et al. 1996:651)

After the second part of the interview and after the end of the experiment, the interpretations as well as the ASR output were reviewed. The accuracy of the ASR output of numbers was measured by the Word Error Rate, as it is a straightforward concept and the standard approach to evaluate the performance of an ASR system (Ali and Renals 2018:20). To correctly calculate WER, the number of errors, i.e., the sum of substitutions (S), insertions (I), and deletions (D), were divided by the total number of words. If there are N total words in the reference transcription, then the word error rate WER is computed as follows (ibid):

$$WER = (I+D+S) / N \times 100$$

It must be noted, however, that the ASR system used in this experiment only recognizes numbers and not all words or complete sentences, which is why WER can in this case be replaced with 'recognized number error rate' but calculated the same way.

The interpretation was only examined in terms of fluency and accuracy – in particular, what numbers were said by the interpreters and whether or not they were in line with those said in the source speech – since other criteria are beyond the scope of this research paper. If the interpreters had any difficulties during the interpreting task or were faced with challenges, they were asked about the coping strategies that they used. These strategies were analyzed and in reference to Gile's (2009:201) coping tactics mentioned in Chapter 3.2.1.

While there are hardly any explicit definitions of the concept of 'accuracy' in interpreting and what it consists of, according to Seleskovitch (1968, cited by Tiselius 2015:3), an accurate interpretation is brought about when an interpretation provides a communicative effect in the target language that is equivalent to the understanding achieved by the original listeners. As Tiselius (2015:3) states, accuracy is measured in two complementary manners: first, when the sum of the different parts that build the interpretation (such as omissions, additions, and fluency) is used to measure the product; and secondly, holistically, when the interpretation is measured as an intrinsic whole. In this respect, the interpretation was assessed in terms of how many, if any, omissions, additions, etc. were used. The recordings of the interpretation of numbers were analyzed taking into account the following questions: has the number been interpreted correctly, and if not, what kind of error was made? The errors were categorized following Braun and Clarici (1996:92):

Table 1: Error categorization of numbers following Braun & Clarici (1996:92)

Error category	**Explanation**
Omission	the number is left out or has been replaced by a general expression
Approximation	the order of magnitude is correct, but the number has been rounded up or down
Lexical errors	the order of magnitude of the number is correct, but some of its components have been changed
Syntactical errors	the order of magnitude is incorrect, even though the right components may be present
Errors of inversion	the tens were inverted with the units
Structurally incorrect numerals	the rules for the correct verbalization of numbers are broken
Phonological errors	the error can be explained by phonological confusion in the SL
Self-corrections	interpreter returns to given response and modifies it
Other mistakes	miscellaneous errors that do not fit any of the other categories, or numbers that combine multiple error types

The interpretations were also examined in terms of fluency of the delivery. While there is no consensus on the exact definition of fluency in this context, as it is a multifaceted concept, Rennert (2010:194) defines fluency as a "prosodic feature of speech that can be viewed as a function of a number of temporal variables. It is the complex interaction of pauses, audible breathing, hesitations, vowel and consonant lengthening, false starts, repairs, repetitions and speech rate that creates the impression of fluency or a lack thereof" (Rennert 2010:194). In this study, fluency was measured with the help of all these elements, explained later in detail in Chapter 5.2.1. She also adds that fluency is more than just a matter of style and may, in fact, affect users' and clients' opinion of the quality of an interpretation in terms of performance and intelligibility (ibid.). These are all key components of natural speech and, thus, simultaneous interpreting.

The reason for choosing the impact of ASR on accuracy and fluency is because these two aspects, according to researchers (Wang and Wang 2019:120; Poechacker 2016:133; Tiselius 2015:3; Rennert 2010:103), are among the most important parameters for interpreting quality, and because the CAI tool is expected to have a greater influence on them. The analysis of the interpretations should help shed some light on the terminological quality of SI performed with the support of a CAI tool with an ASR feature.

4.7 Hypotheses

As mentioned before, CAI tools aim to directly positively influence the interpreting process, its associated cognitive load and ultimately the quality of the outcome. This thesis aims to start to quantify and qualify this claim.

Based on the empirical evidence described in the previous sections, the following hypotheses were proposed:

1. when interpreting with a reference presented by the ASR engine, participants would have an improved rendition of numbers and, thus, a higher accuracy performance, but a lower fluency performance.

However, since a lot of pressure is put on ASR to be able to reach satisfying results, the fact that the feature is tested in a remote setting led to the hypotheses that

2. the tool might be overstrained and not show all the numbers: while the overall recognized number error rate was expected to be relatively low (i.e., lower than 20%), it was proposed that the error type that would most likely be present is Deletion (D)

3. the satisfaction of the interpreters with both their interpretations as well as the tool depended highly on the tool's performance: if the ASR engine performed well, the tool would be perceived as helpful; if it performed poorly (number error rate higher than 20% and high latency), it would be perceived as a distraction

Given the fact that simultaneous interpretation is a demanding task that is in generally performed at the limit of cognitive saturation (Gile 2009: 182, Pisani and Fantinuoli 2021:183), it was assumed that the correct visualization of numbers without the need of manual help would have potentially reduced the overall cognitive load. More precisely, as far as the allocation of cognitive efforts involved is concerned, the experiment hypothesized that

4. the ASR feature would help reduce the interpreters' L and M efforts, as the interpreters would not have needed to retrieve the right term in their memory if it was presented to them, albeit at the expense of increasing the R effort, but leaving more capacity for the P effort and for other tasks, and thus reducing risks of saturation. The task most concentrated on would be P, followed by R, L, and lastly M.

Studies on the stress and performance during RI (Moser-Mercer 2003:14) have shown that all stress values for remote interpreting are above values for live interpreting. Therefore, another hypothesis was that

5. even if remote interpretation with ASR increased overall performance, it would lead to fatigue faster, making participants take more time to recite the alphabet backwards and stumble and make more mistakes after the interpreting task than they did before.

5. Experiment findings

Data generated from the experiment session and the interviews were used to test the hypotheses and explore the rationale of the experimental results. In the following chapters, some of the most telling results will be discussed.

5.1 Analysis of the ASR output

The accuracy of the ASR output of numbers was measured by the precision rate and the recognized number error rate. In this study, all cases in which the last number transcription on the screen was incorrect or not displayed at all, even if a previous correct transcription had been displayed, as well as any deletions, insertions, and substitutions, were all categorized as errors.

Table 2: ASR output of numbers

Participant	Number of displayed numbers	Number of correctly displayed numbers	Precision (%)	Recognized number error rate (%)
1	34	32	94.12	11.11
2	35	32	91.43	11.11
3	35	31	88.57	13.8
4	35	33	94.38	8.33

The source speech contained 36 numerals in total. In 3 out of 4 cases, 35 numerals were recognized as such and displayed by the ASR engine. The precision rates in this experiment were between about 88.5% and 94.4%, similar to those in Defrancq and Fantinuoli's (2020:88) experiment, which were between 85% and 100%, and to the precision rate reported by Fantinuoli and Pisani (2020:189) in their paper, which was 81.43%. These figures show that the ASR output is reliable in most of the cases, even while being used in a simulated remote conference.

In some isolated cases, the numbers were correctly transcribed in a temporary version which lasted for a few seconds, but the final version transcribed on the screen was incorrect, for example, '123%' was at first shown as such for approximately three seconds, then changed to '120' and lastly to '3%.' It must be noted that in these cases, there was a chance for the experimental group to read the correct temporary transcription, even if it changed after, without letting the following incorrect transcription affect their performance. In two other cases, the

correct number was only displayed for a second and quickly disappeared right after, without being replaced by any other number.

There was one particular figure 'four' in one and the same sentence of the source speech that was not recognized by the program at all, in any of the cases: "(…) raising our outlook four times over the course of the year." Although the actual reason for that is unclear, it can be assumed that this error was a phonetic misinterpretation, since the number was followed by the word 'times,' the system might have interpreted the number 4 as the word 'for.' In other cases, for instance, the word 'success' was sometimes understood by the engine as the number '6' and displayed as such, but then quickly disappearing again.

In some cases, the real-time correction ability of the ASR system caused the transcription to change several times until the final result stabilized. For instance, 2021 which was pronounced as 'twenty twenty-one' was once transcribed as '2.021,' followed by '20' and '21' as two separate numbers, then finally '2.021' again.

In other cases, the real-time correction of the transcribed number created confusion by quickly and repeatedly changing the displayed digit. In particular, '1 euro and 45 cents' were presented by the engine inaccurately to all participants, with 'one euro' being shown as '1_year,' which was promptly deleted and followed by either '45,' '40,' or '5' or finally turning into '0.45':

Table 3: Display of '1 euro and 45 cents' in all experiments

Participant 1	1_year > 45 > 5
Participant 2	1_year > 5 > 0.45
Participant 3	1_year > 45 > 4 > 0.45
Participant 4	1_year > 40 > 45 > 0.45

This caused some confusion, as the number was interpreted inaccurately by two out of four participants and omitted altogether by one interpreter. However, it is important to emphasize that if the figure had been pronounced by the speaker as 'one point forty-five euros' in the source speech instead of 'one euro and forty-five cents,' it would have led to other display results and perhaps less confusion among interpreters and, thus, possibly more accurate interpretations.

Thus, the most common ASR error type in this experiment was deletion where a number in the source speech was completely missed in the automatic transcription. There were also a

few substitutions where a word was transcribed as a number or a word was transcribed as a different word, such as ‘year’ instead of ‘euro.’

All numbers were presented as Arabic numerals. ‘Million’ and ‘billion’ were always shown orthographically, as well as the word ‘year,’ with an underscore symbol between the number and the word, e.g. ‘1_year,’ ‘1.5_billion.’ The percent sign ‘%’ always promptly appeared after the percentages. Four-digit numbers that represented years, such as 2019, were shown with a dot between the first and second digit denoting the year: ‘2.019,’ which reportedly caused some irritation among interpreters (see Chapter 5.3.2). Participant 2 suggested that even though, “of course, these numbers are probably the least of [their] problems, but still, the way it portrayed the numbers sometimes looked a bit strange” (P2:200). Furthermore, if a six-digit number contained three zeros at the end, such as 150,000 and 300,000, the last three zeros were presented vertically underneath the first three numerals in all cases, whereas the first three digits were highlighted on top, as shown below:

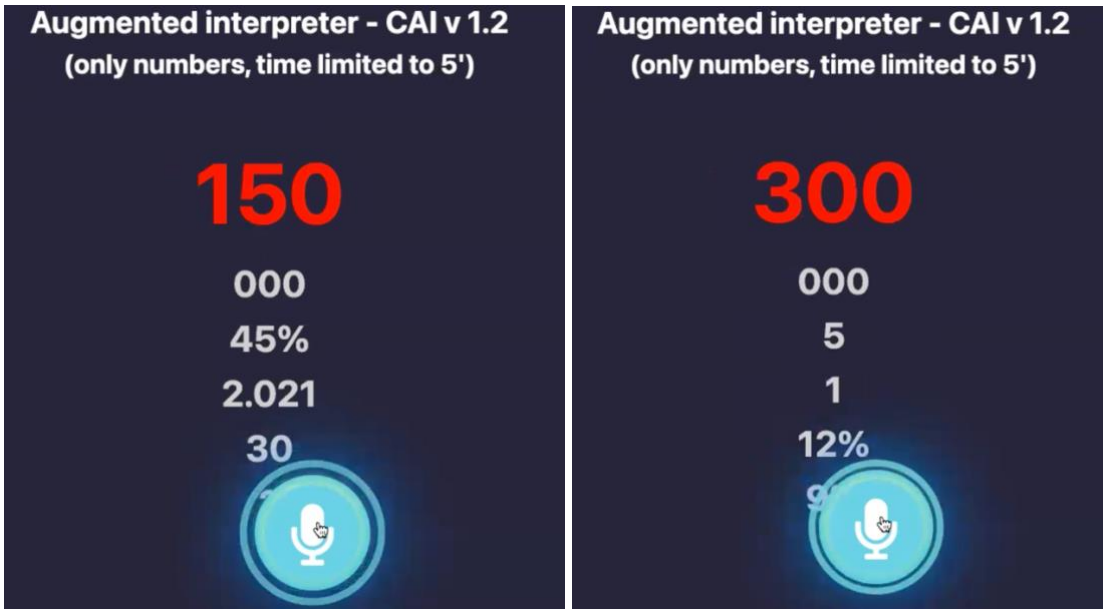

Figure 6: Screenshot of ASR output of six-digit numbers. Augmented interpreter - CAI v 1.2 2022

This seems to have confused some interpreters, as 2 out of 4 of them interpreted these figures as 150 and 300 instead of 150,000 and 350,000 as was stated in the source speech.

5.2 Analysis of the interpretations

While the quality of an interpretation is considered to be an “essentially relative and multi-dimensional concept which can and must be approached with different evaluation methods from a variety of perspectives” (Poechhacker 2016:173), in this paper, the quality of the

interpretation was assessed with regard to the fluency of the delivery and the accuracy of numbers. Grammatical errors did not form a subject of investigation, therefore no special attention was paid to them.

5.2.1 Fluency

To be able to gain some insights into how the CAI tool may have affected the interpreters' fluency, the interpretations of the parts of the speech that contained many numbers, i.e., where the CAI tool displayed numbers and would have been used, were compared to those parts of the speech that did not contain any. The first two minutes of the speech contained 7 numerals, followed by two minutes which more the most number-dense, containing 21 numbers, and finally the last minute of the speech that contained 8 numerals.

In the evaluation of the delivery of the interpreting product, special attention was paid to the number of pauses in speech that were longer than 3 seconds, unintelligible words, non-verbal utterances, hesitations, false starts, self-corrections, repetitions, and words uttered at an increased speech rate. These features were counted as 'flaws' that negatively affected fluency, for lack of a better term. If a hesitation was followed by a long pause, it counted as one 'flaw.' The recordings of the interpretations of each participant were carefully examined, divided into three sections, with each flaw being noted as such and counted in the end. The Russian interpretations are transcribed in the Appendix verbatim, i.e., including all these features that indicated in bold and underlined.

Participant 1's delivery contained several flaws in the first two minutes of the speech, mostly consisting of combinations of words that were pronounced at increased speech rates, amounting to 15 flaws in total. The next two minutes which contained the number-dense part of the speech included 20 flaws, most of which were hesitations followed by longer pauses. These pauses mostly occurred when financial terminology was uttered in the source speech, which leads to believe that the interpreter needed more time to remember and produce a specialized term in the TL, but the numbers were pronounced with no issue. The last minute of P1's speech contained about 10 flaws, most of which were words uttered in an increased tempo.

Overall, Participant 2 delivered a very smooth speech, with only very few flaws during the first two minutes and the last minute of the speech. Most flaws (9 out of 14) could be heard during the number-dense section of the speech, the majority of which were longer pauses and restarts and a couple of self-corrections.

The interpretation provided by Participant 3 contained several flaws that were most present mid-speech, during the two minutes that contained the most numerals. Most of these flaws could be categorized as interruptions and longer pauses followed by hesitations. During that part of the speech, Participant 3 uttered several words with an increased speed of delivery, causing them to sometimes stumble over words and restart. The first two minutes contained several pauses and a few restarts. The last minute of the interpretation, however, apart from very few short pauses, barely had any flaws.

Like the results of the other three participants, the target language speech provided by Participant 4 included most flaws in the middle part of the speech. At the beginning, Participant 4 had a few restarts and repairs, followed by some sounds of hesitation, interruptions, repairs, restarts and even two unintelligible utterances during the number-dense section of the text. The last minute of the speech delivered by Participant 4 was the most fluent with only very few repetitions and sounds of hesitation.

Table 4: Number of flaws per speech segment and participant

Speech segment	**Participant 1**	**Participant 2**	**Particpant 3**	**Participant 4**
First 2 minutes	15	2	15	11
Next 2 minutes	20	9	22	18
Last 1 minute	10	3	6	7
Total	45	14	43	36

All in all, apart from the above-mentioned flaws, all interpreters were able to deliver a coherent speech in the target language that was pleasant to the ear, with clear intonations and enunciations and little to no serious flaws, such as excessive long pauses and incomprehensible utterances. However, apart from Participant 2 who barely had any, sounds of hesitation and longer pauses were present in all participants' speeches (see Appendix for the transcription of the interpretations including these features). The analysis showed that the parts of the speech that contained the most figures and specialized financial terminology seemed to have caused the most trouble for all participants. This leads to believe that the use of the CAI tool during simultaneous interpreting had a negative impact on interpreting fluency. The first two minutes of the speech were also interpreted with lower fluency scores by 3 out of 4 interpreters while the last minute included less flaws in all four cases, meaning the interpreters' performance was better towards the end, which could possibly be attributed to the fact that the interpreters needed some time to adjust or to generally get used to the actual task of interpreting.

5.2.2 Accuracy

The participants' interpretations, as well as the source speech, were recorded and manually checked for number accuracy by comparing the numbers in the source speech with those uttered by the interpreters in the target language. The assessment was conducted by calculating the error rate, i.e., comparing how many numbers were correctly and incorrectly interpreted by each participant, and categorizing the errors in line with Braun and Clarici (1996:92).

Participant 1 accurately interpreted 21 out of 36 numerals, which corresponds to an error rate of 41.67%. Most of the errors (11 out of 15) were omissions, meaning that the number or even the whole sentence containing the number was left out completely or replaced with a general expression, with a few exceptions:

Table 5: Error type analysis P1

Source text	Target text	Error category
1.15	1 and… 1.15%	self-correction and wrong unit of measurement
150,000	150	syntactical error
at one of our subsidiaries	in various branches	other; wrong quantity
300,000	300	syntactical error

Participant 2 correctly interpreted 28 out of 36 numerals, which corresponds to an error rate of 22.22%. Similar to that of Participant 1, all of the errors committed by Participant 2 were omissions, except for one approximation, in which case '19.1 billion' was "rounded" to '19 billion.' At one point, 16% was interpreted as 60% but then quickly corrected to 16%, which may have been a phonological error, as the numbers are near-homophones in English.

Participant 3 correctly interpreted 27 out of 36 numbers, which corresponds to an error rate of 25%. The error types were in almost all cases omissions, except the following:

Table 6: Error type analysis P3

Source text	Target text	Error category
compared to two years ago	compared to the past years	approximation
19.1 billion	20	approximation; although 'billion' was not uttered
1 euro and 45 cents	45 cents	syntactical error

Participant 4 correctly interpreted 25 out of 36 numbers, corresponding to an error rate of 30.55%. While four error types could be considered omissions, as the numbers were completely missing in the target speech, the rest can be categorized as follows:

Table 7: Error type analysis P4

Source text	Target text	Error category
11.5%	15.5%	lexical error
1.15	1015	syntactical error
8.2 billion	8.2 billiard	structurally incorrect; may have been a slip of the tongue
30th	3rd	syntactical error
150,000	150	syntactical error
1 euro and 45 cents	0.45%	syntactical error; wrong unit of measurement
300,000	300	syntactical error

Thus, the average error rate was 29.86%. There were 43 number errors in total, and the most frequent error type among all participants were omissions, followed by syntactical errors and approximations.

Table 8: Total number of errors per error category

Error category	Participants	Total number of errors
omissions	P1, P2, P3, P4	28
syntactical error	P1, P3, P4	8
approximation	P2, P3	3
lexical error	P4	1
structurally incorrect	P4	1
other	P1	2

5.3 Analysis of the interviews

As part of the qualitative content analysis (cf. Mayring 2016), the data obtained from the interviews were subjected to a "text-reducing thematic analysis" following Froschauer and Lueger (2003:158). The aim of this pragmatic evaluation is to summarize an extensive amount of material in a result-oriented and manageable way, so that the essentials of the manifest textual content can be presented in a structured form. This method of analysis makes it possible to provide an overview of the topics addressed in the interviews (Froschauer and Lueger 2003:159). For this purpose, the data obtained from the interviews were divided into thematic sections, which refer to the different theoretical sections of this paper, as also provided for in the interview guideline. The following chapters present the results obtained in both parts of the interviews, whereas Chapter 5.3.1 refers to the questions asked before the interpreting experiment and the following Chapters 5.3.2 and 5.3.3 refer to questions asked after.

5.3.1 RI and CAI tools – general views

After a brief acknowledgment and expression of thanks for their time and effort to participate in this study, the interpreters were asked a few general questions about their professional experience. Participant 1 has been a professional interpreter for 15 years, Participants 2, 3, and 4 for over 25 years.

First, the participants were asked about their experience with remote interpreting. All of them have indicated[3] that they have interpreted remotely before, but not prior to the recent pandemic. These responses lead to the assumption that the only reason the interpreters in this study have worked in remote mode before is solely due to its inevitability brought about by the recent pandemic. However, P3 and P4 stated that they have done some telephone interpreting on several occasions before, albeit without video, which according to P4 "cannot be considered complete remote" (P4:12), perhaps because they would be in the same vicinity as one of the parties.

All of them named several video-conference platforms that they had worked with before, namely *Skype*, *Zoom*, *Webex*, *Microsoft Teams*, *Interprefy*, and *KUDO*. P4 mentioned that they mainly worked with US courts, and the US judicial system decided that they will use only *Zoom* and occasionally *Webex* for their sessions, so their resources are quite limited. P2 mentioned that before the pandemic hit, they were once offered to work in a hub, but it never took place:

> *Actually, I was approached once before the pandemic to do a conference via a hub located somewhere in Stuttgart. It was medical. It didn't come through, but they were planning a medical conference with some live presentations, some surgeries done in Hamburg and simultaneously to be interpreted in many languages. So, it was a hub, you know. Since Corona, a lot of these hubs have sprung up all over Europe and Germany, but they had existed before Corona, obviously. But I didn't get to work there.* (P2:36-41)

When asked about their views on RI and to specify positive and negative aspects of this particular mode of interpreting, 3 out of 4 participants stated that there were mainly negative, apart from P1, who claimed that they "love remote interpreting more than (…) on-the-site

[3] To make it easier to find the quotes taken from the interview, every fifth line of the transcription was numbered. Therefore, the references are line references and not page references. For better legibility, an exact line citation was only added for direct quotes and not for indirect ones. Participants were abbreviated by the letter P and coded in the transcripts as P1-P4. In order to avoid gender references, the gender-neutral pronoun "they" was used for all participants.

interpreting" (P1:24) and that it was a "very, very nice way to spend the pandemic times" (P1:35). P1 did mention, however, that this mode of interpreting has its drawbacks as well.

All of the participants noted that there was too much stress related to mostly to technical issues, such as bad quality of the sound or poor Internet connection. P1 even mentioned one instance when there was a lightning strike and a couple of districts in their area were completely shut down, in which case they did not even have a chance to call the customer over the phone to let them know what had happened. Both P2 and P3 added that one of the major concerns is that usually there is no discipline on the other side, such as when the speakers start talking amongst each other or when they do not use a proper headset. Another frequent problem is when the speakers do not use proper microphones, but instead they use the default microphone on their phone, despite being told otherwise, resulting in poor sound quality. P2 added that some interpreting platforms have a technical support team that is responsible for assisting the interpreters during their work, which sometimes would be more of a hassle than help:

> *I did a conference which lasted 2 weeks and there was always a technician assigned to the Russian booth, and they were monitoring my performance all the time. "You got too close to the mic," "you're too far away from the mic," "we hear some air flow, we hear some background sounds."* (P2:52-55)

As mentioned before, RI allows interpreters to work from home or from anywhere else without having to travel long distances. P2 stated that they felt that RI is more stressful than live interpreting, and that they had expected more comfortable experiences: "like, you know, you sit in your pajamas with a cup of coffee, you don't have to travel, to commute. But there was too much stress related to mostly technical issues (…)" (P2:48).

Moreover, P3 added that when working with a colleague, instead of being physically next to each other in the booth and helping each other by writing down notes for each other, for instance, in an RI scenario, it is necessary to have other means of communication, such as having a WhatsApp chat open and helping each other via messaging. This, however, can sometimes also be distracting and take more time, as the interpreters are so "engrossed in the process" (P3:43) and often do not even have time to type in the chat or look at their phone or their messages. These restrictions have a negative impact on something that is considered to be a key element by many simultaneous interpreters, which is having a colleague in the booth.

P2 added that in RSI, more technical gear is used. They explained that when they interpret in the booth, they only have one laptop, whereas in RSI they have access to a number of screens, using one desktop screen to read texts or any other references relevant to the event

and another screen with the actual conference. P3 stated that it is difficult to manipulate the computer with the glossaries and search the relevant information while also concentrating on the speaker and on the interpretation, making the whole process more intricate. P2 pointed out that in RSI, there are several things that require attention:

> *One of the problems I experienced (…) was that I had to divide my attention between too many functions, and I'm not very good at multitasking. In addition to the usual buttons, for instance, when we were using Interprefy or any other platform like KUDO, I had the microphone buttons, the outgoing and incoming language buttons, I had a special a chat for communicating with my boothmates and Interprefy Technical Support. I had another chat below for the participants and the chair of the meeting and all kinds of documents they were sending. So, my attention was really split, and if I get another screen and another keyboard, I guess this would even increase this multitasking burden. So, I'm not sure which is better, right?* (P2:333-341)

Another aspect that was mentioned by P3 and P4 is that in RI, very often interpreters have no choice but to start to interpret consecutively, even if the original assignment was to provide a simultaneous interpretation. These types of unpredictable conditions seem to occur more often in court RI settings, as P4 states, adding that there have been cases where judges would randomly choose to use a videoconference platform instead of being physically present on court premises:

> *And another very serious drawback of the system is that usually, judges are in a great hurry, because there is a backlog of cases. They want to compress the timing, so even if there would be a possibility to do a proper simultaneous translation, physically it's not possible, because the judges won't let you do this. So it's consecutive, and very limited consecutive. In the Immigration Court, they're mainly using Webex now. So, at some point, it's a little bit easier, but of course it depends. Because sometimes if you are already on the court premises and the judge is using Webex only for his own purposes, you are not hearing it in the headphones, but via some kind of system the judge is using, then again, that's not good.* (P4:19-27)

What makes it even harder is that according to P4, judges mumble and usually speak very fast. Not only is an increased speech rate a major constraining factor during SI (Seeber 2011:186), it most certainly is even harder to deal with when interpreting remotely, given all other restrictions related to this specific mode of interpreting. As P4 reports, even to be able to interpret in court sessions and pass their tests, the courts target their speech rates to at least 150 words per minute, so "you have to do what you can" to "survive" (P4:199).

When asked about their usual way of preparing for an RSI assignment, participants indicated that it does not differ much from that of preparing for a live interpretation. None of them specified any guidelines, such as AIIC or ISO standards that were mentioned in the earlier chapters. The only difference is that there are more double-checks in terms of technicalities, such as making sure that the Internet connection is stable, and all their technical gear is functioning properly. P1 stated that they always restart the computer before every assignment and make sure that all devices, such as their cell phone and tablet, are fully charged, in case anything steps in, such as a power outage, especially because such videoconference platforms as *Zoom* are quite consuming in terms of power. Furthermore, P1, P2 and P3 mentioned that they always have materials and glossaries at hand:

> *I always have a glossary. It doesn't matter if the interpretation live or remote, it's like a tradition for me. Even if it's a topic that I know quite well, even if people will be talking about, like, peace and friendship, I will have a glossary of those two words, just to make sure. Because (…) even if you didn't have accurate information about the topic that's going to be discussed – like if they say it's construction and you make a glossary of all the terms like beams, slabs, and then they talk about economics, like the financial aspects of construction – even in that case it works better when you have prepared.* (P1:60-66)

When asked to specify whether they prefer having their glossary in digital form or printed out, P1 responded with:

> *It depends, really. There's no rule of thumb. I sometimes write them by hand. Normally l will type them because it makes the search a bit easier. You don't have to keep a big heap of different papers and notes to search through*. (P1:72-74)

The lack of information about an interpreting assignment was also mentioned by P4 who said that it is hard for them to prepare for an interpreting task, no matter what mode of interpreting it is, because in court surroundings, they usually never have the possibility to know what exactly they are going to interpret, not even the case:

> *I'm never informed if it is going to be some kind of domestic battery [case] or just divorce, because I could get a letter D, so letter D would mean 'domestic.' What kind of domestic, I will never know. So, I have to go and just dive into it. (…) And I cannot even ask the attorney about the case, because that would be considered biased. Because if I will talk to the attorney prior to the case, I cannot then interpret in the trial.* (P4:37-45)

Although all of them have heard about automatic speech recognition in other contexts before, such as in dictation tools, when asked about their familiarity with CAI tools before the present experiment, P2 stated that they had very vague notions and that they "may have tried a couple of software packages and they were demo versions" (P2:84) and that they did not remember the names, but the tools had had to do with glossaries. P4 said that they have heard of CAI tools before from their colleagues, but they have never had any experience with them, neither used nor tested them.

Apart from that, none of the interpreters could name any CAI tools, let alone say that they have ever used any. In fact, when asked how much they knew about CAI tools, some of the participants seemingly had misconceptions about what the exact definition of CAI tools is, as they had confused them with videoconference platforms such as *Zoom* or *Webex*, and other features. For example, P1 was asked about how aware they were of existing CAI tools, but they answered by mentioning an instance when they interpreted at a webinar using the videoconference tool *Microsoft Teams* that had a built-in feature that also works with ASR and provides real-time captions. While there was a running transcript, automatically subtitling what the speaker was saying, similar to CAI tools with such an ASR feature, it is not a tool designed specifically for interpreters, so it cannot be considered a CAI tool. When asked the same question, P3 also mentioned working with captions when interpreting at seminars where *YouTube* videos are shown, as it likewise has an option to generate captions, both in the SL and TL. Again, this feature is also not a CAI tool designed for interpreters, so that cannot be considered an affirmative answer either. Both P1 and P3 also stated even though these automatically generated captions would sometimes be helpful, they would still be rather deceptive in some cases, not showing accurate results, so there was no possibility to completely rely on them for help and it was important to watch out not just to translate them as they were shown. After being informed about what the interviewer actually meant by the term 'CAI tools,' both participants stated that they could not name any. When asked specifically about *InterpretBank*, none of the four participants had heard about it prior to the present experiment, except for P2, who said that the name "rings a bell, but a very distant one" (P2:116).

Before mentioning working with *Teams* and its built-in closed captioning feature, when asked about their familiarity with CAI tools, P1's initial response was that they have "most certainly heard about them," but have not worked with them quite extensively, because they do not work with written translations so much that they would need them. This means that they were not only unfamiliar with the term 'CAI tools' but they also misunderstood the question,

likely due to the way the question was formulated, which was: "CAT or computer-assisted translation tools have been around for quite some time now and they are becoming more and more popular and widely used. How familiar are you with computer-assisted interpreting tools?" I:76. After correction and clarification that the question was about CAI tools, P1, like P3 and P4, said that they have never heard of any or used any before.

P3 stated that they were not aware of such tools because they are not very "tech savvy" (P3:101) and not too interested in investing their time to learn about them, as they were not sure "with the time left (…) for the simultaneous interpreter to exist, if it's worth it to get into it" (P3:105) and that they would only consider it if somebody showed them how to use such tools properly. P2 stated that they did not particularly like the software designed for collecting glossaries that they had tried out, and added:

> *Well, to tell you the truth, I'm an old fogy in this respect, so I'm not using those newfangled ideas. A bit old fashioned in this one*. (P2:90-91)

5.3.2 RI and CAI tools – views after experiment

After completing the backwards alphabet task once again after interpreting, the results of which will be discussed in Chapter 5.4, the first question that the participants were asked was whether seeing the written numbers helped them or distracted them during interpreting. P1 expressed that they found it to be helpful:

> *Yeah, in 70% of the cases it did help a lot. In some cases, I wondered why there is a dot between the 2 and the rest of the numbers denoting the year. I don't know why. So, in most of the cases, this so-called precision information was quite accurate, and it was helpful. It was almost like having a presentation in front of you, in a very important sense. Not the bullet points, but yeah, it was helpful.* (P1:195-199)

Interestingly, P1 even added that apart from being very helpful in most of the cases, it was like having "someone sitting next to you and they're writing down some numbers. Like a good colleague. (…) Someone who doesn't have to drink coffee, who doesn't have to have a break, something you need, you always need to have some fresh air. This software is really helpful" (P1:210). P2 found the software to be useful as well, giving the following response:

> *Actually, it was helpful. Although it would have been more helpful if I had had some access to this program [beforehand]. Because, I mean, the interface screen, I saw it for the first time, but I was able to concentrate on the figures, they were in red. But there were some other numbers on the*

> *margins, so that might have been a distraction. (…) And with more practice and more familiarity, I think it could be a very useful tool.* (P2:186-193)

Likewise, P4 claimed they had a positive experience with the software, similar to that of P1 and P2, stating the following:

> *It definitely helped. Because at that moment, I was able to concentrate more on the whole content and to try to express myself freely. Because, initially, of course, I was very tired and I probably was making some mistakes, and the level of translation wasn't as smooth as I would like to it to be. But overall, I would like immediately to say that was very helpful, very helpful.* (P4:140-144)

However, unlike the other participants, P3 had a different experience and opinion about the software, claiming that they found it to be more of a distraction, referring to it as "disturbing" (P3:209). P3 stated that they felt that the numbers would disappear too early and not provide much help:

> *I can tell you straight away, it was actually disturbing in a way. Because very often, by the time I get to [a number], when I'm lagging a couple of seconds, it's already gone. And you think it would help you, and then suddenly it's gone, and you are actually left hanging. And had I written it down for myself, I would have it in front of me.* (P3:209-212)

When asked whether they found the transcription to be reliable, P3 stated that they did, but if they had written down those figures themselves as they usually do, they would rely on those more and not look at the ones appearing on the screen, as they were sometimes giving them "false hope, in a way" (P3:331). When asked the same question, P2 said the following:

> *I think most of the times, the system accurately captured the numbers. There was one, I think, at least one mistake. (…) I think it was 1.5 and it showed like 0.45 or something. At least it clashed with what I heard, but on the whole, it was very useful.* (P2:189-192)

P4 agreed that the transcription of the figures was reliable and added that the numbers were "quite clearly represented on the screen, and it helped" (P4:176). P1 also found that the transcription was mostly reliable:

> *Again, in most cases yes. It was quite prompt, it was quite handy in almost all the cases. I think it didn't get all the years right, if I noticed it correctly. But it would be worth delving into it.* (P1:204-206)

This shows that despite being aware of the system's limitations, some interpreters still continued to consult it for help with numbers. Both P1 and P2 indicated that when the years

were uttered in the source speech, such as fiscal 2021 or 2022, the numbers denoting the year appeared in a slightly strange way, with a dot between the first and second digit, which confused and surprised them at first, but at some point, they got "used to it" (P1:323) or "got the hang of it" (P2:200) and it did not throw them off after that. However, P3 used a different approach, adding that if there were short or "simple figures like 1.5" that they believed they did not require assistance for, they would not look at the ASR output at all and completely ignore the number presented by the tool.

When asked to specify how many numbers they used that popped up on the screen, P1 said they used most of them, "about 70%" (P1:225). P2 said that they "lost some of them" (P2:211), but that they got at least 50%. They stated that they at least took a glance at all the numbers, but sometimes they did not have time to register or say them. P3 said that even though they "used a lot of them" (P3:240), sometimes they noticed that while they were thinking how to translate a certain difficult term and "the number was on the screen, by the time [they] translated it, it had already disappeared" (P3:241), so they immediately went further "without keeping them in mind" (P3:272). It was not until towards the end of the interview that P3 realized that the numbers did not actually disappear, but only vertically went down, and that they "didn't figure it out" in time (P3:405). Thus, just like in the experiment conducted by Pisani and Fantinuoli (2021:193), the participants in this study also stated that they could have made better use of the software if they had had the opportunity to practice beforehand.

P4 said that they used "practically all" (P4:190) of the numbers presented by the ASR engine, mostly thanks to the decent speech rate of the speaker:

> *Because with your speed of speech, it was very helpful. I had a possibility to control what [I saw] on the screen. If the speaker would use another speed, something like 180 words per minute, probably I wouldn't be able to do this and then I would start to panic. But with your pace, it was very helpful.* (P4:190-193)

When asked if they could remember any particular numbers that they used or that the ASR helped them with, P4 noted that because they were in the same pace as the source speech and it was quite "simultaneous in all senses" (P4:221), they could not remember or name any in particular.

The participants were also asked to specify whether they enjoyed using the tool. P1 responded that they certainly did and that they believe that it "has a lot of future" (P1:287). P2 responded that "it was OK" and "a bit new and unfamiliar, but useful, definitely" (P2:301).

After a short pause, P3 said they "might grow into it" (P3:354). P4 stated that they enjoyed using the software as well.

Upon being asked whether they feel that the software has potential, the participants stated that it "definitely" (P3:350) and "most certainly" (P1:220) did and that it already "yielded some important results" (P1:220). However, P3 added that usually, this type of information that the ASR provides is already on the slide of a presentation, including all the relevant pieces and terms that they were missing from the software in this case, such as "return on investment," and the ASR tool would actually be diverting their attention away from the slides, because they believe that "slides are the most important thing" (P3:392) and even "more helpful than this tool" (P3:396). But, if there are no slides or other relevant documents with such information, this tool might "a potential" help (P3:426). P4 stated that in general, if they were in the booth and they had this tool and were able to see the figures, then it would be "incomparable" (P4:213), in a good way.

When asked if they see themselves experimenting with CAI software in the future, whether with or without an ASR feature, P1 said that they "most certainly, 100 percent" (P1:330) would. P2 delivered quite a different response, stating that they were generally not interested in such tools, and only if the market picks up and there is some recovery and "we finally overcome Corona and maybe some live events come back" or they get more RSI work, then maybe they "would be using these tools more actively" (P2:369), but for now, "the outlook is not very good, not very promising" for them and that they may "get out of the market completely or retire" (P2:370). They added that they were "of the older persuasion" and "not one of those digital natives" (P2:345), but not as bad as some other colleagues they know that "are challenged by the simple console" and cannot even "cope with the microphone button" (P2:352).

Participant 3 specified that they "would need some persuasion" to see how "other interpreters manage it" (P3:384). P4 gave a more thorough response, indicating that even if they wanted to, in their field of work, there would probably not be any possibility to use such software in the interpreting settings they usually work in:

> *Well, unfortunately, in the foreseeable future, in the environment where I am mainly working with the courts, I cannot say that I would be able to use it. But (…) I would definitely love to. But unfortunately, there are no booths, no equipment, and so it's mainly a combination of, to put it mildly, home-made simultaneous interpreting, interrupted with consecutive. That's what it is.*

> *It's not even whispering, it's not even chuchotage, because you stand next to the judge, and you have to whisper into the ear of the defendant and then, sadly, to interrupt yourself and start to do consecutive.* (P4:313-319)

P4 also added that usually, in court settings, the judge would tell the interpreters and attendees of the trial at the beginning that they only have a limited timeframe for the particular case. However, the only possible scenario to incorporate the use of CAI tools into the workflow is if there is a planned schedule that is followed in the trial:

> *Because if it is a trial and they try to follow some kind of procedure – there is an opening statement, there's a closing statement, there are testimonials – then there is some possibility to use some kind of tool, for consecutive at least. Because testimonials are always [consecutive]. It's the law, by the way, that you do it only in consecutive. But if it is not a proper trial, if there's just conference or a preliminary hearing, then forget about any tools, you cannot do that.* (P4:326-331)

The participants were asked what their ideal ergonomic setup would be when interpreting remotely with such a tool. They were asked to specify how many screens they would use, and if they had two or even more screens, how they would make sure that they are listening and looking at the speaker and that they do not miss any important information. P1 responded with the following:

> *If it's ideal… My screen is not big, it's 13 inches, so I would like to have a presentation, if there's any, on the screen, and (…) the speaker, like it's almost always the case in Zoom, given a proper setup. In my ideal world, I would like this software to run on my cell phone in a separate app, if possible, which will stand pretty close to my main screen, the screen of the laptop. So that I wouldn't have to turn my head around, it should be right there. Somewhere at hand. (…) It's always a matter of seconds, or milli-milli-milliseconds.* (P1:295-304)

P3, on the other hand, assured that they would only use one screen, because they "don't have time to go between two screens" (P3:374). P2 gave the following response:

> *Here I'm a bit torn between. Because, on the one hand. it would be good, probably, to have two screens: one screen for the main interface, the speaker, and another screen for tools like that or for documents. But, on the other hand, in the heat of the battle, additional screens or additional buttons are a challenge, for me at least.* (P2:330-333)

P4 proposed that their ideal setup would involve having the speaker shown on one screen and the figures "in a tiny window somewhere in the corner, left corner or upper corner, depending on the setting" (P4:276). They added that if they were interpreting in a booth and they could

see the speaker through the glass window of the booth, using such a tool and seeing the figures simultaneously appear somewhere in the corner would be "a mass help" (P4:267). They assured, however, that they would not prefer looking "at the screen instead of the speaker" (P4:268) by any means.

In the end of the interview, the participants were asked if they wanted to add anything. P1 decided to summarize their experience in a few words by stating the following:

> *It's been exciting, first of all. It's been really a bit tough because normally you would start any event with just some words of welcome, "I'm happy that you are all here" and so on and so forth. And here we had a very short introduction, so I was not on top of my performance. So, as you understood, I only woke up after the short session. In general, it was very, very helpful, and I loved my experience with this. Although it was quite short, I could get the taste of it. That's why, if you ask me, I would be willing to pay for something like this right now. (…) The next thing I'll do when we are done with this session, I will just look up online if there are options to buy such software, because I really got very enthusiastic about it. (…) People normally have trial periods right now in our day. But if they require me to pay for this trial period, I would. (…) So, I think it tells something about my experience.* (P1:350-364)

Furthermore, some participants made suggestions for improvement of such software. Both P1 and P3 said that it would be more helpful to have not just the figures, but the terms they are referring to, and some "key concepts," as P1 put it:

> *If we just added some keywords, whether or not they will be heard correctly by the app, by the software. As I said, it can go both ways. So, it can be misleading, such as in Microsoft Teams, sometimes. It was several months ago, so maybe it's gotten better over the time. It's worth a try. But I love having some key concepts in front of my eyes, both when I'm delivering a presentation myself or hearing a speech, I will never have a whole text with all the words, I won't read it out. It is difficult to do and to listen to. I will have some bullet points and that's what the software probably will come to. (…) Or you'll have several choices. Someone would prefer having only the numbers, someone will have the dates, someone will have the numbers, the dates, and the keywords, and the names.* (P1:368-380)

As mentioned earlier, the participants were asked to rate how satisfied they were with their interpreting performance on a scale of one to five, with one being not satisfied at all and five being very satisfied. P1 rated their performance a "three" (P1:274). P2 said even though they were "not very satisfied" and ranked their performance with "maybe three or two" (P2:275), it still was not a "complete disaster" (P2:279). P3 also gave their performance a "two," claiming

that they were "not satisfied at all" (P3:319) and the interpretation was "not the best of [their] performance" (P3:247), explaining that they lost a lot of information and would thus have "lots of complaints from the listeners" (P3:325). P4 rated their performance "between three and four" (P4:283). Thus, all participants were generally not satisfied with their own interpreting performance.

5.3.3 RI and CAI tools – cognitive efforts

In order to gain some insight on the effects of CAI tools on the cognitive processes in interpreting, the experimental group was asked to specify what they were mostly looking at and most concentrated on during the interpreting session.

P1 said they were "looking at the numbers, if anywhere" (P1:238). P2 likewise said that their visual attention was mostly directed at the numbers on the screen. P3 said that they were looking at the shared screen and but did not even notice the speaker, although they were shown in a smaller frame in the top right corner, because the speaker is "not the most important thing" (P3:282). P4 said that initially, they were so much concentrated on recording that for the first few seconds they did not even look up at the screen at all. However, it is important to note at this point that the first minute of the speech only contained two numerals. Then, as soon as P4 looked up at the screen, they saw that there were numbers, and the number that was used most recently was highlighted in red, or in "orange" as P4 put it (P4:233), and that is what they were mostly looking at, "always concentrating on the top line" which they found to be "helpful, definitely" (P4:236). They were, however, more focused on looking at the speaker rather than at the ASR output. P4 also added that they did not take any notes during interpreting:

> *No, I didn't make any notes. The first note I made, it was 1.5 percent, the very first figure. When I started to make them, but then I realized that I could see it, I stopped making notes and only followed it, and that was it.* (P4:227-229)

As mentioned before, the participants were asked to rank their four efforts according to Daniel Gile's Effort Model from one to four in order of importance, with #1 being the effort most concentrated on to #4 being the effort least concentrated on. P1 responded with the following:

> *I'll try to be really honest about this. Listening was not a problem, because you're a dream of speaker, I have to tell you. I would dream of having speakers like you all the time. So, this is number 4, definitely, because it didn't take a lot of effort, listening. Then, I think, producing was the most*

> *challenging task. Then, short term memory and then reading, probably. I hope I got it right, (…) I mean true to what happened*. (P1:259-267)

Thus, P1 was mostly concentrated on producing, followed by memorizing, reading, and finally listening. Though their efforts were allocated slightly differently, P2 said that they were likewise mostly concentrated on the P effort:

> *Probably producing, because… This is not always the case, but now that I'm bit out of practice, especially in the finance and economics area, (…) I haven't done any work in this area for a long time, so it was kind of difficult to bring all those terms that I used to have at my fingertips. (…) And of course, sometimes it's difficult to apportion, to allocate the effort between listening and speaking, but normally when I have this feeling, when I know the term, it is an actively used term and it comes to me without any effort, then, of course, the listening part is also easy. Because I don't have to concentrate on dragging it out of my memory, it's already here. And I can program my sentence in the back of my head. This pre-programmed sentence is already there waiting to be produced and I can concentrate on the listening part, you know what I mean? (…) Therefore, I had to think hard, and I probably wasn't listening very well and that just made me feel I expended more effort also on the listening part. But producing was more difficult.* (P2:236-259)

To sum up, P2 was mostly concentrated on producing, followed by listening. Since they had visual support, they felt they "did not have to remember" or "to memorize too much," because they had these numbers "ready in front of them" (P2:269), which means more effort was used on reading than on short-term memory. P3 said that because they did not use the ASR transcription for some numbers, they were mostly focused on listening to what was uttered by the speaker, followed by producing the interpretation output and reading the numbers. Lastly, P3 reported that they were least concentrated on memorizing, because they would immediately move on if they did not hear a number correctly or if it was not accurately presented by the ASR engine.

P4 noted that because they were trying to additionally listen to themselves in order to avoid any "slip of the tongue" (P4:252) or using expressions that were not ideal and, if necessary, to correct themselves in the continuation of their speech, they were highly concentrated on listening and producing. As far as memorizing is concerned, they explained that it "goes as a combining factor for all this" (P4:259), and that they were also focusing on watching the speaker, as they consider it to be "essential" (P4:257). Thus, P4 confirmed that they were most concentrated on listening, then producing, followed by memorizing, and lastly reading.

Table 9: Allocation of efforts as reported by participants

Participants	Allocation of efforts (most to least concentrated on)
P1	P M R L
P2	P L R M
P3	L P R M
P4	L P M R

Chapter 3.2.1 presented several tactics for dealing with problem triggers or other challenges and difficulties that arise during simultaneous interpreting, characterized by Gile (2009:201) as 'comprehension tactics,' 'preventive tactics,' and 'reformulation tactics.' The participants were asked to specify how they coped with any difficulties or issues, for example, if they misheard a word or a number, if any concrete strategies or tactics were applied. P1 said the following:

> *Normally we generalize things. So, if you have misheard something or if you feel you are not sure that you will really reproduce it the way it was said, the only way is to stay on the general side of things, so make it more generic. Probably the safest strategy.* (P1:280-282)

Thus, P1 mostly used what Gile (2009:201) refers to as 'reformulation tactics,' replacing words or speech segments with a superordinate term or a more general statement. Because they consulted the ASR output for help and used about "70%" (P1:225) of the numbers presented by it, meaning they referred to it when dealing with issues related to numbers, this can be considered a comprehension tactic, if the ASR tool in this case can be understood as a replacement for 'boothmate.' Upon analysis of P1's interpretation, it became clear that another comprehension tactic that they used was to stall and delay their response up to a few seconds, seemingly in order to win some more time for thought, causing them to utter the rest of the sentence or speech segment in the TL in an increased speech rate. There were about 10 instances in the target speech where a longer pause was followed by text that was produced at an increased speech rate.

P2 insisted that they have always been bad with numbers and proper names and that being able to interpret them is a "special skill" (P2:306) that it could be trained, and the earlier, the better. They added that they did not get exposure to such exercises with numbers until they were already an active interpreter, and "in the early days, it was a different set up" (P2:316), meaning they were not many technical conferences, but "a lot of blah blah blah and numbers didn't matter" (P2:317). When dealing with numbers and other problem triggers in this experiment, P2 said they used generalization, in addition to self-correction and anticipation:

> *I may have used generalization a couple of times, but I can't give you a specific example. And I also realized a couple of times that I made a mistake, so I had to correct myself. Either it was some grammatical mistake, like cases, declinations, or once, for instance, I thought I heard a term and but then I thought better of it. I thought I heard the Sharpe ratio, but probably it wasn't there. Because it was a financial term. (…) For some reason, I heard this Sharpe ratio, interpreted it, and then I realized, I had to correct myself probably again. But other than that, I don't know if I used any special tricks or whatever.* (P2:285-296)

Thus, P2 mostly used reformulation tactics, in addition to using their knowledge of the language, the subject, and the situation, i.e., their extra-linguistic knowledge, to better understand what was said the in the source speech, which is a comprehension tactic. Moreover, when dealing with difficult economic and financial terms, P2 also mentioned that they used the researcher and a practitioner Ghelly Chernov's theory of probability prediction, according to which an interpreter's redundancy-based anticipation of sound patterns, grammatical structures, semantic structures, and message sense is the main psychological mechanism behind simultaneous interpreting (Poechhacker 2016:93).

P3 said that if they were faced with any challenges, such as if they missed a certain number or pieces of information, their general strategy was to simply "omit it" (P3:336). If a figure was followed by an alternative one referring to the same unit, such as when growth of income was presented in both euros and a percentage, P3 stated they would only interpret the latter, which in this case is how many percent the growth amounted to. This can be considered both a preventative tactic and a reformulation tactic, as the order of elements in the enumeration in the ST was reformulated, thus freeing the short-term memory from the previous information, and then moving on to the other elements. Nevertheless, the tactic used here is also a reformulation tactic, as P3 decided to deliberately not render the information in the TL and omit it completely, either because it had disappeared from short-term memory, as they mentioned they were not much focused on memorizing, or because it appeared irrelevant or less valuable than other upcoming information that would otherwise have been lost.

P4 expressed that if they understood a term but did not quickly have available the appropriate equivalent in the TL, they would explain it rather than translate it, which is considered a reformulation tactic. Furthermore, they stated they would rephrase any sentences or expressions that they were not satisfied with:

> *I can't say that there were tactics, but there was immediately a flashlight in my mind that I didn't use a very smooth expression [that would] probably be*

> *awkward, or there was a slip of the tongue, then I will at least try to repeat this in another sentence in order to clarify it somehow. But in the very short piece, there were not so much space for me to use any other tools, than just simply to try to rephrase it later.* (P4:289-293)

Although in the present experiment, the interpreters only had access to numbers and not the entire speech, some participants explicitly noted that as a matter of fact, it was very beneficial. P1, for instance, prefers only having numbers presented to them rather than the transcript of the whole speech, as they would otherwise find it rather distracting:

> *I would definitely appreciate those numbers being shown on the screen as it was the case right now. And that's most certain. But unlike many other colleagues, I don't like having the full text of the speech in front of me. We just discussed it yesterday with the colleagues and I insisted that it's a distraction for me. I prefer to maintain contact with the speaker and having a text, even if they're reading it out, I have to adapt to their pace of reading, which is always more of a strain rather a benefit for me. Although people differ.* (P1:319-324

P4, who is of the same opinion, added that their concern was that if they had looked at the screen and seen the whole text, it would have most definitely been a "huge distraction," but because there were only figures, they were able to "adapt" to the system quite quickly (P4:168). They also mentioned that it is always more important for them to be able to look at the speaker rather than the text:

> *I can tell you, I tried on several occasions before this court job to use a computer when there were speeches of the presenters provided in advance. And usually, I found it very distracting. Because for me, just to listen and to look into the eyes of the speaker is essential. It's more important for me to look at the speaker than to read his speech. I don't know if it's my personal feature or if it happens very often, but for me, that's it.* (P4:254-258)

Even though some interpreters may feel that the text is a hindrance, others find that their performance is better when they are provided with a written text of the speech, as was the case in the experiment conducted by Cammoun et al. (2009:10). Participant 3, for instance, mentioned that in RI, they enjoy having the transcript of the whole text presented as captions, provided that the ST is "organized and structured in good language" (P3:81). However, if, for instance, the speech to be interpreted is an interview between two or more parties and the participants talk over each other and interrupt each other, then "you can forget about it" (P3:96), as the text would be rather useless.

5.4 Fatigue/alertness exercise

The purpose of this exercise was to try and compare the level of alertness or fatigue among interpreters. To create level playing field, the participants were asked to recite the English alphabet backwards, even though they were all Russian native speakers.

P4 had some difficulties performing the task, stating that it was too hard and too confusing for them to say it backwards and that they were simply not able to do it. For this reason, they were first offered to say the Russian alphabet backwards. However, this did not seem to be any easier, which is why they were then finally given permission to say the English alphabet in the right order instead of backwards. All participants were somewhat nervous to perform the task at first, so they were assured that they will not be graded for their performance or corrected if they made any mistakes. Results of this exercise showed that all the interpreters, apart from P4, were able to recite the alphabet backwards approximately 10 seconds faster after the interpreting session than before.

Table 10: Alphabet exercise

Participant	Time before interpreting (in seconds + milliseconds)	Time after interpreting (in seconds + milliseconds)	Number of mistakes & stumbles before	Number of mistakes & stumbles after
1	29,28	20,94	0	0
2	33,42	25,77	2	4
3	35,05	26,26	4	2
4	11,99	17,05	1	4

Data collected from this analysis indicated that despite interpreting a number-dense speech remotely and with the help of an ASR tool, 3 out of 4 participants were more alert and were able to perform a task quicker after the interpreting session. One participant even added that they "woke up" (P1:190) when reciting the alphabet backwards for the second time. P1 and P2 added that they felt more confident in reciting the alphabet afterwards due to having had some practice beforehand. After expressing that they were in a "no-working mood right now" (P3:159) before the interpreting task, P3 mentioned that they were more "alert" (P3:226) afterwards. It should be noted, however, that mistakes were still made and the recital of Participants 2, 3, and 4 was not flawless, containing a few pauses, restarts and self-corrections.

6. Discussion

In this study, several conclusions could be drawn. First, as far as experiences with RI are concerned, it must be noted that restrictions introduced by COVID-19 forced many clients to adapt to a technology-intensive operational model, causing a rapid growth in the conduction of online meetings (Kudyba 2020:284). Due to its inevitability in the past couple of years, RI has been the only option of providing interpreting services, which is why "it sort of loses sense" (P3:27), as expressed by one of the participants in this study, and one can argue that it has now become mainstream. Thus, interpreters may have gotten more used to it in comparison to the experiments conducted more than a decade ago and have indeed become more accepting of it.

Nevertheless, when the probands were asked about the negative and positive aspects of RI, all of them, except for P1, could only name negative ones. Similar to results obtained from various empirical tests with remote conference interpreting as mentioned in Chapter 2.1.1, most of the problems were associated with and brought about by technical issues. These include instances when interpreters had difficulties understanding what the speaker was saying due to their lack of microphone discipline and overall poor sound quality. Another drawback mentioned by one of the participants was the lack of physically having a colleague in the booth, which is a key element in SI. Furthermore, several respondents stated that preparing for an RSI assignment is not any easier or much different than preparing for a live interpretation due to inaccurate information, or lack thereof, about the event from the client. This response adds on the findings of the survey conducted by Corpas Pastor and Fern (2016:25), in which one of the most frequent answers was that interpreters very often do not receive sufficient information about the assignment from the client beforehand, which hinders them from adequately preparing for a task.

As far as the professional interpreters' familiarity and awareness about the existence of CAI tools is concerned, one major conclusion that could be drawn is that they were initially either generally unaware of such options or had a false idea of what CAI is. Furthermore, several researchers have mentioned that professional interpreters have been reluctant to use CAI tools, and interestingly, this idea was corroborated in some interviews of this study as well. Particularly, even participants who have heard of such tools used such terms as 'not very tech savvy,' 'old fogy,' 'of the older persuasion' and 'not one of those digital natives' when describing their attitude towards technology in interpreting. Some even expressed that in their age, when they are 'not looking for any kind of progress' in their career, they only refer to

using ‘whatever you have, so this is it.’ This may, in fact, underline Pym’s idea that professional interpreters that gained their mystique with an old technology are rather prone to and not enthusiastic about using new technology due to the general fear of technologically induced changes and to the desire to defend “old accrued power” (Pym 2011:4).

After gaining some insight into the functionalities of such software with the help of the present experiment, most participants stated that although they found it to be helpful, they most likely will not be using such tools in the future. However, there was an exception: one participant was very impressed by the software, stating that their experience with it made them very enthusiastic about the topic and that they would be willing to pay for such tools. They also mentioned that having this particular software at hand during SI was like having a good boothmate. Since the cooperation between boothmates, such writing down numbers and terms, is generally considered to be very important among interpreters and – when silent and discrete – not considered a source of distraction, this answer says quite a lot about this participant’s experience with the software. The positive experience of three out of the four participants in this experiment also indicates that the tool could be considered helpful.

It was hypothesized that if the ASR engine performed well, the tool would be perceived as helpful, and if it performed poorly, it would be perceived as a distraction. The tool was perceived as helpful by three out of four participants, whereas the one participant who found it distracting claimed that this was due to the numbers disappearing too early. The average ASR number precision rate in this experiment was 92%, and the average recognized number error rate was 11%, i.e., below 20% as was proposed in the hypothesis. These figures show that the ASR output is reliable in most of the cases, and as was also proposed in the hypothesis, the most frequent error type was deletion. Despite being used in a remote setting where the numbers were presented to the interpreters by means of the screensharing feature, as well as being in its trial version, the software worked faultlessly and did not report any issues, nor did it present any lags or latency problems that may have been related to the remoteness. Because the interpreters were muted in *Zoom*, it was able to catch everything the speaker was feeding into their microphone, even if there were mild background noises. This shows that the software is still functional and can be used even when another program, in this case *Zoom,* is running at the same time. It remains to be seen, however, how the software performs if the interpreters leave their microphones on and feed another channel in *Zoom*, which can be done by activating the built-in interpreting feature, in particular, whether or not that would cause the system to lag, report issues or crash altogether.

Results of the study conducted by Cammoun et al. (2009:125) showed that most of the participants said that they used the text reference either ‘always’ or ‘sometimes.’ In the present study, most participants claimed they used ‘most of them’ or ‘a lot of them.’ Whether or not they enjoyed using the text is another topic that was not addressed by the authors. To gain more insight into the experience of professional interpreters with transcription of numbers during SI, the experimental group was asked to specify whether they enjoyed using the tool. Most participants expressed that they enjoyed using the software, except for P3 who did not give an explicit yes or no answer but responded by stating that they believed the software has potential, and this indicates that it was able to fulfill all the interpreters’ needs and demands in this particular case, which was dealing with numbers in simultaneous interpreting, though not completely, because as P3 and P1 said, it would be helpful to have key words next to the number transcription, such as the noun or unit of measurement the figure refers to. As *InterpretBank* is currently still in its experimental phase, it is said to include more features in the future that would presumably fulfill this particular need.

In addition, as mentioned in the above chapters, previous experiments with RSI have shown that if interpreters need to be able to see the speakers’ and participants’ body language and other non-verbal communication aspects and the technology implementations do not meet this need, there will be anxiety and loss of quality and productivity. Therefore, the participants were asked what their ideal ergonomic setup would be when interpreting remotely with such a tool. All respondents, except for P3 who prefers using only one screen, stated that they would have the conference with the speaker and, if applicable, PPT slides running on one main screen, and the software on another smaller screen right next to the main interface, ensuring that all content is easily visually accessible.

When interpreters were faced with difficulties, they reported that they used such strategies as generalization, anticipation, omission, and paraphrasing. All these interpreting strategies, except for anticipation, entailed an increase in lag, which was evident in the analysis of their interpreting fluency. The fact that all interpreters’ deliveries included the most prosodic features, such as hesitations, increased speech delivery rates, and self-corrections during the most number-dense part of the speech is clear indication that numbers represented the biggest challenge and had a significant effect on fluency of the delivery. While there was little to no hesitation between the figure and the unit of measurement, most difficulties seem to have occurred between the number and the financial term associated with it, such as ‘revenue,’ ‘net income,’ or ‘book-to-bill ratio.’ Most of these technical terms were generalized or omitted in

the target text (TT). Referring back to the Effort Model (Gile 2009; also see Chapter 3.1), one possible explanation for the low fluency scores during the number-dense parts of the speech could be that the efforts required for reading the numbers and coordinating this with the other efforts of the cognitive process for interpreting, such as remembering specialized terminology, would cause more hesitations and pauses in the interpreter's speech, leading to the consequence that they perform poorly in terms of fluency. Therefore, it can be concluded that the effect of the CAI tool on speech fluency is rather negative, but also quite complex: even though P1 found the tool to be very useful, they still had quite a few fluency problems and accuracy mistakes.

As far as accuracy and the interpretation of numbers is concerned, the average error rate was 29.86%. The analysis of how wrong transcriptions were interpreted by the interpreters highlighted that 6 out of the total 43 errors committed by all interpreters could be ascribed to an incorrect or inexistent transcription by the software, as shown in Chapters 5.1 and 5.2.2, which indicates that in those few cases, the interpreters were either confused by the ASR output or relying too much on the software. However, there were several cases in which the interpreter said the correct number in the TL, despite it being incorrectly presented by the ASR engine or not being displayed at all, such as the number 123% being presented as 3% or the number 'four' which did not show up in any of the experiments but was correctly interpreted by 3 out of 4 participants. Similar to the results obtained by Fantinuoli and Pisani (2020:190), this shows that in these cases they were able to use the ASR system as a supportive tool, rather than as a replacement of their skills. Although there was no comparison to how the numbers would have been interpreted without the use of an ASR tool, the fact that all participants stated that they used most of the numbers provided by the tool in their interpretations, especially for complex numbers, and the error rate was rather low leads to believe that the tool had a rather positive impact on number accuracy.

Interestingly, the results of this experiment that used an actual state-of-the-art automatic speech recognition system are very similar to the results of Desmet et al. (2018) who used a simulated transcription and manual synchronization between suggestions and original speech. Their study showed that technological support improves overall accuracy on numbers, being most helpful in reducing errors on complex numbers and decimals, the two categories that are most often interpreted incorrectly (Desmet et al. 2018:25). This was evident in the present experiment as well, which provides an answer to the question raised by Defrancq and Fantinuoli (2020:97) whether professional interpreters benefit from this software to the same

extent as students who were examined in the study Desmet et al. (2018). Thus, the hypothesis that the ASR tool would have a positive impact on accuracy but a negative impact on fluency could be confirmed.

According to the paper by Cammoun et al. (2009), paying attention to additional written material during SI has the potential to be useful, but it is also very demanding. Needless to say, the concept of SI is based on the fact that everything is happening at the same time, hence 'simultaneously,' yet the process involves a number of 'efforts' that come to play to deliver a product of interpretation. With respect to interpreters' perceptions of the cognitive constraints of dual input in such an interpreting scenario, two major conclusions may be drawn. First, this cognitive constraint is difficult to manage when the numbers generated by the ASR engine are shown in a confusing manner or when they disappear quickly. Second, when the numbers are presented accurately and in a way that is easier to understand, it helps reduce the short-term memory capacity requirements. According to the analysis of the participants' reports, the effort that benefits most from the use of ASR in RI is the short-term memory effort. Reading the numbers did not seem to have caused a cognitive overload for any of the participants, as it was the second to last task that they were concentrated on, in all cases. All participants apart from P1 reported that they were still highly focused on listening and producing the TT.

These data indicate that the use of an ASR tool has different effects on the efforts involved in RSI. For most participants, the effort that benefitted most was the short-term memory effort. This can be attributed to the fact that if numbers used in the source speech are visually presented to the interpreters during SI, it is no longer necessary for them to store these numbers in their short-term memory, thus providing earlier relief to it. The added task of reading the transcribed numbers did not seem to have risen any difficulties, as the numbers were highlighted, and the last number would always appear at the top. All participants except P1 expressed that they were highly concentrated on listening, meaning that the ASR tool did not benefit the listening effort as much as was expected in the hypothesis. Moreover, all participants reported that producing the TT also required a high level of concentration and cognitive resources. Thus, these results show that the ASR tool left more capacity for the L and P tasks. When asked to specify what they were mostly looking at during the interpreting session, all participants said that their visual attention was mainly directed at the numbers.

One conclusion drawn from the experiment conducted by Cammoun et al. was that "in simultaneous interpretation, the text is always a friend" (Cammoun et al. 2009:129). However,

the results obtained in this study put this claim into question. In the present experiment, the interpreters only had access to numbers and not the entire transcript of speech, which some participants found much more favorable. Two participants expressed that they do not like having the whole speech text in front of them, because according to their reports, large amount of information would otherwise slow down delivery and place a burden on their mental processing during interpreting.

Similar to the results obtained in Pisani and Fantinuoli's study (2021:193), all participants were generally quite dissatisfied with their own performance. The highest rating that was given was "between three and four" (P4:283) out of five. The average rating was rather low (between two and three), even though some of the participants, in fact, interpreted more numbers correctly than others and showed higher fluency rates. Nevertheless, it is important to note that an interpreting performance does not only include fluent delivery or the correct reproduction of numbers, but also many other aspects, such as their context, the discursive parts, and a good output in the TL, as stated by Pisani and Fantinuoli (2021:193). For that reason, the results of the analysis of number rendition and fluency cannot be linearly compared to the interpreters' satisfaction of their performance during the experiment, since the analyzed data examine only two of the different aspects of satisfactory interpreting performance (ibid.). It remains to be seen how satisfied they would have been with their performance if they had had the opportunity to practice with the software beforehand.

The data obtained from the experiment conducted by Moser-Mercer in 2003, comparing interpreters' fatigue levels when working remotely and on-site, showed that interpreters became tired more quickly when interpreting in remote mode than in live mode. Although there was no comparison between live and remote interpreting in this study, the results obtained here still show the opposite. Data analysis showed that all the interpreters, apart from P4, who also recited the alphabet in the right order instead of backwards, were able to recite the alphabet backwards approximately 10 seconds faster after the interpreting session than before. In addition to the assessment of their interpretations which showed that their performance improved towards the end of the speech, this is another indicator that they were more alert after interpreting than before. It should be noted, however, that mistakes, pauses, hesitations, restarts and self-corrections were still made. Nevertheless, this disproves the above-mentioned hypothesis in this study that interpreting remotely with the ASR tool would make them perform this task slower after interpreting than before.

As Ziegler and Gigliobianco (2018:120) state, the improvement of technology, hardware and software in the past years has made it possible to conduct high-quality online meetings. This, in turn, may have had a positive effect on the physiological and psychological problems related to RI. Furthermore, RI has had a major influence on the professional market in recent years, which may have led to desensitization to a certain degree and a higher level of acceptance among interpreters. Another assumption that can be made is that all participants seemed to be somewhat nervous when asked to recite the alphabet backwards for the first time, i.e., before interpreting, which may have also been the reason why some of them took more time to complete the task so as to avoid making many mistakes. However, it remains to be seen how much faster or slower the task would have been performed if no ASR support had been offered to the interpreters.

6.1 Study limitations

Like all research, this study has certain limitations. Due to the small number of participants, the data and results obtained in this study are not quantitatively significant but may serve as a feasibility test for a larger qualitative study in the future. Even if the sample size of this experiment is limited to only four participants, the results and the observations made by previous studies seem to fit within the framework proposed here. However, a broader theoretical base, more participants and more data are needed to define more precisely the factors involved in numbers processing in RSI as well as their impact on the delivery of the interpretation among professional interpreters.

The method chosen in this study to help assess the interpreters' level of fatigue or alertness was a reversed recitation approach. Although it is one of the oldest and most widely used methods in neuropsychological tests, a more thorough analysis of interpreters' stress or exhaustion levels can be conducted by using other, more sophisticated techniques, such as taking baseline stress measurements (stress hormones) before and after the experiment and comparing them, etc. Therefore, due to the methods used in this experiment, it must be noted that not a lot of light can be shed on the cognitive load of interpreters. While information obtained from the interviews was useful, cognitive load measurement consists of more than asking interpreters how they feel and observing them complete one task.

Moreover, as mentioned before, this experiment was simplified in terms of ergonomics. The results obtained here showed that the software can be used even when another program, in

this case *Zoom*, is running at the same time. However, in a more realistic RSI scenario, the microphone of the interpreter is not muted but feeds another channel on the conference platform whilst the ASR input tool would need to be connected to the microphone of the speaker to be able to pick up the outgoing speech and present numbers used in the ST. Furthermore, this study tested the trial version of the software that only transcribes numbers and is limited to 5 minutes. It would therefore be necessary to conduct further experiments with the full version of the software to establish its performance and applicability in such a scenario.

Nevertheless, some previous experiments were conducted with a mockup system, such as that of Desmet et al. (2018). These types of experiments have the clear advantage of guaranteeing better control of the variables at stake by simulating a controlled and stable transcription, both in terms of error rates and latency. However, as pointed out by Pisani and Fantinuoli (2021:182), using a real-life tool has the advantage of improving the validity of the experiment. In this context, the typical issues of ASR, such as false transcriptions or lagging and latency problems, which are not eliminated by the experimental design, make it possible to draw some conclusions regarding both the product of the interpreting process as well as the potential and limits of the state-of-the-art of ASR technology (ibid).

7. Conclusion

This paper presented a study aimed at investigating the effect of an ASR-enhanced CAI tool on the cognitive processes of RSI and on the quality of professional interpreters' renditions in terms of fluency and accuracy. In particular, the research question examined in this study was whether *InterpretBank*'s ASR feature was able to provide help and could be considered by professional interpreters as a reliable 'electronic boothmate' in a remote setting or if it proved itself to be more of a distraction. Results obtained in the empirical part of this study were able to deliver conclusive answers to this question.

One major general conclusion that can be drawn from the results of the present study is that interpreting with an ASR feature enhances interpreting performance in terms of number accuracy, as shown by the rather low error rate, but negatively influences fluency. As far as the performance of the ASR tool in a remote interpreting setting is concerned, the low latency, high precision rate, and low recognized number error rate indicate that in most cases, the ASR output was reliable and effective in providing support during the interpretation of a speech dense in numbers, and its functionality was not constrained despite being used in remote mode.

Furthermore, although the professional interpreters' perceptions of the tool were mixed, it was clear that the majority considered it be useful and the transcription reliable. One of the participants described the availability of the system by using a metaphor and comparing it to having a 'good colleague.' According to their self-reports, 3 out of 4 participants stated that they found the tool was helpful, expressing that they referred to the tool for help in most of the cases when interpreting numbers. With that, one exception was a participant who found it to be more of a distraction, referring to it as 'disturbing.' However, all participants constituted that if they had had the opportunity to practice and make themselves familiar with the tool, its interface, and functionality beforehand, they would have been able to use it more efficiently, perhaps delivering a better interpretation. Therefore, future evaluations should elucidate if distraction can be reduced with increased familiarity with the CAI tool, or different ways of presenting it, such as providing the noun that the figure refers to or any other accompanying information, such as the unit of measurement. Moreover, since this was not analyzed in the present study, further research should also investigate the rendition of information and terminology used in combination with numbers, as it is known that interpreters tend to direct so many of their attentional resources to numbers that errors also frequently occur in this context (Gile 2009).

After interpreting with the ASR tool, three out of four participants were able to perform the alphabet task faster than before, stating that they felt more alert after the short interpreting session. This was also evident after assessment of their interpretations in terms of linguistic and prosodic aspects, which showed that their performance improved towards the end of the speech. Results of the analysis of the interpretation fluency showed that overall, the participants were able to deliver an interpretation with little to no serious flaws, which are in this study considered to be such features as extremely long pauses and an excessive amount of sounds of hesitation, incomprehensible words, or unintelligible utterances. The analysis also showed that the parts of the speech that contained the most figures and specialized financial terminology seemed to have caused the most trouble for all participants. One assumption was that the main reason for the low fluency scores during the number-dense parts of the speech was because the efforts necessary to locate the needed transcription of the number as well as to remember the financial term associated with it would cause the interpreter to hesitate and make more pauses in their speech. Nevertheless, according to participants' self-reports, the effort that benefited most from the ASR tool was short-term memory, meaning that the tool left more capacity for the listening and producing tasks.

It must be borne in mind, however, that the results obtained in this study are valid only in a simulated conference situation where the interpreter does not have their ASR engine activated and cannot be generalized to other groups, so more large-scale studies in this area with more participants would be welcome to test and corroborate these findings.

In the future, with the expected increase of ASR quality, the proposed technology may be good enough to be used in more difficult settings, including such issues as mispronunciations, heavy accents, background noise or natural sound, etc. An idea for future research would therefore be to test how well an ASR-enhanced software would perform given such circumstances. Furthermore, the speech in this experiment was only five minutes long, which leads to the assumption that other circumstances would provide other results, such as leading to fatigue faster than in this experiment. Another way of testing the effect of the software on the interpreters' performance in terms of accuracy and fluency could be by comparing their performance with the use of the tool and without it.

Moreover, the results of the survey conducted by Corpas Pastor and Fern in 2016 also showed that one of the biggest problems interpreters are faced with is the lack of information about the assignment they receive from the client beforehand, which hinders them from

adequately preparing for the task. As this problem was also mentioned by the participants in this study, this could be used as a motivation to develop a certain tool or a software application that helps clients provide the interpreter with the necessary material beforehand to fix this issue. Another suggestion made by some participants in this study to enhance an ASR tool for number transcription is to enable it to additionally show small amounts of targeted information that a figure refers to. Thus, the way forward is to improve existing tools and create new ones aimed at professional interpreters to facilitate their job and provide them with the necessary information before and during the interpreting task.

Research has shown that professional interpreters are still rather reluctant to using new interpreter-specific technology tools. The respondents in this study were asked their opinion as to whether technology tools could be of use to them in the future. Most of them, even those who found the software to be helpful and that it can positively impact the quality of the interpretation, stated that they would likely not use such tools in the future, albeit because of the lack of time they are constantly engaged with and the complexity of the remote interpreting process itself. This corroborates findings of previous research. Nevertheless, with the parallel increase of online meetings and, thus, remote interpreting due to recent developments linked to COVID-19, if CAI software continue to deliver more satisfying results, interpreters may need to adapt to these new trends and develop different strategies to approach and solve any potential technology-related difficulties in interpreting.

Computer-assisted translation tools are increasingly becoming better, and the quality of machine translation continues to improve, facilitating the job of translators, which, as mentioned before, will most likely soon result in translators spending most of their time post-editing (Desmet et al. 2018:13). However, as far as interpreting tools are concerned, despite their continuous development, it is safe to say that by no means will computer tools be able to completely replace interpreters, at least not in the foreseeable future. The interpreting process is just too complex and only the human mind is able to analyze, synthesize, and string words together to make a meaningful sentence in a way that no computer would be able to. As one of the respondents in the survey conducted by Corpas Pastor and Fern said, “sometimes our minds work faster than a computer” (Corpas Pastor and Fern 2016:35). Besides, the technological transformation will not equally affect all modes and settings of interpreting, as it is not a monolithic profession, as Fantinuoli (2019:9) puts it, and certain forms of interpreting will most likely continue to be performed without technological support, such as confidential political meetings interpreted in consecutive mode. Therefore, there is no need to fear that this

transformation will be a threat to the profession, and a more meaningful approach would be to see it as an opportunity to modernize it and benefit from it in order to ultimately lighten and facilitate interpreters' work.

As mentioned before, many researchers believe that the future of interpreting lies in technology tools. The purpose of this paper was to contribute to raising awareness about the usage of CAI tools and help better understand the way they influence the interpreting process. The results of the present study provide fairly unambiguous answers to the research question, but also raise new questions. There is still so much that needs to be examined and discussed that would help better understand the influence technology has on the future of interpreting, as well as to ensure that CAI is incorporated not only into interpreting teaching but also into professional practice in the future, and researchers propose that the interpreter community start talking about this now. This will help start a conversation and address more research questions in this fairly recent and complex field of interpreting studies.

Bibliography

AIIC Taskforce on Distance Interpreting. 2020. AIIC Guidelines for Distance Interpreting. https://aiic.org/document/4418/AIICGuidelinesforDistanceInterpreting(Version1.0)-ENG.pdf (17.11.2021)

Ali, Ahmed and Renals, Steven. 2018. Word Error Rate Estimation for Speech Recognition: e-WER. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Short Papers)*. Melbourne, pages 20–24.

Augmented Interpreter – CAI v 1.2. 2022. https://interpretbank.com/app/asr-num/index (20.01.22)

Berber-Irabien, Diana-Cristina. 2010. Information and Communication Technologies in Conference Interpreting. PhD thesis, Tarragona, Universitat Rovira i Virgili.

Böhm, Julia, Eberhardt, Angelika and Luppold, Stefan. 2017. Simultandolmetschen: Erfolgsfaktor für Internationale Events. Springer Fachmedien Wiesbaden GmbH, Wiesbaden.

Braun, Susanne and Clarici, Andrea. 1996. Inaccuracy for Numerals in Simultaneous Interpretation: Neurolinguistic and Neuropsychological Perspectives. In *The Interpreters' Newsletter* (7), pages 85–102.

Cammoun, Rawdha, Davies, Catherine, Ivanov, Konstantin and Naimushin, Boris. 2009. Simultaneous Interpretation with Text. Is the Text 'Friend' or 'Foe'? Laying Foundations for a Teaching Module. Université de Genève. Master d'études avancées.

Corpas Pastor, Gloria and Fern, Lily May. 2016. A Survey of Interpreters' Needs and Practices Related to Language Technology. Technical report, University of Málaga.

Corpas Pastor, Gloria. 2017. VIP: Voice-Text Integrated System for Interpreters. In *Proceedings of the 39th Conference Translating and the Computer*. Editions Tradulex, Geneva, pages 7–10.

Corpas Pastor, Gloria. 2018. Tools for Interpreters: The Challenges that Lie Ahead. *In Current Trends in Translation Teaching and Learning E*, 5, pages 157–182.

Costa, Hernani, Corpas Pastor, Gloria and Durán Muñoz, Isabel. 2014. Technology-Assisted Interpreting. In *Multilingual* 143, 25(3), pages 27–32.

Defrancq, Bart and Fantinuoli, Claudio. 2020. Automatic Speech Recognition in the Booth: Assessment of System Performance, Interpreters' Performances and Interactions in the Context of Numbers. In *Target: International Journal of Translation Studies*. John Benjamins Publishing Company, pages 73–102.

Desmet, Bart, Vandierendonck, Mieke and Defrancq, Bart. 2018. Simultaneous Interpretation of Numbers and the Impact of Technological Support. In *Interpreting and Technology*. Language Science Press, Berlin, pages 13–27.

Errattahi, Rahhal, El Hannania, Asmaa and Ouahmane, Hassan. 2018. Automatic Speech Recognition. Errors Detection and Correction: A Review. In *Procedia Computer Science* 128, pages 32–37.

European Union Institutions. 2019. New Technologies and Artificial Intelligence in the Field of Language and Conference Services. Host paper by the European Union Institutions for the 2019 International Annual Meeting on Language Arrangements, Documentation and Publications (IAMLADP), 27 – 29 May 2019, Brussels.

Fantinuoli, Claudio. 2012. InterpretBank – Design and Implementation of a Terminology and Knowledge Management Software for Conference Interpreters, PhD diss., Universität Mainz, Germany.

Fantinuoli, Claudio. 2016. InterpretBank. Redefining Computer-assisted Interpreting Tools. In *Proceedings of the 38th Conference Translating and the Computer*. London, pages 42–52.

Fantinuoli, Claudio. 2017. Speech Recognition in the Interpreter Workstation. In *Proceedings of the 39th Conference Translating and the Computer*. Editions Tradulex, Geneva, pages 25–34.

Fantinuoli, Claudio. 2018. Interpreting and Technology: The Upcoming Technological Turn. In Fantinuoli, C. (ed.), *Interpreting and technology*. Language Science Press, Berlin, pages 1–12.

Fantinuoli, Claudio. 2018a. The Use of Comparable Corpora in Interpreting Practice and Training. In *The Interpreters' Newsletter* (23), pages 133–149.

Fantinuoli, Claudio. 2018b. Computer-Assisted Interpreting: Challenges and Future Perspectives. In *Trends in E-Tools and Resources for Translators and Interpreters*. Leiden, Brill, pages 153–174.

Fantinuoli, Claudio and Prandi, Bianca. 2018c. Teaching Information and Communication Technologies: a Proposal for the Interpreting Classroom. In *Transkom,* 11.2. Frank & Timme, Berlin, pages 162–182.

Fantinuoli, Claudio. 2019. The Technological Turn in Interpreting: The Challenges that Lie Ahead. In Baur, W. & Mayer, F. (eds.): *Übersetzen und Dolmetschen 4.0: Neue Wege im Digitalen Zeitalter. Tagungsband der 3. Internationalen Fachkonferenz des Bundesverbands der Dolmetscher und Übersetzer e. V. (BDÜ)*. BDÜ Fachverlag, Berlin, pages 334–354.

Froschauer, Ulrike and Lueger, Manfred. 2003. Das Qualitative Interview: Zur Praxis Interpretativer Analyse Sozialer Systeme (UTB 2418). Facultas Verlags- und Buchhandels AG, WUV-Universitätsverlag, Vienna.

Gacek, Michael. 2015. Softwarelösungen für DolmetscherInnen. Master thesis, University of Vienna.

Gile, Daniel. 2009. The Effort Models of Interpreting. In *Basic Concepts and Models for Interpreter and Translator Training: Revised edition*. John Benjamins Publishing Company, Amsterdam, 2nd edition, pages 157–190.

Gile, Daniel. 2021. CIRIN Bulletin No. 62. http://cirin-gile.fr/bulletins/Bulletin-62-Jul-2021.pdf (22.01.2022)

Hale, Sandra and Napier, Jemina. 2013. Research Methods in Interpreting: A Practical Resource. Bloomsbury Academic, London.

InterpretBank. 2021. https://www.interpretbank.com/ (07.07.2021)

Kalina, Sylvia. 2007. "Microphone Off" – Application of the Process Model of Interpreting to the Classroom. In *Kalbotyra,* 57(3), pages 111–121.

Kudyba, Stephan. 2020. COVID-19 and the Acceleration of Digital Transformation and the Future of Work. In *Information Systems Management*, Vol. 37, No. 4. Taylor & Francis Group, pages 284–287.

Liu, Minhua, Schallert, Diane, and Carroll, Patrick. 2004. Working Memory and Expertise in Simultaneous Interpreting. In *Interpreting* 6 (1), pages 19–42.

Mayer, Horst Otto. 2013. Interview und Schriftliche Befragung: Grundlagen und Methoden Empirischer Sozialforschung, 6th revised edition. De Gruyter, Munich.

Mayring, Philipp. 2016. Einführung in die qualitative Sozialforschung: Eine Einleitung zu qualitativem Denken, 6th revised edition. Beltz Verlag, Weinheim, Basel.

Moser-Mercer, Barbara. 2003. Remote Interpreting: Assessment of Human Factors and Performance Parameters. In *Communicate!* Summer 2003.

Mouzourakis, Panayotis. 2006. Remote Interpreting: A Technical Perspective on Recent Experiments. In *Interpreting* 8 (1), pages 45–66.

Ortiz, Luis Eduardo Schild and Cavallo, Patrizia. 2018. Computer-Assisted Interpreting Tools (CAI) and Options for Automation with Automatic Speech Recognition. In *TradTerm*, Vol. 32, pages 9–31.

Pisani, Elisabetta and Fantinuoli, Claudio. 2021. Measuring the Impact of Automatic Speech Recognition on Number Rendition in Simultaneous Interpreting. In *Empirical Studies of Translation and Interpreting.* Routledge, London, 1st edition, pages 181–197.

Poechhacker, Franz. 2016. Introducing Interpreting Studies, 2nd edition. Routledge, London/New York.

Prandi, Bianca. 2015. The Use of CAI Tools in Interpreters' Training: A Pilot Study. In *Proceedings of the 37th Conference Translating and the Computer*. London, pages 48–57.

Prandi, Bianca. 2017. Designing a Multimethod Study on the Use of CAI Tools during Simultaneous Interpreting. In *Proceedings of the 39th Conference Translating and the Computer*. Editions Tradulex, Geneva, pages 76–88.

Prandi, Bianca. 2018. An Exploratory Study on CAI Tools in Simultaneous Interpreting: Theoretical Framework and Stimulus Validation. In Fantinuoli, C. (ed.*) Interpreting and technology.* Language Science Press, Berlin, pages 28–59.

Prandi, Bianca. 2020. The Use of CAI Tools in Interpreter Training: Where Are We Now and Where Do We Do From Here? In *inTRAlinea Special Issue: Technology in Interpreter Education and Practice*.

Pym, Anthony. 2011. What Technology Does to Translating. In *Translation & Interpreting*, Vol. 3, No. 1, pages 1–9.

Rennert, Sylvi. 2010. The Impact of Fluency on the Subjective Assessment of Interpreting Quality. In *The Interpreters' Newsletter,* 15. University of Trieste, pages 101–115.

Seeber, Kilian G. 2011. Cognitive Load in Simultaneous Interpreting: Existing Theories – New Models. In *Interpreting* 13 (2), pages 176–204.

Tiselius, Elisabet. 2015. Accuracy. In Pöchhacker, F. (ed.), *Routledge Encyclopedia of Interpreting Studies*. Routledge, London, pages 3–4.

Tripepi Winteringham, Sarah. 2010. The Usefulness of ICTs in Interpreting Practice. In *The Interpreters' Newsletter,* 15. University of Trieste, pages 87–99.

Wang, Xinyu and Wang, Caiwen. 2019. Can Computer-assisted Interpreting Tools Assist Interpreting? In *Transletters. International Journal of Translation and Interpreting* 3, pages 109–139.

Will, Martin. 2015. Zur Eignung Simultanfähiger Terminologiesysteme für das Konferenzdolmetschen. In *trans-kom*, 8 (1). Frank & Timme, Berlin, pages 179–201.

Williams, Mark, LaMarche, Judith, Alexander, Ronald, Stanford, Lisa, Fielstein, Elliot and Bol, Thomas. 1996. Serial 7s and Alphabet Backwards as Brief Measures of Information Processing Speed. In *Archives of Clinical Neuropsychology*, Vol. 11, Issue 8, pages 651–659.

Ziegler, Klaus and Gigliobianco, Sebastiano. 2018. Present? Remote? Remotely Present! New Technological Approaches to Remote Simultaneous Conference Interpreting. In Fantinuoli, C. (ed.), *Interpreting and technology*. Language Science Press, Berlin, pages 119–139.

Appendix

Appendix 1: Interview guideline

Thank you very much for your willingness to participate in my study! I greatly appreciate your time and effort. Before we begin with the actual interpreting session, I would like to ask you a few general questions about your personal/professional experience.

First part of interview:

1) How long have you been a professional interpreter for?
2) Have you ever interpreted in remote mode?
   a. If so, how often?
   b. Have you interpreted in remote mode prior to the pandemic?
   c. And how has your experience with RI been so far – any negative or positive aspects you can think of in particular?
   d. How do you usually prepare for a remote interpreting assignment? Any specific guidelines you follow?
3) How do you usually prepare for a live interpretation?
4) CAT tools are popular and widely used and have been around for a few years now, but how familiar are you with computer-assisted interpreting tools? Can you specify any?
5) Have you ever worked with any computer-assisted interpreting tools before?
   a. If yes, what specific tools?
   b. What was your experience like?
   c. Have you ever used a CAI tool in a remote interpreting setting?
   d. Have you ever worked with an Automatic Speech Recognition feature in interpreting before?
6) Have you ever heard of InterpretBank?

Part of my study is to test how interpreting with an ASR tool affects the overall alertness or fatigue level of interpreters. Therefore, as an attempt to comparatively assess your perceived level of fatigue before and after interpreting with the ASR tool, I will ask you to perform a small task, once before and once after the interpreting session. Please say the alphabet backwards while I time you.

Second part of interview:

1) Do you feel that seeing the written text helped you or did you find it to be more of a distraction? Please give a thorough explanation.

2) Did you find the transcription of numbers reliable?

3) Did you use all/some/many of the numbers that popped up?

   a. If so, can you remember any in particular?

   b. If not, why not?

4) What were you looking at most of the time – the speaker, the ASR output, something else?

5) Needless to say, the concept of SI is based on the fact that everything is happening at the same time, hence "simultaneously," yet the process involves a number of so-called 'efforts' that come to play to deliver a product of interpretation, consisting of the Listening and Analysis Effort, Short-term memory Effort and Speech production Effort. According to Daniel Gile's Effort Model, the segments of speech are first heard and analyzed, then stored in memory for a short time, and finally reproduced in the target language. If interpreters are given access to the speaker's text, or at least parts of it, a new Reading Effort is added. What part do you feel you were most concentrated on – listening to the source speech, producing the text (i.e., interpreting), reading the numbers, or trying to remember the information? Please rank each of the efforts in order of importance with #1 being the effort most concentrated on to #4 being the effort least concentrated on.

6) On a scale of 1-5, one being 'not satisfied at all' and 5 being 'very satisfied,' how satisfied are you with your interpretation?

7) How did you deal with any difficulties, e.g., if you misheard a certain word or figure, did you apply any specific strategies?

8) Even if it was just for a short time, did you enjoy using the ASR tool?

9) What would your ideal ergonomic set-up be when interpreting remotely with such a tool – can everything fit on one screen, or would you need two or more screens?

   a. If two or more, how will you make sure you do not miss important parts of the meeting (e.g., will you still focus on the speakers)?

10) Are you generally interested in CAI tools and do you think you will be using CAI tools, with or without an ASR feature, in the future?

11) Please feel free to add anything.

Thank you very much for your time and effort that make this research study possible. Please make sure you send me the audio file with your interpretation.

Appendix 2: Transcript of the source speech

Source: https://assets.new.siemens.com/siemens/assets/api/uuid:e95e9fb1-6801-4817-8105-8e05e04aa145/Presserede-Q42021-EN.pdf

Ladies and gentlemen,

We're very pleased that you've taken time to be here, and we'd like to welcome you – both in person and online – to our Annual Press Conference.

One year ago, Siemens started a new chapter. As a focused technology company, we support our customers in successfully shaping their sustainable digital transformation. For this purpose, we focus on highly attractive growth markets – industry, infrastructure, mobility and healthcare.

As we speak, world leaders are meeting in Glasgow to discuss how we can achieve the targets of the Paris Agreement, which concern the future of humanity. We must succeed in ensuring that the world will still be livable in the decades to come.

To achieve this goal, we need to do everything we can to limit the temperature increase to 1.5 degrees Celsius. However, we'll only succeed if we implement all the technologies at our disposal now and with full force and if we bring new technologies to the market quickly.

Siemens is successfully contributing to this effort with a clear strategy and the ambition to further accelerate progress. It's with great pleasure that I can tell you today that we delivered on all our promises in fiscal 2021.

2021 was – in every sense – an outstanding fiscal year, in which we got off to a very successful start as a focused technology company. We achieved or – in some cases – even exceeded our guidance, while, as you know, raising our outlook four times over the course of the year 2021.

Comparable revenue grew an impressive 11.5 percent.

Orders from customers were up 21 percent year-over-year on a comparable basis. Here, we saw double-digit growth across all our businesses.

The book-to-bill ratio was 1.15 – a figure which bodes very well for the start of fiscal 2022.

The sharp increase in net income to 6.7 billion euros is a clear indication of how much our performance has improved.

I’m even more pleased with our historic, record-setting free cash flow of 8.2 billion euros. This is an all-time high and corresponds to an extraordinary cash conversion rate of 123 percent.

I’d like to highlight at this point that all our businesses are again operating considerably above their pre-pandemic levels of fiscal 2019. In fiscal 2021, comparable revenue at one of our subsidiaries was up between 5 percent and 6 percent compared to two years ago.

As of September 30, 2021, total shareholder return was at a strong 45 percent.

In the course of the pandemic, we at Team Siemens have developed a fully new approach to the way we learn and further grow. More than 150,000 employees are already using our interactive digital learning platform.

Looking ahead, we will be investing around 500 million euros in the advance development of current and future growth areas for companywide technologies. We also expect our business to achieve average annual revenue growth of about 10 percent through fiscal 2025.

Let me conclude with a few words about our figures in the fourth quarter:

At 19.1 billion euros, orders were up by a strong 16 percent year-over-year on a comparable basis.

Revenue climbed 10 percent on a comparable basis to 17.4 billion euros. We recorded growth across all our businesses.

Revenue rose across all regions: Germany was up 17 percent and the U.S. 9 percent. In China, revenue growth totaled 12 percent.

Earnings per share were €1.45.

We owe our success to our over 300,000 colleagues around the world. I’m seeing huge support for our strategy, our transformed culture and our new way of working. But also – and especially – for our focus on what makes Siemens an exceptional company: technology. Technology with purpose.

The commitment and the passion of people to support our customers is incredible. Together, we’ve shown what’s possible when Siemens is at its best.

Thank you.

Appendix 3: Transcript of the interpretations

Transcription notes:

Features	Marked in text as
pauses in speech	dots … (one dot per second)
unintelligible words	dots in parentheses (…) (one dot per second)
non-verbal utterances and hesitations	Cyrillic letter э
half-sentences, false starts, interruptions	dash –
increased speech rate	underlined text

Interpretation of Participant 1

Уважаемые дамы и господа, мы очень рады э что вы посвятили нам сегодня время. Мы рады приветствовать вас э … на нашем, в нашей аудитории, также онлайн. И мы компания, которая фокусируется на технологии. Мы .. также поддерживаем устойчиво цифровую трансформацию. И для этой цели мы .. фокусируемся на очень быстро растущих рынках э таких как э … здравоохранение и других. Мы встречаемся в Глазго для того, чтобы посмотреть, как мы соответствуем целям Парижского соглашения, которые связаны с опытом будущего. И мы надеемся, что э эта цель будет достижима в течение нескольких лет. Э мы должны ограничить э рост температуры на полтора градуса Цельсия, и мы сможем это сделать только в том случае, если мы внедрим все технологии, которые находятся в нашем распоряжении. И если мы сможем добавить к этому другие технологии, вывести их на рынок достаточно быстро. Сименс успешно … э участвует в этом процессе с помощью новых технологий и .. для меня большое удовольствие рассказать вам сегодня о наших возможностях э для . э ... достижения этой цели. Мы в налоговом году 2021 мы начинаем работать над этим, мы достигаем и иногда .. э даже .. повышаем наш прогноз в 4 раза на следующий год. Что касается удивительного роста в 11.5 процента, мы смогли нарастить (..) базу на 21 процента. Здесь мы видим э двузначный рост э по всем нашим подозрениям. э Наш э ... (..) составил 1 и э .. начале года что составило 1.15%, что очень хорошо. Это, в принципе э ... 6 и 7 миллиардов евро, это показывает, насколько хорошие у нас э ... показатели. Мы достигли в конце концов рекордной величины 8 и 2 миллиарда евро. Это ... соответствует достаточно хорошему э . показателю конверсии, это 123 процента. Я хотел бы подчеркнуть, что .. мы .. э работаем с превышением наших показателей до пандемического уровня. Мы .. можем э сравнить– можем сказать, что наш рост составил 5 и 6 процентов в разных филиалах в сравнении с 19 годом. И .. мы смогли достичь э 45 процентов сохранения наших старых клиентов. И мы ... 150 наших клиентов используют нашу систему удаленного электронного обучения. Мы будем также инвестировать 500 миллионов евро в .. текущий и будущий рост э наших компаний. Мы также ожидаем, что наша компания достигнет в среднем … годового роста– среднего годового роста 10 процентов, до 25 года. Давайте я расскажу вам о целях до 4 квартала. .. наш– э наша задача достичь 16 процентного роста в течение года, 10 процентов э чтобы у нас было

17 и 4 миллиарда **..** евро прибыль. Что касается Германии, то здесь будет 17 процентов, в США будет 9 процентов, в **..** Китае будет 12 процентов. Что касается **...** наших **…** что касается наших успехов, у нас 300 коллег по всему миру. Я очень ценю **э** нашу стратегию, нашу новую– нашу новую модель работы, нашу трансформацию. И вот что делает Сименс исключительной компанией, эта технология, технология **..** и приверженность **э** людей к **.** изменениям. И мы хотим показать, что возможно, когда Сименс достигает своего наилучшего **э** темпа работы. Спасибо.

Interpretation of Participant 2

Дамы и господа, мы очень рады что вы нашли время поприсутствовать здесь и хотим вас поприветствовать, очно и заочно, на нашей ежегодной конференции. Год назад Сименс открыл новую главу. Как компания, сфокусированная на технологиях, мы поддерживаем наших клиентов в том, что касается обеспечения .. цифровой трансформации. Мы сосредоточились на . рынках роста: промышленность, инфраструктура, мобильность и здравоохранение. Сейчас, всемирн– лидеры стран встречаются в Глазго и обсуждают о том, как достичь цели Парижского соглашения по климату . что касается будущего человечества. Мы должны обеспечить возможность проживания для будущих поколений в ближайшие десятилетия. Для достижения этой цели необходимо сделать всё возможное для . ограничения температуры на полтора . градуса. Но мы достигнем успеха только в том случае, если мы сможем внедрить все технологии, которые имеются в нашем распоряжении, и .. выведем их на рынок быстрым образом. Сименс успешно вносит вклад в эти усилия на основе чёткой стратегии и целей по ускорению прогресса. С удовольствием сообщаю вам, что мы выполнили все наши обещания в 21 финансовом году. Это был удивительный финансовый год, и . мы успешно начали реализацию наших технологических целей. В некоторых случаях мы даже превысили наши плановые показатели. 4 . раза .. превысили наши перспективные показатели. . Доходы возросли на 11,5%, заказы выросли более чем на 21 э процесс– процент, простите. Мы увидели рост .. двузначных величин по всем … показателям. Коэффициент составил 1– 1,15. Коэффициент э Шарпа … на 2 процента. Я даже . более рад сообщил, что .. поток свободных денежных средств достиг 2– 8.2 миллиарда. Это .. обеспечивает .. уровень конверсии 123 процента. Мы . работаем с повышением уровня по сравнению с 19 годом. В 21 году доходы в одном из достичь– дочерних предприятий выросли на 5 и 6 процента по сравнению с 2-летним периодом– с периодом 2 года тому назад. Общий доход для акционеров повысился на 45%. В ходе . пандемии, Сименс разработал совершенно новый подход к росту. Более 150000 сотрудников используют нашу цифровую платформу обучения. Мы собираемся инвестировать 500 миллионов евро в .. отрасли роста и развития технологии. Мы также ожидаем, что нам удастся обеспечить в 25 году рост на 10 процентов. Теперь по 4 кварталу цифру .... Заказы составили 19 миллиардов, что на 60– на 16% больше, чем предыдущий показатель. Мы .. зафиксировали рост по всем . отделениям бизнеса и по всем регионам. Германия 17 и США 9%, в Китае 12%. Что касается дохода на акцию, составил евро 45 центов. … У нас работают более 300000 сотрудников всему мира– по всему миру. Кроме того, .. мы сосредоточились на том, что делает нашу компанию исключительной, технология. Технология, которая развивается целенаправленным образом. Мы привержены идее поддержки наших клиентов и вместе мы можем показать на что мы способны. Спасибо за внимание.

Interpretation of Participant 3

Дамы и господа, мы очень рады, что вы нашли время присутствовать на совещании. Мы хотим вас приветствовать э как в виртуальном, так и в реальном режиме. Год назад, Сименс открыл новую главу. И мы поддержим своих клиентов э и поддержим их успех в цифровой трансформации. Мы хотим концентрироваться на э .. рынки с потенциальным э ростом здравоохранения. И мировые лидеры встречаются в Глазго чтобы как дости– обсудить как достичь– выполнить цели поставные в Париже (..) для– ради человечества. Мы надеюсь, что через э десятилетие в этом мире также можно .. хорошо жить. И для этой цели мы хотим сделать всё, чтобы э ограничить э подъём глобальной температуры до полтора градусов. Но это мы сможем если будем использовать существующие технологии. И сможем быстро довести и развернуть новые технологии на рынке. И у нас чёткая стратегия для ускорения прогресса. У нас на Сименсе. И я могу вам объявить сегодня, что мы выполнили все наши обещания на бюджетный год, 21 год. 21 год был замечательный бюджетный год, и мы– у нас было отличное начало как у техно– компании, которая ориентируются на технологии. Мы достигли и даже перевыполнили наши цели и … в течение 21 года. э … Доходы вы– э выросли, прирост доходов составил 11.5. И, что касается заказчиков, у нас на 21 процент рост. У нас двузначный рост обозначился по всем сферам бизнеса. .. Соотношение э заказов к э счетам был 1.15. Это хорошее начало для бюджетного года. У нас э повышение выручки 6,7 миллиардов долларов. Это хорошо отражает рост и успех нашего роста. И я особенно .. рад э нашем свободным денежным .. э резервам. 8.2 миллиардов евро. Это просто рекорд. И это соответствует э конверсии, или индексу конверсии 123 процента. Хочу указать на то, что все наши бизнесы работают э .. э на состояние– превышающем состоянии до пандемии. И в 21 году, наш .. э прирост э .. доходов в одном из наших дочерних предприятий э был 5 или 6 процентов по сравнению с предыдущими годами. э Доход– э возврат– дох– для акционеров был 45 процентов. В ходе пандемии, мы в– на Сименсе предприняли новый подход. 150000 сотрудников используют наши э .. учебную платформу, э –интерактивную платформу для обучения. Мы будем инвестировать 500 миллионов евро в . дальнейшую разработку э для работы над наших э .. областей с ускоренным темпом. И мы хотим достичь прирост доходов на 10 процентов э вплоть до 25 года. Хочу ещё раз упомянуть э статистику на 4 … э квартал. На 20– 16 процентов, то есть, э повышение доходов на– до 7.4 миллиардов долларов, у нас рост по всем сферам бизнеса. Рост э доходов по всем регионам. Германия была с повышением 17 процентов, США 9 процентов, Китай 12 процентов. .. Доходы на акцию за один год был 0.45 центов. Мы .. благодарны нашему успеху, благодаря 300000 коллегам по всему миру. Я вижу хорошую поддержку нашей стратегией, нашей культуры поддержки и нового подхода к работе. Особенности новому фокусу э на технологию, то, что обеспечивает успех нашей компании. И ангажированность э наших э всех в поддержку клиентов просто великолепна. Это показывает товар лицом. Спасибо

Interpretation of Participant 4

Дамы и господа, нам доставляет удовольствие за– мы благодарны вам за то, что вы присутствуете здесь сегодня, лично и на– э по виртуальной системе. Год назад, Сименс начал свою работу, и как компания, которая фокусируется на современной технологии, мы помогаем нашим клиентам э .. улучшить их э отношения с цифровыми технологиями. И поэтому у нас– мы создали достаточно привлекательный э рынок э который .. поскольку мы сейчас говорим, наши лидеры встречаются в Глазго, которые обсуждают сейчас вопрос, как можно достичь э выполн– исполнение Парижских соглашений. Мы должны добиться успехов, и для того, чтобы в предстоящих десятилетиях мы могли двигаться вперёд. И соответственно мы должны снизить э температуру, которая возросла на полтора градуса Цельсия. Но мы добьёмся этого только если мы будем использовать все современные виды технологий, которые мы используем сейчас, полностью, полномасштабно, и если мы это быстро сделаем. Сименс достаточно вносит вклад в осуществление этой стратегии и ускорение этого процесса. С большим удовольствием я могу вам сказать, что мы обеспечиваем выполнение всех наших обязательств в бюджетном году 21 века– 21 года, безусловно был выдающимся годом в этом отношении. И поскольку мы как компания фокусируем свое внимание, мы достигли или даже превзошли наши обещания в э 4 раза по сравнению с предсто– предыдущим годом. Соответствующий доход начало– возросло до 15.5%. И заказы клиентов выросли на 21 процент по сравнению с прошлым годом. Здесь мы видели э .. рост нашего .. экономического роста э из состоящих из (.). В предыдущие годы это было пол– .. 1015, что было достаточно хорошей цифрой, для начала 22 года. Резкий рост э нашего оборота достиг э 6.7 миллиона долларов– евро. И я очень доволен тем, что наш э .. оборот денежных средств вырос до 8.2 биллиардов. Это самая высокая цифра, которую мы достигли за это время. И это демонстрирует выдающиеся рост денежного оборота в 123%. Но я должен сказать, что наши компании сейчас работают гораздо выше до пандемии– панде– уровня до пандемии и сравнительный доход одного из наших подразделений вырос от 1– от 5 до 6% по сравнению с 2 годами тому назад. С 3 сентября 21 года, возврат по доходам акционеров вырос на 45%. И в русле пандемии, мы в Сименсе подготовили значительно новый подход, который (..) обеспечил нам то, что 150 наших сотрудников используют новую цифровую платформу. И в будущем мы собираемся инвестировать 500 миллионов ерo– евро на дальнейшие разработки э аспектов, которые связаны с технологическими достижениями нашей компании. Кроме того, мы э рассчитываем, что мы достигнем постоянного роста 10-процентного оборота до 2025 года. Позвольте мне сказать несколько слов о цифрах, которые мы рассчитываем на 4 квартал. От 19. э 1 миллиардов долларов, заказы выросли на 16% по сравнению с прошлым годом. Доход вырос на э 10% по сравнению с– от цифры 17.4 миллиардов. Мы продемонстрировали рост всех нашего бизнеса– э доход вырос во всех наших подразделениях в Германии на 17% и в Соединённых Штатах 9%. В Китае наш доход вырос на 12%. …. э доходность на акцию выросла на 0,45%. И мы обязаны нашим успехам, нашим 300 коллегам, которые работают по всему миру. И я ощущаю большую поддержку нашей стра– стратегии, нашей культуре, и нашем пути работы. Кроме того, и особенно, э нашего– э нашего– концентрации на том, что отличает Сименс как высоко технологическую компанию. Наши обязательства и страсть наших коллег к достижению цехов, они просто невероятны, и поэтому мы продемонстрируем всё, что может сделать Сименс как можно лучше.

Appendix 4: Transcript of the interviews

Interview with Participant 1

I: **First of all, thank you very much for your willingness to participate in this study, I really appreciate your time and effort. And before we begin with the actual interpreting session, I would like to ask you a few general questions about your personal and professional experience with interpreting.**

P1: Sure.

I: **So, how long have you been a professional interpreter for?**

P1: 15 years.

I: **15 years, okay. And have you ever interpreted in remote mode?**

P1: Quite a lot over the past 18 months, about 200 days to put it into numbers.

I: **But have you done it prior to the pandemic, or did it only start once the pandemic hit? Have you had any experience before that?**

P1: No, it started with the pandemic. A couple of months after the pandemic started.

I: **And how has your experience been so far? Any negative or positive aspects that you can think of in particular?**

P1: I love remote interpreting more than I love on-the-site interpreting, really. As for some negative aspects, you have to be very careful with your Internet connection, because sometimes it can get switched off unexpectedly. I have lots of different options, I have my mobile Internet to step in if my Wi-Fi is not working any longer. But I can remember one instance when lightning hit really hard and a couple of districts in my area were completely shut down, so I didn't even have a chance to call my customer, my client, over the phone to let them know I'm OK.

I: **Yeah, that's not good.**

P1: So there was no way, but everyone understands that. It can be an option as well. But overall, it's a very, very nice way to spend the pandemic times.

I: **Do you have any usual way of preparing for a remote assignment? Other than, of course, making sure that your Internet connection is stable, but are there any guidelines that you follow as far as your set up is concerned, regarding the microphone or technology?**

P1: From the technical side?

I: **Yes, anything that you notice in particular, making sure that everything is working?**

P1: Well, I will, of course, reboot my computer, restart my computer. Because if you just close your laptop for a couple of days, then Zoom will sometimes pause for several seconds when you're in a conference. That's what I noticed. I didn't know, I thought something was wrong with my Internet. But the good thing that helped me and some other people out there when I shared the experience with them, is just restarting the computer before every assignment. Other than that, I make sure that all of my devices, my cell phone, my tablet, everything is charged 100 percent, in case anything steps in, like a shutdown, for example, when there's no electricity. So it can happen that I will have to work on my cell phone and Zoom is quite consuming in terms of power.

I: **Yeah, absolutely. And what about live interpretation – how would you usually prepare for that? Are there any specific rules that you follow to make sure that you're ready for a live interpretation?**

P1: I always have a glossary. It doesn't matter if the interpretation live or remote, it's like a tradition for me. Even if it's a topic that I know quite well, even if people will be talking about, like, peace and friendship, I will have a glossary of those two words, just to make sure. Because it happens that even if you didn't have accurate information about the topic that's going to be discussed – like if they say it's construction and you make a glossary of all the terms like beams, slabs, and then they talk about economics, like the financial aspects of construction – even in that case it works better when you have prepared. I don't know, it's some magic that works with me 100%.

I: **And these glossaries, do you have them printed out or are they digital? Like, for example, an excel file, or what do they look like?**

P1: It depends, really. There's no rule of thumb. I sometimes write them by hand. Normally l will type them, because it makes the search a bit easier. You don't have to keep a big heap of different papers and notes to search through.

I: **CAT or computer-assisted translation tools have been around for quite some time now and they are becoming more popular and are widely used. How familiar are you with computer-assisted interpreting tools? Have you ever heard about them, have you ever worked with any of them?**

P1: In fact, I have most certainly heard about them, but I wouldn't say that I worked with them quite extensively. Because I don't work with written translations so much that I would need them, CAT tools. In terms of personal interest, I think they are becoming more and more well-developed.

I: **Are you talking about translation tools or interpreting tools?**

P1: Translation tools. As for interpreting, I have never… Actually, your email was the first instance that I ever heard about such options.

I: **OK, interesting.**

P1: Well, well, no. No, I'm not correct. There was one a customer, we worked on… what do you call the Windows conference, the webinar?

I: **Teams?**

P1: Teams, yes. I don't work with it too often, so I forget the name. It's the only customer that prefers that system, so I'm fine with that. But I had to do a chuchotage and the sound was not good, by any means or any standards. So, the text was very helpful. It was in a room, the English sound was in the room, and there was a running transcript, it was subtitled automatically. In some instances, it did help, in other instances it only prevented me from really getting the correct point.

I: **These subtitles, they were automatic subtitles from Teams, right? It wasn't a special tool just for subtitles?**

P1: No.

I: **So it was just a Teams feature.**

P1: It was a built-in feature, yeah. Sometimes they were helpful, sometimes they were deceptive. In most, like, in 70% of cases they were rather helpful, but you had to watch out not just to translate them as you see them.

I: **Other than these subtitles, any other tools that are actual interpreting tools? Have you ever worked with any of them, have you ever heard of any of them? Other than the one that I wrote you about, InterpretBank?**

P1: No, not really.

I: **So, this was the first time you've heard about [InterpretBank] as well, in my e-mail?**

P1: Exactly.

I: **Part of my study is to test how automatic speech recognition tools affect the overall interpreting performance and the level of alertness or fatigue of interpreters. I have a little task that I'm going to ask you to do, to try to compare your performance before interpreting and after interpreting with the tool. It's simple, please say the alphabet backwards and I will time you to see whether there's going to be any differences between now and afterwards.**

P1: To be frank, I knew it, but I didn't prepare anything, so I don't even have it on the paper. So, the pure experiment is more important for me than performance in this case.

I: **Of course.**

P1: So I will make mistakes

I: **That's fine.**

P1: In whichever language.

I: **We'll just do the English one.**

P1: Sure, or we could do German.

I: **Let's do English.**

P1: Alright. ZYXWVUTSRQPONMLKJIHGFEDCBA.

I: **OK, perfect.**

P1: How many mistakes?

I: **Zero mistakes. So good, very good.**

P1: Wow! We'll have some mistakes [later].

I: **That's fine, we'll see.**

P1: So now I have to get my cell phone ready?

I: **Yes, please do.**

P1: Now I'll just switch on my sound recording.

I. **Yes. So, it's on?**

P1: It is on.

I: **I'm going to have to ask you to mute yourself on Zoom, because otherwise the program will hear your voice as well and it will confuse it, thank you. Let me go ahead and share my screen with you and then I will start with the speech. Can you see my screen now?**

P1: (thumbs up)

speech

I: **Alright, you can unmute yourself now.**

P1: Wow. A bit tough.

I: **I can imagine. But before we go to the second part of the interview, I have to ask you again to say the alphabet backwards like you did just now, earlier.**

P1: Now after some training. ZYXWVUTSRQPONMLKJIHGFEDCBA.

I: **Perfect, it was actually faster than the first time.**

P1: I woke up.

I: **Probably, yeah. So, do you feel that the written text with the numbers helped you or did you find it to be more of a distraction? And explain why or why not.**

P1: Yeah, in 70% of the cases it did help a lot. In some cases, I wondered why there is a dot between the 2 and the rest of the numbers denoting the year. I don't know why. So, in most of the cases, this so-called precision information was quite accurate, and it was helpful. It was almost like having a presentation in front of you, in a very important sense. Not the bullet points, but yeah, it was helpful.

I: **The focus is on the numbers, so don't worry about things like "free cash flow" (laughs). Did you find the transcription reliable**?

P1: Again, in most cases yes. It was quite prompt, it was quite handy in almost all the cases. I think it didn't get all the years right, if I noticed it correctly. But it would be worth delving into it.

I: **Absolutely, I will examine the output as well.**

P1: Yeah, it's my first take. But in most cases, in many cases it was very helpful. It's like someone sitting next to you and they're writing down some numbers. Like a good colleague.

I: **Yeah, like a boothmate, right?**

P1: Yeah, sure. Someone who doesn't have to drink coffee, who doesn't have to have a break, something you need, you always need to have some fresh air. This software is really helpful.

I: **Has potential?**

P1: Yes, sure, most certainly. It already had yielded some important results.

I: **Did you use all of the numbers that popped up, or some, or many of them? Or were there some that you did not use purposely?**

P1: I used most of them. Again, about 70%.

I: **Are there any numbers, other than the years, that you remember that were not accurate?**

P1: I think it showed 2021 all the time, even though in some cases it was…. Yes, it was right when they were talking about 2025, it got it correctly. And again, that dot after 1st digit of the year. So you just get used to it.

I: **Probably. That's just how the software works. So the dot was kind of confusing. What do you feel like you were looking at most of the time, the speaker, the ASR output, or maybe something else?**

P1: I was looking at the numbers, if anywhere.

I: **Needless to say, the concept of simultaneous interpreting is based on the fact that everything is happening simultaneously, at the same time, which means that it's a process that involves a lot of cognitive processes that come to play to deliver a product of interpreting. There is this effort model that was concepted by, I don't know if you've ever heard of him, Daniel Gile.**

P1: I have, sure.

I: **The effort model in simultaneous interpreting consists of the listening and analysis effort, short term memory effort, production effort, but also the coordination effort to coordinate the three other efforts. If an interpreter is given access to the text, there's another reading effort. If I asked you to rank those efforts from one to four, what would you say you were most concentrated on?**

P1: Listening, short term memory, producing, and?

I: **Listening, short term memory, producing, reading. With number one being what you most concentrated on and forth being what you were least concentrated on.**

P1: Sure, I see, I see. I'll try to be really honest about this. Listening was not a problem, because you're a dream of speaker, I have to tell you. I would dream of having speakers like you all the time. So this is number 4, definitely, because it didn't take a lot of effort, listening. Then, I think, producing was the most challenging task. Then, short term memory and then reading, probably. I hope I got it right.

I: **There's no wrong or right.**

P1: Yeah, but I mean true to what happened.

I: **OK, yeah. It's interesting to find out whether reading the text makes it easier to listen by having the text in front of you. Another question is, on a scale of one to five, with one being not satisfied at all and five being very satisfied, how satisfied are you with your interpretation, with your performance?**

P1: I'd say three this time.

I: **And how did you deal with any difficulties, for example, if you misheard a word or a number that you didn't quite understand or didn't have time to understand, how did you deal with these issues, these challenges? Did you apply any specific strategies?**

P1: Normally we generalize things. So, if you have misheard something or if you feel you are not sure that you will really reproduce it the way it was said, the only way is to stay on the general side of things, so make it more generic. Probably the safest strategy.

I: **I know it was just five minutes, so not much time to spend with the tool, but did you enjoy using the tool?**

P1: Sure, I did. It has a lot of future, I think.

I: **If you were to use this tool in your work, what would your ideal ergonomic setup be when interpreting remotely? Would you make sure that you have two screens, or would you put everything on one screen? And if you had two or even more screens, how do you make sure that you are listening and looking at the speaker or that you don't miss any important information?**

P1: If it’s ideal… My screen is not big, it's 13 inches, so I would like to have a presentation, if there's any, on the screen, and I'd like to have the speaker, like it's almost always the case in Zoom, given a proper setup. In my ideal world, I would like this software to run on my cell phone in a separate app, if possible, which will stand pretty close to my main screen, the screen of the laptop. So that I wouldn't have to turn my head around, it should be right there. Somewhere at hand.

I: **Yeah, that makes sense.**

P1: It's always a matter of seconds, or milli-milli-milliseconds.

I: **Of course. Just now, I was sharing my screen, and were you able to see me and the numbers? Or was it just the numbers?**

P1: Yes, I was.

I: **Just making sure, because that's important. Now that you've had a little bit of experience with such a tool, are you generally interested in computer-assisted interpreting tools? Do you think you will be using some in the future, with such an ASR feature or without? There are also some that assist you when you manually look up something in a glossary, or some even hear what the speaker is saying and shows you the translation. Do you see yourself working with something like that? Or do you think it would be too much work, too much additional efforts?**

P1: I would definitely appreciate those numbers being shown on the screen as it was the case right now. And that's most certain. But unlike many other colleagues, I don't like having the full text of the speech in front of me. We just discussed it yesterday with the colleagues and I insisted that it's a distraction for me. I prefer to maintain contact with the speaker and having a text, even if they're reading it out, I have to adapt to their pace of reading, which is always more of a strain rather a benefit for me. Although people differ.

I: **Of course, it's different for everybody, I agree. But in general, would you say you are interested in experimenting more or seeing what they have to offer, what there is on the market?**

P1: Most certainly. 100 percent.

I: **It's an area that's still not much talked about or there haven't really been that many studies with it, because it's relatively new. It hasn't been around as long as computer-assisted translation tools, so it's exciting to see where this whole thing goes.**

P1: And how fast.

I: **Yeah, exactly, and how fast.**

P1: If we remember our cell phones, like, 20 years ago, polyphonic cell phones were like something like a wonder.

I: **Completely. Not everybody had them and now even every child has one.**

P1: Yeah, it can show colors instead of black and white, wow.

I: **You see how far technology has gotten. Is there anything that you would like to add or ask? Something that you feel is important for this experiment?**

P1: Maybe I would summarize my experience in a few words. It's been exciting, first of all. It's been really a bit tough because normally you would start any event with just some words of welcome, "I'm happy that you are all here" and so on and so forth. And here we had a very short introduction, so I was not on top of my performance. So, as you understood, I only woke up after the short session. In general, it was very, very helpful, and I loved my experience with this. Although it was quite short, I could get the taste of it. That's why, if you ask me, I would be willing to pay for something like this right now. I would say I would. If there's an app which sells, I will. So the next thing I'll do when we are done with this session, I will just look up online if there are options to buy such software, because I really got very enthusiastic about it and if there is any option. People normally have trial periods right now in our day. But if they require me to pay for this trial period, I would.

I: **OK.**

P1: So, I think it tells something about my experience.

I: **Absolutely, very much so.**

P1: I would invest in such software. If we just added some keywords, whether or not they will be heard correctly by the app, by the software. As I said, it can go both ways. So it can be misleading, such as in Microsoft Teams, sometimes. It was several months ago, so maybe it's gotten better over the time. It's worth a try. But I love having some key concepts in front of my eyes, both when I'm delivering a presentation myself or making a speech, I will never have a whole text with all the words, I won't read it out. It is difficult to do and to listen to. I will have some bullet points and that's what the software probably will come to.

I: **Yeah, maybe. I think as time goes by, it will get better and better.**

P1: Or you will have an option. You'll have several choices. Someone would prefer having only the numbers, someone will have the dates, someone will have the numbers, the dates, and the keywords, and the names.

I: **Or proper names maybe yeah.**

P1: Sure.

I: **Studies have shown that so far, there has been a reluctance among professional interpreters as far as technology in interpreting is concerned, perhaps because they are used to having their own processes of how to prepare for an assignment or how they interpret. They haven't been very willing to try out new technologies. So, it's refreshing to see that there are colleagues who are, in fact, willing to try out things. OK, I think I asked everything I wanted to ask. As I said, thank you very much for your time and your effort that make this research possible, because there hasn't been, not to my knowledge, any experiences with this particular software among professional interpreters. Maybe**

there has only been one with the developer of the software, but I believe conducting more experiments will lead to further conclusions.

P1: Thank you very, very much, the pleasure has been all mine. And good luck with your further studies, good luck with your wonderful project, which is unlike many things we do at universities. It's quite useful, it's quite helpful, for real. Not just the theory which you put on onto a shelf and forget about, you get your degree and it's "thank God I'm done with that." This will help people.

I: **Hopefully. I hope so, definitely hope so.**

P1: I'm keeping my fingers crossed for you and for your project.

I: **Thank you very much. Also, make sure you send me the audio file of your interpretation. Just an audio file to my email would be best.**

P1: Yeah, sure I will right now. Before I look up the software online.

I: **Alright. Thank you.**

P1: Thank you.

Interview with Participant 2

I: **Thank you once again very much for your willingness to participate in this study, I very appreciate your time and effort. And before we begin the actual interpreting session, I would like to ask you a few general questions about your professional experience. So the first question is, how long have you been in professional interpreter for?**

P2: Well, since the mid 1990s.

I: **Since the mid 90s, so it's been quite some time, you have a lot of experience**.

P2: Yeah, I should caveat that now I'm not sure I can count as an active professional interpreter, because since corona broke out, there's been very little work. And the last gig I did was probably in October, so I'm a bit out of practice and just wanted to caveat that for the sake of your academic rigor.

I: **OK, thank you.**

P2: Maybe it won't impact the statistical power or whatever, but still just wanted to make a note of that.

I: **Thank you. So, have you interpreted in remote mode before?**

P2: Yes, I have experience with some platforms like Interprefy, KUDO, Zoom, and Webex.

I: **Have you interpreted remotely before the pandemic started, or did it only start with the pandemic?**

P2: No, only after the pandemic started.

I: **Because there were no clients that wanted that or just because you never got the chance? Or did only become popular once the pandemic started?**

P2: Yeah, I hadn't heard about this mode before the pandemic.

I: **OK.**

P2: Actually, I was approached once before the pandemic to do a conference via a hub located somewhere in Stuttgart. It was medical. It didn't come through, but they were planning a medical conference with some live presentations, some surgeries done in Hamburg and simultaneously to be interpreted in many languages. So it was a hub, you know. Since Corona, a lot of these hubs have sprung up all over Europe and Germany, but they had existed before Corona, obviously. But I didn't get to work there.

I: **With interpreting remotely, how has your experience been so far? Are there any positive or negative aspects that come to mind in particular when you think about working in remote mode?**

P2: I would say more negative than positive. Because, well, I had expected more comfortable experiences. Like, you know, you sit in your pajamas with a cup of coffee, you don't have to

travel, to commute. But there was too much stress related to mostly technical issues, quality of the sound, and a lot of hassle from techs. I mean, those technicians, especially Interprefy, gave me a lot of grief, because they were very, very… well, I wouldn't say harassing, but they monitored the conference. I did a conference which lasted 2 weeks and there was always a technician assigned to the Russian booth, and they were monitoring my performance all the time. "You got too close to the mic," "you're too far away from the mic," "we hear some air flow," "we hear some background sounds," you know… And, of course, the clients, the customers, they didn't bother to wear a headset or use proper mics. So, I would say more stressful than…

I: **Than live?**

P2: … than I expected, yeah, than live.

I: **When you have such a remote interpreting assignment, is there any specific way of how you prepare for it? Anything that you look at in particular?**

P2: You mean that may differ from preparing for live conferences?

I: **Yeah.**

P2: Not really. I just make sure my Internet connection is OK. Well, otherwise…. no, not really. Maybe I use my screens more, because when I do it from the booth, I just have one laptop. And here I have access to a number of screens and when I have to use a lot of documents, which was the case with that conference, so I used my desktop screen to read some texts and another, an old laptop. So then there may be a difference in terms of more technical gear.

I: **So, how many screens do you have in total then?**

P2: Well now I'm using one, but normally… well, two at the most.

I: **Computer-assisted translating tools have been around for quite some time now and they're quite popular and widely used. But how familiar are you with computer-assisted interpreting tools? Have you ever heard of them? If you have, any tools in particular?**

P2: No, I have very vague notions. I think, I may have tried a couple of software packages and they were demo versions. I don't remember the names, but they had to do with glossaries.

I: **Glossaries, OK.**

P2: They were specially designed, you know, as a collection of glossaries. But I didn't like that particularly well. Well, to tell you the truth I'm an old fogy in this respect, so I'm not using those newfangled ideas. A bit old fashioned in this one.

I: **Do you remember what the software was called? Or do you just vaguely remember how it worked, but there's no name that comes to mind?**

P2: Don't remember, no.

I: **So, it was just one software that you tried out and didn’t like. My next question would have been if you've ever tried working with such a tool in a remote interpreting setting, but I'm guessing you probably haven’t.**

P2: No.

I: **You are familiar with automatic speech recognition, you probably have used it day-to-day, as a dictation software?**

P2: Yes, I've used that a couple of times as a dictation tool.

I: **But as far as interpreting is concerned, have you ever tried it in this context?**

P2: (shakes head) No.

I: **The software that I will be testing today is called InterpretBank. Have you ever heard of it prior to today?**

P2: It rings a bell, but a very distant one.

I: **Before we begin with the interpreting session, part of my study is to also test how the tool affects your level of alertness or fatigue, which is why I'm going to ask you to perform a simple task and I will time you. I will also ask you to do the same after the interpreting session, and I will compare. So just say the alphabet backwards, the English alphabet, unless you feel more comfortable saying the Russian one because it's your native language.**

P2: English is fine. Backwards?

I: **Backwards, yeah. Whenever you're ready.**

P2: I'm ready.

I: **OK.**

P2: ZYXWVUTSRNM … no… RQPO… NMLKJIHFEDCBA.

I: **Perfect. This was just a little exercise, so thank you for that. Now I'm going to be reading out a speech by a CEO that took place, I believe, in November last year, and they are talking about the fiscal year 2021, a financial report on how they were doing as a company. Thus, it contains some financial terms, so please just interpret as you normally would, and I will have the software running in the background. I would ask you, however, to mute yourself on Zoom, because if you were to leave your mic on, the software might be overwhelmed with too much speech input. But I would also ask you to please record your interpretation, for example, on your cell phone or any other recording device and then just send me the audio file after we're done.**

P2: Yeah, I've got this recording application on my iPhone. Would that be OK?

I: **Sure, of course.**

P2: I just have never sent files.

I: **The voice memos?**

P2: So, it should be OK, right?

I: **Yeah, of course, sure. When you click on the file and then you can click on "share" and then that way you can share it by sending an email or WhatsApp. So, I'm going to turn on this software and share my screen. OK, can you see the screen?**

P2: Yes, "Augmented interpreter?"

I: **Yes, now I'm going to have to ask you to mute yourself on Zoom and start recording on your other device. And whenever you're ready…**

speech

I: **So that was a very number dense speech, and with that said, I will be looking at how you dealt with it later. That was also the reason for choosing a speech that was quite difficult, of course. And also, I was trying to make sure that it was five minutes long, that's why I may have spoken a little faster than usual. Whenever you're ready, I will ask you to once again say the alphabet backwards just so that I can compare the two times – before and after.**

P2: OK, I'm ready. ZYXWVUTSRQLMNLKJIHGFECBA.

I: **This time was actually faster, so it looks like you're more awake.**

P2: I had some practice.

I: **Yeah, exactly. Thank you very much for that. Like I said, I realize it was not an easy speech, but unfortunately sometimes that's how it is in the interpreting business, especially if it's something like a financial report. Usually, you would have a presentation. This brings me to my question: did you find that seeing the written text of the numbers helped you or was it more of a distraction? And explain why.**

P2: Actually, it was helpful. Although it would have been more helpful if I had had some access to this program. Because, I mean, the interface screen, I saw it for the first time, but I was able to concentrate on the figures, they were in red. But there were some other numbers on the margins, so that might have been a distraction. And I think most of the times, the system accurately captured the numbers. There was one, I think, at least one mistake. I think it was the earnings per book value or something. I think it was 1.5 and it showed like 0.45 or something. At least it clashed with what I heard, but on the whole, it was very useful. And with more practice and more familiarity, I think it could be a very useful tool.

I: **Except for that one time, when you saw that the output was not right, but other than that, did you find the transcription to be reliable? The numbers that were shown.**

P: Yes, although sometimes, for instance, when you read the years, like fiscal 2021 or 2022, the numbers appeared in a slightly strange way. Like 2.021. First, I was surprised, but then I got the hang of it and then it didn't throw me off after that. Although, of course, these numbers are probably the least of my problems. But still, the way it portrayed the numbers sometimes looked a bit strange. Like all those dots in the middle of the number.

I: **But other than that, would you say it was reliable?**

P2: Yeah.

I: **Did you use all of the numbers, some of the numbers or many of the numbers that popped up? Percentagewise?**

P2: I tried, but I lost some of them, definitely. I'm not sure… I hope I got at least 50%.

I: **OK, but whether you got it right or not, did you use them?**

P2: Yeah, at least I looked at them. Sometimes I didn't have time to say them or register.

I: **What were you looking at most of the time? Were you looking at the numbers, the speaker, or something else maybe? Your notes or something?**

P2: The numbers.

I: **I'm going to try to formulate this question in a way that doesn't sound too complicated. Needless to say, the concept of simultaneous importation is, of course, that everything is happening simultaneously, at the same time, yet the process involves a number of so-called efforts that come to play to deliver a product of interpretation. There is an interpreting researcher Daniel Gile who created this Effort Model, which says that simultaneous interpreting is made up of three core efforts: listening, short term memory, and producing. Also, there's a coordination effort, which coordinates the other three. When you have the text in front of you, that is also a visual input. So, then there are four efforts, which are listening, memorizing, producing, and reading. If I were to ask you what you feel that you were most concentrated on, from one to four, with one being what you most concentrated on, for example listening, and four being what you were least concentrated on, for example producing, how would you rate these four? What would you say you were mostly concentrated on or using most energy or efforts on?**

P2: Probably producing, because…. This is not always the case, but now that I'm bit out of practice, especially in the finance and economics area, I haven't done a lot of work, I haven't done any work in this area for a long time, so it was kind of difficult to bring all those terms that I used to have at my fingertips. It was a bit difficult. And of course, sometimes it's difficult to apportion, to allocate the effort between listening and speaking, but normally when I have this feeling, when I know the term, it is an actively used term and it comes to me without any effort, then, of course, the listening part is also easy, because I don't have to concentrate on dragging it out of my memory, it's already here. And I can program my sentence in the back of my head. This pre-programmed sentence is already there waiting to be produced and I can concentrate on the listening part, you know what I mean?

I: **Yeah, I understand.**

P2: Yeah, which is in line with what we were taught, is the Chernov's theory of (in Russian) *probability prediction*. According to his theory, it was the main psychological mechanism behind simultaneous interpreting. So I can already forecast the saying, because I know the terms and I pre-produce it. This is my subjective feeling. This time it was difficult, because the terms were not…sort of...

I: **Available?**

P2: Yeah, ready, and therefore I had to think hard, and I probably wasn't listening very well and that just made me feel I expended more effort also on the listening part. But producing was more difficult.

I: **That being said, did you then spend more time trying to memorize it, or reading the text, making sure that you're reading the right number?**

P2: Probably reading.

I: **More concentrated on reading?**

P2: Because since I had visual support, I felt I did not have to remember, to memorize too much, because I had these numbers ready in front of me.

I: **Another ranking question, on a scale of one to five, with one being not satisfied at all and five being very satisfied, how satisfied were you with your interpretation performance?**

P2: Maybe three or two. Not very satisfied.

I: **OK.**

P2: I mean, it wasn't a complete disaster.

I: **If there were any difficulties, for example, if you didn't hear a word or figure, or there was some kind of challenge that you were faced with, did you apply any specific strategies, any tactics?**

P2: I may have used generalization a couple of times, but I can't give you a specific example. And I also realized a couple of times that I made a mistake, so I had to correct myself. Either it was some grammatical mistake, like cases, declinations, or once, for instance, I thought I heard a term and but then I thought better of it. I thought I heard the Sharpe ratio, but probably it wasn't there. Because it was a financial term, like "risk ratio," but I don't think it was... What was it?

I: **It was "book-to-bill ratio."**

P2: Ah, no, so at one time, for some reason, I heard this Sharpe ratio, interpreted it, and then I realized I had to correct myself probably again. But other than that, I don't know if I used any special tricks or whatever.

I: **It was obviously a short speech, and you only used this software for five minutes, but would you say you enjoyed using it?**

P2: Yeah. It was OK, yeah. As I said, it was a bit new and unfamiliar, but useful, definitely. Because numbers, I've always been bad with numbers and proper names. And when I was at college, we didn't have these special exercises, because later when I did some courses at Cambridge and elsewhere, I saw that students were given special training for numbers. I wasn't lucky to do that, maybe because I didn't attend an interpreter training course at the time as an undergraduate. So, it just is a special skill, right, numbers?

I: **Yeah, absolutely.**

P2: Yeah, in German they are even more challenging.

I: **Exactly, because they're switched around.**

P2: But it could be trained, and the earlier you do it, the better. But as I said, I didn't get exposure to these exercises until I was already an active interpreter. Except in the early days, it was a different set up. We didn't have technical conferences, we had a lot of blah blah blah and numbers didn't matter. I remember a colleague here who said, "well, I don't really care whether it's a billion or a million," you know, that was the attitude. (laughs)

I: **(laughs) Yeah, I don't know about that, it might be problematic.**

P2: It was a general feeling that numbers were not important. And partly because of the nature of work, you know, we had a lot of liberal art stuff and very, very few technical conferences.

I: **You already mentioned that you usually work with two screens, and if you were working with such a tool at a remote interpreting conference, how would your ideal ergonomic setup be? Would you have two screens, what would you be looking at and how would you make sure that you don't miss out on anything important?**

P2: Here I'm a bit torn between. Because, on the one hand. it would be good, probably, to have two screens: one screen for the main interface, the speaker, and another screen for tools like that or for documents. But, on the other hand, in the heat of the battle, additional screens or additional buttons are a challenge, for me at least. One of the problems I experienced, we're coming back to your earlier interview, to my RSI jobs, was that I had to divide my attention between too many functions, and I'm not very good at multitasking. In addition to the usual buttons, for instance, when we were using Interprefy or any other platform like KUDO, I had the microphone buttons, the outgoing and incoming language buttons, I had a special a chat for communicating with my boothmates and Interprefy Technical Support. I had another chat below for the participants and the chair of the meeting and all kinds of documents they were sending. So, my attention was really split, and if I get another screen and another keyboard, I guess this would even increase this multitasking burden. So, I'm not sure which is better, right?

I: **That makes sense, of course.**

P2: But maybe because I am, as I said, of the older persuasion and I'm not one of those digital natives, right? Maybe for you guys it would be easier to have more screens and then you wouldn't be challenged by these buttons.

I: **I guess it depends on the nature of the person. There are people who just don't like technology, doesn't matter what type it is. I think it's different for everybody.**

P2: Some people are challenged by the simple console. We have colleagues here who couldn't cope with the microphone button. (laughs)

I: **Yeah, that's kind of hard if you're an interpreter (laughs)**

P2: I'm not that bad, but still.

I: **This brings me to my final question. You say it is hard for you to divide your attention, that's why you're not sure about how you feel about these kinds of tools, but in the future, are you generally interested in following up what is going to happen to computer-assisted interpreting tools? Do you see yourself using them in the future? Whether it's with such an ASR feature or without?**

P2: If the trend continues and I'm still in the market, which is not guaranteed, because, as I said, I've had very little work and I even had my AIIC status downgraded from active to associate, because I now don't work, well, not enough to be an active member in my view. So, I don't feel like paying the full fee. But if the market picks up and there's some recovery and we finally overcome Corona and maybe some live events come back or I get more RSI work, maybe, yeah, I would be using these tools more actively. But I don't know. Now, the outlook is not very good, not very promising for me. Or maybe I will get out of the market completely or retire.

I: **So, in general you're not very interested.**

P2: (shakes head) Mh.

I: **Well, that was it. Feel free to add anything if you want.**

P2: I was trying to send you the recording, did you get any? Because I'm not sure.

I: **No, where did you send it? E-mail?**

P2: E-Mail, yeah.

I: **Oh wait, I got something. it's M4A format, so that's an audio format, that's fine, thank you very much.**

P2: If for some reason it didn't come through got gobbled in the process, just drop me a line, I will try to resend it.

I: **I think it was fine, looks good. Well, once again, thank you very much.**

P2: No problem, you're welcome.

I: For your time and your effort that makes this research study possible, because I need participants such as yourself, specialists who have been in the market for so long. Thank

you also for sending me your interpretation. And if you're interested, I can, of course, when I'm done with the research, send you a copy if you'd like to read about it.

P2: Yeah, I'd appreciate it.

I**: OK, thank you very much.**

P2: I wish you success with your academic endeavors.

I: **Thank you, thank you very much.**

P2: Bye bye.

I: **Bye bye.**

Interview with Participant 3

I: **Thank you very much for your willingness to participate in this study. Before we begin with the actual interpreting session, I'm going to ask you a few questions about your professional experience as an interpreter. The first question is, how long have you been a professional interpreter for?**

P3: Well, let's see… since '86.

I: **So quite some time now.**

P3: Yeah.

I: **Have you ever interpreted in remote mode?**

P3: Sure.

I: **But was that prior to the pandemic or did it just start once the pandemic hit?**

P3: There were a couple of times before the pandemic, but it was… well, there was no Zoom. What did I do… Telephone, I think, and Skype.

I: **With video or just on the phone?**

P3: I think without video.

I: **Yeah, that's also remote interpreting.**

P3: But, I mean, the pandemic has been here for quite a while now, it sort of loses sense.

I: **True, it has been with us for quite some time now unfortunately. Can you think of any negative or positive aspects in particular when interpreting in remote mode? Is there anything specific that you really dislike about it or really like about it?**

P3: Well, that's actually the usual thing. First of all, most of the time, you have to do it consecutively, it's not necessarily my favorite mode, so to say. Because also there's no discipline on the other side, they start talking amongst each other, arguing and stuff. If it's simultaneous, of course, it's the sound quality, the speakers using just normal iPhone microphone, although they were told otherwise. And if you're working with a colleague, also you have to have other means of communication, like having a WhatsApp chat open and communicating via that and helping each other via messaging. But it takes more time.

I: **Of course.**

P3: And very often they don't even notice it until later, because they're so much engrossed in the process.

I: **How do you usually prepare for such an interpreting situation, where you're interpreting in remote mode?**

P3: The same way.

I: **The same way as you would for a live interpretation?**

P3: Yes, of course. You have materials, you go through them, you make a glossary, just the usual stuff.

I: **Nothing different from how you would prepare for a live interpretation?**

P3: Yeah, the only problem, well, not the only problem, but it just occurred to me, lots of glossaries are on the computer, but you also have a screen [with the meeting] on the computer, you need a second screen, but actually it's difficult to manipulate both while concentrating on the speaker, on the interpretation. So, it's opening up the glossaries, searching information and all that. That's, of course, more intricate.

I: **Of course. So, I'm sure you've heard about computer-assisted translation tools, which have been around for quite some time now and are gaining popularity, a lot of translators are using them, but how familiar are you with computer-assisted interpreting tools? Have you ever heard of them, and can you name any specific ones?**

P3: I cannot name any, but I've used them already before. For example, during a seminar, they said there will be a Ted talk piece being shown, so what I did is first I checked whether there are translated captions already in existence. Very often there are, for Ted talks, actually, that's the rule. And even if I didn't have to interpret and my colleague did, at the same second as it was shown on the screen for the classroom or for the seminar room, I started with the captions and showed it to my colleague or myself. Which would be perfect, you could just read it off, of course, in mind that there are mistakes, like "son" meaning (in Russian) *"son" or "sun."*

I**: Of course.**

P3: Of course, you have to listen to the original. But that's a very good tool. And sometimes I would also generate captions, and there's also such a tool in YouTube. If it's a YouTube piece. If it's a good lecture and if the source is organized and structured in good language, it's a great tool, I would say.

I: **Captions?**

P3: No, not the captions, the translation.

I: **The translation of it?**

P3: Not just captions, I'm sorry, I meant captions of the translation.

I: **OK, so they are already translated**.

P3: Yeah, so speech recognition plus translation. Yes, that's what I mean. Not the original, but, for example, if it's Larry King taking an interview with some sports celebrity and they use Americanisms and talk in-between each other and interrupt each other, then you can forget about it. But I did use it and I love it.

I: **But this is just a YouTube option, so it's not an actual CAI tool?**

P3: Not an actual tool, no. I'm not aware, I'm not very tech savvy (laughs).

I: (laughs) **So not very familiar with any other tools.**

P3: I'm not very familiar and I'm just thinking, with the time left for me, for the simultaneous interpreter to exist, if it's worth it to get into it. Maybe I will if somebody shows me.

I: **Other than the captions that are translated, you've never worked with any actual CAI tools before?**

P3: No.

I: **What about automatic speech recognition, other than in this context?**

P3: No, I haven't. Although my son is actually working in that field, artificial intelligence, speech recognition with multilingual translation.

I: **Oh, so he's active in that area.**

P3: But it's actually more science and theory and experiments rather than practical.

I: **But still, you've heard about it before.**

P3: Of course, yeah.

I: **The software that I will be testing today, which I only have the test version of, is InterpretBank, but have you heard about it prior to this?**

P3: No, just from you.

I: **Part of my study is to test how interpreting with such a tool that recognizes numbers affects not only the performance of the interpreter, but also their level of alertness or fatigue. I will try to draw some comparisons by having you perform a small task, then I will time you to see if there's a difference between how you did before interpreting and after interpreting. I would ask you to say the alphabet backwards while I time you and compare it.**

P3: English? English alphabet or which alphabet?

I: **English, unless you feel it's easier for you to do Russian? Whichever**.

P3: Doesn't matter.

I: **Then we'll do English, because the other participants will also be doing English, so that it's not unfair.**

P3: Oh wow, backwards. I've never done that.

I**: It's OK, I'm not going to correct you (laughs). It's just a small test.**

P3: Oh, should I do it now?

I: **Sure, whenever you're ready.**

P3: ZYXWVT… no, USTSRQPONMLJH… no, KJHGFEDCBA.

I: **Thank you.**

P3: I'm taking it easy, I'm actually in a no-working mood right now, because I'm interpreting, by the way today, for real later on.

I: **Well then, it's a little bit of practice, I guess**.

P3: Yeah.

I: **I'm about to read a speech that was made last year, it must have been November 2021, by the CEO of the company Siemens, a financial report on how they did in 2021. It will contain some numbers, so please just interpret how you usually would, I will then examine your interpretation afterwards.**

P3: Can I use my pen like we usually do?

I: **Sure, just interpret like you usually do. I will, however, be sharing my screen with the tool. Can you see it now? It should say –**

P3: "I am your digital boothmate," yeah.

I: **I'm about to turn it on and start reading my speech. I will ask you, however, beforehand, to record, if you have an audio recording device, or just your phone.**

P3: Just a second, let me turn on my phone so that the screen doesn't accidentally turn off. (gets up) Oh, just a second, I have to open the door, I'm sorry.

I: **No worries.**

P3: (comes back) So, should I start recording?

I: **Yeah, you can. Sure, start recording and then mute yourself on Zoom.**

P3: OK.

I: **Let me open this the tool once again and share my screen. You should be able to see it, yeah?**

P3: (nods)

speech

P3: (in Russian) *that was not easy.*

I: **Yeah, was it a little hard?**

P3: Yeah, that was a bit hard. Because usually there are slides of the presentation and you know what they're talking about.

I: **With zero preparation, of course, it's different. Part of my research is to see if the software can help you in that way, if you were not handed any documents, for example.**

P3: I know what you mean. I can tell you straight away, it was actually disturbing in a way. Because very often, by the time I get to it, when I'm lagging couple of seconds, it's already gone. And you think it would help you, and then suddenly it's gone, and you are actually left hanging. And had I written it down for myself, I would have it in front of me.

I: **So, you found it distracting, OK. But before we get to the second part of the interview, I'm going to ask you once again to say the alphabet backwards. The same thing that you did earlier. I'm going to time you again.**

P3: OK. ZYXWVUTSRQPONMLJH…KJHG…GFEDCBA.

I: **That was almost 10 seconds faster than the first time.**

P3: Because I'm more…hm…

I: **More alert?**

P3: More alert now yeah.

I: **So, you said that it was more of a distraction than it was a help. And what bothered you the most? Was it the fact that the numbers came up too late or too early?**

P3: No, they were not late, they disappeared too early.

I: **OK, I can see why that might be a distraction. But did you find the transcription reliable? Do you think it was accurate?**

P3: No, the transcription was reliable, no problem with that.

I: **And did you use any of the numbers that popped up? Maybe some or not at all?**

P3: Yes of course, I used a lot of them. But sometimes I noticed that, for example, the terms for "revenue per share" or "return for shareholders," while I was thinking how to say that and the number was on the screen, by the time I translated it, it had already disappeared.

I: **I understand. Like I said, this was a hard speech, but that was the point. You're not going to be graded for it (laughs).**

P3: Yeah, absolutely. It was not the best of my performance, I'll tell you straightaway.

I: **OK, no problem.**

P3: But what I also would like to say, what would help more are not just figures, but, even if it's in English, if it said the terms, like revenue for share, what it is in particular.

I: **The thing is, I was actually going to test the software in full, because it does have a feature where you can upload your glossary and it automatically hears the word, for example, "revenue" and then it shows you the English word and the translation right away. It also works with automatic speech recognition.**

P3: My son is also working on that aspect, by the way.

I: **I see, but this is just a test version, so it's not the full version of it.**

P3: Can you give me the name? Because I'll tell him the name of it.

I: **InterpretBank.**

P3: InterpretBank, OK.

I: **Can you remember any numbers in particular that you used or something that was very helpful?**

P3: No, I don't remember because I just immediately went further without keeping them in mind. Because they were on the screen.

I: **You can't remember, no problem. And what were you mostly looking at, the numbers, the speaker, or maybe something else? Your notes?**

P3: I didn't see the speaker. It was maybe on the screen, but I was looking at the shared screen.

I: **I was sharing the screen –**

P3: But I wouldn't see you anyway because that's not the most important thing.

I: The next question I will try to formulate it in a way that doesn't sound too confusing. I don't know how familiar you are with the researcher Daniel Gile who constructed the effort model. Needless to say, the concept of simultaneous interpreting means that everything is happening simultaneously, at the same time: listening, trying to understand what they're saying, trying to remember it, and trying to produce the text, which is all happening at once, meaning a lot of cognitive processes at the same time. He divides simultaneous interpreting into the three main efforts: listening, short term memory, and production, producing what you heard.

P3: The output.

I: **The output, yeah. When you have the text in front of you of the speaker, then another effort is added, which is reading. So, it's a dual input of listening and reading, but also producing and trying to remember what was said. If I asked you to rank these four things, what do you feel that you are most concentrated on – listening to the to the speaker,**

producing the text, reading the numbers, or trying to remember the information? #1 is what you were most concentrated on and #4 which you were least concentrated on?

P3: So memory, no. Just very little, because, for example, if it's reverse order, I had to employ it. Reading, hm… Actually, I've noticed that when it's simple figures like 1.5 or whatever, it's not just long ones, I would even ignore that. And it would be even giving me false hope, so to say. So I would say, reading probably like 4 and 3. And the ranking for the rest, listening and producing…

I: **So, you were most concentrated on listening?**

P3: Yes.

I**: So, listening and then producing and then reading and memory was the least.**

P3: Yeah.

I: **Another scale question. On a scale of one to five, how satisfied are you with your interpretation? With one being not satisfied at all.**

P3: Not satisfied at all.

I: **How would you say, from one to five, one is not satisfied and five is very satisfied, how would you rank it?**

P3: Maybe 2. Because I really lost a lot of things. I would have lots of complaints from the listeners.

I: **This was not the most realistic scenario, so obviously it would be different.**

P3: Yes, I know, of course. And actually, I think I would be writing down those figures and I would be more relying on them than looking at the screen. The screen was actually giving me false hope, in a way.

I: **OK, interesting. And how did you deal with difficulties. for example, if you misheard a word or a figure, or too much information? Did you have any strategies?**

P3: I would omit it. I would omit, or if there's an alternative coming up, like, the growth was so and so many euro or so many percent, I'd just catch up and say how many percent.

I: **I understand.**

P3: Well, omitting, basically.

I: **So, omitting was the general strategy.**

P3: Yeah.

I: **Obviously, this was a very short five-and-a-half-minute speech, and it was a bit hard., but would you say you enjoyed using the tool?**

P3: It has potential, definitely, it has potential.

I**: But in this case, did you enjoy it or not, what would you say?**

P3: I might grow into it.

I: **OK (laughs)**

P3: But the thing is, I would concentrate on the screen and stop writing to myself. And writing other stuff rather than figures, other than figures, would also help.

I: **And if you were to work with such a tool, if you had, for example, not just the number output, but also the glossary that would show you the technical terms, what would your ideal –**

P3: It doesn't have to be a glossary, it could be the English.

I: **Just the English, just the source speech**?

P3: The speech, yeah

I: **And what would your ideal setup be? Would you have two screens, would you have one screen? What would you be looking at?**

P3: One screen, because I don't have time to go between two screens.

I: **Now you've tried this tool out a little bit. Of course, there are different computer-assisted interpreting tools, and not all of them have automatic speech recognition, some of them, for example, help you while you're in the booth to look for a certain term, others help when you're getting prepared for an assignment to look for corpora and other texts. Are you generally interested in these kinds of tools? Do you think you will be using them in the future? Whether it's without with or without an ASR feature, do you see yourself using them?**

P3: I would need some persuasion, I would say. Just to see how other interpreters manage it. But I'll tell you, most of what you've been showing right now is what is usually on the slide, with the pieces like what I've been missing, like return on investment, and it's already there on the slide.

I: **Of course, yeah, with slides, it's much easier.**

P3: So, this other thing would be actually diverting my attention away from the slides, and slides are the most important thing.

I: **Yeah, because you have the reference right in front of you and that's very helpful.**

P3: Yes, and it's more helpful than this tool.

I: **Feel free to add or ask me anything if you have any questions.**

P3: As I said, for the numbers… oh, maybe I missed it, because the numbers didn't disappear, they just went down.

I: **Yeah, they went down.**

P3: And I didn't figure it out.

I: **So, if you would have had more practice with the tool, maybe it would have been more helpful?**

P3: Yeah, I didn't know, because I thought it disappeared but actually, it was still there.

I: **Yeah, the new information was on top.**

P3: Yeah, I see now.

I: **It's called InterpretBank, you can go ahead and maybe try it out once again, just to see how it works. Maybe it has potential so that you might use these kind of tools in the future, maybe not, it is different for everybody**.

P3: Do you use it?

I: **I haven't yet, because I don't have the full version of it, I have the test version and I've been testing it. But I do find it in some cases very helpful. But I've never actually used it during an interpreting session, during a real interpreting job.**

P3: As I said, it might be a potential [help], if there are no slides.

I**: Exactly, if you have zero documents, then maybe that would be helpful, I agree. That brings us to the end. Once again, thank you very much for your time that even makes this research study possible, because I need as many professional interpreters as I can get as probands. And also, the audio file that you just recorded, go ahead and send that to me, by email would be the best option.**

P3: OK, I have your email, just a second. Yeah, I am going to send it to you by email.

I: **Thank you.**

P3: Can you give me your address quickly? I'll just do it right now.

I: **Sure, I'll write in the chat. I'm just writing my address. it's […].**

P3: As I said, that wasn't the best effort.

I: **No worries, you're not going to get graded on it, it's just for analyzing (laughs).**

P3: OK, I think it is done, just see if you received it.

I: **OK, one second, I'll check, let me turn this off really quick.**

Interview with Participant 4

First of all, thank you very much for your willingness to participate in my study, I really appreciate your time and effort. And before we begin with the actual interpreting session, I just would like to ask a few general questions about your professional experience as an interpreter. So, how long have you been a professional interpreter for?

P4: Since 1993.

I: **Since 1993, so it's been quite some time now. And have you ever interpreted in remote mode? I mean, in the pandemic I'm guessing you probably have, but before that?**

P4: No, it only started with the pandemic. Telephone interpreting, yes, there were several occasions, but it cannot be considered complete remote.

I: **And how has your experience been so far with remote interpreting? Are there any negative or positive aspects in particular that you can think of?**

P4: Frankly, mainly negative. But, you know, it depends on the situation. Because I work mainly with US courts, and the US judicial system decided that they will use only Zoom and occasionally Webex, so we are quite limited with the resources. And another very serious drawback of the system is that usually, judges are in a great hurry, because there is a backlog of cases. They want to compress the timing, so even if there would be a possibility to do a proper simultaneous translation, physically it's not possible, because the judges won't let you do this. So it's consecutive and very limited consecutive. In the immigration court, they're mainly using Webex now. So, at some point, it's a little bit easier, but of course it depends. Because sometimes if you are already on the court premises and the judge is using Webex only for his own purposes, you are not hearing it in the headphones, but via some kind of system the judge is using, then again, that’s not good.

I: **So mainly just negative aspects.**

P4: Unfortunately, yes.

I: **How do you usually prepare for a remote interpreting assignment as opposed to live interpretation, is there a difference?**

P4: Well, not really. Because in court surroundings, I never ever have any possibility to know what I'm going to interpret at all. Even the case. Because I'm never informed if it is going to be some kind of domestic battery or just divorce, because I could get a letter D, so letter D would mean “domestic.” What kind of domestic, I will never know. So I have to go and just dive into it.

I: **So usually there's no way for you to prepare, because you don't even know what's coming, correct?**

P4: Absolutely. And I cannot even ask the attorney about the case, because that would be considered biased. Because if I will talk to the attorney prior to the case, I cannot then interpret in the trial.

I: **Wow, that's harsh.** (laughs)

P4: (laughs)

I: **So, computer-assisted translation tools have been around for quite some time now, and translators are working with them more and more. But how familiar are you with computer-assisted interpreting tools? Have you ever heard of them, can you name some, any specific ones?**

P4: I only heard that my colleagues who work in the International Conference settings use them, but I've never had any experience, neither used nor tested them.

I: So, no specific names or anything like that that come to mind?

P4: No, no.

I: **The software that I will be testing today is InterpretBank. Have you ever heard of it, prior to my email?**

P4: No, not prior to your email, no.

I: **Part of my study is to test how interpreting with such an ASR tool affects the overall alertness or fatigue of an interpreter. Therefore, I'll ask you to complete a small task, once before the interpreting session and once after, to say the alphabet backwards and I will time you. Then I will compare the two, how you did before and how you did after interpreting. So, whenever you're ready, I'll start timing you. Just say that the alphabet backwards.**

P4: Z Y X. Well… (laughs) I was really not prepared for this task.

I: **No worries.**

P4: X Y … (long pause) V U S R T. Whoa. (in Russian) *No, this is confusing, I can just do it in the other order, I just read it in my mind twice in the normal order, I can't…*

I: **OK, you can do Russian, if it's easier for you.**

P4: (in Russian) *saying it backwards, I've never practiced that, wait. YA YU, E… No, I can't do this right now, honestly. I've never done this. I've counted backwards before, that's different, but not the alphabet. So, I can't do it, I can't do it backwards.*

I: **Let's see, then maybe I can think of another one.**

P4: (in Russian) *I can count from 100, easy. 100, 99, 98, 97…*

I: (laughs) **We don't have that much time.**

P4: (in Russian) *I can do it from 50 then, 50, 49, 48, 47 …*

I: **Let me think. I don't know what else I could ask you to do. I guess if you do it the right way maybe, I don't know** (laughs). **Better than nothing, I guess.**

P4: (in Russian) *Why it is easier for me to count backwards is because in medical settings, they make you count backwards when giving anesthesia, so that why I am used to that, but alphabet… I have never done it backwards in my life.*

I: (in Russian) ***I guess I'll have to think of something else.*** **Maybe just say it in the right order, I guess.**

P4: ABCDEFG HIJKLMNOPQRSTUV .. VXYZ. (in Russian) *Oh, I forgot W*.

I: (laughs) **No worries. Thank you very much for that, it's just a little exercise. I am now going to read a speech that was performed by the CEO of Siemens in November last year. He's talking about the financial report of the fiscal year 2021. It is a text that contains some specialized financial terminology. So please just interpret as you normally would. There's no pressure, you're not going to be graded for your performance like at school, let's just say that.**

P4: (laughs)

I: **If you have a cell phone device or something to record your interpretation with, I will have to ask you also to mute yourself on Zoom, because otherwise the software will also pick up what you're saying, and then submit it to me after. I will now open the engine and share my screen. Can you see it? Thumbs up if you can see my screen, it should say "augmented interpreter."**

P4: (nods, thumbs up)

I: **Perfect. Whenever you're ready, I will start the speech.**

P4: I'm not supposed to press on anything, I just mute myself and listen to you?

I: **Yes, you mute yourself on Zoom, but you record your interpretation on your phone.**

speech

I: **You can unmute yourself as soon as you're done.**

P4: First of all, I would like to mention that initially, I was so much concentrated on recording that I didn't even look at the screen. So, in the first couple of minutes, maybe three to five, you will see that I didn't look at the screen. Later I realized that all this information that I was pronouncing out loud was on the screen, and it definitely helped. Because at that moment, I was able to concentrate more on the whole content and to try to express myself freely. Because, initially, of course, I was very tired and I probably was making some mistakes, and the level of translation wasn't as smooth as I would like to it to be. But overall, I would like immediately to say that was very helpful, very helpful.

I: **OK.**

P4: I will do my best to send you this recording. I don't know how I will succeed it with email, or how large this file is going to be. So if I won't be able to send it via email, then we will find something else.

I: **Of course, we'll figure it out. But it should be fine. I think if it's only 5 minutes, it shouldn't be a problem. Now, before I ask you a few questions, because we did the task with the alphabet, I'll ask you to do that one more time, to see if you were faster or slower.**

P4: ABC DE F… ABCDEF …. stop…. ABCDEFGHIJKLMNOP RSQUVWXYZ.

I: **Thank you for that.**

P4: (in Russian) *I got it mixed up.* (laughs)

I: (laughs) **No worries, it's not being graded like in school. So, you just said that you found it to be more helpful than a distraction, correct?**

P4: Yes, absolutely. Well, surprisingly to myself, I was able to abstract myself from the whole thing I was listening to. Because the figures which were just popped up, I was able to concentrate on them. Because my concern was that if I will only look at the screen and there will be the whole text, that would be definitely a large distraction, huge distraction. But because there were only figures, I was able quite quickly to adapt to this. At least I hope that was the case.

I: **If that's your perception, then of course, that's all that matters. And did you find the transcription of the numbers reliable?**

P4: I think, yes. Because I remember there were 16.7 billion, there were 17.2, you know, all these things. They were quite clearly represented on the screen, and it helped. But percentages were simple, because, of course, 9, 10, 12% were not a difficult thing. But these things about the growth, the revenue, the income per share, that was the important thing to remember.

I: **You said that the first few seconds, you might have missed any numbers that were coming, but did you use–**

P4: I'm not sure that I missed, but I didn't even think about that, that it would be helpful. I was only trying to record it.

I: **Actually, the first minute almost does not contain any figures, so you probably didn't even miss any. But as soon as you noticed them, did you feel like you used all of them, some of them, or many of them? How many?**

P4: Practically all of them. Because with your speed of speech, it was very helpful. I had a possibility to control what was on the screen. If the speaker would use another speed, something like 180 words per minute, probably I wouldn't be able to do this and then I would start to panic. But with your pace, it was very helpful.

I: **Yeah, I was trying to stay between 115 to 130 maximum words per minute, that's how I timed my speech.**

P4: Yeah, that's very, very helpful. And, by the way, again, in court, whenever you pass their tests, they usually target at 150 at least.

I: **Oh, that's fast.**

P4: Very fast. And that's how judges are mumbling. So you have to do what you can.

I: **Absolutely, yeah.**

P4: You're just surviving, that's it. (laughs)

I: **Yeah, it's like survival mode.**

P4: Anyway, so what I'm doing right now is so different from what I used to do in the past. Maybe I just lost some kind of tempo, or pace, or whatever, I don't know. In general, if I would be in the booth and I would have this tool, and I could see the figures, oh… incomparable. Particularly, when I remember interpreting the annual meeting for gold mining companies, and we were talking about different figures, not only output, but shares and the technicalities, like the tons of different chemical processes. That would be something.

I: **So, you used almost all numbers, but do you remember any in particular that specifically helped you?**

P4: Well, no. I think I was just in the pace with your speech, so it was quite simultaneous, in all senses. That was my feeling.

I: **And what were you looking at most of the time, the speaker, the numbers, or maybe something else? Notes or something like that?**

P4: No, I didn't make any notes. The first note I made, it was 1.5 percent, the very first figure. When I started to make them, but then I realized that I could see it, I stopped making notes and only followed it, and that was it.

I: **What were you looking at most of the time, the numbers then?**

P4: The first line of the numbers. There were columns, but in orange it was the one you used recently. So, that's why I said I was in the same pace as you. Because all the figures were already pronounced before that. At least I hope I did. I was always concentrating on the top line when I was doing this. And it was helpful, definitely.

I: **The next question I'm going to try to formulate and not make it sound too confusing. Needless to say, simultaneous interpreting is based on the fact that everything is happening at the same time, simultaneously, but the process involves a lot of so-called efforts that come into play when interpreting. According to a model that was concepted by Daniel Gile, who is also researcher in linguistics and translation studies, simultaneous interpreting consists of three core efforts: listening, memorizing, and producing. There is also the coordination effort which coordinates the other three. When you have a speech in front of you, whether it's the whole transcript of the speech or just numbers, another effort is added, which is reading. If I were to ask you to rank your efforts, what you felt like you were most concentrated on, listening, reading, producing, or memorizing, #1**

being which you most concentrated on and #4 being which you were least concentrated on, how would you rank them?

P4: Listening, producing, because, obviously, when you try to listen to yourself, you understand that there could be some kind of slip of the tongue or not really good expression, and you're trying to correct in the continuation of your speech and make it more clear what you are doing. Reading, I can tell you, I tried on several occasions before this court job to use a computer when there were speeches of the presenters provided in advance. And usually I found it very distracting. Because for me, just to listen and to look into the eyes of the speaker is essential. It's more important for me to look at the speaker than to read his speech. I don't know if it's my personal feature or if it happens very often, but for me that's it. Memorizing, of course, it goes as a combining factor for all this. Listening and producing. I would say that's in the background. But listening, producing, and watching the speaker.

I: **So, if I understood it correctly, you would say that you were most concentrated on listening, then producing, then memorizing, and then reading? If it's just these four?**

P4: Yep, absolutely. But if, for example, in front of me, in the window of my booth, I will see the speaker, and at the same time somewhere in the corner I will see the figures which would appear, only figures, that would be a mass help, absolutely. But what I don't want is to look at the screen instead of the speaker, that's for sure. That's not my choice.

I: **One of the questions is actually what your ideal ergonomic setup would be if you were interpreting remotely with such a tool, how many screens would you want to have and how would you make sure that you don't miss anything. So, if I understood it correctly, your ideal setup would be having the speaker on the screen and then somewhere on the corner of your eye, or on a second screen, this kind of machine?**

P4: Like in a tiny window somewhere in the corner, left corner or upper corner, depending on the setting, to see these figures. But definitely not the text.

I: **Not the text, but just key factors, such as figures. And another scale question, on a scale of 1 to 5, with one being not satisfied at all and five being very satisfied, how satisfied were you with your interpretation just now?**

P4: Just now, between three and four, I would say.

I: **And if there were any difficulties, for example, if you misheard a word or a figure, or any other kind of challenges, how did you deal with them? Did you apply any certain strategies or tactics?**

P4: I can't say that there were tactics, but there was immediately a flashlight in my mind that I didn't use a very smooth expression [that would] probably be awkward, or there was a slip of the tongue, then I will at least try to repeat this in another sentence in order to clarify it somehow. But in the very short piece, there were not so much space for me to use any other tools, than just simply to try to rephrase it later.

I: **So, rephrasing. Or self-correction?**

P4: Not exactly self-correction, because you don't have time for said “sorry, I didn't say that,” but just try to rephrase it a bit later, including the previous information.

I: **So, rephrasing it, and making it sound better, I understand.**

P4: Exactly.

I: **This was a short speech, and you only were able to use this tool for five minutes, but would you say you enjoyed using the tool?**

P4: Yes, I did, I did.

I: **Now that you've had a little taste of it, are you generally interested in computer-assisted interpreting tools, or do you see yourself using some in the future? Whether it's with such an ASR feature, for making glossaries or for looking up terms in the booth?**

P4: Well, unfortunately, in the foreseeable future, in the environment where I am mainly working with the courts, I cannot say that I would be able to use it. But I would love, I would definitely love to. But unfortunately, there are no booths, no equipment, and so it's mainly a combination of, to put it mildly, home-made simultaneous interpreting, interrupted with consecutive. That's what it is. It's not even whispering, it's not even chuchotage, because you stand next to the judge, and you have to whisper into the ear of the defendant and then, sadly, to interrupt yourself and start to do consecutive.

I: **That might be hard, although there are, I believe, for consecutive interpreting, some tools for notetaking, but I'm guessing if you do simultaneous/consecutive in one session, you probably don't have much time to take many notes.**

P4: Particularly, if the judge from the beginning will tell you “I have 25 minutes for your case.” So that’s the end of the story. Because if it is a trial and they try to follow some kind of procedure – there is an opening statement, there’s a closing statement, there are testimonials – then there is some possibility to use some kind of tool, for consecutive at least. Because testimonials are always [consecutive]. It’s the law, by the way, that you do it only in consecutive. But if it is not a proper trial, if there's just conference or a preliminary hearing, then forget about any tools, you cannot do that.

I: **Feel free to add anything, about your experience or anything in general, because that was actually my last question.**

P4: Well, when I was doing conference interpreting and in the proper settings, with booths, with sound technicians who would assist you, of course, that was a totally different story. And I miss it, I miss it very much. And I’m trying to keep my skills at a certain level, but definitely I have to crop it up to the real situation, because I cannot rely on all the support I used to have working in the conference settings. And in my age, when I'm not looking for any kind of progress in my career, it's only whatever you have, so this is it. (laughs)

I: **Thank you very much for your time and effort to make this research study possible. I need professional interpreters such as yourself who have been in the market for quite some time. And like I said earlier, please make sure that you send me the audio file with your interpretation.**

P4: I will do my best. And accidentally I even recorded a piece of our conversation, but I will see what happens.

I: **No problem, as long as I have the actual interpretation, that's all that matters.**

P4: Well, from my side, I wish you all the best in your thesis, in the preparation of your thesis, in the defending of your thesis. And, of course, I'm sure you will have a beautiful career in front of you, whatever you will decide, the career of a researcher, the career of a conference interpreter, whatever you will decide. I wish you all the best and lots of luck with that.

I: **Thank you, thank you very much.**

Abstract (English)

While computer-assisted translation tools have been around and widely used in written translation for decades, technological support for interpreting has only recently been introduced through a limited number of computer-assisted interpreting (CAI) tools. One of the newest features that has been proposed to enhance CAI tools and possibly optimize the workflow of professional interpreters is Automatic Speech Recognition (ASR). One such software solution that features an ASR function is *InterpretBank*. This thesis investigated the usability and applicability of this tool within a remote simultaneous interpreting (RSI) setting, which is increasingly becoming the most frequently used setting at present. The tool's impact on interpreting quality with regard to accuracy of numbers and fluency of the delivery was assessed. The thesis also investigated the way the tool was perceived by professional interpreters. The experiment consisted of testing the software with the help of four participants who were individually asked to interpret a short English speech containing numbers into Russian via online meeting. Two semi-structured interviews were conducted with the participants before and after the experiment. During the speech, the interpreters were given access to the ASR output by means of the screensharing feature. An analysis of the interpreters' performance was carried out, and both the ASR output as well as the interpreting product were taken into consideration in the examination and data evaluation. To assess the participants' level of alertness or fatigue, they were asked to perform the task of saying the English alphabet backwards before and after interpreting with the ASR tool. The results were then triangulated against the feedback given by the participants at the end of the experiment. Results showed that the use of this ASR tool in RSI had a positive effect on the accuracy of number rendition, as shown by the rather low error rate, but a negative impact on the fluency of the interpretation delivery. The tool was able to deliver an adequate performance, low latency, and high precision rates while being used in remote mode and proved itself to be effective in providing support during the interpretation of a speech dense in numbers. Three out of four participants found the software to be useful and helpful, expressing that they referred to the tool for help in most of the cases when interpreting numbers, and one participant found it distracting. However, even including the ones who had a positive experience with the tool, only one participant showed enthusiasm and interest in using such tools in the future. Three out of four participants were able to perform the alphabet task faster after interpreting than before, stating that they felt more alert after the short interpreting session.

Abstract (Deutsch)

Während computergestützte Übersetzungstools bei der schriftlichen Übersetzung bereits seit Jahrzehnten bekannt und weit verbreitet sind, wurde die technologische Unterstützung für das Dolmetschen erst vor kurzem durch eine begrenzte Anzahl an computergestützten Dolmetschtools (CAI) eingeführt. Eine der neuesten entwickelten Funktionen, um CAI-Tools zu verbessern und den Arbeitsablauf professioneller Dolmetscher zu optimieren, ist die automatische Spracherkennung (ASR). Eine Softwarelösung, die eine ASR-Funktion bietet, ist *InterpretBank*. In dieser Arbeit wurde die Nutzbarkeit und Anwendbarkeit dieses Tools im Rahmen des Remote-Simultandolmetschens (RSI) untersucht, das derzeit immer häufiger zum Einsatz kommt. Die Auswirkung des Tools auf die Qualität des Dolmetschens im Hinblick auf den Redefluss und die Genauigkeit der Zahlen wurde bewertet. In der Arbeit wurde auch die Wahrnehmung dieses Tools durch professionelle DolmetscherInnen untersucht. Das Experiment bestand darin, die Software mit Hilfe von vier TeilnehmerInnen zu testen, die einzeln gebeten wurden, eine kurze englische Rede, die viele Zahlen enthielt, per Online-Meeting ins Russische zu dolmetschen. Vor und nach dem Experiment wurden Interviews mit den TeilnehmerInnen geführt. Während der Rede hatten die DolmetscherInnen mit Hilfe der Screensharing-Funktion Zugriff auf den ASR-Output. Eine Analyse der Leistung der DolmetscherInnen sowie deren Angaben im Interview wurde durchgeführt. Bei der Untersuchung und Datenauswertung wurden sowohl der ASR-Output als auch die Verdolmetschung berücksichtigt. Um den Grad der Wachsamkeit beziehungsweise Müdigkeit der TeilnehmerInnen zu beurteilen, wurden sie gebeten, das englische Alphabet vor und nach dem Dolmetschen mit dem ASR-Tool rückwärtszusprechen. Die Ergebnisse wurden dann mit dem Feedback der TeilnehmerInnen am Ende des Experiments abgeglichen. Die Ergebnisse zeigten, dass der Einsatz dieses Tools bei RSI eine positive Auswirkung auf die Dolmetschleistung im Hinblick auf die Genauigkeit der Zahlen hatte, wie die relativ niedrige Fehlerquote ergab, jedoch den Redefluss in der Verdolmetschung negativ beeinflusst hatte. Das Tool erwies sich als effektiv und lieferte eine angemessene Leistung, niedrige Latenzzeiten und hohe Präzisionsraten im Remote-Modus. Drei von vier TeilnehmerInnen empfanden die Software als nützlich und hilfreich und gaben an, dass sie beim Dolmetschen von Zahlen meistens auf das Tool zurückgriffen, während einE TeilnehmerIn das Tool als störend empfand. Drei von vier TeilnehmerInnen konnten nach dem Dolmetschen das Alphabet schneller rückwärts rezitieren als vorher und gaben an, dass sie sich nach der kurzen Dolmetschersitzung wacher fühlten.