universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The Impact of Poor Audio Quality on Simultaneous Interpreting "

verfasst von / submitted by

Tatjana Roos, BA

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

Master of Arts (MA)

Wien, 2023 / Vienna 2023

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. Dr. Franz Pöchhacker

Acknowledgement

I would like to thank all those who have supported me throughout the writing of this thesis.

I would like to extend my greatest thanks to the supervisor of my thesis, Prof. Dr. Pöchhacker, for his valuable contributions.

I would also like to thank all the students at the Centre for Translation Studies at the University of Vienna who have kindly helped me with my thesis by offering their interpretations.

I would also like to thank my family for their support throughout my studies and for believing in me, always. Quero agradecer especialmente à minha mãe, por todo o apoio e inspiração que ela sempre me proporcionou. Und auch ein Danke an meinen Vater, der mir immer wieder zeigt, dass man das Leben nicht allzu ernst nehmen muss.

Finally, I would also like to express my gratitude to all my friends who have shared this journey with me.

Table of contents

List of figures

List of tables

0. Introduction

Remote simultaneous interpreting (RSI) has been part of interpreting work settings since the late 1960s. The first major experiments with RSI were carried out in the 1970s by UNESCO, both in 1976 with the Paris-Nairobi ('Symphonie Satellite') experiment and the New York-Buenos Aires experiment in 1978 (Moser-Mercer 2003). Ever since, many more experiments have been carried out to test several variables related to this interpreting setting, such as the quality of the interpreters' output as well as to what degree this setting affected their motivation and stamina (Moser-Mercer 2003; Kurz 2000). Most studies showed that interpreters liked remote interpreting less than on-site interpreting but also indicated that the setting had no negative impact on the quality of the interpretation, despite interpreters believing this to be the case. The recent pandemic forced the entire world to stand still and restructure their day-to-day lives, both privately and professionally. Among the affected were interpreters, who now, more than ever, were dependent on online remote settings in order to continue to work. This applies both to professional interpreters and to interpreting students, who now had to shift to an online teaching setting. This would impose new challenges on the education of future interpreters, one of them being the challenge of having to work with reduced or overall poor audio quality. The importance of visual input for both simultaneous and consecutive interpreting has been widely researched and tested (Rennert 2008; Saeed et al. 2022). However, the effect of poor sound quality is much less focused on. The present thesis aims to outline the cognitive and physical repercussions of poor audio quality for interpreters as well as how these can be measured in terms of the interpreter's output. The underlying theoretical framework of this paper is Gile's Effort Model for simultaneous interpreting, which describes the four cognitive efforts necessary to perform an interpreting task.

Chapter 1 will offer a more detailed definition of the necessary terminology, including that of simultaneous interpreting and remote simultaneous interpreting (RSI), and describe the development of RSI and other forms of remote interpreting from the 1960s up to the recent start of the Covid-19 pandemic. It will also outline standards and requirements set for this practice and further elaborate on why these requirements are important for successful interpreting practice. Further, it will describe disadvantages and possible health risks of working with poor audio quality over an extended period of time. Chapter 2 will focus mainly on the cognitive aspect of

simultaneous interpreting. Firstly, an explanation of Gile's Effort Model will be provided, followed by an attempt to correlate poor audio quality and listening effort with increased cognitive demand. The effect of degraded sound on audio comprehension and on short-term memory will be outlined in this chapter and several studies carried out in this field will be described. Chapter 3 describes further variables that can affect the outcome of an interpretation, focussing on factors related to the source text, to the speaker and to the interpreting environment. The fourth chapter offers a description of disfluencies and explains how they can be a measurable variable linked to cognitive effort. Chapter 5 introduces the experiment carried out for the empirical part of the present thesis and outlines the material and methodology used as well as describing the participants of the study, the procedure of the experiment and the data analysis that followed the experiment. In this chapter, the empirical material is analysed regarding its validity for the experiment that is carried out and also concerning the source-text-dependent factors that affect the interpretation as described in Chapter 3. The experiment described in Chapter 5 involved six interpreting students from the University of Vienna, who were all at an advanced stage of their MA in Translation with a focus on conference interpreting. The participants were asked to interpret approximately nine minutes of a YouTube video which showed a panel discussion held via the platform Zoom. Five panellists were invited to the discussion, three of whom were not using a headset. For one of the panellists this resulted in poor audio quality while he was speaking. The aim of the experiment was to test whether the interpreting students produced a higher number of disfluencies while interpreting the segment with poor audio quality compared to the segment with good audio quality. According to Gile's Effort Model, the listening and analysis of the source text would require additional cognitive resources, leaving less capacity for the production effort and output control. The hypothesis was that the rendition of the poor audio quality segment would include more disfluencies. The sixth chapter is divided into six subchapters, one for each participant of the experiment, and contains the analysis of the findings. Chapter 7 offers a discussion of the findings of the experiment and a conclusion. This chapter focuses on the overall trend of the results and reveals whether they confirm or disprove the hypothesis made for the thesis. The limitations of the thesis and the experiment are outlined in this chapter, offering suggestions for further research to be done in this field.

1. Remote simultaneous interpreting (RSI)

1.1 Definition of key terms

1.1.1 Simultaneous interpreting

Simultaneous Interpreting (SI) can be defined as an interpreting mode in which the interpreter reproduces a speech that is being delivered by a speaker into another language with a time lag of a few seconds (Diriker 2015). In this mode, interpreters mostly work in soundproof interpreting booths with equipment that aims to prevent acoustic overlap between the original speech, heard by the interpreter via headphones, and its simultaneous interpretation spoken into a microphone. Further, SI can be practiced without an interpreting booth, such as is the case for whispered interpreting, also called *chuchotage,* or through a mobile system, called *bidule* (Diriker 2015:383). Simultaneous interpreting can also be used between spoken and signed languages, in which case it is referred to as 'bimodal interpreting', or between two signed languages.

Simultaneous interpreting dates back to the early 1920s, when Edward Filene and Alan Gordon-Finlay used telephone technology to develop the first telephonic interpreting equipment. It was not, however, until the Nuremberg Trials following World War II in 1945 that the interpreting mode became known all around the world. Up to this point, consecutive interpreting had been the main interpreting mode used for multilingual conferences. This meant that interpreters had to take notes during the delivery of a speech and, once the speech was finished, reproduce it in another language. Nowadays, simultaneous interpreting has become the norm during international meetings, especially if more than two languages are involved (Seeber 2015).

1.1.2 Remote simultaneous interpreting

Both the International Organisation for Standardisation (ISO) and the International Association of Conference Interpreters (AIIC) define remote simultaneous interpreting as an act which is supported by information and communications technology, and which consists of simultaneously interpreting a speaker that is in a different location

from the interpreter. While ISO refers to this as distance interpreting, some organisations denominate this setting as remote interpreting (ISO 2017). AIIC differentiates between distance and remote interpreting, defining the former as a setting in which the speakers are in one location and the interpreter in another, and the latter as a setting in which the interpreter is dislocated both from the speaker and the audience and works with either only sound, audio remote interpreting, or sound and image, so-called video remote interpreting (AIIC 2020a). Braun defines remote interpreting (RI) as the use of communication technologies to gain access to a language professional that is in a different room, building, city or country. In this setting, according to Braun (2015:352), “a telephone line or videoconference link is used to connect the interpreter to the primary participants, who are together at one site.” The setting in which interpreting is carried out by telephone is also called telephone interpreting or over-the-phone interpreting. It was proposed in the early 1950s and is now widely used for interlingual dialogues. Pöchhacker (2016:184) further elaborates that this particular form of interpreting was first implemented in Australia in the 1970s, more specifically within an Emergency Interpreter Service. A decade later, telephone interpreting spread across the US and has since grown rapidly worldwide. Remote interpreting by videoconference, on the other hand, is mostly only referred to as remote interpreting when used in connection with spoken-language interpreting. Braun refers to this as videoconference-based interpreting. In the field of sign language interpreting, the term video remote interpreting has established itself (Braun 2015:352). In this setting, the interpreters and the communicating parties are connected via audio and video, meaning that the interpreter and their clients all see each other (Nimdzi 2020). This differs from remote simultaneous interpreting, because in RSI the listeners only receive an audio, whereas the interpreter receives both audio and visual input from the communicating parties. As the name suggests, RSI always refers to simultaneous interpreting. Both video remote interpreting and remote simultaneous interpreting fall under the category of remote interpreting. A further distinction drawn by Braun in reference to remote interpreting is that of two-way and three-way telephone or videoconference connections. The first type consists of both speakers being in one location and the interpreter elsewhere and in the second type both speakers are at different locations and the interpreter is in a third location, hence the denomination of three-way connection.

Remote interpreting is linked to many advantages, such as a reduction of costs since interpreters no longer have to travel to the location of a conference and sleep in a hotel (Ziegler & Gigliobianco 2019). It also offers environmental advantages and can be a solution to space constraints in conferences where interpreting into several languages is necessary (Mouzourakis 2006:46). Remote Interpreting also enables interpreters to offer their services to more than one client per day and take advantage of different time zones, which would be a great advantage to interpreters covering 'exotic' languages. International organisations, such as the UN and the EU, which have staff interpreters located at several of their duty stations, including New York, Geneva, Vienna or Nairobi, would profit from remote interpreting as it could provide for a more efficient use of human resources (Mouzourakis 2006). Mouzourakis saw a potential for development in this type of interpreting up to the point of which interpreters might work from "a dense network of decentralised interpreting stations linking all the major cities of the globe." (2006:47). Barbara Moser-Mercer (2003), on the other hand, outlines the disadvantages related to remote interpreting, referring to psychological aspects of stress management due to being far away from the speakers, cognitive stress caused by the multitude of information to be processed, dependency on a screen to retrieve visual information and a feeling of isolation and lack of motivation. Ziegler and Gigliobianco (2019) further describe a range of physiological factors that have prevented remote interpreting from gaining acceptance within the interpreting community, naming fatigue, stress and further symptoms such as headaches and concentration problems. Overall, interpreters also feel an overall sense of unease due to "not being there" (Mouzourakis 2006:56). Most of these factors are attributed to the fact that interpreters only had a limited view of the speaker and the audience (Ziegler & Gigliobianco 2009:120).

1.2 Pre- and post-pandemic development of RI

Although RI has experienced a strong commercial boom in recent years, the technology itself is not entirely new (Moser-Mercer 2003). Braun (2015) explains that the development of videoconference-based interpreting was driven by the interest of supra-national institutions, naming the UN and the EU as two examples. The first

significant experiments to test the feasibility of remote interpreting were run in the 1970s. Among the most relevant experiments is the 'Symphonie Satellite' experiment carried out by UNESCO in 1976, connecting interpreters and speakers from Paris to Nairobi, and the New York - Buenos Aires experiment by the United Nations in 1978 (Moser-Mercer 2003; Braun 2015:354). The first UNESCO experiment had the main objective to reduce conference-related costs. It connected the headquarters in the French capital with a conference centre in Nairobi and included three interpreting methods: remote interpreting by telephone, remote interpreting by video link and interpreting in a videoconference between Paris and Nairobi. All interpreters were in Paris (Sommerlad 1977; Ziegler & Gigiobianco 2019:123). The second UNESCO experiment was organised in 1978 for a conference in Buenos Aires, where interpreters were active both on-site and remotely, working from New York. The group working remotely received audio-visual signals through a satellite connection. The experiment showed that there was a possibility for interpreters to work remotely and still achieve high-quality interpretation (UNESCO 1987:26). However, the interpreters working on-site did perform better due to more in-depth preparation and more extensive knowledge about the conference (Chernov 2004:82-90). The European Commission conducted several experiments in this field in the mid-90s and the European Telecommunications Standards Institute (ETSI) led a pilot study in 1993 which tested ISDN video telephony for conference interpreters. Further experiments were conducted by the European Commission in 1997 (Zaremba 1997) and 2000 (European Commission 2000), the European Parliament in 2001 (European Parliament 2001) and the United Nations in 1999 (United Nations 1999) and 2001 (United Nations 2001). Although some of the available reports do not distinguish between remote and teleconference interpreting, they indicate that RI was seen as a challenge or even unacceptable, whereas interpreting in a videoconference link seemed more feasible (Braun 2015:354). Mouzourakis reported in 2006 that the studies conducted by the different supra-national institutions used a variety of technical conditions. The studies conducted by the ETSI, the International Telecommunications Union (ITU) in collaboration with the École de traduction et d'interprétation (ETI) and the first European Commission study all used ISDN connections based on H.320, which was the encoding and transmission standard developed by the ITU for ISDN-based videoconferencing. However, the technology used was found to be unacceptable for simultaneous interpreting as it did not meet the ISO 2063 standard

for simultaneous interpreting in terms of sound quality. The experiments conducted by the United Nations used a different ISDN technology, consisting of connections with non-standard encoding to achieve a better audio signal. The later studies carried out by the institutions of the European Union were based on yet another technology: coaxial or fibre optics cable connections to avoid a loss of both audio and visual quality. Furthermore, the equipment used was not the same between the studies. Mouzourakis (2006:52) explains that it was difficult to attribute the range of physiological and psychological problems observed in the participants to a particular technical setup or even a specific set of working conditions provided by a particular organisation due to the lack of consistency between the studies. According to Mouzourakis, the problems observed were most likely caused by the overarching condition of remoteness.

The 'Code for the use of new technologies in conference interpretation' published by the International Association of Conference Interpreters (AIIC) in 2000 clearly reflects the resistance towards remote interpreting in the community of professional interpreters. They warned that "the temptation to divert certain technologies from their primary purpose e.g. by putting interpreters in front of monitors or screens to interpret at a distance a meeting attended by participants assembled in one place (i.e. tele-interpreting), is unacceptable." (AIIC 2000). Some institutions were discouraged by this position. They had hoped that remote interpreting was a way to improve interpreter availability by reducing time and travel expenses. They also thought, from the 1990s, to have found a solution to meet the logistical and linguistic challenges imposed by the expansion of the European Union and its consequent shortage of interpreting booths in meeting rooms (Mouzourakis 2003). Braun and Taylor (2012) conducted a survey in which 200 legal interpreters were questioned on their attitudes towards videoconference-based interpreting. Despite the fact that many interpreters saw the implementation of this type of interpreting as an attempt for clients to reduce costs, some interpreters also had a positive attitude towards remote interpreting, specifically in relation to its potential for improving access to interpreting services and fairness of justice. The survey also revealed links between the interpreters' views towards videoconferencing and the quality of the equipment available in the country they worked in and the overall working conditions (Braun 2015).

The European Parliament carried out a more extensive study in 2004, which showed little evidence that RSI, when performed from a booth with high-quality video screens, had a negative impact on the interpreters' performance quality (Roziner &

Shlesinger 2010). Contradicting this, the study also showed that interpreters, in addition to eye strain and drowsiness, experienced higher levels of burnout in terms of mental and physical exhaustion, cognitive fatigue and mental stress when interpreting from a remote location. The authors deduce that the interpreters' feeling of discomfort and alienation cannot be caused by the differences in environmental conditions. As found in previous studies (e.g. United Nations 2001), the impact on interpreters when interpreting remotely seems to be mainly psychological.

More recently, several European countries have implemented video conference facilities in courtrooms based on the ITU's newer H.323 standard for video conferences, which delivers better video and audio quality because it uses the internet as opposed to ISDN-based systems. The objective behind this was to encourage and promote the use of videoconferencing in legal proceedings within the EU. These systems can provide better support for videoconference-based interpreting than their predecessors if used in combination with high-end peripheral equipment such as cameras and microphones. Nevertheless, the availability and use of web- or cloud-based videoconference services, which are linked to a varying and unstable sound and image quality, and their accessibility via tablets or other mobile devices, more specifically in healthcare settings, cast doubt upon the practicability of remote interpreting using these very systems.

With the Covid-19 pandemic and the consequent shift towards a home office culture, a new debate about RSI has been sparked among the interpreting community. Reviews on dozens of RSI platforms and guidelines and recommendations have been published by associations and standardisation bodies, as will be discussed further in Chapter 1.3. According to Caniato (2020), the quality of sound received little attention in these publications. Instead, whenever sound was discussed, the main concern revolved around the prevention of acoustic shock and the use of correct headsets or specific devices that would be attached to interpreters' computers to remove any sudden, loud noises.

The challenge in adjusting to virtual meetings was felt across organisations. Everyone started feeling a type of fatigue resulting from prolonged exposure to digital compression used in web conferencing, which Kate Murphy (2020) refers to as "Zoom fatigue". The unresolved issues of sound quality, cognitive load and quality of service, which had been exposed by previous studies, were once again the reality for interpreters worldwide. While clients were aware of the poor audio quality in online

conferences, most participants attended meetings only irregularly and if they failed to understand or hear parts of the conversation, they were in a position to interrupt the speaker to clarify or even to disregard their lack of understanding. Interpreters, on the other hand, did not have the same luxury and had to work with compressed sound every day (Seeber 2021:264).

1.3 ISO standards and AIIC requirements for RSI

The International Organisation for Standardisation (ISO) is a federation of national standards bodies, the ISO member bodies, which acts and is recognised worldwide. ISO technical committees prepare the International Standards in a given subject and if an ISO member body is interested in a specific subject for which a committee has been established, they have the right to be represented in that committee (ISO 2016). The development of ISO standards is done in collaboration with international organisations, both governmental and non-governmental. There is also a close collaboration between ISO and the International Electrotechnical Commission (IEC) in matters relating to electrotechnical standardisation (ISO 2016).

Within the field of interpreting, ISO standards have been developed to regulate technical requirements for interpreting booths (ISO 2016), technical setups with reference to equipment used by interpreters and clients (ISO 2017), technical requirements for interpreting platforms (ISO/PAS 2020) as well as interpreting practices in several settings, such as legal (ISO 2019) or healthcare (ISO 2020) interpreting, among others. Within the context of this paper, the most relevant standards are those which refer to technical requirements for interpreting booths and equipment, which ultimately affect the quality of the sound input available to the interpreters. With this in mind, the two main ISO standards under discussion are ISO 20108:2017 and ISO 20109:2016. ISO 20108:2017 standardises the technical requirements for the quality of transmission of sound and image input to interpreters and specifies the characteristics of the audio and video signals. In addition to all requirements referring to on-site interpreting, in which all conference participants and the interpreters are located in one place, ISO 20108:2017 also specifies requirements

for distance interpreting settings. The focus of the requirements described in this ISO standard lies on the frequencies that need to be transmitted by the equipment used both by the interpreter and the speakers as well as on technical conditions for the interpreting equipment that directly affect the sound quality, such as echo cancelling properties and protection against acoustic shock. ISO 20108:2017 defines that all sound input to the interpreters should automatically be adjusted so as to minimise the volume difference between the remote source, the floor and all languages, specifying an input level between nominal and ± 12dB. Further, the ISO standard stipulates that the Speech Transmission Index, that is, the quality or intelligibility score of the speech, should be no lower than 0.64 as set by IEC 80268-16 (ISO 2017). With reference to the microphones used by the speakers, they should have the same technical characteristics as those inside the interpreter console, further defined in ISO 20109:2016. When no audio input is switched on, one or more microphones should be turned on in order to provide the interpreter with the ambience of the conference room. The ambient microphone(s), however, should be adjustable by the interpreter. In a distance interpreting setting, any sound deriving from a distant site or being transmitted to a distant site should have a delay of no more than 100ms. Appropriate measures should be put in place for the quality of the connection to the distant site to prevent audible artefacts. Additionally, acoustic echo cancelling should be set up on all sites.

ISO 20109:2016 specifies the requirements for the equipment used for simultaneous interpreting. According to the standard, the audio processing should be entirely digital and overall latency between the input into the microphone and the output through the headphones should be no longer than 10ms. The ISO standard states that all sound pressure levels (SPL) mentioned should be based on a nominal input level, which is to be between 85 dB_{SPL} and 115 dB_{SPL} at a speaking distance of 30 cm. For the use of passive headphones, the throughput should be between 95 dB_{SPL} and 115 dB_{SPL} per milliwatt and the impedance should be 32 ohms (ISO 2016). The frequency reproduced by the interpreting system, excluding the microphones and headphones, should be “between at least 125 Hz and 15 000 Hz ± 3 dB, high-pass with attenuation of at least 12 dB per octave for frequencies below 125 Hz in order to improve speech intelligibility.” (ISO 2016). For microphones and headphones, the audio-frequency reproduction range should be the same, but ± 10 dB. For optimal intelligibility, the interpreting system must be free of any perceptible distortion and the total harmonic distortion (THD) should be below 1%. Furthermore, there should be no audible noise

or hum within the interpreting system, with an end-to-end signal-to-noise ratio of at least 95 dBA at maximum input level. For the interpreter's safety, an audible hearing damage warning must be activated when the mean SPL is higher than 80 dBA_{SPL} for more than 60 seconds. For the same reason, there should also be an inbuilt limit to loud sounds, for which 94 dBA_{SPL} is the maximum output level for any sounds that last longer than 100ms. The interpreter's microphone must have an adjustable stem so that it can be oriented in any desired acoustical direction. Its polar pattern should be directional, so that it can simultaneously supress any external noise, while also offering high sound quality, even if the speaker or the interpreter changes their position relative to the microphone. It should additionally be mounted in a way that does not transmit any contact noises. Lastly, it should also not pick up an audible interference from any electromagnetic sources that might be nearby. The abovementioned requirements apply not only to microphones attached to an interpreting console, but also to any mobile microphones, such as handheld, lapel or neckband microphones as well as microphones that are built into a computer or laptop.

The International Association of Conference Interpreters states on its website that it "works to safeguard interpreters' health, to give advice on technical standards and to negotiate terms and conditions with the major international organisations such as the EU and UN." (AIIC n.d.). In reference to distance interpreting, AIIC has composed its own set of guidelines and checklists for interpreters and clients to abide by. In these recommendations, reference is made to the abovementioned ISO standards, specifically concerning the technical requirements, such as the audio frequency reproduced by headphones and microphones as well as the technical requirements for the interpreting booth and the interpreting console, both on-site and within RSI platforms. In addition to the technical standards set by the ISO, AIIC also offers a set of guidelines for interpreters and speakers on topics such as microphone management, the use of remote interpreting platforms (AIIC 2020b) and on interpreting from home during the Covid-19 pandemic (AIIC 2020a). For remote speakers, for example, the Association compiled a list of tips on how to work with remote interpreting platforms, adding elements such as replication of eye contact by looking into the camera, turning off all sound notifications, using microphones that comply with the technical requirements set out by the ISO, avoiding touching or moving the microphones while speaking as well as advising speakers to talk directly into their microphones at a distance of 30cm to 50cm and to turn their microphones off once

they finish speaking (AIIC 2020b). The AIIC recommendations also outline the rights of the interpreters, notifying clients that the audiovisual quality is imperative for an interpreter's performance and in case of pixelation, freezing of the image or partial or even complete loss of audio input, the interpreter can no longer perform their task. In such cases, the interpreter must indicate to the technician that the sound is inaudible and may suspend their interpretation until the audio and/or visual quality is deemed sufficient to continue (AIIC 2020b). The AIIC interpreter checklist for remote interpreting assignments from home, which was published during and because of the Covid-19 pandemic, entails further technical considerations for all stakeholders (AIIC 2020c). It includes requirements for the employer, client or organiser regarding the RSI system or platform to be used. The latter should have the necessary technology to enable booth partnering during a conference, specifically having a handover system and including an easily visible and accessible communication channel for all team members, including the booth partner, the chair, the meeting coordinator and technicians. The interpreting platform must also offer a preselection of all relevant input and output channels and adjustable volume configurations. For interpreters' safety, the RSI platform must also offer adequate protection against acoustic shock in accordance with ISO 20109:2016. The checklist also states the necessary preparations before the start of a conference, suggesting a full test run to be carried out involving all parties present in the event. Furthermore, the interpreters should have full visual access to the active speaker or signer and all presentations and slides. Another audio-related point in the checklist is whether the employer or client assumes liability for any sudden acoustic shock or peak loads that might damage the interpreter's hearing (AIIC 2020c). A further section in the AIIC checklist refers to pre-meeting training for interpreters and speakers. This training includes presenting documentation in a videoconference, which speakers might not be familiar with. Lastly, it is recommended that training should be provided to customers on the topic of microphone management, covering matters such as how to speak into a microphone, when and how to mute it, how to prevent microphone feedback and acoustic shock, for example by not tapping on or dropping the microphone.

1.4 Poor sound quality and its impact on interpreters' physical and mental health

The growing exposure of interpreters to acoustic threats, which can lead to symptoms such as tinnitus, hyperacusis or hearing loss, is becoming more and more known within the conference interpreting industry. AIIC is currently conducting a study on 'acoustic shocks' in the booth following reports that a growing number of interpreters who work remotely for the Canadian Parliament is experiencing the abovementioned symptoms, which indicates that the RSI setting could increase an interpreter's chances of developing hearing problems (Caniato 2020; Guiducci 2020; AIIC 2020d).

Symptoms similar to those experienced by interpreters who have been exposed to remote interpreting for a prolonged period of time can be observed in call centre worker populations. Sudden spikes in sound pressure levels (SPL) cause reflexes in the human ear, which may lead to syndromes such as 'Acoustic Shock Disorder' (ASD). This is caused by a high input of sound pressure 'shocks' on an otherwise healthy ear. It is the reason why ISO and AIIC standards recommend SPL output limiters. These limit the number of decibels a headset can produce, meaning that no sudden noise will reach a SPL high enough to cause an 'acoustic shock'. Unfortunately, having to extract meaning from poor sound with a low signal-to-noise ratio, that is, with excessive levels of background noise and microphone gain as well as from distorted sound has become a regular occurrence in RSI settings, regardless of the quality of the interpreter's headphones (Caniato 2020). In order to extract meaning out of the degraded sound, interpreters often turn up their volume to maximise the amount of 'useful' signal. Over longer periods of time, this exposure to high volume acoustics fatigues the middle ear muscle, leading to detrimental developments in interpreters' hearing even in the absence of sudden SPL spikes (Caniato 2020). When referring to 'acoustic shock' most definitions describe an unexpected, loud sound which can cause a range of symptoms, such as pain, tinnitus, vestibular disturbance and hyperacusis. The term does not apply to increased volumes of sound, such as that of an airplane engine, nor does it suggest that these high SPL sounds instantly shatter the eardrum of the listener. According to Hooper (2014) this type of hearing damage would only occur if the SPL were to reach 120dB, which with today's technology is highly unlikely as it is, by default, limited in this regard. In fact, the ear canals and eardrums of people who have reported to be suffering from 'acoustic shock' appear to be healthy and

undamaged (Westcott 2008). Further, Acoustic Shock Disorder symptoms are known to arise in individuals that have never before experienced a sudden and painful spike in SPL, but do have a long, cumulative exposure to headset use (Westcott 2006). The strenuous task of extracting words from compressed sound strains the middle ear muscles, which continuously adjust the impedance of elastic membranes to chase 'load bearing' frequencies that are either corrupted by interfering noise, have been masked or removed or that are distorted or replaced with surrogates during compression, which happens often in RSI. At the same time, these muscles stiffen the very same membranes, that is, reduce their elasticity, to protect the inner ear as a reflex to the poor-quality acoustics. As a result of degraded acoustics, interpreters turn up the volume of the incoming audio, while also raising their own voices as a consequence (Wu et al. 2018), creating a vicious cycle that will expose the ears to many more decibels that would normally be needed. If the hearing is subjected to this for a long period of time, it may become damaged, leading to symptoms and syndromes such as ASD.

Guiducci (2020) traces the sound quality issue back to transmission chains. He explains that the transmission chain consists of every signal between the speaker's microphone and the interpreter's headphones and states that any type of sound alteration that leads to poor audio quality is usually introduced within the transmission chain. Therefore, according to Guiducci, no ISO-compliant headphone or microphone would suffice to compensate for the damage induced by an inadequate audio chain. In conference settings, and more so in RSI settings, audio chains are often poorly managed. They oftentimes consist of compressors, poorly tuned equalisers, limiters or feedback prevention mechanisms, which will lead to sound degradation, regardless of whether the interpreter equipment is otherwise ISO compliant. All this happens, according to the author, in times of modern technology which allows for a digital transmission of high-quality sound either on-site or across the globe, with even cheaper setups, reduced cabling and greater language availability, flexibility and scalability. During on-site interpreting, the audio data is sent to the interpreting consoles, converted back to analogue sound by a digital-to-analogue converter (DAC), amplified and sent into the interpreter's headphones. When a distant site is added, however, the data will often travel great distances through the internet and will be affected by factors such as transmission protocol resilience, bandwidth capacity and overall network latency. All these elements will influence the quality of the sound

ultimately reaching the interpreter's headphones (Guiducci 2020). Guiducci associates the occurrence of muffled sound with poor equalisation or an exaggerated use of feedback control mechanisms. Low sampling rates or low quality of real-time compression codecs decrease the overall frequency response available to the interpreter. Further, echo suppression algorithms and noise reduction filters destabilise speech artefacts, as they cut out some of the useful acoustic information in the process.

In addition to the abovementioned physical health consequences, expending effort on listening may also lead to psychological and physiological symptoms, namely increased stress and fatigue, which in the long term may also affect the interpreter's mental health (Hornsby 2013; Pichora-Fuller 2016; Wang et al. 2018). The term 'listening effort' is defined either as (a) understanding a speech signal that is distorted (Francis et al. & Alvar 2016), atypical (McAuliffe et al. 2012) or masked or reverberant (Desjardins & Doherty 2013) or (b) listening while being distracted by competing information (Janse 2012; Koelewijn et al. 2012) or (c) having to simultaneously hold and rehearse unrelated information in mind while listening (Francis 2010; Rudner et al. 2012). When considering these possible definitions, a rephrasing of Pichora-Fuller et al.'s (2016) consensus definition of the same term can be attempted, as "the allocation of cognitive resources to overcome obstacles or challenges to achieving listening-oriented goals" (Francis & Love 2020:3).

2. Cognition and simultaneous interpreting

As discussed in the previous chapter, interpreting under conditions in which sound quality does not meet a minimum standard can lead to long-term negative effects on the interpreters' physical and psychological health. In addition to the repercussions concerning the interpreters' hearing and the development of various hearing disorders and symptoms, poor sound quality can also affect the interpreter in the very moment of producing the interpretation. The increased listening effort may demand more resources from the interpreter and therefore perhaps even affect the quality of the output produced (Gile 2009). AIIC defines cognitive load as the "portion of the interpreter's limited cognitive capacity devoted to performing an interpreting task in a given environment." (AIIC 2020a). While several scholars have developed models to illustrate the cognitive process during simultaneous interpreting, ranging from Lederer (1981) and her eight mental operations, to Setton's cognitive-pragmatic model (1999), which combines a variety of cognitive-scientific research to discuss all relevant aspects of comprehension, memory and production in the interpreting process, this thesis will draw on Daniel Gile's well-known Effort Model of simultaneous interpreting (Gile 2009).

2.1 Gile's Effort Model

Daniel Gile's Effort Model of simultaneous interpreting is based on the assumption that certain mental, non-automatic operations require attention or, as he refers to it, processing capacity, which is only available in limited supply (Gile 2009:159). Gile explains that carrying out non-automatic operations takes processing capacity from this limited supply and that when the processing capacity that is available for any given task is insufficient, performance deteriorates. He applies this to interpreting by stating that the interpreter has a limited quantity of cognitive capacities which they can dedicate to the task of interpreting. Within the interpreting task, there were originally (1985) three so-called "efforts" between which the cognitive capacities needed to be allocated: the 'listening and analysis' (L), 'production' (P) and 'memory' (M) efforts. The naming of the components as "efforts" relates to their effortful nature, as they include deliberate action which involves decisions and requires resources (Gile 2009:160). The basic principle behind the Effort Model was that the sum of the three efforts should not

exceed the overall cognitive or processing capacity available to the interpreter (Gile 1985; Pöchhacker 2016:91). The original formula was therefore:

$(L + P + M) < C_{apacity}$

In a later refinement of his model (1997/2002), Gile added a fourth effort, the so-called 'coordination effort' (C), to represent the effort of having to coordinate the first three efforts. Gile defines the 'listening and analysis' effort as a sum of all comprehension-oriented operations, starting from the subconscious analysis of the sound waves which carry the source-language speech to the interpreter's ears to the identification of the words and finally the understanding of the meaning of the utterance. The 'production' effort refers to the output of the interpreter. In simultaneous interpreting, this effort defines a series of sub-processes, including the mental representation of the message to be delivered, the planning of the speech, the actual speech output as well as self-monitoring and self-correction, if necessary (Gile 2009:163). The 'memory' effort refers to short-term memory operations, which take place one after the other during the interpreting task. These operations result from the time lag between the utterance of the source speech and the output of the interpretation, as the phonetic segments an interpreter hears need to be kept in short-term memory until they can be combined to allow for the identification of a word or phoneme. The short-term memory operations may also be linked to the time necessary to select appropriate words or syntactic structures before producing speech, during which the information or idea that was previously expressed needs to be maintained in memory (Gile 2009:165f.). The relationship between the components of the model can be expressed in a set of formulas as follows (Gile 1997/2002:165):

1. $SI = L + P + M + C$
2. $TR = LR + MR + PR + CR$

The first formula expresses that simultaneous interpreting is the (non-arithmetic) sum of the three main efforts plus a coordination effort. The second formula states that the total processing capacity requirements (TR) are a sum of the individual processing capacity requirements.

3. LA ≥ LR
4. MA ≥ MR
5. PA ≥ PR
6. CA ≥ CR
7. TA ≥ TR

Formulas three to seven mean that the total capacity available to the interpreter for each one of the efforts must be equal to or larger than its requirements for the current task. In other words, a task or effort cannot require more cognitive capacities than are available to the interpreter. This concept is tied to another assumption made by Gile, the “Tightrope Hypothesis” (Gile 1999). According to this assumption, interpreters, particularly during simultaneous interpreting, tend to work at the very limit of either their overall processing capacity or of their individual efforts due to high effort-specific requirements or suboptimal resource allocation to each one of the efforts (Gile 2009:182). If the interpreting task requires a cognitive load higher than what is available to them, performance deteriorates (Gile 2009:158). Gile explains further that without the Tightrope Hypothesis, the natural assumption would be that interpreters have sufficient processing capacity available to cover the requirements for each effort and any mistakes made by them would be due to insufficient linguistic or extralinguistic knowledge as opposed to a result of what he refers to as “chronic cognitive tension between processing capacity supply and demand.” (Gile 2009:182).

Based on the assumption linked to the Tightrope Hypothesis, Gile refers to so-called “problem triggers”, such as proper names, numbers and compound technical terms as well as poor sound, which can lead to “failure sequences” and require special “coping tactics” by the interpreter (Gile 1995). He explains that problem triggers are not always easily identified, as they do not necessarily lead to errors in the interpreter’s output but require a re-allocation of cognitive load towards the different efforts, depending on which effort is affected by the given trigger (Gile 2009:171f.). By implication, this suggests that the interpreter has a limited amount of cognitive resources and constantly works at the very limit of their cognitive ability, which needs to be reallocated according to the difficulties in a given speech. As the cognitive effort necessary during interpreting is, according to Gile, a result of the non-automatic processes taking place in the brain, he emphasises the importance of gradually automating the cognitive operations.

2.2 Degraded sound and short-term memory

Following the previous chapter's outline of mental processes that take place during simultaneous interpreting, this chapter aims to illustrate the consequences of poor audio quality on mental processes by discussing studies that have been carried out in this field.

One of the earliest known studies that tested the effect poor audio quality has on the output of interpreters was David Gerver's experiment in 1974 (Gerver 1974). He tested 12 experienced professionals who interpreted and shadowed short passages of scripted French prose into English, at three different noise levels. Judges then evaluated the interpreters' output by analysing the correctly transmitted words as well as errors, omissions and corrections made by the interpreters. The findings showed that noise had a significant adverse effect on the performance of the interpreting and shadowing tasks. The evaluation of the interpreter's output also showed that higher noise levels led to a poorer rating by the judges involved in Gerver's experiment. The subjects produced more errors and omitted more information in the interpretation than during shadowing at higher noise levels. Gerver also states that the subjects of his experiment seemed to be less able to monitor their own errors at lower signal-noise ratios, deducing that the difficulty in perceiving source language input had led interpreters to have a reduced channel capacity available for the translation and monitoring of their own output (Gerver 1974). In Gile's terms, one could say that interpreters had allocated their cognitive capacities to the listening and analysis effort and therefore had less capacity available for the production effort.

Tommola and Lindholm (1995) carried out a similar experiment, testing eight professionals who interpreted real conference presentations from English into Finnish both with and without the addition of white noise at 5dB. The interpretations were evaluated by two judges in terms of propositional accuracy and showed that poorer sound quality significantly impaired the interpreters' accuracy. While both experiments were carried out over 20 years ago and standards for transmission quality have since been set, the issues of signal quality and noise have re-emerged through the increasing use of teleconferencing and remote interpreting (Pöchhacker 2016:123).

Jonathan Peelle (2018) claims there is a link between acoustic challenge and errors in speech understanding, poor performance on a concurrent secondary task as well as reduced memory for the verbal material. He explains that the observation of

pupils during the processing of degraded speech showed that they would dilate whenever acoustic challenge occurred, indicating an increase in neural activity. According to Peelle (2018:204), acoustic degradation is more than an auditory problem, as it presents listeners with a challenge that requires the reallocation of cognitive resources. During speech comprehension, listeners must quickly match the incoming acoustic stream to stored representations of the words and phonemes in order to successfully extract meaning from an utterance. With acoustic degradation, less information is available to the listener, therefore reducing the quality of speech cues and decreasing the chance for correct comprehension. The listeners must rely to a greater extent on cognitive systems to extract meaning from the acoustically degraded speech than from a clear speech. This so-called listening effort is defined by Pichora-Fuller et al. as “the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task that involves listening” (Pichora-Fuller et al. 2016:92S).

According to Heinrich et al. (2008) listening to degraded sound also affects short-term memory. Even when the speech as a whole is understood, words or syllables that were acoustically degraded will be more difficult to remember. They refer to a number of experiments that were carried out, involving mainly interference related to background noise, people talking or so-called babble, when different people talk simultaneously. Heinrich et al. (2008) attempt an explanation as to why the background noise during a discourse would interfere with the memory abilities and suggest that the noise simply masks the material to be remembered. Both sounds activate the same regions in the brain for sound comprehension and the background noise ‘drowns out’ the relevant information provided by the speech (2008:736). This theory, however, could be refuted if one were to look at Rabbitt’s experiment (1968). It showed that memory for spoken digits in participants with normal hearing was also adversely affected by the presence of noise both when the noise still allowed for recognition of the digits, albeit effortful, and when the noise did not overlap with the spoken digit that was supposed to be remembered (the target) but with the item following the target.
Cousins et al. (2014) replicated Rabbitt’s experiment using word lists. They played 60 lists of seven words each selected at random from the Toronto Word Pool. The length of seven words was used because “medium-length lists permit the most variability in recall strategy” (Cousins et al. 2014:625). The participants were asked to listen to the lists and recall as many of the seven words as possible and in any order. Each list was

prepared under two conditions: the ‘masked condition’ and a ‘control condition’. In the masked condition, one of the seven words was masked by the sound of 20-talker babble, that is the overlapping sound of 20 people simultaneously speaking, at 38dB, which represents a signal-to-noise ratio (SNR) of +2. This is considered to still allow for successful word recognition, although involving perceptual effort. The position of the masked word was varied and chosen at random as being either the 3rd, 4th or 5th word in the list. To reduce the contrast between the one masked word in the list, the remaining words were played with what is considered a low level of babble at 32 dB, resulting in a SNR of +8, which allows for easy audibility. In the control condition list, all words were played at a low-level babble of +8 SNR. Participants did not know which list they would listen to before the experiment. The results showed that the strongest effect on participants’ ability to recall the words was observed when the third word in the list was masked. Cousins et al. (2014) also observed that not only the masked word itself was less likely to be recalled by the participants, but also the word prior to it.

McCoy et al. (2005) refer to the ‘effortfulness hypothesis’ in their work, which denotes that the additional effort that a listener must execute during degraded speech comes at the cost of the processing resources they might otherwise use to encode the speech content in memory. This theory would somewhat align with Gile’s Effort Model. In order to test the hypothesis, they carried out an experiment with two groups of participants, with the same age, education and verbal ability. Both groups only differed in their hearing abilities. One group was composed of people with good hearing and the other with participants that had mild-to-moderate hearing loss. All participants were presented with a list of 15 words and were tested on their ability to remember the final three words of the list. Both groups presented an almost perfect recall of the final three words, but the group of participants with mild-to-moderate hearing loss presented significantly greater difficulty in remembering the nonfinal words of the list (McCoy et al. 2005). The above-mentioned studies provide compelling evidence that acoustic challenge impacts non-acoustic tasks, such as memory for what has correctly been heard. This indicates that domain-general cognitive resources are involved in the comprehension of degraded speech (Peelle 2018:206). Peelle (2018) concludes by stating that listening to degraded speech is challenging and demands additional cognitive resources from the listener to allow for successful comprehension. Effortful listening involves cognitive processes by the listener such as verbal working memory and attention-based performance monitoring. Therefore, acoustic challenge is not

solely an auditory problem, but strongly affects several cognitive operations that are required for both linguistic and non-linguistic tasks.

The involvement of working memory as the primary cognitive resource impacted by degraded speech is also supported by Francis and Love (2020). The authors attempt an explanation as to why working memory suffers during effortful listening by referring to active models by Heald and Nusbaum (2014). According to them, incoming speech signals are turned into mental representations by the listener and then compared against other mental representations of likely linguistic interpretations in a process called hypothesis testing. The result of these comparisons helps to adjust the processing of incoming linguistic signals and the formation of hypotheses as to what the intended linguistic content of a signal may be. Therefore, when a signal is distorted, the allocation of attentional resources needs to be adjusted in order to focus on signal properties that are comprehensible while plausible interpretations are brought from long-term memory into short-term memory storage for comparison (Francis and Love 2020:4). This cyclical and continuous refinement process of reallocating selective attention and creating and evaluating possible hypotheses is what incurs effort in the form of increased cognitive resource demand during effortful listening. When taking into account the impact of degraded sound both on the listening and analysis effort as well as the memory effort, Gile's tightrope hypothesis suggests that because interpreters work close to cognitive saturation, many errors, omissions and/or infelicities (EOIs) may occur. These are not due to a lack of knowledge or linguistic skills by the interpreter, but to insufficient available cognitive resources (Gile 1999; Shao and Chai 2021). Mizuno (2017) explains the very same causality through the concept of Gile's 'imported load'. He explains this with an example of a difficult interpreting task, in which the cognitive load of listening and analysis increases. Due to the increased cognitive demand, the execution of this task or effort takes longer. A result of the delayed listening and analysis effort is that the difficult segment and the subsequent segment uttered by the speaker will stagnate in working memory, causing an increase in memory load. At the same time, the delay of the listening task 'L', would cause a delay in the translation 'T' and production 'P'. If 'T' requires more time due to, for example, a difficulty in retrieving corresponding words or phrases, the memory load 'M' will also increase. If an interpreter must retain an already translated word or phrase in their working memory for some time in order to integrate it with the following part of a sentence due to structural differences of languages, the load of 'M' increases even

further. The accumulated cognitive load in ‘L’ and ‘M’ could lead to interpreting failures, such as errors, omissions and infelicities (EOIs), even if the material itself could be easily processed without the additional load.

3. Factors that impact the interpretation

The difficulty of a given interpreting task can be influenced by a variety of factors. Pöchhacker (2004:122) refers to these influential elements as input variables. He describes them as external factors that impact an interpreter's attention, comprehension and production. The term "input variables" is used as these relate to the nature of the source text, which serves as the input to the interpreter's mental processing operations and therefore affects the interpreter's output. Gile divides what he calls 'cognitive problem triggers' into two categories. One category relates to an overload caused by information density, which can result from fast delivery rates, enumerations or prepared speeches that are read out loud and do not include many pauses or hesitations. This first category also includes external factors such as audio quality, unfamiliar accents, unfamiliar style or argumentation structures, wrong choice of words and other elements that put additional strain on the listening and analysis effort. The second category includes problem triggers in the form of speech segments that do not require much processing capacity but are either short in duration or contain little redundancy. A momentary lack of attention by the interpreter may therefore lead to information loss. This often happens when numbers, short names or acronyms are used (Gile 1997/2002:163-171; 192-194). All of these problem triggers can lead to a deterioration of interpreting quality, but this is not always the case. If a long name is uttered at the end of a sentence and is followed by a prolonged pause, there is sufficient cognitive capacity for the memory and production efforts to process the information at hand. Problem triggers that challenge an interpreter's processing capacity can lead to a deterioration of the output, which can be perceived through errors or omissions as well as affect the presentation of the target text. This can be noticed through the interpreter's voice or intonation (Gile 2009:171f.).

In her master's thesis, Barbara Heinisch-Obermoser (2010) broke the factors that negatively affect an interpretation down into five categories. One category is linguistic factors related to the source text, such as the use of subject-specific terminology, the amount of numbers and enumerations, the use of proper names and acronyms and the complexity of the syntax of the text. The second category is cultural factors, such as the use of quotes or other culture-specific elements in the text. The third category includes speaker-dependent elements, namely their intonation, the

speech rate and accent, among others. Category four refers to environmental factors that play a role, including environmental noise, audio quality, availability of documents, the temperature in the booth, visual access to the speakers, the number of speakers and people in the audience, comfort and technical conditions and many more (Heinisch-Obermoser 2010:56). The fifth and last category are factors that relate to the interpreter. This category takes into account how much previous knowledge the interpreters have, their mental condition, health condition, experience, relationship to team members, skill level, intelligence as well as how much motivation they have, and to what extent they prepared for the interpreting task at hand. In her thesis, Heinisch-Obermoser discusses several factors that can be used to determine how challenging a particular speech is for interpreting. However, not all factors discussed by her are relevant to this thesis.

In her first category, Heinisch-Obermoser discusses terminology, numbers, enumerations, acronyms, syntax, idioms and language register. The author includes terminology in her list, as the use of subject-specific terminology can be challenging for an interpreter. Interpreters are not subject-matter experts and are therefore unable to carry out the necessary word-sense disambiguation of terms they do not know and are more reliant on the context surrounding the term to decipher its meaning (Taylor 1989). Heinisch-Obermoser also explains that numbers pose a challenge to interpreters, because they require cognitive effort to be memorised (Mazza 2000). They cannot be predicted and are usually linked to a wider context, which needs to be stored in working memory alongside the number. Enumerations pose a similar difficulty. They present loose contextless elements of information that can easily be forgotten. They increase the information density of a speech and therefore decrease the available processing capacity of the interpreter (Shlesinger 2003). Similar to numbers, acronyms and abbreviations are speech segments that do not demand much processing capacity but are only short and offer little redundancy (Gile 2009). Proper names, according to Heinisch-Obermoser (2010:38), are a further factor because omitting a name is seen as a grave error in interpreting, as names of any type are considered to be of great importance in conference settings (Taylor 1989). The category of idioms and language register refers to the use of proverbs, idioms, puns, metaphors, idiomatic expressions, colloquial expressions as well as rhetorical questions and often poses a great challenge to interpreters as there are not always equivalents in the target language. While the abovementioned factors are relevant to

determine the difficulty of a source text, they do not apply to the material used in the current thesis. The text used for the experiment does not include any subject-specific terminology, numbers, enumerations, proper names, acronyms or idioms. Therefore, these factors will not be discussed in more detail in this chapter. Instead, the factor syntax, which does apply to the material, will be discussed further in the following subchapter.

In the category of cultural factors, Heinisch-Obermoser (2010) discusses cultural specifics, forms of address and titles and quotes. She explains that certain cultural elements, referred to as realia, do not have equivalents in other cultures. This lack of equivalence may cause problems for the interpreter, who is trying to convey the meaning of this cultural element to the target audience (Weissenhofer 1994:321). Quotes, be they biblical quotes, legal paragraphs or quotes from famous personalities, are part of the professional life of every interpreter and can prove difficult to convey in another language. As the source text does not contain any culture-specific references or quotes, these will not be discussed further. However, the source text does include forms of address and titles and therefore this factor will be described further.

Heinisch-Obermoser's (2010) speaker-dependent factors include prosody, speech rate, accent, nonverbal communication, errors in the source text and slips of the tongue and corrections in the source text. According to Cooper et al. (1982) and AIIC (2002:25), unfamiliar accents are what cause interpreters the greatest amount of source-text-related stress, even more so than high speech rates and reading out pre-scripted texts. Nonverbal communication supports, accompanies and influences verbal utterances. Interpreters need to visually register nonverbal information to gain further information on the meaning of a verbal utterance (Hörmann 1957:311-337). Errors in the source text are a further challenge interpreters must deal with and can come in the form of wrong numbers, names or false facts as well as wrong quotes (Heinisch-Obermoser 2010:54). The factors accent, nonverbal communication and errors in the source text are not relevant for the speech used for the experiment. However, prosody, speech rate and slips of the tongue and corrections in the source text are all relevant and will be described in more detail.

Heinisch-Obermoser's (2010) fourth category of environmental factors includes visual and acoustic input. For the purpose of the current thesis, acoustic input is the most relevant factor and will be described in more detail in this chapter. Visual input is a crucial and broadly researched factor in conference interpreting. It refers to the

interpreter's visual access to the entire conference, the nonverbal communication signals of speakers and the audience, kinesics, turn-taking signals and visual material such as videos, PowerPoint presentations, graphs and other forms of accompanying material (Pöchhacker 2004:126ff.). Visual input is crucial for the interpreter as it facilitates the comprehension of the source message. As the experiment in this thesis allows for visual access to the speakers, this factor will not be discussed further.

The fifth and final category of interpreter-related factors discusses factors such as previous knowledge, the interpreters' health and psychological state, tiredness, level of talent and experience and preparedness for a given task, among many others (Heinisch-Obermoser 2010:57). As the limitations of this thesis do not allow for a physical and mental health check of each interpreter and the participating interpreters are not able to prepare for the interpretation task in advance, this category is also irrelevant in the context of the experiment and will not be analysed further.

The following subchapters will provide a more detailed description of all the input factors that apply to the materials used in this thesis, namely syntax, forms of address and titles, prosody, speech rate, slips of the tongue, corrections in the source text and acoustic input. Heinisch-Obermoser's (2010) categorisation of source-text-related, cultural, speaker-related and environmental factors will be used to divide the input variables in discussion.

3.1 Source-text-related factors

The information content of the source text is considered one of the most challenging variables to measure (Pöchhacker 2004). Measuring the complexity of a source text may be carried out with a variety of approaches. One possible approach consists in analysing the source text on a lexical level, considering factors such as word frequency, subject-specific terminology or lexical variability as well as non-redundant text elements like numbers or names (Gile 1984) and semantic phenomena like false cognates in a given language pair (Shlesinger 2000). While research focussed on the lexical aspects of text difficulty is valuable, source-text complexity cannot be determined solely by difficulties at the lexical level. Other possible approaches used to measure the level of difficulty in a source text are readability measures, such as the

Flesch Reading Ease score or propositional analyses to calculate the information and new concept density in texts.

Tommola and Helevä (1998) carried out an experimental study in which student interpreters worked from English into Finnish with texts of different levels of syntactic complexity. The subjects' output accuracy was measured and showed that higher levels of complexity were linked to lower levels of accuracy. Conversely, a study carried out by Dillinger (1994) with the language combination English-French showed that syntactic variables, that is, clause density and clause embedding, only had a weak effect on the performance of the study's participants. A further possible approach relates to the text type. In Dillinger's experiment, both the professional and untrained subjects showed significantly better performance for narrative passages in comparison to speeches describing procedures. As both speeches had a similar number of words, clauses, propositions and cohesive elements, Dillinger deduced that the findings were a result of the informational structure of the texts and concluded that the narrative text was easier to comprehend due to the sequence of episodes it consisted of, whereas the hierarchical structure of procedures and sub-procedures were more difficult to understand.

According to Minhua Liu and Yu-Hsien Chiu (2011), evaluating a source text's difficulty for interpretation is not an easy undertaking due to the "fleeting nature" (2011:245) of the task. Different factors unrelated to the difficulty of the content to be translated may also interfere with output quality. These include the type of interpreting to be carried out or the extent to which an interpreter is prepared for a given topic. Certain elements of a source text have shown to reduce output quality, such as information density (Barik 1973; Dillinger 1994), syntactic complexity and lexical difficulty (Darò et al 1996; Tommola & Helevä 1998). In their work, Liu and Chiu (2011) use the readability level based on a readability formula, information density as well as new-concept density to measure the difficulty of different source texts. The authors justify their choice of indicators by explaining that the readability level considers both word and sentence length, whereas word length is calculated based on the number of syllables in a word and sentence length can be measured by counting the number of words in a sentence. Both elements indicate the lexical difficulty and the syntactic complexity, as, according to them, words with more syllables tend to be considered more difficult and sentences containing a higher number of words are usually more complex because they contain more subordinate clauses (Liu & Chiu 2011). By

calculating both the average number of syllables per word and the average number of words per sentence, they determine the readability of their texts, using the Flesch Reading Ease Score (Flesch 1948). The formula to calculate the score is:

206.835 – 0.846 X AWL (Average Word Length) – 1.015 X ASL (Average Sentence Length)
(Flesch 1948:225)

The Flesch Reading Ease Score formula was originally created by Rudolf Flesch in 1943 and later revised in 1948. The original formula used McCall-Crabbs *Standard Test Lessons in Reading* (1925) as a criterion. The Flesch Reading Ease Score formula was designed to measure readability for general adults and, according to Klare (1974), gave more attention to abstract words and sentence length, in comparison to other formulas. In 1948, Flesch came to the conclusion that counting affixes in order to measure readability was too time-consuming and that the count of personal references was misleading, which led him to develop two new formulas to replace his previous one. One formula was designed to measure Reading Ease and the other measured Human Interest. While the formula for Human Interest did not gain as much public interest, his Reading Ease formula became one of the most widely used formulas for readability measurement. The revised Reading Ease formula takes into account the average sentence length in words and the average word length in syllables. In order to calculate the average sentence length, the total number of words in a text must be divided by the total number of sentences. For the average number of syllables per word, the total number of syllables of a particular text should be divided by the total number of words. To find the readability score, the average word length must be multiplied by 0.846 and the average sentence length by 1.015. Both values must then be deducted from 206.835, offering the final score. The resulting scores lie between 0 and 100, where the more difficult texts will be on the lower end of the range. According to Rudolf Flesch (1948:230), scores ranging from 0-30 indicate that a text is “very difficult”, a score between 30-50 means “difficult, 50-60 is “fairly difficult”, 60-70 is “standard”, 70-80 is “fairly easy”, 80-90 is “easy” and 90-100 means “very easy”. The Flesch Reading Ease formula is frequently used to measure reading difficulty for adults. Using readability to measure source text difficulty does, however, mean that the information present in a given text is not considered. It has therefore received much

criticism for its limited nature, as it does not include factors such as conceptual complexity, text organisation or a reader's knowledge (Fulcher 1997:501).

In Liu and Chiu's experiment (2011), information density and new-concept density were used alongside the Flesch Reading Ease Score to calculate text complexity. Propositional analysis was used to determine information density, as this method allows for the representation of meaning units in a text. Sentences containing a higher number of words may also be an indicator of information density, as they generally contain more information (Dam 2001:30f.). Lastly, new-concept density, which is also based on propositional analysis, was used by Liu and Chiu (2011) because a higher level of redundancies, that is, repeated information, was suggested as being a factor that facilitates the interpreter's task (Déjean Le Féal 1982). Propositional analysis is used to indicate how many so-called "meaning units" (Liu & Chiu 2011:249) a text contains in the form of propositions. According to Solso (1998:259f.), a proposition is the smallest unit that is meaningful. Typically, a proposition consists of a predicate and one or more arguments. The predicate's function is to specify the relationship among the arguments (Kintsch & van Dijk 1978). To calculate both the informational and new-concept densities, the authors followed Bovair and Kieras' (1985) guidelines. The proportion of the overall number of propositions to the total number of words in the text was calculated to reach a percentage of information density. A higher score indicated that a text was denser in information. Similarly, the number of new arguments was calculated against the number of propositions in each source text to indicate the density of new concepts. Here a higher density also indicated that the text contained a higher number of new concepts and less redundancy.

3.1.1 Linguistic factors

3.1.1.1 Syntax

Sentence structures can either facilitate the understanding of a text or complicate it through the use of complex relative clauses and convoluted sentences. They can also interfere with an interpreter's ability to anticipate what may come next, which also

results in a need for higher processing capacity. Whether syntax is considered a problem trigger or not often depends on the language combination. If source and target language differ syntactically, more resources must be allocated to the short-term memory effort, as the interpreter must retain many pieces of information until they are able to produce a target text. This applies especially to the language combination German-English, as German verbs are often placed at the end of the sentence and the interpreter must wait for the verb to be uttered before being able to start interpreting the sentence into English (Gile 2009:193-195). That is, unless they are able to anticipate the verb that will be used. A further syntax-related challenge is the different use of articles and pronouns between different languages (Jakobson 1959:235ff.). Difficulties in interpreting that result from syntactic structures can either arise due to the complexity of the source speech or due to the strain on short-term memory during the restructuring of a source-text segment into the target language (Setton 1999:102).

It is important to differentiate between spontaneously delivered speeches and texts that are scripted and read aloud. While spoken language may include morphosyntactic errors, written language does not (Crevatin 1989:22). This means that spoken language also includes disfluencies such as filled and unfilled pauses, hesitations, unexpected discontinuity in the middle of a sentence, also referred to as an anacoluthon, as well as self-corrections and syntactical errors (Kalina 1998:137f.; Pöchhacker 2004:105-139). These are all elements of spontaneous speech, which is characterised by its own rhythm. This rhythm includes hesitations combined with periods of fluent speech production. Spontaneous speech does not follow strict syntactical structures or grammatical rules. This leads to higher levels of redundancy and a lower degree of semantic density per unit of information. Bühler (1989:134) explains that many interpreters simplify syntactically complex sentences by omitting adverbs, relative clauses or adjectives. Shlesinger (1989) speaks of the “oral-literate continuum” and indicates five criteria by which a text can be placed within that continuum. The criteria are ‘degree of planning’, ‘shared context’, ‘lexis’, ‘degree of involvement’ and ‘nonverbal features’. The degree of planning refers to the fact that texts which are orally presented, are planned. This criterion includes higher lexical density, such as nominalisations, modifiers and subordinate clauses. Spoken texts are characterised by fragmented syntax and a lower level of cohesion. To understand them, knowledge about the surrounding context is necessary. Another characteristic of spoken texts is that they include more pauses, redundancies, repetitions and

monitoring of information flow. Shared context is the second criterion, as the audience of spoken texts needs to have a better understanding of the context than for written texts, which are more explicit with semantic references. The criterion of lexis refers to the use of colloquial language in spoken texts. The degree of involvement suggests that the level of emphasis on the relationship between the communicating parties is linked to orality. The final criteria are nonverbal features as well as fragmented prosody, which includes hesitations and anacolutha.

Kopczyński (1982:256) offers a categorisation of texts which are usually presented during conferences. One category is non-prepared oral dialogues or monologues, which include toasts and discussions. According to Kopczyński, a speech would be considered a partially prepared monologue with keywords. A welcoming speech, on the other hand, would be a prepared and written monologue which is intended for oral presentation. An example of a read-out monologue prepared in writing and intended for written communication is a resolution or a communiqué.

Pöchhacker (1994:113-176) emphasises that speeches present different levels of orality or literacy. He explains that in order to differentiate impromptu speech from a read text, one must consider the degree of preformulation. Preformulation is characterised by two scales. The first scale indicates to what extent the speaker prepared their speech. The points on the scale are extemporaneous, pre-conceived, recited and read aloud. The second scale refers to written support used by the speaker during their presentation. This can be a memorised outline, a written outline, a fully formulated manuscript or even a printed manuscript. This means that a speaker can speak spontaneously while using a complete speech manuscript for support, but also that they can present a pre-formulated text without written support. It is also common that speakers start a presentation by reading out a written speech but end up summarising certain passages or speaking faster towards the end due to time constraints (Crevatin 1989:22).

3.1.2 Cultural factors

Interpreters act as both language and cultural mediators. Their job is to mitigate possible misunderstandings resulting from cultural barriers by adapting the culturally determined idiosyncrasies of the source text to the culturally determined expectations

of the recipients. Each culture has a specific thinking pattern and only by recognising the specifics of a given pattern are interpreters able to perform well (Kondo et al. 1997:158). Culture does not only refer to traditions, customs and the history of that culture but also art, literature, its people and their collective knowledge and experiences which shape the way they think. This makes culture the foundation of behaviour. Cultural discrepancies can especially be felt when thoughts cannot simply be translated into another language with the same words or images (Crevatin 1989:22).

Luckily for interpreters, international conferences and meetings often involve the exchange of experiences and information within a certain field. While speakers may have different cultural backgrounds, their academic and professional backgrounds will most likely be very similar (Pöchhacker 1994). This is also referred to as “cultural homogeneity of group”, that is, to what extent participants in a particular conference are similar to members of a professional group in regard to their world and professional knowledge.

3.1.2.1 Forms of address and titles

Speech openings can become an issue for interpreters if they are not previously informed of the names of speakers and important members of the audience, including their titles. When it comes to conveying salutations and titles into the target language, cultural conventions play an important role. An example are German titles that often do not exist in other cultures and can therefore not be easily translated. In English, for example, it is uncommon to address someone by their title, whereas in German it would be common. How interpreters cope with these differences depends on their assessment and knowledge of the communicating parties (Pöchhacker 1994:210).

3.2 Speaker-related factors

Speakers are the beginning of the communication chain and their voices, speed of delivery, how they structure their speech and express themselves, their pronunciation and other factors all determine how the interpreter can process the information

provided (Kurz 2005). Speeches that are logical and coherently structured are easier to follow, as are speeches that include pauses and intonation. The segmentation of the speech also determines the time lag between the utterance of the source text and the interpretation into the target language, also referred to as the ear-voice span, which is linked to memory-storage capacity. The longer the span, the more information needs to be retained in working memory. Oftentimes, however, speeches are not coherent nor are they logical, giving the interpreter the additional task to make sense of the "nonsense" they hear.

The speech rate and the length of pauses during the delivery are other factors that determine the interpreter's performance. In addition, unusual linguistic styles or argumentative structures can influence the quality of the output. What can be of help is having a view of the nonverbal communication signals of the speaker that accompany the speech (Gile 2009:193-200; Kalina 1998:117; Pöchhacker 1994:97).

3.2.1 Prosody

Language signals are composed of phonetics and prosody. Phonetics alone are not enough to understand the signal. Interpreters also rely on characteristics such as accent and intonation as well as pauses and delivery speed (Bußmann 1990:559). Prosody, in other words, defines how something is said and helps to structure a speech. It can be divided into four different types of phenomena: tonal, dynamic, durational and mixed phenomena. Tonal phenomena refer to the intonation and pitch used by the speaker, while the volume of speech, the rhythm and the emphasis placed on certain words by pronouncing them faster or slower are considered dynamic phenomena. Durational phenomena include pauses, the speech rate and duration of the speech and, lastly, mixed phenomena refer to the accent of the speaker (Ahrens 2004:77).

One of the earliest attempts to test the impact of prosodically degraded input on the output quality of simultaneous interpreters was carried out by David Gerver (1971). The experiment involved six professional interpreters who were asked to interpret ten short texts from French into English. Half of the source texts had been read on tape with standard prosody and with a speech rate of 100 words per minute. The other half had been recorded with minimal intonation and stress and any pauses of 250 milliseconds or more had been removed from the recording. The experiment showed

that the monotonous half of the source texts resulted in considerably lower scores in accuracy for the interpreters' output. Gerver concluded that prosodic cues, such as intonation, stress and pauses, contributed to the interpreter's segmentation and processing of the source text. Déjean le Féal (1982) confirmed this finding in her own study, where she showed the link between intonation patterns and the perception of input speed.

3.2.2 Speech rate

High speech rates are problematic for interpreters, regardless of the interpreting mode. While in consecutive interpreting this means that notes must be taken faster, simultaneous interpreters reach a certain limit at which they cannot increase the speed of their target-text production (Li 2010). The greatest challenge in this regard is syllable density (Gran &Taylor 1990:246). Shlesinger (2003) found that interpreters produce more complete renderings of the source text when the speech rate was increased, as the time span in which information stored in working memory would get lost, decreased. Using Gile's Effort Model, an increased delivery rate means that more cognitive effort must be dedicated to the listening and analysis of the source text and less capacity is available for target-text production or output monitoring. This means that fast delivery may result in a higher number of errors or omissions in the target text. If speakers deliver their presentation quickly in a language other than their native language, this leads to even greater difficulty for the interpreter (Li 2010). The strategies to cope with fast speakers vary. If possible, interpreters can ask the speaker to slow down. However, this is either not always possible or, even if it is, the speaker might remember to speak at a slower pace for only a short period of time and then return to the original speed. Another strategy is to increase the target-text production rate. As previously mentioned, however, this has its limits. A third and very common strategy is to summarise while interpreting. This is not always possible, especially when speeches are high in information density and contain complex arguments that, if shortened, might lose their logic in the target text. In the worst of cases, if any of the previous strategies do not work, interpreters will stop interpreting under unreasonable working conditions (Li 2010). Fast delivery of speeches is often a problem if speakers are either very nervous or if they have written out their speeches and are reading them

off their notes. Conversely, speeches that are delivered at a very slow pace can take cognitive effort away from the listening and analysis and production efforts but put additional strain on the memory effort. Interpreters must retain information in their short-term memory for longer periods of time, which is also a cognitive burden (Gile 2009:200). Interestingly, the real speech rate is not always equal to the perceived speech rate. Sometimes an interpreter's target text sounds stressed even though the original is presented at an adequate pace. The perceived speech rate can be extremely subjective. The less understanding an interpreter has of a given topic, the more they must actively listen to and process the source speech, often giving the impression that the pace is faster than it actually is (Déjean Le Féal 1980:161).

Pöchhacker (2004) observes that an interpreter's perception of the delivery rate might vary depending on the intonation and pauses used by the speakers. Déjean Le Féal (1982) showed in her dissertation in 1978 that speeches produced with monotonous intonation and short pauses were perceived to be faster and more challenging to interpret compared to speeches which contained more intonation and longer pauses. Déjean Le Féal (1980:161) further elaborates by explaining that the redundancy of a given text determines the subjective perception of the speech rate. She also states that sentence segmentation plays an important role as shorter segments facilitate following the speaker's thoughts, unlike preformulated texts that are read aloud (Déjean Le Féal 1980:166).

While the speed of source-text production is relevant regardless of the mode of interpreting, a stronger focus has been placed on its impact during SI (Pöchhacker 2004:124). An AIIC symposium in 1965 suggested that a speed of 100 to 120 words per minute was considered optimal for simultaneous interpreting, a finding which was later confirmed by Gerver (1969/2002). Gerver's experiment involved ten professional interpreters working from French into English, which was their A language. The results of the experiment showed that when the speech rate surpassed 95 to 120 words per minute, the participants had more errors in their output and showed an increase in their ear-voice span and pausing. Gerver also concluded that there was a limit to the extent to which interpreters were able to accelerate their output rate as the input rate kept increasing. Once interpreters reached this limit, they would produce a steady flow of output "at the expense of an increase in errors and omissions" (Gerver 1969/2002:66).

Pöchhacker (2004) explains the difficulty in measuring input speed. While one can choose between counting either words or syllables per minute, other factors may

interfere with the effect of speech rate, such as the frequency and duration of pauses. This factor is used to calculate the articulation rate by considering both speech bursts and pauses. Pöchhacker (2004) further describes the difficulty in measuring the articulation rate using word count, given the morphological differences between different languages and the various pause criteria which range between 70 and 600 milliseconds. It is therefore difficult to compare different studies carried out in this regard.

3.2.3 Slips of the tongue and corrections in the source text

Slips of the tongue are instances of speakers saying something different from what they intended to say (Söderpalm Talo 1980:81) and are a feature of speech production (Nooteboom 1980:87). Causes for slips of the tongue can be similar sounding words, semantic similarity between two terms and other factors. Oral speech is marked by interruption, new formulations, changes of argumentative structures and syntactic and lexical blends. Hesitations, pauses and expletives are often signs of changes in utterance planning. Slips of the tongue can take different forms, for example through anticipations or preservations as well as blends, omissions and many more (Pöchhacker 1994:134ff.). An example of an anticipation is when a speaker says “alsho share” instead of “also share”. The beginning of a word or a syllable is more prone to errors than the end of a word. Errors can be observed on a segment, sound, syllable, morpheme or lexical level. Many of these errors or their corrections go unnoticed by the speakers, as they are primarily focused on conveying the meaning of their message (Pöchhacker 1994:134ff.).

Slips of the tongue can be divided into corrected and uncorrected slips of the tongue. Many are corrected by the speaker, especially anticipations. However, errors in the source text do not necessarily lead to errors in the interpretation. Target-text errors made by the interpreter are caused by a cognitive overload linked to finding a solution to the erroneous source-text segment. The more complex the error, the higher the cognitive demand for error-free target-text production. Even when the speaker corrects themselves, interpreters must concentrate more, and this can cause errors in the interpretation (Kalina 1998:197f.)

3.3 Environmental factors

Environmental factors consist mainly of the acoustic perception of the source text and the visual access to the meeting taking place. Other factors that fall under this category are the temperature and oxygen levels in the booth, which can also impact an interpreter's performance. The number of participants in a conference that rely on the interpretation, the attitude of the organisers and conference participants towards interpreters and the availability of preparation material also play a role in the quality of the interpretation (Setton 1999:99).

3.3.1 Acoustic input

Background noise and audio quality also fall into the category of external factors. Simultaneous interpreters who are, according to Gile (2009), constantly working at the very limit of their cognitive capacity, can easily be distracted by background noise, either in the conference room or inside the booth, by poor audio quality and even by the sound of their own voice. Although background noise poses a greater challenge to consecutive interpreters than it does to simultaneous interpreters, the nature of their interpreting mode allows them to either change their positioning towards the speaker or to address the issue directly (Kalina 1998:71-73; Pöchhacker 2004:126-128). Environmental noise stemming from whispers in the conference room or noise produced by a booth partner demands a stronger focus on listening and analysis, both for simultaneous and consecutive interpreting (Gile 2009:193). Poor audio quality can also be caused by poor sound insulation of the interpreting booth. It can affect the interpreter's performance, as was shown by Gerver's experiment (1974).

4. Disfluencies

While the terms fluency and disfluency are frequently and freely used in linguistics, authors agree that there is no clear definition of either one of them (Rennert 2019; Tissi 2000). Most people might be able to say whether they perceive the delivery of a speech to be ‘fluent’ or ‘disfluent’; however the criteria on which this perception is based, are in no way consistent. Aguado Padilla (2002), Rennert (2019) and other authors agree that the difficulty in defining fluency is linked to its everyday use and ambiguous meaning. Both ‘fluency’ and ‘disfluency’ carry different meanings according to the disciplines in which they are used. For example, they may refer to a speaker’s command of a foreign language, or just their overall eloquence (Rennert 2019:28). Lennon (2000) distinguishes between ‘higher-order fluency’, referring to the overall command of a language, and ‘lower-order fluency’ in terms of variables such as speech rate, pauses, false starts and hesitations.

4.1 Parameters

In an attempt to objectify and quantify fluency, scholars in the field of interpreting as well as authors in linguistics have resorted to temporal parameters such as pauses, speech rate and hesitation sounds to measure fluency. Even though the number of parameters to be considered differs slightly according to different authors, there is a group of variables which are named in a majority of the literature available. For Hedge (1993:275) “frequent pauses, repetitions and self-corrections” are characteristics of non-fluent speech production by non-native speakers, and Chambers (1997) focuses mainly on pauses and hesitations, including their frequency and length within speech production. Kurz & Pöchhacker (1995) perceive fluency as an interaction of several temporal factors, such as speech rate, hesitation, pauses and false starts. They note, however, that it is unknown to which extent each one of the factors affects the impression of fluency. Tissi (2000:112) uses a detailed analytical method in her study of unfilled pauses and disfluencies in simultaneous interpreting. She proposes the umbrella term ‘non-fluencies’, which are divided into ‘silent pauses’ and ‘disfluencies’. Silent pauses are broken down into grammatical and/or communicative pauses and non-grammatical pauses. Disfluencies are further divided into two subcategories: filled

pauses, which include vocalised hesitations and the lengthening of vowels and consonants, and interruptions, which include repetitions, self-corrections and false starts (see Figure 1).

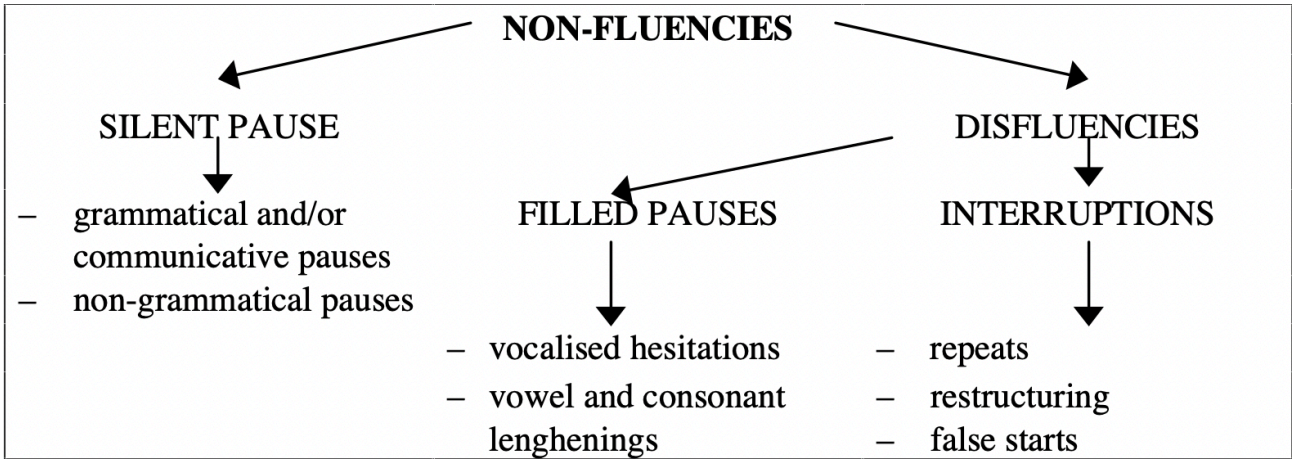

Figure 1: Breakdown of non-fluencies according to Tissi (2000:112)

Sylvi Rennert (2019), on the other hand, focuses on the variables pauses, hesitations, false starts, self-corrections and speech rate to assess the fluency of a speech.

4.2 Pauses

The variable 'pauses' defines interruptions in the flow of speech or "an interruption in the acoustic signal of the uttered sound continuum" (Ahrens 2004:102). Such interruptions can be unfilled pauses (also referred to as silent or mute pauses) or filled pauses, in which hesitation sounds or sound prolongations may interrupt the speech. According to Rennert (2019), such a definition can be problematic, as pauses might be perceived by listeners even when there is no actual interruption of the acoustic signal, for example when sounds are stretched before or as a substitute for a pause (Pompino-Marschall 1995:237). Several authors have set a minimum length for pauses to be considered a result of cognitive overload. These range from 0.10 seconds (Hieke et al. 1983) up to 0.25 seconds (Goldman-Eisler 1968:12). A clear minimum value for pause length is difficult to establish, however, as the threshold at after which a pause is perceived depends on its position. Butcher (1981:205) states that pauses at syntactic positions are only perceived by listeners starting from a length of 220ms, whereas non-syntactic pauses are perceived at 80ms. Pauses made between

intonation or information units delimit these from each other and therefore fulfil an important function in segmenting the flow of speech (Ahrens 2004:104f.; Goldman-Eisler 1968:13). Therefore, Ahrens (2004:159f.) distinguishes between sense-supporting and disruptive pauses according to their position. Pauses that support the meaning of the utterance occur at the end of a phrase or sentence. When a pause is made within a grammatical structure that does not usually include pauses, it is considered disruptive, as it hinders the listener's comprehension process. Filled pauses are usually filled with hesitation sounds ("uh", "hmm", "um") or stretched sounds. They are signs of spontaneous speech planning processes and are mainly used to gain time for speech planning (Arnold & Tanenhaus 2011).

The durational aspect of pauses can also be problematic because pauses can be perceived by the listener when there is in fact no interruption in the speech signal, as is the case when certain sounds, such as occlusives, are articulated. Whereas filled pauses can be observed through hesitation markers or vowel prolongations, unfilled pauses are more difficult to identify. Royé (1983:63f.) suggests that unfilled pauses should be divided into pauses with and without breathing. Ahrens (2004:102), however, disagrees with this distinction, arguing that interpreters can also breathe during filled pauses before or after producing the hesitation sound. Ahrens also argues against Royé's proposal of a third type, namely "combined pauses" (Royé 1983:64), which are pauses consisting partially of silent periods with or without breathing, followed by periods of filled pauses. Ahrens (2004) believes that the combined pauses could simply be considered filled pauses. She further explains that this depends on the transcription method used. If a transcript shows a hesitation marker and additionally notes unfilled pauses before and after the hesitation sound, one could consider that different types of pauses can be observed. As "silent phases" occur before and after hesitation sounds, it would be more logical to classify the entire pause, including the hesitation sound, as a filled pause, as most linguists do (Mead 2002:235).

There is much debate within interpreting studies as to the minimum length of pauses to be considered for analysis. Royé (1983:37) suggests three different categories of pause lengths: short pauses that are up to 0.4 seconds in length, normal pauses between 0.4 and 1.0 seconds, and long pauses, which are more than 1 second long. According to this categorisation, Goldman-Eisler's (1958:64) breathing pauses between 0.5 and 1.0 seconds in length fall under the category of normal pauses. In her experiments, the author only considers pauses that are longer than 0.25 seconds

(Goldman-Eisler 1961:220; 1968:12). Experimental research conducted by Coblenzer & Muhar (1976:109) has shown that pauses shorter than 0.25 seconds only serve as breathing pauses.

4.3 Interruptions

Other types of interruptions of speech are also considered disfluencies, such as repetitions, false starts or self-corrections, referred to as 'restructuring' by Tissi (2000). Repetitions consist of repeating words, syllables or parts of sentences, as long as these do not fulfil any rhetorical function (Tissi 2000:114). Word repetitions can, however, fulfil two different functions: on the one hand, they can serve to delay and thus plan speech; on the other hand, the repetition of the last word uttered before a pause can serve to tie in with what was said before (Hieke 1981:152). False starts occur when a sentence is interrupted and a new one is started without correcting or completing the previous one. Self-corrections include the corrections of slips of the tongue as well as grammatical, structural, content or stylistic errors. Some plan changes, where the structure is changed within a sentence, can also be counted among self-corrections (Pöchhacker 1994:135f.; Tissi 2000:114). In contrast to hesitation sounds, these types of disfluencies are not a sign of gaining time or language planning, but rather of correction of errors or as a result of conscious or unintentional changes of plan or false starts (Rennert 2019:34).

According to Rennert (2019:66), the most detailed quantitative evaluation of prosodic features of SI to date was carried out by Barbara Ahrens (2004), who evaluated three interpretations of the same source speech. In addition to her analysis of prosodic phenomena of intonation and final pitch, Ahrens also analysed the position, number and duration of the pauses in the interpretations in comparison to the original speech.

4.4 Disfluency and cognitive load

In 1961, Goldman-Eisler related the incidence of pauses to the cognitive effort required by the linguistic activity carried out. She demonstrated that the number of pauses diminishes with the progressive automatization of the task. This aligns with Daniel Gile's efforts, which are only considered as such due to their non-automatic nature.

A more recent experiment to measure cognitive load by analysing disfluencies was carried out by Ewa Gumul (2021). While the occurrence of disfluencies in interpreting students was not the main focus of her experiment, its analysis was used to investigate the correlation between explicitation and increased cognitive load in SI by interpreting students. She referred to the analysis of the disfluencies, that is, of the final product of the interpreters, as 'performance method'. Gumul analysed the interpreters' output by observing three types of disfluencies: hesitation markers, false starts and unfilled anomalous pauses exceeding two seconds, which are considered to be problem indicators. These types of infelicities are considered to suggest an increased cognitive load for the interpreter, which may lead to greater cognitive effort on their part (Gumul 2021:52).

The hesitation markers analysed by Gumul consist of filled pauses or fillers, that is, "non-lexical fillers in the form of meaningless strings of prolonged sounds" (Gumul 2021:56). In the experiment, false starts were considered to be cases of word truncation, which happen either when interpreters produce part of a word, usually the first syllable, and then repeat the entire word (e.g., 'si…silly ideas') or when they pronounce the first syllable of a word to then produce another word altogether (e.g., 'si…stupid ideas') (Gilquin 2008). The last category of disfluencies analysed in Gumul's experiment were unfilled silent pauses of more than two seconds in length. According to Pradas Macías, two-second pauses are considered to be the threshold value for anomalous pauses (Pradas Macías 2006). Defrancq et al. (2015) link pauses and hesitations to the cognitive load imposed by possible problem triggers in the source text. These problem triggers result in processing difficulties by the interpreter. According to the authors, disfluencies such as pauses, hesitation markers and false starts are indicators of cognitive difficulty and could suggest cognitive overload and cognitive effort experienced by the interpreter. In the study, Gumul tested 40 advanced interpreting students from two different universities in Poland. All participants

interpreted two source texts (STs) of approximately 20 minutes each, one from English into Polish and one from Polish into English. All participants had Polish as their A and English as their B language. Both interpretations were carried out on two separate days, which were at least one week apart. Both texts had been presented during a medical conference and were thematically similar. Apart from the same subject matter, the STs were also comparable in terms of the degree of orality: all of them appear to have been pre-written texts delivered orally. The source texts were recorded by native-speaker lecturers in laboratory conditions to ensure that the speed of both deliveries would be equivalent, that is, comparable in syllable count per minute. Both STs were also slightly adapted regarding their 'explicitating potential', by including the same number of metaphors in each speech. The results of Gumul's experiment showed that there is a direct link between explicitation shifts and cognitive load. The disfluencies were measured before and after the explicitation shift and results showed that explicitation may be caused by and act as a cause of increased cognitive load, as the number of disfluencies observed was high before and after the explicitation was made. From Gumul's findings, one can also conclude that increased cognitive load leads to an increased occurrence of disfluencies.

5. Methodology

This chapter describes the material and methodology used for an experiment carried out to test the connection between poor audio quality and disfluencies such as those described in the previous chapter. The underlying hypothesis of the experiment is that poor audio quality leads to increased cognitive effort by the interpreters who, in turn, will have less cognitive capacity available to produce or monitor their output, which would therefore show a higher number of disfluencies.

5.1 Research design

The empirical part of the present thesis is inspired by Gerver's experiment (1974), but is not a replication of his method. Instead of testing different levels of sound degradation or using a control and a test group, this experiment involved one group of six participants who interpreted the same texts. The video material used for the study included footage that offered objectively good audio quality and also poor audio quality. All six participants were asked to interpret the same video, meaning that all were subjected to the same conditions. Instead of manipulating the same source text and artificially creating sound degradation, the material chosen for the study was a text that naturally included different levels of sound quality. This decision was made because the participants were to be tested under realistic interpreting conditions instead of creating an artificial interpreting situation. This also meant that other variables will have affected the outcome of the experiment, namely the fact that the source texts of each level of sound quality were different in length, in content, in delivery speed and even, to a certain extent, in difficulty. While the speech rate and difficulty levels of the source texts are further analysed in the following chapter to attempt a comparison between both texts, the difference in length and content between both texts are variables that could not be eliminated and will be taken into account during the analysis of the data in Chapter 6 and the discussion of the findings in Chapter 7.

5.2 Material

A YouTube video from the channel ‘Österreichische Gesellschaft für Europapolitik’ (Austrian Society for European Politics) was chosen for the experiment (Österreichische Gesellschaft für Europapolitik 2022). The video shows an online discussion panel consisting of five people who, over the course of approximately 60 minutes, discuss the possible developments of the conflict between Russia and Ukraine. The video was released on February 2nd, 2022, and therefore the potential of a possible invasion of Ukraine by Russia is still being discussed as a hypothetical development. The invasion of Ukraine by Russian troops took place on February 24th, 2022.

The discussion panel in the video consists of Claudia Major, Head of the Research Group at the German Institute for International and Security Affairs, Gerhard Mangott, Professor at the University of Innsbruck and Head of the Department of International Relations, Walter Feichtinger, Head of the Centre for Strategic Analysis in Vienna, Wolfgang Sporrer, Fellow at the Hertie School of Governance in Berlin, and Paul Schmidt, General Secretary of the Austrian Society for European Politics and moderator of the panel discussion. In order to test the hypothesis, the first eight minutes and fifty seconds of the video were chosen as material for the experiment. This is because the beginning of the video contextualises the discussion to be interpreted and it includes two different speakers. Paul Schmidt, the moderator, introduces the topic and the panel members for approximately five minutes, before addressing Gerhard Mangott with the first question. Mr Mangott then replies to the question posed to him for approximately two and a half minutes. While Paul Schmidt used in-ear headphones that were connected to his computer via cable and could be heard very clearly, Gerhard Mangott decided to participate in the discussion using the microphone and speakers built into his laptop, resulting in poor audio quality. When he speaks, the sound reverberates in the room he is in, and it becomes increasingly difficult to acoustically understand what he is saying. The contrast becomes even more noticeable when Paul Schmidt continues to speak and other guests share their assessment of the situation throughout the online meeting, as all other guests can be heard clearly, although two other guests, namely Walter Feichtinger and Wolfgang Sporrer, are also not wearing a headset, as can be seen in Figure 2.

Figure 2: Illustration of speakers

The main objective of using the first eight minutes and fifty seconds of the video was to allow the interpreters to be introduced to the subject matter of the discussion and to start with a speaker who can be heard well. They then experience the contrast in acoustic quality from Mr Schmidt to Mr Mangott. It is expected that a higher number of disfluencies appear during the interpretation of Mr Mangott's speech, which might be linked to the audio quality of the speaker. The disfluencies can, however, also be linked to the content of the speech itself, as the content of what is said differs between Mr Schmidt and Mr Mangott. Whereas Mr Schmidt introduces the subject matter and offers an overview of the situation, including the developments that led to a point where a Russian invasion of Ukraine is a possible scenario, Mr Mangott offers an expert's view on the matter. He analyses political strategies and positions as well as possible future developments depending on which side – the West and Russia are placed in juxtaposition in his analysis – decides to give in to the other side's demands first. The following subchapter will address other factors that could have affected the interpreters' performance as described in Chapter 3 and apply them to the source text used for the experiment of the current thesis.

5.2.1 Source-text difficulty

As described in Chapter 3, several factors can impact an interpreter's performance. In order to test whether one variable, namely audio quality, has a particular influence on

the interpreters' output, other variables must be either eliminated or taken into account in the analysis of the source and target texts. As the experiment for this thesis was not carried out in a lab with controlled variables, but rather in a natural setting, other elements relating to the content as well as the delivery of the source text must be discussed in order to pinpoint the role of audio quality in the outcome of the experiment, which is analysed later in this thesis. In reference to the linguistic aspects of the source text, it must be mentioned that the video excerpt did not contain any subject-specific vocabulary that would need to be prepared in advance. The terminology used was familiar to the interpreters as the topic of the discussion was very present in the media and news and similar videos revolving around the same topic had been previously interpreted by the participants. The terminology is therefore assumed not to have been a challenge to the interpreters. While the interpreters were not given time to prepare the topic ahead of the interpretation, they were given a list of all the relevant speakers of the video. As the video used for the experiment is available on YouTube, the participants were allowed to read the description below the video and write down all the names of participants of the discussion panel, as Paul Schmidt, the moderator, does name all the guests and their titles and institution names at the beginning of the video. By allowing the participants to note down all personal and institution names, the additional difficulty of an enumeration, proper names and forms of address in the source text was eliminated.

In the material used for the experiment, both speakers are native German speakers from Austria, albeit from different regions of the country. Neither speaker had a pronounced accent while speaking German. Taking into account that all interpreters who participated in the experiment are Austrian and therefore familiar with regional varieties of the language, it can be assumed that the accents would not have been a factor that affected the interpretation in the context of this experiment. Nonverbal communication elements were also not crucial for the understanding of the text. No visual elements or supporting material were used by any of the participating speakers and they could all be seen from their head down, allowing for facial expressions and, to some extent, gestures, to be captured by the camera. A further element that must be taken into consideration is the content of the source text and the speed at which it was delivered. For the purpose of this thesis, the speeches of Paul Schmidt and Gerhard Mangott are seen as two separate source texts which are to be compared with one another in terms of readability, information density and new-concept density.

Adopting Liu & Chiu's (2011) approach, propositional analysis was used to calculate information and new-concept density, and the Flesch Reading Ease Score served to measure the readability of the source text used for the experiment. Additionally, the average speech rate in words per minute was calculated for both speakers. During the transcription of the source text, it became evident that Paul Schmidt paused more frequently during his speech, which might also impact the interpretation. Consequently, the articulation rate, that is, the number of words per minute excluding all silent pauses, was calculated for both speeches (Ahrens 2004:101; Pöchhacker 1994:131). By calculating and comparing both texts in terms of speech rate, articulation rate and level of difficulty, this thesis aims to assess whether any of these factors could have significantly influenced the quality of the interpretation or, more precisely, the occurrence of disfluencies. Should both texts show a considerable difference in speech and/or text difficulty, one could assume that audio quality was not the only or even the dominant factor that resulted in the disfluencies analysed in Chapter 6. If, however, the values for these variables are similar between both texts, it is more likely that the disfluencies will primarily have resulted from the lack of audio quality during Gerhard Mangott's speech.

Table 1: Information density and new-concept density of source material

Speaker	Number of words	Number of propositions	Information density (%)	Number of new arguments	New-concept density (%)
Paul Schmidt	665	228	34.3	170	75
Gerhard Mangott	450	162	36.0	133	82

Table 1 shows the number of propositions and new arguments per text as well as the values for information and new-concept density. As can be seen, both speeches are comparable in information density, with a difference of less than 2%. Although the difference in new concept density between the speeches is seven percent, indicating that Gerhard Mangott's speech presents more new arguments and is less repetitive than Paul Schmidt's, this discrepancy does not necessarily mean that Mr Mangott's speech is much more difficult to interpret. It does, however, indicate that his speech should be slightly more difficult due to a lower degree of redundancy.

Table 2: Readability of source material in Flesch Reading Ease Score

Speaker	Word length	Sentence length	Reading Ease Score	Difficulty Level
Paul Schmidt	2.00	15.11	21.80	Very difficult
Gerhard Mangott	2.03	18.00	16.57	Very difficult

Table 2 shows the Flesch Reading Ease scores of both texts based on the average number and length of words present in each of them. Gerhard Mangott's speech shows a lower score, indicating that his speech should be slightly more difficult to interpret. However, both speeches received a score below 30 and are therefore both considered to be "very difficult", meaning they still fall into the same category and might therefore not differ considerably.

Table 3: Speech and articulation rates

	Speech rate (words per minute)	Articulation rate (words per minute)
Paul Schmidt	123.5	134.4
Gerhard Mangott	128.5	138.5

Table 3 shows the speech and articulation rates of both speakers, measured in words per minute, and indicates that Gerhard Mangott spoke slightly faster than Paul Schmidt. As found by Gerver (1969/2002), the ideal speech rate for interpreting is 100 to 120 English words per minute. Above this rate, source-text production is considered to be above optimum speed. The minimal difference between Mr Schmidt's and Mr Mangott's speech rates is therefore most likely not a decisive factor when it comes to assessing the difficulty in interpreting both speakers. The values for the articulation rate of both speakers show that, on average, Gerhard Mangott had a faster articulation rate. The pause analysis of the source text showed that Mr Mangott paused less frequently while delivering his speech in comparison to Mr Schmidt. The articulation rate reflects what the average speech rate also shows, which is that Mr Mangott spoke faster overall.

The calculations analysing source material difficulty, readability scores and the speech and articulation rates of the speakers show that Gerhard Mangott's speech is overall more difficult to interpret when compared to Paul Schmidt's speech, regardless of any additional audio quality issues. Nevertheless, most differences found do not

seem substantial enough to conclude that any disfluencies produced during the interpretations of both speeches would primarily result from the increased difficulty of Mr Mangott's speech, rather than caused by the poor audio quality while he speaks.

A final consideration to be taken into account is the content of the speeches to be interpreted. The online Zoom conference takes place before February 24th, 2022. The conversation consists of a discussion as to how likely an invasion of Ukraine by Russia is. The interpreters of the experiment, however, interpreted the video after February 24th, 2022, at which point the invasion had already taken place. Considering as well, that arguments against the possibility of an invasion are brought up, the interpreters had to interpret speculations concerning a situation whose outcome they already knew. In other words, they heard arguments they knew to have proved false, creating a phenomenon originally described by Leon Festinger (1957), namely, "cognitive dissonance". According to his theory, if an individual were to hold two or more elements of knowledge which are mutually relevant but inconsistent with one another, an unpleasant state of discomfort is created, resulting in so-called "dissonance". Festinger (1957) also theorised that the same individual would be motivated by this state of mental discomfort to engage in "psychological work" in an attempt to reduce the inconsistency. This psychological work will typically support the cognition most resistant to change. In the case of this experiment, the knowledge that the invasion took place would be more resistant to change than the speculation that an invasion was unlikely.

5.3 Participants

The experiment involved six student interpreters from the Centre for Translation Studies at the University of Vienna. The participants all had German as their A language and English as their B language and were enrolled in the MA in Translation and Interpreting with a focus on conference interpreting. They were all in their fourth semester of the MA curriculum and therefore had similar levels of experience and interpreting skills. None of the participants had passed their final interpreting exams at the time of the experiment and all had taken several simultaneous interpreting classes

throughout their studies. They were also all enrolled in “Simultaneous Interpreting II English”, a course aimed at more advanced students.

5.4 Procedure

The purpose of the experiment was to observe whether the interpreting students produced a higher number of disfluencies while interpreting Gerhard Mangott, as the audio quality during his speech was poor because he was not using headphones with an external microphone. By carrying out this experiment, the thesis aimed to show how poor audio quality can negatively affect the interpreter’s output and lead to a target text which contains a higher number of disfluencies.

The participants of the study were asked to simultaneously interpret an excerpt from the YouTube video from German into English. They were not given any material prior to their interpretation. They did, however, have five minutes before the start of the task to read the description below the YouTube video to understand in what context the conversation was taking place and to note down any important names of participants and the institutions at which they worked. The aim of the study was to find disfluencies in the interpreters’ output linked to the audio quality and not caused by difficulties relating to names and titles. The interpretation was also carried out within the context of a simultaneous interpreting class in which similar videos and similar topics had been used as practice material before.

The participating students had all previously worked with poor audio quality, as videos containing poor sound quality had been used during previous classes. All the participating students also started their master’s degree during the Covid-19 pandemic and had experienced three online semesters, during which they had to interpret via Zoom lectures. They had all also completed the course “Konferenzsimulation” (Mock Conference), in which students are required, among other things, to participate in an online mock conference and to interpret using an RSI platform. This means that they were all somewhat familiar with RSI. The interpreters were not given any time to listen to the video before starting to interpret but were asked to start from the beginning, which was approximately two minutes into the video and had to interpret for a duration of eight minutes and fifty seconds. For the course in which the experiment was carried out (“Simultaneous Interpreting II English”), all students were required to be able to

interpret for a minimum of ten minutes at the beginning of the semester and a minimum of twenty minutes by the end, meaning that the duration of the video used for this experiment was well within the students' skill range. Since the experiment was carried out in a lecture hall at the Centre for Translation Studies, the students were able to perform the task inside conference interpreting booths. Inside the booths, they had a monitor that displayed the video, over-ear headphones and a microphone attached to a control panel. Through the control panel, the interpreters were able to adjust the volume of the video as well as the bass and the treble. The experiment was carried out on May 13th, 2022, at the Centre for Translation Studies. The participants were not told which aspect of their output would be analysed for the thesis. They were only informed about the topic of the video and that it contained poor audio quality. The participants were instructed to interpret from the beginning of the discussion for approximately nine minutes and record themselves, using their own smartphone devices, which include a speech recording function.

5.5 Data analysis

Following the experiment, the recordings of all participants were collected and transcribed. During the transcription, filled and unfilled pauses and hesitations were annotated for further analysis. As defined in Chapter 4.2, filled pauses are characterised by hesitation sounds or sound prolongations, and unfilled or silent pauses can be identified when there is a disruption to the acoustic signal of uttered sound. No minimum pause length was used for silent pauses. Instead, all pauses that were perceived as being disruptive throughout the interpretation and that indicated cognitive distress were noted in the transcript. To measure the length of the unfilled pauses, the recordings were uploaded into the computer programme Audacity. The programme displays a spectrogram which clearly shows where interpreters were not producing any sound, as the soundwaves displayed flattened at the frequency level 0 (Figure 3).

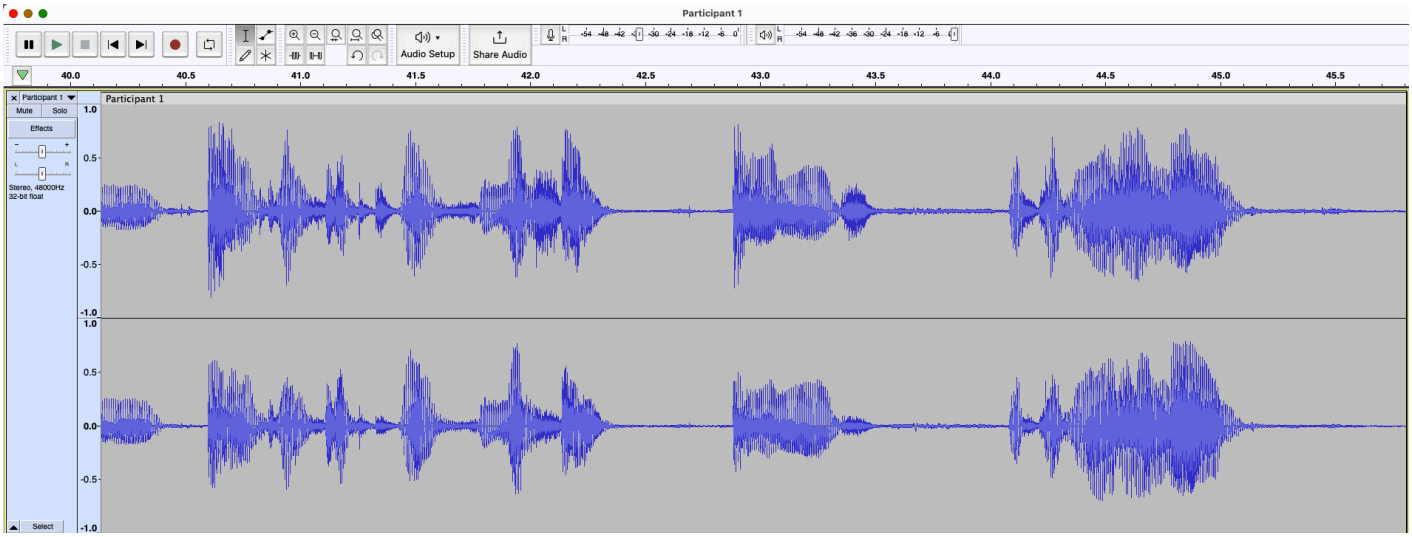

Figure 3: *Spectrogram as shown on Audacity*

Using the transcripts, which contained an annotation of all hesitation pauses, the length of each hesitation pause was measured in Audacity and noted in the transcript. Where interpreters paused before and after hesitation sounds or sound prolongations, in other words, where unfilled pauses preceded and followed a filled pause, they were measured as separate pauses if they were perceived to be disruptive.

The length of the pauses varied but was no shorter than 0.23 seconds. Any initial pauses, that is, the silence before target-text production started, were not considered, as SI requires a certain time lag between the utterance of the speaker and the start of the interpretation, which can therefore not be considered a non-fluency (Tissi 2000:113). Further, any pauses used for prosodic purposes were also not considered in the analysis. Pauses used by the interpreters for breath intake were also not considered. In the transcripts, pauses were displayed within square brackets, noted down in seconds and rounded up to the nearest hundredth. For the analysis, the source text was also transcribed. While other disfluency markers have not been annotated in the source-text transcript, pauses were measured and annotated. This is to enable a more detailed analysis of the causes of silent pauses in the target text. Prolonged pauses in the source text might have led to pauses in target-text production, given that the interpreter was waiting for further input before continuing.

In the target-text transcripts, hesitation sounds such as “uhm” and “uh” were included to show where interpreters produced them. Prolonged syllables were indicated with the “=” symbol, based on Barbara Ahren’s (2004) transcription method. Following the transcription of the six recorded interpretations, including the different

types of disfluencies, a quantitative analysis was carried out. The aim was to determine the average length and total number of pauses, and the frequency of other disfluency indicators for each participant during the different audio conditions. The hypothesis was that the interpreters would present a higher number of disfluencies and longer pauses on average during the excerpt with poor audio quality, compared to the passage in which the speaker can be heard clearly.

6. Results

The results obtained from data analysis are described in this chapter. For easier reference, Paul Schmidt's source text will be referred to as Speech 1 and Gerhard Mangott's speech as Speech 2 and the speakers will be referred to as Speaker 1 and Speaker 2 as well as by their names. It is to be understood that Speech 1 is the video excerpt during which the audio quality was considered to be good, and Speech 2 was the speech with poor audio quality. For the illustration of the results, the pause analysis will be displayed in one table and the remaining disfluency phenomena in a separate table. This is because pauses were analysed with regard to their frequency, their overall and average length and their length relative to the duration of the interpretation. For other disfluency phenomena, only the frequency was counted.

In order to calculate the relative pause length, the overall duration of the interpretations was measured for each text as well as the sum of the length of all pauses during the interpretation. The relative length shows, in per cent, how much of the interpretation consisted of unfilled pauses. The second table displays the number of hesitation sounds, of vowel and consonant prolongations, referred to as prolonged sounds in the table, the number of false starts, of repeats and of self-corrections. To refer to repairs (self-corrections), the term "structure shifts" (Pöchhacker 1995) is used in the sense of Tissi's (2000) term restructuring. The presentation of the results in tables should facilitate the comparison of the values for each text. The average pause length is displayed up to the second decimal, as research focused on pause analysis often measures pauses up to two decimals (Goldman-Eisler 1968). The percentage of relative pause length will be displayed up to the first decimal as it is uncommon to display more than one decimal in percentages.

6.1 Participant 1

Table 4: Pause analysis Participant 1

Speech	Number of pauses	Mean pause length (in seconds)	Relative pause length (in %)
Speech 1	75	0.74	17.4
Speech 2	66	0.82	27.4

Table 5: Disfluencies Participant 1

Speech	Hesitation sounds	Prolonged sounds	False starts	Repeats	Structure shifts
Speech 1	6	14	1	2	1
Speech 2	3	11	0	0	3

As can be seen in Tables 4 and 5, the interpreter produced more pauses in their rendition of Speech 1 compared to Speech 2. However, it should be considered that the length of Paul Schmidt's source text is approximately 100 seconds longer than that of Gerhard Mangott. This difference can be seen in the column illustrating the pause length relative to speech length. Although the interpretation of Speech 1 had more pauses, it had a share of 17.4% of pauses, whereas the interpretation of Speech 2 had a share of 27.4% of pauses, considerably more. Additionally, the average length of the pauses produced while interpreting Speech 2 is slightly higher, with 0.82 seconds in Speech 2 compared to 0.74 seconds in Speech 1. Summarising the results presented in Table 4, Participant 1 paused more often during their interpretation of Speech 1, but the pauses for Speech 2 were longer and therefore make up a larger proportion of the interpretation. When considering the difference in speech length, Participant 1, however, did pause more often during Speech 2.

In addition to pauses, other types of disfluencies were also observed in the transcript of the participant's rendition of the speeches, such as vocalised hesitations. Where silent pauses were observed before or after the utterance of a hesitation sound, they were measured separately and included in the pause analysis above. Participant

1 produced six hesitation sounds during their rendition of Paul Schmidt's speech and three during Gerhard Mangott's speech. Even if the length of both texts were to be taken into account, the interpreter produced more hesitation sounds during Speech 1.

Further, disfluencies in the form of vowel and consonant lengthenings as well as false starts, repetitions and self-corrections were observed in the target text. The frequency of vowel and consonant lengthenings was minimally higher during Mr Schmidt's speech, with 14 lengthened sounds overall, compared to 11 during Mr Mangott's speech. The interpreter produced one false start while interpreting Speaker 1 and none while interpreting Speaker 2. Participant 1 also presented two repetitions in the interpretation of the first speaker and none during the interpretation of Speech 2. The one type of disfluency that was observably more frequent during the rendition of the poor audio quality passage was that of structure shifts. In this aspect, the transcript shows three self-corrections in the rendition of Speech 2 and only one in the rendition of Speech 1.

6.2 Participant 2

Table 6: Pause analysis Participant 2

Speech	Number of pauses	Average pause length (in seconds)	Relative pause length (in %)
Speech 1	46	0.86	12.7
Speech 2	49	1.45	35.6

Table 7: Disfluencies Participant 2

Speech	Hesitation sounds	Prolonged sounds	False starts	Repeats	Structure shifts
Speech 1	14	20	2	4	7
Speech 2	7	22	5	2	1

As can be seen in Table 6, Participant 2 produced 46 pauses during their interpretation of Speech 1 and 49 pauses for Speech 2. Considering that Speech 1 is longer than Speech 2, the results show that many more pauses were produced for Speech 2. In the case of this participant, the frequency, the average length and the relative length of pauses were all considerably higher for Speech 2 than they were for Speech 1. The pauses analysed for the interpretation of Speech 2 account for almost 36% of the overall interpretation of the speech, which is almost three times more than the pauses analysed in Speech 1. While the average pause length for Speech 2 is 1.45 seconds, it must be noted that there were passages in the interpretation during which the interpreter started a sentence, did not complete the sentence and, following a long period of silence, started a new sentence. This resulted in a few pauses of more than five seconds in length, which considerably affected the average pause length and the pause length relative to speech length.

The other disfluency phenomena analysed for Participant 2 show a similar tendency, with the exception of hesitation sounds. Interestingly, Participant 2 resorted to hesitation sounds more frequently during Paul Schmidt's source text, with 14 hesitation sounds observed, compared to only seven during Mr Mangott's speech.

During their interpretation of Speech 1, the interpreter presented 20 cases of vowel or consonant lengthening, compared to 22 while interpreting Speech 2. The prolonged sounds were often followed by either hesitation sounds or periods of unfilled pauses. The false starts for Speech 2 consist of incomplete sentences, followed by long pauses. The interpreter would then omit large parts of the source text and continue interpreting as soon as the speaker started his next sentence. The occurrence of repetitions was equal for both texts, with two cases observed for each. During Mr Schmidt's speech, Participant 2 had to restructure their sentences seven times and re-start the sentence twice, whereas, for the poor audio quality speech, they restructured their sentence once, but restarted their sentence five times.

6.3 Participant 3

Table 8: Pause analysis Participant 3

Speech	Number of pauses	Average pause length (in seconds)	Relative pause length (in %)
Speech 1	64	0.90	17.4
Speech 2	47	1.34	30.5

Table 9: Disfluencies Participant 3

Speech	Hesitation sounds	Prolonged sounds	False starts	Repeats	Structure shifts
Speech 1	4	29	1	0	10
Speech 1	4	14	5	2	3

The third participant showed similar results to the previous two interpreters. During their interpretation of Speech 1, 64 disruptive pauses could be observed. The average pause length was 0.9 seconds, meaning that during the 331 seconds of the student's interpretation, the pauses made up 17.4%. The participant's interpretation of Speech 2 was 206 seconds long and contained 47 unfilled pauses. Similarly to Participant 2, Participant 3 produced some pauses of more than four seconds in length while interpreting Gerhard Mangott, which affected the average pause length and the relative pause length, resulting in values of 1.34 seconds and 30.5%, respectively. Shortly after Paul Schmidt asks Gerhard Mangott a follow-up question as to whether a compromise can be found for the conflict between Russia and Ukraine, and the speakers change, the interpreter continues interpreting Gerhard Mangott's speech but seems to have difficulties following the source text. This results in a silent pause of 11.52 seconds, which is much longer than all other pauses produced by the interpreter, which range from 0.23 seconds to 2.98 seconds. This one prolonged unfilled pause affects the overall average pause length significantly. The long pause leads to the omission of a

large part of Mr Mangott's speech and Participant 3 continues their interpretation without including the information missed during the 11 seconds of silence.

The interpreter did not voice many hesitation sounds in either one of the interpretations, producing four sounds in each text. It is interesting to observe that Participant 3 produced noticeably more vowel and consonant lengthenings during their interpretation of the speech that contained good audio quality. In fact, their interpretation of Speech 1 contained more than twice as many prolonged sounds as their interpretation of Speech 2. Equally, the number of self-corrections in the interpretation of Speech 1 is more than three times as high as the number of the same disfluency observed in their rendition of Speech 2. This indicates a difficulty with Paul Schmidt's source text. The results for false starts and repetitions, however, indicate the opposite. Participant 3 produced more false starts and repetitions for Speech 2 than for Speech 1. The target text for Speech 2 included five false starts and two repetitions, whereas the output for Speech 1 only had one false start and no repetitions at all.

6.4 Participant 4

Table 10: Pause analysis Participant 4

Speech	Number of pauses	Average pause length (in seconds)	Relative pause length (in %)
Speech 1	55	0.78	13.3
Speech 2	38	0.84	15.5

Table 11: Disfluencies Participant 4

Speech	Hesitation sounds	Prolonged sounds	False starts	Repeats	Structure shifts
Speech 1	7	12	1	4	4
Speech 2	3	7	1	4	5

Tables 10 and 11 display the results for Participant 4. The interpreter produced 55 silent pauses during their rendition of Speech 1 and only 38 pauses during their rendition of Speech 2. The interpretation of Speech 1 is approximately 300 seconds long and Speech 2 is approximately 200 seconds long, meaning that Speech 1 is one-third longer than Speech 2. While interpreting Speech 2, Participant 4 produced 38 pauses within 200 seconds. If they were to continue producing pauses at the same intervals for another 100 seconds, it can be estimated that Participant 4 would have produced 58 pauses within 300 seconds. In other words, the interpreter almost produced the same number of pauses for both texts. The average pause length and the relative pause lengths only differ marginally, with slightly higher values for Speech 2 in both cases. This distinguishes Participant 4 from the previously analysed interpreting students, as all other participants presented much more and much longer pauses for Speech 2 than they did for Speech 1.

Regarding the other types of disfluency shown in Table 11, Participant 4 shows either higher or equal numbers for almost all types of disfluencies while interpreting Speech 1. While interpreting Paul Schmidt and being subjected to better sound quality, Participant 4 produced seven hesitation sounds and 12 sound prolongations. During their interpretation of Speech 2, both disfluencies were only observed seven and three times, respectively. For both false starts and repeats, the values are equal for both speeches and self-corrections present the only exception, with one more self-correction in the interpretation of Speech 2 than there is in the rendition of Speech 1.

6.5 Participant 5

Table 12: Pause analysis Participant 5

Speech	Number of pauses	Average pause length (in seconds)	Relative pause length (in %)
Speech 1	52	0.98	16.1
Speech 2	51	1.06	23.6

Table 13: Disfluencies Participant 5

Speech	Hesitation sounds	Prolonged sounds	False starts	Repeats	Structure shifts
Speech 1	3	18	2	8	8
Speech 2	3	14	0	5	7

The results for Participant 5 are displayed in Tables 12 and 13. As was seen for Participants 1, 2 and 3, the pause phenomena indicate greater difficulty in the interpretation of Speech 2. During their rendition into English, Participant 5 produced almost the same number of pauses for Speech 1 as they did for Speech 2. Considering, again, that Speech 1 is longer than Speech 2, this means that the interpretation of Speech 2 contains relatively more pauses than that of Speech 1. Contrary to the findings for other participants, this interpreting student did not produce longer pauses for one text than for the other. The average pause length for both texts differs only minimally, with an average of 0.98 seconds for Speech 1 and 1.06 seconds for Speech 2. All pauses produced by Participant 5 were approximately equal in length. Therefore, no exceptionally long pauses affected the average length. As the number of pauses and the average pause length are relatively higher for Speech 2, the pause length relative to speech length for Speech 2 is considerably higher than it is for Speech 1, with 23.6% compared to 16.1%, respectively.

With the exception of hesitation sounds, which can be observed three times in both interpretations, all other disfluencies were more frequent in the English rendition of Speech 1. Prolonged sounds, repeats and self-corrections were all found in both interpretations, but more frequent in the interpretation of Paul Schmidt's speech. However, the difference in frequency is not substantial. Participant 5 produced 18 sound prolongations, eight repetitions and eight structure shifts while interpreting Paul Schmidt and 14 sound prolongations, five repetitions and seven structure shifts during Gerhard Mangott's speech. False starts, however, were not observed at all during the rendition of Speech 2 but were produced while interpreting Speaker 1.

6.6 Participant 6

Table 14: Pause analysis Participant 6

Speech	Number of pauses	Average pause length (in seconds)	Relative pause length (in %)
Speech 1	42	0.78	9.7
Speech 2	36	1.26	21.2

Table 15: Disfluencies Participant 6

Speech	Hesitation sounds	Prolonged sounds	False starts	Repeats	Structure shifts
Speech 1	5	11	0	2	5
Speech 2	5	22	2	2	3

As shown in Tables 14 and 15, the number of pauses produced by Participant 6 for both texts was very similar, with 42 pauses in the interpretation of Speech 1 and 36 pauses for Speech 2. Considering the difference in length, Participant 6 produced more pauses within the time frame of Speech 2. The pauses were also longer, leading to an average length of 1.26 seconds for Speech 2, considerably more than the average of 0.78 seconds observed for Speech 1. The length and frequency of the pauses resulted in an interpretation with a share of 21.2% of pauses, more than twice as much as observed for Speech 1, where silent pauses made up 9.7% of the interpretation.

The findings for other types of disfluency show that Participant 6 produced 22 prolonged sounds while interpreting Speaker 2, compared to only 11 while interpreting Speaker 1. In their rendition of Speech 2, the interpreter also produced two false starts, repeated themselves twice, corrected three sentences and produced five hesitation sounds. While interpreting Speech 1, they had no false starts at all, two repeats, five self-corrections and five hesitation sounds. This means that repetitions and hesitation sounds were observed in equal measure in both texts and self-corrections even slightly more frequently during the interpretation of Speech 1.

6.7 Aggregate results

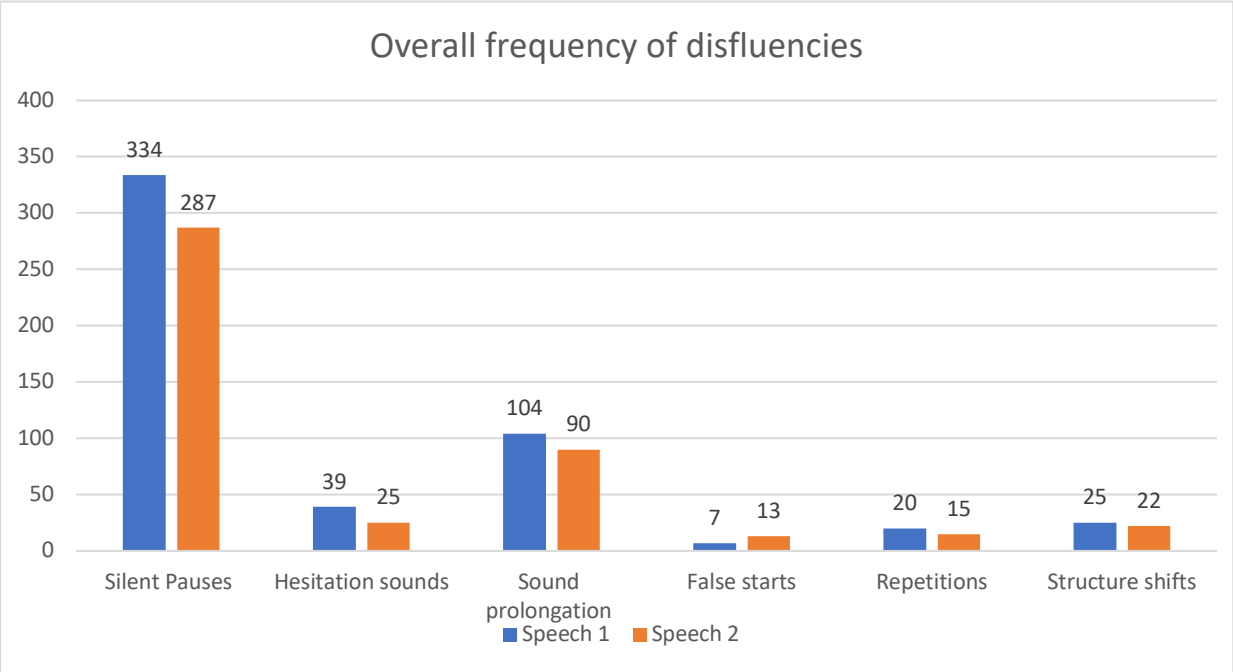

Figure 4: Aggregate results disfluencies

Figure 4 offers an overview of the aggregate results obtained in the experiment. The most frequent type of disfluency analysed during the interpretations was silent pauses. Overall, all six participants produced 334 pauses while interpreting Speech 1 and 287 while interpreting Speech 2. The second most common type of disfluency was sound prolongations, which occurred 104 times while interpreting Speaker 1 and 90 times during the interpretation of Speaker 2. In third place are hesitation sounds, which were observed 39 and 25 times in Speeches 1 and 2, respectively. Structure shifts, which were the fifth most frequent type of disfluency observed, were noted 25 times in the English rendition of Speech 1 and 22 times for Speech 2. They are closely followed by repetitions, which were produced 20 times while interpreting Paul Schmidt and 15 times while interpreting Gerhard Mangott. The least frequent type of disfluency was false starts. These were observed only seven times in total across all six interpretations of Speech 1 and 13 times in all interpretations of Speech 2.

Since silent pauses, hesitation sounds and sound prolongations are all considered pauses (Tissi 2000), it can be said that pauses were the most frequent type of disfluency observed and unfilled pauses were more frequent than filled pauses. Disfluencies categorised as interruptions, namely repetitions, structure shifts and false

starts were, in comparison, less frequently observed. While the bar chart shows that for all types of disfluency, with the exception of false starts, the interpretations for Speech 1 included more disfluencies than the interpretations of Speech 2, it should be reiterated that Speech 1 and 2 differ in length and this trend is to be expected. Speech 1 was 323 seconds long and Speech 2 was 204 seconds long, meaning that Speech 1 is approximately one-third longer than Speech 2. In order to compensate for this difference, the results found for Speech 2 were recalculated for the same length and illustrated in Figure 5. For the calculations, it was assumed that the interpreters would continue producing disfluencies at the same rate for the remainder of the time.

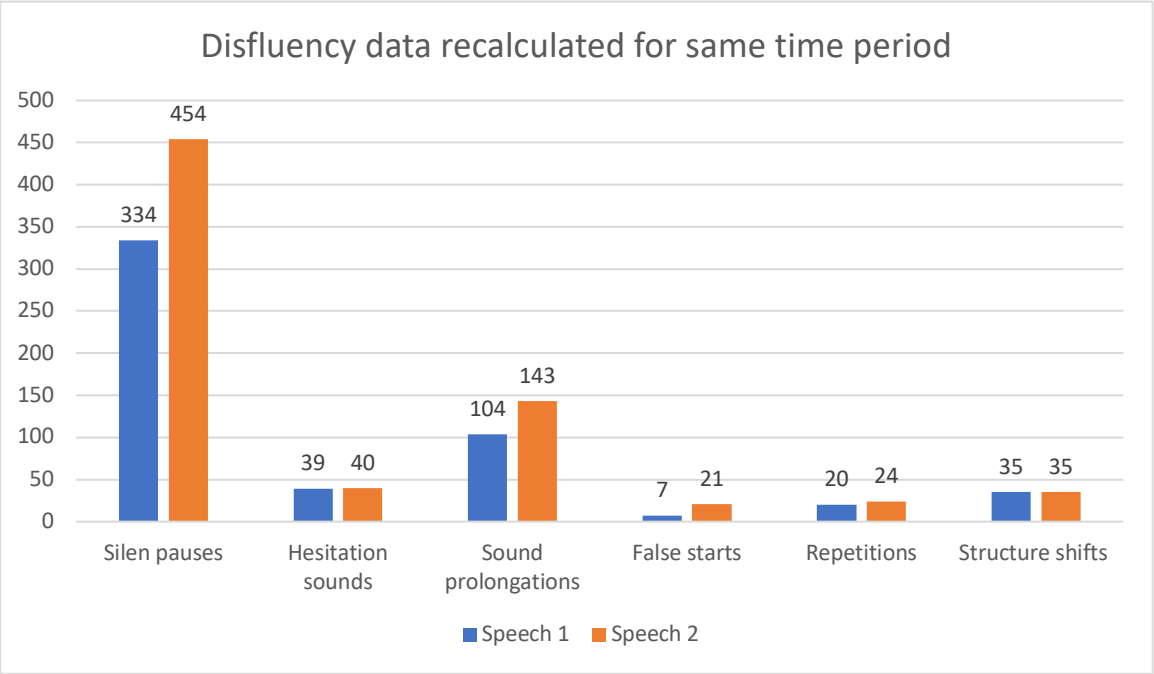

Figure 5: Aggregate data recalculated

As seen in Figure 5, silent pauses and sound prolongations were considerably more frequent during the interpretation of Speech 2. For a speech length of 323 seconds, the participants produced 334 silent pauses for Speech 1 and 454 pauses during Speech 2. Sound prolongations were observed 104 times during Speech 1, and 143 times during Speech 2. False starts were also more frequent during the interpretation of Gerhard Mangott, with the data showing that interpreters produced 21 false starts during their rendition of Speaker 2 and only seven false starts while interpreting Speaker 1. For hesitation sounds and repetitions, the discrepancy is much smaller.

The participants of the experiment produced 39 and 40 hesitation sounds for Speech 1 and Speech 2, respectively. For Speech 1, 20 repetitions were counted across all six interpretations for Speech 1 and 24 for Speech 2. Structure shifts were the only type of disfluency that had the same values for both speeches, appearing 35 times in the interpretations for each speech. Overall, the data shows that if both speeches had the same length, all disfluencies, except for structure shifts, were more frequently observed for Speech 2 than for Speech 1.

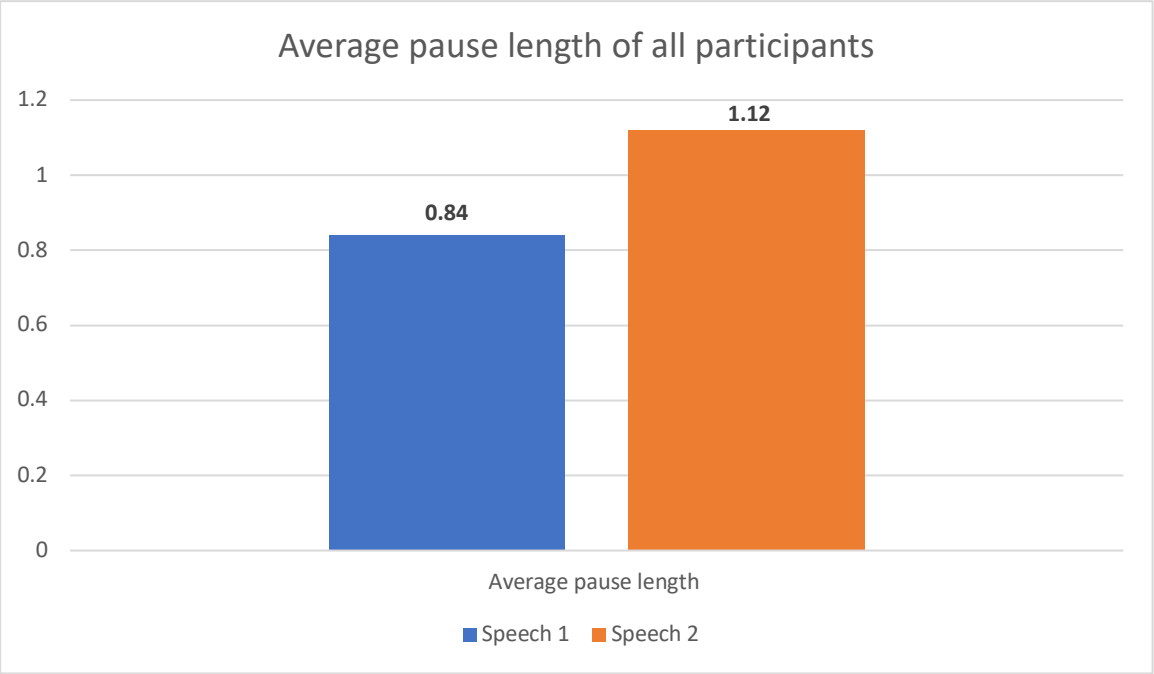

Figure 6: Average pause length

Figure 6 shows the average pause length of all six participants for each speech. As can be seen, the pauses for Speech 2 were much longer, lasting 1.12 seconds on average, compared to 0.84 seconds for Speech 1. All six participants produced longer pauses for Speech 2 than for Speech 1. Four out of six of the participants produced pauses at an average length of above 1 second. The other two interpreters had an average length of around 0.80 seconds, which was still longer than their pauses produced during Speech 1. For Speech 1, all interpreters had an average pause length between 0.74 seconds and 0.98, meaning that no interpreter surpassed the 1-second mark.

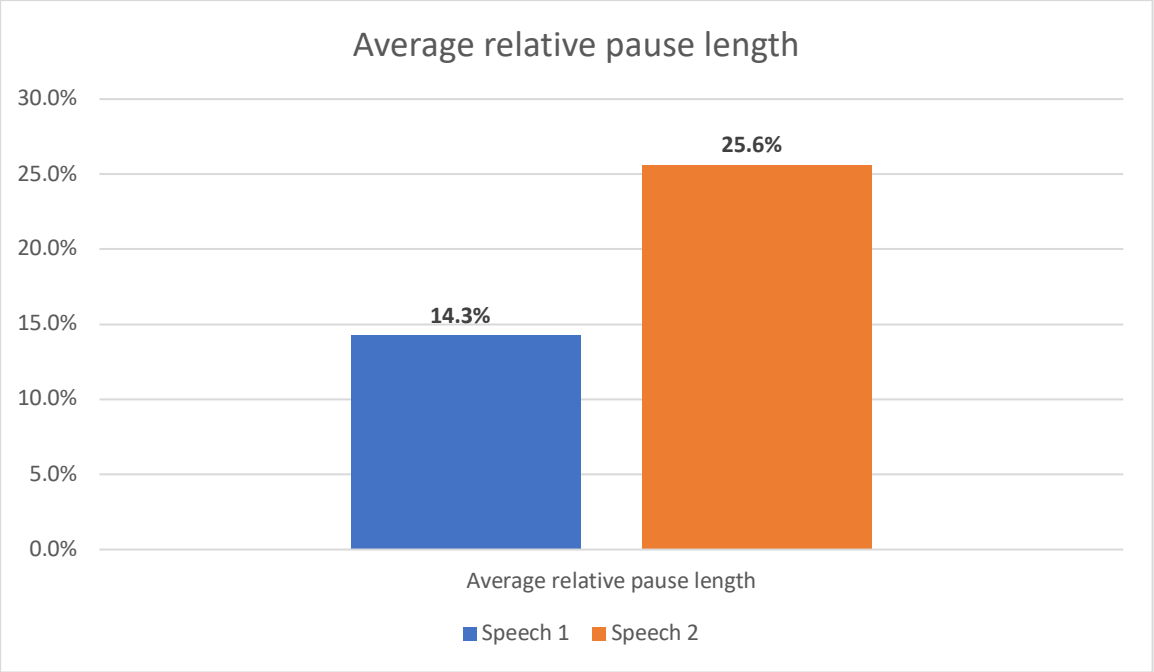

Figure 7: Average relative pause length

Figure 7 illustrates the substantial difference in relative pause length for Speech 1 and Speech 2. All six participants displayed a considerably higher percentage of pause length relative to speech length for Speech 2, resulting in an overall average that clearly reflects these findings. The average relative pause length for Speech 2, at 25.6%, is almost twice as high as that for Speech 1, for which the average is 14.3%.

7. Discussion and conclusions

The present thesis has explored the link between increased cognitive demand caused by poor sound quality and the occurrence of disfluencies in interpreters' outputs. While the hypothesis for the experiment predicted that all participants would produce more disfluencies during their interpretation of Gerhard Mangott's speech, which contained poor audio quality, than during Paul Schmidt's text, this was not the case. Although the results were mixed and the findings for some participants partially sustained the hypothesis, no clear link between poor audio quality and the occurrence of disfluencies can be found. Speech 2 resulted in a higher number of unfilled pauses for all participants. Silent pauses were also, on average, longer for Speech 2 than the silent pauses produced for Speech 1. Filled pauses, namely hesitation sounds, and sound prolongations were also more frequent in all interpretations for Speech 2. Therefore, the frequency of filled and unfilled pauses supported the hypothesis and indicated higher cognitive strain during the video excerpt with poor audio quality. Nevertheless, other types of disfluency, namely self-corrections, false starts and repetitions showed no clear pattern. Overall, the interpreting students produced slightly more disfluencies during their rendering of Speech 2, but the difference was not large enough to indicate a correlation between the audio quality and the disfluencies observed.

A possible explanation for this outcome could be related to the source text. As shown in Chapter 5.2.1, Gerhard Mangott's speech is higher in information density, new-concept density, and is also delivered at a faster speech rate and articulation rate. All this indicates that Speaker 2 is overall more difficult to interpret and should, as the hypothesis predicted, lead to a higher number of disfluencies. However, Paul Schmidt's speech and articulation rates show that he speaks more slowly and that he pauses more often during his speech. The increased number of pauses, both filled and unfilled, during the interpretations of his speech might be due to the fact that interpreters were waiting for further input from the speaker before continuing their interpretation. The number of pauses in Paul Schmidt's speech also mean that interpreters had to store information in their working memory for longer and the additional cognitive load imposed on the memory effort could have affected the interpreter's output production. Similarly, Gerhard Mangott's increased speech rate could have led to an increased speech rate by the interpreters, resulting in fewer

pauses. As explained in Chapter 3.2.2., interpreters who are confronted with increased speech rate will accelerate their target-text production in order to cope with the faster input. It could be argued that, while producing a target text that accommodated the accelerated delivery rate, interpreters would be more prone to making mistakes and being forced to resort to self-corrections and other types of disfluency, but this was not the case with the interpreters involved in this experiment. A further factor that could have contributed to the disfluencies observed for Speech 1 is the structure of both speeches. The first part of Paul Schmidt's speech, which accounts for a majority of his input, consists of an introduction to the topic discussed at the online meeting and an introduction of all the guests, followed by a question addressed to Mr Mangott. It is likely that this part of his speech was prepared in advance and was read aloud, either from a pre-formulated script or from a note containing keywords. Gerhard Mangott's speech, however, was produced spontaneously and was therefore perhaps easier to follow, as it sounded more natural and spontaneous.

In the analysis of the results, silent pauses were the only type of disfluency that was observed more frequently in all six renditions of Speech 2. The average pause lengths for each speech also indicated greater difficulty in interpreting Speech 2. Here it must be noted that for some participants exceptionally long pauses throughout their interpretation resulted in a high average pause length. This was the case for Participants 2 and 3. Participant 2 often stopped interpreting in the middle of a sentence and then continued following an extended period of silence, resulting in pauses of more than five seconds in length. These prolonged pauses affected the average pause length considerably. In the case of Participant 3, one extremely long pause of 11.52 seconds during Gerhard Mangott's speech had the same impact on the average pause length.

For all disfluencies, the data analysis takes into account that the speeches in the experiment differ in length. To compensate for this, the values found for Speech 2 were recalculated for the same speech length of Speech 1. While this provides an estimate for the value of a given disfluency for a longer time period, it is based on the assumption that the interpreters would produce disfluencies at the same rate as they did for the first 204 seconds of interpretation. Since the calculations are carried out based on an assumption, they do not lead to exact values. Interpreters are not machines that produce a certain output at a stable rate throughout their interpretation and therefore the values used should be considered with caution. While the analysis

of all types of disfluency was affected by the difference in length of both audio conditions, the pause length relative to speech length was the only measured variable that takes the different speech lengths into account. For every one of the six participating interpreting students, the relative pause length was higher for Speech 2 and for most participants, apart from Participant 4, the difference was substantial. This resulted in an average relative pause length for Speech 2 that is almost twice as high as the value for Speech 1. These results further confirm that all participants paused more frequently during the excerpt containing poor audio quality and that the pauses were also considerably longer, which is a further indicator of cognitive strain, supporting the hypothesis made for this thesis.

In the analysis of hesitation sounds the findings varied among the participating interpreters. While Participants 3, 5 and 6 produced more hesitation sounds during the poor audio quality condition, the remaining three participants vocalised hesitation sounds more often during their rendition of Speech 1. Therefore, the aggregate results show that, overall, the interpretations of Speech 2 only show a minimally higher frequency of hesitation sounds. The results for sound prolongations were similar. Participants 1, 2, 5 and 6 produced more sound prolongations while interpreting Speech 2. However, Participants 3 and 4 lengthened vowels and consonants more often during their rendition of Speech 1. While the findings for those participants who were more challenged by Speech 2 were predicted by the hypothesis, it is worth analysing what might have led Participants 3 and 4 to their increased number of sound prolongations during Speech 1. As in the case of silent pauses and hesitation sounds, the vowel and consonant prolongations could have been a result of the lower speech and articulation rates of Speaker 1. As the speaker paused more frequently, the interpreters could have used sound lengthenings to gather enough input to continue the interpretation. As most participants still presented findings that supported the hypothesis, the overall result shows an increased number of sound prolongations for Speech 2, as expected.

It is worth noting that for one participant in particular, namely Participant 5, silent pauses and sound prolongations were the types of disfluency that were observed most frequently. While pausing, lengthening sounds or even during hesitation sounds, interpreters plan their speech and use these types of disfluency as strategies to gain time for speech planning. False starts, repetitions and self-corrections all occur when an interpreter does not wait before producing speech and subsequently needs to

correct or repeat their utterance. For Participant 5, it can be observed that they show a tendency towards disfluencies that serve to gain time for speech planning before continuing their interpretation. As a result, it is less likely that they produce an output that needs to be repeated or corrected. Thus, it could be speculated that the higher number of pauses and prolonged sounds led to a lower number of other types of disfluency.

The findings relating to false starts supported the expected outcome for most interpreters. Participants 1 and 5 were the only two who produced more disfluencies during Speech 1, hence contradicting the hypothesis. In both cases, the interpreters produced no false starts at all during Gerhard Mangott's speech and only one or two false starts during Paul Schmidt's speech. Considering that both interpreters only produced a very low number of false starts throughout their interpretations, it could be argued that they do not resort to this type of disfluency during cognitive strain. The remaining four interpreting students did produce more false starts during their interpretations of Mr Mangott's speech and so the overall results for this type of disfluency still support the hypothesis underlying this thesis.

The analysis of repetitions included in all interpreters' outputs showed that one half of the participants produced more repetitions in Speech 1 and the other half in Speech 2. Participants 1, 2 and 5 repeated themselves more frequently while interpreting Speaker 1 and Participants 3, 4 and 6 produced more repetitions while interpreting Speaker 2. Nevertheless, the students who repeated themselves more often during Speech 1 only showed a minimally higher frequency of repetitions during this speech. The results for Participants 3, 4 and 6, however, showed a greater difference between repetitions for Speech 1 and Speech 2. Thus, the overall findings show slightly more repetitions for Speech 2 than for Speech 1.

The results for structure shifts during the interpretations were the most surprising. Similarly to the results found for repetitions, half of all participants produced more structure shifts for Speech 1 and the other half for Speech 2. However, unlike for repetitions, the difference in the number of structure shifts for each speech was the same for both groups, meaning that the final results show that this type of disfluency was observed in equal measure for both speeches, indicating that no speech was more or less straining than the other, which refutes the hypothesis of the thesis.

Although previous experiments, such as the one carried out by David Gerver in 1974, showed a link between poor audio quality and performance deterioration in simultaneous interpreting, the experiment in the current thesis was unable to produce conclusive results. The lack of a clear causal effect of poor sound quality on disfluencies in the participant's interpretations can be linked to several limiting factors concerning the experiment carried out in this thesis. In his experiment, Gerver (1974) used different levels of sound degradation to test if the output quality deteriorated. Tommola and Lindholm (1995) added white noise at 5dB to the sound they used for a similar experiment. Due to constraints relating to the experiment for this thesis, the level of sound degradation during Gerhard Mangott's speech was not measured. It is possible that the level of sound degradation present during his speech was not acute enough to cause perceptible cognitive strain for the interpreters. Perhaps there is a minimum level of sound degradation, white noise or length of exposure necessary to trigger cognitive overload during SI and only then will interpreters produce noticeably more disfluencies during their rendition. It would be interesting to replicate the experiment of the current thesis with different levels of sound degradation or exposure length in order to test whether the expected results were produced at a certain value. However, testing different degradation levels would mean that the same source text would need to be artificially altered, resulting in an interpreting scenario that would not be realistic. Testing the impact of sound quality at a certain exposure length would be an easier undertaking, as the plethora of online audiovisual material could include a video that contains longer excerpts of poor audio quality.

To further build on this experiment, future research could include a questionnaire or an interview as a second step following the experiment, to learn how the interpreters perceived both texts and whether they felt that the poor audio quality posed a noticeable challenge during the interpretation. This would allow for further analysis as to which of the texts was thought to be more difficult and whether poor audio quality is, in fact, perceived as being cognitively more challenging. The pause analysis carried out with the target-text transcripts only allowed for a subjective analysis, as only pauses that were perceived to show hesitation by the interpreter were annotated. It could be that interpreters felt challenged throughout the excerpt displaying poor audio quality but did not produce any type of disfluency.

What must also be considered is that both texts, albeit adequate for the skill level of all the participants, were very fast-paced and informationally dense. While the

audio quality was the sole focus of the experiment, it must be considered that the fast speech rate, complex subject matter and dense information flow will also have contributed to disfluencies by all interpreters and therefore not all disfluencies can be solely traced to the poor audio quality. If the experiment were to be repeated, one could either search for texts that are produced at lower speech rates or are objectively easier or less dense in information. Alternatively, the same text could be used for two different groups and the audio quality manipulated for the test group, to analyse whether this group produces more disfluencies than the control group. While this would be an effective way to exclude other variables that may affect the results, such as a difference in the content of the materials used for the experiment, the artificial manipulation of sound would mean that the control group would interpret a naturally produced text, whereas the test group would be working with an artificial text, which could pose other difficulties during the interpretation. This approach was not used in the current master's thesis because the aim in analysing disfluencies produced during poor audio quality interpretation was to use authentic material that an interpreter would realistically be confronted with in the current market and draw attention to the importance of adequate sound quality for an interpreter's performance and ultimately for the communication between both parties that rely on the interpretation in the first place.

The link between poor audio quality and disfluencies in the target text has been the focus of only a few studies in the field. While an experiment with this focus offers a valuable approach to test the impact of poor audio quality on the cognitive processes during interpretation, it does not allow for a complete analysis and might overlook other phenomena in the interpretation that would also suggest cognitive overload. In the current thesis, for example, the focus on disfluencies meant that any other observable impacts on the target text were not considered. In this experiment, for example, some participants, namely Participants 2, 3 and 6, omitted parts of the source text while interpreting Speech 2. These omissions are only reflected in the average pause lengths for each participant as the interpreters were silent for extended periods of time, but do not show to what extent they were challenged by the audio condition they were subjected to. A more detailed analysis of the target text, specifically of the errors, omissions and infelicities (EOIs), would offer a more complete picture of the impact the audio quality might have had on the interpretation. The participants in the experiment did not only produce omissions and errors in their interpretations but also presented other signs that indicated they were cognitively strained, such as long sighs of

frustration or formulating statements as if they were questions because they were unsure whether their output was accurate. All these phenomena indicate that the cognitive resources were allocated primarily to the listening and analysis effort and fewer resources were available to the production effort and, as a result, to output monitoring.

A further point to be taken into consideration when interpreting the findings of this thesis is the limited representativeness of the data set. Although each interpreter's output was transcribed and thoroughly analysed, the data analysis focuses on quantitative parameters and not on qualitative features of interpreting performance. Considering that the experiment follows a quantitative approach, six participants and the findings for their target texts are not sufficient to fully confirm or disprove a hypothesis. In order to produce more solid findings, the experiment would have to be replicated using a much higher number of participants to truly allow for a detailed analysis of a pattern among all participants that might then be extrapolated to a larger group.

Even though the findings of this thesis do not fully confirm a causal effect of poor audio quality and disfluencies produced during simultaneous interpreting, the experiment as well as other experiments described throughout this thesis show the importance of further research on audio quality and its impact on simultaneous interpreting. The effect poor audio quality has on interpreters' short-term memories, their output quality and their mental health has been shown in several experiments in past years. This shows the importance of further research to be conducted on this very impactful input variable, so that more awareness is raised around the topic of audio input and interpreting. Further research could mean that a specific value for sound degradation or exposure length is found at which poor audio quality starts affecting the cognitive capacities of a conference interpreter and this finding could improve the working conditions of future interpreters.

Bibliography

Aguado Padilla, Karin (2002). *Imitation als Erwerbsstrategie. Interaktive und kognitive Dimensionen des Fremdsprachenerwerbs*. PhD Thesis, Universität Bielefeld.

Ahrens, Barbara (2004). *Prosodie beim Simultandolmetschen*. Frankfurt/Main: Peter Lang.

AIIC (2000). Code for the Use of New Technologies in Conference Interpretation. https://www.staff.uni-mainz.de/fantinuo/class/files/cai/aiic_code_interpreting%20technologies.pdf [Accessed 15 July 2022]

AIIC (2002). Interpreter Workload Study – Full Report https://aiic.org/document/468/AIICWebzine_FebMar2002_7_AIIC_Interpreter_workload_study_full_report_WLS_Full_Report.pdf [Accessed 7 April 2023]

AIIC (2020a). AIIC Guidelines for Distance Interpreting (Version 1.0). https://aiic.org/document/4418/AIIC%20Guidelines%20for%20Distance%20Interpreting%20(Version%201.0)%20-%20ENG.pdf [Accessed 15 July 2022]

AIIC (2020b). AIIC Best Practices for Interpreters During the Covid-19 Crisis. https://aiic.org/document/4840/AIIC%20best%20practices%20for%20interpreters%20during%20the%20Covid-19%20crisis%20-%20ENG.pdf [Accessed 15 July 2022]

AIIC (2020c). AIIC Interpreter Checklist: Performing Remote Interpreting Assignments from Home in Extremis During the Covid-19 Pandemic. https://aiic.org/document/4845/AIIC-Interpreter-Checklist.pdf [Accessed 15 July 2022]

AIIC (2020d). Acoustic Shocks Research Project. https://aiic.org/uploaded/web/Acoustic%20Shocks%20Research%20Project.pdf [Accessed 15 July 2022]

AIIC (n.d.). AIIC Unwrapped. https://aiic.org/site/world/about [Accessed 15 July 2022]

Arnold, Jennifer E. & Tanenhaus, Michael K. (2011). Disfluency Effects in Comprehension: How New Information Can Become Accessible. In: Gibson,

Edward & Perlmutter, Neal (eds.) *The Processing and Acquisition of Reference. Cambridge*, MA: MIT Press, 197– 217.

Barik, Henri C. (1973). Simultaneous Interpretation: Temporal and Quantitative Data. *Language and Speech* 16 (3), 237–270.

Bovair, Susan & Kieras, David E. (1985). A Guide to Propositional Analysis for Research on Technical Prose. In: Britton, Bruce K. & Black, John B. (eds.) *Understanding Expository Text*. Hillsdale: Erlbaum, 317-362.

Braun, Sabine (2015). Remote Interpreting. In: Mikkelson, Holly & Jourdenais, Renée (eds.) *The Routledge Handbook of Interpreting*. New York: Routledge, 352 – 368.

Braun, Sabine & Taylor, Judith L. (eds.) (2012). *Videoconference and Remote Interpreting in Criminal Proceedings*. Cambridge/Antwerp: Intersentia.

Bühler, Hildegund (1989). Discourse Analysis and the Spoken Text – a Critical Analysis of the Performance of Advanced Interpretation Students. In: Gran, Laura & Dodds, John (eds.) T*he Theoretical and Practical Aspects of Teaching Conference Interpretation. First International Symposium on Conference Interpreting at the University of Trieste*. Udine: Campanotto Editore, 131-137.

Butcher, Andrew (1981*). Aspects of the Speech Pause. Phonetic Correlates and Communicative Functions*. Kiel: Institut für Phonetik.

Bußmann, Hadumod (1983). *Lexikon der Sprachwissenschaft. 2. Aufl.* Stuttgart: Kröner.

Caniato, Andrea (2020). Acoustic Shocks are a Red Herring. A Different, Almost Silent Threat is Slowly Poisoning the Interpreter's Ears. *LinkedIn*. https://www.linkedin.com/pulse/acoustic-shocks-red-herring-different-not-so-silent-threat-caniato/ [Accessed 15 July 2022]

Chambers, Francine (1997). What Do We Mean by Fluency? *System* 25 (4), 535–544

Chernov, Ghelly V. (2004). *Interference and Anticipation in Simultaneous Interpreting: A Probability-Prediction Model*. Amsterdam/Philadelphia: John Benjamins.

Coblenzer, Horst & Muhar, Franz (2006). *Atem und Stimme: Anleitung zum guten Sprechen*. 20. Aufl. Vienna: öbv & hpt.

Cooper, Cary L., Davies, Rachel & Tung, Rosalie L. (1982). Interpreting Stress: Sources of Job Stress Among Conference Interpreters. *Multilingua* 1 (2), 97–107.

Cousins, Katheryn A.; Dar, Hayim; Wingfield, Arthur and Miller, Paul (2014). Acoustic Masking Disrupts Time-Dependent Mechanisms of Memory Encoding in Word-List Recall. *Memory & Cognition 42 (4)*, 622–638.

Crevatin, Franco (1989). Directions in Research Towards a Theory of Interpretation. In Gran, Laura & Dodds, John (eds.) *The Theoretical and Practical Aspects of Teaching Conference Interpretation: First International Symposium on Conference Interpreting at the University of Trieste.* Udine: Campanotto, 21-22.

Dam, Helle V. (2001). On the Option Between Form-Based and Meaning-Based Interpreting: The Effect of Source Text Difficulty on Lexical Target Text Form in Simultaneous Interpreting. *The Interpreters' Newsletter* 11, 27–55.

Darò, Valeria; Lambert, Sylvie & Fabbro, Franco (1996). Conscious Monitoring of Attention During Simultaneous Interpretation. *Interpreting* 1 (1), 101–124.

Defrancq, Bart; Plevoets, Koen & Magnifico, Cédric (2015). Connective Markers in Interpreting and Translation: Where Do They Come From. In: Romero-Trillo, Jesús (Ed.) *Yearbook of Corpus Linguistics and Pragmatics*, *3*. Dordrecht: Springer, 195–222.

Déjean Le Féal, Karla (1980). Die Satzsegmentierung beim freien Vortrag bzw. beim Verlesen von Texten und ihr Einfluss auf das Sprachverstehen. In: Kühlwein, Wolfgang & Raasch, Albert (eds.) *Sprache und Verstehen. Band I. Kongressberichte der 10. Jahrestagung der Gesellschaft für Angewandte Linguistik GAL e.V.* Tübingen: Narr*,* 161-168.

Déjean Le Féal, Karla (1982). Why Impromptu Speech is Easy to Understand. In: Enkvist, Nils Erik (ed.) *Impromptu Speech: A Symposium*. Åbo: Åbo Akademi, 221–239.

Desjardins, Jamie L. & Doherty, Karen A. (2013). Age-Related Changes in Listening Effort for Various Types of Masker Noises. *Ear and Hearing* 34 (3), 261–272.

Dillinger, Mike (1994). Comprehension During Interpreting: What Do Interpreters Know That Bilinguals Don't? In: Lambert, Sylvie & Moser-Mercer, Barbara (eds.) *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. Amsterdam: John Benjamins, 155–189.

Diriker, Ebru (2015). Simultaneous Interpreting. In: Pöchhacker, Franz (ed.) *Routledge Encyclopedia of Interpreting Studies.* London: Routledge, 382-385.

European Commission (2000). *Rapport concernant les tests de simulation de téléconférence au SCIC en janvier 2000*. Unpublished internal report. Brussels: EC/SCIC.

European Parliament. (2001). Report on Remote Interpretation Test: 22-25 January, 2001 Brussels. https://www.europarl.europa.eu/interp/remote_interpreting/ep_report1.pdf [Accessed 15 July 2022]

Festinger, Leon (1957). *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

Flesch, Rudolf (1948). A New Readability Yardstick. *Journal of Applied Psychology* 32, 221-233.

Francis, Alexander L. (2010). Improved Segregation of Simultaneous Talkers Differentially Affects Perceptual and Cognitive Capacity Demands for Recognizing Speech in Competing Speech. *Attention, Perception, & Psychophysics* 72 (2), 501–516.

Francis, Alexander L. & Love, Jordan (2020). Listening Effort: Are we Measuring Cognition or Affect, or Both? *Wiley Interdisciplinary Reviews. Cognitive Science*, *11* (1), e1514.

Francis, Alexander L.; MacPherson, Megan K.; Chandrasekaran, Bharath & Alvar, Ann M. (2016). Autonomic Nervous System Responses During Perception of Masked Speech may Reflect Constructs Other Than Subjective Listening Effort. *Frontiers in Psychology* 7, Art. 263.

Fulcher, Glenn (1997). Text Difficulty and Accessibility: Reading Formulae and Expert Judgment. *System* 25 (4), 497–513.

Gerver, David (1969/2002). The Effects of Source Language Presentation Rate on the Performance of Simultaneous Conference Interpreters. In: Pöchhacker, Franz & Shlesinger, Miriam (eds.) *The Interpreting Studies Reader*. London: Routledge, 53–66.

Gerver, David (1971). *Aspects of Simultaneous Interpretation and Human Information* Processing. PhD dissertation, University of Oxford.

Gerver, David (1974). The Effects of Noise on the Performance of Simultaneous Interpreters: Accuracy of Performance. *Acta Psychologica* 38 (3), 159–167.

Gile, Daniel (1984). Les noms propres en interprétation simultanée. *Multilingua* 3 (2), 79–85.

Gile, Daniel (1985). Le modèle d'efforts et l'équilibre en interprétation simultanée. *Meta* 30 (1), 44–48.

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

Gile, Daniel (1997/2002). Conference Interpreting as a Cognitive Management Problem. In: Pöchhacker, Franz & Shlesinger, Miriam (eds.) *The Interpreting Studies Reader*. London: Routledge, 163–176.

Gile, Daniel (1999). Testing the Effort Models' Tightrope Hypothesis in Simultaneous Interpreting – a Contribution. Hermes. *Journal of Linguistics* 23, 153–172.

Gile, Daniel (2009). *Basic Concepts and Models for Interpreter and Translator Training*. 2nd edn. Amsterdam: John Benjamins.

Gilquin, Gaëtanelle (2008). Hesitation Markers Among EFL Learners: Pragmatic Deficiency or Difference? In: Romero Trillo, Jesús (Ed.), *Pragmatics and Corpus Linguistics: A Mutualistic Entente*. Berlin: Mouton de Gruyter, 119–149.

Goldman-Eisler, Frieda (1958). Speech Analysis and Mental Processes. *Language and Speech* 1 (1), 59-75.

Goldman-Eisler, Frieda (1961). A Comparative Study of Two Hesitation Phenomena. *Language and Speech* 4 (1), 18-26.

Goldman-Eisler, Frieda (1968). *Psycholinguistics. Experiments in Spontaneous Speech*. London / New York: Academic Press.

Gran, Laura & Taylor, Christopher (1990). Closing Statement. In: Gran, Laura & Taylor, Christopher (eds.) *Aspects of Applied and Experimental Research and Conference Interpretation. Round Table on Interpretation Research, November 16, 1989.* Udine: Campanotto Editore, 245-253.

Guiducci, Cristian (2020). Headsets Won't Work Miracles: Here is How Digital Sound Gets Degraded in the 21st Century. *LinkedIn*. https://www.linkedin.com/pulse/headsets-wont-work-miracles-here-how-digital-sound-gets-caniato/ [Accessed 15 July 2022]

Gumul, Ewa (2021). Explicitation and Cognitive Load in Simultaneous Interpreting Product- And Process-Oriented Analysis of Trainee Interpreters' Outputs. *Interpreting* 23 (1), 45–75.

Heald, Shannon & Nusbaum, Howard (2014). Speech Perception as an Active Cognitive Process. *Frontiers in Systems Neuroscience* 8, Art. 35.

Hedge, Tricia (1993). Key Concepts in ELT (Fluency). *ELT Journal* 47 (3), 275–276.

Heinisch-Obermoser, Barbara (2010). *Der interinstitutionelle Aufnahmetest für DolmetscherInnen bei den EU-Institutionen : eine Korpusanalyse der Prüfungsreden*. Master‘s thesis, University of Vienna.

Heinrich, Antje; Schneider, Bruce A. & Craik, Fergus I.M. (2008). Investigating the Influence of Continuous Babble on Auditory Short-Term Memory Performance. *Quarterly Journal of Experimental Psychology* 61 (5), 735–751.

Hieke, Adolf E. (1981). A Content-Processing View of Hesitation Phenomena. *Language and Speech* 24 (2), 147–160.

Hieke, Adolf E.; Kowal, Sabine & O’Connell, Daniel C. (1983). The Trouble with ‘Articulatory’ Pauses. *Language and Speech* 26 (3), 203–214

Hooper, R.E. (2014). Acoustic Shock Controversies. *The Journal of Laryngology & Otology* 128 (S2), S2–S9.

Hörmann, Hans (1957). *Meinen und Verstehen. Grundzüge einer psychologischen Semantik.* Frankfurt am Main: Suhrkamp.

Hornsby, Benjamin W. (2013). The Effects of Hearing Aid Use on Listening Effort and Mental Fatigue Associated with Sustained Speech Processing Demands. *Ear and Hearing* 34 (5), 523–534.

ISO (2016). *ISO 20109: 2016. Simultaneous Interpreting – Equipment – Requirement*. https://www.iso.org/standard/67063.html [Accessed 15 July 2022]

ISO (2017). *ISO 20108:2017. Simultaneous interpreting — Quality and transmission of sound and image input — Requirements. https://www.iso.org/standard/67062.html* [Accessed 15 July 2022]

ISO (2019). *ISO 20228:2019. Interpreting services — Legal interpreting — Requirements. https://www.iso.org/standard/67327.html* [Accessed 15 July 2022]

ISO (2020). *ISO 21998:2020 Interpreting services — Healthcare interpreting — Requirements and recommendations.* https://www.iso.org/standard/72344.html [Accessed 15 July 2022]

ISO/PAS (2020). *ISO/PAS 24019:2020 Simultaneous interpreting delivery platforms — Requirements and recommendations.* https://www.iso.org/standard/77590.html [Accessed 15 July 2022]

Jakobson, Roman (1959). On Linguistic Aspects of Translation. In: Brower, Reuben A. (ed.) *On Translation*. Cambridge/Massachussets: Harvard University Press, 232-239.

Janse, Esther (2012). A Non-Auditory Measure of Interference Predicts Distraction by Competing Speech in Older Adults. *Aging, Neuropsychology, and Cognition* 19 (6), 741–758.

Kalina, Sylvia (1998). *Strategische Prozesse beim Dolmetschen. Theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen*. Tübingen: Narr.

Kintsch, Walter & van Dijk, Teun A. (1978). Toward a Model of Text Comprehension and Production. *Psychological Review* 85 (5), 363-394.

Klare, George R. (1974). Assessing Readability. *Reading research quarterly* 10(1), 62–102.

Koelewijn, Thomas; Zekveld, Adriana. A.; Festen, Joost M.; Rönnberg, Jerker & Kramer, Sophia. E. (2012). Processing Load Induced by Informational Masking is Related to Linguistic Abilities. *International Journal of Otolaryngology* 2012, 1–11.

Kondo, Masaomi; Tebble, Helen; Alexieva, Bistra; Dam, Helle V.; Kata, David; Mizuno, Akira; Setton, Robin & Zalka, Ilona (1997). Intercultural Communication, Negotiation, and Interpreting. In: Gambier, Yves; Gile, Daniel & Taylor, Christopher (eds.) *Conference Interpreting: Current Trends in Research. Proceedings of the International Conference on Interpreting: What do we know and how? Turku, August 25-27, 1994.* Amsterdam/Philadelphia: John Benjamins, 149-166.

Kopczyński, Andrzej (1982). Effects of Some Characteristics of Impromptu Speech on Conference Interpreting. In: Enkvist, Nils Erik (ed.) *Impromptu Speech: A Symposium.* Abo: Abo Akademi, 255-266.

Kurz, Ingrid (2000). Tagungsort Genf/Nairobi/Wien: Zu einigen Aspekten des Teledolmetschens. In: Kadric, Mira, Kaindl, Klaus & Pöchhacker, Franz (eds.) *Translationswissenschaft: Festschrift für Mary Snell-Hornby zum 60. Geburtstag*. Tübingen: Stauffenburg, 291–302

Kurz, Ingrid (2005). Akzent und Dolmetschen – Informationsverlust bei einem nichtmuttersprachlichen Redner. *Bulletin suisse de linguistique appliquée* (81), 57-71.

Kurz, Ingrid & Pöchhacker, Franz (1995). Quality in TV Interpreting. *Translatio – Nouvelles de la FIT – FIT Newsletter* 15 (3/4), 350–358.

Lederer, Marianne (1981). *La traduction simultanée–Expérience et théorie*. Paris: Minard Lettres Modernes.

Lennon, Paul (2000) The Lexical Element in Spoken Second Language Fluency. In: Riggenbach, Heidi (ed.) *Perspectives on Fluency*. Michigan: University of Michigan Press, 25–42.

Li, Changshuan (2010). Coping Strategies for Fast Delivery in Simultaneous Interpretation. *The Journal of Specialised Translation* 13. http://www.jostrans.org/issue13/art_li.pdf [Accessed 1 May 2023]

Liu, Minhua & Chiu, Yu-Hsien (2011). Assessing Source Material Difficulty for Consecutive Interpreting. *Interpreting* 11 (1), 244–266.

Mazza, Cristina (2000). Numbers in Simultaneous Interpretation. https://www.openstarts.units.it/bitstream/10077/2450/1/05.pdf [Accessed 18 February 2023]

McAuliffe, Megan J.; Wilding, Philippa J.; Rickard, Natalie A. & O'Beirne, Greg A. (2012). Effect of Speaker Age on Speech Recognition and Perceived Listening

Effort in Older Adults with Hearing Loss. *Journal of Speech, Language, and Hearing Research* 55, 838–847.

McCall, William A. & Crabbs, Lelah M. (1925). *Standard Test Lessons in Reading*. New York: Bureau of Publications, Teachers College, Columbia University.

McCoy, Sandra L.; Tun, Patricia A.; Cox, Clarke L.; Colangelo, Marianne; Stewart; Raj A. and Wingfield, Arthur (2005). Hearing Loss and Perceptual Effort: Downstream Effects on Older Adults' Memory for Speech. *The Quarterly Journal of Experimental Psychology* 58 (1), 22–33.

Mead, Peter (2002). Évolution des pauses dans l'apprentissage de l'interprétation consécutive. PhD Thesis, Université Lumière Lyon.

Mizuno, Akira (2017). Simultaneous Interpreting and Cognitive Constraints. *紀要/青山学院大学文学部* (Minutes / Aoyama Gakuin University Faculty of Literature) 58, 1–28.

Moser-Mercer, Barbara (2003). Remote Interpreting: Assessment of Human Factors and Performance Parameters. https://aiic.org/document/516/AIICWebzine_Summer2003_3_MOSER-MERCER_Remote_interpreting_Assessment_of_human_factors_and_performance_parameters_Original.pdf [Accessed 15 July 2022]

Mouzourakis, Panayotis (2003). That Feeling of Being There: Vision and Presence in Remote Interpreting. *The AIIC Webzine 23 https://aiic.org/document/520/AIICWebzine_Summer2003_7_MOUZOURAKIS_That_feeling_of_being_there_Vision_and_presence_in_remote_interpreting_EN.pdf* [Accessed 25 March 2023]

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

Murphy, Kate (2020). Why Zoom Is Terrible. *The New York times*. https://www.nytimes.com/2020/04/29/sunday-review/zoom-video-conference.html [Accessed 15 July 2022]

Nimdzi (2020). The Virtual Interpreting Landscape. https://www.nimdzi.com/virtual-interpreting-landscape/ [Accessed 25 March 2023]

Nooteboom, Sieb G. (1980). Speaking and Unspeaking: Detection and Correction of Phonological and Lexical Errors in Spontaneous Speech. In: Fromkin, Victoria A. (ed.) *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press, 87-95.

Österreichische Gesellschaft für Europapolitik (2022). Europa Club Live: Krisen- und Kriegsszenarien um die Ukraine und die Rolle der EU. [Videoclip 8‘50“] https://www.youtube.com/watch?v=w3QQnDfVGaE [Accessed 28 March 2023]

Peelle, Jonathan E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear and Hearing* 39 (2), 204–214

Pichora-Fuller, Kathleen M. (2016). How Social Psychological Factors may Modulate Auditory and Cognitive Functioning During Listening. *Ear and Hearing* 37 (Suppl 1), 92S–100S.

Pichora-Fuller, Kathleen M.; Kramer, Sophia E.; Eckert, Mark A.; Edwards, Brent.; Hornsby, Benjamin W.; Humes, Larry E.; Lemke, Ulrike; Lunner, Thomas; Matthen, Mohan; Mackersie, Carol L.; Naylor, Graham; Phillips, Natalie, A.; Richter, Michael; Rudner, Mary; Sommers, Mitchell, S.; Tremblay, Kelly L. & Wingfield, Arthur (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing* 37 (Suppl 1), 5S–27S.

Pöchhacker, Franz (1994). *Simultandolmetschen als komplexes Handeln*. Tübingen: Gunter Narr.

Pöchhacker, Franz (1995). Slips and Shifts in Simultaneous Interpreting. In Tommola, Jorma (ed.) Topics in Interpreting Research, Turku: University of Turku Centre for Translation and Interpreting, 73–90.

Pöchhacker, Franz (2004). *Introducing Interpreting Studies*. London/New York: Routledge.

Pöchhacker, Franz (2016). *Introducing Interpreting Studies*. 2nd edn. London: Routledge.

Pompino-Marschall, Bernd (1995). *Einführung in die Phonetik*. Berlin: De Gruyter.

Pradas Macías, Esperanza, M. (2006). Probing Quality Criteria in Simultaneous Interpreting: The Role of Silent Pauses in Fluency. *Interpreting 8* (1), 25–43.

Rabbitt, Patrick M. (1968). Channel-Capacity, Intelligibility and Immediate Memory. *Quarterly Journal of Experimental Psychology* 20 (3), 241–248.

Rennert, Sylvi (2008). Visual Input in Simultaneous Interpreting. *Meta* 53 (1), 204-217.

Rennert, Sylvi (2019). *Redeflüssigkeit und Dolmetschqualität. Wirkung und Bewertung*. Tübingen: Narr Francke Attempto.

Royé, Hans-Walter (1983). *Segmentierung und Hervorhebungen in gesprochener deutscher Standardsprache. Analyse eines Fernseh-Polylogs*. Tübingen: De Gruyter.

Roziner, Ilan & Shlesinger, Miriam (2010). Much Ado about Something Remote: Stress and Performance in Remote Interpreting. *Interpreting* 12 (2), 214–247.

Rudner, Mary; Lunner, Thomas; Behrens, Thomas; Thorén, Elisabet S. & Rönnberg, Jerker (2012). Working Memory Capacity May Influence Perceived Effort During Aided Speech Recognition in Noise. *Journal of the American Academy of Audiology* 23 (8), 577–589.

Saeed, Muhammad A.; González, Eloy R.; Korybski, Tomasz; Davitti, Elena and Braun, Sabine (2022). Connected yet Distant: An Experimental Study into the

Visual Needs of the Interpreter in Remote Simultaneous Interpreting. In: Kurosu, Masaaki (ed.) *Human-Computer Interaction. User Experience and Behavior*. Cham: Springer, 214-232

Seeber, Kilian G. (2015). Simultaneous Interpreting. In: Mikkelson, Holly & Jourdenais, Renée (eds.) *The Routledge Handbook of Interpreting.* London: Routledge, 79-95.

Seeber, Kilian G. (ed.) (2021). *100 Years of Conference Interpreting*. Newcastle-upon-Tyne: Cambridge Scholars.

Setton, Robin (1999). *Simultaneous Interpretation. A Cognitive-Pragmatic Analysis*. Amsterdam/Philadelphia: John Benjamins.

Shao, Zhangminzi & Chai, Mingjiong (2021). The Effect of Cognitive Load on Simultaneous Interpreting Performance: an Empirical Study at the Local Level. *Perspectives* 29 (5), 778–794.

Shlesinger, Miriam (1989). *Simultaneous Interpretation as a Factor in Effecting Shifts in the Position of Texts on the Oral-Literate Continuum*. MA thesis, Tel Aviv University.

Shlesinger, Miriam (2000). *Strategic Allocation of Working Memory and Other Attentional Resources*. PhD dissertation, Bar-Ilan University.

Shlesinger, Miriam (2003). Effects of Presentation Rate on Working Memory in Simultaneous Interpreting. http://www.openstarts.units.it/dspace/bitstream/10077/2470/1/02.pdf [Accessed 18 February 2023]

Söderpalm Talo, Ewa (1980). Slips of the Tongue in Normal and Pathological Speech. In: Fromkin, Victoria A. (ed.) *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press, 81-86.

Solso, Robert L. (1998). *Cognitive Psychology*. 5th edn. Boston: Allyn and Bacon

Sommerlad, E. Lloyd (1977). The 'Symphonie' Experiment: Unesco's First Teleconference by Satellite. *The UNESCO Courier* 30 (4), 32.

Taylor, Christopher (1989). Textual Memory and the Teaching of Consecutive Interpretation. In: Gran, Laura and Dodds, John (eds.) *The Theoretical and Practical Aspects of Teaching Conference Interpretation: First International Symposium on Conference Interpreting at the University of Trieste*. Udine: Campanotto, 177-184.

Tissi, Benedetta (2000). Silent Pauses and Disfluencies in Simultaneous Interpretation: a Descriptive Analysis. *The Interpreters' Newsletter 10* (4), 103-127.

Tommola, Jorma & Helevä, M. (1998). Language Direction and Source Text Complexity: Effects on Trainee Performance in Simultaneous Interpreting. In: Bowker, Lynne; Cronin, Michael; Kenny, Dorothy & Pearson, Jennifer (eds.) *Unity in Diversity: Current Trends in Translation Studies*. Manchester: St. Jerome, 177–186.

Tommola, Jorma & Lindholm, Johan (1995). Experimental Research on Interpreting: Which Dependent Variable? In: Tommola, Jorma (ed.) *Topics in Interpreting Research*. Turku: University of Turku Centre for Translation and Interpreting, 121–133.

UNESCO (1987) Management of Interpretation Services in the United Nations System. http://unesdoc.unesco.org/ images/0007/000732/073286eo.pdf [Accessed 15 July 2022]

United Nations (1999). A Joint Experiment in Remote Interpretation. UNHQ-UNOG-UNOV. Geneva: United Nations, Department of General Assembly Affairs and Conference Services.

United Nations (2001). Remote Interpretation. *The Interpreters' Newsletter* 11, 163–180.

Wang, Yang; Naylor, Graham; Kramer, Sophia E.; Zekveld, Adriana A.; Wendt, Dorothea; Ohlenforst, Barbara & Lunner, Thomas. (2018). Relations Between Self-Reported Daily-Life Fatigue, Hearing Status, and Pupil Dilation During a Speech Perception in Noise Task. *Ear and Hearing* 39 (3), 573–582.

Weissenhofer, Peter (1994). Zur Rolle der terminologischen Begriffslehre in der Translationswissenschaft. In: Snell-Hornby, Mary; Pöchhacker, Franz & Kaindl, Klaus (eds.) *Translation Studies. An Interdiscipline.* Amsterdam/Philadelphia: John Benjamins, 319-326.

Westcott, Myriam (2006). Acoustic Shock Injury (ASI). *Acta Oto-Laryngologica* 126 (556), 54-58.

Westcott, Myriam (2008). Tonic Tensor Tympani Syndrome – an Explanation for Everyday Sounds Causing Pain in Tinnitus and Hyperacusis Clients. Poster IXth International Tinnitus Seminar, Gothenburg

Wu, Yu-Hsiang; Stangl, Elizabeth; Chipara, Octav; Hasan, Syed Shabih; Welhaven, Anne & Oleson, Jacob. (2018*).* Characteristics of Real-World Signal to Noise Ratios and Speech Listening Situations of Older Adults With Mild to Moderate Hearing Loss*. Ear and Hearing* 39 (2), 293-304.

Zaremba, H.-D. (1997). *Evaluation des questionnaires Beaulieu. Expériences en automne 1995*. Unpublished internal report. Brussels: European Commission - SCIC.

Ziegler, Klaus & Gigliobianco, Sebastiano (2019). Present? Remote? Remotely Present! New Technological Approaches to Remote Simultaneous Conference Interpreting. In: Fantinuoli, Claudio (ed.) *Interpreting and Technology*. Berlin: Language Science Press, 119-137.

Abstract – English

The aim of this master's thesis is to investigate the impact of poor audio quality on simultaneous interpreting, measured by the frequency of disfluencies present in the target text. The first part of the thesis offers a theoretical framework which describes the key concepts discussed throughout the thesis as well as the cognitive processes that take place in an interpreter's brain during simultaneous interpreting. The first chapter offers definitions of key terms and describes the development of remote interpreting and remote simultaneous interpreting throughout the 20th and 21st centuries. In this chapter, technical requirements referring to the audio input during simultaneous interpreting, both on-site and remote, are listed and explained. Chapter 2 describes Daniel Gile's Effort Model (2009), followed by examples of empirical studies that have tested the relation between poor audio quality and cognitive processes during simultaneous interpreting. The third chapter describes further factors that can impact an interpreter's performance, including input variables related to the source text, the speaker and the interpreting environment. Chapter 4 explains what disfluencies are and how they have previously been the focus of empirical research in interpreting studies.

The second part of the thesis describes and discusses an experiment partially based on Gerver's (1974) experiment to test the influence of poor audio quality on output quality. Six advanced student interpreters from the Centre for Translation Studies at the University of Vienna were involved in the experiment. They were asked to interpret a video under two audio conditions, which were good and poor audio quality. Both audio conditions were present within one, nine-minute-long video, which all six interpreters were asked to interpret from German into English. The participants' output was transcribed and analysed for disfluencies. The hypothesis was that interpreters would produce more disfluencies in the form of unfilled pauses, hesitation sounds, sound prolongations, self-corrections, false starts and repetitions during their rendition of the source text that presented poor audio quality. The results of the data analysis, however, did not confirm the hypothesis as the interpreters' target texts showed no clear pattern in this regard.

Abstract – Deutsch

Die vorliegende Arbeit verfolgt das Ziel, die Auswirkungen von schlechter Tonqualität auf das Simultandolmetschen anhand der Häufigkeit von Unflüssigkeiten im Zieltext zu untersuchen. Der erste Teil der Arbeit bietet einen theoretischen Rahmen, der die in der Arbeit diskutierten Konzepte sowie die kognitiven Prozesse beschreibt, die im Gehirn von Dolmetschenden während des Simultandolmetschens ablaufen. Das erste Kapitel definiert grundlegende Begriffe für diese Arbeit und beschreibt die Entwicklung des Ferndolmetschens und des Remote Simultaneous Interpreting im Laufe des 20. und 21. Jahrhunderts. In diesem Kapitel werden ebenfalls Normen für die technische Ausstattung, die für das Simultandolmetschen notwendig ist, erklärt. Kapitel 2 erläutert die kognitiven Prozesse des Simultandolmetschens anhand Daniel Giles Effort Model (2009). Ebenfalls werden in diesem Kapitel mehrere empirische Studien beschrieben, die bereits den Zusammenhang von schlechter Tonqualität und kognitiven Prozessen beim Simultandolmetschen erforscht haben. Kapitel 3 beschreibt weitere Faktoren, die einen Einfluss auf die Dolmetschqualität haben und teilt diese Faktoren in drei Kategorien ein: ausgangstextbezogene Faktoren, sprecherabhängige Faktoren und Umweltfaktoren. Im vierten Kapitel werden Unflüssigkeiten genauer definiert und bereits durchgeführte Studien in Hinblick auf die Produktion von Unflüssigkeiten unter verschiedener Dolmetschbedingungen besprochen.

Der zweite Teil dieser Masterarbeit beschreibt ein Experiment, das sich zum Teil an Gervers Experiment (1974) stützt und den Einfluss von schlechter Tonqualität auf die Zieltextqualität testet. An dem im Rahmen dieser Masterarbeit durchgeführten Experiment nahmen sechs fortgeschrittene Dolmetschstudierende des Zentrums für Translationswissenschaft der Universität Wien teil. Sie wurden gebeten, ein neunminütiges Video aus dem Deutschen ins Englische zu dolmetschen, das teils gute und teils schlechte Tonqualität enthielt. Die Dolmetschungen wurden anschließend transkribiert und die Unflüssigkeiten analysiert. Die Hypothese war, dass die Dolmetschungen des Videoausschnittes mit schlechter Tonqualität häufiger Unflüssigkeiten in der Form von stillen Pausen, Häsitationslauten, Selbstkorrekturen, Fehlstarts und Wiederholungen aufweisen. Die Datenanalyse konnte diese Hypothese jedoch nicht bestätigen. Die Auswertung der Dolmetschungen zeigte hinsichtlich der Häufigkeit von Unflüssigkeiten kein klares Muster.