

Chapter 6

Present? Remote? Remotely present! New technological approaches to remote simultaneous conference interpreting

Klaus Ziegler

AIIC Technical Committee

Sebastiano Gigliobianco

SDI München

Since the 1970s, there have been several approaches to test and implement remote interpreting as a complementary interpreting modality in addition to the traditional and proven interpretation on site. The reasons for experimenting with remote interpretation in conference settings are manifold and can generally be classified by economic aspects, availability issues or organizational matters. In this paper, we discuss the preliminary results of a pilot study aimed at exploring how the limitations of remote interpreting described by the literature could be overcome using new technological advances in Information and Communication Technology. We discuss challenges and technological solutions for remote simultaneous conference interpreting from an interdisciplinary perspective and sketch out what the future workspace for conference interpreters might look like.

1 Introduction

Since the 1970s, there have been several approaches to test and implement remote interpreting as a complementary interpreting modality in addition to the traditional and proven interpretation on site, with all the parties involved (speakers, audience and interpreters) being present in the same room, thus communicating in a more or less face-to-face scenario. The reasons for experimenting with remote interpretation in conference settings are manifold and can generally be classified by economic aspects (e.g. reduced travel costs for interpreters and/or



speakers and/or audience), availability issues (e.g. no local interpreters available for a specific language combination) or organizational matters (e.g. an interpreter team can be hired within a shorter period of time, the room design does not allow for interpreting booths or booths are simply not wanted to be seen in the room).

Despite the encouraging results of studies, tests and experiments carried out throughout history, e.g. at UNESCO in 1976 (Kurz 2000), the United Nations in 1978 (Chernov 2004), in 1982 (UNESCO 1987), and in 2001 (Mouzourakis 2006), the European Union in 1992 (Kurz 2000), in 2001 (European Parliament Interpretation Directorate 2001) and 2005 (Roziner & Shlesinger 2010) using different technologies to transmit of audio and video signals from and to interpreters, there have always been two main factors preventing the large scale implementation of remote conference interpretation: technological limitations (due to insufficient availability of bandwidth for the synchronized transmission of sound and image with the necessary quality when transmitting via the Internet or telecommunication network, or very high costs when using satellite communication, either exclusively or along with terrestrial transmission technologies) and the more or less general refusal of the use of the so-called “new technologies” by conference interpreters. Apart from measurable physiological factors, like fatigue and stress leading to symptoms such as headaches and concentration problems, interpreters used to complain about the unease they were feeling because of not “being there” Mouzourakis (2006: 56), not having the possibility to get the right feel for the situation and not being able to interact directly with the other participants of the event. These psychological symptoms were mainly attributed to the limited view of the speaker and the audience.

In the last few years, general conditions for conference interpreting have been changing constantly not only due to globalization and altered market needs, but also due to digitalization and extremely fast developing information and communication technologies. The availability of hardware and software for dynamic monitoring and controlling of important parameters, such as lip synchronization, latency, video resolution and frequency response, as well as network infrastructures that allow for simultaneous transmission of high definition video and high quality audio signals via the Internet, combined with latest video, virtual reality and augmented reality (AR) technologies might offer possibilities to overcome existing technological, physiological and psychological problems.

2 Interdisciplinary and terminological challenges

One major challenge when discussing “remote interpreting” as a method for the delivery of interpreting services is the fact that there are a lot of different con-

cepts being used in practice by the different stakeholders when referring to this method. We can observe that there is still no harmonized terminology being used even by technical experts and researchers. This linguistic phenomenon can be explained to a certain extent by the fact that interpreters, in general, are not experts in this technical field. Therefore, they tend to use technical concepts without knowing exactly what the technical background for certain scenarios is and what the implications of a certain technical setup are. Technically speaking, a video conference and video conferencing can be defined as

a live, visual connection between two or more people residing in separate locations for the purpose of communication. At its simplest, video conferencing provides transmission of static images and text between two locations. At its most sophisticated, it provides transmission of full-motion video images and high-quality audio between multiple locations (TechTarget Network 2017).

This definition shows that the same concept is being used for a wide variety of technical setups.

When it comes to including interpreters to facilitate the necessary translation between the languages spoken by the participants in a communicative event, things become even more diffused. The term ‘remote interpreting’, one of the most widely used concepts, covers a whole range of technologically different setups. These setups range from a traditional presence-based scenario where interpreters, main speakers and the audience are concentrated at one event location, and one or several secondary speakers are connected from a distance for a limited duration, to a situation where none of the actors within the triad, speaker-listener-interpreter, are at the same location as the others.

Braun takes up this terminological challenge by saying that

Two main uses of telephone and videoconference communication can be distinguished in connection with interpreting. One of these, remote interpreting (RI), refers to the use of communication technologies to gain access to an interpreter in another room, building, town, city or country. In this setting, a telephone line or videoconference link is used to connect the interpreter to the primary participants, who are together at one site. (Braun 2015: 1)

This definition excludes a setup where the participants are located at different locations. For this case, she introduces the concept of “teleconference interpreting to cover both telephone and videoconference communication” (Braun 2015: 2). Further on, she introduces as a separate term

teleconference interpreting to cover both telephone and videoconference communication (ibid.)

For disambiguity purposes, she then introduces the terms “telephone-based interpreting” and “videoconference-based interpreting” (ibid.), but adds that there are a lot of additional concepts being used in practice.

For the purposes of this article, we broaden up the perspective and adhere to the definition given in the recently published ISO 20108:2016, introducing the term of “distance interpreting” (with “remote interpreting” as admitted term), giving the definition of “interpreting of a speaker in a different location from that of the interpreter, enabled by information and communications technology (ICT)”.

Analyzing technological, communicative, cognitive, physiological and psychological aspects, it becomes clear very soon that every single dislocation of one of the emitting or receiving elements (speaker, listener or interpreter) to a different location (thus becoming a distant location) has a considerable impact on the technological setup and the components and transmission channels needed to enable communication between the different parties involved.

Furthermore, communicative aspects like verbal and non-verbal communication are altered by the rearrangement of the setting, as elements like gestures, facial expressions, cannot be perceived directly anymore, and have to be captured, transmitted and reproduced again at the distant site in order to be made available.

As for cognition, the cognitive load is also being influenced significantly by every alteration of the setting (Moser-Mercer 2005). Such a load is generated by additional, or at least altered, receptive and productive tasks related to the sub-processes of listening to a source text, processing its content and re-producing that content in the target language. Research has shown that cognition is intimately linked to physiological processes that take place in the human body, especially the brain as the controlling unit, and that a variation of acoustic and visual input to interpreters has an impact on vital functional systems such as the respiratory system and metabolism, leading to stress, early onset of fatigue and other phenomena (Moser-Mercer 2003).

Last but not least, there is also a wide range of psychological factors that have to be considered when approaching distance interpreting from an interdisciplinary point of view. In an ideal setting, communication taking place with all participants being at the same location, in the same room and with no obstacles impeding direct mutual perception, allows the participants in the communicative event to make use of at least four out of the five basic human senses: touch, smell,

sight, hearing. Most of the related parameters, such as neuronal stimulation for haptic feedback, composition of ambient air for smelling, audio frequencies for hearing and light waves for vision can easily be measured, quantified and evaluated. However, other phenomena such as energy fields (sometimes referred to as mental, astral/emotional and etheric/physical bodies) that affect the balance of both the physical body and the non-physical mind are much more difficult to capture and evaluate, although they might have an influence on human interaction and, by that, on communication.

Whereas many interesting aspects related to distance interpreting have been addressed based on other interpreting specializations, such as legal interpreting (see *AVIDICUS* projects 1–3), there have not yet been similar large-scale projects directly or indirectly related to distance simultaneous conference interpreting. However, several aspects have been studied on a smaller scale, such as stress and performance in remote interpreting (Moser-Mercer 2003; Roziner & Shlesinger 2010), perception of remote interpreting by interpreters (Mouzourakis 2006) or visual input (Rennert 2008; Luisetto 2016).

3 Solutions in the past

3.1 Tests and experiments: UNESCO and the UN

Since the seventies, due to the rapid development in telecommunication technology, big international institutions, such as the United Nations and the European Union have begun testing new interpreting solutions to reduce the cost of conferences (UNESCO 1987: 26).

One of the first experiments with remote interpreting took place in 1976 when the UNESCO organized its General Assembly in Nairobi. The interpreters, however, were asked to work from Paris, which was connected to the capital city of Kenya through a satellite connection which provided an audio-video connection quality equal to that of a standard TV broadcast (Mouzourakis 1996: 30). The interpreters were not satisfied with their performance and stated that they were more tired and stressed than usual (Kurz 2000: 294).

A second experiment with remote interpreting was organized in 1978 by the United Nations. The conference was held in Buenos Aires and there were interpreters working both in Buenos Aires and in New York, the latter received audio-video signals through a satellite connection. The results showed that it was possible for the interpreters working in New York to achieve high quality interpretation (UNESCO 1987: 26), although the interpreters working in Buenos Aires

were able to deliver a better performance due to their more intensive preparation and knowledge about the conference (Chernov 2004: 82–90).

Another experiment with satellite connection took place in Vienna in 1982 during the United Nations conference on the Exploration and Peaceful Use of Outer Space. At this conference, the interpreters worked in Vienna as well, but not in the same building (Andres & Falk 2009: 10). The communication was a success, but the interpreters complained about increased stress (UNESCO 1987: 26).

Although these experiments with satellite connection between the 70s and the 80s demonstrated that remote interpreting was possible, the huge costs and the increased level of stress led to the conclusion that this technology was too expensive and immature to be used yet (*ibid.* 26).

Thanks to the ongoing advancement of ICT and the growth of the European Union, new possibilities for remote interpreting arose. From the beginning of the 90s the European Union was the main driving force behind this development with the aim to create a central hub where the interpreters could work (Braun & Taylor 2011: 3).

One of the first positive results was achieved in 1999 in Vienna. ISDN technology was used during a UN Inter-Agency Meeting on Language Arrangements, Documentation and Publications, which was held in Geneva and interpreted from Vienna. During this experiment, the transmitted audio signals were based on a frequency going up to 7 kHz for the first time. The conference hall in Geneva was recorded by three cameras and the pictures were projected in Vienna using a projection screen with a transmission rate of 384 kbps (Mouzourakis 2006: 63). The audience was satisfied with the performance of the interpreters, and for the first time the interpreters were pleased with the quality of the audio transmission, although they criticized the video quality (Andres & Falk 2009: 11). Two months later, a second experiment was organized together with the International Telecommunication Union (ITU) and the École de Traduction et d'Interprétation (ETI). For this experiment, two French booths were installed: one in the conference hall and a second one in a remote location. In this case, the audio signal was encoded in MP3 format and the video signal was transmitted at a rate of 382 kbps (Mouzourakis 2006: 63). The audience received the signal from the local and remote booth alternately and was happy with the results. The interpreters, on the other hand, perceived the physical distance to the conference hall as negative and felt as if they were losing control over the situation. Interestingly, saliva samples were taken from both booths before and after the conference; contrary to what the interpreters stated, no relevant difference was noted in stress hormones (Moser-Mercer 2003).

A new experiment was conducted in 2001 in New York by the United Nations in which both ISDN and satellite connections were tested. The conference was recorded using three cameras and a 42- and 25-inch plasma screens were installed in front of the booths. With this experiment, the UN laid down the following minimum requirements for remote interpreting:

14 kHz sound (requiring 128 kbps) for sending floor sound to the booths (14 kHz) and 10 kHz sound (at 64 kbps) for sending interpretation back to the floor (10 kHz); 512 kbps for the image of the speaker plus 384 kbps for the floor/podium image. (Mouzourakis 2006: 63)

Two new studies took place in 2001 and 2005 at the European Parliament to test a connection through optic fibre (*ibid.* 64). The results of the first experiment showed that the interpreters were satisfied with the audio and video quality. However, they criticised the selection of pictures and the fact that they did not have a comprehensive view of the conference room (European Parliament Interpretation Directorate 2001: 19–21). Moreover, they stated that they felt uncomfortable and that the remote interpretation setting was overall more tiring (*ibid.* 22–23). The second experiment, which lasted five weeks, brought similar results: the interpreters were pleased with the audio and video quality although a complete view of the audience was missing. Moreover, they felt alienated and isolated from the conference. The screens caused eye-burning, headaches, lower concentration and higher tension and fatigue. They also stated that they felt their performance to be of inferior quality while working remotely. Medical examinations, however, found no evidence of increased stress, and a performance evaluation confirmed that the interpreters' remote performance was slightly inferior, but not enough to reach statistical relevance (Roziner & Shlesinger 2010: 225–243).

4 Current solutions

4.1 LAN-based solutions

An example of a remote interpreting solution, which is based on a local network is currently being used by the Directorate General for Interpretation of the European Commission (DG SCIC). In this setting, interpreters work from a conventional permanent booth located in another room at the same location, although mobile booths are also used occasionally, in front of which 4 high-resolution screens are placed. Interpreters received the following images via the screens:

- on a 50-inch full HD screen with split view, an overview of the entire meeting table from two different angles is shown; thanks to the big screen and the high resolution, the interpreters are able to see all the participants' faces
- an image of the active speaker is transmitted on two 22-inch full HD screens which are placed laterally to the 50-inch screen
- a static shot of the presidency is displayed on a third 22-inch full-HD screen which is placed on top of the 50-inch screen

In order to transmit the images from the meeting room to the interpreting room, optic fibre cables are used (Technical Compliance Team, DG Interpretation, European Commission 2016).

4.2 External-data-transmission-based solutions

With the growing demand for a more flexible and less costly delivery of interpreting services in a globalized world, several attempts were made in the last decade to make use of technologies, thus overcoming the restrictions of the traditional presence-based scenario with a direct wired connection of all components of the conference system (see §3). For the purpose of this article and from a merely technological point of view, these technologies can be divided into solutions including the transmission of either audio signals only, or the transmission of both audio and video signals to the interpreters.

4.2.1 Audio conferencing solutions

One approach to overcoming existing limitations of bandwidth and thus the impossibility for transmitting sound and image to the interpreters with the required quality (see §3) consists of reducing the transmitted content exclusively to audible signals. The most accessible and therefore most frequently used technology for this transmission relies on the use of telephony. In technical terms, a first rough distinction has to be made between traditional landlines and mobile telephony, as the audible frequencies are mainly being transmitted either via wired connection (landline) or as waves through the air. The use of copper cable and the non-existence of fibre optic connections in certain areas, but also the necessity of handling billions of those connections at a time, still lead to a considerable decrease of the audio frequencies being transmitted to a maximum of 8 kHz, even

if landlines today are more often based on digital VoIP technology.¹ Whereas for consecutive interpreting the range of up to 8.000 Hz might be enough (often named ‘wideband audio’) because of the clear separation of the reception (hearing) and production (talking) processes on the interpreter’s side, simultaneous interpreting with listening and talking at the same time is considerably affected by the loss of frequencies due to the masking effect generated by the interpreter’s own voice, making incoming frequencies on the same level not audible anymore (Jumpelt 1984). As a result, the physical and cognitive effort of listening will increase. Depending on the frequency of the incoming voice of the speaker and the interpreter’s own voice (note that e.g. female and male voices are located in different frequency ranges), the interpreter’s output quality will necessarily decrease despite all efforts to compensate for the missing input by interpreting strategies like variation of the voice-to-ear-span or additional pausing. It should not be forgotten that the interpreter has no possibility to compensate for the absence of audible input by visual input when using audio-only solutions. Even if these technologies have improved a lot and specially designed audio bridges allow controlling incoming and outgoing signals with a high level of quality of service even in multilingual scenarios, the above described acoustic parameters cannot be neglected.

4.2.2 Video conferencing solutions

As far as solutions for transmitting sound and image at the same time are concerned, one of the most frequently used technologies is based on video conferencing solutions. These solutions rely on the principle of separately capturing image and sound at the source with camera and microphone, encoding the data with a certain algorithm, transmitting the data packages using one or more parallel lines and decoding the packages again at the destination, making sound and image audible and visible again. These solutions necessarily depend on the availability of adequate and compatible equipment at all source and end points of the communication and the use of a common standard for encoding and decoding of signals (codec). One of the critical points with these video conferencing solutions in terms of quality of simultaneous interpreting is the fact that standard codecs use a compressed file format for transmission of audio signals and compression is carried out either by cutting out certain frequencies, or by other procedures causing similar effects as described before for the audio only solutions.²

¹For a comparison of VoIP coding algorithms, see Singh & Mian (2016).

²For more information on techniques for video compression, see Wiegand et al. (2003).

5 Technical requirements for remote interpreting

As described above, remote interpreting is a specific method of (conference) interpreting and covers a variety of scenarios of a speaker at a different location from that of the interpreter, enabled by information and communication technology. The main aspect in these scenarios is that the interpreter is not physically present in the same room as all the other actors of the communicative event, thus not having a direct view of either the conference room the conference room, the speaker and/or the audience. To enable interpreters to perform adequately and deliver a quality service, some technical requirements need to be fulfilled. As we have seen before, different experiments have been carried out in order to identify the minimum technical requirements to assure adequate performance quality.

One of the first attempts to define minimal requirements for remote interpreting was the *Code for the use of new technologies in conference interpreting*, which was published by the International Association of Conference Interpreters (AIIC) in 2000 together with other national interpreters' associations, the European Parliament, the European Court of Justice and the joint conference and interpreting service of the European Union and the World Customs Organisation (Korak 2010: 31). AIIC states that interpreters' work conditions in remote settings must comply with the requirements set out in ISO standards 2603 and 4043 (1998 editions, withdrawn in 2016 and replaced with a new set of ISO standards 2603, 4043, 20108 and 20109) which define the work environment of interpreters in mobile and permanent booths. Moreover, the following requirements are to be fulfilled:

- all frequencies between 125 Hz and 12500 Hz are to be transmitted
- interpreters must receive high definition images of the speaker and other participants
- the interpreter shall work no more than two hours per day (AIIC 2000: 2)

A more comprehensive study commissioned by the European Commission was conducted in 2010 by the *Fraunhofer Institut* to evaluate the minimal requirements of video and audio quality for simultaneous interpretation. During this study, conference interpreters were asked to evaluate different audio and video signals to assess the impact of transmission quality on their performance. This study resulted in a guideline:

- all frequencies between 125 Hz and 12500 Hz must be transmitted, although frequencies starting at 75 Hz should also be included in the range

- video quality should be at least 1280×720 at 50 Hz with a ratio of 16:9
- audio must be synchronized (lip synchronization) with the video track with a maximum value of -25 ms or $+95$ ms

The work of the *Fraunhofer Institut* was used as a starting point to draft the new ISO-Standards 20109 “*Simultaneous interpreting – Equipment – Requirements*” and 20108 “*Simultaneous interpreting – Quality and transmission of sound and image input – Requirements*”, which not only raised the minimum requirements, but also added new ones concerning sound and image transmission:

- all frequencies between 125 and 15000 Hz ± 3 dB must be transmitted (ISO 20109:2016: 3)
- image quality must be good enough to avoid blurring and freezing of the video³
- audio must be synchronised with the images with a maximum delay of 45 ms or advance of 125 ms
- latency (from the source to the interpreters) must be lower than 500 ms (ISO 20108:2016: 7–8)

6 Future workspace

Based on the technical parameters described in §5, the authors started an experimental research project with the objective of modelling a future workspace for conference interpreters while performing remotely in simultaneous mode. The framework that was chosen for the experimental study was based on a standard working environment for conference interpreters, including, amongst others, a soundproof simultaneous interpreting booth and a hardware interpreter’s console as an audio/video interface. The experimental setting was also characterized by some specific markers in terms of interactivity and communication patterns, generally assigned to conference interpreting in literature, such as a mostly monological discourse pattern, interpreting into one language, no possibility for

³The draft of the ISO-standard 20108 contained more detailed specifications regarding video and transmission requirements: video quality must be at least 720p at 50 Hz or 1080p at 25 Hz, and the signal must be compressed using at least H.264 at 1152 kbps. Moreover, the packet loss value should not exceed 0.2%, jitter should be lower than 15 ms and the latency (roundtrip) in the system shall not exceed 200 ms.

interaction with the speaker during his/her intervention, a symmetric communicative setting with speakers on the same educational level, the same linguistic code used by both speakers, and no variation of speaker's registers (Angelelli 2000: 582–583). All tested scenarios in the experiment were designed for possible implementation in conference interpreting hubs with a basic technical setup similar to the one that can usually be found at conference venues.

6.1 The experiment

The experiment took place in Düsseldorf in collaboration with PCs Professional Conference System GmbH. This experiment aimed to test three different remote interpreting settings:

- using a 65-inch screen with a picture-in-picture function
- using a camera remotely controlled by the interpreter
- using a 360-degree camera and virtual reality glasses

Common to all three parts of the experiment was the text which the interpreters had to interpret in simultaneous mode. Two speakers were asked to talk for a total length of about 10 minutes per scenario. The dialogue was not prepared, but rather improvised and the only rule the speakers had was that the first five minutes of the speech had to be informal and the last five of a more technical nature. According to the Effort Model of Gile (2009) interpreters need to distribute their concentration among the different tasks of simultaneous interpreting, the different degree of difficulty of the speeches aimed to test whether interpreters were still able to operate the new device in the booth under different degrees of stress. The interpreters were not informed about the nature of the text they would have to interpret; this was done to eliminate preparation as a variable from the equation and to assure that they would concentrate on the interpreting effort as well. Moreover, since the interpreters had direct control of the video signals being transmitted, two speakers were intentionally selected in order to force the interpreters to adjust the video settings based on which speaker was talking in order to assure that the new device was actively used during the experiment. The dialogue was held in German and each of the two interpreters was working from their A into their B language. Since the length of each dialogue was only ten minutes and pauses between the scenarios were planned, each interpreter worked alone in the booth while the other was waiting in a separate room to guarantee equal conditions. Moreover, having each interpreter working alone assured that they were forced to operate the additional technological feature while

interpreting to determine whether and to what extent this would have an impact on their performance. After the experiment, the interpreters were asked to fill out a questionnaire to evaluate the different settings and technological setups. The questionnaire consisted of 28 questions divided into two parts: first, a more generic part to create a profile of the subjects with questions about their age, experience, language combination and whether or not they already had experience with remote interpreting, and then a second more specific one, in which for each scenario the interpreters were asked to summarize their personal opinion about these new solutions. For example, they were asked whether they could operate the extra devices in the booth, whether they felt they had to put extra effort into it and whether these new solutions were better or poorer than a standard video and audio transmission which they could not control. Moreover, for each scenario they were asked to evaluate the quality of the video transmission on a scale from 1 (poorest) to 5 (best), to share any personal opinion, comment, or criticism and finally to state whether or not they would like to see this solution implemented in the future.

Both the subjects in this study were professional conference interpreters holding a degree in conference interpreting. Both started their career more than 16 years ago and stated having worked as conference interpreters between 51 and 100 days per year, and can therefore be qualified as ‘experienced’ professional interpreters. In terms of the internal validity of the study, it is also important to mention that they stated having had very little or no experience at all with remote interpreting in conference settings. Due to the very small sample of this study, the results of this experiment cannot be generalized. The primary aim of this pilot study was to test the feasibility of these new technologies. To achieve external validity, the study would have to be reproduced on a larger scale and conducted with an appropriate methodological approach.

6.1.1 Picture-in-picture

The picture-in-picture solution was the first scenario being tested. For this part of the experiment, the speakers were in the conference room, in which two different cameras, each pointed at one speaker, were recording the event. The audio signal was recorded with a wireless microphone and fed together with the video signal to the Extron SMP 351 recording and streaming processor, which streamed the audio and video tracks over the local network of PCs to the interpreting booth. To guarantee lip synchronisation, the audio signal was delayed by 275 ms before being fed to the Extron. In another room, a 65-inch monitor was placed 95 cm from the booth, guaranteeing a distance of roughly 155 cm between the screen and the eyes of the interpreters. According to Causo’s guidelines (2011: 2), the

booth was placed in such manner that the window' frames of the booth would not obstruct the view of the interpreters. The video signal from the conference room was shown to the interpreters using a picture-in-picture function, meaning that both signals were transmitted simultaneously and that one image occupied a larger portion of the screen while the second video feed was displayed in a smaller format in the bottom right corner and the interpreters were able to switch between the two. According to the draft of the ISO standard 20108⁴ (2016: 6), the video signal was reproduced with a resolution of 1920×1080 pixels at 30 frames per seconds (fps).

The results showed that both participants reacted positively to this solution, and they found it to be better than a static video signal, upon which they have no influence. None of the interpreters noticed any increase in workload due to the more technical nature of the second half of the text and stated that they had no problem in using the new device in the booth to switch between the two pictures. Both interpreters stated that they would like to see this solution implemented in the future.

6.1.2 Remote-controlled camera

During the second part of the experiment, the speakers were recorded by only one camera placed in the middle of the conference table. The interpreters were able to remotely move the camera 270 degrees horizontally and 90 degrees vertically on its axes from within the booth. The camera was directly connected to the internal local network of PCs and the interpreters were able to move it using a mouse connected to a laptop in the booth. The audio signal was captured separately using a wireless microphone and delayed by 275 ms before being transferred to the interpreter's booth using a cable. With a resolution of 1080p by 25 fps, this camera respected the indications of the draft of the ISO standard 20108 as well.

The results of the questionnaires showed that this solution was also welcomed by the interpreters. They quickly got used to the control of the camera via mouse and stated that this could rapidly become an automatism. Furthermore, this solution was considered to be better than a non-controllable video transmission. None of the interpreters noticed the difference between the first and second part of the dialogue and had no problem with the camera control.

⁴When the experiment was conducted, the final version of the ISO-Norm 20108 had not been published yet. The draft was therefore used for reference.

6.1.3 Virtual reality VR glasses

For the last part of the experiment, the interpreters still worked in a normal interpreting booth wearing Elegiant VR-glasses which use a smartphone (in this case an Apple iPhone SE) to reproduce images. The smartphone was directly connected through Wi-Fi to the 360-degree camera used to capture the dialogue between the two speakers. The audio was captured with a wireless microphone and fed to the interpreting booth using a cable after being delayed by 500 ms. In order to fully exploit the potential of the camera and the VR glasses, the speakers stood in front of each other in a big hall and were asked to move around while talking as well as making use of gestures and mimics.

The interpreters had very different opinions on this solution. The first interpreter found the VR glasses an interesting alternative to a monitor although not very comfortable to wear for a longer period. The subject also stated that, after a short time, one gets used to blindly operating the console in the booth. The second interpreter, on the other hand, was not happy with this solution, finding it tiring for the eyes – it must be noted that the second interpreter wore glasses and because of the design of the VR glasses it was not possible to wear both correction and VR glasses – and complained that it was very difficult if not impossible to operate the console in the booth or use a personal computer for terminology research.

At this point, the authors need to clarify that especially this last part of the test was of extremely experimental nature and that the results are not particularly conclusive, also due to the small number of participants. The main goal of this experiment, however, was not to test whether the selected model of VR glasses is ready for a practical application in interpreting booths, but rather to make a first test on operability of VR glasses combined with traditional equipment for simultaneous interpreting in soundproof booths. The results of this test shall be used for additional and more elaborate studies in the future.

6.2 From virtual reality to augmented reality solutions

The experimental study described in §6.1 suggests that the lack of a direct view of the speaker during an event might be compensated for, at least to a certain extent, by the use of technologies that allow the interpreter to control the video input if they feel the need to process non-verbal elements of the content produced by the speakers. Nevertheless, a one-dimensional screen reproducing one or several images will always reduce the possibilities of perception of the setting captured with cameras, as there is still a very clear separation of the interpreter in their

remote environment and the room where the original event is taking place. The feeling of separation from the action, often referred to as the feeling of not “being there”, might still be big enough to prevent the interpreter from overcoming this psychologically relevant issue.

The last step in the experiment described above was to dive into a virtual reality scenario, making the interpreter feel immersed in the situation, e.g. making similar movements with their eyes and/or their head and body as they would be doing if seated in the room with the speaker and the audience, looking for the necessary visual information to complete the audible content and to render the entire intended message to the listeners. The use of a screen with a double video feed, controlled by the interpreter and reproduced as a picture-in-picture image on a wide screen was accepted by both subjects in the experiment and the additional control task didn't seem to have any negative effect in terms of cognitive overload. However, the use of virtual reality glasses clearly showed that the physical separation from the real world and the traditional equipment placed in it (console for controlling audio input and output, and laptop for document and knowledge management) raises several cognitive and ergonomic issues, although the much more dynamic and self-controlled setting was considered positive as such.

Following this line, the combination of both real and virtual elements in an augmented or mixed reality scenario, where computer-generated images are superimposed on the user's view of the real world, would be the next logical step towards a practical solution for the challenges that remote simultaneous interpreting imposes. AR glasses such as Microsoft's HoloLens, for example, allow projecting images and possibly other virtual images into the vision field of the interpreter, while they can still see and control the real hardware components they need for the delivery of the interpreting service. Any software application used for document and knowledge management could be moved from the real world into the virtual world, projecting only the image of the respective interface into the field of vision and allowing the interpreter to virtually manipulate the application with their hands. Hence, ergonomic aspects such as weight and wearing comfort of the device, cognitive aspects such as real or perceived additional workload and, of course, compliance with the technical parameters set out for distance interpreting, as well as the processing capacity of the processing unit will have to be studied in terms of usability and feasibility.

6.3 Interpreting hub solution

Workspace for conference interpreters working remotely in simultaneous mode should take into account the relevant parameters as stated in existing ISO standards 2603/4043, as far as sound insulation, ventilation and ergonomics are concerned, ISO 20109, as far as equipment for simultaneous interpreting is concerned, and ISO 20108, as far as quality and transmission of sound and image input to interpreters is concerned. Therefore, workspace solutions for distance interpreting should be designed accordingly, assuring that the most important parameters that allow for quality simultaneous interpreting are met.

Even if for economic reasons a solution with the interpreter working from their home office might seem the most obvious and easiest solution, there are several issues that need to be considered when envisaging such solutions. First of all, a dedicated internet connection with assured availability of the necessary bandwidth for transmitting high definition images together with ISO-compliant sound to the interpreter is either not available or hardly affordable for individuals. In addition, home office rooms would have to be equipped either with an appropriate interpreting booth or with components assuring compliance with the main parameters of ISO 2603/4043 in terms of e.g. insulation and ventilation. A lack of possibilities for dynamic control of the Internet connection while interpreting, as well as data protection and confidentiality issues would also have to be resolved if home office workspace were to be used.

Particularly for multilingual events with more than two languages being spoken and interpreted at the same time, a promising approach in terms of Quality of Service both from technical and interpreting performance views are solutions with interpreting being performed in specially designed hubs, where interpreters would find all the necessary working conditions to deliver quality interpreting. Traditional ISO compliant booths as a minimum standard, or a specifically designed workspace ensuring the basic requirements. This should be equipped with state-of-the-art digital interpreting equipment, dedicated internet connections with permanently available bandwidth, controlled ambient conditions with active regulation of air supply and carbon dioxide levels as well as immersive 3D environments with high resolution projection screens. Interpreters should be able to choose the desired visual input out of several video feeds, AR components; technical support available on site would set the appropriate technical framework. In addition, offices with access to online and offline information sources for preparation before and during the event, rest rooms for relaxation and lounge areas for the necessary professional and social exchange within the interpreting team, would help to overcome the alienation perceived by interpreters working

in remote scenarios and environments that have not been developed specifically for this purpose.

Apart from these more technical aspects, any kind of hub solution would also have to address psychological factors such as measurable or perceived stress. Moser-Mercer (2005: 15) states that “it appears that (...) interpreters seem to be under increased psychological stress when working away from the conference room, mostly because they experience a lack of control of the situation”. Even if there are no studies available yet that would explain exactly which kind of control interpreters are missing,⁵ aspects such as those given below are what interpreters feel they need to control

- availability of a technician in case of technical problems
- the interaction with team mates working in the same interpreting booth or across booths in multilingual conferences
- the (sometimes very limited or even non-existing) possibility of talking to speakers and audience before the conference or during breaks
- self-control of the direction of sight, focusing on speaker audience, presentation or any other visual input available

The design of an interpreting hub as described above could easily cover the first two aspects. A hub would need the presence of a technician onsite in the hub to manage the technical equipment and the whole team (if interpreting takes place from one hub only) or at least the booth mates would be working together at the same place. The interaction and communication with speakers and audience would, of course, require a specific solution, but could be handled, e.g. considering ‘institutionalized’ briefing and Q&A sessions before, during and after the meeting.

Control of the view on the speaker, audience and additional visual input, such as presentations, in presence-based interpreting usually performed by head movements and eye focusing would need further development based on the technologies described in §6.1. Considering the rapid development of image captioning, transmission and reproduction, it seems only a matter of time until these technologies will be affordable and adaptedn to the needs of remote simultaneous conference interpreting.

⁵Note that, apart from the interpreter’s console with standard control buttons for the incoming sound, an interpreter in a presence-based conference setting also has only limited control of the actions in the room.

7 Conclusions

In times of growing demand for flexible, accessible and customer-oriented digitalized communication services in a globalized world, technological solutions for quality simultaneous interpreting services will have to be developed, taking into account several aspects related to the organization and delivery of those services. Existing standard solutions developed for other purposes, such as video conferencing or web conferencing without interpreting, or solutions for consecutive interpreting in specializations other than conference interpreting appear not to be sophisticated enough to meet the special requirements that distance interpreting imposes in terms of sound and image input to interpreters in bilingual and multilingual conference settings. Technological enhancements in the field of virtual reality and augmented reality, as well as immersive communication environments, may offer the possibility to overcome existing constraints.

One of the major challenges for interpreting studies in the field of distance interpreting will be finding a more interdisciplinary and future-oriented approach, building teams of researchers in the technical, medical and psychological field, to name only a few of them, and to combine these different research disciplines in multidisciplinary projects that can actively lead to designing the future workspace for conference interpreters in the first place, and for other specialized interpreting services as well. It goes without saying that fellow (conference) interpreters need to be prepared for remote simultaneous interpreting during their training, as this modality is experiencing a growing demand in different interpreting specializations, including tele- and videoconference interpreting (Braun 2015). Apart from the integration of training modules designed specifically for this modality, addressing cognitive, communicative and technical aspects, amongst others, this would also require adequate equipment of training facilities with the appropriate features (audio/video conferencing hard- and software and connectivity, to mention just a few of them).

References

- AIIC. 2000. *Code for the use of new technologies in conference interpretation*. http://www.staff.uni-mainz.de/fantino/class/files/cai/aiic_code_interpreting%20technologies.pdf, accessed 2017-10-31.

- Andres, Dörte & Stefanie Falk. 2009. Information and communication technologies (ICT) in interpreting: Remote and telephone interpreting. In Dörte Andres & Sonja Pöllabauer (eds.), *Spürst Du wie der Bauch rauf-runter? Fachdolmetschen im Gesundheitsbereich* (InterPartes), 9–27. München: Martin Meidenbauer.
- Angelelli, Claudia. 2000. Interpretation as a communicative event: A look through hymes' lenses. *Meta* 45(4). 580–592.
- AVIDICUS project. 2008–2011. *AVIDICUS 1 (2008–11)*. http://wp.videoconference-interpreting.net/?page_id=13, accessed 2018-5-12.
- AVIDICUS project. 2011–2013. *AVIDICUS 2 (2011–13)*. http://wp.videoconference-interpreting.net/?page_id=16, accessed 2018-5-12.
- AVIDICUS project. 2014–2016. *AVIDICUS 3 (2014–16)*. http://wp.videoconference-interpreting.net/?page_id=154, accessed 2018-5-12.
- Braun, Sabine. 2015. Remote interpreting. In Holly Mikkelson & Renée Jourdenais (eds.), *Routledge handbook of interpreting*. Abingdon/New York: Routledge.
- Braun, Sabine & Judith L. Taylor. 2011. Video-mediated interpreting: An overview of current practice and research. In Sabine Braun & Judith L. Taylor (eds.), *Videoconference and remote interpreting in criminal proceedings*, 25–57. Guilford: University of Surrey.
- Causo, José Esteban. 2011. Technical guidelines for remote interpretation. In Sabine Braun & Judith L. Taylor (eds.), *Videoconference and remote interpreting in criminal proceedings*, 301–306. Guilford: University of Surrey.
- Chernov, Ghelly V. 2004. *Interference and anticipation in simultaneous interpreting: A probability-prediction model*. Amsterdam/Philadelphia: John Benjamin.
- European Parliament Interpretation Directorate. 2001. *Report on remote interpretation test, 22-25 January 2001*. Brussels. http://www.europarl.europa.eu/interp/remote_interpreting/ep_report1.pdf, accessed 2017-12-29.
- Gile, Daniel. 2009. *Basic concepts and models for interpreter and translator training*. Amsterdam/Philadelphia: John Benjamins.
- ISO 20108:2016. 2016. *Simultaneous interpreting – quality and transmission of sound and image input – requirements*. Standard. International Organization for Standardization.
- ISO 20109:2016. 2016. *Simultaneous interpreting – equipment – requirements*. Standard. International Organization for Standardization.
- ISO 2603:2016. 2016. *Simultaneous interpreting – permanent booths – requirements*. Standard. International Organization for Standardization.
- ISO 4043:2016. 2016. *Simultaneous interpreting – mobile booths – requirements*. Standard. International Organization for Standardization.

- Jumpelt, Walter. 1984. Die Arbeitswelt des Konferenzdolmetschers nach neuen ISO- und IEC-Normen. *Lebende Sprachen* 29(1). 1–5.
- Korak, Christina. 2010. *Remote interpreting via Skype: Anwendungsmöglichkeiten von VoIP-Software im Bereich Community Interpreting: Communicate everywhere?* Berlin: Frank & Timme.
- Kurz, Ingrid. 2000. Tagungsort Genf/Nairobi/Wien: Zu einigen Aspekten des Teledolmetschens. In Midra Kadric, Klaus Kaindl & Franz Pöchhacker (eds.), *Translationswissenschaft: Festschrift für Mary Snell-Hornby zum 60. Geburtstag*, 291–302. Tübingen: Stauffenburg Festschriften.
- Luisetto, Martina. 2016. Visueller Input beim Remote Interpreting: Eine empirische Untersuchung zum Einfluss des visuellen Inputs auf die Qualität des Dolmetschprodukts beim Remote Interpreting anhand des Sprachenpaars Deutsch-Italienisch. In Ursula Gross-Dinter (ed.), *Dolmetschen 3.0 – Einblicke in einen Beruf im Wandel*, 51–92. Berlin: Frank & Timme.
- Moser-Mercer, Barbara. 2003. *Remote interpreting: Assessment of human factors and performance parameters*. <https://aiic.net/page/1125/remote-interpreting-assessment-of-human-factors-and-pe/lang/1>, accessed 2017-10-31.
- Moser-Mercer, Barbara. 2005. Remote interpreting: Issues of multi-sensory integration in a multilingual task. *Meta* 50(2). 727–738.
- Mouzourakis, Panayotis. 1996. Videoconferencing: Techniques and challenges. *Interpreting* 1(1). 21–38.
- Mouzourakis, Panayotis. 2006. Remote interpreting: A technical perspective on recent experiments. *Interpreting* 8(1). 45–66.
- Rennert, Sylvi. 2008. Visual input in simultaneous interpreting. *Meta* 53(1). 204–217.
- Roziner, Ilan & Miriam Shlesinger. 2010. Much ado about something remote: Stress and performance in remote interpreting. *Interpreting* 12(2). 214–247.
- Singh, Hardeep & M Mian. 2016. Comparative study and analysis of various VoIP coding algorithms. *International Journal of Computer Applications* 141. 6–10.
- Technical Compliance Team, DG Interpretation, European Commission. 2016. *Internal archives of the European Union*.
- TechTarget Network. 2017. <http://searchunifiedcommunications.techtarget.com/definitions/page/7>, accessed 2017-10-31.
- UNESCO. 1987. *Management of interpretation services in the United Nations system. Report of the United Nations joint inspection unit*. <http://unesdoc.unesco.org/images/0007/000732/073286eo.pdf>, accessed 2018-5-12.
- Wiegand, Thomas, Gary J. Sullivan, Gisle Bjøntegaard & Ajay Luthra. 2003. Overview of the H.264/AVC video coding standard. In IEEE Transactions (ed.), *Circuits and systems for video technology*, 560–576.