Distance conference interpreting

Kilian G. Seeber and Brian Fox

Definitions

Distance interpreting (DI), defined by the International Standardization Organization as, “interpreting of a speaker in a different location from that of the interpreter, enabled by information and communications technology (ICT)” (ISO 2017a: 2), is a blanket term covering different technical arrangements enabling multilingual events whereby participants—including interpreters—need not share the same physical space. Importantly, the notion of distance is not specifically defined and can range from a few metres (when interpreters are located in a room adjacent to the meeting room) to a few thousand kilometres (when interpreters are located in a different country or on a different continent). Although the nomenclature used to refer to specific types of distance interpreting set-ups has varied over the years, recently there has been a convergence on four discrete notions, as they capture what have been identified as the principal features of DI (Braun 2015; Seeber 2018). The International Association of Conference Interpreters (AIIC) defines four types of distance interpreting (AIIC Executive Committee 2018): Videoconference Interpreting (VCI), Audioconference Interpreting (ACI), Video Remote Interpreting (VRI) and Audio Remote Interpreting (ARI). These four notions allow a distinction to be made based either on the medium of communication (i.e. whether the interpretation is based on an audio or audio-video signal) or on the distribution of the participants of an event (i.e. whether the interpreters share the same physical space of none or at least some of the conference attendees). Consequently, the former distinction puts the emphasis on a signal-processing aspect while the latter focuses on the interpreters’ workspace, although the two parameters are often intertwined, as our discussion will show. Depending on context, a distinction by medium of communication or by distribution of participants might be more relevant to the discussion. Importantly, while these four types of DI relate to distinct technical modalities enabling interpretation during multilingual events, they do not specifically address the issue of mode, meaning that in principle, they are applicable both to consecutive and simultaneous interpreting.
The medium of communication

Unlike sign language interpreting, which is intuitively associated with visual input, spoken conference interpreting is often seen primarily as the processing of auditory signals—after all interpreters listen and speak, as opposed to translators who read and write. This reductionist view is at stark odds with the importance practising conference interpreters attribute to visual input (AIIC Executive Committee 2018) and with research findings corroborating the impact of visual input during language processing in general, and simultaneous interpreting in particular (Rennert 2015; Seeber 2017; but see Gieshoff 2018, discussed in Chmiel, Chapter 33, in this volume). Seeing the impact that the medium of communication can have on the interpreting process, it seems warranted to differentiate DI modalities according to the medium used. This would then allow for a distinction between audio-mediated interpreting and video-mediated interpreting, where the former includes audioconference interpreting and audio remote interpreting, and the latter includes videoconference interpreting and video remote interpreting (see Figure 35.1).

The distribution of participants

The location of the different participants (speakers, listeners, interpreters) has an impact on the type of visual input interpreters receive along with the medium of communication. Traditionally, this parameter was synonymous with the location of (permanently installed or mobile) interpreting booths and the resulting impact on the interpreters’ ability to see speakers, their props and their audience (allowing for the processing of additional visual input and feedback). With the introduction of DI modalities, however, the question about who has a direct view of what has become more nuanced, as the meeting room is no longer the only source of the conference proceedings. It is therefore not inconsequential to differentiate between teleconference interpreting and remote interpreting. The former includes audio and video-conference interpreting, implying that interpreters share the physical space of at least some (usually the majority) of the conference participants. The latter comprises audio and video remote interpreting.
Distance conference interpreting

interpreting, and refers to scenarios where the interpreters are not located with any other conference participants, and indeed, not even necessarily with each other (see Figure 35.2).

Other terminology

While the aforementioned nomenclature is comprehensive in that it describes all major types of DI regardless of the interpreting mode, i.e. consecutive and simultaneous, the term Remote Simultaneous Interpreting (RSI) was used as early as 1998 following an experiment at the Stanford University School of Medicine. Although this term has been widely adopted by Simultaneous Interpreting Delivery Platforms (SIDPs), it makes no distinction between systems providing audio only as opposed to audio-visual input for conference participants including interpreters. For all intents and purposes, RSI could be seen as a subcategory of ARI or VRI, depending on the availability of visual input.

A short history of distance interpreting

Early conference interpreters, often diplomats and civil servants themselves, usually sat in the meeting room with the negotiating parties. It was not until the introduction of simultaneous interpreting equipment at the Geneva headquarters of the International Labour Organization and the Moscow-based Comintern in 1926 (Chernov 2016)—both based on re-purposed telephone and gramophone technology—that interpreters were relegated from the main podium to the regular seats and eventually, to a corner, often in the back of the room. This development was part of a much larger evolution of Information and Communication Technologies (ICTs). In 1927, for instance, AT&T’s Bell Labs successfully implemented a two-way audio and one-way video connection between Secretary of Commerce Herbert Hoover in Washington, DC, and AT&T president Gifford in New York (Senft 2019) while, in 1936, Leipzig and Berlin were connected through the first visual telephone system or Gegenseh-Fern sprechanlage (Deutsches Fernsehmuseum Wiesbaden 1938).

The first distance interpreting solutions were the telephone interpreting systems conceived in the late 1950s, such as Nestler’s (1957) Tel-interpret, which added a switch-array to an

Figure 35.2 DI modalities by distribution of communicators

Source: Adapted from Avidicus (2018).
existing phone line, giving interpreters control over the line and allowing them to add their interpretation to the flow of conversation. It would take more than a decade before telephone interpreting services and thus the first type of audio remote interpreting, would become a staple service for telecommunication providers in large, sparsely populated countries. In Australia, for instance, telephone interpretation was offered as a free service by the Department of Immigration, starting in 1973, while in the United States the company Communication and Language Line (later acquired by AT&T) started offering a paid telephone interpreting service in 1981 (Mikkelson 2003). Although telephone interpreting has continued to grow over the years, it is principally used in Public Service Interpreting (PSI) settings (Kelly 2008).

About the same time, the first large-scale teleconferencing experiment was organized by UNESCO with a satellite link connecting the conference centre in Nairobi to the organization’s headquarters in Paris for their General Conference in 1976. Teams of interpreters were located both in Paris and in Nairobi interpreting conference debates and simulated proceedings based on radio-transmitted audio feeds and so-called visiophone technology (UNESCO 1976). Consequently, both ARI and VRI were tested, showing the technical feasibility of remote interpreting, but also flagging some of the inherent challenges. The organization’s chief interpreter summarizes these in his report, suggesting a direct link between the lack of visual information, quality of output, and stress and fatigue.

There was no doubt that a lack of the supporting and guiding elements which come from presence in the meeting room does make interpreting much more difficult and requires a far greater degree of concentration resulting in either a drop in quality, or a sharp increase in stress and fatigue.

(UNESCO 1976: 31)

As for the conclusions drawn with regard to video remote interpreting, the report indicates a more mitigated picture. It was

less of a strain, but the video screen was too small and the image of only head and shoulders of the speaker was generally not adequate for interpretation purposes. It was not possible they said to follow a discussion without seeing everything—the person speaking, the persons addressed and the others who may be preparing to intervene.

(UNESCO 1976: 31)

Although the experiment was deemed “reasonably successful” (1976: 32), and dozens of small and larger tests were carried out during the decades to follow, including milestone experiments at the International Telecommunications Union (ITU) in 1999 (Moser-Mercer 2003), the European Parliament (EP) in 2004 (Roziner & Shlesinger 2010) and the International Federation of Football Associations (FIFA) in 2014 (Seeber et al. 2019), systematic implementation of audio or video remote interpreting did not materialize—at least not on a large scale—as most of the challenges identified in 1976 seemed to persist.

One noteworthy exception is the European Commission, where video remote interpreting was introduced more than 15 years ago. The first use of remote interpretation for EU Heads of State and Government (HoSG) was at the express wish of the British Prime Minister at Hampton Court, UK, during the 2005 London European Council. While it was normal practice for the plenary meetings of the European Council to be interpreted from and into all the (then) 22 official languages, working lunches were interpreted from 22 into 5 languages as dining rooms generally offered limited space in the room. However, the UK authorities’
choice of venue in Hampton Court’s mediaeval Banqueting Hall could not accommodate any interpreting booths and it was also a rigorously protected National Heritage site which allowed no modification or temporary alteration. Consequently, mobile booths had to be installed in a provisional structure erected in the courtyard adjacent to the building.

The HoSGs expressed their satisfaction with the new arrangements and asked that they be made permanent. The interpreters, while not persuaded that the benefits outweighed the difficulties, felt they could not go against the wishes of the EU’s highest instance and the interpreting arrangements continue to this day. On that basis, SCIC (the Directorate-General for Interpretation at the European Commission) considered that, in the common interest, the other EU interpreting services, staff representatives and AIIC should be included in the ensuing discussions on the use of remote interpreting. The resulting Inter-Institutional Agreement (IIA) on Remote Interpreting was reached in October 2007. Importantly, session time duration for remote was shorter than for traditional simultaneous assignments because of the added strains involved, resulting notably from the sensation of “not really being there”, i.e. feeling somewhat estranged from the reality and the resulting need for even greater mental focus than usual.

**Distance interpreting today**

As we have seen, distance interpreting is far from being a novel development. However, the implementation of (at least certain) DI modalities in the field of multilingual conference interpreting, defined by Diriker as “the rendering of speeches delivered in one language into another at formal and informal conferences and in conference-like settings” (2015: 78), has been relatively slow (Braun 2015). This might partly be explained by AIIC’s long-standing firm stance against the use of most DI modalities, which only softened with the release of the Position Paper (AIIC Executive Committee 2018) and its Guidelines for distance interpreting (AIIC Task Force on Distance Interpreting Executive Committee 2020). This is reflected in what at the time of writing can be considered the most comprehensive data set on the use of DI in conference interpreting commissioned by AIIC in 2018 (Seeber 2020), reflecting the Covid-19 pandemic is likely to impact some of the parameters observed and accelerate the pace of some of the developments identified (AIIC Executive Committee 2020).

**Penetration**

Data from 2018 suggest that many AIIC members were not complete novices to the use of different DI modalities even before the Covid-19 pandemic (80 per cent have experience with VCI, 37 per cent with ACI, 26 per cent with VRI and 17 per cent with ARI, Seeber 2020). Penetration of DI, however, was rather heterogeneous and differed both by modality and by country. The reasons for such disparity are likely a combination of several factors, including the historical development of different technologies, the legal requirements applying to particular settings as well as the technical complexity and confidentiality needs associated with certain events. Table 35.1 very clearly shows that while in some countries (e.g. Germany, France, Switzerland, and the United States) audio- and video-mediated DI modalities appeared to be approximately equally prevalent (with differences of not more than 2 percentage points between them), this was not true for other countries, like Canada, where there seemed to be a clear preference for audio-mediated DI (26 per cent) over video-mediated DI (13 per cent), or Brazil and Belgium, which only made the top 6 list for audio-mediated and video-mediated DI respectively.

A similar, cumulative analysis of the penetration of DI modalities by location of participants (see Table 35.2) reveals that while in some countries (e.g. Germany and Switzerland)
teleconferencing set-ups were much more widely used than remote interpreting set-ups, in others (e.g. France, Canada and the United States) both set-ups were used to similar extents (with differences below 4 percentage points).

Finally, at the bottom of the (top 10) list of reported DI frequency (see Table 35.3), we see countries like Italy (with around 4 per cent of penetration reported for three of the DI modalities, with ARI being even lower) and Spain (with a penetration of about 5 per cent reported for all DI modalities with the exception of ACI, which is below 3.25 per cent). Importantly, although this might be linked to the sample of participants that contributed to the 2018 survey, which in Italy was relatively low (13.8 per cent of membership), it is not a plausible explanation for the discrepancy observed in Spain, where the sample was proportionally higher (31.4 per cent of membership)—higher even than that of Switzerland (23.6 per cent of membership) and almost as high as that of Germany (35.6 per cent of membership).

<table>
<thead>
<tr>
<th>Country</th>
<th>Audio mediated</th>
<th>Video mediated</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEU</td>
<td>28.03</td>
<td>29.69</td>
</tr>
<tr>
<td>CAN</td>
<td>26.15</td>
<td>23.75</td>
</tr>
<tr>
<td>FRA</td>
<td>22.99</td>
<td>16.39</td>
</tr>
<tr>
<td>CHE</td>
<td>17.23</td>
<td>16.12</td>
</tr>
<tr>
<td>USA</td>
<td>15.03</td>
<td>13.02</td>
</tr>
<tr>
<td>BRA</td>
<td>14.62</td>
<td>12.44</td>
</tr>
</tbody>
</table>

Source: Based on AIIC’s 2018 DI survey (Seeber 2020).

<table>
<thead>
<tr>
<th>Country</th>
<th>Teleconference</th>
<th>Remote</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEU</td>
<td>34.17</td>
<td>23.55</td>
</tr>
<tr>
<td>FRA</td>
<td>23.43</td>
<td>23.30</td>
</tr>
<tr>
<td>CHE</td>
<td>21.52</td>
<td>21.61</td>
</tr>
<tr>
<td>CAN</td>
<td>17.57</td>
<td>17.12</td>
</tr>
<tr>
<td>USA</td>
<td>12.69</td>
<td>14.78</td>
</tr>
<tr>
<td>BEL</td>
<td>10.60</td>
<td>14.16</td>
</tr>
</tbody>
</table>

Source: Based on AIIC’s 2018 DI survey (Seeber 2020).

The state-of-the-art vs the state of the practice

The AIIC survey from 2018 also marked a noteworthy political change of an association which, until then, had been rather reluctant to accept the possibility of conference interpreters working in DI modalities. Since then, its dedicated Taskforce on Distance Interpreting (TFDI) has produced a series of guidelines and recommendations, informed by the data gathered in the survey and grounded in the technical specifications laid down in a series of relevant standards by the International Organization for Standardization (ISO). AIIC’s stated objective is one of providing “clearly defined working conditions adapted to each specific modality” given that, “Distance Interpreting modalities are perceived by conference interpreters as fundamentally different” and that, “different [D]istance [I]nterpreting scenarios provide interpreters with varying quantity and quality of sensory input, with differing possibilities for effective...
teamwork, and that such forms of distance interpreting may present both technological and human challenges” (AIIC Executive Committee 2018: 1–2).

**Workplace features**

Of the four DI modalities, the two types of teleconference interpreting, i.e. audioconference interpreting and videoconference interpreting, do not generally impact the interpreters’ workplace insofar as interpreters continue working in their usual environment, e.g. an interpreting booth in a conference room. The two remote interpreting modalities, audio remote interpreting and video remote interpreting, on the other hand, hold the potential of fundamentally changing interpreters’ work environment, as they include setups ranging from rather rudimentary single-device setups, e.g. working from a laptop computer, qualified as *in extremis* by AIIC, to highly sophisticated and permanently installed interpreting hubs. AIIC strongly recommends that, “all interpreters be in the same room or space (also referred to as a DI hub)” drawing particular attention to the fact that, “[t]he interpreter must be able to work with their language team and other language teams seamlessly (e.g. communication, collaboration, turn-taking)” (AIIC Executive Committee 2018: 1) and that, “[a]t least one qualified conference technician shall be present on site throughout the event to assist interpreters with the correct functioning of the equipment” (AIIC Executive Committee 2018: 2), given that managing the technical and physical aspects of a meeting “whilst interpreting with a reduced quality and quantity of relevant sensory inputs, increases the cognitive load on the interpreter, and can be a source of additional stress and fatigue” (AIIC Taskforce on Distance Interpreting, Technical Committee, 2020a: n.p.). The professional association’s insistence on the use of hubs is corroborated by the interpreters’ preference expressed in a recent study on remote conference interpreting, which concluded that even if technology were to allow working from their own homes, working from a central hub was preferable as it facilitated teamwork, promoted team spirit and ensured the support from qualified technicians (Seeber et al. 2019). Data show that in practice, VRI is still mostly performed from a conference venue (albeit a different room), in the physical presence of booth mates (86 per cent) and team mates (72 per cent), while those providing ARI principally do so from a private location, and are often not located in the same place as their booth mates (44 per cent) and even less often share the same location as their team mates (30 per cent) (Seeber 2020). Taken together, these data seem to suggest that at present, the idea of dedicated conference interpreting hubs is as yet insufficiently developed and that most video

### Table 35.3 Penetration of all four DI modalities (%)

<table>
<thead>
<tr>
<th>VCI</th>
<th>ACI</th>
<th>VRI</th>
<th>ARI</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEU</td>
<td>15.07</td>
<td>DEU</td>
<td>19.11</td>
</tr>
<tr>
<td>FRA</td>
<td>12.05</td>
<td>FRA</td>
<td>11.38</td>
</tr>
<tr>
<td>CHE</td>
<td>10.55</td>
<td>CHE</td>
<td>10.98</td>
</tr>
<tr>
<td>BEL</td>
<td>7.34</td>
<td>CAN</td>
<td>10.98</td>
</tr>
<tr>
<td>CAN</td>
<td>6.59</td>
<td>USA</td>
<td>6.10</td>
</tr>
<tr>
<td>USA</td>
<td>6.59</td>
<td>BRA</td>
<td>5.69</td>
</tr>
<tr>
<td>ESP</td>
<td>4.71</td>
<td>ITA</td>
<td>4.07</td>
</tr>
<tr>
<td>GBR</td>
<td>4.52</td>
<td>GBR</td>
<td>3.66</td>
</tr>
<tr>
<td>BRA</td>
<td>3.95</td>
<td>NLD</td>
<td>3.66</td>
</tr>
<tr>
<td>ITA</td>
<td>3.58</td>
<td>BEL</td>
<td>3.25</td>
</tr>
</tbody>
</table>

Source: Based on AIIC’s 2018 DI survey (Seeber 2020).
remote interpreting is provided using temporary installations at conference venues while most audio remote interpreting is provided from private locations.

**Workspace features**

On the face of it, the impact of DI on the interpreters’ workplace, in other words where they exercise the profession, would seem to be among the most visible changes, after all, while the introduction of simultaneous interpreting technology merely physically removed them from the centre of multilingual conferences, distance interpreting aims at physically removing them from event venues altogether—when such a venue even exists. The changes affecting conference interpreters’ workspace, however, in other words the immediate technical environment in which they work, is likely to change in equal measure if not more. In fact, the conference interpreter’s physical removal from the meeting room raises the question about the need for what has become standard (simultaneous) conference interpreting equipment, from soundproof interpreting booths to consoles and receivers ensuring adequate sound quantity and quality.

For bilingual meetings that can be serviced by a small team of two or three interpreters (see AIIC Basic Tests, 2014), or when interpreters are not located in the same physical space, it is conceivable to work without a sound-proof booth, as no acoustic insulation from the meeting room (which is physically removed) or the output of other interpreters is needed. However, even when working in such conditions, a sound-absorbing environment (e.g. a sound-proofed room) is advisable so as to minimize reverberation, echo and other unwanted acoustic artefacts. For the much more frequent scenario encountered in multilingual conference interpreting, with several, often up to a dozen or more languages provided for one meeting, working without sound-proofing equipment is impossible when interpreters share the same physical space (e.g. in a mobile or permanently installed hub). Pending significant advances in sound-cancelling or absorbing technology that would allow different interpreters working alongside without insulation without interference and noise pollution, booths (or sound-proofing materials that fulfil the same function) are therefore likely to remain a fixture of conference interpreting hubs.

Research on hub environments furthermore suggests that their physical layout including the orientation of booths need to be carefully considered and adapted to the nature of the meeting and the type of interpreting provided, including but not limited to the use of relay, for which direct or indirect visual contact with interpreters providing relay, is particularly relevant, as they allow interpreters to anticipate upcoming turns, etc. (Seeber et al. 2019).

The question of the auditory interface between interpreters and the conference room is also being raised, as the traditional physical console (often paired with infrared or radio receivers) is no longer the only way in which interpreters can interact with event participants. Of late, along with online interpreting platforms, computer-emulated soft consoles have emerged. These might well replace the tested-and-tried hardware consoles that, admittedly, have changed very little in over 50 years with regard to their principal features and functionalities. And while AIIC recommends that even in DI environments, “[i]nterpreting consoles shall be in line with ISO 20109 requirements”, referring to the international standard from 2016 that outlines the technical features of interpreting consoles, it allows for the use of “software-based interpreter interfaces” so long as they “offer the same basic functionalities” and provided that these soft consoles be made available on a separate computer (AIIC Taskforce on Distance Interpreting 2019: 2). Recent data suggest that, with the exception of ARI, which is often provided over the phone, conference interpreters still primarily use physical consoles when providing DI (Seeber 2020).
**Auditory input**

Of the four DI set-ups, the two audio-mediated modalities, i.e. ACI and ARI, reduce the conference interpreters’ input to one single channel (see Figure 35.1). Although, “specific measurement models based on key performance indicators defined specifically for the use in environments where simultaneous interpreting takes place are not yet available” (ISO 2017b: 11), so-called perceptual evaluation of audio quality, in other words, a subjective assessment of said quality on a four-point continuous scale, has been recommended to ascertain the perception of auditory impairments. The frequency response, on the other hand, is clearly defined as, “correctly reproducing audio-frequencies between at least 125 Hz and 15000 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: 2). As operators of traditional copper wire-based telephony services limit the bandwidth of calls to a range of 400 Hz to 3400 Hz, phone calls do not meet the standard set out for conference interpreting. Voice over Internet telephone protocols (VoIP), on the other hand, suffer from significant bandwidth variability and consequently, can result in the loss or delay of received data packets (i.e. “jitter”) as well as noticeable latency. Even providers offering so-called wideband audio, often referred to as HD audio, normally only cover the frequency range from 50 Hz to 7000 Hz. Interestingly, Ziegler and Gigliobianco suggest:

> whereas for consecutive interpreting the range of up to 8.000 Hz might be enough … 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.

(2018: 6)

Against this background, it is not surprising that insufficient audio quality has been a recurring issue flagged in many DI tests and experiments (Braun 2015; Roziner & Shlesinger 2010), especially those involving only auditory input, such as audioconference interpreting and audio remote interpreting (Kelly 2008; Rosenberg 2007). What might be surprising, on the other hand, is that almost 60 per cent of the interpreters surveyed report always or often providing ACI services over the phone. For ARI, this percentage is even slightly higher at 65 per cent (Seeber 2020). These numbers suggest that guidelines on audio quality for the purpose of audio-mediated conferences are often not implemented.

**Visual input**

The two video-mediated DI modalities, i.e. VCI and VRI, provide the conference interpreter input both on the auditory and the visual channel (see Figure 35.1). The quantity and quality of visual provided for the purpose of conference interpreting have been debated since the first experiments at UNESCO. This highlighted the multi-modal nature of the interpreting task and has been addressed in various subsequent tests and experiments, few of which have implemented the recommendations issued by AIIC, according to which

> the active speaker should be displayed on at least 65% of the main screen with the correct aspect ratio. Other participants, including the chair/bureau, and a general view of the event room should be displayed on the remaining part of the screen or on separate screens.

(AIIC Taskforce on Distance Interpreting 2019: 4)
The official ISO standard goes further and specifies that “the camera should zoom in so that the speaker is visible from the waist up and all lip and hand movements are also visible”, that any visual materials used by speakers “should be displayed using the direct signal of the room … on a separate screen” and gives clear specifications of camera angles, movements and latencies (ISO 2017a: 10). A recent study on video remote conference interpreting suggests that the speaker’s face, lip movements and gestures were visible more often during the DI set-up under scrutiny than during ordinary face-to-face set-ups. Conversely, however, it also highlights that the speaker’s hand movements (for the purpose of pointing at slides, for example) were visible less often than had they been physically present in the conference room (Seeber et al. 2019). We conclude that while the close-up and thus most likely static capture of one speaker with one camera might already work to conference interpreters’ satisfaction, it is the coordination among and toggling between camera views as well as the visual angles that are currently not sufficiently adapted to what interpreters want and need to see, a control issue already highlighted in earlier studies (Moser-Mercer 2003). In actual practice, conference interpreters report receiving visual input more often on giant screens when working in VCI than in VRI (Seeber 2020). This most likely means that in VCI, the same giant screen situated in the conference room allowing participants to follow the video conference, is also the interpreters’ source of visual input. Another challenge that has been closely linked to the quality and quantity of visual input is that of presence or immersion (Mouzourakis 2006), in other words the feeling of being there.

**The new reality**

The Covid-19 pandemic, with its wide-ranging restrictions on large gatherings and travel, temporarily put a halt to traditional multilingual on-site meetings. Within only a few weeks, the latter went from being the norm to becoming the exception. Conversely, the number of meetings mediated through online conferencing tools (e.g. AdobeConnect, Google Hangouts Meet, Skype, Webex, Zoom, etc.) skyrocketed, although reliable figures quantifying this increase are not available. While it can be assumed that many meetings have been transformed from multilingual to monolingual—most likely English-only—events, others have availed themselves of different solutions to provide interpretation into several languages. With large parts of the population (including conference interpreters) confined at home, the scenario clearly identified as least optimal by AIIC prior to the outbreak of the pandemic has been gaining traction as conference interpreters are working from their own homes. In their advice on performing interpreting assignments from home in extremis, the association highlights the shortcomings of makeshift or re-purposed VRI solutions and underscores the importance of using tailored SIDPs as a starting point.

Single-channel cloud platforms were not designed for simultaneous interpreting from a home studio, with some exception given to sign languages. A multichannel videoconference system or RSI platform that allows for all language channels provided to be controlled as separate audio channels is the preferred option for remote simultaneous interpreting. Multi-channel platforms should develop an interpreter interface with console functionalities, in accordance with ISO.

*(AIIC Taskforce on Distance Interpreting, Technical Committee, 2020b: 7)*

Even the use of SIDPs specifically designed for multilingual conference interpreting from a home office are currently fraught with many of the workplace and workspace problems identified above. As all data are communicated through the Internet, issues such as frequency range,
stability and latency are conditioned by the available bandwidth, which can fluctuate depending on each participant’s individual connection. The effective control of parasitic noise is difficult in a home environment and, according to AIIC, only realistically possible with a mobile podcast studio or a mobile interpreting booth (AIIC Taskforce on Distance Interpreting, Technical Committee, 2020b). Visual input, on the other hand, is also severely constrained in such a set-up as reproducing a view of “non-speaking participants, chair and boothmate … is more difficult in home settings using a web platform” (AIIC Taskforce on Distance Interpreting, Technical Committee 2020b: 8).

Because these constraints are most likely not conducive to a high-quality service and might not be inconsequential for conference interpreters’ load and thus well-being (Zekveld, Kramer & Festen 2011), DG SCIC, the largest employer of conference interpreters, developed its own solutions. The so-called ‘Enhanced Videoconference’, or ‘VC+’ scenario, foresees interpreters working in conventional booths on-site (in ISO2306-compliant booths in European Commission buildings) in meeting rooms with on-site participants. A SIDP would then be used to disseminate both the proceedings and the interpretation among remote participants (see Figure 35.3). This set-up neatly fits into VCI, with some participants and the interpreters physically present at the meeting venue and some participants attending remotely.

The so-called ‘Virtual Conference’, or ‘VC++’ scenario, would have interpreters work on-site in conventional ISO2306-compliant booths, while all participants would be located remotely and follow proceedings and the interpretation through a SIDP (see Figure 35.4). Virtual conferences, therefore, are an instance of VRI, as interpreters are not co-located with any of the other participants in the meeting, which no longer has a physical venue. Similar solutions, essentially based on the transformation of existing, ISO-compliant conference interpreting workplaces into Distance Interpreting hubs, are being explored by interpreting services of other international organizations.

**The knowns and unknowns of distance interpreting**

In almost fifty years of tests, experiments and—depending on the modality—either tentative (in the case of VRI) or relatively systematic (in the case of VCI) implementation of
Distance Interpreting, we have collected some—but far from enough—evidence allowing us to understand how these technology-mediated distance modalities affect both the conference interpreting task and conference interpreters.

**Technical specifications**

Ever since the first attempts to implement DI modalities in the 1950s, quality parameters pertaining to the transmission of sound and image were flagged as the major factors affecting the work of conference interpreters. Today, many technical specifications for DI are laid down in a series of relatively recent ISO standards and specifications and fairly comprehensively cover features and requirements relating to the quality and transmission of sound and image (ISO 20108, 2017b), conference systems equipment (ISO 22259, ISO 2019), simultaneous interpreting equipment (ISO 20109, ISO 2016), as well as simultaneous interpreting delivery platforms (ISO PAS 24019, ISO 2020). The chief signal-related specifications include, but are not limited to, the minimum frequency range (from 125 Hz to 15 KHz), the synchronization of audio with video (from -125 ms to +45 ms) and the overall latency of the transmission of the original (less than 500 ms) and provide a reliable baseline for the implementation of DI solutions.

**Acceptance**

When it comes to human factors, in other words, the interaction of humans with these systems, a fundamental question concerns professional conference interpreters’ acceptance of ICTs in general and DI modalities in particular. Technology has continuously shaped conference interpreting through what Fantinuoli (2018) calls the two technological turns: first, the introduction of simultaneous in the 1920s and second, the emergence of the Internet in the 1990s. Computer-assisted interpreting (CAI) tools have made inroads into the profession (see Fantinuoli, Chapter 36, in this volume). Technological advances with the potential of de-professionalization of the industry, however, are deemed to elicit a “general attitude of aversion and skepticism by professionals” (Fantinuoli 2018: 8).
Perhaps DI in conference settings is a simple but inevitable extension of technological progress, or the third technological turn, as Fantinuoli (2018) calls it. And yet, until recently, it has been reported that DI is much more widely accepted in public service, legal and medical settings than in conference settings and that conference interpreters specifically viewed these modalities rather sceptically (Braun 2015). This alleged scepticism might well be related to the fact that conference interpreting is generally perceived as more prestigious and more professionalized than other types of interpreting (Diriker 2015) and that conference interpreters can (and readily do) refer to a robust body of guidelines and best practices developed over the years by their own international professional associations, by key institutional employers and by the International Organization for Standardization (ISO).

**Stress and fatigue**

The largest studies specifically designed to quantify and qualify physiological and psychological stress associated with DI revealed a curious mismatch between objectively measured stress and reported fatigue (Moser 2003; Roziner & Shlesinger 2010). Since these studies were carried out, doubts have been raised about the salivary cortisol measures used to quantify stress, as “it is still elusive to interpret whether the changes in circulating levels of stress mediators such as cortisol can reflect the acute, chronic, or diurnal variations” (Lee et al. 2015: 209). It is not inconceivable, therefore, that the dissonance between objective and subjective measures was due to a construct validity issue. It is interesting to note that the Covid-19 pandemic, which has led to an explosion in the global number of videoconferences (regardless of mono- or multilingual) has once more highlighted the fatigue generated by videoconferences, this time from the vantage point of participants, not interpreters. Press articles on “Zoom fatigue” abound (e.g. Harvard Business Review 2020; National Geographic 2020; World Economic Forum 2020) describing the phenomenon and ascribing it to a range of factors, ranging from the increased demands to process non-verbal cues, body language but also tone and pitch as well as the awareness of being watched and the fact of constantly watching oneself. While targeted studies into videoconference-induced fatigue are still missing, the acute rise in the use of this technology by a non-negligible part of the population might have brought to the fore something conference interpreters have been experiencing for some time.

**Cognitive load**

A construct very closely related to the previous two, i.e. stress and fatigue, is that of cognitive load (see also Riccardi, Chapter 27, in this volume). It is widely accepted that conference interpreting, especially simultaneous conference interpreting, is one of the most cognitively complex language tasks of which the human brain is capable (Grosjean 2011). The extent to which different DI modalities generate higher cognitive load than in-situ interpreting, if at all, has received a lot of attention of late but is currently unknown. For the time being, the evidence for an increase in load is subjective and often merely anecdotal, based on conference interpreters’ accounts of their experience with different types of DI. This is reminiscent of the early days of simultaneous interpreting: the alleged additional load generated by simultaneous interpreting over consecutive interpreting, which is at the origin of the “30-minute turn” and AIIC’s guidelines governing team strength (AIIC Basic Texts 2014), was initially but an observation made during the first few attempts at implementing the back then revolutionary interpreting mode at the International Labour Organization (ILO) in the late
Kilian G. Seeber and Brian Fox

1920s (Baigorri-Jalón 2011). The experimental evidence corroborating the claim came almost 75 years later (Moser et al. 1998).

Presence

Another issue that has been discussed in relation to DI modalities is the phenomenon of alienation or lack of presence perceived by conference interpreters when they do not share the same physical space as other conference participants. Interestingly, this has been consistently reported in conference interpreters’ anecdotal accounts from the very first experiments with DI (UNESCO 1976) to the most recent studies of video remote conference interpreting (e.g. Seeber et al. 2019). It is important to underline that the reported feeling of limited control or alienation does not appear to be related to conference interpreters’ level of acceptance of new technologies, suggesting that it might indeed be related to DI’s (physical and cognitive) ergonomics rather than being a construct of a pre-existing bias.

Training for distance interpreting

Seeing the increasing use of DI modalities, calling for their integration in conference interpreter training curricula makes intuitive sense. This call was recently reiterated by Ziegler and Gigliobianco (2018: 137). In keeping with a skills-based training approach (Ericsson 2000; Moser-Mercer 2008), this raises the fundamental question of what the principal differences are between on-site and distance interpreting and how these can be trained. Here we differentiate between the operation of the technology, e.g. the use of the hardware or software interface, on one hand, and the performance of the task itself, on the other.

If the tools required to perform DI are indeed different from the customary interfaces interpreters are used to (e.g. physical consoles in the case of simultaneous interpreting), training aspiring, but also practising, interpreters is necessary in order to reduce or eliminate possible apprehension when interacting with them. Having said that, this type of technical knowledge about the interface is only relevant when the platform is paired with a soft console. In fact, many platforms can be coupled with a traditional hard console (see Figures 35.3 and 35.4) which would not necessitate any additional training. The point could be made that interpreting trainees simply require more exposure to an environment where they only follow conference proceedings by audio or video feed. Yet most interpreter training programmes have long integrated pre-recorded speeches from proprietary or public sources. Many have built up their own stock of recorded materials and the European institutions have given them access to their Speech Repository replete with crafted and real-live recordings. Whenever these materials are used in an on-site learning environment (with mobile or permanent booths equipped with one or multiple screens), this essentially re-creates a DI workspace (see also Deyssel & Lesch 2018). To those not involved in the training of conference interpreters, therefore, it might not have occurred that to a non-negligible extent, the next generation of interpreters has already been training for the DI set-ups most recommended by the professional association, i.e. specifically designed interpreting hubs. Importantly, as training installations work with closed wired circuits and a controlled environment, both the quantity and quality of the audio-visual signals made available to trainees most likely far exceed that of many DI set-ups.

Currently, the main difference between on-site and distance interpreting could therefore be found in the quantitatively and qualitatively impoverished signal (in audio-mediated interpreting, this concerns the auditory channel only while in video-mediated interpreting it might well concern both channels) interpreters work with. Rather than developing skill-based
training approaches for such adverse conditions, which might not be in the interest of trainees’ health and well-being, more efforts should be deployed to ensure that existing DI standards and recommendations are actually implemented.

A tentative outlook

Although the Covid-19-related restrictions conditioning the types of multilingual meetings conference interpreters usually service are likely to be a temporary phenomenon, they have the potential to act as a catalyst, further accelerating the development of DI.

Unsurprisingly, these developments will benefit some to the detriment of others. For example, it is not inconceivable that, owing to the complexity and (at least as compared to hard-wired solutions) still relatively low reliability of many DI solutions, more meetings might shift from a multilingual to a monolingual regime, increasing the predominance of English (see Albl-Mikasa, Chapter 39, in this volume). Although this might be welcomed by some as it cuts costs, it will deprive some of equitable access provided only by the implementation of multilingualism.

In places where conference interpretation is considered essential rather than a part of discretionary spending (be it for communicative or for political reasons), the service might be increasingly viewed as a less exclusive commodity. After all, DI allows the sourcing of providers (i.e. interpreters) around the globe without repercussions on cost, as no travel is involved. Such a development has the potential of benefiting conference interpreters living in low-wage countries, that are not generally centres of multilingual conferences, to the detriment of those living in high-wage countries, who might simply be unable to compete over price owing to their relatively higher living expenses.

Similarly, as the average duration of meetings might well decrease (owing to general audio- or videoconference fatigue), the current business model, which for (spoken language) conference interpreters adhering to AIIC’s principles is based on full non-fractionable days, is likely to change. Conversely, as the absolute cost for conference interpreting services goes down, more conference organizers might consider offering interpreting services. At the same time, the concomitant changes to the current business model might have more far-reaching repercussions in terms of the volume of work necessary to generate the same income. With a decrease in rates, the margin conference interpreters can reasonably apply to compensate for their preparation time will equally diminish. In the best-case scenario, we might see renewed growth in a segment that is generally viewed to be stagnating if not shrinking (see Pöchhacker in press). In the worst-case scenario, we might see a cheaper, but most likely also qualitatively inferior service.

Notes

1 SIDPs are cloud-based conferencing platforms allowing for (or specifically designed to support) simultaneous interpretation. They often use soft (i.e. virtual) interpreting consoles but can often be connected to traditional hard (i.e. physical) consoles.

2 Most organizations in the UN system work symmetrically (i.e. out of and into) six languages while the institutions of the European Union apply an asymmetrical (i.e. needs-based) system of up to 24 languages.

3 This applies primarily to spoken languages, as signed languages entail a natural separation of channels.

Further reading

Fox, Brian 2018. The technical feasibility of remote interpretation in the 4 main conference-serving duty stations UNHQ, UNOG, UNOV, UNON. Report for DGACM, UN. New York: UN.


References


506
Distance conference interpreting


Seeber, Kilian G. 2018. Distance interpreting survey: Answers to seven questions. Available at: https://members.aiic.net/p/8640 (accessed 17 May 2020).


