

Remote Interpreting: Issues of Multi-Sensory Integration in a Multilingual Task

Barbara Moser-Mercer

Processus et cheminements en traduction et interprétation
Processes and Pathways in Translation and Interpretation
Volume 50, numéro 2, avril 2005

URI : <https://id.erudit.org/iderudit/011014ar>
DOI : <https://doi.org/10.7202/011014ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

[Découvrir la revue](#)

Citer cet article

Moser-Mercer, B. (2005). Remote Interpreting: Issues of Multi-Sensory Integration in a Multilingual Task. *Meta*, 50 (2), 727–738.
<https://doi.org/10.7202/011014ar>

Résumé de l'article

Cet article vise à mettre en évidence le rôle central de l'intégration multi-sensorielle dans l'interprétation simultanée. L'absence de présence virtuelle apparaît être un des principaux facteurs responsables des plus mauvaises performances dans l'interprétation à distance par opposition à l'interprétation simultanée en direct. Cette détérioration de la qualité apparaît être due aux premiers signes de fatigue, lesquels semblent être la conséquence de l'allocation de ressources cognitives supplémentaires aux processus de compréhension, privant ainsi les autres processus, notamment la production, des ressources nécessaires au maintien d'un bon niveau de performances pendant la durée normale d'un tour de parole.

Remote Interpreting: Issues of Multi-Sensory Integration in a Multilingual Task

BARBARA MOSER-MERCER

University of Geneva, Geneva, Switzerland

barbara.moser@eti.unige.ch

RÉSUMÉ

Cet article vise à mettre en évidence le rôle central de l'intégration multi-sensorielle dans l'interprétation simultanée. L'absence de présence virtuelle apparaît être un des principaux facteurs responsables des plus mauvaises performances dans l'interprétation à distance par opposition à l'interprétation simultanée en direct. Cette détérioration de la qualité apparaît être due aux premiers signes de fatigue, lesquels semblent être la conséquence de l'allocation de ressources cognitives supplémentaires aux processus de compréhension, privant ainsi les autres processus, notamment la production, des ressources nécessaires au maintien d'un bon niveau de performances pendant la durée normale d'un tour de parole.

ABSTRACT

This article seeks to present evidence for the pivotal role of multi-sensory integration in simultaneous interpreting. The lack of virtual presence has emerged as one of the major factors determining poorer performance in remote as opposed to live simultaneous interpreting. This deterioration of quality appears to be based in early onset of fatigue, which in turn seems to be a consequence of allocating additional cognitive resources to comprehension processes during simultaneous interpreting and therefore depriving other parts of the process, notably production, of the resources necessary to maintain a high level of performance during normal turn time.

MOTS-CLÉS/KEYWORDS

remote interpreting, virtual presence, multi-sensory integration

Introduction

Feasibility and success of multilingual communication depend largely on the competence of the speakers and listeners on the one hand and, increasingly, the availability of human-machine interfaces that can facilitate rapid access to information from a variety of sources – auditory, visual, tactile – on the other. The burgeoning developments in technology make it feasible to present information at a distance not only by text but also by speech, computer-animated agents, gesture, and even by touch. In addition to standard desktop computers, these advanced interfaces are being implemented in telephones, mobile phones, and small handheld devices. Research has convincingly demonstrated that comprehension and communication in a monolingual setting usually succeed when listeners integrate information from several sources in an optimal way. For example, information from the face improves speech intelligibility of the message and visible body language complements auditory information perceived from the same source, provided the information is time-aligned and not contradictory.

The value of multiple sources of information in communication is even more apparent in a bilingual or multilingual setting. Listeners as comprehenders in this situation operate under a variety of constraints such as less than adequate proficiency in the language/culture of the speaker, which may increase the need to integrate information from several sources as information from any one individual source may be inadequate for successful comprehension. Even expert communicators, such as conference interpreters, are not immune to constraints. Distant communication across languages and cultures in a virtual space adds yet another layer of complexity and a new challenge, that of not being in the same place at the same time. This usually leads to a feeling of alienation, as well as to the need to communicate with the help of multiple media, in itself not always an easy task and usually one requiring considerable cognitive resources. Comprehension in a foreign language relies heavily on redundancy and active discourse construction to offset the difficulties inherent in listening in another language. The development of technological support systems to facilitate communication in a multilingual environment has certainly been beneficial in terms of making more information available to the multilingual user. Grafted onto traditional tasks or work processes, however, they have often not met with unconditional approval on the part of users. One of the reasons is that we do not yet understand how novices and experts process multiple sources of information in a media rich environment. Research in the field of expertise has highlighted the importance of creativity and innovative approaches to the development of true high-level performance, such as simultaneous interpreting. Routine expertise, impressive as it can be, is derailed the moment the task environment or the task sequence changes. Adaptive expertise, on the other hand, looks at the task from a variety of perspectives, is not wedded to a single definition of the problem, and is ready to explore entirely new approaches to getting the job done. A more thorough understanding of the underlying causes of expert breakdown in novel interpreting environments could indeed lead to improved interface designs.

Multi-sensory integration

Perceptual as well as behavioral processes are influenced by simultaneous inputs from several senses. Speech perception is a prototypical situation in which information from the face and voice is seamlessly processed to impose meaning in face-to-face communication. Audible and visible speech are complementary in that one source of information is more informative when the other source is less so. Humans possess an impressive array of specialized sensory systems that allow them to monitor simultaneously a host of environmental cues. This “parallel” processing of multiple cues not only increases the probability of detecting a given stimulus but, because the information carried along each sensory channel reflects a different feature of that stimulus, it also increases the likelihood of its accurate identification. In many circumstances, events are more readily perceived, have less ambiguity, and elicit a response far more rapidly when signaled by the coordinated action of multiple sensory modalities. Sensory systems have evolved to work in concert, and normally, different sensory cues that originate from the same event are concordant in both space and time. The products of this spatial and temporal coherence are synergistic intersensory interactions within the central nervous system, interactions that are presumed to enhance the

salience of the initiating event. Thus, for example, seeing a speaker's face makes the spoken message far easier to understand (Sumbly and Pollack: 1954), which is why conference interpreters insist on a direct view of the speaker and the meeting room. And rightly so, their insistence finds its justification in the special nature of our sensory systems.

Multisensory neurons, which receive input from more than a single sensory modality, are found in many areas of the central nervous system (Stein and Meredith: 1993). These neurons are involved in a number of circuits, and presumably in a variety of cognitive functions. Multisensory neurons in the cortex are likely participants in the perceptual, mnemonic, and associative processes that serve to bind together the modality-specific components of a multisensory experience. Other multisensory neurons, positioned at the sensorimotor interface, mediate goal-directed orientation behavior. These neurons have been studied extensively and serve as the model for deciphering how multiple sensory cues are integrated at the level of the single neuron. Visual, auditory, and somatosensory inputs converge on individual neurons in the superior colliculus (SC), where each of these modalities is represented in a common coordinate frame. Thus, the modality-specific receptive fields of an individual multisensory neuron represent similar regions of space. When presented simultaneously and paired within their receptive fields a visual and auditory stimulus result in a substantial response enhancement, well above the sum of the two individual responses. The timing of these stimuli is critical, and the magnitude of their interaction changes when the interval between the two stimuli is manipulated, with the interval or "temporal window" being on the order of several hundred milliseconds. The multisensory interactions that are observable at the level of the single neuron are reflected in behavior (Stein, Meredith, Huneycutt, and McDade: 1989). The ability to detect and orient toward a visual stimulus, for example, is markedly enhanced when it is paired with an auditory cue at the same position in space. However, if the auditory cue is spatially disparate from the visual, the response is strongly degraded. The ability of SC neurons to respond to different sensory stimuli depends on projections from a specific region of neocortex (Wallace and Stein, 1994). Without input from cortex these neurons no longer exhibit their synergistic interactions that characterize multisensory integration. We can thus conclude that higher-level cognitive functions of the neocortex play a substantial role in controlling the information-processing capability of multisensory neurons in the SC and the latter are highly sensitive to temporal alignment for successfully supporting multisensory integration.

Presence

The effectiveness of virtual environments has often been linked to the sense of presence reported by users of those environments. Presence can be defined as the subjective experience of being in one place or environment, even when one is physically situated in another (Witmer and Singer: 1998). Since face-to-face communication is a multimodal process it involves complex interactions between verbal and visual behaviors. As people speak, they gesture for emphasis and illustration, they gaze at listeners and visually monitor the environment, their facial expressions change and their body posture and orientation shift as they talk. Likewise, listeners look at speakers, as the speakers talk. Listeners monitor speakers' facial expressions and gestures, they nod

their heads to show assent, and their facial expressions and physical posture change depending on their interest in and attitude to the speaker's utterance. Furthermore, as people interact they orient to, gesture at, and manipulate physical objects in the environment they share.

Despite the multimodal nature of face-to-face communication, the most pervasive and successful technology for communicating at a distance is the telephone, which relies solely on the voice modality. Attempts at supplementing the voice modality by adding visual information have not been very successful and technologies such as videophone, web cams and videoconference occupy a relatively small share of the telephone market. This goes to show that the role of visible information in communication is both complex and subtle and that we need a more detailed theoretical understanding of the precise functions of visible information in communication. From a practical point of view we need to understand how visible information is vital for communication so that technologies can be designed that exploit visible information to provide more effective remote communication than is currently available. Much of video-mediated communication has been based on the assumption that visible information will necessarily benefit interaction, but there have not been any specific hypotheses about how these benefits will come about.

From the interpreter's perspective, one of the fundamental problems with human communication is that the literal meaning of an individual utterance underspecifies the speaker's intended meaning (Clark and Marshall: 1981, Grice: 1975). Interpreters have to infer the speaker's intended meaning by supplementing what was said with contextual information external to the utterance (see also Setton, 1999). A second problem for interpreters is to determine the effect of the speaker's utterance on the audience/delegates in the meeting room and whether he, the interpreter, drew the correct set of inferences from what was being said. Feedback mechanisms are crucial to this part of the interpreting process: speakers provide listeners, and thus interpreters, with frequent opportunities to offer feedback about what was just said (O'Conaill, Whittaker and Wilbur 1993) – delegates listening to the interpreter's nod their head and thus accept the interpreter's rendition of what the speaker said and meant, other delegates reply to the speaker's intervention and confirm thereby that they have understood. These feedback processes take place on a moment-by-moment basis (Clark and Brennan: 1991) and are crucial to the fast-paced flow of simultaneous interpreting. They are largely responsible for successful semantic anticipation in ongoing discourse without which both the consecutive and simultaneous interpreters' task becomes much more demanding.

Communication is not restricted to the exchange of propositional information, but it encompasses also the affective state or interpersonal attitude of the participants. This social information about participants' feelings, emotions, and attitudes to the other delegates and to what is being discussed is of vital importance to the interpreter as it constitutes the general framework that defines a communication event (Diriker: 2004). As with conversational intentions, participants generally do not make this information verbally explicit, so it usually has to be inferred. Access to affective information is important as it can change the outcome of verbal exchanges in situations where emotion plays a critical role, such as in negotiations (Short, Williams, and Christie: 1976).

There are several types of visible information that are used to support some of the features of face-to-face communication: gaze, or the way people extract visible information from their environment; gesture, or the set of dynamic movements and shapes formed by a person's hands and arms during communication; facial expressions conveyed by the eyes, eyebrows, nose, mouth, and forehead; and posture, or the inclination and orientation of a conversational participant's body, in particular their trunk and upper body. The latter feature is less dynamic than the former three.

Research has shown that not even high-quality audio and video replicate face-to-face processes. As an explanation I suggest retaining the hypothesis that current technology does not allow for accurate simulation of the presentational aspects of face-to-face interaction, nor does it allow for rich cross-sensory stimulation as discussed in the section on multisensory integration, and that spatial audio and video may therefore be needed to replicate communication processes. Another hypothesis to retain would be that certain types of information are substitutable across different media, whereas others are not. In face-to-face communication, cognitive and process information is partially transmitted by head nods, eye gaze, and head turning (Walker: 1993), but might also be transmitted effectively by other nonvisual cues. However, the removal of the visual channel changes the outcome of tasks that require access to affect suggesting that there is no non-visual substitute for transmitting affective information.

How can we describe the nature of the experience of presence and how can we summarize the factors that influence presence? The feeling of presence in a virtual environment (VE) depends on the extent to which attention is shifted from the physical environment to the VE, but does not require the total displacement of attention, as attention is typically divided between the physical world one works in and the VE. This division of attention, or the allocation of attentional resources to achieve the feeling of presence may vary across a range of values, but it is safe to say that interpreters have to pay an "attentional resource" price for feeling present while working remotely. Not only must the interpreter divide his attention between his physical world (the booth he is working in) and the VE (the remote meeting room), in order to experience presence in a remote setting he must be able to focus on one meaningfully coherent set of stimuli (in the VE) to the exclusion of unrelated stimuli in the physical location. The more the stimuli in the physical location fit in with those in the remote location, the less contradiction the interpreter has to resolve and the greater his chance of integrating the information across modalities to form a meaningful whole. If the interpreter is able to focus his attention on a coherent set of stimuli then he will also feel involved and immersed in the VE. Involvement and immersion produce higher levels of presence. If fully immersed one perceives to be interacting directly, not indirectly or remotely, with the VE. Thus, immersion would definitely help interpreters feel part of the VE.

Of the factors that are hypothesized to contribute to a sense of presence four are of particular importance: Control factors, sensory factors, distraction factors and realism factors. Time and again interpreters participating in remote interpreting experiments have complained about a lack of control. Indeed, the less control a person has over his task environment or in interacting with the VE, the poorer the experience of presence will be. Control relates not only to providing input on the design of the physical and the VE, it also means being able to anticipate or predict what will

happen next (Held and Durlach: 1992), a crucial factor in simultaneous interpreting and a strategy that is of paramount importance for resource allocation and savings (Moser: 1978).

Sensory factors relate directly to the different senses that need to be stimulated in order for the interpreter to develop the feeling of presence. Not surprisingly, visual information strongly influences presence, but multisensory stimulation provides an even better environment in which to develop the feeling of presence. Multimodal information needs to be consistent, i.e. it has to describe the same objective world. If the information received through one of the senses (auditory) differs from that received through another channel (visual), presence cannot develop properly – worse even, the interpreter needs to allocate additional resources to resolve the contradiction.

Distraction from the VE makes it more difficult to develop a feeling of presence in that same VE. In all of the remote interpreting tests so far interpreters were working in standard booths with the remote site being “brought to them” on either a computer screen in their booths, or a computer screen in front of their booths, or via one or several large picture panels in the (empty) conference room they were working in. Thus, none of them have really been completely isolated from the “real world” of their booths and the empty conference room these booths were located in. But it is precisely this isolation from the real world that helps those who work in VE to immerse themselves in the virtual world, as it enables them to ward off distractions from the real world and minimize the level of divided attention between the two worlds.

In the process of experimenting with different ways of presenting the virtual environment to interpreters most have concluded that large image panels that represent the real world as accurately as possible would support interpreters’ feeling of presence. Indeed, VE scene realism is critical to interpreters feeling connected to the proceedings in the conference room. This realism might, however, also have its drawbacks as some users of VE environments have experienced disorientation after leaving the virtual world (Witmer and Singer, 1998). The paradox might just be that the more presence you feel the more disoriented you will feel once you have completed your assignment.

To sum up, we can say that presence is a subjective sensation that is not easily amenable to objective physiological definition and measurement. Various characteristics of virtual environments may support or interfere with the experience of presence. Results of past remote interpreting experiments seem to make it clear, however, that presence is vital to good performance in the booth. We have seen in our discussion above that some of the factors responsible for presence support processing in simultaneous interpreting in that they reduce the amount of additional cognitive resources an interpreter needs to allocate to become fully immersed in the remote event.

Empirical evidence for presence

The first major remote interpreting experiments were carried out in the 1970s: the Paris-Nairobi (“Symphonie Satellite”) experiment by UNESCO in 1976 and the New York-Buenos Aires experiment by the United Nations in 1978. A series of experiments was conducted by the European Commission in 1995 (Studio Beaulieu), and a pilot study on ISDN video telephony for conference interpreters was carried out by

the European Telecommunications Standards Institute in 1993. The European Commission launched another test in 1997 (*Zaremba, 1997*) and yet another in 2000 (*European Commission: 2000*, <http://www.euoparl.eu.int/interp/remote_interpreting/scic_janvier2000.pdf>); the European Parliament launched two in 2001 (*European Parliament: 2001*, <http://www.euoparl.eu.int/interp/remote_interpreting/ep_report1>, <http://www.euoparl.eu.int/interp/remote_interpreting/ep_report2>), The European Council carried out a test in 2001 (<http://www.eurparl.eu.int/interp/remote/sg_conseil_avril2001.pdf>), and the United Nations explored the issue again in 1999 (*United Nations: 1999*) and in 2001. The International Telecommunication Union and the Ecole de traduction et d'interprétation launched the first controlled experiment in remote interpreting in 1999 (<http://www.aiic.net/ViewPage.cfm?page_id=1125>; Moser-Mercer, in press).

The following table provides the result of a meta-analysis of parameters that influence the feeling of presence as described in the section above. In all of the tests used for this meta-analysis, questionnaires had been handed out to participating interpreters on a daily basis. The duration of these tests ranged from three days to two weeks. The variable length of the tests should, however, not influence the first three parameters, whereas motivation is certainly in part dependent on how long the interpreter was participating in a given test. The values for the various parameters, including motivation, represent the average of all values collected for the entire test period for a given (n). Most tests used a continuous scale from - 5 to + 5, with 0 representing the neutral value, i.e. the "live" interpreting situation. In the case of the ITU/ETI study, where such a scale was not used, the data from the test were converted to facilitate comparability.

TABLE 1
A meta-analysis of parameters influencing the feeling of presence

	UN 4/1999 n=20	ITU/ETI 4/1999 n=12	SCIC 1/2000 n=43	EP 3/2001 n=34	EP 12/2001 n=33
View of the conference room	-3.61	- 2.0	-3.5	-2.48	-3.21
View of speaker	-0.28	+1.0	+0.6	+0.93	+1.30
Alienation from conference room	-3.8	- 2.5	-3.4	-3.5	-4.03
Motivation	-3.3	Not studied as separate parameter	-3.4	-2.52	-3.79

Legend:

UN 4/1999 – United Nations, A joint experiment in remote interpretation (UNHQ-UNOG-UNOV)
 ITU/ETI 4/1999 – International Telecommunication Union/École de traduction et d'interprétation, Université de Genève, Remote interpreting test
 SCIC 1/2000 – European Commission (SCIC), Tests de simulation de téléconférence
 EP 3/2001 and 12/2001 – European Parliament, Remote interpreting test

The above analysis provides clear evidence of interpreters' inability to develop a feeling of presence due to the fact that they could not obtain a realistic view of the conference room. These values correlate strongly with the feeling of alienation expressed by a significant majority of interpreters. While the positive values assigned to the view interpreters generally obtained of the speaker, values that are even above

the “norm” for live interpretation, may contribute to interpreters’ ability to obtain information from multiple modalities, audio and image, this advantage is most likely offset by the technical problems of complete synchronization of image and sound. Synchronization was not studied as a separate parameter in most of the above tests, but usually mentioned as free comment on the questionnaires returned by participating interpreters. As we know from our discussion of multisensory integration above, the temporal window for multiple streams of information to come together and thus be successfully integrated in the brain, is on the order of several hundred milliseconds. The delays in synchronization, depending on the transmission technology used (with satellite transmissions resulting in the largest lag between audio and video), will usually exceed the value of this temporal window. Thus, we can conclude that although an improved view of the speaker is noted positively, this certainly does not outweigh the disadvantages of not being able to form an impression of the remote room.

The EP 1/2001 test also provides values for a number of significant correlations, with two being of particular interest to our discussion of presence here: Ease of concentration/feeling of participation in meeting ($r = 0.55$), and motivation/feeling of participation in meeting ($r=0.68$). In the first set of correlated questions we find clear evidence of the difficulty of piecing together of a distant reality, the virtual environment, and the constant division of attention between the real world, i.e. the interpreter’s booth in an empty meeting room, and the virtual environment. This has indeed a very negative influence on focused attention (concentration) and additional cognitive resources need to be deployed to counteract the negative impact this might have on performance. If we add to that the additional resources to be deployed in order to re-align asynchronous input, i.e. when the image lags behind the sound, and the resultant loss of benefits from multi-sensory integration, we can confirm the two hypotheses indicated above: 1. *Research has shown that not even high-quality audio and video replicate face-to-face processes as current technology does not allow for accurate simulation of the presentational aspects of face-to-face interaction, nor does it allow for rich cross-sensory stimulation as discussed in the section on multisensory integration.* 2. *Certain types of information are substitutable across different media, whereas others are not. In face-to-face communication, cognitive and process information is partially transmitted by head nods, eye gaze, and head turning, but might also be transmitted effectively by other non-visual cues. However, the removal of the visual channel changes the outcome of tasks that require access to affect suggesting that there is no non-visual substitute for transmitting affective information.*

Conclusion

A number of quotes from test reports included in this meta-analysis should serve to illustrate and corroborate the theoretical assumptions that introduced this paper.

“Le fait que les interprètes ne soient pas présents dans la salle leur donne une sensation de déstabilisation (manque de points de repère) et d’aliénation, ce qui a provoqué un sentiment de démotivation très prononcé” (SCIC/European Council 6/2001)

“L’absence de l’interaction délégué-interprète ne permet pas de voir les réactions des clients et de vérifier l’efficacité de l’interprétation, comme par exemple de voir si le

résultat est clair et compréhensible, s'il n'y a pas de problèmes avec la terminologie." (SCIC/European Council 6/2001)

"Toute la gestuelle échappe. Il devient impossible de suivre réactions et interactions des délégués, d'identifier le prochain intervenant, d'anticiper la langue qu'il va parler, de se rendre compte d'éventuels problèmes techniques et de proposer des remèdes aux délégués... Le choix de la cible des caméras ne peut pas être fait par un cameraman" (ITU/ETI, 1999)

"The experiment revealed something unexpected. Although interpreters need to see what is taking place in the meeting room, they can lose their concentration when certain views of the room are selected for them by a camera at a time not of their choosing. When they are in the meeting room, interpreters select various views of the speaker and their audience only when these views do not interfere with the simultaneous processing of all the other information required for interpretation. Obtaining the visual message is one of the many tasks which interpreters handle simultaneously. The selection and timing of video-images by a camera operator cannot substitute for the interpreter's own selection and timing of the images they need, the exception being the view of the speaker" (UN, 4/1999).

"Interpreters felt alienated from the reality they were being asked to interpret. They lost their motivation and felt that their performance was not as good as it normally is. This in turn contributed to the stress, anxiety and overall loss of motivation" (UN, 4/1999).

"Much more difficult to keep up concentration compared to "normal" meetings. The slightest noise/movement in the booth becomes a distraction" (EP 1/2001)

"As I became more tired, I gradually lost the illusion of participating in the meeting, and instead, I just found myself watching the TV. An agent became a viewer. This is not good for motivation." (EP, 1/2001)

The last quote in particular is revealing in the sense that the interpreter was clearly unable to immerse himself in the virtual environment, he did not work there, he just watched from the outside, became unmotivated and could not develop the feeling of presence which in turn further increased fatigue as he had to deploy even more resources to ensure high-quality performance. Other quotes confirm the difficulty of transmitting affective information in an environment that is visually poorer compared to the real conference room. Others again underscore the difficulty of re-aligning asynchronous input when working with non-aligned audio and video signals. For multisensory integration to work it appears that none other than the interpreter himself must choose what he is looking at any given moment in time. Equipping the remote site with large panels that simulate a complete view of the meeting room might indeed contribute to successful multisensory integration. It appears to be a step in the right direction. For true presence to develop, however, the distance between the real world and the virtual environment still needs to be bridged.

Still, interpreting from a remote site will never be the "real" thing. Why, then, are we so intent on re-creating a reality that will be next to impossible to achieve? Why can't interpreters adapt? Isn't adaptive expertise what is required to survive in the 21st century? Are interpreters routine experts, unable for any extended period of time to adapt to a new work environment? Haven't millions of workers had to do so over the last 50 years? Braun (2004) is rather optimistic:

Insgesamt bestätigen die Untersuchungsergebnisse nachdrücklich, dass unsere (menschliche) Kommunikationsfähigkeit die Fähigkeit zur Bewältigung von Kommunikationsproblemen sowie zur Anpassung an neue Formen der Kommunikation einschliesst. Somit lässt sich die in bezug auf die technisch vermittelte Kommunikation prävalente “Restriktionshypothese” weitgehend entkräften. Ihr kann entgegengesetzt werden, dass Kommunikation immer unter bestimmten situativen, thematischen und eben technischen Bedingungen stattfindet – durch die sie stets auch geprägt ist, aber an die wir uns in der Regel anpassen können. (Braun, 2004: 337)

However, I cannot share her optimism. While I agree that interpreters do adapt successfully for a limited period of time, they also seem to be paying a price for it in terms of increased fatigue. The process of simultaneous interpreting is highly complex. Even an accomplished expert faces multiple challenges. Using a new machine, or a new tool, flying a new type of plane, all require retrofitting work processes. But it appears that in all these examples experts have some margin, they can re-deploy resources that are no longer required for carrying out the new task. Interpreters working remotely, however, need to continue carrying out the task of simultaneous interpreting without being able to change either the input (speakers) or the output (performance quality), yet having to face the additional challenge of “retrofitting the process” in order to overcome deficiencies created by the new environment. It is as if there were all of a sudden lots of bugs in the software that once worked perfectly well and only extremely limited cognitive resources to fix them. In this constellation nobody is willing to “yield”: the speakers won’t speak more slowly, the delegates won’t accept lesser quality. It remains to be seen if indeed a new generation of computer-savvy students arrives with acquired cognitive processes that seem more amenable to the task at hand.

REFERENCES

- AIIC (2002): *Code for the Use of New Technologies in Conference Interpretation*, Retrieved from: <http://www.aiic.net/community/print/default.cfm/page120>.
- AROCHA, J.F. and V.L. PATEL (1995): Construction-Integration Theory and Clinical Reasoning, in Ch. A. WEAVER, S. MANNES, and C.R. FLETCHER (eds.), *Discourse Comprehension. Essays in Honor of Walter Kintsch*, p. 359-381, Hillsdale, NJ, LEA.
- BÖCKER, M., and D. ANDERSON (1993): Remote Conference Interpreting Using ISDN Videotelephony: A Requirements Analysis and Feasibility Study, in *Proceedings of the Human Factors and Ergonomics society, 37th annual meeting*, p. 235-239.
- BRAUN, S. (2004): *Kommunikation unter widrigen Umständen? Fallstudien zu einsprachigen und gedolmetschten Videokonferenzen*, Tübingen, Gunter Narr Verlag.
- CASSIDY, T. (1999): *Stress, Cognition and Health*, London: Routledge.
- CLARK, H., and BRENNAN, S. (1991): Grounding in Communication, in L.B. RESNICK, J. LEVINE, and S. TEASLEY (eds.), *Perspectives on Socially Shared Cognition* p. 127-149, Washington, DC, APA.
- CLARK, H., and C. MARSHAL. (1981): Definite Reference and Mutual Knowledge, in A. JOSHI, B. WEBBER, and I. SAG (eds.), *Elements of Discourse Understanding*, p. 10-63, Cambridge, Cambridge University Press.
- CONWAY, M.A. (1997): *Cognitive Models of Memory*, Hove, Psychology Press.
- CHMIEL, N. (1998): *Jobs, Technology and People*, London, Routledge.
- DIRIKER, E. (2004): *De-/Re-contextualizing Conference Interpreting. Interpreters in the Ivory Tower?*, Amsterdam, John Benjamins.
- ERICSSON, K.A., and KINTSCH, W. (1995): Long-Term Working Memory, *Psychological Review*, p. 102, 211-245.

- ETSI. (1993): *Study of ISDN Videotelephony for Conference Interpreters. Draft Final report*, Sophia Antipolis, France: European Telecommunications Standards Institute.
- EUROPEAN COMMISSION (2000): *Rapport concernant les tests de simulation de téléconférence au SCIC en janvier 2000*, Unpublished internal report, Brussels, EC/SCIC.
- EUROPEAN PARLIAMENT (1/2001): *Report on Remote Interpretation Test, 22-25 January, 2001*, Brussels, Unpublished internal report, Brussels, European Parliament, Working Party on New Technologies of the Interpretation Directorate.
- EUROPEAN PARLIAMENT (12/2001): *Report on Remote Interpretation Test, 3-6 December, 2001*, Brussels, Unpublished internal report, Brussels, European Parliament, Working Party on New Technologies of the Interpretation Directorate.
- FINN, K.E., SELLEN, A.J., and S.B. WILBUR (1997): *Video-Mediated Communication*, Mahwah, NJ, LEA.
- GREASSER, A.C. (1981): *Prose Comprehension Beyond the Word*, New York, Springer Verlag.
- GREASSER, A.C., and L.F. CLARK (1985): *Structures and Procedures of Implicit Knowledge*, Norwood, NJ, Ablex.
- GREASSER, A.C., LANG, K.L., and R.M. ROBERTS (1991): Question Answering in the Context of Stories, *Journal of Experimental Psychology, General*, p. 120, 254-277.
- GRAESER, A.C., and R.A. ZWAAN (1995): Inference Generation and the Construction of Situation Models, in Ch.A. WEAVER, S. MANNES, and C.R. FLETCHER (eds.), *Discourse Comprehension. Essays in honor of Walter Kintsch*, p. 117-140, Hillsdale, NJ, LEA.
- GRICE, H.P. (1975): Logic and Conversation, in P. COLE and J. MORGAN (eds.), *Syntax and Semantics 3: Speech acts* (p. 225-242), New York, Academic Press.
- HELD, R., and N. DURLACH (1992): Telepresence. *Presence: Teleoperators and Virtual Environments 1(1)*: p.109-112.
- ITU/ETI/SWISSCOM. (1999): Background paper to interim report on remote interpretation, International Telecommunications Union, École de traduction et d'interprétation, Swisscom, Report published for IAMLAPD (1999), Rome, WFP, Geneva, ITU/ETI.
- KIMBALL, D.R., and K.J.HOLYOAK, Transfer and Expertise, in E. TULVING and F.I.M. CRAIK (eds.), *The Oxford Handbook of Memory* (p. 109-124), Oxford, OUP.
- KINTSCH, W. (1988): The Role of Knowledge in Discourse Comprehension, A Construction-Integration Model, *Psychological Review* 95, p. 163-182.
- KINTSCH, W. (1998): *Comprehension. A Paradigm for Cognition*, Cambridge, Cambridge University Press.
- MASSARO, D.W. (1987): *Speech Perception by Ear and Eye. A Paradigm for Psychological Inquiry*, Hillsdale, NJ, LEA.
- MOSER, B. (1978): Simultaneous Interpretation: A Hypothetical Model and Its Practical Application, in D. GERVER and H.W. SINAÏKO (eds.), *Language, Interpretation, and Communication* (p. 353-368), New York: Plenum Press.
- MOSER-MERCER, B. (1997): *Visual Contact in Simultaneous Interpretation*, Unpublished manuscript, Geneva, École de traduction et d'interprétation.
- MOSER-MERCER, B. (2000): Cognitive Issues in Interpreting Research, Paper presented at the colloquium *Linguistik und Translation* (21.-22.11.2000), Hamburg, Universität Hamburg.
- MOSER-MERCER, B. (2002): Situation Models: The Cognitive Relation Between Interpreter, Speaker and Audience, in F. ISRAËL (ed.), *Identité, altérité, équivalence? La traduction comme relation. Actes du Colloque International tenu à l'ESIT les 24, 25 et 26 mai 2000 en hommage à Marianne Lederer* (p. 163-187), Paris, Lettres Modernes Minard.
- MOSER-MERCER, B. (2003): *Remote Interpreting: Assessment of Human Factors and Performance Parameters*. Retrieved from: <<http://www.aic.net/community/print/default.cfm/page1125>>.
- MOSER-MERCER, B. (in press): Remote Interpreting: The Crucial Role of Presence, to appear in *Bulletin Suisse de linguistique appliqué* (Sondernummer: Empirische Untersuchungen zum Übersetzungs- und Dolmetschprozess 2005).
- MOSER-MERCER, B. (in preparation): *Developing Expertise in Interpreting*.

- MOSER-MERCER, B., KÜNZLI, A., and M. KORAC (1998): Prolonged Turns in Interpreting: Effects on Quality, Physiological and Psychological Stress (Pilot study), *Interpreting* 3-1, p. 47-64.
- O'CONNAILL, B., WHITTAKER, S., and S. WILBUR (1993): Conversations over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication, *Human-Computer Interaction* 8, p. 389-428.
- MOUZOURAKIS, T. (2003): *That Feeling of Being There: Vision and Presence in Remote Interpreting*, Retrieved from <<http://www.aiic.net/community/print/default.cfm/page1173>>.
- SETTON, R. (1999): *Simultaneous Interpretation. A Cognitive-Pragmatic Analysis*, Amsterdam, John Benjamins.
- SHORT, J., WILLIAMS, E., and B. CHRISTIE (1976): *The Social Psychology of Telecommunications*, London, Wiley.
- STEIN, B.E., MEREDITH, M.A., HUNEYCUTT, W.S., and L. MCDADE (1989): Behavioral Indices of Multisensory Integration, Orientation to Visual Cues Is Affected by Auditory Stimuli, *Journal of Cognitive Neuroscience* 1, p. 12-24.
- SUMBY, W.H., and I. POLLACK (1954): Visual Contribution to Speech Intelligibility in Noise, *Journal of the Acoustic Society of America*, 26, p. 212-215.
- UNITED NATIONS (1999): *A Joint Experiment in Remote Interpretation*. UNHQ-UNOG-UNOV, Geneva, United Nations, Department of General Assembly Affairs and Conference Services.
- VAN DIJK, T.A. (1994): Discourse and Cognition in Society, in D. CROWLEY and M. MITCHELL (Eds.), *Communication Theory Today* (p. 107-126), Oxford: Pergamon Press.
- VAN DIJK, T.A. (1995): On Macrostructures, Mental Models and Other Inventions, A Brief Personal History of the Kintsch-van Dijk Theory, in Ch.A. WEAVER, S. MANNES, and C.R. FLETCHER (eds.), *Discourse Comprehension. Essays in honor of Walter Kintsch*, p. 383-410, Hillsdale, NJ: LEA.
- VAN DIJK, T.A. and W. KINTSCH (1983): *Strategies of Discourse Comprehension*, New York, Academic Press.
- WALKER, M. (1993): *Information Redundancy in Dialogue*, Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.
- WALLACE, M.T. and B.E. STEIN (1994): Cross-Modal Synthesis in the Midbrain Depends on Input from Cortex, *Journal of Neurophysiology* 71, p. 429-432.
- WITMER, B.G., and M.J. SINGER (1998): Measuring Presence in Virtual Environments: A Presence Questionnaire, *Presence* 7, p. 225-240.
- ZAREMBA, H.-D. (1997): *Évaluation des questionnaires Beaulieu. Expériences en automne 1995*, Unpublished internal report, Brussels, European Commission – SCIC.