

## Putting SmartTerp to Test

### A tool for the challenges of remote interpreting

Francesca M. Frittella

Shanghai International Studies University, PhD Candidate

[frittella@shisu.edu.cn](mailto:frittella@shisu.edu.cn)

Susana Rodríguez

SmarTerp, Activity Leader

[susana@smarterterp.eu](mailto:susana@smarterterp.eu)

**ABSTRACT:** Remote simultaneous interpreting (RSI) and computer-assisted interpreting (CAI) tools powered by automatic speech recognition (ASR) and artificial intelligence (AI) are both technological developments in the interpreting profession propelled by the COVID-19 pandemic. Given the additional complexity of operating a user interface (UI) during simultaneous interpreting, we may consider UI design and overall system usability to be of crucial importance for successful RSI. However, to our knowledge, no previous article presented the evaluation of an RSI platform from the perspective of its usability and users' requirements. In this article, we present a recent evaluation study of the RSI platform SmarTerp. Reflecting a general trend in the industry, SmarTerp is one of the first RSI systems to integrate an ASR/CAI tool. This paper presents the first evaluation study of a CAI-tool integrated RSI platform. The study drew on the usability engineering evaluation methods of expert appraisal and field trial. Eight high-level conference interpreters tested SmarTerp in a simulated RSI conference based on a real debate at the European Parliament. The interpreters were divided into four booths (DEU, FRA, ITA, SPA) and ENG was the relay language. After the test, they completed three tasks gathering feedback from the perspective of (1) the individual interpreter, (2) the booth, and (3) the group of professionals. After presenting

the SmarTerp project, the paper defines the concept of ‘usability’ and details the study method. The discussion of the results sheds light on interpreters’ needs and requirements for RSI systems.

**KEYWORDS:** remote simultaneous interpreting, computer-assisted interpreting, automatic speech recognition, interpreting technology evaluation research, usability

**RESUMEN:** La interpretación simultánea remota (ISR) y las herramientas de interpretación asistida por ordenador (IAO) basadas en el reconocimiento automático del habla (ASR, por sus siglas en inglés) y la inteligencia artificial (IA) son dos desarrollos tecnológicos en la profesión de la interpretación impulsados por la pandemia del COVID-19. Dada la complejidad adicional del manejo de una interfaz de usuario durante la interpretación simultánea, podemos considerar que el diseño de la interfaz de usuario y la usabilidad general del sistema son de importancia crucial para el éxito de la ISR. Sin embargo, hasta donde sabemos, ningún artículo anterior ha presentado la evaluación de una plataforma de ISR desde la perspectiva de su usabilidad y los requisitos de los usuarios. En este artículo, presentamos un reciente estudio de evaluación de la plataforma de ISR SmarTerp. En la estela de una tendencia general en la industria, SmarTerp es uno de los primeros sistemas de ISR que integra una herramienta de ASR/IAO. Este artículo presenta el primer estudio de evaluación de una plataforma de ISR integrada con una herramienta IAO. El estudio se ha basado en los métodos de evaluación de ingeniería de la usabilidad mediante una valoración experta y pruebas de campo. Ocho intérpretes de conferencia de alto nivel utilizaron la plataforma de ISR SmarTerp para interpretar una conferencia simulada basada en un debate real en el Parlamento Europeo. Los intérpretes se dividieron en cuatro cabinas (DEU, FRA, ITA, SPA) y el ENG fue el idioma en el que se daba relay. Tras la prueba, los intérpretes completaron tres tareas en las que se recogía información desde la perspectiva (1) del intérprete individual, (2) de la cabina y (3) del grupo de profesionales. Tras presentar el proyecto SmarTerp, el artículo define el concepto de “usabilidad” y detalla el método de estudio. La reflexión sobre los resultados arroja luz sobre las necesidades y requisitos de los intérpretes para los sistemas de ISR.

**PALABRAS CLAVE:** interpretación simultánea remota, interpretación asistida por ordenador, reconocimiento automático del habla, investigación sobre la evaluación de tecnología de la interpretación, usabilidad

## 1. Introduction

The outbreak of the COVID-19 pandemic in 2020 and the subsequent restrictions on travel and in-person gathering propelled the spread of remote simultaneous interpreting (RSI). RSI is a technology-enabled interpreting modality (Braun, 2019) in which the interpreting service is provided at a distance in the simultaneous mode. As a form of “distance interpreting” based on the definition of the International Organization for Standardization, RSI consists of “interpreting of a speaker in a different location from that of the interpreter, enabled by information and communications technology (ICT)” (ISO, 2017). The first attempts to deliver RSI started in the 1970s, but the technical limitations (bandwidth, in particular) preventing the implementation of RSI could only be overcome in the second decade of the 21st century (Ziegler & Gigliobianco, 2018).

Today, RSI may be delivered both through generalized web conferencing tools and specialized software solutions known as “RSI (delivery) platforms” (ISO, 2020). The latter include functions specific to RSI and the actions that interpreters must perform in the course of their activity, such as changing turns with their boothmate, communicating with him/her, taking relay from another booth, etc. Some fundamental aspects of RSI platforms’ UI design and technical requirements are currently regulated by an ISO standard (ISO, 2020).

Putting RSI platforms within the broader perspective of the evolution of the interpreting profession, they may be regarded as one manifestation of the growing technologization of the interpreting work environment and work processes, which has been termed the “technological turn” (Fantinuoli, 2018b). RSI platforms have been categorized as “setting-oriented technologies” (Fantinuoli, 2018a), i.e., which change the environment in which interpreting is performed but do not alter its underlying cognitive processes. At the same time, from a situated cognition perspective (e.g., Strohner, 1995), which considers humans and their environments to form a cognitive ecosystem, RSI platforms may have an impact also on the interpreters’ cognition, as already argued for the use of ICTs in translation (cf. Risku, 2010; Risku, 2002). Early studies on RSI found that professional interpreters reverted to novice-like behaviours in this modality, suggesting that the modality itself (or, possibly, the

technical constraints of the systems used in the studies) prevented professionals' "cognitive adaptive behaviour" (Moser-Mercer, 2008).

The possible impact of RSI platforms on interpreters may be understood if we consider RSI as a form of human-computer interaction (HCI) (e.g., Card et al., 1986) requiring interpreters to operate a digital user interface whilst performing the challenging task of SI. The growing inclusion of computer-assisted interpreting (CAI) tools<sup>1</sup> into RSI platforms, albeit potentially beneficial for interpreters, contributes to increasing the "crowdedness" of platforms' UI. Because in HCI even a seemingly minor detail can make a major difference to users and impact their performance (cf. Nielsen, 1993), we may consider the UI design and overall usability of RSI platforms to be crucial for a successful interpretation, even more so in the case of complex systems with an integrated CAI tool. However, research on the usability and UI design of ICTs for interpreters has been scarce and is only now starting to emerge as a possible research strand within the sub-field of interpreting studies dedicated to interpreting technology research (cf. Frittella, *in press*, for a detailed review).

In this paper, we present an evaluation study of the ASR/CAI-tool integrated RSI platform SmarTerp<sup>2</sup>. The aim of the evaluation was to identify possible usage problems in interpreters' interaction with the system and deepen our understanding of their needs and requirements. The system was evaluated with eight high-level conference interpreters in a mock RSI conference simulating a real-life assignment. Data was gathered through researchers' observations during the conference and a series of feedback activities after the test using the user questionnaire and focus group methods.

After introducing SmarTerp and clarifying the central role of user research within the project, we define usability and related concepts that are central to the present study. We then detail the study methods, present

---

1 CAI tools are software solutions specifically designed to support interpreting sub-processes such as preparation and interpretation (cf. Fantinuoli, 2018b). In order to support the interpreting process, CAI tools draw on automatic speech recognition (ASR) and artificial intelligence (AI) technology to provide interpreters with real-time visual aids (cf. Fantinuoli, 2017). In other words, these CAI tools (also called ASR/CAI tools) display numbers, specialized terms, acronyms and named entities on interpreters' laptop screen in real time. The elements displayed by CAI tools are linguistic items associated with heightened cognitive load and above-average error rates during interpreting and are hence described in the interpreting studies literature as "problem triggers" (Gile, 2009).

2 [www.smarter-interpreting.eu](http://www.smarter-interpreting.eu)

the results of our data analysis, and conclude with a discussion of the study limitations, its key outcomes and their possible implications.

The paper may be of interest to readers for the innovative character of its methods used to evaluate a novel technological setup that began to emerge after the COVID-19 pandemic. Furthermore, abstracting from the case study, its findings may deepen our understanding of interpreters' needs and requirements on RSI platforms with CAI tool integration, with implications for future research and UI design.

## 2. The SmarTerp Project

SmarTerp is an Innovation Activity funded by the European Union in the framework of the EIT Digital BP2021 grant. The aim was to develop a remote simultaneous interpreting (RSI) platform with an integrated ASR- and AI-powered CAI tool. The project was initiated by the Spanish conference interpreter Susana Rodríguez, the second author of this paper. Under her coordination, an interdisciplinary team started working on SmarTerp in the autumn of 2020. As a first step, the Activity Leader Susana Rodríguez organized a series of think tanks with several stakeholders to begin preparing a document on interpreters' requirements for the development of the CAI tool and the RSI platform. The design, development and research activities began in January 2021 and continued until the end of the year. Francesca Maria Frittella, the first author of this paper, started collaborating on the project in January 2021 as a researcher. Following a user-centred design, or usability engineering, approach (cf. Nielsen, 1993), she was responsible for testing the SmarTerp prototypes of increasing sophistication to integrate interpreters' feedback into its development. The usability engineering process and the research activities conducted to ensure the interpreter-centred development of the CAI tool are described in detail in a forthcoming book with the academic publisher Language Science Press (Frittella, *in press*).

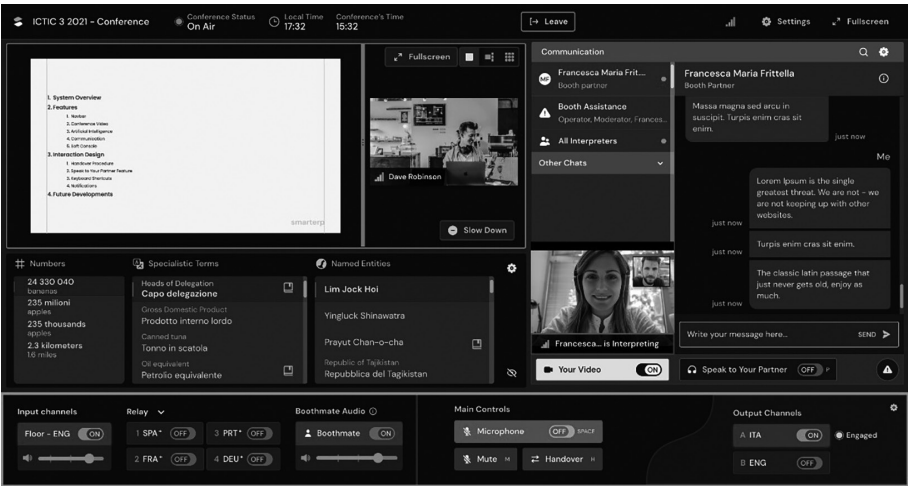
The SmarTerp interface (Fig.1) comprises two components:

An RSI system, aimed at creating the ideal conditions for a high-quality RSI service through (a) ISO-compliant audio and video quality, (b) communication options with all actors involved in the assignment (technician,

operator, conference moderator, other booths—via chat—and the boothmate—via chat and a direct audio and video channel), and (c) an RSI console allowing interpreters to perform all key actions required by this interpreting mode (change input and output channel, control their microphone, listen to their boothmate, pick up relay).

An ASR/CAI tool, supporting interpreters in the rendition of common problem triggers, i.e., named entities, acronyms and specialised terms, and numbers.

Figure 1: SmarTerp’s UI: CAI tool-integrated RSI platform at the time of the study



At the time of the test presented in this paper, the RSI platform and the CAI tool were moving from the Technology Readiness Level<sup>3</sup> (TRL) 4 (technology validated in the lab) to TRL 7 (system prototype demonstration in operational environment). At the time of writing (April 2022), the CAI tool reached TRL7 and TRL8 and is expected to reach TRL9 by the end of 2022, and both SmarTerp components are being further improved through ongoing R&D activities related to system performance and accessibility.

3 ESA - Technology Readiness Levels (TRL)

### 3. Defining Usability

In this section, we will define the concept of ‘usability’, which is central to our investigation. For a more detailed discussion, see Frittella (in press). The International Standardization Organisation defined usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO, 2018). Starting from this definition, we can isolate some key tenants of usability and its evaluation. First, the usability of a product in action or a functioning prototype must be evaluated in connection with *specified users*, i.e., the test participants must be representative of the group of people for whom the product was designed. Well before the testing phase, usability engineers employ a variety of methods (e.g., contextual inquiry, interviews, etc.) to identify users’ needs and requirements and define the basic functions and features of the product (Nielsen, 1993). But the insights gained in the subsequent testing phase can help deepen the understanding of users leading to a revision of the original concept (Nielsen, 1993). During the test, the users must try to accomplish *specified goals*, i.e., perform the tasks that the product aims to support, in a context that resembles that of the intended use. Use of the product should have both a positive impact on users’ performance and should be perceived as satisfactory. Through the analysis of users’ performance, it becomes apparent whether users succeed in performing intended goals (i.e., the product is effective) with a reasonable time expenditure (i.e., the product is efficient). Patterns of error or bottlenecks in users’ interaction with the product, called *usage problems* (Lewis, 2012), signify that something is wrong with the product: either its technical specifications (how the product works) or its user interface (UI, how it is designed) systematically lead to a suboptimal interaction. At the same time, even if a product may be used successfully, users will not choose to use it if it implies negative feelings of, for instance, effort and irritation. In other words, the usability of a product may impact both the outcome of the interaction and users’ emotional response to the product. For this reason, the usability of a product is often evaluated through the combined analysis of users’ performance and perception. The intended outcome of the evaluation is recommendations for the further improvement of the product. A rigorously conducted test

can serve as a case study contributing to forming a general understanding of the needs and requirements of a specific group of users and their systematic response to specific design features in a given context of use. In the specific case of this study, we aimed to improve the interface of SmarTerp. At the same time, we hoped to generate knowledge about interpreters' needs and platform requirements that may inform the work of others and contribute to the nascent research strand of interpreting technology usability.

## **4. Methods**

### **4.1 Aims and Research Questions**

The study presented in this paper was the first test of the working prototype of SmarTerp in action in a simulation of a real-life interpreting assignment. To our knowledge, it is also the first study to report on the evaluation of an RSI system with an integrated ASR/CAI tool. The aim of the study was to deepen our understanding of expert users' (high-level conference interpreters with extensive RSI experience) perception of the product when used in a naturalistic mock RSI assignment. The research questions were:

Do usage problems emerge when the tool is used in action? Are there UI features or technical specifications that should be adjusted?

Are test participants satisfied with SmarTerp? Why (not)?

Are the solution and all its components perceived as complete, effective and usable?

What are the key strengths and aspects requiring improvement in participants' views?

### **4.2 Research Approach**

A combination of usability engineering evaluation methods was chosen to maximize the benefits of our unique setting and reduce the risk of possible



biases.

We drew on the *field study* method (Farrell, 2016), involving the observation of users interacting with the product in a naturalistic setting to obtain ecologically valid insights. In our case, SmartTerp was not ready to be tested in an actual assignment, but we chose to simulate real-life conditions to obtain findings that could be as reflective as possible of a real interaction.

We also drew on the *expert focus group* method (Nielsen, 1993), which is used to assess user needs and feelings emerging from the dialogue amongst peers. The choice of experts to constitute the focus group allows researchers to gain insights into how the product compares to competitor products and how users map product features onto their background knowledge of the task (in our case, RSI) and its requirements.

Finally, we decided to employ *individual feedback questionnaires* as well as *pair* and *group notetaking* methods (Farrell, 2017) to probe into different perspectives, obtain a comprehensive understanding and reduce the risk of group bias, which is a possible limitation of the focus group method.

### 4.3 Timeline

The study took place between October and December 2021. The major tasks involved in the study are represented in the timeline below (Fig. 2). After the data was analyzed by the first author of this paper, she drafted a report that was discussed internally with the User Experience Design (UXD) and the development teams to develop recommendations for the further development of SmartTerp.

Figure 2: Study timeline



4.4 Participants

Study participants were high-level conference interpreters and full members of a selected professional association, the name of which may not be disclosed. Participants volunteered to participate in the study and were allowed participation if they fulfilled the following selection criteria:

- Must have one language amongst DEU, FRA, ITA and SPA in combination with ENG; two interpreters were required for each combination, one to interpret into ENG (i.e. as a ‘B’ language) and one from ENG (i.e. as a ‘C’ language).
- Must be a full member of the professional association of choice for the study.
- Had at least 10 RSI assignments in the 12 months preceding the test.

To check that all criteria were met and gather general information about their background, participants completed an enrolment questionnaire during the enrolment phase. To confirm their participation, they signed an informed consent which detailed all information about the test, tasks and deadlines, data collection and analysis procedure. Each participant received compensation of €600,00 for their participation in the study.

Table 1 provides an overview of the booth composition. A self-chosen pseudonym is used instead of the participants’ actual first name.

Table 1: Study participants – Booth arrangement

Booth	ENG B Interpreter	ENG C Interpreter
<i>French</i>	Oli	Emma
<i>German</i>	Pippi	Ar
<i>Italian</i>	Lina	Freddy
<i>Spanish</i>	Kai	Voice

The countries from which participants joined the test are Belgium (1 participant), Canada (1), France (1), Netherlands (1), Germany (2), and Italy (2). Their age ranged from below 35 (1 participant) to above 60 (4 participants). The

Figure 3: Participants’ most common RSI location

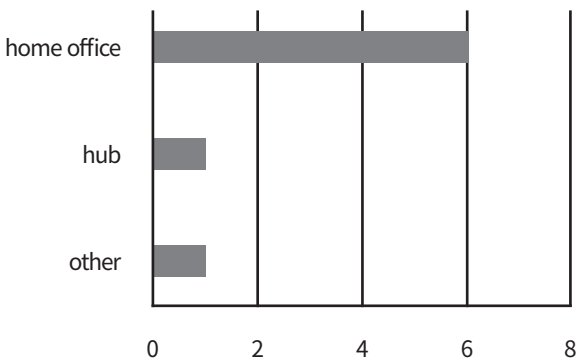
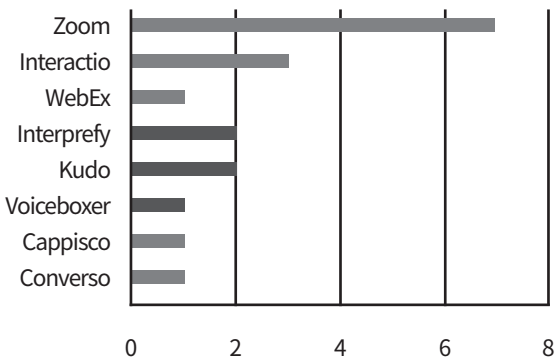


Figure 4: RSI platform most commonly used by participants



graphs below show the most common remote location they interpreted from and the most used platforms amongst participants. None of the participants had ever used an ASR- and AI-powered CAI tool in the booth before. One interpreter declared that he occasionally used ASR to generate an automatic transcript. Four participants declared that they commonly query digital glossaries during SI, one does so occasionally, and three never do so.

### 4.5 Training

Before the test, participants completed a self-paced interactive e-course (approx. 2h) developed by the first author of this paper. After the study, the course was made publicly available as an open-access educational resource

through the platform *Interpremy*<sup>4</sup>. The course comprises the following units:

Fundamental information about each component of *SmarTerp*, delivered in the form of written text and instructional videos.

Instruction on key procedural aspects of operating *SmarTerp* (e.g. step-by-step guide to testing your system before the assignment begins), provided as flowcharts, checklists and further resources.

Interactive software simulations, allowing participants to practice fundamental operations (setting input and output channels in a variety of scenarios, microphone control, handover) before actual use of *SmarTerp* during the test.

A CAI tool-assisted interpreting exercise, for participants to familiarise themselves with the tool before the test.

Table 2: Overview of study procedures and methods

Phase	Duration	Activity	Aims
Mock RSI Conference	30 min	Sound Check	Testing the communication functions and system operators' control.
	90 min	Interpretation	Testing the RSI console and the CAI tool integration, identifying usage problems.
Feedback Session	20 min	Interpreter's Questionnaire	Tapping into the perspective of the individual user on <i>SmarTerp</i> .
	30 min	Booth Feedback	Evaluating <i>SmarTerp</i> from the perspective of booth collaboration.
	30 min	Interpreter Team Feedback (2 groups)	Evaluating <i>SmarTerp</i> from the perspective of the team.
Break	10 min		
Plenary Discussion	30 min	Final Presentation and Discussion	Confirm key outcomes of the study with participants.

4 <https://interpremy.com/courses/smarterp-full-course/>

## 4.6 Test Procedures

The test was conducted on Monday 6 December 2021 from 14:00 to 18:00 CET and moderated by the authors of this paper. The first author, Francesca Maria Frittella, contributed to the study as an external researcher in the SmarTerp project and Susana Rodriguez was present as SmarTerp's activity leader. As detailed below, the study comprised a mock RSI conference followed by feedback activities from (a) the individual interpreters, (b) the booth, and (c) the whole interpreting team. Table 2 shows an overview of the test procedures and methods.

### 4.6.1 Mock RSI conference

The mock RSI conference lasted for two hours. It was conducted on the SmarTerp system operating live. Like a real assignment, the conference opened with a 30-minute technical check, which was performed by the second author of the paper in the role of system operator. After the technical check, participants interpreted on SmarTerp for 90 minutes. They occupied four booths (DEU, FRA, ITA, SPA) with English being the shared language. The test video was taken from the online repository of EU Parliament debates<sup>5</sup> and trimmed to create a series of presentations in all booths' languages. CAI tool assistance was provided, and terminology was extracted from interpreters' glossaries as well as SmarTerp's built-in glossary. A fifth Esperanto dummy booth was set up for the researchers to listen to participants' delivery during the assignment and observe interpreters' comments via the platform's chat. The aim of these observations was to identify problems encountered by participants in the use of the system. The first author of the paper recorded these observations during the test.

### 4.6.2 Feedback activities

The mock conference was followed by a series of feedback activities for a total duration of two hours. This part of the study was led by the first author

---

5 <https://www.europarl.europa.eu/plenary/en/debates-video.html>

of this paper and conducted via Zoom. The aim of the activities was to glean a subjective evaluation of SmarTerp from different perspectives: that of the individual users operating the tool, that of the booth team collaborating on SmarTerp, and that of the whole team.

The individual interpreter's perspective was explored through an *interpreter's questionnaire*. The digital questionnaire was filled out by each interpreter straight after the RSI mock conference. It comprised 14 questions related to participants' satisfaction with (a) SmarTerp overall, (b) the RSI console, and (c) the CAI tool. Closed questions aimed to gain a measure of participants' perception (on a 7-point Likert scale, which was later scaled -3 to +3) and were followed by open questions probing into the main drivers of those perceptions. The CAI tool section of the questionnaire was taken from Frittella's (in press) usability test.

The in-booth collaboration in SmarTerp was explored through a *booth feedback task*, where one break out room was created for each booth (i.e. four rooms in total with two interpreters in each). Participants were asked to discuss and then answer together in writing three questions about their experience working as boothmates in SmarTerp. The written questions were then sent to the researchers. The researchers switched between break-out rooms and noted significant quotes they overheard in the discussions.

A *group feedback task* made it possible to explore the perception of the interpreting team from a broader perspective. Participants were divided into two breakout rooms of four so that all booths would be represented in each group. Participants were asked to discuss and then answer together in writing five questions about their overall perception of SmarTerp, its key strengths and shortcomings. The written questions were then sent to the researchers. The researchers switched between break-out rooms and noted significant quotes they overheard in the discussions.

After a break, the first author of the paper summarized the key outcomes of the study in a *plenary discussion* based on the notes she made during the conference and feedback activities. She asked participants whether they felt that their point of view had been correctly understood and if they felt that their main concerns had been identified. Participants were then given time to make comments and ask questions.

## 5. Results

The results are divided as follows:

1. Mock conference, reporting detected bugs and usage problems.
2. Interpreter's perspective, emerging from feedback activity 1 (the interpreter's questionnaire) and comprising the aspects (a) overall satisfaction, (b) satisfaction with the RSI platform, (c) satisfaction with the CAI tool.
3. Booth perspective, emerging from feedback activity 2 (pair discussion and notetaking amongst booth colleagues).
4. Group perspective, emerging from feedback activity 3 (group discussion and notetaking in two groups of four).

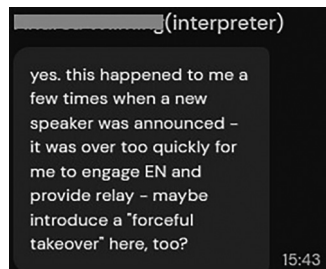
Within each category, numerical results (on a -3 to +3 scale) are followed by a qualitative analysis of positive and negative themes and representative quotes.

### 5.1 Mock Conference Observations

#### 5.1.1 Relay output channel

At the time of the test, the output channel used for relay could be selected by interpreters only when this was not already 'engaged' (i.e. used by another interpreter). This meant that, whenever an interpreter's turn to give the relay finished, s/he should immediately turn the channel off to allow his/her boothmate to take over. During the test, we observed that interpreters

Figure 5: Interpreters' chat comment concerning relay output channel



recurrently forgot to do so, causing a problem for all other booths. In the post-task questionnaire as well as the feedback discussions, this was mentioned as a major shortcoming of the system. This is exemplified in Figure 5, a screenshot of the message sent in the ‘all interpreters’ chat during the mock conference:

### 5.1.2 Communication channels

At the time of the test, interpreters could live message other parties involved in the assignment through the following communication channels:

1. *Technician*, for problems on the speaker’s side.
2. *Operator*, for problems on the interpreter’s side.
3. *Booth assistance*, to talk to technician, operator, boothmate and conference moderator simultaneously.
4. *Boothmate*, via chat, audio and video
5. *All interpreters* in the assignment
6. *Another interpreter* in another booth

Several critical incidents were observed during the mock conference that pointed to problems in the use of these communication channels. In the idle time before the sound check prior to interpreting, the researchers sent a welcome message to each interpreter individually. Only 1/8 replied to it whereas the others did not visualize it. When issues emerged during the sound check, interpreters reported them in the ‘all interpreters’ chat, although they had been requested to message the technician directly.

The post-feedback activities confirmed that users found the communication channels difficult to use:

In the interpreter’s questionnaire, one interpreter wrote: “Communication channels are confusing (too many, all the same color).” In the booth feedback activity, one booth commented: “Only one of us used the group chat because it was difficult to keep an eye on all chats.” During the group feedback activity, one group commented: “While you’re interpreting, you do not have capacities to wonder ‘who should I report this problem to?’ and select the right chat!” During the group feedback activity, one interpreter thought that there was no



option to communicate with the conference organizer. When she expressed this concern, the group was unsure whether the option was available and brought this up as a question to the moderators in the final wrap up session.

5.1.3 Console control

A further problem discussed extensively in one group during the group feedback task was that participants could not use the console proactively. The console imposed specific procedures on the participants, in particular, the output channel of the passive interpreter could not be modified before the change of turn, but only after. The comment below (from one questionnaire) explains the problem:

Also, on/off selection of mike/input channel//output channel/relay channel was counter-intuitive somehow (or just takes getting used to?) it seemed the system took some decisions for me - f. ex. I am not able to choose my output channel as a precaution (because I know which direction will be next in a few seconds) unless my mike is switched on. With a hard console in the booth, I would pre-select the channel I need next way before it is actually my turn (reducing cognitive load at handover moment).

5.2 Interpreter’s Questionnaire

5.2.1 Overall satisfaction with SmarTerp

Figures 6 and 7 show participants’ overall satisfaction with SmarTerp and their self-reported perception of individual usability criteria:

Figure 6: Column chart: Overall satisfaction with SmarTerp

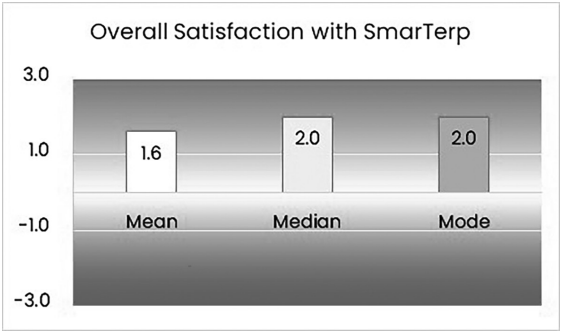
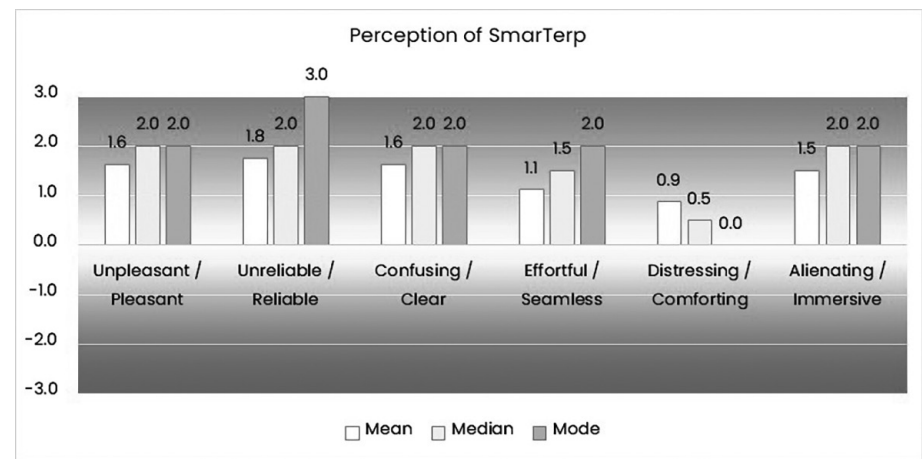


Figure 7: Perception of SmarTerp by usability criteria



The main themes that emerged from the questionnaires and that explain the results above are reported in Table 3:

Table 3: Drivers of users' positive and negative perception of SmarTerp

Drivers of Users' Perception	
Positive	Negative
Interpreters' needs are the main concern	Complex
Comprehensive	CAI tool is overwhelming
Well organised	Bugs need to be fixed

The following quote summarizes the perceived advantages and disadvantages:

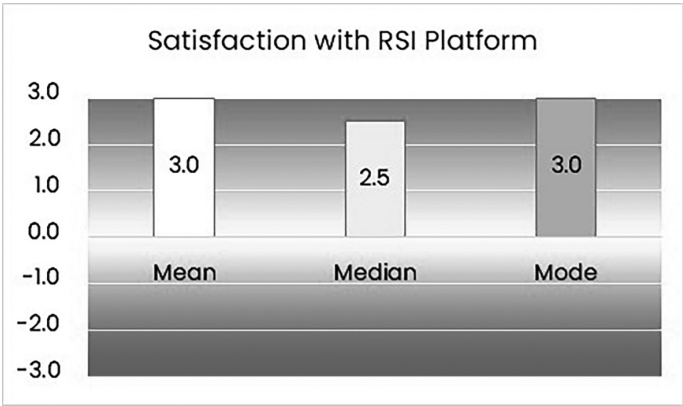
Wow! Equally distributed between wow-overwhelming and wow-this-is-next-gen-interpreting.

Wow, how much work and brainpower went into this! [...] The interface reminds me of sitting in a cockpit of a plane - so many buttons to keep an eye on...

5.2.2 Satisfaction with the RSI platform

Figure 8 shows participants’ response to the question ‘*what is your satisfaction with SmarTerp’s RSI platform?*’

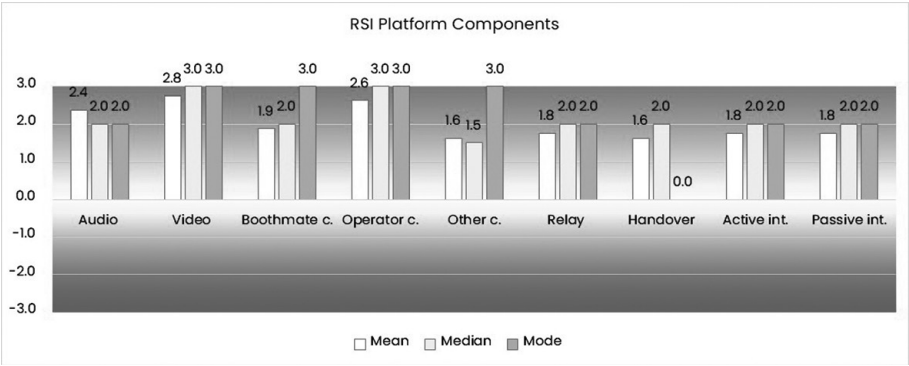
Figure 8: Column Chart: Participants’ satisfaction with SmarTerp’s RSI platform



The main driver of users’ satisfaction that emerges from the questionnaire responses are the comprehensive and integrated communication options with the boothmate, mentioned as a main strength of SmarTerp:

“It’s a step up from the other platforms because it has communication between terps (audio, video and text!), between the terp and the operator, and between the terp and the technician.”

Figure 9: Column chart: Participants’ satisfaction with SmarTerp’s RSI platform components



“The main point for me is: I can hear boothmate and speaker (or several channels, in general) via one interface and (!) I can adjust the volume for both independently! This is brilliant.”

The perception of individual platform components is shown in Figure 9 above.

5.2.3 Satisfaction with the CAI tool

Figure 10 shows participants’ satisfaction with SmarTerp’s CAI tool.

Figure 10: Column chart: Participants’ satisfaction with SmarTerp’s CAI tool

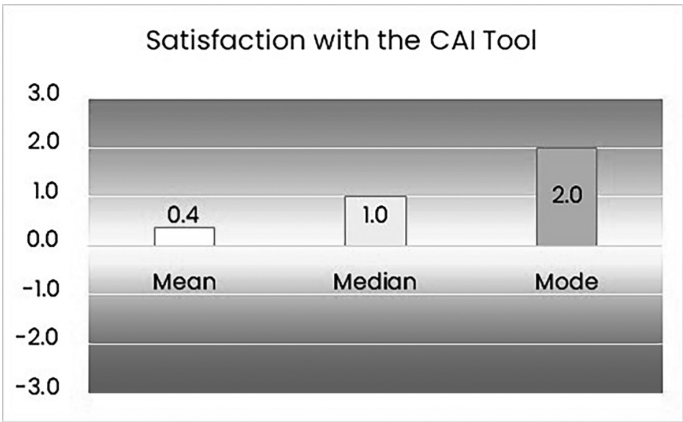
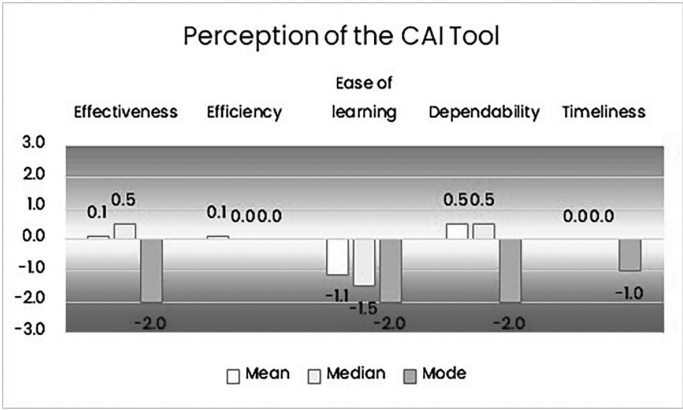


Figure 11 in the next page shows users’ evaluation of the following tool factors:

- 1. *Effectiveness*: The CAI tool helped me improve the quality of my delivery.
- 2. *Efficiency*: I could successfully use the CAI tool without unnecessary effort and confusion.
- 3. *Ease of learning*: The CAI tool can be used successfully without training.
- 4. *Dependability*: I felt that I could trust the CAI tool at all times.
- 5. *Timeliness*: The CAI tool’s support was timely.

Participants who rated these selections positively mainly did so on the

Figure 11: Column chart: Participants' satisfaction with SmarTerp's CAI tool by usability quality



ground that they could recall some instances when the tool was helpful to them, or felt “reassured” knowing that the CAI tool was provided:

It is comforting even to get confirmation and with names I LOVE IT. I find it so hard to get the names right, all you have to do here is learn to buy some time and you know that the correct name will come!

Participants who reported a negative experience said that they found the tool “overwhelming” to the point that they were unable to use it effectively:

The idea is great, but it overwhelmed and confused me, my interpretation was really bad when I tried to make use of it.

Another aspect perceived as negative was the tool’s speed:

I need help when it’s dense, fast and full of numbers. That’s when the tool was slowest.

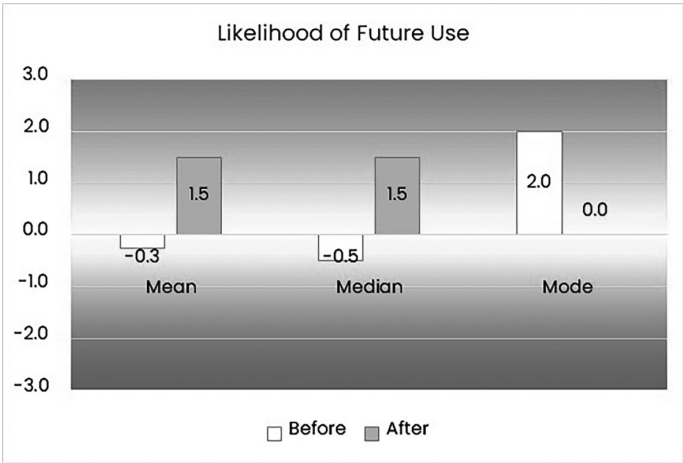
Moreover, the following design issues were mentioned:

- 1. Would prefer number + referent together
- 2. Would prefer everything in one column

3. Would prefer running transcript

Before the test (i.e. in the enrolment questionnaire) and after the test, participants were asked to respond to the question “*How likely are you to use an ASR-powered CAI tool (such as SmarTerp) during SI in the future?*”. Figure 12 reports participants’ response to this question before and after:

**Figure 12: Column chart: Participants’ self-reported likelihood of using SmarTerp’s CAI tool in the future (before and after the test)**



Despite the shortcomings highlighted by participants, the mean and median self-reported likelihood to use the tool in the future increased by 1.8 and 2 points respectively. The self-reported likelihood increased for 7 participants and decreased for only 1 of them. This change appears to be independent of the age group the participant belonged to and the self-reported attitude to technology. The participant for whom the self-reported likelihood decreased was the one who provided relay most often, as the conference moderation was in German and she was designated as an English-B interpreter. Future research may or may not find this to be a relevant factor.

### 5.2.4 Booth feedback

The booth discussion revealed the main features perceived as facilitating in-booth collaboration:

1. It is great that one is automatically muted if the other one is speaking.
2. Video is great for smoother handover.
3. Seamless, easy, very convenient to listen to my boothmate. The audio+video+text chat between boothmates is fantastic.

When asked “What were the main problems you encountered in in-booth collaboration?” participants replied:

We found that one of us only used the booth-internal chat and the other one was on the general team chat, so that resulted in us hardly ever using the chat to cooperate; booth recommended dividing boothmate chat from other chats.

When asked “How could we enhance in-booth collaboration?” participants replied:

1. Make boothmate’s video larger.
2. I want to be able to select my own output channel, regardless of boothmate’s channel and regardless of the mike status.
3. See both chats at the same time (team & booth), different background colors, maybe different location on the screen.

We also inquired into the extent to which the speak-to-your boothmate channel had been used prior to and during RSI. Participants’ responses show that:

1. All booths used it before the conference.
2. One booth used it during interpreting and found it positive because it

provides a “faster way to communicate while interpreting.” Even booths who did not use it during RSI perceived it as an asset: “it is great that this option exists!” One interpreter explained that this feature was somewhat hidden in the interface and proposed that it be made easier to access: “In Converso you have this feature ‘share a word with your boothmate’ right underneath the video. It makes it much easier to say something quickly when in need.”

### 5.2.5 Group feedback

The group discussion first revolved around the question “What are the main strengths of SmarTerp?”. The following characteristics emerged as the top three aspects:

1. In-booth communication through audio, video, and text: it is comprehensive and facilitates collaboration.
2. Operator support: you feel reassured knowing that you can count on the operator.
3. A CAI tool is provided: it would be very helpful if it was faster.

The two groups were then asked to respond to the question “How does SmarTerp help provide a professional service?”:

1. Through the operator’s support.
2. Through the CAI tool, which, one group suggested, may even be a sales argument for consultant interpreters, who could show their clients that they have a “robust infrastructure helping interpreters provide high-quality service”.

The groups were then asked “What are the main features to be improved or added?”:

1. The boothmate video should be bigger. In this respect, one participant



mentioned the RSI platform Converso as a good example.

2. The CAI tool should be faster.
3. A slow-down button for speakers should be added.
4. It should be possible to contact the conference organiser.
5. Speaker's name and key information should be displayed permanently, together with a list of participants (like in Zoom) to be hidden and unhidden.
6. There should be a dedicated field to suggest words to boothmate; many different chats are difficult to keep track of.

## 6. Discussion

Before summarizing the study findings and discussing their possible relevance, several limitations must be addressed. First, the mock conference *simulated* a real RSI assignment but was *not* a real RSI assignment. It is possible that interpreters' perception may change (for the good or the bad) and different use patterns may emerge when SmarTerp is used in a real conference. Second, the mock conference simulated an extreme case of complex interaction, as RSI assignments only rarely have four booths with relay. Third, the European Parliament debate chosen presented a very high speech pace, which may have negatively affected interpreters' perception of the CAI tool's usefulness. Finally, as it is known in usability engineering, "users are not designers" (Nielsen, 1993). Our study participants expressed views on the UI and recommended possible changes; however, their views should be considered as an expression of a need that is currently not fulfilled or only partially fulfilled by the system rather than as recommendations that should be directly incorporated into the system's design.

Coming to the key outcomes, the study tapped into participants' perception of SmarTerp's RSI platform with integrated CAI tool used in a mock RSI assignment. We can summarize the key UI features and system's technical specifications that influenced the system's usability and participants' satisfaction with the tool.

First, participants perceived SmarTerp's multiple chat channels as superfluous and even disruptive. They were found not to use chat channels

except for the ‘all interpreters’ chat. They requested for the boothmate chat to be isolated from the other channels.

Second, a major strength of the system in participants’ view was the possibility of communicating with the boothmate via audio and video.

Third, interpreters expressed a need for agency when they identified the constraints of the system in terms of selection of the output channel (which can only happen after the change of turn and not before) as a major drawback. They felt that they were limited in their choice of a strategic approach consisting in selecting the right settings before turn-taking so as to focus entirely on listening after the change of turn.

Fourth, participants found the CAI tool reassuring but overwhelming. They defined it as a major asset of the RSI system, which may contribute to a professional service. The self-reported likelihood of future use increased for all but one participant at a before-after-test comparison. However, they also expressed the need to “get used to it”—as reported also in other tests of the SmarTerp ASR/CAI tool with different participants’ cohorts (Frittella, *in press*). Their perception of the tool’s usefulness during the test varied based on the speed of the speech and the individual interpreter’s adaptivity. Interestingly, the participant who gave *retour* into English most often (i.e., the English B interpreter in the German booth) was the only one who reported a lower likelihood to use the tool after the test compared to her before-test estimate. Could this depend on the greater responsibility of her role, which decreased her tolerance for a possible disruption of the interpreting activity because of using an unfamiliar tool? Or could it depend on the fact that she had to give *retour* for brief speech passages (i.e., the introduction of a new speaker) very rapidly not to overlap with the next speech, which may have made her more sensitive to the tool’s latency? These are questions that should be addressed in future inquiries.

What do these findings tell us in general about users’ needs and requirements on RSI systems?

First, interpreters need simplicity. In order to perform their task effectively, they need the UI interface of RSI platforms to make it possible to identify all functions at a glance. To increase the usability of such systems, designers should henceforth focus on reducing the number of steps that users must make to accomplish their tasks inherent to RSI. In our example, forcing

users to open and scroll through a drop-down menu to find a needed chat channel proved to be an ineffective design choice.

Second, interpreters need an RSI system to provide the conditions for a naturalistic interaction with the boothmate. Our test participants attached great importance to functions enabling non-verbal communication with the boothmate. Platform designers should consider this need of users and identify ways to make an effective in-booth collaboration possible.

Third, RSI platforms should not obstruct interpreters' strategic behavior. Although there is no previous research to back up this hypothesis, it is most likely that, with experience, interpreters learn to operate the technological equipment needed for the job in a way that does not interfere with the core subprocesses of interpreting (listening and analyzing the source speech, producing the target speech, etc.) Notwithstanding possible technical constraints, RSI platforms should be designed to favor these strategic choices of users, hence making it possible for them to manage their cognitive resources effectively.

Fourth, interpreters may benefit from training on how to use ASR/CAI tools effectively. Our study participants had completed an unguided exercise session before testing SmarTerp but would have needed more familiarization. Furthermore, the effectiveness of the use of ASR/CAI tools seems to depend on a number of factors (including speech pace and idiosyncratic interpreter's characteristics). Finally, just as interpreters use RSI platforms strategically, it is likely that they may need to develop strategies in the use of ASR/CAI tools.

From our discussion, it becomes apparent that several aspects of interpreters' use of technological tools are still under-researched. A major research gap derives from the fact that neither the use of RSI platforms nor of ASR/CAI tools has been examined in real interpreting assignments to gain insights into how functions and features support or hinder the interpreting process. Another gap that is particularly significant for UI design work is that the work of interpreters has never been analyzed with the explicit aim to identify the strategic actions that interpreters need to be able to perform when interacting with their tools through dedicated methods such as cognitive task analysis. The first author of this paper is currently adopting CTA within her doctoral research work with the aim of informing the development of research-based training on ASR/CAI tools. Similar analysis methods should be employed to inform the design of tools.

## 7. Conclusion

The recent COVID-19 pandemic gave impetus to the rise of RSI, and it is likely that part of the interpreting market will consist of online services even after the pandemic. Given the additional complexity of operating a user interface during simultaneous interpreting, UI design and overall system usability are of crucial importance for successful RSI. However, to our knowledge, no previous paper presented the evaluation of an RSI platform from the perspective of its usability and users' requirements. This paper presented the evaluation study of the ASR/CAI tool-integrated RSI platform SmarTerp. To our knowledge, this represents the first study conducted on this complex technological setup. We simulated a realistic RSI assignment, in which eight high-level conference interpreters tested SmarTerp while interpreting a recording of a European Parliament plenary debate. Researchers' observations, as well as users' feedback after the test, led to the identification of tool features that may be obstructive to the interpreting process. In our discussion, we suggested that the usage problems we identified may be reflective of general needs to be considered in the design of RSI platforms. We concluded with some remarks on research gaps that must be addressed to inform the interpreter-centered design of technology.

## Funding

The study presented in this paper was conducted as part of the EIT Digital grant for the Innovation Activity 'SmarTerp'.

## References

- Braun, Sabine. (2019). Technology and interpreting. In Minako O'Hagan (Ed.), *The Routledge Handbook of Translation and Technology* (pp. 271-288). Routledge. <https://doi.org/10.4324/9781315311258>
- Card, Stuart K., Thomas P. Moran and Allen Newell. (1986). *The Psychology of Human-Computer Interaction*. CRC Press.

- Fantinuoli, Claudio. (2018a). Interpreting and technology: The upcoming technological turn. In Claudio Fantinuoli (Ed.), *Interpreting and Technology* (pp. 1-12). Language Science Press. <https://langsci-press.org/catalog/book/209>
- Fantinuoli, Claudio. (2018b). Computer-assisted interpreting: Challenges and future perspectives. *Trends in E-tools and Resources for Translators and Interpreters*, 153-174.
- Fantinuoli, Claudio. (2017). Speech recognition in the interpreter workstation. *Proceedings of the Translating and the Computer*, 39, 25-34.
- Farrell, Susan. (2017). *Group Notetaking for User Research*. Retrieved May 15, 2022 from <https://www.nngroup.com/articles/group-notetaking/>
- Farrell, Susan. (2016). *Field Studies*. Retrieved May 15, 2022 from <https://www.nngroup.com/articles/field-studies/>
- Frittella, Francesca M. (in press). *Designing Interpreting Technology for Usability: A Case Study of SmartTerp*. Language Science Press.
- Gile, Daniel. (2009). *Basic Concepts and Models for Interpreter and Translator Training* (Revised ed.). John Benjamins.
- ISO. (2020). *Simultaneous Interpreting Delivery Platforms—Requirements and Recommendations* (ISO/PAS Standard No. 24019:2020). <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/70/67062.html>
- ISO. (2018). *Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts* (ISO Standard No. 9241-11).
- ISO. (2017). *Simultaneous Interpreting—Quality and Transmission of Sound and Image Input—Requirements* (ISO Standard No. 20108:2017). <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/70/67062.html>
- Lewis, James R. (2012). Usability testing. In Gavriel Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (4th ed., pp. 1267-1312). John Wiley & Sons. <https://doi.org/10.1002/9781118131350.ch46>
- Moser-Mercer, Barbara. (2008). Skill acquisition in interpreting: A human performance perspective. *The Interpreter and Translator Trainer*, 2(1), 1-28.
- Nielsen, Jakob. (1993). *Usability Engineering*. Academic Press.
- Risku, Hanna. (2010). A cognitive scientific view on technical communication and translation: Do embodiment and situatedness really make a difference? *Target. International Journal of Translation Studies*, 22(1), 94-111.
- Risku, Hanna. (2002). Situatedness in translation studies. *Cognitive Systems Research*, 3(3), 523-533.
- Strohner, Hans. (1995). *Kognitive Systeme: Eine Einführung in die Kognitionswissenschaft*

[Cognitive systems: An introduction to cognitive science]. Verlag für Sozialwissenschaften.

Ziegler, Klaus and Sebastiano Gigliobianco. (2018). Present? Remote? Remotely present! New technological approaches to remote simultaneous conference interpreting. In Claudio Fantinuoli (Ed.), *Interpreting and Technology* (pp. 119-139). Language Science Press. <https://doi.org/10.5281/zenodo.1493299>

## Professional Profiles

Francesca Maria Frittella is a conference interpreter for Italian, English, German and Mandarin Chinese, an interpreter trainer, and researcher. She is a PhD candidate at Shanghai International Studies University and collaborated in the SmarTerp project as external consultant for user research and training.

Susana Rodríguez is a conference interpreter for Spanish, Italian, French, English, and Portuguese, and an independent researcher. She is the Activity Leader of the Innovation Activity SmarTerp ([www.smarter-interpreting.eu](http://www.smarter-interpreting.eu)) funded by the EIT Digital, supported by the European Institute of Innovation & Technology, a body of the European Union.