Presented at iLRN'16, University of California, Santa Barbara, USA, 27th June 2016

The 21st Century Interpreter: Exploring the use of smart-glasses for technology-augmented interpreting

Chantel Dan Chen[1], Florence Myles[1] and Victor Callaghan[1]

[1]University, of Essex
cchenn, fmyles, vic @essex.ac.uk

Abstract. In this 'work in progress' paper we set out the case for how smart-glasses can be used to augment and improve live Simultaneous Interpreting (SI). We do this through reviewing the relevant literature and identifying the current challenges faced by professional interpreters, such as cognitive load, memory constraints and session dynamics. Finally, we describe our experimental framework and the prototype smart-glasses based system we are building which will act as a testbed for research into the use of augmented-reality smart-glasses as an aid to interpreting. The main contributions of this paper are the review of the state of the art in interpreting technology plus the smart-glass experimental framework which act as an aid to Simultaneous Interpreting (SI). Later papers will report of other phases of our work.

Keywords: Simultaneous Interpreting, Translation, Languages, Augmented Reality, Smart Glasses, Meta, glossary-building, term extraction, multi-media learning, multitasking

1 Introduction

Interpreting is to orally translate the spoken words in language 'A' into language 'B'. Modern interpreting gained its professional status as early as the establishment of League of Nations, the forerunner to the United Nation [1], where interpreters were required to render oral languages between French and English, the two working languages of the organization.

Interpreters work in two different modes: consecutive and simultaneous. A consecutive interpreter listens to the source spoken language and renders it into the target language when the speaker stops for interpreters to deliver the messages to the listeners. A simultaneous interpreter renders the spoken language into the target language to the listeners in real-time while the speaker is delivering a speech. In this paper, we will only discuss simultaneous interpreting, as the smart-glasses will be applied to simultaneous interpreting only. Nowadays, simultaneous interpreters work in many different settings. International organizations, such as the United Nations and the European Commission, employ their own in-house interpreters, managed by a specific department (United Nationals DGACM n.d.), which oversees management of interpreting services for their on-going programme of international conferences and meetings.

Interpreting services are considered an ancillary service of the Meeting Incentives Conferences Exhibitions (MICE) industry [3]. Along with the development of MICE industry around the world [4], in order to engage multi-national participants in conferences and meetings, there is a growing need of professional interpreters. As such, there are already a large number of freelance interpreters, especially in the mega cities, providing interpreting services to international conferences, seminars and multi-language meetings.

The growing trend and demand are reflected by the university education system. In China alone, more than 100 universities have master level interpreters' education programmes. In the UK, the U.S and the European countries more and more universities provide master level interpreters' education. In order to provide a near-native working environment, universities invest large amount of funding in building interpreters' lab with a conference setting with a large conference table and delegate positions. The conference participants listen to the interpretation at the delegate positions through headsets.

2 Simultaneous interpreters' technical working environment

2.1 Inside the simultaneous interpreter's booth

The physical working environments of simultaneous interpreters are fixed and mobile booths. Simultaneous interpreters usually work in pairs in a booth (Fig.1). Each booth is set up with two user consoles (Fig. 2), which are each provided with a microphone and a headset. Interpreters listen to the source language through the headset and deliver the interpretation via the microphone at the same time. The interpreters take turns to interpret at every 20 – 30 minutes. The listeners outside booth listen to the interpretation from the wireless receivers or at the delegate positions. All the audio feeds are connected to a mixing console which is controlled by an audio-visual technician on site.

Fig. 1. Interpreters working in pair in a booth **Fig. 2.** Interpreter's console (the Interpreting Lab in the University of Essex)

In order to maintain the quality of an interpreters' working environment, ISO - standards [5] have been established for both mobile booths and fixed booths. The European Commission [6] has also published a technical specification for booths in conference rooms. The standards and specifications require a booth technician onsite to guarantee the two-way communication in and outside the booth. Three core metrics aim to reduce unnecessary cognitive load on the interpreters' thereby improving their performance:

- The input sound quality (to provide clearer speech)
- The quietness of the booth (so interpreters can concentrate), and
- A good view of the conference/meeting proceedings.

Interpreters also bring their own technological devices such as a laptop, tablet computer and/or smart phone to booth. Such personal devices are used to (1) display session materials (i.e. agenda, presentation files) plus a self-prepared glossary and (2) facilitate searches on the Internet.

2.2 Alternative conference interpreting equipment

In recent years, alternative equipment has been used in conference venues, mainly to reduce the cost of equipment. For example, the Tourguide system with one-way communication channel is sometimes used for small scale conferences/meetings. With this system, booths, interpreters' consoles and the mixing console are not required. Audiences listen to the interpretation through wireless receivers. To have good audio reception, interpreters need to sit near the loud-speakers or near the human speakers. Though it saves the cost of equipment hiring, such a working environment can greatly affect the interpreters' performance due to uncontrollable audio input.

A recent innovation was the introduction of a mobile phone application which, together with Bluetooth, is used to transmit interpretation services to individual listeners, replacing the wired equipment [7]. Audio input and output for both interpreters and audiences are controlled by the application. The application claims to ease the job of *conference equipment manager*, not that of the interpreters, however.

2.3 Multimedia learning context at conferences/meetings

Conferences and meetings often have a theme or correlated themes. Invited speakers talk around the theme with the aid of presentation files, often in one of the two formats PowerPoints or pdf. The introduction of the theme, the speakers and the speakers' topics are presented on the conference/meeting agenda. The purposes of conferences and meetings are to disseminate information and exchange ideas. The process of dissemination and interaction is actually a learning process for the participants. Therefore, interpreters work not just across different subject knowledge, topics and cultures but also in different learning contexts. Recent years have seen large ad-

vances in the provision of technological support for conferences and meetings. Compared with 20 years ago, conference speakers no longer use transparent plastic slides but instead use computer based presentation files, large rich multimedia displays (i.e. screen panels), fancy lighting, and more reliable and clearer sound systems help to enhance the multimedia learning experience of the conference/meeting participants.

Along with the development of software and applications, it becomes much easier and faster to design and create graphical information. Presenters add audio and video clips, complex diagrams, and figures to their presentations for better demonstration and explanation and to compress complex ideas within their presentations. The multimedia display of information and the more complex content in a presentation constitute a "*multimedia cognitive load*" for interpreters [8]. The implication is that while comprehending the presenter's messages in real-time as well as delivering it in the another language, interpreters will have to make use of much or all of the limited capacity of their working memory to comprehend, process and express the message in another language. There will be very little capacity left for interpreters to follow up the presenter-designed learning process for audiences.

To facilitate comprehension of a particular presentation, interpreters study the text and diagrams on slides to form understanding of the speaker's presentation and main ideas prior to the conference/meeting. In order to accurately render the speech and maintain a good flow of delivery, good views of the presentation file and the conference proceedings are essential for interpreters in the booth at the conference/meeting.

3 The role of the glossary for simultaneous interpreters

While preparing for an interpreting task, an interpreter usually compiles a bilingual glossary, which is formatted as two parallel columns, with one column presenting language-A and the other the equivalent word or phrase in language-B. The glossary usually contains unfamiliar words, technical terms and proper names extracted from *the speakers' presentation files, conference/meeting agenda and relevant readings* during the preparation phase. Professional interpreters, including the interpreters from the Association Internationale des Interprètes de Conférence (International Association of Conference Interpreters AIIC), consider glossaries to be of paramount importance.

AIIC is a global association of conference interpreters with over 3,000 professional members from across the world. The organization was established more than 60 years ago. Their web magazine regularly publishes articles about hot issues in the interpreting world, glossaries being one of the popular topics. The association has given guidance on glossary building in their *Practical Guide for Professional Conference Interpreters* [9]. This guide suggests the process of glossary building is a learning process which helps the interpreter to understand and remember terminologies and concepts.

A recent article in AIIC [10] presented the results of "*A survey of glossary practice of conference interpreters*". The results confirmed the importance of the learning process during glossary building, describing the process as one to "*learn about issues and concepts*". In the survey, professionals agreed that most of the glossary comes from presentations, the agenda and information linked to the agenda [10]. Moreover, the survey indicated that instantaneously retrieving the glossary from (1) the interpreter's memory or (2) a glossary list, are the only ways to use the prepared terms in the process of real-time rendition and delivery. This survey, not only emphasized the significance of the glossary list, the presentations, the agenda and interpreter's memory, but also illustrated a dynamic relationship and links between them.

3.1 Technologies for extracting terms and build up glossary

The ways to search for accurate translations of terminologies and proper names have changed from using traditional dictionaries to online dictionaries, and/or massive cloud services and databases [11, 12]. Xu and Sharoff [13] reviewed methods using comparable corpora to extract terminologies from conference documents and web content. They claim when the accuracy of the generated term lists is high, the use of automatic term lists could improve the preparation efficiency of interpreters.

More applications are also available to interpreters. Costa et-al [14] reviewed the available software for interpreter's terminology management to be used prior to an interpreting task. They also described "unit conversion" applications for mobile phones which are helpful when converting between currencies and measuring units.

3.2 Are technologies assisting interpreters in the right way?

This is a serious question raised by researchers and practicing interpreters [12, 15]. Technologies can be helpful, but with conditions and constraints. Various issues raised include how much time interpreters might spend on finding the resources and trainings required to learn and adapt to the new technologies, the familiarity required to use the new technologies, and the cognitive capacity available when working for using these technologies. For example, when an interpreter works in the booth, with a laptop to read the slides, a tablet showing terminologies, and a mobile phone at hand ready for looking up new terms, the interpreter will have to shift attention and increase processing capacity when using different media to search for information.

4 Challenges to Interpreters

4.1 Cognitive challenges

Cognitive challenges are also widely acknowledged and discussed theoretically by researchers and practicing interpreters. The last two decades has seen considerable

discussions concerning the cognitive challenges faced by interpreters, firstly from a linguistic perspective [16–19], and secondly from a psychological perspective [20–22]. This research has shown that modern presenting methods and rich-media contexts bring additional cognitive challenges, the extent of which are dependent on the content in the presentation files and on the nature of the technological environments.

Brook Macnamara [23] from Princeton University reviewed all the cognitive aptitudes required of an interpreter, and identified the cognitive functions required for interpreting. She used five complex diagrams to illustrate the required skills, abilities, intelligence, and memory from "operational, perspicacity, processing, and second language learning" perspectives (see Macnamara's paper for details), which in turn evidently reflects the cognitive challenges often experienced in interpreting.

4.2 Multitasking, attentional control and memory

Simultaneity of cognitive tasks (listening, processing and speaking) is known as multi-tasking, which is a foundational skill of Simultaneous Interpreting (SI). Attentional control allows interpreters to appropriately allocate attentional resources: (1) to attend to the useful stimuli to "logically reason, analyse and store information in memory", (2) to activate a functional working memory for processing information and form renditions in the target language [23]. With the additions of presentation files, the use of glossary list and other conference/meeting materials, the interpreters also need to allocate attentions to visual aids so as to assist comprehension and rendition. Technological advances in the personal devices are intended to support the interpreters with better management and easy alignment of additional visual information. However, the diversified applications and formats of the conference materials require the interpreter to allocate cognitive capacity and shift attentional resources for managing and processing different visual materials. For example, in a case when an interpreter needs to find a term in the glossary (prepared from the presentation materials), the interpreter's attention shifts to finding the term in the long list of glossary.

As suggested by Macnamara [23], in the process of simultaneous interpreting, attention is allocated to different tasks simultaneously. Familiarity of tasks reduce cognitive load. The extreme development of familiarity is automation (as cited in [23]). In the previous case of 'term searching' in the glossary, an automated search for terms in the glossary illustrates one form of automation. Later in this paper we will present a system (hypothesis) which explores both opportunities for reducing cognitive load through use of automation and a better designed Human Computer Interaction (HCI).

4.3 Challenges caused by the location of booth

We will illustrate the challenges facing interpreters by studying one of the settings of our training facilities in the University of Essex. LTB6 (Lecture Theatre for teach-

ing) in the University of Essex was built with fixed booths. This facility is used to host mock conferences to train interpreters. The venue comprises a large lecture room with a capacity for 300 people. The booths are fixed on one side of the upper floor (see Fig. 5).

When the interpreters go into the booths to setup the workstation, they turn on a laptop which displays a glossary list together with the speakers' presentation. In this particular context, the interpreters need to constantly check the main auditorium screen to follow the presenter's speech. As the screen concerned is about 30 meters' to one side of the booth (Fig. 3 and Fig. 5), the interpreters have difficulty reading text on the screen. To have a view of the conference proceeding, the interpreters need switch their gaze from the main auditorium to their personal laptop from time to time. Another difficulty is that the interpreter is not always able to realize immediately when the presenter changes slides, especially when the display on the projector is unclear (Fig. 4). In cases where speaker's jump slides, there is a risk of negative psychological effects on interpreters who feel they have lost track of the presentation.

Fig 3. Interpreter looks at the projecter from booth

Fig 4. Projector's view from booth

Fig 5. Booth position in LTB6

The pre-prepared glossary list can have thirty (or more) pairs of specialized terms in two languages. When the presenter mentions a term which was included in the prepared list but which the interpreter cannot remember the exact translation of, she/he needs to refer to the glossary list. Finding the term from the glossary list means re-focusing their attention away from the speaker and the list (adding to their cognitive load), until the term is located. In a case when multiple unremembered terms appear within one sentence, the interpreter needs to find all of them from the glossary list, occupying a great amount of the interpreter's cognitive capability and risking delays in interpreting.

Thus, from this setting we argue that cognitive loading (or overloading!) of an interpreter is a major factor in determining how well an interpreter performs. In particu-

lar, for any technology to be adopted by interpreters it needs to lower, rather than increase their cognitive load. The two most important aspects of cognitive loading for interpreters is 1) their working memory, and 2) their speed of reasoning. The first of these can be supported by creating computer supported glossaries of terms, with fast search methods to access them (essential extending working memory) and the second of these can be improved by good human-computer interaction design making information and control simple and intuitive (essentially simplifying any reasoning activities). By way of a theoretical basis, for the first we are building on the concept of working memory, for the second we build on the notion of elementary mental discriminations, or the Stroud number. Exploring how technology, and in particular smart-glasses, could positively augment an interpreter's capability is the aim of our research. Our approach to this is described in the following section

5 Interpreting in booth with augmented reality glasses

Fig 6. Chantel (interpreter) in a booth

Fig 7. Chantel (interpreter) wearing meta-1 glasses

As was explained in the previous section, we have set out to explore how smart glasses may be used to reduce the cognitive load on interpreters, in order to improve their performance. Thus, a project was initiated in the University of Essex to undertake research on potential solutions to the challenges described in the previous section for 21st century interpreters using augmented reality smart-glasses. At this stage we are hypothesising that smart-glasses can overcome the problems we have described, so our mission is to characterize the challenge (one of the purposes of this paper), create some theoretical models for the pedagogy and computer architecture (another aim of this paper) and then finally test the hypothesis by experimenting with a real system (an aim of a future paper). Our hypothesis is not simply a binary question (does it hold or not) but rather an exploration of the variables at work especially regarding HCI parameters such as size, position, colour and mode of control of the interpreting session data. Thus our experimental architecture seeks to accommodate as much customisation as possible, allowing the interpreters to change as much of the appearance and operation of the system as is practical. Explaining this in another way,

we are arguing that by placing a pre-prepared glossary, together with other session information in the interpreter's field of view (Fig. 8) using augmented reality glasses (with appropriately designed Human Computer Interaction), interpreters will be able to reduce their cognitive effort and concentrate more on rendering information and messages from different sources.

At this stage we are prototyping the system, starting with an electronic mock-up of the user interface which is shown in the diagram below:

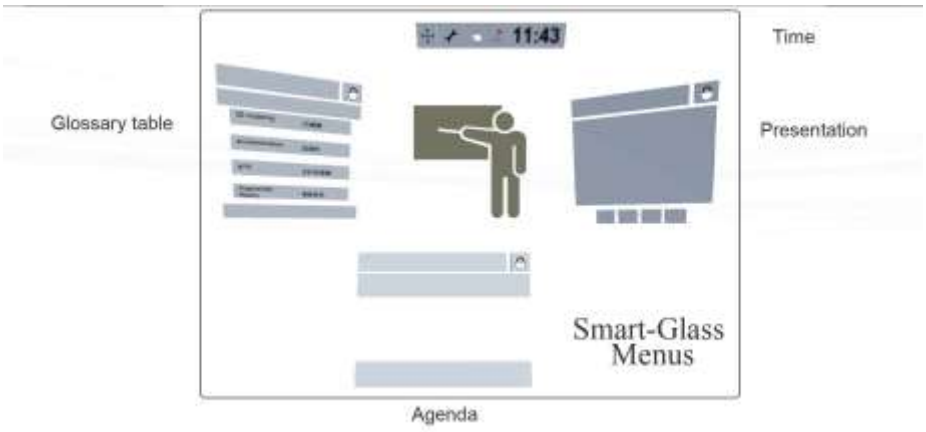

Fig 8. AR-Language Interpreting smart-glasses screen

We envisage the smart-glasses will be worn by the interpreters during live sessions allowing them to simultaneously view the real event and virtual screens containing supplementary materials positioned to one side of their field of view. The virtual screens are relatively large (a metre or so at a distance of a few meters) and contain information such as the glossary of terms, the agenda, the presenters' slides, the time and an auxiliary window that could, for example, be used by the supporting (second) interpreter who could provide additional and unplanned information. We also envisage that the second interpreter would wear a set of smart-glasses which they could use to manipulate information at key moments; to assist the main interpreter (eg undertake an online search for unknown vocabulary arising from a Q&A with the audience). This is very much an experimental system, and so one of its purposes is to allow the interpreter to customize the environment as much as possible so new research data can be gathered from how the system is personalized or used in live interpreting sessions. Thus there are many hidden functionalities concerned with personalizing the environment.

This framework forms a model for interpreting that we call SmARTI (Smart Augmented Reality Technology for Interpreters). The Meta glasses we are using were designed for individuals to wear, but have proved to be little heavy for prolonged use. Thus, one of the ideal specs for of smart-glasses for interpreters would be lightness; other features being no wires (not tethered), fashionable appearance, excellent sound, long battery life (at least a half day) etc. The current state-of-the-art in wearable AR

glasses has some way to go before they would meet an ideal specification for interpreters since they are tethered, a little on the heavy side for prolonged use, and the geeky appearance might not be appealing to all interpreters! To popularize the use of this technology, interpreters will require further hardware improvements which this work will also aim to throw light on.

6 Summary & Reflections

This paper introduced the booth environment for simultaneous interpreters. It argued that insufficient assistance is given to the interpreters in booths to reduce the cognitive load caused by the increasing use of technology and the ever-increasing complexity of contexts at conferences and meetings. In particular, we identified that extending working memory and easing reasoning tasks were key areas where technology might be used to improve an interpreter's performance. We also proposed that wearable smart-glasses might provide a useful simultaneous interpreting environment and, have described some preliminary studies we are undertaking using Meta-1 augmented-reality glasses. This is a work-in-progress project and at this stage we have framed the problem space through a literature review, identified the research issues to be explored, proposed a solution (with hypothesis), created an operational model (SmARTI – Smart Augmented Reality Technology for Interpreters) and built a simple prototype all of which we have reported on in this paper. Our longer-term aim is that we hope to be able to create what is, in effect, a virtual (and wearable) interpreting booth that is designed in such a way as to reduce the cognitive load on interpreters, thereby improving their mobility and performance. Our aim is to refine this design through ongoing work, further exploring the issues and reporting on those at later conferences.

Acknowledgements

The authors wish to thank Thomas Hayward and Ahmed Alzahrani for providing the hardware and software assistance in designing the interpreter's interface. We also acknowledge the University of Essex's UROP fund which enabled us to implement this prototype. Finally, we wish to thank the Department of Language and Linguistics University of Essex for supporting the project and interpreter lab facilities described in this paper.

References

[1] Whispering and the Origin of Simultaneous Interpreting (cont.). http://unprofessionaltranslation.blogspot.co.uk/2012/12/whispering-and-invention-of.html. Accessed 12 Apr 2016

[2] United Nations DGACM. http://www.un.org/depts/DGACM/interpretation.shtml. Accessed 12 Apr 2016

[3] (2014) CBI Product Factsheet : MICE from the EU.
[4] Vijay S (2015) What's trending in the MICE Industry? http://www.thesmartcube.com/insights/strategy/item/what-s-trending-in-the-mice-industry-2. Accessed 14 Apr 2016
[5] ISO (1998) ISO 2603:1998(en), Booths for simultaneous interpretation — General characteristics and equipment. https://www.iso.org/obp/ui/#iso:std:iso:2603:ed-3:v1:en. Accessed 18 Apr 2016
[6] European commission (2016) Main Technical Specifications for Conference Rooms.
[7] Interactio (2016) New Technologies and Conference Interpretation. http://webcache.googleusercontent.com/search?q=cache:8rdUrhH5dPwJ:www.port.ac.uk/media/contacts-and-departments/slas/events/tr08-farwick.pdf+&cd=11&hl=en&ct=clnk&gl=uk. Accessed 9 Apr 2016
[8] Sorden SD (2012) The cognitive theory of multimedia learning. In: Brown G, Lara-Alecio R (eds) Handb. Educ. Theor. Information Age, pp 1–31
[9] AIIC (2004) Practical Guide for Professional Conference Interpreters. http://aiic.net/page/628#22. Accessed 18 Apr 2016
[10] Jiang H (2015) A survey of glossary practice of conference interpreters. http://aiic.net/page/7151/a-survey-of-glossary-practice-of-conference-interpreters/lang/1. Accessed 18 Apr 2016
[11] Kelly N (2009) Moving toward machine interpretation. Transl. localization
[12] Farwick JS (2010) New Technologies and Conference Interpretation Judith S . Farwick. In: Lee-Jahnke, Hannelore Prunc E (ed) Am Schnittpkt. Von Philol. Und Transl. Festschrift Zu Ehren Von Martin Forstner. Peter Lang, pp 70–79
[13] Xu R, Sharoff S (2014) Evaluating Term Extraction Methods for Interpreters. In: 4th Int. Work. Comput. Terminol. pp 86–93
[14] Costa, Hernani; Pastor Gloria Corpas; Munes ID (2014) Technology-Assisted Interpreting. Multilingual 25:27–32.
[15] Pym A (2011) What technology does to translating. Int J Transl Interpret 3:1–9.
[16] Roy CB (1993) A sociolinguistic analysis of the interpreter's role in simultaneous talk in interpreted interaction. Multilingua 12:341–364. doi: 10.1515/mult.1993.12.4.341
[17] Darò V, Fabbro F (1994) Verbal memory during simultaneous interpretation: Effects of phonological interference. Appl Linguist 15:365–381. doi: 10.1093/applin/15.4.365
[18] Gile D (2008) 2008 Local Analysis Forum. In: Forum 6. pp 2.59–77.
[19] Hanaoka O (2002) JAIS Proper Names in the Media : Problems for Translators / Interpreters and L2 Learners. Interpret Stud 2:28–42.
[20] Gerver D (1975) A Psychological Approach to Simultaneous Interpretation. Meta J des traducteurs 20:119–128. doi: 10.7202/002885ar
[21] Moser-Mercer B, Künzli A, Korac M (1998) Prolonged turns in interpreting: Effects on quality, physiological and psychological stress (Pilot study). Interpreting 3:47–64. doi: 10.1075/intp.3.1.03mos
[22] Kurz I (2003) Physiological Stress During Simultaneous Interpreting: a comparison of experts and novices. Interpret Newsl 12:51–67
[23] Macnamara BN (2012) Interpreter Cognitive Aptitudes. J Interpret 19:9–31.