Cognitive load in simultaneous interpreting: Measures and methods

Article *in* Target International Journal of Translation Studies · March 2013

DOI: 10.1075/target.25.1.03see

1 author:

Kilian G. Seeber
University of Geneva

John Benjamins Publishing Company

Cognitive load in simultaneous interpreting

Measures and methods

Kilian G. Seeber
University of Geneva, Switzerland

The mental effort required to perform a simultaneous interpreting task or the cognitive load generated by it has attracted the interest of many a researcher in the field. To date, however, there is little agreement on the most suitable method to measure this phenomenon. In this contribution, I set out to discuss four of the most common methods of measuring cognitive load and the way in which they have been applied in interpreting research, providing examples for each and highlighting their respective advantages and disadvantages. The main focus of the contribution will be on pupillometry, a psycho-physiological method I deem to be among the most promising approaches to objectively measure cognitive load during simultaneous interpreting in real time.

Keywords: simultaneous interpreting, cognitive load, mental effort, analytical methods, subjective methods, performance methods, psycho-physiological methods, methodology, pupillometry

1. Introduction

Among interpreting scholars, the list of those who consider simultaneous interpreting a cognitively tasking activity (e.g., Gile 1995; Hyönä et al. 1995; Massaro and Shlesinger 1997; Moser-Mercer 1997; De Groot 2000) seems to extend beyond that of those who consider such statements as nothing but "primitives or clichés" (Setton 2003, 37). Indeed, Setton argues that concurrent sub-tasks during simultaneous interpreting can be performed "comfortably if they are all sharing the same representation" (2001, 5). Over the years, the notion of cognitive load generated by the interpreting task, or the amount of cognitive effort necessary to perform it, has generated a substantial amount of interest and has been addressed by scholars from within and outside the paradigm who believe that such investigation might be very fruitful (de Groot 1997). The amount of empirical evidence gathered to

Target 25:1 (2013), **18–32**. DOI 10.1075/target.25.1.03see

corroborate theories and claims about the amount of cognitive load generated by the task, however, would appear to be inversely proportional to the strength of the assertions put forward. As the following discussion will show, this imbalance may partially be explained by the difficulty of finding an appropriate paradigm within which to test hypotheses, coupled with a methodology capable of identifying, isolating and measuring the phenomenon as directly as possible. The purpose of this article is to provide an analysis of the potential and limitations of some of the most widely used methods for investigating cognitive load in simultaneous interpreting and the metrics (or measures) they employ, with a special focus on pupillometry.

2. Measuring cognitive load

The multidimensional nature of cognitive load makes it difficult to define.[1] On the one hand, this construct represents the load imposed on the performer by a particular task (Paas and Merrienboer 1993), on the other hand, it represents the perceived effort invested by a performer during the execution of that task (Yin et al. 2008). For the purpose of the present analysis, cognitive load will be defined as the amount of capacity the performance of a cognitive task occupies in an inherently capacity-limited system (Miller 1956). Paas et al. (2003) and Schultheis and Jameson (2004) describe four discrete categories of methods for the assessment of cognitive load, all with their respective advantages and disadvantages: analytical methods, subjective methods, performance methods and psycho-physiological methods.

Analytical methods attempt to estimate cognitive load relying on subjective data, often elicited through expert opinion, and analytical data, often generated with mathematical models or task analysis (Paas et al. 2003). The advantage of this approach is that the cross-tabulation of subjective and analytical data can take place at a purely theoretical level thus avoiding sometimes cumbersome empirical testing; its major shortcoming is that it relies exclusively on prior knowledge both about the task and about the subjects and is therefore unable to take into account individual performance differences.

Subjective methods use self-reported data as a means to quantify phenomena that are perceived as difficult or impossible to assess objectively, such as cognitive load. Data are generated using introspection, as well as retrospective and concurrent verbalization, and are reflected in metrics such as rating scales. The advantage of these methods over analytical methods is that they involve task performers, who have been shown to be able to provide an appreciation of their perceived mental load (Paas et al. 2003). Their drawback is a possible contamination of data by memory and consciousness effects, seeing that the response is usually time-delayed. With the exception of concurrent verbalization (i.e., think-aloud), self-rating and assessment takes place after the performance of the task.

Performance methods usually involve the simultaneous performance of a primary task and a secondary task with the goal of identifying the extent to which the latter affects the former (Haapalainen et al. 2010). The advantage of these methods is that they allow tasks to be studied without the need to de-compose them. Their major drawback is that uncontrolled processes might confound the causal relationship between the two tasks at hand.

Psycho-physiological methods, finally, assess the physiological processes known to co-vary with changes in cognitive load. In doing so, they provide a more direct and, given that these physiological responses are not subject to voluntary control, more objective measure of cognitive load. The drawback of these methods is their complexity; they can be invasive (to varying degrees) and therefore might interfere with the task itself.

What follows is an attempt to illustrate, by means of a few examples, how these methods have been used to qualify and quantify cognitive load phenomena in simultaneous interpreting.

2.1 Cognitive load in interpreting research: Analytical methods

Among the best-known analytical approaches to conceiving of cognitive load in interpreting are Gile's effort models (1995). The underlying framework of this "conceptual framework" (Gile, 2008: 62) is similar to Kahneman's (1973) single resource theory that postulates the existence of a single pool of finite processing capacity to fuel all cognitive tasks. Summarizing more complex operations under the heading of four efforts, i.e., listening and analysis, production, memory, and coordination, Gile proposes a simple architecture to describe the amount of effort invested in the simultaneous performance of these operations to get from the input "I" to the total amount of invested effort "T" (see Figure 1). The appeal of this conceptual framework is its striking simplicity. Reducing the complex simultaneous interpreting task to an uncomplicated mathematical formula (which, as the author points out, should not be understood as a simple arithmetic sum) is a noteworthy feat. One of the examples Gile provides illustrates the processing capacity demands involved in the simultaneous interpretation of the sentence given at the bottom of Figure 1. The schematic convincingly illustrates how a local increase in information density (i.e., the section in italics) causes a knock-on effect eventually leading to load being exported to a subsequent processing stage. However, this model's main strength, i.e., its ability to capture the complex task of simultaneous interpreting and provide a simple and concise account of its intrinsic cognitive demands, constitutes a potential weakness for its application beyond the realm of teaching. It is possible that some of the parallel cognitive processes recruit the same resources. It is also possible, however, that some tasks involved in an online bilingual language processing task such as simultaneous interpreting are constrained by limitations going beyond those inherent to the three principal component tasks. Such limitations might only come to bear when the component tasks are combined into a more complex one. In other words, in terms of cognitive load, simultaneous interpreting might be more or less than the sum of its parts. It is conceivable that this incongruity cannot merely be accounted for by the coordination effort, which is assumed to be ever-present and thus not specifically represented in Gile's model (see Figure 1).

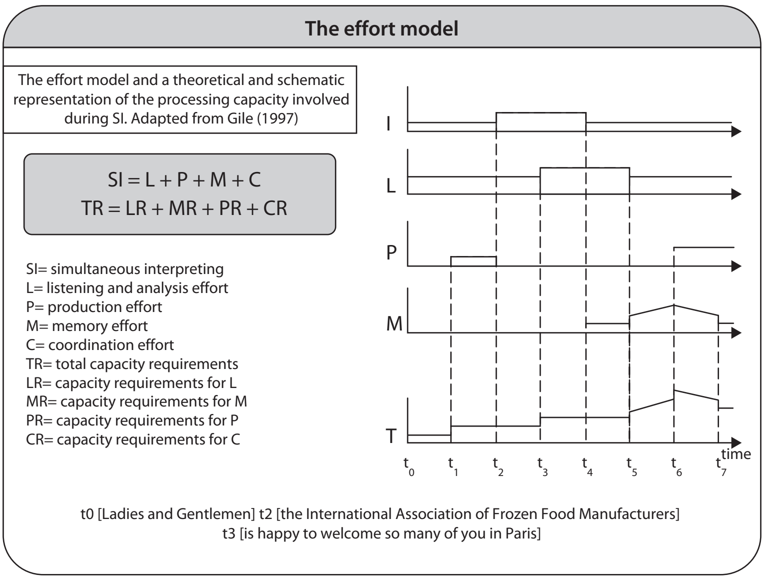

Figure 1. Adaptation of Gile's (1997) effort model

What is more, empirical findings do not lend unequivocal support to Kahneman's (1973) single resource theory. On the contrary, they suggest that structurally similar tasks interfere more with each other than structurally dissimilar ones do (Wickens 2002). I attempted to address some of these limitations by introducing the Cognitive Load Model as a competing account to the Effort Model (Seeber 2011). The former is based on Wickens' (1984, 2002) Multiple Resource Theory and assumes a finite amount of task-specific processing capacity as well as a certain amount of task-interference depending on the structural proximity of the tasks involved[2] (see Figure 2).

The strength of the Cognitive Load Model is its ability to illustrate cognitive load taking into account both the input and output — something the Effort Model

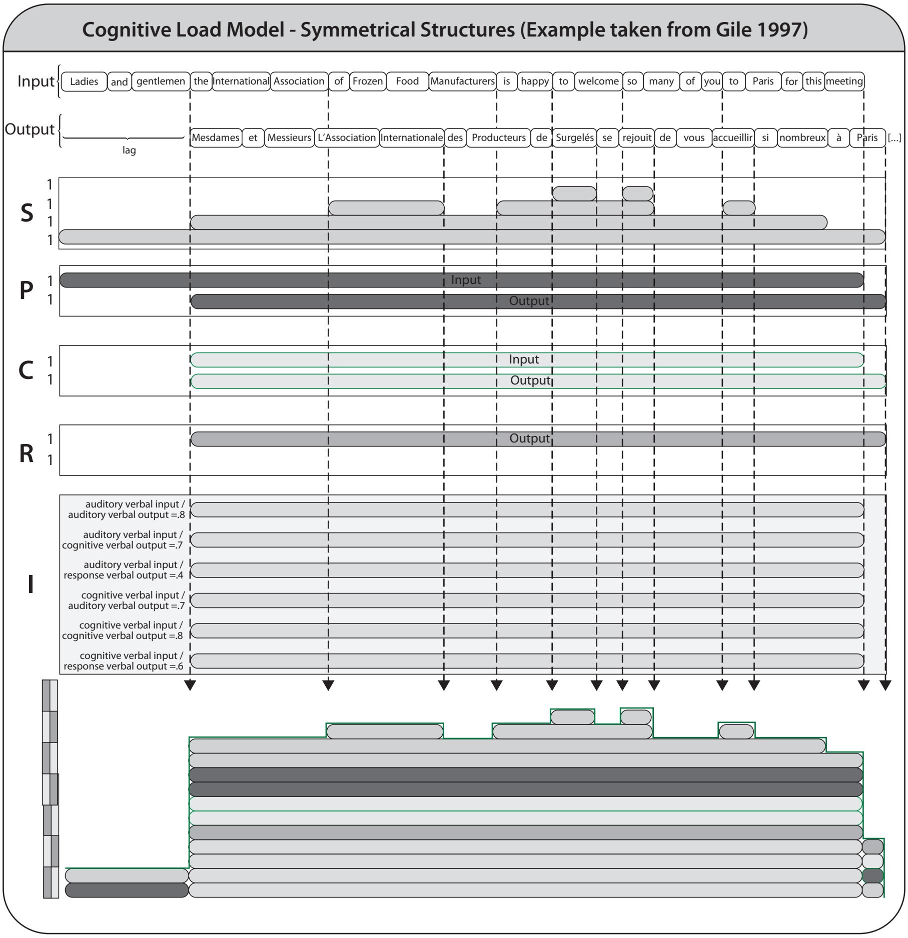

Figure 2. Seeber's (2011) Cognitive Load Model for symmetrical structures (S = storage, P = perceptual auditory verbal processing, C = cognitive verbal processing, R = verbal response processing, I = interference)

seems to fall short of. As the maximum local load[3] is intrinsically linked to the amount of parallel processing (and thus interference generated by different tasks), as well as the amount of time for which elements need to be stored, the predictions of Gile's Effort Model appear somewhat arbitrary. The Cognitive Load Model for the same sentence, for instance, illustrates one (of several) ways of interpreting the sentence that keeps cognitive load fairly constant, without presupposing any particular background knowledge by the interpreter[4] (the latter limitation is shared with Gile's Effort Model and is due to the nature of the measure used). These two examples of analytical measures of cognitive load illustrate the constraints inherent to an approach unable to take into account individual differences.

2.2 Cognitive load in interpreting research: Subjective methods

A review of the recent literature suggests that unlike analytical methods, subjective methods do not seem to be widely used to assess cognitive load in interpreting research. This might be attributable to two reasons. On the one hand, there appears to be little agreement on the reliability of self-reports of cognitive load. While some (e.g., Gopher and Braune 1984; Paas et al. 2003) found them to be relatively reliable, others (e.g., Mital and Govindaraju 1999) found evidence to the contrary. On the other hand, it is conceded that only concurrent verbalization can accurately reflect the mental status of participants (Bernardini 2001), and that post-hoc verbalization after long tasks may lead to incomplete descriptions (Ericsson and Simon 1984) and distortions. These limitations prompted Shlesinger to conclude that "for all intents and purposes, TAPs, in the ordinary sense, are not a viable tool for us" (2000: 3). In a more comprehensive methodological discussion of verbal methods in interpreting research, Ivanova (2000) suggests using transcripts of interpreters' output as retrieval cues during post-hoc verbalizations. This approach encourages participants to recall rather than to reconstruct cognitive processes, but does not solve the problem of a close interaction between participant and experimenter, which is argued to contaminate data (Bernardini 2001). Lamberger-Felber, who compares 12 interpreters' subjective assessment of source-text difficulty using Pöchhacker's (1994) discourse parameters, reports "high scoring variability in almost all individual parameters investigated" (Lamberger-Felber 2001: 47). While these findings most likely indicate the need for further validation of Pöchhacker's taxonomy for the purpose of assessing the difficulty of source materials in simultaneous interpreting, they might also simply serve as a reminder to use caution when subjectively estimating the amount of load likely to be generated by a particular task.

Finally, subjectivity also makes its way into scientific inquiry through the backdoor of material selection for experimental purposes. Scholars select and match source materials for experiments on simultaneous interpreting and related tasks (i.e., shadowing, sight translation, simultaneous interpreting with text, etc.) based on a subjective assessment of difficulty. Readers are told, e.g., that the selected materials are "of roughly equal difficulty" (Moser-Mercer et al. 2000, 115) and thus left to trust the (interpreter-)researchers in their subjective assessment.

It would appear, then, that subjective methods might not provide the necessary objectivity to reliably assess cognitive load in simultaneous interpreting.

2.3 Cognitive load in interpreting research: Performance methods

Performance measures have a long tradition in psychology and made their way into interpreting research when the first psychologists started showing interest in this novel object of study. In line with the tradition at the time, the experimental analyses of the task focused on performance speed and performance accuracy. Oléron and Nanpon (1965), for example, quantified the time lag (i.e., EVS or Ear to Voice Span) between simultaneous interpreters and speakers, whereas Barik (1969) quantified errors and omissions in simultaneous interpreting.

It would appear that performance measures have lost little of their appeal as they are still frequently used measures in the experimental study of simultaneous interpreting (e.g., Mazza 2000; Gile 2008; 2011; Tauchmanová 2011). Although contributing to the generation of a non-negligible amount of data, it is interesting to see that one of the principal challenges inherent to performance methods (identified by Woodworth as early as 1899) has not often been addressed, let alone solved, in almost 50 years of experimental research on simultaneous interpreting: the trade-off between speed and accuracy. It is generally believed that, when carried out quickly, tasks will suffer in terms of accuracy, and when carried out accurately, they will suffer in terms of speed. The crux with simultaneous interpreting is that unlike in psychological experiments entailing arguably simpler tasks (like giving a true-or-false response), where participants can be instructed and given a regulatory focus, researchers do not usually attempt to control interpreters' regulatory focus. This means that the regulatory focus within one and the same interpreter might change from one performance to the next or even within one and the same performance. Furthermore, as simultaneous interpreting already consists of simultaneously executed tasks (i.e., language comprehension and language production), the traditional approach of having a main task carried out simultaneously with a secondary task (Paas et al. 2003; Haapalainen et al. 2010) is not practicable. In the absence of a traditional secondary task paradigm, establishing a causal relationship between performance speed or performance accuracy and cognitive load at any given point during the process becomes problematic. This problem is best exemplified by Pym's (2008) re-interpretation of Gile's (1999) data where, using a purely theoretical approach, Pym convincingly illustrates how varying priorities among participants can be invoked as an equally plausible explanation of the observed phenomena.[5]

Seeing that the aforementioned performance methods to measure cognitive load in simultaneous interpreting allow competing (and potentially mutually exclusive) interpretations of the results, it would seem prudent to combine them with other, more objective measures.

2.4 Cognitive load in interpreting research: Psycho-psychological methods

The fourth approach used to measure cognitive load relates to psycho-physiological techniques, i.e., methods allowing the measurement of cardiac, hematic, electro-dermal, ocular, muscular and cerebral responses. Many of these techniques are continuous and allow for a moment-to-moment analysis of events, which is of crucial importance for language processing, not only because cognitive load is assumed to fluctuate locally, but also because the individual operations inherent in language processing are performed very rapidly and can be measured in fractions of seconds (Mitchell 2004). Another advantage is that most physiological responses are controlled by the sympathetic nervous system, and as such, cannot be consciously influenced, but rather constitute an objective measure. The main drawback of psycho-physiological measures is the difficulty of identifying and determining what is actually measured. Furthermore, several techniques are invasive, and many of them highly complex and expensive. In order to be considered suitable for the measurement of cognitive load in simultaneous interpreting, any method must be evaluated against the following four requirements: noise-resistance[6], non-invasiveness[7], temporal resolution and affordability. Given these constraints, it is not surprising that to date only very few psycho-physiological measures have been applied to the study of simultaneous interpreting. Petsche et al. (1993), for example, used electroencephalography (EEG) to compare brain activation as modulated by directionality (from or into the native language) during shadowing and simultaneous interpreting tasks. In order to avoid artifacts, however, both tasks had to be performed covertly (i.e., without articulation). Price et al. (1999) and Rinne et al. (2000) both used positron emission tomography (PET) to compare brain activation during translation and interpreting tasks. While Price et al. (1999) did not find any evidence for an increase in activation, Rinne et al. (2000) did. In both experiments, the method required the intravenous administration of $^{15}O\text{-}H_20$, a positron-emitting tracer, making it rather invasive. More recently, Hervais-Adelman et al. (2011) used functional magnetic resonance imaging (fMRI) and successfully identified the neural substrates underlying the simultaneous processing of two languages during simultaneous interpreting. Among the chief limitations of this method is its temporal resolution (in the range of 2 to 3 seconds), which makes it impossible to time lock certain cognitive load phenomena. Pupillometry, finally, a method developed several decades prior to the advent of modern brain-imaging techniques, would appear to have better temporal resolution, be less invasive and more affordable than the aforementioned methods. To substantiate these claims the rest of the discussion will focus on this method and how it applies to the measurement of cognitive load in simultaneous interpreting.

3. Pupillometry

3.1 The pupil

If we exclude the effect of drugs, which can cause constriction (as is the case with alcohol, opioids, and antipsychotics) or dilation (in the case of the central nervous system stimulants and hallucinogens), the pupil, broadly speaking, reacts to three kinds of stimuli: luminosity (Clarke et al. 2003), emotions (Stelmack and Mandelzys 1975) and cognitive activity (Kahneman et al. 1969). These three stimuli can cause the pupil size to vary from 1.5 mm in bright light to up to 8 to 9 mm in dim light (Andreassi 2000). In order to make sense of the pupillary reactions and, more specifically, of the peak amplitude of the pupil, it is important to consider some of its fundamental physiological characteristics. First of all, pupil response can occur as quickly as 200 ms after stimulus presentation (Lowenstein and Loewenfeld 1962), although dilation as a response to cognitive load usually seems to begin after 300 to 500 ms (Beatty 1982 and Hoeks and Levelt 1993). Lowenstein and Loewenfeld (1962) furthermore observed that pupil diameter is largest in rested individuals, whereas it decreases with fatigue. Similarly, it has been suggested that pupillary response decreases with age, weakening the correlation between cognitive load and pupil dilation (van Gerven et al. 2003). Another phenomenon to be kept in mind when interpreting pupillometric data is the manifestation of cognitive overload, i.e., when the task exceeds the cognitive resources available to perform it. Although Peavler (1974) suggests that once the capacity threshold has been reached, pupil dilation will stabilize, Poock (1973) and Granholm et al. (1996) found that pupil dilation decreases rapidly once cognitive overload was reached.

3.2 The technology

Since the initial interest in pupillometry in the 1960s, the technology used to measure pupil dilation has made considerable progress both in terms of accuracy and user-friendliness. In the early days of eye tracking[8], subjects had to be "firmly positioned in an adjustable head-holder" (Bradshaw 1968: 266), or use a "chin rest and bite board" (Schluroff 1982: 137) to maintain the distance between the eye and the tracker. Pupil size was then recorded with an external video camera, and the film was either plotted onto graph paper or projected onto a larger surface frame by frame; the pupil diameter was measured manually (see Bradshaw 1968). Today, pupillometry is considered a relatively simple, affordable and non-invasive method for assessing autonomic function (Bär et al. 2005), and is applied in psychophysiology, pharmacology, neurology, and psychiatry. While until recently, only fixed (or head-mounted) eye trackers were deemed suitable for the measurement of pupil dilation, Klingner et al. (2008) convincingly replicated some of the classic cognitive pupillometry results (e.g., Kahneman and Beatty 1966) using a remote eye tracker.[9]

3.3 Using pupillometry to measure cognitive load in simultaneous interpreting

From the preceding discussion one might conclude that pupillometry is ideally suited to measure cognitive load during a complex task like simultaneous interpreting. It might surprise, then, to see that it has only been applied to the interpreting paradigm very rarely. In a groundbreaking experiment, Tommola and Niemi (1986) used pupillometry to study the effect of directionality-contingent syntactic complexity on mental load, for the first time demonstrating the feasibility of the method. Almost a decade later, Hyönä et al. (1995) conducted a more systematic analysis of cognitive load comparing and contrasting the load generated during different (arguably related) language processing tasks (i.e., listening comprehension, shadowing, and simultaneous interpreting). The results indicated an increase of mean pupil dilation, and consequently cognitive load, from listening comprehension to shadowing, and simultaneous interpreting. Seeber and Kerzel used the same method to measure online (i.e., real time) cognitive load during simultaneous interpreting and gathered evidence suggesting that German verb-final structures generate more cognitive load than German verb-initial structures when interpreted into English (Seeber and Kerzel 2012). Their data also suggests that simultaneously interpreting sentences without context generates more cognitive load than interpreting sentences (ceteris paribus) embedded in context. These examples illustrate the great potential of pupillometry as a method and t TEPRs as a measure of cognitive load in simultaneous interpreting.

3.4 Potential and limitations of pupillometry in simultaneous interpreting

Like all research methods, pupillometry has its potential and limitations. Its potential, largely due to its limited invasiveness and its high temporal resolution, has already been discussed in the previous sections. As for its limitations, they mainly stem from the intrinsic nature of the measures it provides. Mean dilation, peak dilation and mean latency are revealing when applied to short auditory stimuli (e.g., at the phrase and sentence level), but much less so when applied to long stimuli (e.g., at the discourse level). In fact, Schultheis and Jameson (2004) found no difference in mean pupil dilation across (340-word long) texts of different (subjective) difficulty. Their conclusion that "pupil size may differ between easy and difficult conditions only in certain periods of a task" (2004: 233) is supported by

Haapalainen et al. (2010), who found that median heart flux and median electrocardiogram are more reliable measures of task difficulty than median pupil dilation with a trial length of three minutes. These findings, it seems, are related to the very nature of the measure that reflects moment-to-moment variations of load and maps several sources of load onto one measure. Averaging changes in pupil dilation over long periods (e.g., of several minutes) might thus cancel out the changes in cognitive load reflected in them. It is no coincidence that the experiments in which this method has been applied successfully (both outside and within the paradigm) were tightly controlled and used isolated stimuli and short periods of interest (e.g., Hyönä et al., Seeber and Kerzel 2012). Th is suggests that, provided the necessary methodological rigor is applied, pupillometry might be a reliable method to measure local cognitive load in simultaneous interpreting.

4. Conclusion

In this article I attempted to briefly illustrate, by means of a few examples, how different methods have been applied to researching cognitive load in simultaneous interpreting. Although each of the four approaches has its unique advantages and disadvantages, this analysis suggests the use of pupillometry as a way to observe more objectively how much cognitive load is generated during the interpreting task as it unfolds. Having said that, even this measure currently comes with as of yet unresolved challenges. In fact, it would appear that while reliable as a means to assess local cognitive load at or below the sentence level, this method might not be indicated to quantify average cognitive load across long stimuli. What is more, the signal-to-noise ratio[10] requires comprehensive data preparation (see Klingner et al. 2008) while the many-to-one mapping still precludes us from attributing measured load to individual component tasks. Much as with other limitations identified in this overview, they are not to suggest that the method is unsuitable or invalid. However, it would appear that we ***do*** require more "research into research" and that, "the day when we can spend more time discussing ***what*** we found, and less time agonizing over ***how*** we found it or whether we went about it the right way" (Shlesinger 2000, 13) has not yet come.

Notes

1. See Seeber (2011) for a more comprehensive discussion on the definition of cognitive load and how it relates to simultaneous interpreting.

2. For a more detailed description, see Wickens (1984, 2002) as well as Seeber (2007, 2011).

3. The maximum amount of load generated in a particular period of interest.

4. As pointed out above, analytical approaches are unable to take into account individual differences: for the sake of the exercise it is thus assumed that the interpreter is unfamiliar with the name of the organizations he is working for or the venues he is working at and has to process the input incrementally without being able to resort to anticipation.

5. For a comprehensive discussion see Pym (2008).

6. The extent to which measurements are influenced and falsified by artifacts.

7. The extent to which a method invades the physical integrity of the body.

8. Pupil dilation is usually measured using an eye tracker, i.e. an infrared camera measuring the movement of the eyes.

9. See Klingner (2010) for a detailed description of the methodology including data processing.

10. The level of desired signal compared to the level of background noise (or artifacts).

References

Andreassi, John L. 2000. *Psychophysiology: Human Behavior and Physiological Response*, 4th ed. Mahwah, NJ: Lawrence Erlbaum.

Bär, Klaus-Jürgen, Michael K. Boettger, Silke Till, Julia Dolicek, and Heinrich Sauer. 2005. "Lateralization of Pupillary Light Reflex Parameters." *Clinical Neuropsychology* 116: 790–798.

Barik, Henri C. 1969. *A Study of Simultaneous Interpretation*. Unpublished PhD dissertation, University of North Carolina, Chapel Hill.

Beatty, Jackson. 1982. "Task Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources." *Psychological Bulletin* 91 (2): 276–292.

Bernardini, Silvia. 2001. "Think-Aloud Protocols in Translation Research: Achievements, Limits, Future Prospects." *Target* 13 (2): 214–263.

Bradshaw, John L. 1968. "Load and Pupillary Changes in Continuous Processing Tasks." *British Journal of Psychology* 59 (3): 265–271.

Clarke, Robert J., Hongyu Zhang, and Paul D.R. Gamlin. 2003. "Characteristics of the Pupillary Light Reflex in the Alert Rhesus Monkey." *Journal of Neurophysiology* 89: 3179–3189.

De Groot, Annette M.B. 1997. "The Cognitive Study of Translation and Interpretation. Three Approaches." In *Cognitive Processes in Translation and Interpreting*, ed. by Joseph H. Danks, Gregory M. Shreve, Stephen B. Fountain, and Michael K. McBeath, 25–56. Thousand Oaks, London, New Delhi: Sage Publications.

De Groot, Annette M.B. 2000. "A Complex-skill Approach to Translation and Interpreting." In *Tapping and Mapping the Processes of Translation and Interpreting*, ed. by Sonja Tirkkonen-Condit and Riitta Jääskeläinen, 53–68. Amsterdam: John Benjamins.

Ericsson, Anders K., and Herbert A. Simon. 1984. *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: Bradford Books/MIT Press.

Gile, Daniel. 1995. *Regards sur la recherche en interprétation de conférence*. Lille: Presses universitaires de Lille.

Gile, Daniel. 1997. "Conference Interpreting as a Cognitive Management Problem." In *Cognitive Processes in Translation and Interpreting*, ed. by Joseph H. Danks, Stephen B. Fountain, Michael K. McBeath, and Gregory M. Shreve, 196–214. Thousand Oaks, London, New Delhi: Sage Publications.

Gile, Daniel. 1999. "Testing the Effort Models' Tightrope Hypothesis in Simultaneous Interpreting — A Contribution." *Hermes* 23: 153–172.

Gile, Daniel. 2008. "Local Cognitive Load in Simultaneous Interpreting and its Implications for Empirical Research." *Forum* 6 (2): 59–77.

Gile, Daniel. 2011. "Errors, Omissions and Infelicities in Broadcast Interpreting: Preliminary Findings from a Case Study." In *Methods and Strategies of Process Research*, ed. by Birgitta Englund Dimitrova, Cecilia Alvstad, Adelina Hild, and Elisabet Tiselius, 201–218. Amsterdam: John Benjamins.

Gopher, Daniel, and Rolf Braune. 1984. "On the Psychophysics of Workload: Why Bother with Subjective Measures?" *Human Factors* 26: 519–532.

Granholm, Eric, Robert F. Asarnow, Andrew J. Sarkin, and Karen L. Dykes. 1996. "Pupillary Responses Index Cognitive Resource Limitations." *Psychophysiology* 33: 457–461.

Haapalainen, Eija, SeungJun Kim, Jodi F. Forlizzi, and Anind K. Dey. 2010. "Psycho-Physiological Measures for Assessing Cognitive Load." In *Proceedings of the 12th International Conference on Ubiquitous Computing*, 301–310. Copenhagen.

Hervais-Adelman, Alexis G., Barbara Moser-Mercer, and Narly Golestani. 2011. "Executive Control of Language in the Bilingual Brain: Integrating the Evidence from Neuroimaging to Neuropsychology." *Frontiers in Psychology* 2 (234): 1–8.

Hoeks, Bert, and Willem J.M. Levelt. 1993. "Pupillary Dilation as a Measure of Attention: A Quantitative System Analysis." *Behavior Research Methods, Instruments and Computers* 25: 16–26.

Hyönä, Jukka, Jorma Tommola, and Anna-Mari Alaja. 1995. "Pupil Dilation as a Measure of Processing Load in Simultaneous Interpreting and Other Language Tasks." *The Quarterly Journal of Experimental Psychology* 48A (3): 598–612.

Ivanova, Adelina. 2000. "The Use of Retrospection in Research on Simultaneous Interpreting." In *Tapping and Mapping the Processes of Translation and Interpreting*, ed. by Sonja Tirkkonen-Condit and Riitta Jääskeläinen, 27–52. Amsterdam: John Benjamins.

Kahneman, Daniel. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, Daniel, and Jackson Beatty. 1966. "Pupil Diameter and Load on Memory." *Science* 154: 1583–1585.

Kahneman, Daniel, Bernard Tursky, David Shapiro, and Andrew Crider. 1969. "Pupillary, Heart Rate, and Skin Resistance Changes during a Mental Task." *Journal of Experimental Psychology* 79 (1): 164–167.

Klingner, Jeff. 2010. *Measuring Cognitive Load during Visual Tasks by Combining Pupillometry and Eye Tracking*. Unpublished PhD dissertation, Stanford University Computer Science Department.

Klingner, Jeff, Rakshit Kumar, and Pat Hanrahan. 2008. "Measuring the Task-Evoked Pupillary Response with a Remote Eye Tracker." ETRA 2008: *Proceedings of the 2008 Symposium on Eye Tracking Research and Applications*, 69–72. Savannah, Georgia: ACM.

Lamberger-Felber, Heike. 2001. "Text-Oriented Research into Interpreting: Examples from a Case-Study." *Hermes* 26: 39–64.

Lowenstein, Otto, and Irene E. Loewenfeld. 1962. "The Pupil." In *The Eye*, Vol. 3, *Muscular Mechanisms*, ed. by Hugh Davson, 231–267. New York: Academic Press.

Massaro, Dominic W., and Miriam Shlesinger. 1997. "Information Processing and a Computational Approach to the Study of Simultaneous Interpretation." *Interpreting* 1 (1/2): 13–53.

Mazza, Cristina. 2000. *Numbers in Simultaneous Interpretation*. Unpublished graduation thesis, Universita degli Studi di Bologna, SSLMIT, Forli.

Miller, George A. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information." *Psychological Review* 63: 81–97.

Mital, Anil, and Majorkumar Govindaraju. 1999. "Is It Possible to Have a Single Measure for all Work?" *International Journal of Industrial Engineering Theory* 6: 190–195.

Mitchell, Don. 2004. "On-line Methods in Language Processing: Introduction and Historical Review." In *The On-line Study of Sentence Comprehension: Eyetracking, EPRs and Beyond*, ed. by Manuel Carreiras and Charles Clifton Jr., 15–32. Hove: Taylor & Francis.

Moser-Mercer, Barbara. 1997. "Beyond Curiosity. Can Interpreting Research Meet the Challenge?" In *Cognitive Processes in Translation and Interpreting*, ed. by Joseph H. Danks, Stephen B. Fountain, Michael K. McBeath, and Gregory M. Shreve, 176–195. Thousand Oaks, London, New Delhi: Sage Publications.

Moser-Mercer, Barbara, Ulrich Frauenfelder, Beatriz Casado, and Alexander Künzli. 2000. "Searching to Define Expertise in Interpreting." In *Language Processing and Simultaneous Interpreting*, ed. by Birgitta Englund Dimitrova and Kenneth Hyltenstam, 107–132. Amsterdam: John Benjamins.

Oléron, Pierre, and Hubert Nanpon. 1965. "Recherches sur la traduction simultanée." *Journal de psychologie normale et pathologique* 62: 73–94.

Paas, Fred G.W.C., and Jeroen J.G. Merrienboer. 1993. "The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures." *Human Factors* 35: 737–743.

Paas, Fred G.W.C., Juhani E. Tuovinen, Huib K. Tabbers, and Pascal W.M. van Gerven. 2003. "Cognitive Load Measurement as a Means to Advance Cognitive Load Theory." *Educational Psychologist* 38 (1): 63–71.

Peavler, Scott W. 1974. "Pupil Size, Information Overload, and Performance Differences." *Psychophysiology* 11: 559–566.

Petsche, Hellmuth, Susan C. Etlinger, and Oliver Filz. 1993. "Brain Electrical Mechanisms of Bilingual Speech Management: An Initial Investigation." *Electroencephalography and Clinical Neurophysiology* 86: 385–394.

Pöchhacker, Franz. 1994. *Simultandolmetschen als komplexes Handeln*. Tübingen: Gunter Narr.

Poock, Gary K. 1973. "Information Processing vs. Pupil Diameter." *Perceptual and Motor Skills* 37: 1000–1002.

Price, Cathy J., David W. Green, and Roswitha Von Studnitz. 1999. "A Functional Imaging Study of Translation and Language Switching." *Brain* 122 (12): 2221–2235.

Pym, Anthony D. 2008. "On Omission in Simultaneous Interpreting: Risk Analysis of a Hidden Effort." In *Efforts and Models in Interpreting and Translation Research: A Tribute to Daniel Gile*, ed. by Gyde Hansen, Andrew Chesterman, and Heidrun Gerzymisch-Arbogast, 83–105. Amsterdam: John Benjamins.

Rinne, Juha O., Jorma Tommola, Matti Laine, Bernd J. Krause, Daniela Schmidt, Valtteri Kaasinen, Mika Teräs, Hannu Sipilä, and Marianna Sunnari. 2000. "The Translating Brain: Cerebral Activation Patterns During Simultaneous Interpreting." *Neuroscience Letters* 294: 85–88.

Shlesinger, Miriam. 2000. "Interpreting as a Cognitive Process: How can we know what happens?" In *Tapping and Mapping the Processes of Translation and Interpreting*, ed. by Sonja Tirkkonen-Condit and Riitta Jääskeläinen, 3–15. Amsterdam: John Benjamins.

Schluroff, Michael. 1982. "Pupil Responses to Grammatical Complexity of Sentences." *Brain and Language* 17: 133–145.
Schultheis, Holger, and Anthony Jameson. 2004. "Assessing Cognitive Load in Adaptive Hypermedia Systems: Physiological and Behavioral Methods." In *Adaptive Hypermedia and Adaptive Web-Based Systems: Proceedings of AH 2004*, ed. by Wolfgang Nejdl and Paul De Bra, 225–234. Berlin: Springer.
Seeber, Kilian G. 2007. "Thinking outside the Cube: Modeling Language Processing Tasks in a Multiple Resource Paradigm." *Interspeech* 2007, Antwerp, Belgium. 1382–1385.
Seeber, Kilian G. 2011. "Cognitive Load in Simultaneous Interpreting: Existing Theories — New Models." *Interpreting* 13 (2): 176–204.
Seeber, Kilian G., and Dirk Kerzel. 2012. "Cognitive Load in Simultaneous Interpreting: Model Meets Data." *International Journal of Bilingualism* 16 (2): 228–242.
Setton, Robin. 2001. "Translation Studies and Cognitive Science: Do we Need each other?" *CTIS Occasional Papers* 1: 113–126.
Setton, Robin. 2003. "Models of the Interpreting Process." In *Avances en la investigación sobre la interpretación*, ed. by Angela Collados Aís and José Antonio Sabio Panilla, 29–91. Granada: Editorial Comares.
Stelmack, Robert M., and Nathan Mandelzys. 1975. "Extraversion and Pupillary Response to Affective and Taboo Words." *Psychophysiology* 12: 536–540.
Tauchmanová, Jana. 2011. *Daniel Gile´s Effort Model in Simultaneous Interpreting*. Unpublished MA thesis, Institute of Translation Studies, Charles University, Prague.
Tommola, Jorma, and Pekka Niemi. 1986. "Mental Load in Simultaneous Interpreting: An On-line Pilot Study." In *Nordic Research in Text Linguistics and Discourse Analysis*, ed. by Lars S. Evensen, 171–184. Trondheim: Tapir.
Van Gerven, Pascal W.M., Fred Paas, Jeroen J.G. van Merrienboer, and Henk Schmidt. 2003. "Memory Load and the Cognitive Pupillary Response in Aging." *Psychophysiology* 41: 167–174.
Wickens, Christopher D. 1984. "Processing Resources in Attention." In *Varieties of Attention*, ed. by Raja Parasuraman and David R. Davies, 63–102. New York: Academic Press.
Wickens, Christopher D. 2002. "Multiple Resources and Performance Prediction." *Theoretical Issues in Ergonomics Science* 3 (2): 159–177.
Woodworth, Robert S. 1899. "The Accuracy of Voluntary Movement." *Psychological Review* 3 (Suppl. 13): 1–119.
Yin, Bo, Fang Chen, Natalie Ruiz, and Eliathamby Ambikairajah. 2008. "Speech-Based Cognitive Load Monitoring System." *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2041–2044. Las Vegas, Nevada.

Author's address

Kilian G. Seeber
University of Geneva
40 bd du Pont D'Arve
CH-1205 GENEVA
Switzerland

Kilian.Seeber@unige.ch