



Audiovisual translation as multimodal mediation

Jan-Louis Kruger

Collaborators: Sixin Liao, Erik Reichle, Lili Yu, Natalia Wisniewska

*“The soul, fortunately, has an interpreter –
often an unconscious but still a faithful
interpreter – in the eye”.
(Charlotte Brontë 1847)*

Introduction

- Digital revolution of VoD and streaming services like Netflix, Stan, Prime Video, etc. makes VHS and DVD a distant memory.
- Multiple devices and contexts.
- Resulting in boom in use of subtitles – way majority of the population in many countries engage with media.
- Subtitling increasingly automated.
- Fansubbing has an increasingly prominent role.
- Remains an important tool in multimodal intercultural mediation.



But what has this meant for AVT?

The cost of universal access

- Universal access comes at a cost:
 - Amateur subtitlers, audio describers & multinational companies and content providers prioritise economy, volume and efficiency over quality and immersion.
 - How good is good enough? What happens to the viewer experience?
- Let's look at verbatim subtitles.
 - Fast
 - Variable
- “Speed in SDH is as much a technical matter as it is economic (broadcasters, service providers), political and ideological (deaf associations)” (Romero-Fresco, 2009:110).
- The demand for verbatim subtitles is not limited to SDH - many L1&2 viewers also demand full and accurate transcripts of the dialogue.
- This is often a rather charged debate, but requires more research.
- Recent studies have started to engage with the impact of presentation rate on processing.

The age of access

- We are witnessing:
 - Astonishing volumes of subtitled material produced daily
 - Increasing role of technology
 - Increasing variability in quality
 - But, *increasing access*.
- With increasing access and volume, need arises to look at how subtitling functions as multimodal mediation.
- Subtitles represent only one component of film as multimodal text.
- The majority of advances in the “automation” of subtitling, however, tends to shift the emphasis entirely onto the transcription of speech.
- This has resulted in an increase in the presentation rate of subtitles in many cases (leaving other quality issues aside for the time being).

Aims

-
- To provide a brief introduction to eye-tracking research.
 - To provide an **overview** of some studies on **subtitle processing**.
 - To introduce a **multimodal integrated-language framework**.
 - To share the **methodology** of a study designed to determine the impact of presentation speed on viewers (comprehension, attention allocation, and reading).
 - To present the **findings** this study.

AVT and multimodality

-
- In AVT we simultaneously process spoken and written language, soundtrack and visual information.
 - Information in subtitles depends on, supplements, or repeats what is available in the other modes.
 - NB: Film is dynamic – unlike in reading a book, viewers have to prioritise different sources of information at different times.

Background on eye tracking in AVT

-
- Eye tracking has gained a lot of ground in media accessibility research.
 - Originated in the 1980s through the work of Gery d'Ydewalle and colleagues, it now includes a range of different measures in diverse research designs, and inspired novel avenues of research and application.
 - Early studies focussed mainly on attention distribution between subtitles and images, and amount of visual attention (dwell time) in the subtitles.
 - More recently, ET has been used to interrogate (among other things):
 - the impact of translation strategies on subtitle reading,
 - the impact of shot changes, text segmentation, placement and presentation speed on subtitle processing,
 - the impact of subtitles on the effectiveness of visual processing of film,
 - the amount of cognitive load induced by subtitles.

How are subtitles read?

- Previous studies on subtitle reading were concerned with
 - viewer responses (cf. Braverman & Hertzog, 1980; Jensema, 1998; Romero-Fresco, 2009; 2015)
 - attention distribution (d'Ydewalle, Muylle & Rensbergen, 1982; D'Ydewalle & Gielen, 1992; Koolstra, Van Der Voort & d'Ydewalle, 1999; Jensema, Danturthi & Burch, 2000; Winke, Gass & Sydorenko, 2013)
 - cognitive processing (cf. Perego, Del Missier, Porta & Mosconi, 2010; Bisson, Van Heuven, Conklin & Tunney, 2012; Szarkowska, Krejtz, Pilipczuk, Dutka & Kruger, 2016; Szarkowska, Krejtz, Klyszejko & Wieczorek, 2011; Szarkowska A, Gerber-Moro, 2018)

The limits of visual perception

-
- We can only see a very small window in enough detail at any time to extract meaningful information from that part of the visual scene.
 - Need to shift gaze continually in routines adapted to type of scene.
 - Subtitles must be read to extract meaning, but gaze is also attracted to movement, scene changes, etc..
 - What do we know about these eye movements?

Exercise

Just how much can we see with peripheral vision?

Just how much can we see with peripheral vision

Just how much can we see with peripheral vision

Just how much can we see with peripheral vision

Image and text



the kangaroo one of Australia's most recognizable marsupials

Foveal vision (Tobii.com)

<https://www.slideshare.net/AcuityETS/eye-tracking-in-usability>



MACQUARIE
University

tobii
www.tobii.com

What do we actually see?

- The 1-2° large area of foveal vision is about the size of a thumbnail on an arm lengths distance
- The fovea is less than 1% of the retina but takes up over 50% of the visual cortex in the brain.
- Peripheral vision is mainly good at picking up movements and contrasts



Why eye tracking?

The importance of research questions

- Without an **objective measure like eye tracking**, hard to prove that subtitle reading improves **comprehension**, reduces **CL**, increases **immersion**, etc.
- ET reveals where viewers are looking and how they read subtitles.
- Remains indirect if we only look at global measures:
 - average DT
 - fixation count
 - fixation duration
 - crossovers, etc.
- We need to answer questions such as:
 - How does L1vs.L2/layout/speed impact on global measures and attention distribution?
 - How many of the words are processed? (skipping rate)
 - How are words/phrases/subtitles processed? (skimming, refixations, regressions)
 - How does video and soundtrack presence/complexity/redundancy impact on subtitle reading?
- Many questions are hard to answer without word-level data.

Introduction to measures

Some ET measures used to study processing of subtitles:

- Global measures
 - Common (average time spent on subtitles vs. video, fixation counts and durations, dwell time)
 - Less common (word skipping rates during first-pass reading and refixation probabilities)
- Local measures
 - word-level measures to study frequency effects, word-length effects, and wrap-up effects.

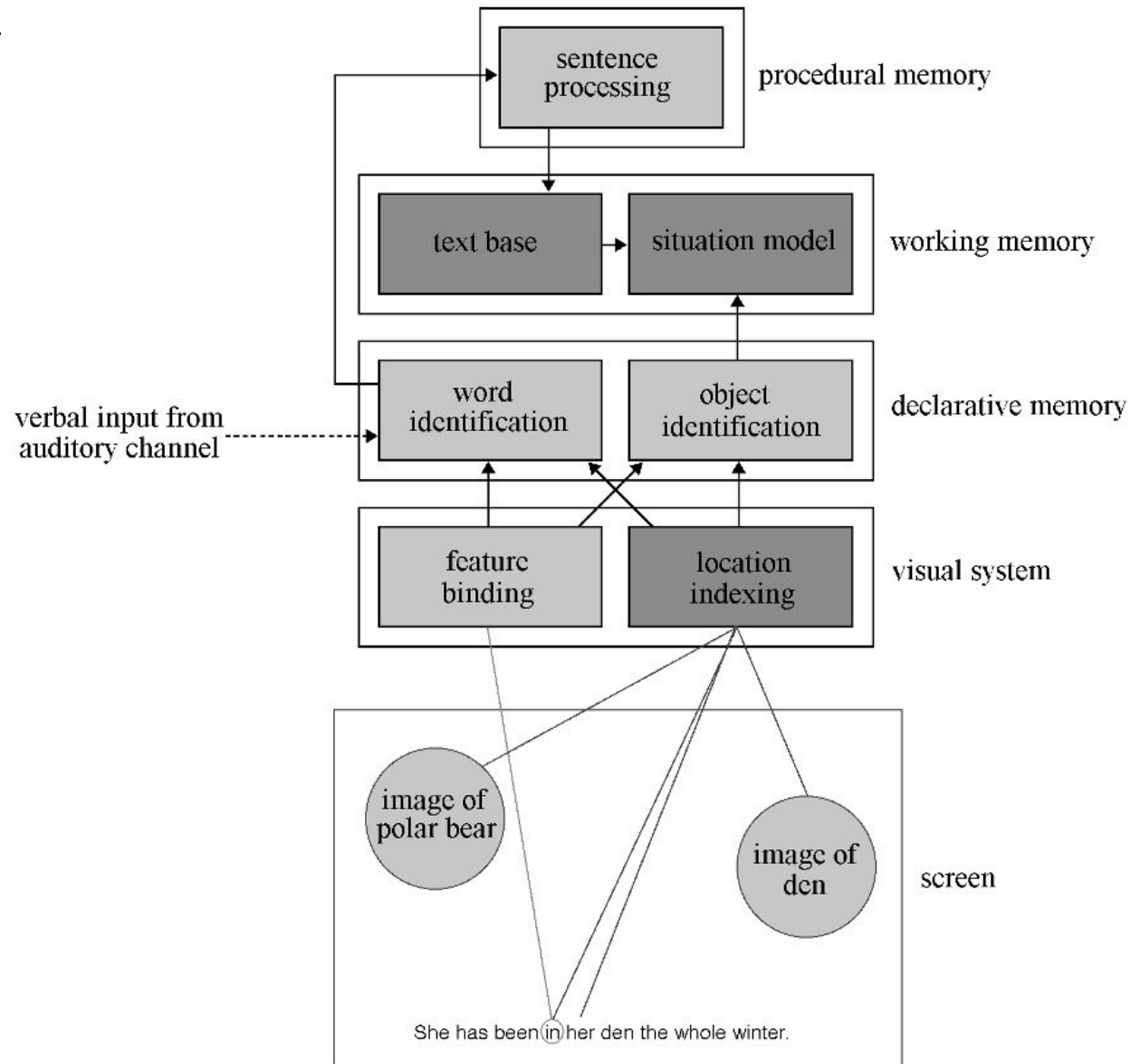
Skilled reading

-
- Skilled reading is cognitively demanding (cf. Dehaene, 2009).
 - Readers have to coordinate systems responsible for vision, attention allocation, language processing, working and long-term memory, and eye movement control.
 - When we read, it's hard to do other things.
 - Reading subtitles: constantly shift our gaze from word to word, from subtitle to screen.
 - Can't read and look at the screen simultaneously: temporary "blindness".
 - "Obstacles" such as less familiar or long words or grammatical structures, ambiguous line segmentation or punctuation, or errors interrupt reading.
 - I.e. reading can be tripped up – even more so when it has to compete for attention with dynamic elements of the screen. And vice versa.

Visual routines

-
- Reading and scene perception are two independent channels calling for radically different visual routines.
 - Reading subtitles is a truncated reading routine due to the presentation in chunks and the alternation between the two routines.
 - This fundamentally changes the reading routine with shorter fixations and fewer regressions.

A multimodal integrated-language framework



Automaticity and efficiency

- D'Ydewalle and Van Rensbergen (1989) found that viewers read subtitles automatically, regardless of condition (i.e. standard or reversed).
- This has been confirmed in a number of studies (e.g. d'Ydewalle et al., 1991, d'Ydewalle and de Bruycker, 2007; Bisson et al. 2012).
- A common theme in these studies is that viewers pay more attention to subtitles when they need them (i.e. when the audio is in a foreign language and the subtitles in L1).
- Nevertheless, even redundant subtitles (SLS) are read.
- Furthermore, d'Ydewalle and colleagues found that subtitles are processed more efficiently than spoken language particularly in demanding content.

Processing multimodal redundancy

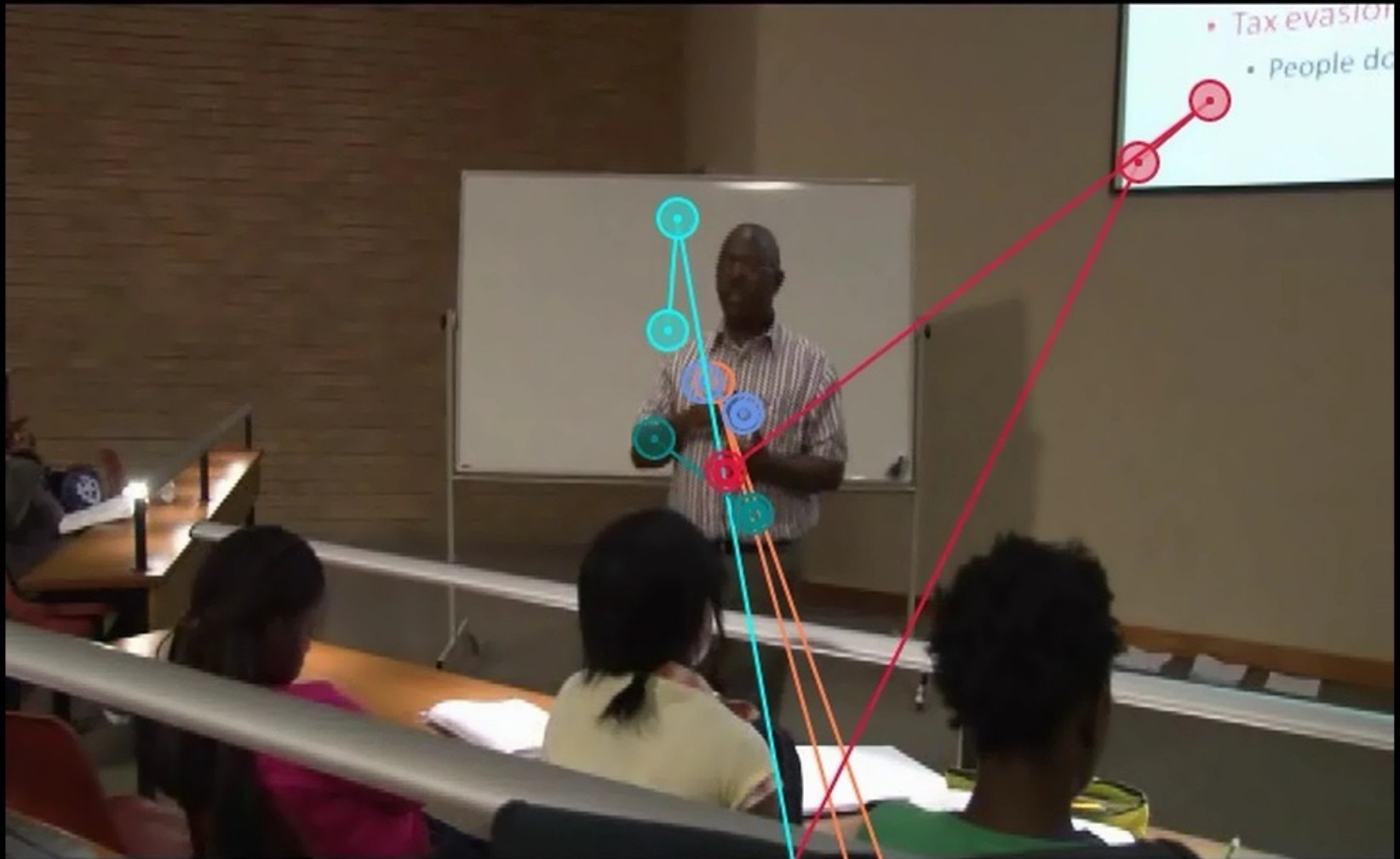
- Viewers can process multimodal redundancy with astonishing efficiency.
- Often triggers involuntary checking or comparison – errors pop out.



Cognitive effectiveness and adaptive processing

- Viewers can adapt to new conditions.
- The mind can assess relative importance of sources & prioritise those that are most critical to an understanding of the text.
- Perego et al. (2010): viewers can process subtitles effectively without compromising the processing of visuals.
- Krejtz et al. (2013): even when subtitles go over shot changes, viewers either carry on reading the subtitles, or correct themselves very quickly should they have shifted their gaze to the beginning of the line after the shot change.

Multiple sources



Most of those are tax deductible.

Saliency

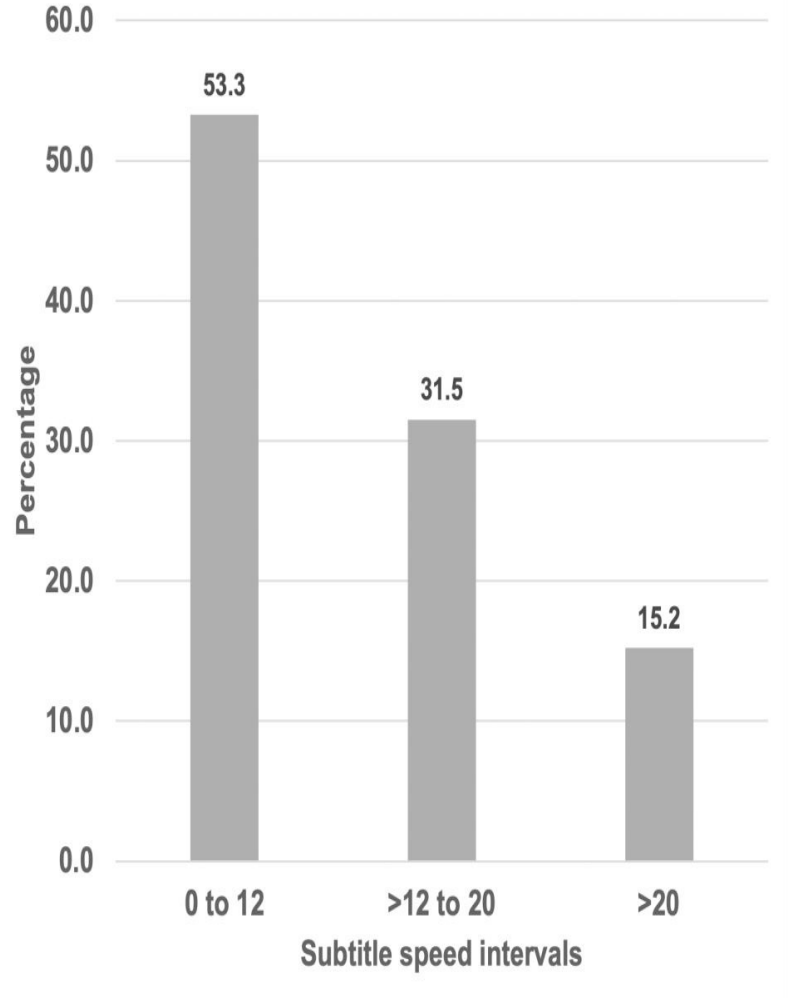
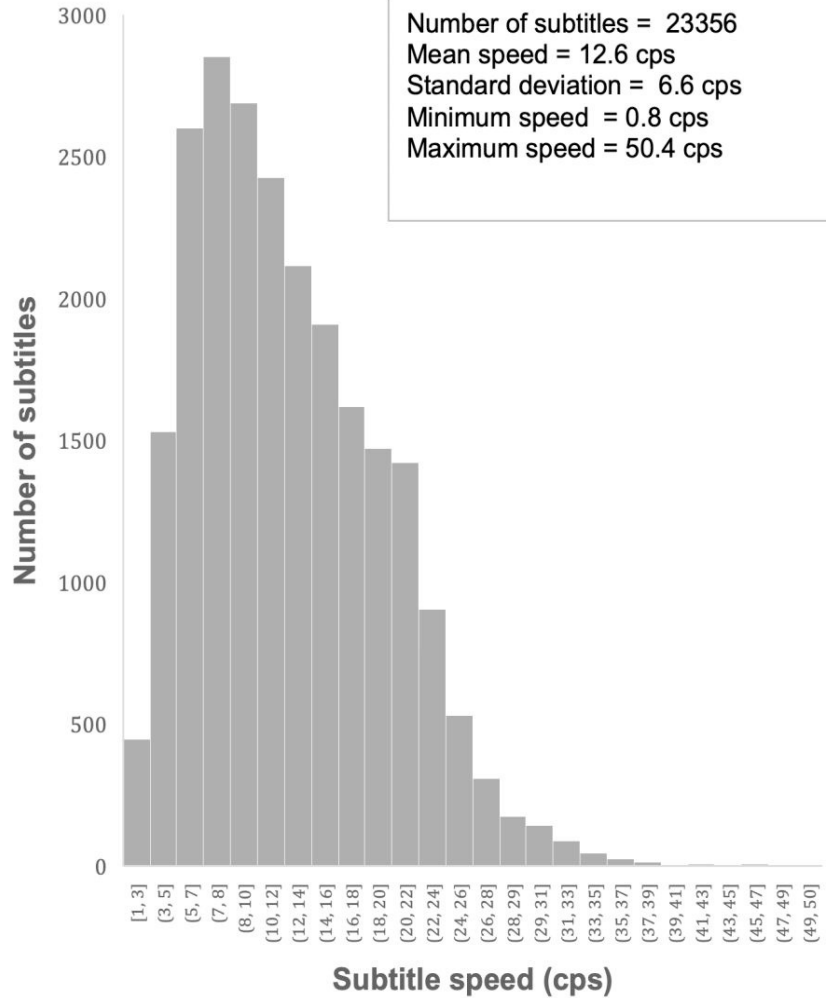


Example: Presentation rate or subtitle speed

-
- Subtitle speed is a contentious issue.
 - Jensema (1998) found that most viewers (20 years ago) are comfortable with 145wpm (12cps), and most can even handle speeds of up to 170wpm (14cps).
 - Szarkowska and colleagues (e.g. 2011, 2016) investigated the processing of verbatim, standard and edited subtitles and found that for hearing audiences, there was very little difference in the amount of time spent in the subtitles.
 - Szarkowska and Gerber Morón (2018) found that viewers can process subtitles at 20cps and still follow the images.
 - Verbatim subtitles have variable speed.

Distribution of subtitle speed across a sample of 11 popular films on Netflix

Number of subtitles = 23356
Mean speed = 12.6 cps
Standard deviation = 6.6 cps
Minimum speed = 0.8 cps
Maximum speed = 50.4 cps

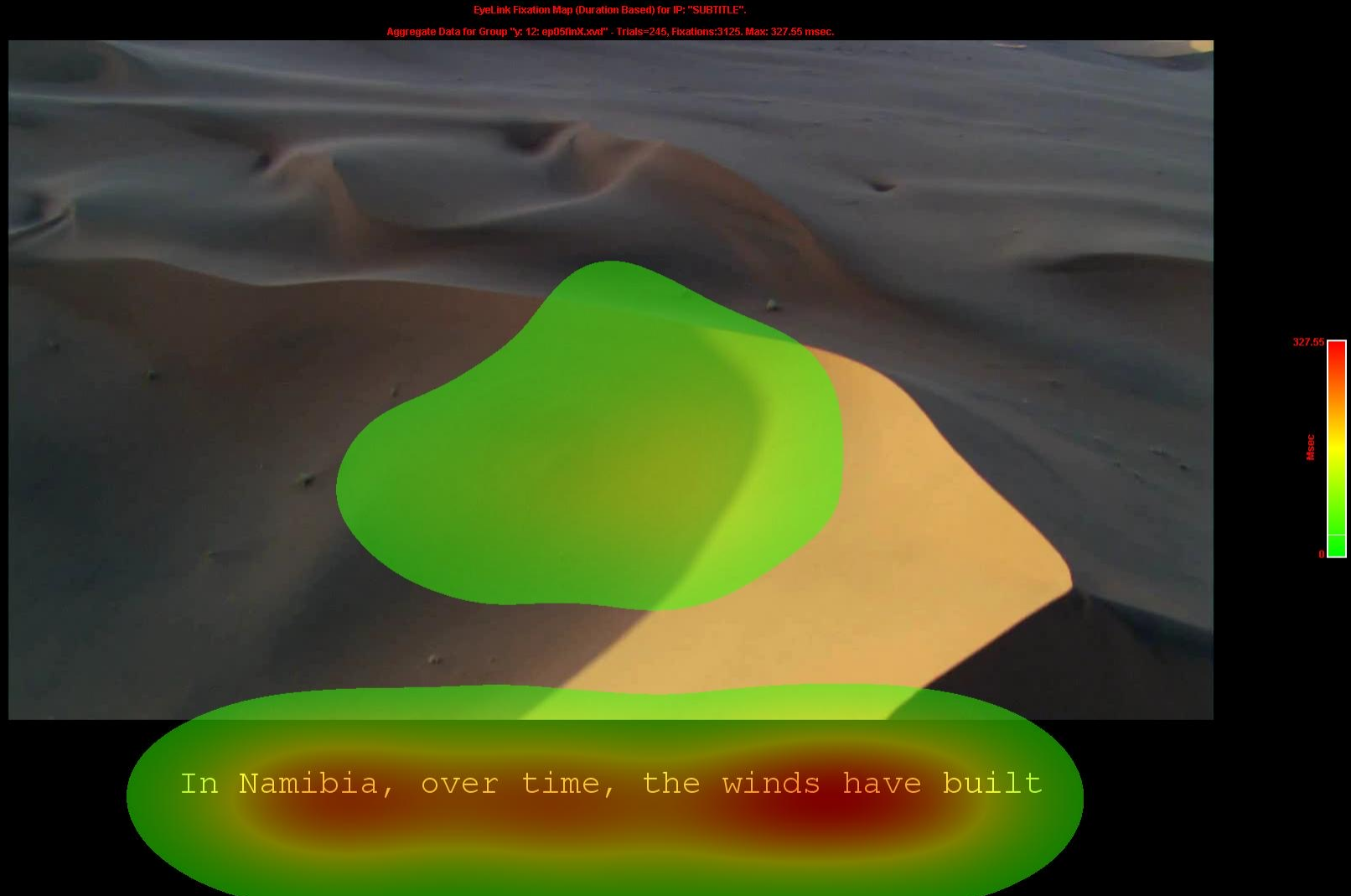


Experimental design

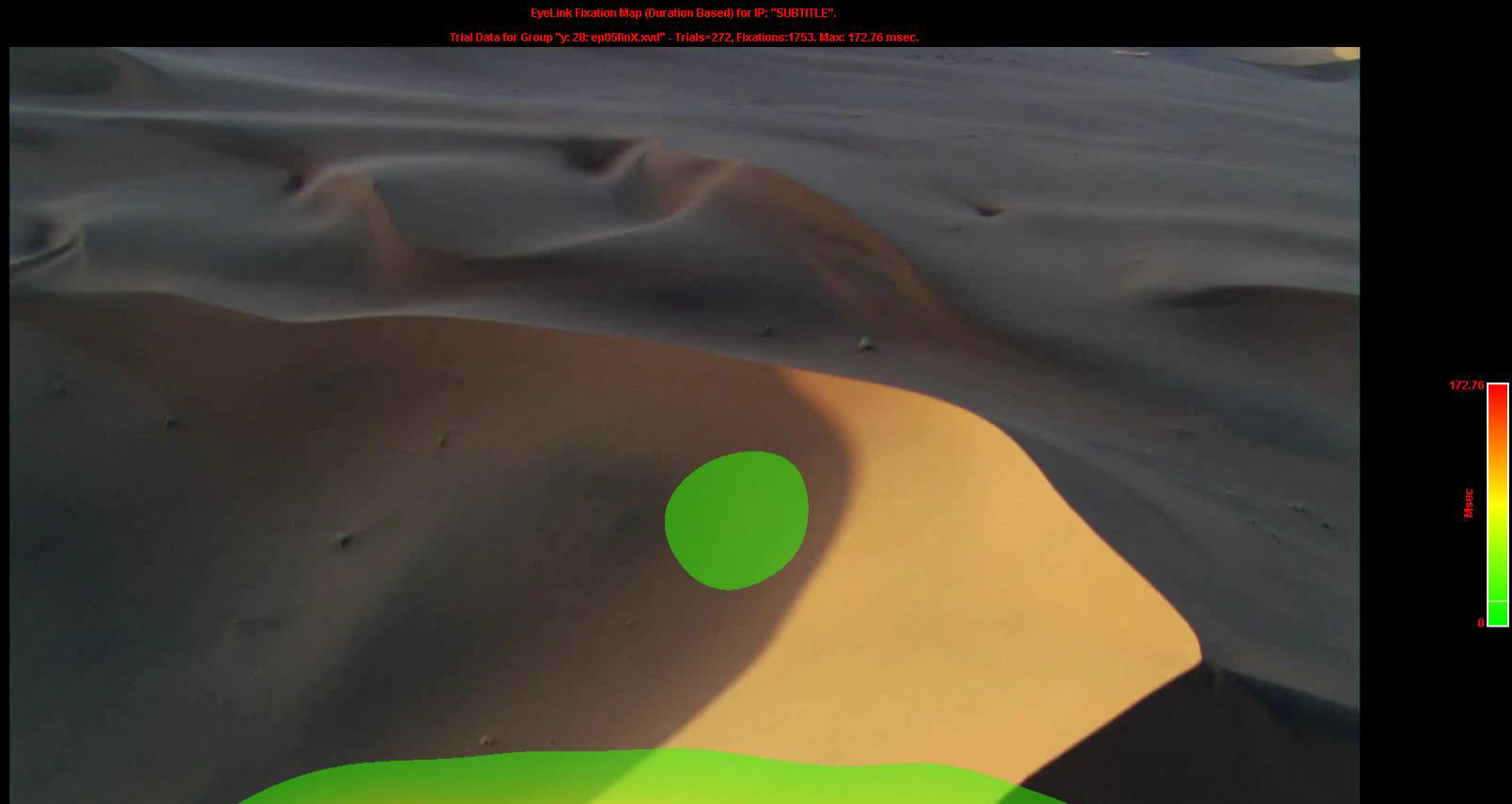
- What happens when subtitles are faster?
- To answer this question, we designed a study in which we kept the volume of text per subtitle stable, and also the layout (only one-liners), and only varied the duration on screen (12, 20 and 28cps).
- Since sound is an important factor as an auditory cue for the onset of the subtitle, we removed the sound.
- We also presented the subtitles either with or without video to investigate the impact of video on subtitle processing at the different speeds, and to have a baseline reading condition.

-
- 2 (**Video conditions:** with video, no video) * 3 (**Subtitle speed conditions:** 12cps, 20cps, and 28cps).
 - No sound was used and the volume of text was kept identical across all three speeds
 - 6 videos were used for each experiment, thus, each of the conditions included 1 video and its subtitles (~80).
 - 8 questions related to the **subtitle only** were presented by the end of each video, thus, each participant need to answer $8*6=48$ questions in total.

Heatmap at 12cps



Heatmap at 28cps



In Namibia, over time, the winds have built

12cps



~~So they lick saliva onto their forearms.~~

28cps

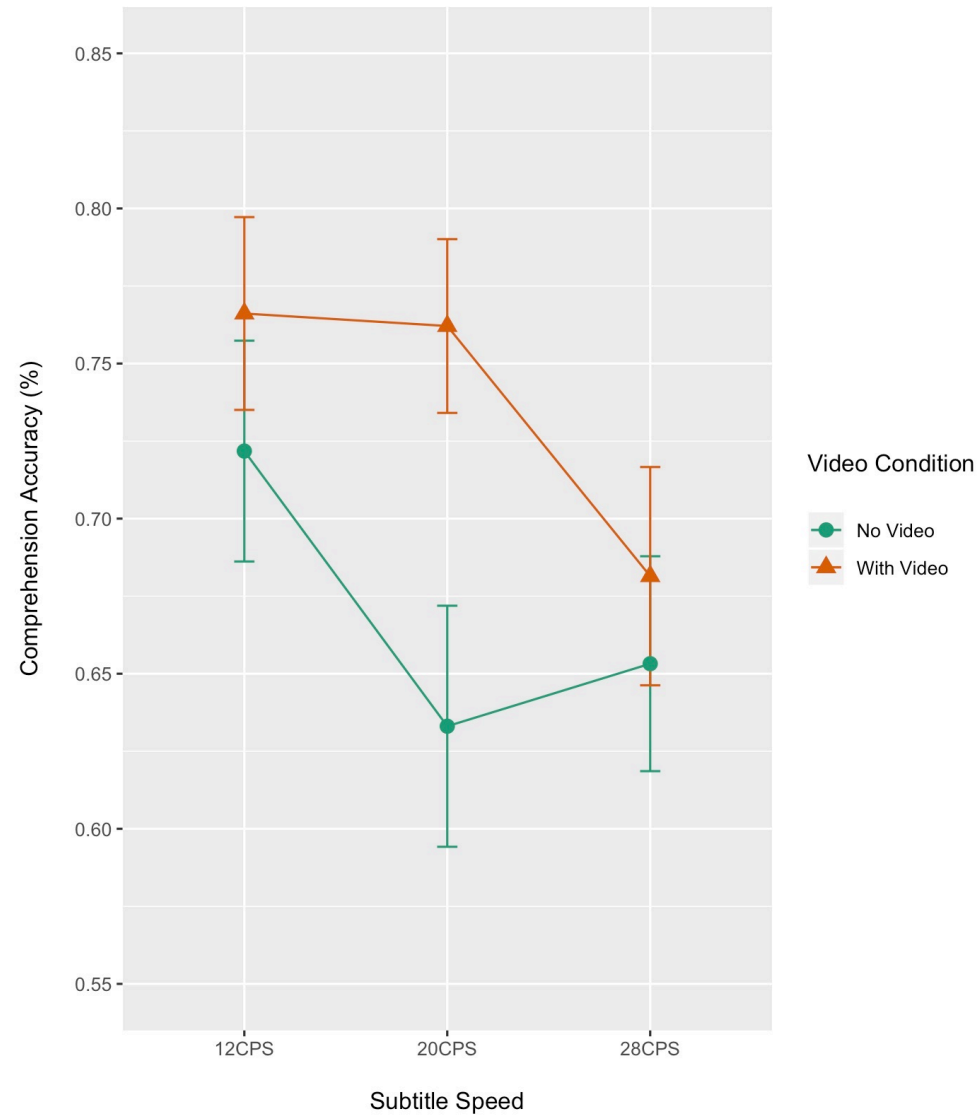


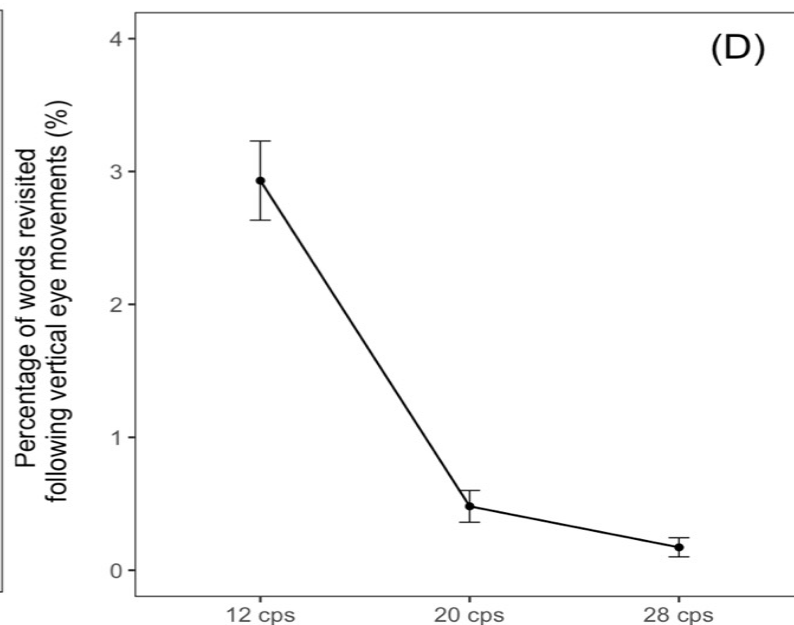
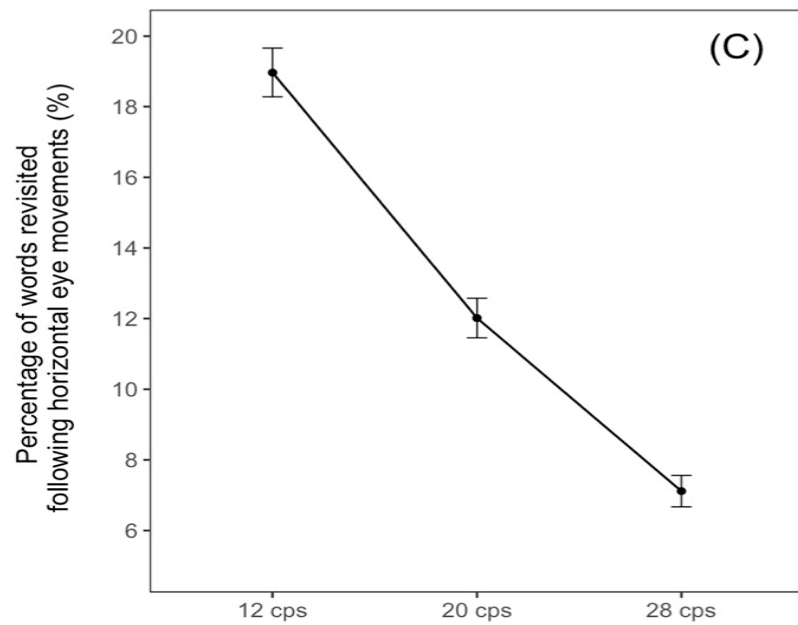
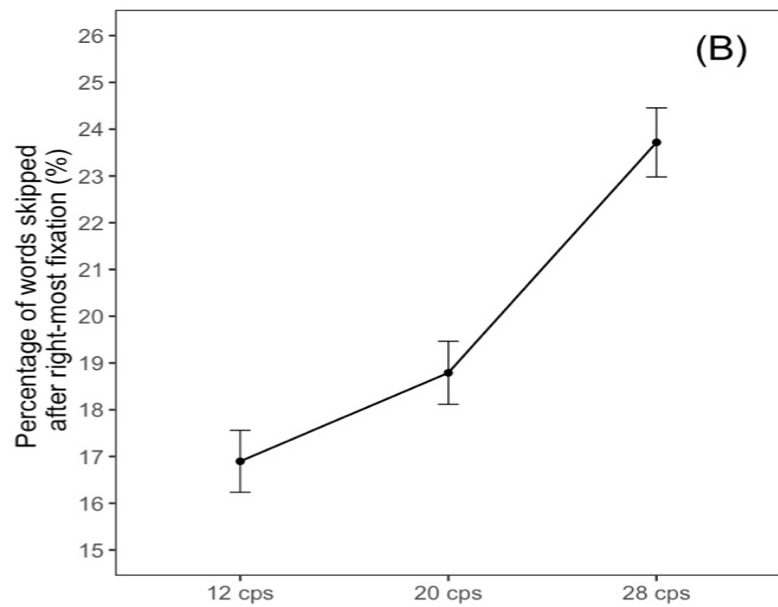
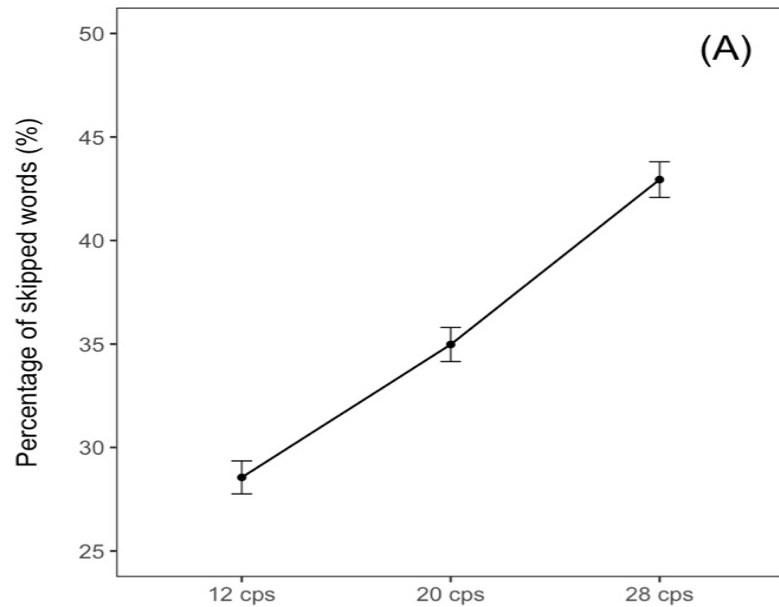
So they lick saliva onto their forearms.

On-line eye tracking measures

- We examined how video and presentation speed influenced two robust and commonly used markers of lexical processing:
 - Word-length effect (shorter words identified more easily and skipped more often due to perceptual span)
 - Word-frequency effect (the forms and meanings of high frequency words are easier to access or retrieve)
- Because eye movements provide important clues about the mental processes during reading and other tasks (for a review, see Rayner, 1998), we also examined a number of standard eye-movement measures (e.g., saccade length and fixation duration).
- But first, comprehension...

Comprehension





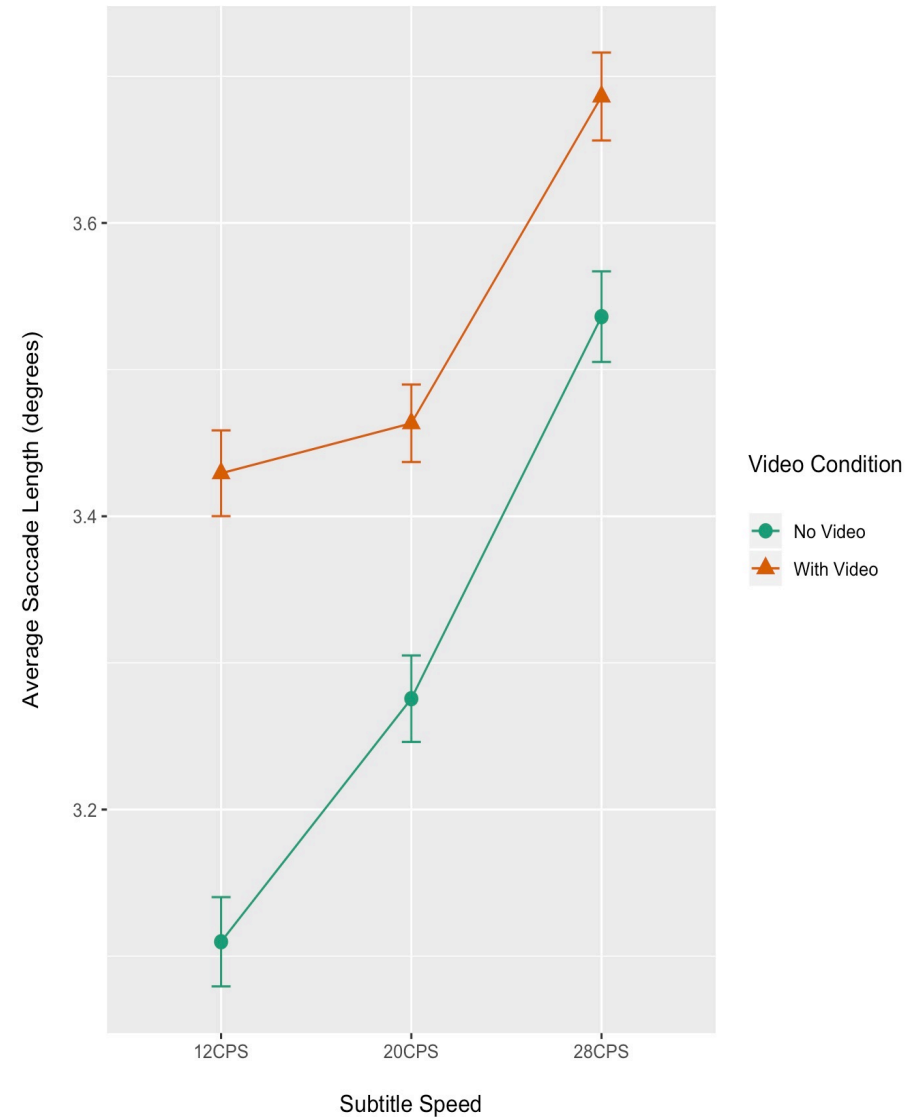
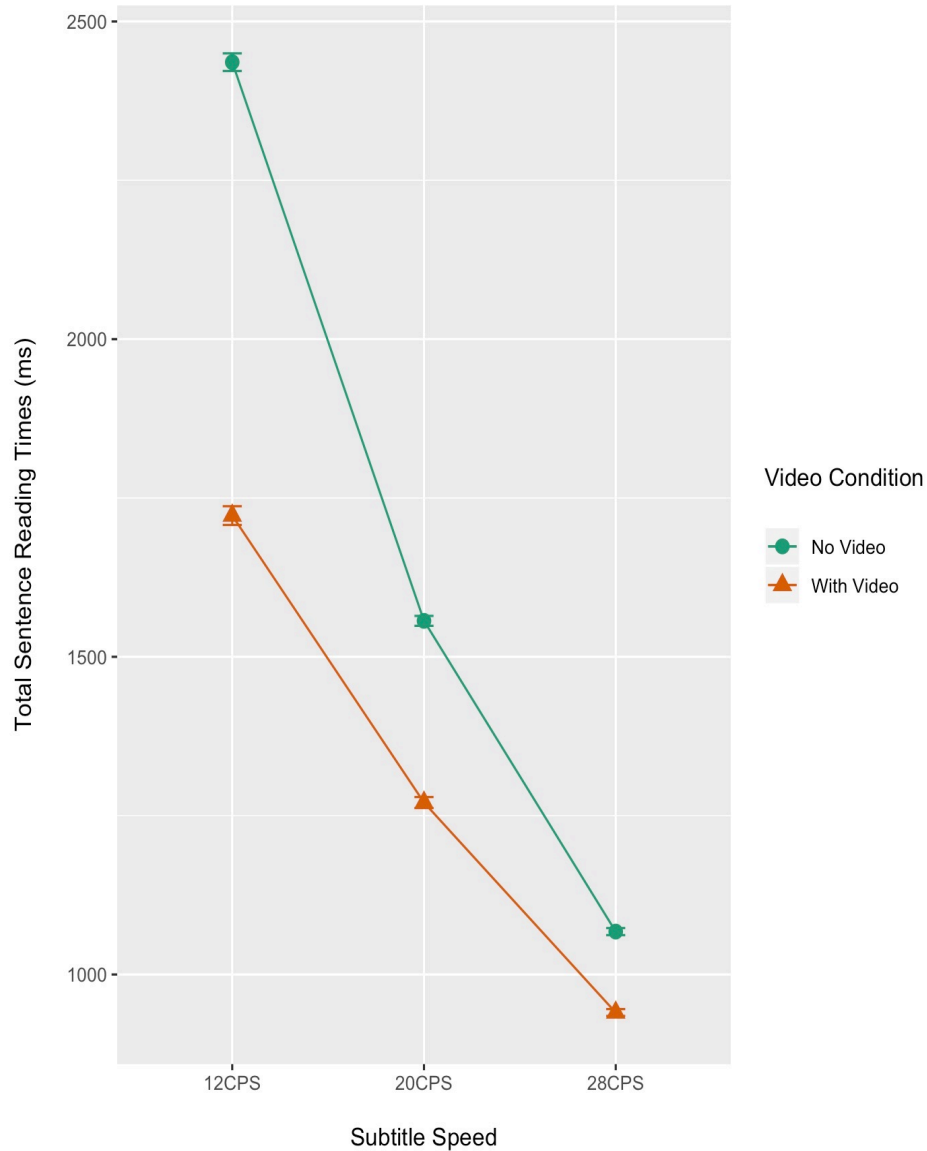
Subtitle Speed

Subtitle Speed

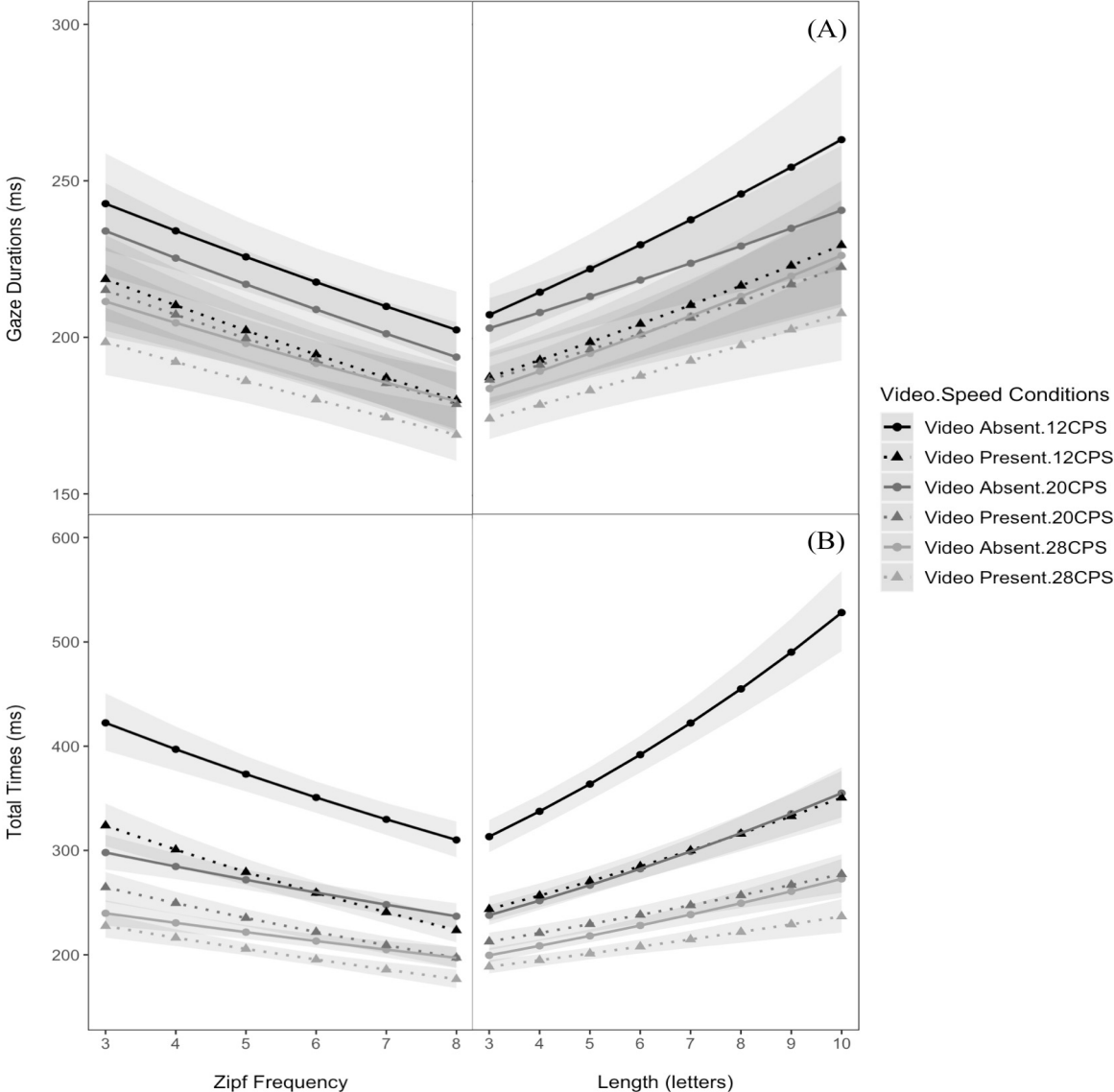
Global subtitle reading

* Total sentence reading times

* Average saccade length

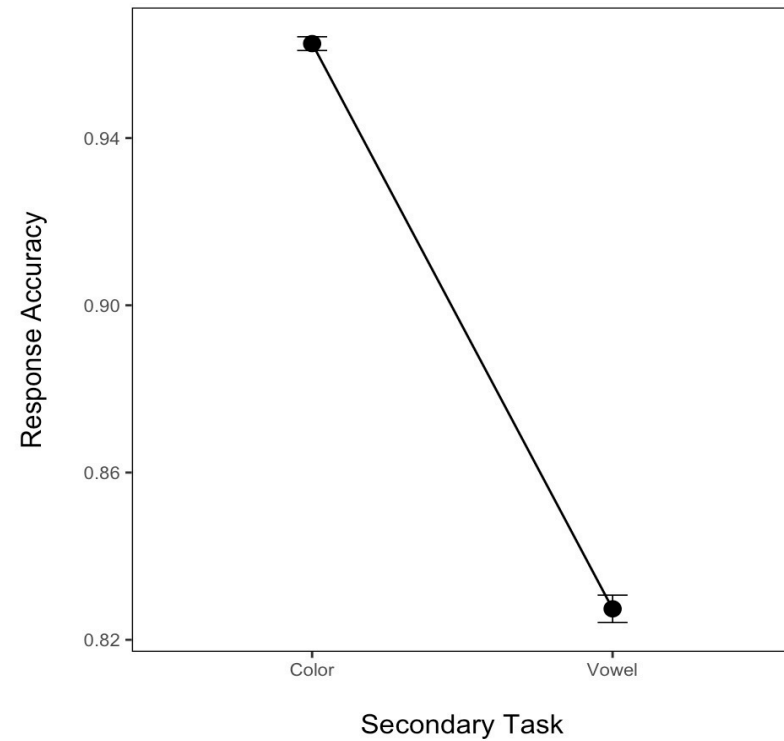
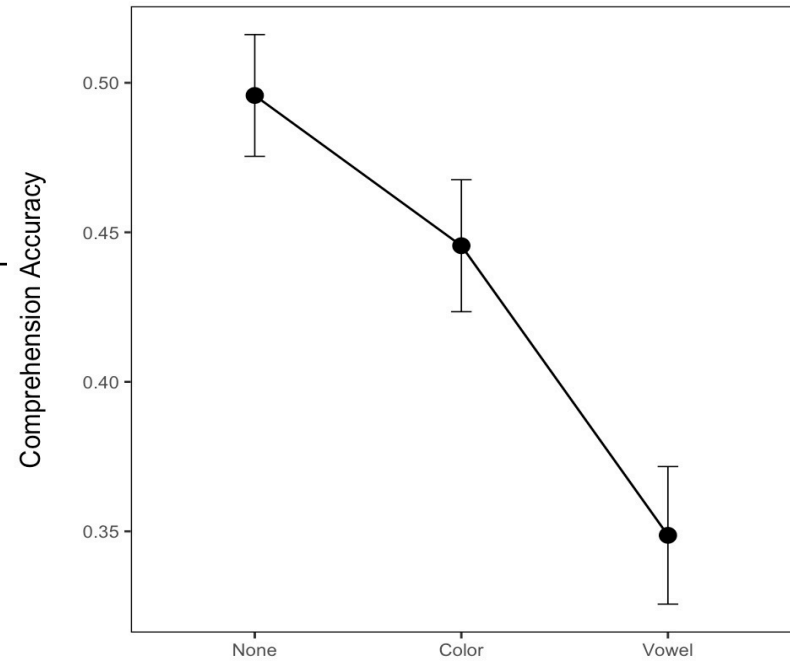
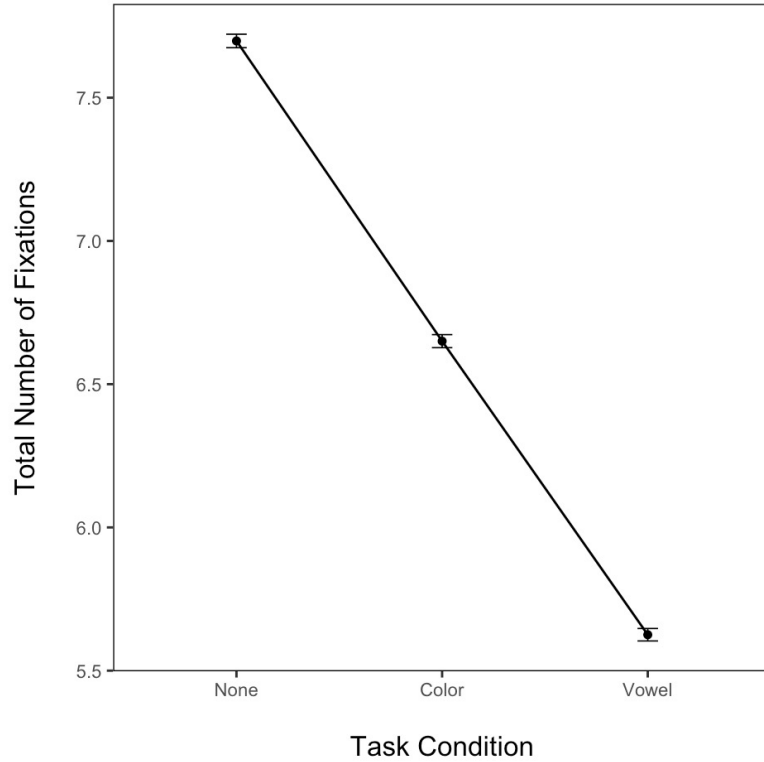


Local word reading: Word frequency and word length



Dual-task experiment

- 59 PP in final analysis



Meaningful processing

- Pedagogical and usability implications:
 - meaningful processing of captions is compromised when the presentation rate increases to levels around 28cps.
 - Faster subtitles cause more superficial processing: fewer, shorter fixations, longer saccades, more skipped words & more subtitles not read to completion.
 - Increased film complexity causes lower processing and comprehension of subtitles, lower response accuracy on secondary task

-
- It seems, therefore, that subtitles are still processed when the demands increase, **but less thoroughly.**
 - It is reasonable to assume that persistently high demands when viewers have to process fast subtitles or complex visuals will have two important consequences:
 1. Reading will take priority over scene perception (unless abandoned)
 2. Reading will become more superficial (fewer, shorter fixations and more skipping)
 3. Comprehension will suffer

In closing

-
- The human mind is highly adaptable, ensuring effective cognitive processing of multimodal texts such as subtitled film.
 - We found evidence of changes in visual behavior as demands increase (video and speed).
 - These changes ensure that attentional resources are managed to prioritise and maximise processing.
 - However, there are indications that these increasing demands impact negatively on performance.
 - In other studies (e.g. Sixin Liao's PhD), we are also looking at the impact of language in subtitles and audio on subtitle processing.
 - More studies are clearly needed to gain a better understanding of how AVT mediates in multimodal texts.
 - Hopefully this will also shape AVT practices in future to ensure that we preserve the richness of audiovisual texts by allowing viewers access to all of the multimodality.

Select references

- Chan, W., Kruger, J.L. & Doherty, S. (2019). Comparing the impact of automatically generated and corrected subtitles on cognitive load and learning in a first- and second-language educational context. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*. 18: 237-272.
- Kruger, J.L. (2018). Eye tracking in audiovisual translation research. In Luis Perez-Gonzalez, ed. *The Routledge Handbook of Audiovisual Translation*. London: Routledge. Chapter 22. pp. 366-382.
- Kruger, J.L., Doherty, S., Fox, W., and de Lissa, P. 2018. Multimodal measurement of cognitive load during subtitle processing: Same-language subtitles for foreign-language viewers. In I. Lacruz and R. Jääskeläinen (eds.), *Innovation and expansion in Translation Process Research*. London, UK: John Benjamins. Chapter 12. pp 267-294. doi:10.1075/ata.18.12kru
- Kruger, J.L., Soto-Sanfiel., M. T., Doherty, S., and Ibrahim, R. 2016. Towards a cognitive audiovisual translatology: Subtitles and embodied cognition. In. Ricardo Muñoz (ed.), *Reembedding Translation Process Research*. pp.171-194. London: John Benjamins Publishing Company.
- Kruger, J.L., Wisniewska, N., Liao, S. Why subtitle speed matters: evidence from word skipping and revisits. *Applied Psycholinguistics*. Manuscript in Press
- Liao, S., Kruger, J.L., Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation*. Issue 33: 70-89 .
- Liao, S., Yu, L., Reichle, E.D., Kruger, J.L. (2020) Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading*. DOI: 10.1080/10888438.2020.1823986.
- Reichle, E.D., Yu, L., Liao, S., Kruger, J.L. Using simulations to understand the reading of rapidly displayed subtitles. *Proceedings of the 43rd Annual Conference of the Cognitive Science Society*. Manuscript in press.
- Szarkowska, A., & Bogucka, L. (2019). Six-second rule revisited: An eye-tracking study on the impact of speech rate and language proficiency on subtitle reading. *Translation, Cognition & Behavior*, 2(1), 101-124.
- Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. *Plos One*, 13(6).

Language processing in multimodal contexts

**Adaptation of keynote delivered at the *Media 4 All* conference
in Stockholm in 2019**

