Reception of Machine-Translated and Human-Translated Subtitles: A Case Study

Frederike Schierl
Tampere University, Finland

frederike.schierl@tuni.fi

Abstract

Accessibility and inclusion have become key terms of the last decades, and this does not exclude linguistics. Machine-translated subtitling has become the new approach to overcome linguistic accessibility barriers since it has proven to be fast and thus cost-efficient for audiovisual media, as opposed to human translation, which is time-intensive and costly. Machine translation can be considered as a solution when a translation is urgently needed. Overall, studies researching benefits of subtitling yield different results, also always depending on the application context (see Chan et al., 2022; Hu et al., 2020). Still, the acceptance of machine-translated subtitles is limited (see Tuominen et al., 2023) and users are rather skeptical, especially regarding the quality of MT subtitles. In the presented project, I investigated the effects of machine-translated subtitling (raw machine translation) compared to human-translated subtitling on the consumer, presenting the results of a case study, knowing that HT as the gold standard for translation is more and more put into question and being aware of today's convincing output of NMT. The presented study investigates the use of (machine-translated) subtitles by the average consumer due to the current strong societal interest. I base my research project on the 3 R concept, i.e. response, reaction, and repercussion (Gambier, 2009), in which participants were asked to watch two video presentations on educational topics, one in German and another in Finnish, subtitled either with machine translation or by a human translator, or in a mixed condition (machine-translated and human-translated). Subtitle languages were English, German, and Finnish. Afterwards, they were asked to respond to questions on the video content (information retrieval) and evaluate the subtitles based on the User Experience Questionnaire (Laugwitz et al., 2008) and NASA Task Load Index (NASA, 2006). The case study shows that information retrieval in the HT conditions is higher, except for the direction Finnish-German. However, users generally report a better user experience for all languages, which indicates a higher immersion. Participants also report that long subtitles combined with a fast pace contribute to more stress and more distraction from the other visual elements. Generally, users recognise the potential of MT subtitles, but also state that a human-in-the-loop is still needed to ensure publishable quality.

1. Introduction

Accessible information is a pivotal element in today's society and audiovisual media have become a crucial channel of information, not only for everyday needs, but also, among others, in an educational context (Gernsbacher, 2015; Negi and Mitra, 2022, see also Zhang, 2005). Subtitles are nowadays one of the tools of choice to foster accessibility to information and knowledge, more and more also with the aid of machine translation (Castilho et al., 2017; Hu et al., 2020). Still, the overarching question remains how the consumer reacts to this rather new mode of subtitling (ST), compared to the 'traditional' human-translated subtitling (HTST). This can be particularly relevant when using machine-translated subtitles (MTST) in an educational context, in which language and topics can be ambiguous and complex; aspects, which can be still a challenge for the current state-of-the-art MT systems (see Bender, 2010; Bywood et al., 2017).

Reception studies have gained more and more attention over the last decades, especially regarding the investigation of the user's need facing audiovisual translation (Tuominen, 2013, 2018). The use of instruments for quantitative measurements, eye-tracking for instance, has also contributed to the growing interest in the topic and gives more insights into reading processes (cf. Orrego-Carmona, 2015; Lång, 2016; Hu et al., 2020). Investigating the reception of subtitles started in the late 80s and early 90s with studies by d'Ydewalle et al., (1987), d'Ydewalle et al. (1991), showing that reading subtitles cannot be avoided although the original audio language is known to the viewer.

With the evermore increasing use of videos as source of learning, not only as a complement but sometimes even a substitute to in-classroom teaching, the interest has grown in knowing how users (the viewers) perceive the material in terms of cognitive treatment, the cognitive load (cf. Paas et al., 2003; Sweller et al., 2011), redundancy effect (Kalyuga, 2012), and attention-split (Ayres and Sweller, 2005). This is especially of interest when adding the additional source of information that subtitling may offer and has been discussed whether it is an actual aid or a nuisance, considering the redundancy effect, which states that the same information given via various channels at the same time hinder the learning effect. Although a study by Diao and Sweller (2007) reports a concurrence between a written presentation and the verbatim spoken presentation in a study with first-year university students, other studies on the topic cannot confirm the findings, therefore subtitles do not pose a hindrance per se (see Perego et al., 2010; Kruger et al., 2014; Gernsbacher, 2015; Liao et al., 2020). For instance, a study investigating reception of subtitles in an educational context (Chan et al., 2022) has shown that watching subtitles in L1 improves comprehension, but subtitling in L2 also creates a higher cognitive load to the audience.

Studies investigating the reception of machine-translated subtitles are still scarce. In a study from 2023, Tuominen et al. (2023) researched users' perception of MT subtitles and concluded that those can be useful, but users are still rather skeptical and would not rely on machine translation alone but prefer human-translated subtitles or at least subtitles checked by a human translator.

2. Research interest and Methodological Framework

The presented study aims to build on previous research on subtitling reception in general and the comparison of MT and HT subtitles in particular, focusing on the perception of the subtitles (the act of paying and sharing attention) but also the cognitive processing (resulting in gaining information/learning from the material).

The research hypotheses are based on the 3 R concept (Gambier 2009): response, reaction, and repercussion. In this concept, response refers to the act of perceiving the subtitles visually (perceptual decoding), the reaction entails the cognitive treatment, how the subtitles are processed cognitively, and repercussion comprises a socio-cultural component, including expectations and attitudes. Therefore, the following hypotheses shall be investigated:

- H1 (for response): Informants in the conditions of MTST will pay more attention to the subtitles than to the image (compared to HTST) since they concentrate more on the written text when the subtitling quality is suspected to be low.
- H2 (for reaction): Informants in the conditions of MTST will report a lack of comprehension, which reflects in lower rating scores compared to HTST. The cognitive load will also be higher.
- H3 (for repercussion): Informants in the conditions of MTST report a less positive attitude towards subtitling than in the conditions of HTST.

3. Experimental Setup

3.1. Video Material

The corpus for the research experiment consisted of two videos recorded by teachers from the university in the native language of the respective speaker, i.e. German and Finnish. Both videos had an equal length of 6 to 7 minutes and were presented in a slide-show lecture format. The presentations contained information on a respective research/teaching topic of the speaker, in this case *professional listening in on-site interpreting* and *regional and culture studies in the German-speaking area.* The videos were recorded for the purpose of the experiment to avoid potential copyright issues that may appear when using prefabricated videos on the Internet. The videos were recorded in the video platform *Panopto*, which also supports generation of intralingual subtitling via automatic speech recognition (ASR). The generated speech transcripts were exported, light post-edited (only spelling errors, wrong punctuation and misrecognitions were corrected) and then machine-translated. Subtitling languages of the videos were English and German for the Finnish video and Finnish for the German video. It was considered to include also English subtitles for the German video, but soon was decided against due to the location of the research and the considered limited benefit for the outcomes.

For the machine-translated condition, the open-source software *SubtitleEdit* was chosen since it also has a built-in machine-translation function which operates with *Google Translate*, and this tool supports a direct export of the translated transcript with time codes.

The human-translated subtitles were done by a professional subtitler with 25 years of experience in both subtitling from scratch and with templates. The subtitler had no templates for this job, but created and translated the subtitles based on the provided videos.

3.2. Participants

36 participants took part in the experiment (27 female, 7 male, 1 other [non-binary], and 1 person did not want to give further information) with an age range from 20-69 (20-29: 16, 30-39: 11; 40-49: 6, 50-59: 2; 60-69: 1). They were recruited via the University's intranet and by contacting a German university through the network of the researcher. Participants were not asked directly about their native language, but the language(s) they are most proficient in ("considered native or near-native speaker"). The answers were as follows: 18x Finnish, 20x English, 16x German, 2x Swedish, 1x Spanish, 2x French, 1x Dutch (multiple answers were possible). Therefore, the general high language proficiency concentrates on Finnish, English, and German. Especially English was very often mentioned in addition to another (native) language. Since the videos were in Finnish and German, language proficiency in those languages was asked in particular, yielding the following distribution as seen in Table 1.

	No knowledge	Beginner	Intermediate	Advanced	Native Speaker
Finnish	13	3	1	0	19
German	6	8	5	3	14

Table 1 Distribution of language competencies in Finnish and German

Furthermore, participants were asked for their media consumption habits, especially how many hours they spent on a daily basis with audiovisual material and whether or how often they would consume subtitles when they watch audiovisual material. Five stated less than an hour, 9 participants 1-2 hours, 10 indicated 2-3 hours, and 12 participants said that they would watch more than 3 hours of audiovisual material per day. Asking for their habits of having subtitles on display while watching AV material, participants rated their consumption habits on a scale from 1 (never on display) to 7 (always on display) as shown in Table 2.

1	2	3	4	5	6	7
1x	4x	4x	7x	7x	5x	8x

Table 2 Frequency of subtitles display in AV consumption habits

Overall, the participant group was used to watch subtitles although it must be noted that this was partly to be expected due to the location of the study where the population is exposed to subtitles in the daily life. However, also the general young age of the participants might pose a factor since consumption habits have changed and there is a slight tendency, also in so-called "dubbing countries", to watch audiovisual material (shows on the Internet for example) in the original language with subtitles.

3.3.   Methods

To investigate the reception of MT and HT subtitles in an educational context, the study comprised, on the one hand, ten (10) questions aiming at gaining insight into free recall of the video content, but this part also contained questions on perception of colours and shapes to gain insight of potential elements of attention of the participant. On the other hand, the surveys asked for information on subtitle use, in particular whether participants were dependent on the subtitles to understand the content or how they perceived the length of the subtitles. Furthermore, this part of the survey contained statements from the NASA Task Load Index (NASA, 2006) and User Experience Questionnaire (Laugwitz et al., 2008). The advantage of these questionnaires is their standardisation and their recognition and validation in the field of user experience research.

3.4.   Procedure

The study was conducted online or also on-site at the University's facilities, depending on the availability of the participant. If participants chose to participate online, they were met in a one-to-one synchronous meeting in which the conducting researcher gave instructions on the procedure of the experiment and how to set it up from distance (how to display the subtitles if the participant was not familiar with the function). The study consisted of watching two videos and completing three surveys; two surveys were connected to the videos (video content, user experience evaluation), the last survey contained questions on attitudes and expectations towards subtitling in general and MT subtitles in particular. The general procedure of the experiment was watching the first video, completing the survey, watching the second video, completing the survey, and then completing the final survey. The participants saw the subtitles in a blinded condition, i.e. they did not know beforehand whether they were human-translated or machine-translated. Overall, the experiment duration was approximately one hour.

In the online participation, a link with access to the video material and to the surveys, which were also accessible online, were sent to them at the start of each meeting, also to avoid that the participant would be tempted to watch the videos beforehand. If participants chose to participate on-site, the experiment was set up for them in the right manner, so that they only had to watch the videos and fill in the surveys.

Before the experiment, in every case, participants were provided with an information sheet of the experiment, a privacy notice and an informed consent form, which they had to fill in and send back to the researcher before the beginning of the experiment. In the on-site condition, participants could fill in the form right before the experiment. Before the beginning of the experiment, the procedure of the experiment was explained again orally, what to do and what not to do during the experiment and participants were invited to ask questions if something remained unclear. After finishing the experiment, participants were furthermore invited to comment on the procedure, for instance how they felt while they were watching the videos and the subtitles.

4. Results

4.1. German – Finnish

36 participants saw the German video with the Finnish subtitles, of which 18 saw the MTST version (MT quality) and 18 the HTST version (HT quality). Out of the 36 participants, 18 stated that they have watched the subtitles all the time, 10 stated having them watched sometimes, and 8 stated not having watched them. Furthermore, on a scale from 'entirely' to 'not at all', (entirely, somewhat, a little, almost not, not at all), 12 participants had to rely entirely on the subtitles to understand the content (6 in the HTST condition and 6 in the MTST condition), 5 stated 'somewhat' (1 in HTST in 4 in MTST), 1 had to rely 'a little' (HTST) and 18 (10 in HTST and 8 in MTST) had not to rely on them at all.

Information Retrieval

In the first part of the survey, the section aiming at information retrieval and free recall from the video, 22 points could be reached in total. Table 3 shows a comparison between the two modes of presentation.

	MTST (Average [MD, SD])	HTST (Average [MD, SD])
Correct Response Rate (CRR)	13.41 [14, 2.9]	14.83 [14.75, 1.78]
In %	59.8	67.9

Table 3 Information Retrieval German video with Finnish subtitles

As the comparison shows, HTST outperforms MTST by 1.4 points on average (8.1% per cent) although it must be noted that the median score between MTST and HTST does differ significantly.

Task Load and User Experience

The initial design of the Task Load as well as the User Experience Questionnaire does not contain a direct numerical evaluation scale, but for unification and limits of the survey tool, it was chosen to modify the presentation into this scaling. Furthermore, the initial design of the Task Load Index comprises a 21-bar scale, divided into three main subscales, therefore it was decided to use a 7-point-scaling. Additionally, the UEQ contains a seven-bullet evaluation scale with two adjacent items (for instance good/bad). Due to the limits of the tools, it was chosen to keep the 7-point scaling, but to exclude the opposition due to non-feasibility. It is of note that participants were asked to complete the survey, the evaluation, to the best of their knowledge. There is a certain data loss due to missing data especially from those participants who were not proficient in Finnish.

NASA Task Load Index

	MTST (Average, [MD, SD])	HTST (Average, [MD, SD])
How mentally demanding was the task?	3.65 [4, 1.84]	2.94 [2, 1.47]
How insecure/stressed/annoyed were you during the task?	2.59 [2, 1.18]	1.67 [1, 0.97]

How much time pressure did you feel due to the rate at which the ST occurred?	3.57 [3.5, 1.79]	1.63 [1, 1.48]
How successful were you in accomplishing the task, i.e. understanding the content?	4.53 [5, 1.66]	4.75 [5.5, 2.05]

Table 4 NASA Task Load Index German video with Finnish subtitles

Task Load Index for the condition German video with Finnish subtitles shows overall better scores for HTST: Mental demand was perceived lower, participants reported being less stressed and insecure and they felt less time pressure. Task success was also perceived higher although only marginally.

User Experience Questionnaire

	MTST (Average, [MD, SD])	HTST (Average, [MD, SD])
enjoyable	3.77 [3, 1.24]	5.73 [6, 1.74]
understandable	4.69 [5, 1.65]	6.17 [7, 1.75]
supportive	4.33 [4.5, 1.37]	5.7 [6.5, 1.95]
bad	2.25 [1.5, 1.60]	1.08 [1, 0.29]
unpleasant	2.46 [2, 1.45]	1 [1, 0]
meet expectations	4.58 [4.5, 0.90]	5.92 [7, 1.83]
confusing	2.86 [2.5, 1.61]	1.67 [1, 1.73]
practical	4.5 [5, 1.45]	6.3 [7, 1.06]
efficient	4.09 [4, 1.44]	6 [6, 0.74]

Table 5 User Experience Questionnaire German video with Finnish subtitles

Participants reported a better user experience according to the user experience questionnaire in the HTST condition (Table 5). Of note are in particular the scores for HTST in 'understandability', 'practicality', and 'efficiency', which are close to perfect score. Also, the very low scores for 'unpleasure' and 'badness' are striking.

4.2.   Finnish – German

Of the 36 participants, 20 were presented with a Finnish video and German subtitles. 10 watched the video with HT subtitles and 10 with MT subtitles. Overall, 10 stated having read the subtitles all the time, 4 read them sometimes and 6 stated not having read them. 10 participants stated that they were entirely dependent on the subtitles (4 in the HTST condition and 6 in the MTST condition) and 10 were not dependent on them at all (6 in the HTST and 4 in the MTST condition)

Information Retrieval

In the recall part of the survey, 21 points could be reached in total. Table 6 presents the results in the different modes of presentation.

	MTST (Average [MD, SD])	HTST (Average [MD, SD])
Correct Response Rate (CRR)	16.8 [17.25, 2.07]	14.9 [14.25, 1.91]
In %	80.00	70.96

Table 6 Information Retrieval Finnish video with German subtitles

Results show that there is a higher information retrieval (or free information recall) in the MTST condition.

NASA Task Load Index

	MTST (Average [MD, SD])	HTST (Average [MD, SD])
How mentally demanding was the task?	2.89 [2, 1.9]	3.9 [4, 1.20]
How insecure/stressed/annoyed were you during the task?	2.78 [2, 1.64]	2.3 [2, 0.95]
How much time pressure did you feel due to the rate at which the ST occurred?	3.5 [3, 1.93]	1.71 [1, 1.11]
How successful were you in accomplishing the task, i.e. understanding the content?	5.22 [6, 1.48]	4.6 [5, 1.58]

Table 7 NASA Task Load Index Finnish video with German subtitles

User Experience Questionnaire

	MTST (Average [MD, SD])	HTST (Average [MD, SD])
enjoyable	3.43 [3, 2.23]	5.67 [6, 1]
understandable	4.43 [4, 2.07]	6.38 [6, 0.52]
supportive	4.67 [4.5, 2.07]	6.25 [6, 0.71]
bad	2.14 [1, 1.57]	1.11 [1, 0.33]
unpleasant	2.57 [3, 1.40]	1.33 [1, 0.5]
meet expectations	4.8 [5, 1.64]	5.5 [5, 1.07]
confusing	2.2 [2, 1.30]	1.44 [1, 0.53]
practical	4.83 [4.5, 1.83]	5.5 [6, 1.41]
efficient	4.14 [5, 2.19]	5.75 [5.5, 0.87]

Table 8 User Experience Questionnaire Finnish video with German subtitles

The results (Table 7 and Table 8) indicate that mental demand is perceived lower in the MTST condition, and also that participants felt more successful in accomplishing the task. However, insecurity and stress as well as time pressure are lower in the HTST condition. Furthermore, the UEQ shows better results in all aspect compared to MTST.

4.3.   Finnish – English

16 of the 36 participants were presented with the Finnish video and English subtitles: 8 with MTST and 8 with HTST. 8 stated having read the subtitles all the time, 4 sometimes and 4 said not having read them. In addition to the reading, 6 stated being entirely dependent on the ST to understand the content (3 in the HTST condition and 3 in the MTST condition), 1 participant indicated 'somewhat' dependent (MTST), 2 indicated 'almost not' (MTST), and 7 stated not being dependent on the subtitling at all (5 in the HTST condition and 2 in the MTST condition).

Information Retrieval

	MTST (Average [MD, SD])	HTST (Average [MD, SD])
Correct Response Rate (CRR)	14.63 [14.5, 1.79]	15.75 [16.25, 2.96]
In %	69.64	75

Table 9 Information Retrieval Finnish video with English subtitles

Regarding information retrieval, results show that HTST outperforms MTST by 1.12 points on average and higher information gain by approx. 5 per cent.

NASA Task Load Index

	MTST (Average [MD, SD])	HTST (Average [MD, SD])
How mentally demanding was the task?	4.25 [5, 1.58]	4.5 [4.5, 1.20]
How insecure/stressed/annoyed were you during the task?	3.25 [3, 1.67]	2.38 [2, 1.19]
How much time pressure did you feel due to the rate at which the ST occurred?	3.86 [4, 1.57]	2.75 [2.5, 2.05]
How successful were you in accomplishing the task, i.e. understanding the content?	4.38 [4, 1.51]	4.88 [5, 1.81]

Table 10 NASA Task Load Index Finnish video with English subtitles

The Task Load Index shows better results throughout HTST, however it is of note that mental demand and perception of task success are slightly higher. Insecurity or stress and time pressure are perceived lower in the HTST condition as well.

User Experience Questionnaire

	MTST (Average [MD, SD])	HTST (Average [MD, SD])
enjoyable	3.86 [4, 1.86]	4 [4, 2]
understandable	5.43 [6, 2.07]	5.43 [6, 1.51]
supportive	4.57 [4, 1.81]	4.83 [5, 2.31]
bad	2 [1, 1.53]	1.5 [1, 1.22]
unpleasant	2.29 [1, 1.89]	2.14 [1, 1.68]
meet expectations	4.43 [5, 1.51]	5.67 [6, 1.51]
confusing	4.14 [5, 1.86]	3.14 [3, 2.34]
practical	4.86 [4, 1.77]	5.29 [5, 1.50]
efficient	4.43 [5, 2.37]	4.71 [4, 1.70]

Table 11 User Experience Questionnaire Finnish video with English subtitles

Results for Task Load and User Experience show that mental demand was slightly lower in the MTST condition, but participants reported better scores for other aspects of the task load. Regarding user experience, HTST outperforms MTST in all items but for 'understandability', in which scores are equal.

4.4. Perception of Quality

Next to the evaluation of the subtitles, participants were asked to evaluate the quality on a scale from 1 – 7 (one being the lowest score, 7 the highest) and to comment on their perception of quality. The results are presented in Table 12.

	GE-FI	FI-GE	FI-EN
Machine Translation	4.3 [4, 0.67]	4 [4, 1.73]	5.29 [5, 1.50]
Human Translation	6 [6, 1.04]	6.17 [6, 0.41]	5.25 [5, 1.04]

Table 12 Evaluation of quality perception; Scale from 1-7

Further, participants were invited to comment on their perception of quality or how they define (high) quality, knowing that 'quality' is a highly subjective topic. Among the most often mentioned aspects were 'correctness' (spelling and grammar), 'shortness' and 'precision' (the message should be conveyed) and good subtitles should be well-timed.

4.5. Attitudes and expectations towards (machine-translated) subtitling

Generally, participants in the experiment stated that they have a positive attitude towards subtitling, stating that they pose an aid and are sometimes a crucial element in understanding the content of the AV material. Also, they recognise the potential of machine-translated subtitles (for example to have a gist of the content), however the informants very often stated that they would not trust the machine translation alone, but it needs a human-in-the-loop to reach good publishable quality.

5. Discussion

The presented study investigated differences in reception of MT and HT subtitles, based on quantitative (numeric evaluation scales) and qualitative (comments and open questions on attitutdes and expectations) research methods, which rely on the subjective percepetion of the participant. Each participant watched a video in German and Finnish and responded to a survey after each video. The ST language of the German video was only Finnish, the subtitle languages of the Finnish video were either German or English, depending on the assigned participant number. The research interest was based on Gambier's (2009) 3 R concept: response, reaction, repercussion. The results yielded a general difference in reception: For the combination German video and Finnish subtitle, HTST outperformed MTST in every measured aspect: information retrieval, task load, and user experience. For the combination Finnish video and German subtitle, results showed a better information retrieval in MTST, which was unexpected. Also, mental demand was perceived lower and task success was perceived higher with MT subtitles, for which a correlation cannot be excluded. There is no obvious explanation for this phenomenon, but an assumption is that Finnish-speaking participants did not read the ST consciously (at least they report for the majority that they did not read them) due to lacking language proficiency and speed of the ST, therefore they were less interested and focused more on the audio. At the same time, German-speaking participants (with no to little Finnish proficiency and therefore depending more on the ST) could gain enough information from the ST. This explanation, however, needs further investigation. User experience was again always perceived better in the HTST condition and clearly outperforms user experience with MTST. The experimental condition Finnish video and English subtitles yielded again mixed results: Participants reported a better information in the HTST condition and also in perception of task load. However, it is of note that mental demand and success rate were only perceived slightly higher in the HTST condition. Furthermore, investigating user experience, the aspect of 'understandability' reached equal scores in both conditions, and HTST is also perceived slightly better in the items of 'enjoyability', 'support', 'pleasure', and 'efficiency'.

Taking these results into account, there is no evidence in this study with the presented methods that participants will spend more time looking at the subtitles when they are machine-translated and will be therefore more distracted (H1). Furthermore, cognitive load was perceived higher in the conditions GE-FI and FI-EN, but not in FI-GE, therefore H2 could not be completely confirmed either. Of note is also that participants commented (either in written or oral form) that they felt stressed by the speed of the ST in the MT condition, especially when being dependent on the ST; therefore, they reported not being able to spend much time on other aspects of the video. User experience is, for one exception, always perceived better in the HT condition, which resulted in higher rating scores.

Participants were also asked to evaluate the quality, again based on personal perception. The results show a better score for the HT subtitles, except for MT, which scores marginally higher, which is therefore not significant. Lastly, it was assumed that participants watching MTST will report less satisfaction and lower expectations and attitudes (H3). This hypothesis could not be confirmed either since participants generally expressed a recognition for the potential of MT subtitles, with the condition that they still need improvement, that there should be a human-in-the-loop, and that the output of the MT, hence also the applicability, depends on the context. Some comments remarked, for example, that MT is unsuitable for a context with complex educational information.

6. Conclusion

The presented study aimed to contribute to the further research of user reception of HT and MT subtitles. The results of the study contributed to the potential of this mode of translation, especially in the context of accessibility at an efficient level (cost and time-efficient for an acceptable result), focussing among others on the aspect of pure information gain on the one hand and the overall perception (the whole entertainment) on the other. I would like to argue that the research results also contributed to a topic that might pose an easy question ('Are HT subtitles better than MT subtitles?') to which the answer is more complex than initially thought. The study has shown that MT subtitles can make AV content more accessible. Still there is a long way to go to reach an MT output that is qualitatively convincing and better applicable. Furthermore, there is still the average consumer who must be convinced but the study has also shown that expectations are not bad as long as the consumer is also aware that the ST was produced (mainly) by a machine. One may remember when MT entered the translation market in the early 2000s and everybody was rather skeptical about it and nowadays nobody can imagine a life without the aid of MT. After all, despite good results in information retrieval, the study does not answer the question of whether consumers would prefer the information gain alone or whether it must be seen as a whole entity in which user experience is part of the viewing process. The study had several limitations. Firstly, it must be noted that cognition and memory are highly individual, and some participants were maybe able to retain more information than others. The study had no baseline, i.e. they did not watch a video without subtitles (there was no control group in general) or it was not measured with a pre-test how well their working memory was. This was not considered in the initial planning of the experiment and would have been difficult to perform in the online setup of the experiment. The experiment also had no test round in which the participants could get familiar with the experimental setup, the video and the subtitling as well as potential questions. Since the overall duration of the main experiment was approximately one hour, a test round would have contributed to an extension in time and potentially less motivation and/or earlier fatigue, which could have had an influence of the results. Secondly, I am aware that the notion of quality is highly individualised and that there is no general definition. It was therefore important that participants could express their own understanding of quality next to the quantitative perception. It also must be taken into account that all participants had a background of higher education, which might have influenced their perception of quality as well. Thirdly, the experimental setup, which included always a combination of the Finnish video with the German one, limited data to this point: 36 participants watched the German video (18 HT/ 18 MT), but participants had to be split further due to the nature of two sets of subtitles for the Finnish video (German: 10 HT/10 MT; English 8 HT/8 MT), which leads to a stronger reliability for the German presentation with Finnish subtitles.

It is planned as further research to investigate more the perceptual decoding of the subtitles, which aims at a more detailed look into the first research hypothesis, particularly where the participants look when watching a video with subtitles. A further study will include eye-tracking, which intends also to have a closer look at results from the presented study.

References

Ayres, Paul, and John Sweller. 2005. 'The Split-Attention Principle in Multimedia Learning'. In *The Cambridge Handbook of Multimedia Learning*, edited by Richard Mayer, 1st ed., 135–46. Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.009.

Bender, Oliver. 2010. 'Robust Machine Translation for Multi-Domain Tasks'. PhD Thesis, Aachen, Techn. Hochsch., Diss., 2010.

Bywood, Lindsay, Panayota Georgakopoulou, and Thierry Etchegoyhen. 2017. 'Embracing the Threat: Machine Translation as a Solution for Subtitling'. *Perspectives* 25 (3): 492–508. https://doi.org/10.1080/0907676X.2017.1291695.

Chan, Win Shan, Jan-Louis Kruger, and Stephen Doherty. 2022. 'An Investigation of Subtitles as Learning Support in University Education'. *Journal of Specialised Translation*, no. 38: 155–79.

Diao, Yali, and John Sweller. 2007. 'Redundancy in Foreign Language Reading Comprehension Instruction: Concurrent Written and Spoken Presentations'. *Learning and Instruction* 17 (1): 78–88. https://doi.org/10.1016/j.learninstruc.2006.11.007.

D'Ydewalle, Géry, Caroline Praet, Karl Verfaillie, and Johan Van Rensbergen. 1991. 'Watching Subtitled Television: Automatic Reading Behavior'. *Communication Research* 18 (5): 650–66. https://doi.org/10.1177/009365091018005005.

D'Ydewalle, Géry, Johan Van Rensbergen, and Joris Pollet. 1987. 'Reading a Message When the Same Message Is Available Auditorily in Another Language: The Case of Subtitling'. In *Eye Movements from Physiology to Cognition*, 313–21. Elsevier. https://doi.org/10.1016/B978-0-444-70113-8.50047-3.

Gambier, Yves. 2009. 'Challenges in Research on Audiovisual Translation'. In *Translation Research Projects 2*, edited by Pym, Anthony and Alexander Perekrestenko, 17–25. Tarragona: Intercultural Studies Group.

Gernsbacher, Morton Ann. 2015. 'Video Captions Benefit Everyone'. *Policy Insights from the Behavioral and Brain Sciences* 2 (1): 195–202. https://doi.org/10.1177/2372732215602130.

Hu, Ke, Sharon O'Brien, and Dorothy Kenny. 2020. 'A Reception Study of Machine Translated Subtitles for MOOCs'. *Perspectives* 28 (4): 521–38. https://doi.org/10.1080/0907676X.2019.1595069.

Kalyuga, Slava. 2012. 'Instructional Benefits of Spoken Words: A Review of Cognitive Load Factors'. *Educational Research Review* 7 (2): 145–59.

Kruger, Jan-Louis, Esté Hefer, and Gordon Matthew. 2014. 'Attention Distribution and Cognitive Load in a Subtitled Academic Lecture: L1 vs. L2'. *Journal of Eye Movement Research* 7 (5). https://doi.org/10.16910/jemr.7.5.4.

Lång, Juha. 2016. 'Subtitles vs. Narration: The Acquisition of Information from Visual-Verbal and Audio-Verbal Channels When Watching a Television Documentary'. In *Eye-Tracking and Applied Linguistics*, 59–81. Berlin: Language Science Press. http://langsci-press.org/catalog/book/108.

Laugwitz, Bettina, Theo Held, and Martin Schrepp. 2008. 'Construction and Evaluation of a User Experience Questionnaire'. In *Symposium of the Austrian HCI and Usability Engineering Group*, edited by Andreas Holzinger, 63–76. Springer.

Liao, Sixin, Jan-Louis Kruger, and Stephen Doherty. 2020. 'The Impact of Monolingual and Bilingual Subtitles on Visual Attention, Cognitive Load, and Comprehension'. *The Journal of Specialised Translation*, no. 33: 70–98.

NASA. 2006. 'NASA TLX: Task Load Index'.

Negi, Shivsevak, and Ritayan Mitra. 2022. 'Native Language Subtitling of Educational Videos: A Multimodal Analysis with Eye Tracking, EEG and Self-reports'. *British Journal of Educational Technology* 53 (6): 1793–1816. https://doi.org/10.1111/bjet.13214.

Orrego-Carmona, David. 2015. 'The Reception of (Non) Professional Subtitling.' PhD Thesis, Universitat Rovira i Virgili.

Paas, Fred, Juhani E. Tuovinen, Huib 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. https://doi.org/10.1207/S15326985EP3801_8.

Perego, Elisa, Fabio Del Missier, Marco Porta, and Mauro Mosconi. 2010. 'The Cognitive Effectiveness of Subtitle Processing'. *Media Psychology* 13 (3): 243–72. https://doi.org/10.1080/15213269.2010.502873.

Sweller, John, Paul Ayres, and Slava Kalyuga. 2011. *Cognitive Load Theory*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-8126-4.

Tuominen, Tiina. 2013. *The Art of Accidental Reading and Incidental Listening. An Empirical Study on the Viewing of Subtitled Films*. Tampere University Press.

Tuominen, Tiina. 2018. 'Multi-Method Research: Reception in Context'. In *Benjamins Translation Library*, edited by Elena Di Giovanni and Yves Gambier, 141:69–90. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/btl.141.05tuo.

Tuominen, Tiina, Maarit Koponen, Kaisa Vitikainen, Umut Sulubacak, and Jörg Tiedemann. 2023. 'Exploring the Gaps in Linguistic Accessibility of Media: The Potential of Automated Subtitling as a Solution'. *Journal of Specialised Translation*, no. 39: 77–89.

Zhang, Dongsong. 2005. 'Interactive Multimedia-Based E-Learning: A Study of Effectiveness'. *American Journal of Distance Education* 19 (3): 149–62. https://doi.org/10.1207/s15389286ajde1903_3.