Journal of Audiovisual Translation

Can Viewers Identify Raw Machine Translation in Subtitles, and What It Means for Reception?

Juerong Qiu ✉
University of Melbourne

Citation: Qiu, J. Can Viewers Identify Raw Machine Translation in Subtitles, and What It Means for Reception? Journal of Audiovisual Translation, 8(1), 1–21. https://doi.org/10.47476/jat.v8i1.2025.316

Editor(s): J. Pedersen

Received: April 26, 2024

Published: April 11, 2025

Abstract

This research compares the audience's reception of fansubs and raw machine-translated subtitles, aiming to investigate the possibility of integrating raw machine translation into subtitling Chinese period dramas into English. The data-collecting instruments include think-aloud protocols, screen recordings, and interviews. Eighteen native English-speaking participants were invited to explore their screen-based activities triggered by each mode of subtitles, their ability to differentiate between the two subtitle modes and their emotional involvement. The results show that raw machine-translated subtitles did not trigger more screen-based activities than fansubs did during the viewing process because the viewer's interaction with the video was not limited to subtitle-related issues. No participants could differentiate raw machine-translated subtitles from fansubs, but seven participants identified the presence of machine translation. Consequently, the seven participants had lower degrees of self-reported involvement compared to those who did not notice the presence of machine translation. Still, the awareness of machine translation did not significantly affect the participant's willingness to continue watching.

Key words: subtitling, machine translation, audience reception, fansubs, screen-based activities, involvement.

✉ Juerong.qiu@unimelb.edu.au, https://orcid.org/0000-0002-1011-7988

Introduction

The development of neural machine translation (NMT) and its improvement in accuracy have led to increasing implementation in diverse industries, including audiovisual translation (AVT). As public broadcasters experience a growing demand for multilingual content to support linguistic accessibility, automated subtitles can be a solution to provide real-time subtitle services and reduce workload (Tuominen et al., 2023). Further, AI-driven subtitling tools encompass automatic speech recognition (ASR), shot change detection, auto-spotting, audio-informed segmentation, and NMT. These tools not only offer individual users access to audiovisual content with multilingual subtitles, democratising the act of translation, but also allow viewers to engage with video content across linguistic boundaries without external assistance, fostering a more inclusive and interconnected global community. These technological developments are unfolding a brand-new landscape, while their practicality and reception remain largely unexplored from the end-user's perspective.

This study investigates the use of raw machine-translated subtitles from the viewer's perspective. Using the English translation of the Chinese period drama *Nirvana in Fire* (2015) as a case study, I compare the reception effects of raw machine-translated subtitles and fansubs on Viki (an Asian-based video-on-demand service provider) on native-English-speaking viewers. The two subtitle modes are integrated into one video with a complete storyline to examine the contextual use of MT. Participants can engage in screen-based activities throughout the viewing process, including pausing, going back, and fast-forwarding. This design not only tries to mimic the actual viewing experience on streaming platforms, but also gives participants a second chance to scrutinise specific subtitle lines or scenes, potentially increasing the likelihood of identifying subtitle-related issues.

This study aims to explore: (1) whether viewers can discern the presence of machine-translated subtitles and identify the part subtitled by MT before being notified, (2) whether the two subtitle modes trigger different frequencies of screen-based activities during viewing, and (3) whether the identification of MT has negative effects on audience involvement. This study also investigates the elements in subtitles and the viewer's perceptions leading to their identification of subtitle modes. Further, I explore whether the (in)correct identification affects their receptive behaviours and whether this identification potentially causes the abandonment of the show.

1. Chinese Shows "Going Global" and Their Receptions Overseas

Chinese drama exports have surged recently, fuelled by the expanding broadcasting coverage and growing demand in the film and drama market (Li, 2022). The Report on the International Distribution of Chinese TV Dramas reveals the total export value of mainland Chinese dramas in 2021 was USD 56.83 million (as cited in Ji, 2022). According to the Report (Ji, 2022), from 2015 to 2021, 52 of the top 70 Chinese dramas in annual international distribution revenue were period dramas. This genre has received official endorsement from the Chinese government for overseas promotion, aiming to boost global audiences' interest and a deeper understanding of Chinese history and culture (Li, 2022).

Asia is the largest market for Chinese dramas, largely due to the cultural proximity among Asian countries. However, Chinese dramas are still in their infancy in English-speaking markets. One of the main reasons is the lack of quality subtitles (Chen & Chang, 2024). YouTube and Viki are the two major online streaming platforms for Chinese mainland dramas to reach a global audience (Chen & Chang, 2024). Major Chinese streaming platforms, such as Youku, IQiyi, and Tencent, have their own YouTube channels. Many Chinese shows on these channels either do not provide English subtitles or rely solely on YouTube's auto-captions. Thus, it is relevant to explore the effect of these English auto-captions on audience reception.

Viki offers multilingual subtitles produced by dedicated fansubbing teams (Li, 2022). Unlike many fansubbing sites, Viki operates through legitimate content licensing and generates revenue from advertising and subscriptions (Zhang, 2022). Moreover, the turnaround time on Viki is significantly shorter than that on YouTube. A Viki subtitle team can finish the English subtitles for an episode within 24–36 hours after its broadcast in China because Viki's subtitling platform allows four translators to translate simultaneously (Zhang, 2022). In contrast, based on the author's observation, YouTube often takes around a week to update an episode.

Figure 1

Screenshots taken on Viki (left) and YouTube (right)

In this study, Viki subtitles are chosen for three main reasons. First, the show is available for free on both Viki and YouTube worldwide. It has received a rating of 9.6/10 from 9,800 votes on Viki and 1,000 comments on the first episode on YouTube (last accessed on July 10, 2024). Second, YouTube subtitles appear beneath the original Chinese, leading to some overlap that may distract viewers, while Viki's subtitles are designed to be more viewer-friendly (see Figure 1). Lastly, character identifications and location details are often untranslated in YouTube subtitles. Thus, Viki subtitles are selected for their popularity, availability, and viewer-friendliness.

How Chinese shows are received by global audiences on video streaming platforms has also gathered scholarly attention. Corpus-driven approaches analyse comments, posts on fan forums, and reviews to investigate the role of interlingual subtitles in the receptive process (e.g., Chen & Chang, 2024; Li,

2022; Wu & Chen, 2022). The three studies agree that the accuracy of English subtitles is not the viewers' primary concern. The availability of English subtitles is far more important than their linguistic accuracy or their presentation, such as the number of lines. Moreover, factors like storyline, acting, characters, cinematography, and special effects shape viewers' overall receptive experience. Wu and Chen (2022) highlight that corpus analysis often falls short of explaining the reasons for these preferences. This limitation suggests a need for empirical methods, such as surveys and interviews, to gain deeper insights into audiences' receptive behaviours and experiences. I thus use different data-gathering methods to explore diverse aspects of reception and reveal the nuanced ways individual viewers engage with the stimulus.

2. Reception Studies at the Crossroads

2.1. Convergence Between Media Studies and Translation Studies

The initial phase of reception studies in media research, marked by Stuart Hall's encoding/decoding model (1973), established a strong empirical foundation. The research focus has shifted from audience interpretations of a media text to a broader understanding of media culture, such as how audiences perceive themselves in the mediascape and how media shape culture, ideology, and society (Alasuutari, 1999).

In recent years, the empowerment of the audience, fuelled by participatory culture and advancements in digital technologies, has introduced numerous concepts and models to media studies. These include but are not limited to transportation (Green & Brock, 2000), identification (Cohen, 2001), and immersion (Nilsson et al., 2016). However, these concepts overlap to some extent. For instance, both transportation (Green & Sestir, 2017) and immersion (Murray, 1997) describe deep absorption in a story, creating the sensation of being immersed in an alternate reality. Furthermore, the interrelationships between these concepts remain unclear, requiring the integration and application of conceptual and empirical advancements across different theoretical frameworks and academic disciplines (Frey, 2018). Another issue in audience research is the neglect of a translation's effect on reception. As transregional and global audiences participate in the reception process, we need to develop methods to understand audience reception more comprehensively (Hill, 2018).

Unlike media reception studies, theoretical models in translation studies focus on the role of translation in the reception process, such as the 3Rs (response, reaction, and repercussion) model first proposed by Chesterman (2007) and introduced to audiovisual translation by Gambier (2018). However, the 3Rs model may fall short of providing a solid understanding of why audiences react and respond to audiovisual products in a particular manner, especially when multiple channels of information are at play. Therefore, Tuominen (2019) argues that translation studies should incorporate insights from media studies to better examine individual differences and the dynamic

ways audiences interact with audiovisual content. In this study, I incorporate concepts of from audience research in media studies to gain a contextualised understanding of audience involvement.

2.2. Audience Reception of Fansubs and Automated Subtitles

Interlingual subtitling involves not only translating messages from one language to another but also converting spoken language into written text (Pedersen, 2011). Fansubbing is a mode of interlingual subtitling, created by fans and for fans (Díaz Cintas & Muñoz Sánchez, 2006). Some research has delved into the analysis of various practices within fansubbing communities, such as translator's notes, and additional texts to explain culture-specific elements in the audiovisual text (Díaz Cintas & Muñoz Sánchez, 2006). The reception of the translator's notes is often assessed through questionnaires and eye-tracking (e.g., Caffrey, 2012; Künzli & Ehrensberger-Dow, 2011). Further, Orrego-Carmona (2019) assesses fansubs from a functional quality perspective, which consists of the production conditions and the reception conditions. He finds that audiences enjoy the variety of forms and styles that fansubs offer. Audiences' preference for fansubs is not necessarily subject to linguistic accuracy or technical aspects. Instead, the deciding factors are the turnaround time and the availability of translated subtitles.

The reception of automated subtitling has gained more scholarly attention than that of fansubbing. Hu, O'Brien, and Kenny (2020) examined how machine-translated subtitles for MOOCs are received by Chinese undergraduates. Through eye-tracking and questionnaires, they tested raw MT subtitles, human-translated (HT) subtitles, and post-edited machine-translated (PEMT) subtitles. The results indicate that participants provided with full PEMT subtitles scored higher on the comprehension test compared to those provided with raw MT subtitles, although the difference was not statistically significant. Conversely, participants provided with HT subtitles did not score higher than the other two groups on the comprehension test and self-reported evaluation. Most participants maintained a positive attitude towards the subtitles despite the subtitle mode shown. Then, Calvo-Ferrer (2023) explored viewers' ability to distinguish between ChatGPT-generated subtitles and human-created subtitles extracted from *The Office* (Gervais & Merchant, 2001) through questionnaires. The participants were 66 first-year students majoring in Translation and Interpreting and 53 final-year students. Results show: (1) Despite participants showing a tendency to associate lower subtitle quality with ChatGPT-generated translations, differentiating between the two remained difficult for them. (2) Experience with ChatGPT did not make it easier for participants to distinguish between the two subtitle modes. (3) Final-year participants demonstrated a greater likelihood of correctly identifying ChatGPT-generated subtitles compared to their first-year counterparts, indicating the significance of translation training in recognising ChatGPT-generated subtitles.

Tuominen et al. (2023) explore the potential of automated subtitling to support linguistic accessibility for public broadcasts through focus groups and questionnaires. They incorporated an ASR system and NMT output to examine English viewers' comprehension, acceptability, and self-reported cognitive load when watching clips of Finnish-language news. They found that participants

understood the gist of a program with automated subtitles, but out-of-synchrony subtitles and MT errors caused disruptions, resulting in significant cognitive load. The finding also shows that participants' interest and motivation influenced their acceptance of automated subtitles, indicating greater tolerance for lower quality when genuine interest or need to access the material was high.

3. Research Design

3.1. Research Question and Hypotheses

The research questions concern the possible differences that raw machine-translated subtitles and fansubs bring to audience reception during viewing and after viewing:

1. Do viewers notice the presence of two modes of subtitles in the video?
2. Do machine-translated subtitles trigger more screen-based activities than fansubs during viewing?
3. If viewers identify the presence of raw machine translation, does it lead to a decrease in their involvement compared to those who are unaware of machine translation?

The initial hypotheses are as follows:

1. Participants can distinguish between raw machine-translated subtitles and fansubs.
2. Machine-translated subtitles trigger more screen-based activities than fansubs do.
3. Awareness of machine-translated subtitles correlates negatively with audience involvement.

Hypothesis 1 posits that machine-translated subtitles can be distinguished from fansubs due to the presence of translator's notes in fansubs and the higher frequency of linguistic errors in machine-translated subtitles (see 4.2 below for details). Participants with prior knowledge of fansubbing conventions and MT may be more likely to make a correct identification. These elements will be considered in hypothesis testing. Hypothesis 2 is based on research findings that errors and awkward features in subtitles can distract viewers' attention and shift their focus to the error (Hu et al., 2020; Tuominen et al., 2023). Pauses and revisits are seen as indicators of disruption. It is assumed that participants are so distracted by errors that they focus on individual subtitle lines instead of the overall viewing experience. Hypothesis 3 concerns the relationship between audience involvement and perceptions of machine-translated subtitles. Unlike previous studies measuring audience enjoyment (e.g., Orrego-Carmona, 2015), this research focuses on self-reported involvement and the willingness to watch more episodes. Since the participants' involvement in the video was measured through two variables, Hypothesis 3 is divided into two sub-hypotheses: (1) the participants who noticed the existence of machine-translated subtitles initially rate their levels of involvement lower than those who were not aware of it, and (2) the participants who noticed the existence of machine-translated subtitles initially are less likely to express willingness to continue watching than those who

were not aware of it. This assumption stems from participants associating identified errors with machine-translated subtitles, prompting doubts about translation accuracy, as observed in Calvo-Ferrer (2023). These doubts may, in turn, increase attention to subtitles, reducing overall involvement.

3.2. Research Material

The period drama, *Nirvana in Fire* (Lang Ya Bang, 琅琊榜) (Hou, 2015), tells the story of Lin Shu, who returns to the capital of Liang to seek revenge for a twelve-year-old false treason charge against his family. The show contains 54 episodes, each lasting around 45 minutes.

The research material is a 9-minute-and-19-second video extracted from Episode One. This segment introduces the drama's background, includes most of the main characters' first appearances, and ends with a flashback, providing a relatively complete storyline. The video allows participants to grasp the plot, promoting more comprehensive and rational feedback on their involvement. The approximately ten-minute viewing time can also give participants more opportunities to initiate screen-based activities and verbalise their thoughts if needed.

The video was divided into two parts, each featuring a different subtitle mode. The division occurs at the video's halfway point, during a shot change (04:28). Both subtitle modes use the same font, size, and colour (as shown in the Viki subtitle on the left of Figure 1), created via the Aegisub subtitling tool.

The machine-translated subtitles, generated by DeepL Translator, were chosen for their translation accuracy, as reported by Hu and Li's research (2023). The machine-translated version contains 33 linguistic errors, covering improper lexicon, incorrect pronouns, wrong tenses, ambiguity, and a lack of cohesive devices (see Appendix 1). The fansubbed version (see Appendix 2) includes one typographical error, two generalisation issues, and one improper tense, but they are not severe enough to alter the meaning of the subtitle lines or cause possible misinterpretation.

Table 1 shows that, on average, the length of fansubs is greater than that of MT subtitles. Additionally, MT subtitles are generally faster than fansubs in terms of characters per second (cps), and the variation in speed is higher for MT subtitles compared to fansubs.

There are nine translator's notes in fansubs, the first seven in the first half of the video and the remaining two in the second half. These notes offer explanations for extralinguistic cultural references (Pedersen, 2007), character identifications, and location details (see Table 2). These notes appear one line below the subtitles on Viki, maintaining the same font and size but with italicisation. The machine-translated version does not have these notes.

Table 1

Information on the Nature of the Two Subtitle Modes

	Machine-translated subtitles		Fansubs	
	First half	Second half	First half	Second half
Number of subtitles	67	68	71	66
Subtitle length	3–58	2–51	3–54	3–48
(character per line)	M=24.3	M=21.8	M=31.7	M=27.7
Subtitle speed	3–38 cps	5–36 cps	3–29 cps	2–20 cps
	M=16.4 cps	*M*=16.2 cps	*M*=10.6 cps	*M*=11.2 cps
Number of errors	18	15	2	2

Table 2

Nine Translator's Notes in the Fansubbed Version

No.	Time	Source text	Fansubs
1	00:03	十文	10 wen. *T/N: wen-ancient unit of currency*
		那 这个呢	Oh.
2	00:05	言豫津 国舅府公子	*Yan Yujin – Son of the Emperor's brother-in-law*
3	00:11	萧景睿 宁国侯府世子	*Xiao Jingrui – Eldest sone of the Marqius[1] of Ning*
4	00:18	什么时候才能回到金陵啊	when will we be able to return to Jinling? T/N: Jinling is the capital city of Da-Liang
5	00:48	景瑞 那是大渝王族的战旗吧	Jingrui, isn't that the battle flag of Da-Yu's Imperial family? *T/N: Da- honorific, lit. the Great*
6	03:21	你苏哥哥就要丢下你去金陵了	Your Su-gege is about to abandon you and go to Jinling. T/N: "-gege" – endearment for older males.

[1] This typographical error originally exists in the fansubs.

7	03:38	大梁都城金陵	The capital of Da-Liang, Jinling City
8	04:32	霓凰姐姐	Nihuang-jiejie. T/N: "-jiejie" endearment for older females
9	04:48	穆霓凰　云南穆府郡主	*Mu Nihuang – Princess of the Imperial Mu Household of Yunnan.*

3.3. Participants

Eighteen students from the University of Melbourne, aged between 18 and 30, took part in the experiment. They had no prior knowledge of Mandarin and no background in translation or interpreting. The participants were recruited through convenience and snowball sampling. Four participants reported that they had not watched any Chinese audiovisual products before, A05 and B06 claimed that they had watched Chinese audiovisual products over 2 hours per week, and the remaining participants had all watched less than one hour per week. A05 mentioned having watched *Nirvana in Fire* before.

The eighteen participants were randomly assigned to two groups. Group A watched the video with fansubs in the first half and machine-translated subtitles in the second half. Group B watched in the reverse sequence. Participants were not informed of the existence of two subtitle modes during viewing. These arrangements were designed to explore whether the two modes of subtitles trigger different reactions during viewing and whether the participants would notice the shift of subtitle modes.

3.4. Research Methods

The experiment was conducted with each participant on Zoom and was audio-recorded. This research has been approved by the ethics committee at the University of Melbourne (ID: 2057044).

3.4.1. Screen Recordings

Screen recordings capture all changes on the computer screen, documenting participants' interactions with the subtitles (pausing, fast-forwarding, or rewinding) and other viewing behaviours. This method enables a close examination of the viewing process from the participant's perspective. Participants initiated screen-sharing when ready to watch the provided video and could have screen-based activities during viewing. To ease potential anxiety, a demonstration with a practice session was conducted, addressing Hansen's (2013) concerns regarding experiment awareness. During video watching, interruptions were minimised to facilitate full engagement.

3.4.2. Think-Aloud Protocols

Think-aloud protocols also serve as side notes for checking the participants' screen-based activities in real time. More importantly, the participants' statements provide clues about what they are thinking at a particular point in time. In this process, the researcher acted as a listener while participants voiced their thoughts, fostering ongoing interaction. Participants had no time limit to watch the video, allowing them to verbalise their thoughts during playback or pause and reflect before sharing.

Simple non-directive prompts (i.e., "yes," "oh," "okay") and nonverbal cues (nodding, smiling) were used to encourage more verbalisations. To mitigate minimal verbalisation, participants were instructed during the demonstration to articulate their thoughts as much as possible. Verbalisation was entirely voluntary. No prompts were given following moments of silence.

3.4.3. Interviews

The interview session started after each participant finished watching the video. The interviews explored questions about the participants' attitudinal issues related to the video and subtitles, delving into the audience's involvement.

The semi-structured interview lasted 10 to 20 minutes. Open-ended questions were asked, along with follow-up questions based on participants' responses. The first question is: "Who do you think subtitled the video and why?" Participants were unaware that the video contained two subtitle modes and were provided options such as professionals, fans, and MT to indicate the possibility of multiple subtitle modes combined.

Afterwards, I informed participants about the two non-standard subtitling modes and asked them to identify which half of the video was subtitled by MT. Then, participants rated their level of involvement on a scale from zero to ten and explained their chosen score. Additionally, participants were asked whether they considered watching the show in their spare time and provided reasons for doing so.

3.5 Data Analysis

3.5.1. Quantitative Data

SPSS and Microsoft Excel were used to analyse numerical data and create figures. Screen recordings provided data on the number of screen-based activities and their timelines. T-tests were conducted to examine Hypothesis 2, exploring the relationship between screen-based activities and the two subtitle modes.

The only numerical data in the interviews is the participant's self-reported involvement, measured on a scale from zero to ten (totally not involved to deeply involved). There were two sets of categorical data in the interview phase. The first was each participant's answer to the question: "Do you want to keep watching a few more episodes?" Their answers included "Yes", "Probably", and "No". The second is the participant's answer to the question: "Who do you think subtitled the video?" I divided their answers into two major categories: (1) awareness of machine-translated subtitles and (2) non-awareness of machine-translated subtitles. If the participants thought the video or part of the video was generated by raw MT or PEMT, they were in the former category. The remaining participants were then put into the latter category. This categorisation is used for testing Hypothesis 1. For Hypothesis 3, a t-test and a chi-squared test were conducted to explore the relationship between awareness of machine translation and the indicators of audience involvement (self-reported involvement and willingness to continue watching).

3.5.2. Qualitative Data

All participants' verbalisations were transcribed and organised according to the video's chronological order. This presentation format can locate what each participant was commenting on and how their understanding of the video evolved or changed.

A thematic approach was used to code the statements using NVivo 12. Since participants' statements about the same scene often focused on different aspects, codes were developed based on the themes present in the participants' statements. Each code corresponded to a specific theme. For example, as shown in Table 3, participants A03 and A04 both commented at 01:44 but addressed different elements of the scene, resulting in separate codes. Statements addressing the same theme were grouped under a single code, and new codes were created when additional themes emerged. This iterative process was applied across all statements.

In the interviews, although most participants provided detailed answers, some responded briefly, and others gave irrelevant or off-topic responses. To address this, a combined deductive and inductive approach was used for coding and analysis. In the deductive phase, top-level codes were derived from the interview questions, and second-level codes emerged through multiple readings of the data. For example, the top-level code "Continue watching or not" had three second-level codes: "Yes," "No", and "Probably". Varied reasons for a decision led to third-level codes like "viewing habits" and "motivation". Thus, a text segment could be coded at different levels.

Table 3

Examples of Codes in Think-Aloud Protocols

Participant	Statement	Code
01:44 [Scene]	Lin Chen checks Mei Changsu's pulse	
A03	Weird hairstyle. I guess it's traditional.	Costumes
A04	Is he [Mei Changsu] sick or something?	Assumption
06:28 [Fansubs]	I asked him to come to the capital to recuperate.	
B02	[rewind from 06:30]: I have to catch that last bit again.	Subtitle speed
06:28 [MTS]	I invited him to rest in the Capital	
A05	What is wrong with his health? Is it because he had a broken heart?	Confusion

4. Results and Discussion

4.1. Identifying Machine-Translated Subtitles

Since participants were not informed about the presence of machine-translated subtitles beforehand, their responses varied significantly when asked who subtitled the video. No participants identified that the video included two subtitle modes. This lack of recognition can be attributed to their limited experience with Chinese audiovisual content. As mentioned previously, only two participants watched Chinese audiovisual content relatively frequently. Without prior knowledge of machine translation or subtitled Chinese shows, most participants lacked a benchmark to make informed judgments about the subtitle modes.

Five participants believed that they watched fansubs. Five participants said the subtitles were "created by a machine with human correction later on", so they were categorised under "Post-editing". Three participants thought the video was subtitled by professionals. Two participants believed the subtitles were generated by MT without any human intervention, thus classified as "Raw MT". Two participants believed human translators subtitled the video but could not determine whether they were professionals or non-professionals, so they were placed in the "Not MT" category. One participant speculated that the researcher created the subtitles, stating, "I thought it was you only because it seems those subtitles are intentionally messed up".

In the follow-up question, the participants needed to specify which half was subtitled by machine translation and justify their decision. Six participants in Group A and seven in Group B chose the correct half. Although the two groups watched the video with subtitles in reversed sequences, a chi-square test showed no significant association between the subtitle sequence and correct identification of subtitle modes ($p = 0.60$). This indicates that Group A's exposure to more translator's notes did not aid their distinction between the subtitle modes. However, the small sample size may limit the detection of statistically significant differences.

The correct identification mainly relies on noticing the translator's notes or spotting problems in machine-translated subtitles. The translator's notes served as a basis for identification. For example, B05 mentioned, "Where there'd be notes, it would have been translated by fans". MT errors concerned lexical, fluency, and contextual issues that do not require prior knowledge of Chinese to identify. First, they noticed "nonsensical" errors, such as "Mansion of Xie" becoming "House of Thanks" ("谢" *xie* can be a person's surname or mean "to thank"). Second, fluency and readability were major metrics participants paid attention to. For example, A08 commented that "the first half (fansubs) is more fluent, so fans would do this part", and A09 said, "The MT part feels a bit convoluted..." These comments suggest that participants often associated a natural flow in language with human effort. Last, participants recalled the parts where they had difficulty following the content and then assigned machine-translated subtitles to those parts. According to B07, "I think it started to make more sense in the second half. In the first half, the machine loses the sense part and just gets the words".

A02 and B03 incorrectly identified the subtitle mode. A02 did not know the function of a translator's note, assuming these notes were similar to glossary descriptions and had been automatically added by the algorithm. This misunderstanding could be attributed to unfamiliarity with subtitling conventions or a lack of machine translation literacy. B03 did not find any errors in the video, so she made a random guess.

Three participants said, "I don't know". A04 admitted that she could not tell the difference between the subtitles in the two halves. A07 recalled seeing translation issues in both halves of the video, so she could not determine which half used MT. B04 stated that she did not have enough knowledge of MT to make an informed decision.

The participants' answers to this question not only reflected their opinions on subtitle accuracy but also manifested their perceptions of different modes of subtitling shaped by their viewing experience.

Professional subtitling was considered error-free and they believed that the use of translator's notes was uncommon in this form of subtitling. When it comes to fansubs, participants mentioned three main aspects. First, participants observed that fansubbers often include translator's notes to clarify complex cultural references. Second, they perceived fansub translation quality as inferior to professional subtitling, such as omissions of complex expressions. Third, fansubbed versions were

considered more colloquial than professional subtitles. Concerning machine-translated subtitles, participants' familiarity was primarily limited to using YouTube auto-captions and Google Translate, as noted by A01, B01, and B03.

Revisiting Hypothesis 1, we can conclude that participants cannot distinguish machine-translated subtitles from fansubs, and less than half of them recognised the presence of machine-translated subtitles. First, participants were not prompted to pay special attention to the subtitles, so searching for errors was not their priority during viewing. Additionally, no prior knowledge of the source language made them unable to identify certain major translation errors that distort the original meaning but still make sense in the target language. Thus, the differences in translation accuracy between the two subtitle modes were not apparent enough to be easily distinguished. On the other hand, participants' perceptions of different subtitle modes guided their decisions, but these perceptions were shaped by the limited use of MT and an underestimation of MT accuracy. Thus, when informed which part was machine-translated, they made comments like "It was quite well done for a machine" (A01).

4.2. Screen-Based Activities and Subtitle Modes

The second hypothesis explores the relationship between screen-based activities and subtitle modes. Two t-tests show that there is no significant correlation between the two variables in both the first and second halves of the video (first half: $p = 0.27$; second half: $p = 0.26$).

Of the eighteen participants, eight did not initiate any screen-based activities and B02 paused six times, which was the largest number of screen-based activities. As shown in Table 4, the most common reason for participants to pause or rewind was re-reading the translator's notes.

Table 4

Five Main Reasons That Trigger Screen-Based Activities

Reasons	Number of screen-based activities
Translator's notes	7
Subtitle speed	6
Missing context	5
Translation errors	3
Interest in on-screen items	2

Six subtitles were considered fast by A06 and B02. They paused the video to read the subtitles. The subtitle speed of the six subtitles ranges from 14 to 24, two fansub lines and four machine-translated subtitles. However, the six subtitles are not the fastest ones among all subtitles, as shown in Table 1.

A potential explanation, as suggested by A06, is that the challenge stemmed from the sequencing: "That was a really long sentence, and the next one, though shorter, was still quite long. Having them one after the other made it hard to keep up." Another potential factor could be scene changes, as rapid transitions between shots may make it more difficult to process the information in the subtitles.

Three translation errors in machine-translated subtitles triggered pauses and going back:

1. Incorrect translation of name: mistranslating "Mr Su Zhe" as "Mr Sutzer". Although the mistranslated name did not cause misunderstanding, it failed to follow translation norms.
   A101 [going back from 08:04 to 08:02 and pausing 2 seconds]: "Mr Sutzer", hahaha. Definitely not a Chinese name.
2. Capitalisation problem: "the royal family of Yu" becoming "the royal Family of yu". The capitalisation issue was minor and did not lead to misunderstanding.
   B09 [going back from 00:50 to 00:41]: Wait… that subtitle had weird capitalisation… "royal family" and then, so that's weird capitalisation…
3. Misuse of pronoun: mistranslating "your pulse" as "my pulse". The original Chinese sentence "**我**诊完脉可是什么都没说" uses ellipsis, which is common in pro-drop languages like Chinese. The correct translation is, "After I checked your pulse, I didn't say anything". The MT system failed to "recognise" the omitted pronouns due to the pro-drop nature of Chinese.
   B09 [going back from 02:19 to 02:14]: "I finished my pulse but didn't say anything"? That's really weird. I'm not quite sure what's going on in that conversation. It's a mistranslation…

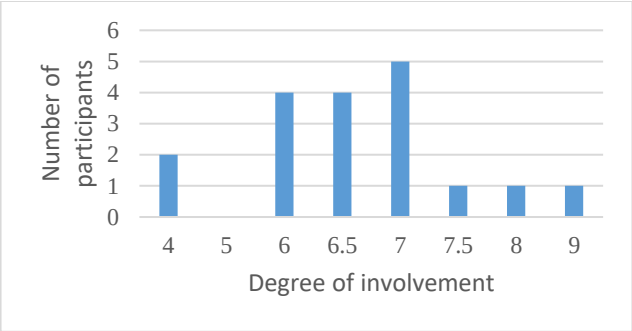
Hence, screen-based activities were triggered by numerous elements in the video and were not restricted to questionable subtitles. Even when strictly considering subtitle-related issues, the pauses and rewinds were not limited to machine-translated subtitles. Thus, the number of screen-based activities does not correlate with the modes of subtitles. Subtitles are merely a component of the target text, and viewers engage with the audiovisual text in its entirety. This highlights the holistic nature of viewer interaction with the audiovisual text, where translation is just one of many factors influencing the overall experience.

4.3. Audience Involvement

Audience involvement was measured through self-reported involvement and the willingness to continue watching. In self-reported involvement, some participants did not provide a specific number but a range, such as 6 to 7. In those cases, the number was counted as 6.5 for calculation purposes. A06 and B07 gave the lowest number (4/10), and A04 gave the highest number 9/10 (M = 6.44, see Figure 2). Since A05 had watched the show before, her data was eliminated.

Figure 1

The Eighteen Participants' Self-Reported Degrees of Involvement



When the participants explained what features affect their involvement, the following four elements received both positive and negative comments:

1. Subtitles: A01 and B04 praised the quality of the translated subtitles, but A05, A09, and B06 found them too fast to follow. The remaining seven participants expressed that improved subtitle accuracy would enhance their involvement.
2. The pace of the video: B08 found the pace comfortable, while A09, B01, B06, and B07 felt the pace was unsmooth, introducing too much information too quickly.
3. Action scenes: A01, A03, and A08 enjoyed the fighting scenes, finding the martial arts movements "cool", while A02 deemed the flying and bouncing movements "unrealistic".
4. Acting: A01 and A07 criticised the acting as wooden and the emotions as unreal, whereas B05 praised the acting.

Participants were later asked whether they would continue watching the show. Four participants said "Yes", driven by curiosity about the follow-up plot. Six participants stated "probably", explaining that they needed to watch more episodes to decide. Seven participants indicated no intention to continue watching. They mentioned that it was not "their type of genre"; they preferred reality TV shows or modern-day dramas because these genres were more relatable than period dramas. Their preferences have been shaped through long-term viewing experiences and habits, which are difficult to change through exposure to a period drama.

Hypothesis 3 looks at the relationship between audience involvement and awareness of machine-translated subtitles. As discussed in 4.1, MT errors are one of the main factors that helped participants identify the subtitle modes. Put another way, the association between audience involvement and awareness of machine-translated subtitles concerns the effect of problematic subtitles on involvement.

For the first sub-hypothesis, a t-test shows a statistically significant difference in self-reported involvement between the seven participants aware of machine translation (M = 5.57, SD = 1.10) and the ten participants unaware of it (M = 6.95, SD = 0.86; p < 0.001). This implies that those who

detected machine-translated subtitles tended to feel less involved in the video. However, regarding the second sub-hypothesis, there was no statistically significant difference between the participants who noticed the presence of MT and those who did not in the willingness to continue watching (X^2(2, 16) = 3.66, p = 0.16).

One possible explanation for the divergence is that self-reported involvement and willingness to continue watching represent two different states of involvement: the former being more immediate and the latter a more sustained, graded state. MT errors can act as barriers that create instant disruptions. Upon identifying these errors, participants became increasingly suspicious of the content's accuracy, leading them to be more critical and attentive to subsequent errors. This heightened scrutiny can interrupt the immersive experience, as found by Tuominen et al (2023). However, these disruptions were not severe enough to produce a lasting effect that might prompt participants to stop viewing—an outcome influenced, to some extent, by the controlled nature of the experimental setting.

The graded state of involvement reflects the extent to which participants desire to further engage with the narrative. This state of involvement requires strong curiosity about plot development and genuine interest in the genre shaped by long-term viewing experiences. Thus, instant disruptions may affect immediate involvement; the deeper, more sustained involvement is supported by an ongoing interest in the narrative.

4.4. Audience-Text Interactions

Self-reported involvement and willingness to continue watching inform us about the factors influencing audience involvement and the outcomes of being (not) involved. Further, the analysis of verbalisations showcases what constitutes an immersive experience.

Firstly, an immersive experience entails the audience's engagement in the world created by cinematography, leading to a sense of transportation, the feeling of being absorbed in the narrative world (Green & Brock, 2000). Although participants were only shown nine-minute extracts from *Nirvana in Fire*, deep emotional immersions can be observed:

> A04 [at 09:19] Oh, no! I want to know more.
>
> B02 [at 01:25]: I'm interested to know what show it is.
>
> B05 [at 09:19]: Very dramatic and interesting. Where can I watch it?

Holland (1975) mentions the experience of "switching" between feeling totally engrossed and being aware of oneself. In the experiment, the absurdity of physics-defying martial arts skills might temporarily pull participants out of the narrative world, especially if Western viewers are not accustomed to such conventions. Later, their curiosity about the interpersonal relationships in the

video could lead them to re-immerse themselves in the narrative. This switching experience may explain why participants felt a sense of complete involvement at certain points during viewing, but their post-hoc responses indicated that they did not have strong intentions to continue watching.

Second, the audience-text interaction goes beyond the involvement in the narrative world. Viewers immerse themselves in a chosen character to navigate through the challenging situations depicted in the show. They internalise the character's views and attitudes, seeing the world as another person sees it. This can be reflected in how viewers address the characters. Some participants use a blend of first and third-person perspectives, known as free indirect discourse. For example, when B01 saw the male lead hiding in the carriage and glancing at his fiancée, she said, "This guy is making a decision. He's like, I'm going to regret this, but I have to. I can't meet her now."

Audiovisual shows often contain characters that viewers compare to each other and to whom they come to feel close: Viewers can not only want to be in their favourite character's shoes, but also try to approach those characters in a manner like the people they encounter in real life (Sheldon et al., 2021). Therefore, some viewers' verbalisations use the second person to address the character, just like talking to a friend in a conversation:

> B03 [03:21]: Yes man! You should blow smoke up his ass. Give him all the expensive elixirs!
>
> B09 [04:34]: You can't get on a horse like that.

Interestingly, although the conversations are inherently one-sided—no real conversational give-and-take is happening—viewers sometimes respond to the characters emotionally as if they were real people (Sheldon et al., 2021). This para-social interaction may involve an affective connection with characters and storylines that enhances a sense of presence in the viewer-character relationship (Tukachinsky & Sangalang, 2016).

Narrative-related and character-related involvement operate at cognitive and emotional levels, with participants contemplating details, dialogue, and contextual clues in the show while aligning with the characters' perspectives and emotions. At this level, translated subtitles become the actual words spoken by the characters, transcending their role as mere translations—a concept described by Pedersen as "a contract of illusion" (2011, p. 22). As mentioned, when MT errors occur, they break this illusion and momentarily disrupt the immersive experience. However, if the narrative is compelling and the character is attractive enough, it can swiftly recapture viewers' attention, drawing them back into the story. This dynamic interaction suggests that the narrative and character development have a more profound impact on audience engagement than subtitle accuracy. In other words, the robustness of the narrative and the emotional connection with characters serve as the primary drivers of sustained involvement, overriding the influence of occasional translation inaccuracies. Translation, therefore, plays a limited role in influencing audience involvement compared to the broader impact of storytelling and cinematography.

5. Conclusion

This study compares the audience's reception of fansubs and raw machine-translated subtitles by investigating eighteen participants' ability to distinguish between the two, differences in screen-based activities triggered by these subtitle modes, and the relationship between awareness of MT and audience involvement.

Results show that no participant sensed the existence of two subtitle modes or the shift between them. However, of the eighteen participants, seven correctly identified that the video contained machine-translated subtitles. Although most participants found the machine-translated subtitles acceptable, only two reported being satisfied with the translated subtitles. Three problems in subtitles stood out: awkward lexical choices that did not fit the context, unreadable texts, and subtitles that did not make sense in the target language. Raw machine-translated subtitles do not necessarily trigger more screen-based activities as viewer attention extends beyond translation errors. Other elements, including context, on-screen objects, translator's notes, and fast subtitles also triggered screen-based activities. These activities may sometimes disrupt immersion but indicate instances of interest in the audiovisual product, such as viewers' curiosity about unfamiliar objects.

In terms of audience involvement, participants who noticed MT reported feeling less involved compared to those who were unaware of it. Increased attention to subtitles prompted by detected errors may correspond to a higher likelihood of spotting more errors, causing immediate disruption in audience involvement. As more errors are identified, participants may be less able to feel involved in the story again. However, no evidence suggests awareness of machine-translated subtitles leads to disinterest in continuing the show. This decision is influenced by viewers' viewing habits and interests, rather than the accuracy of the subtitles. Narrative-related and character-related involvements were observed during the participants' viewing process. This indicates that the given video presents a captivating story that can bring involvement to viewing experiences, regardless of the subtitle modes presented to the viewers. However, the impact of subtitles, as an important information channel in audiovisual texts, requires further investigation due to the small sample size and the limited duration of the research material.

Although accurate subtitles can enhance the contract of illusion between the viewer and the narrative, a deeper and more nuanced engagement requires a compelling depiction of the story world and personal interest. Transparency is thus important when using machine translation; this may help viewers manage their expectations regarding subtitle accuracy (Tuominen et al., 2023). Further, language service providers can provide information about automation and its current stage of development to support the informed and reflective use of automated subtitles.

References

Alasuutari, P. (1999). Introduction: Three phases of reception studies. In P. Alasuutari (Ed)., *Rethinking the media audience: The new agenda* (pp. 2–21). Sage.

Caffrey, C. (2012). Using an eye-tracking tool to measure the effects of experimental subtitling procedures. In E. Perego (Ed.), *Eye tracking in audiovisual translation* (pp. 223–258). Aracne.

Calvo-Ferrer, J. R. (2023). Can you tell the difference? A study of human vs machine-translated subtitles. *Perspectives*. https://doi.org/10.1080/0907676X.2023.2268149

Chen, L., & Chang, K. (2024). A corpus-based approach to the reception of Chinese television dramas streamed overseas. *Humanities and Social Sciences Communications*, *11*(1), 417. https://doi.org/10.1057/s41599-024-02897-7

Chesterman, A. (2007). Bridge concepts in translation sociology. In W. Wolf & A. Fukari (Eds.), *Constructing a sociology of translation* (pp. 171–186). John Benjamins.

Cohen, J. (2001). Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication and Society*, *4*(3), 245–264. https://doi.org/10.1207/S15327825MCS0403_01

Díaz Cintas, J., & Muñoz Sánchez, P. (2006). Fansubs: Audiovisual translation in an amateur environment. *The Journal of Specialised Translation*, *6*, 37–52.

Frey, F. (2018). The experiential mode of media reception: A holistic framework concept. *Communication Theory*, *28*(4), 487–510. https://doi.org/10.1093/ct/qty010

Gambier, Y. (2018). Translation studies, audiovisual translation and reception. In E. Di Giovanni & Y. Gambier (Eds.), *Translation studies, audiovisual translation and reception* (pp. 43–66). John Benjamins. https://doi.org/10.1075/btl.141.04gam

Gervais, R., & Merchant, S. (2001). *The office.* BBC.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*, 701–721. https://doi.org/10.1037/0022-3514.79.5.701

Green, M. C., & Sestir, M. (2017). Transportation theory. In *The international encyclopedia of media effects* (pp. 1–14). John Wiley & Sons. https://doi.org/10.1002/9781118783764.wbieme0083

Hall, S. (1973). *Encoding and decoding in the television discourse*. Centre for Contemporary Cultural Studies, University of Birmingham.

Hansen, G. (2013). The translation process as object of research. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 88–101). Routledge.

Hill, A. (2018). Media audiences and reception studies. In E. Di Giovanni & Y. Gambier (Eds.), *Translation studies, audiovisual translation and reception* (pp. 3–20). John Benjamins.

Holland, N. N. (1975). *The dynamics of literary response*. Norton.

Hou, H. (2015). *Nirvana in Fire* (Season One). Daylight Entertainment. https://www.viki.com/tv/22943c-nirvana-in-fire

Hu, K., O'Brien, S., & Kenny, D. (2020). A reception study of machine translated subtitles for MOOCs. *Perspectives*, *28*(4), 521–538. https://doi.org/10.1080/0907676X.2019.1595069

Hu, K., & Li, X. (2023). The creativity and limitations of AI neural machine translation: A corpus-based study of DeepL's English-to-Chinese translation of Shakespeare's plays. Babel. https://doi.org/10.1075/babel.00331.hu

Ji, Y. (2022). More high-quality Chinese TV dramas see success around the world. *Global Times.* https://www.globaltimes.cn/page/202212/1282428.shtml

Künzli, A., & Ehrensberger-Dow, M. (2011). Innovative subtitling: A reception study. In Alvstad, C., Hild, A., & Tiselius, E. (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 187–200). John Benjamins.

Li, J. (2022). How much does a subtitle say?: A critical reception study of Chinese television dramas streamed overseas. In M. Samuel & L. Mitchell (Eds.), *Streaming and screen culture in Asia-Pacific* (pp. 197–215). Springer. https://doi.org/10.1007/978-3-031-09374-6_11

Murray, J. B. (1997). *Hamlet on the holodeck: The future of narrative in cyberspace*. Free Press.

Nilsson, N. C., Nordahl, R., & Serafin, S. (2016). Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology*, *12*(2), 108–134. http://dx.doi.org/10.17011/ht/urn.201611174652

Orrego-Carmona, D. (2015). *The reception of (non)professional subtitling* [Unpublished doctoral dissertation]. Universitat Rovira i Virgili.

Orrego-Carmona, D. (2019). A holistic approach to non-professional subtitling from a functional quality perspective. *Translation Studies*, *12*(2), 196–212. https://doi.org/10.1080/14781700.2019.1686414

Pedersen, J. (2007). Cultural interchangeability: The effects of substituting cultural references in subtitling. *Perspectives, 15*(1), 30–48. https://doi.org/10.2167/pst003.0

Pedersen, J. (2011). *Subtitling norms for television*. John Benjamins.

Sheldon, Z., Romanowski, M., & Shafer, D. M. (2021). Para-social interactions and digital characters: The changing landscape of cinema and viewer/character relationships. *Atlantic Journal of Communication*, *29*(1), 15–25. https://doi.org/10.1080/15456870.2019.1702550

Tukachinsky, R., & Sangalang, A. (2016). The effect of relational and interactive aspects of para-social experiences on attitudes and message resistance. *Communication Reports*, *29*, 175–188. https://doi.org/10.1080/08934215.2016.1148750

Tuominen, T. (2019). Experiencing translated media: Why audience research needs translation studies. *The Translator*, *25*(3), 229–241. https://doi.org/10.1080/13556509.2019.1615694

Tuominen, T., Koponen, M., Vitikainen, K., Sulubacak, U., & Tiedemann, J. (2023). Exploring the gaps in linguistic accessibility of media: The potential of automated subtitling as a solution. *The Journal of Specialised Translation*, *39*, 77–98.

Wu, Z., & Chen, Z. (2022). Towards a corpus-driven approach to audiovisual translation (AVT) reception: A case study of YouTube viewer comments. *Journal of Specialised Translation*, *38*, 128–154.

Zhang, S. (2022). Exploring how Chinese TV dramas reach global audiences via viki in the transnational flow of TV content. *Journal of Transcultural Communication*, *2*(1), 69–89. https://doi.org/10.1515/jtc-2022-0014