

Automatic Subtitling and Machine Translation in Aiding  
English to Thai Simultaneous Interpretation



Mr. Nutdanai Sornchai

An Independent Study Submitted in Partial Fulfillment of the  
Requirements  
for the Degree of Master of Arts in Translation and Interpretation  
Field of Study of Translation and Interpretation  
FACULTY OF ARTS  
Chulalongkorn University  
Academic Year 2021  
Copyright of Chulalongkorn University

การใช้คำบรรยายได้ภาพอัตโนมัติและคำแปลอัตโนมัติเพื่อช่วยในการล่ามพูดพร้อมจาก  
ภาษาอังกฤษเป็นภาษาไทย



สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาอักษรศาสตรมหาบัณฑิต  
สาขาวิชาการแปลและการล่าม สาขาวิชาการแปลและการล่าม  
คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2564  
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Independent Study Title      Automatic Subtitling and Machine Translation in Aiding  
English to Thai Simultaneous Interpretation  
By                                      Mr. Nutdanai Sornchai  
Field of Study                      Translation and Interpretation  
Thesis Advisor                      Assistant Professor NUNGHATAI  
RANGPONSUMRIT, Ph.D.

---

Accepted by the FACULTY OF ARTS, Chulalongkorn University in Partial  
Fulfillment of the Requirement for the Master of Arts

INDEPENDENT STUDY COMMITTEE

----- Chairman  
(Assistant Professor SONGSAK MADSALAE, Ph.D.)

----- Advisor  
(Assistant Professor NUNGHATAI  
RANGPONSUMRIT, Ph.D.)

----- Examiner  
(Assistant Professor ATTAPOL  
THAMRONGRATTANARIT, Ph.D.)



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

ณัฐดนัย ศรีไชย : การใช้คำบรรยายได้ภาพอัตโนมัติและคำแปลอัตโนมัติเพื่อช่วยในการล่ามพูดพร้อมจากภาษาอังกฤษเป็นภาษาไทย. ( Automatic Subtitling and Machine Translation in Aiding English to Thai Simultaneous Interpretation) อ.ที่ปรึกษาหลัก : ผศ. ดร.หนึ่งหทัย แรงผลสัมฤทธิ์

เทคโนโลยีการรู้จำเสียงอัตโนมัติ (ASR) และเทคโนโลยีการแปลด้วยเครื่อง (MT) ได้พัฒนาอย่างมีนัยสำคัญในช่วงไม่กี่ปีที่ผ่านมา ทำให้นักวิจัยต่าง ๆ พยายามนำเทคโนโลยีดังกล่าวเพื่อช่วยในการล่ามพูดพร้อม ผู้วิจัยจึงได้จัดทำงานวิจัยนี้เพื่อศึกษาประเด็นวิจัยข้างต้น งานวิจัยนี้มีผู้เข้าร่วมวิจัยสี่คนทำการล่าม (จากภาษาอังกฤษเป็นภาษาไทย) 3 ครั้ง ได้แก่ ล่ามโดยมีคำบรรยายได้ภาพอัตโนมัติ ล่ามโดยมีคำแปลอัตโนมัติ และล่ามโดยไม่มีคำบรรยายใด ๆ จากนั้นจึงให้ผู้ฟังให้คะแนนการแปลและตอบแบบสอบถามความเข้าใจ ผู้เข้าร่วมวิจัยได้ให้ความเห็นว่าคำบรรยายอัตโนมัติช่วยให้ตรวจสอบการแปลได้ ในขณะที่ MT ให้วิธีการคิดในระหว่างแปล ส่วนความเห็นจากผู้ฟังนั้นทำให้สรุปได้ว่าคำบรรยายอัตโนมัติมีข้อได้เปรียบเล็กน้อย อย่างไรก็ตามข้อได้เปรียบดังกล่าวอาจถือได้ว่าไม่มีนัยสำคัญทางสถิติและควรทำการศึกษาเพิ่มเติม ซึ่งในอนาคตยังอาจทำการวิเคราะห์เพิ่มเติมในด้านความถูกต้อง คุณภาพการล่าม และจิตวิทยาของภาษากายในระหว่างการล่ามได้อีกด้วย งานวิจัยนี้ให้ข้อมูลว่าเทคโนโลยี ASR และ MT อาจเป็นประโยชน์เมื่อคนทำการล่ามพูดพร้อมได้อย่างไร

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

สาขาวิชา การแปลและการล่าม

ลายมือชื่อนิติ

ปีการศึกษา 2564

.....  
ลายมือชื่อ อ.ที่ปรึกษาหลัก

.....

# # 6388028022 : MAJOR TRANSLATION AND INTERPRETATION

KEYWORD simultaneous interpretation, ASR, machine translation, English to Thai interpretation

Nutdanai Sornchai : Automatic Subtitling and Machine Translation in Aiding English to Thai Simultaneous Interpretation. Advisor: Asst. Prof. NUNGHATAI RANGPONSUMRIT, Ph.D.

With the development of automatic speech recognition (ASR) and machine translation (MT) technology in recent years, their uses in simultaneous interpretation (SI) have been a subject of interest. This research, conducted to study these technologies in aiding the SI process, has four participants interpreted (English to Thai) three comparable speeches with automatic subtitling, with MT, and without subtitling. An audience then rated the interpretation renditions and answered a comprehension questionnaire. According to the participants, automatic subtitling allowed them to cross-check their translation, while MT provided a way to decrease the cognitive effort. According to the audience, automatic subtitling has a slight advantage; however, this could be considered statistically insignificant, and further studies are needed. Further analysis could also be made concerning the accuracy, interpretation qualities, and psychoanalysis of body language during the interpretation session. This research provides information on how ASR and MT technologies could be beneficial regarding human SI.



Field of Study: Translation and Interpretation  
Academic Year: 2021

Student's Signature  
.....  
Advisor's Signature  
.....

## ACKNOWLEDGEMENTS

This research would not have been possible without the help of countless people, whom I would like to take this opportunity to express my gratitude.

Thank you very much, Asst. Prof. Nunghatai Rangponsumrit, Ph.D., my main research advisor. Thank you for your invaluable guidance you provided for me, and the extreme patience you have with me. Many times, you pointed me to places and paths I did not even know exist. Without you, this research would have been very much impossible.

Thank you very much, Asst. Prof. Songsak Madsalae, Ph.D. and Asst. Prof. Attapol Thamrongrattanarit, Ph.D., members of the Special Research committee. Your guidance, insights, and recommendations have been immensely helpful to this research.

Thank you very much, Nipaporn Angkawanich, Ph.D. for assisting me through many steps of completing this research.

Thank you very much, the participants and the audience for this research. A special mention to Asst. Prof. Nattaporn Tonanon, D.Eng. and Lect. Nonthawat Prachantasena, my former project advisor and my mentor respectively, who were members of the audience for the research. For everyone else, it has been a great opportunity to work with you all, one of whom I call my little sister.

Finally, thank you very much, my mother and father, my family, my partner, and my friends, for your endless moral support along the way.

Nutdanai Sornchai

## TABLE OF CONTENTS

	<b>Page</b>
.....	iii
ABSTRACT (THAI) .....	iii
.....	iv
ABSTRACT (ENGLISH) .....	iv
ACKNOWLEDGEMENTS .....	v
TABLE OF CONTENTS .....	vi
1 Rationale .....	2
2 Previous Works .....	3
2.1 ASR Technology and Its Uses in Interpretation .....	3
2.2 MT Technology and Its Uses in Interpretation .....	3
2.3 Interpretation Quality Assessment .....	4
3 Objectives .....	5
4 Methodology .....	5
4.1 Design .....	5
4.2 Participants and Material .....	5
4.3 Procedure .....	6
5 Results .....	8
5.1 Interpreters' Feedback .....	8
5.2 Audience Response .....	8
5.2.1 Audience's Comprehension Quizzes Responses .....	8
5.3 Anecdotal Comments Regarding Interpreters' Body Language .....	11
6 Conclusion .....	12
Appendix A: Audience's Questionnaire Sample .....	13
REFERENCES .....	15
VITA .....	18



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**



To  
Senior Colonel Pitsawong Sornchai  
an Arts student herself, who succumbed to lung cancer and never had the chance to  
see two of her grandchildren graduate from the faculty she once studied in.

Grandma, this one is for you.



## 1 Rationale

It is undeniable that simultaneous interpretation (SI) has a vital role in international communication. Though there have been many developments in automatic SI systems albeit with a number of issues (Wu & Wang, 2018), as of 2022 human SI remains in widespread use. Thus, improvements in human interpretation have always been a subject of interest.

Developments in automatic speech recognition (ASR) and automatic subtitling have been very extensive. Although as of 2022 these technologies still leave room for improvement, automatic subtitling has already seen widespread use, such as in meeting minutes transcription (e.g., Amazon Transcribe), video streaming (e.g., CaptionHub (Jameson & Pengelly, 2019), YouTube), and conferences. These automatic subtitling are often useful in aiding the understanding of the speakers. Thus, there have been studies and attempts in exploiting ASR in aiding and reducing the load of the interpreters (Defrancq & Fantinuoli, 2020) (Rodríguez et al., 2021). Nonetheless, studies regarding the use of ASR and automatic subtitling during the interpretation process are rare.

Regarding machine translation (MT), developments as of 2022 are advanced enough that they are often used in many contexts, including document translation and early development in speech interpretation, the latter of which has had moderate results (Alves, 2021). The development of MT has also lent its hand in automatically translating videos. This technology is most prominently used in video streaming websites, with YouTube being the most well-known (Microsoft Corporation, n.d.), though there have studies showing that MT of automatic subtitling, particularly on YouTube, still suffers from frequent translation errors (Laksana & Putri, 2018), such as grammatical errors and poor word choices (Manurung, 2021).

Regarding the Thai language, Pewnim (2017) concluded that English-Thai Machine Interpreting was still unusable in practical application and identified several shortcomings in the ASR and MT technologies. With the advancements in these technologies since then, this research was conducted to gain insights into how well ASR, automatic subtitling, and MT can aid simultaneous interpreters.

## **2 Previous Works**

### **2.1 ASR Technology and Its Uses in Interpretation**

Over the past few decades, the development in ASR has been vast, and the technology has seen widespread use in many cases, such as in subtitles in conferences and aiding hearing-impaired users. For example, Zhao et al. (2006) have developed an automatic subtitling system in telemedicine to further improve accessibility for hearing-impaired patients. Moreover, ASR has also seen heavy uses in video streaming as automatic subtitling, such as on YouTube and other video streaming platforms. This automatic subtitling could be beneficial regarding the accessibility of these videos, though it was shown that YouTube automatic subtitling still has significant errors, affecting both the automatic subtitling's quality itself and the subsequent translation (Laksana & Putri, 2018).

There have been studies and attempts in using the ASR technology in reducing interpreters' workload. Gaber and Corpas-Pastor (2021) have proposed the use of ASR in training interpreters by creating an ad-hoc corpus and aiding these interpreters in the preparation and documentation phase before the interpretation process. Many ASR tools were studied and compiled, and it was expected that the system would reduce the cognitive load of the interpreters. Gaber et al. (2020) have also designed a procedure for using ASR technology in the conference preparation of the interpreters, and also in the extraction of terminology. Nonetheless, studies regarding the use of ASR and automatic subtitling in aiding interpreters during SI are scarce.

### **2.2 MT Technology and Its Uses in Interpretation**

MT sees significant improvements in the past decade thanks to developments in artificial intelligence. Nowadays, automatic translation services are ubiquitous, such as Microsoft Translator, Google Translate, and Apple translation function. These developments have led to many researchers attempting to create an automatic interpretation system. Wu and Wang (2018) have created a prototype system in English for Japanese SI. Though their research was still in the early stages, it saw promising results in translation quality and matching the speed of the speakers. Sawaf (2012) has also designed a system using ASR technology with human edited-MT for the translation of live Arabic speakers on television into English. The system mentioned in the research was able to provide the translation within a relatively short period of time of less than one hour, or less than 30 minutes in some cases, from the initial television broadcast.

Developments since 2016 in direct speech-to-text translation without the use of ASR. Berard et al. (2016) explored a preliminary end-to-end system for French-English translation from speech that does not use the transcription of the source. Bansal et al. (2017) used "unsupervised term discovery (UTD) to cluster repeated patterns in audio" and employed these said patterns to train a system of Spanish-English MT model that pairs the patterns with the translation directly.

Nevertheless, studies have shown that, as of 2022, automatic interpretation systems still have drawbacks that in some cases render the system unsuitable for practical use. Research by Pewnim (2017) has shown that the automatic interpretation system from English to Thai using Google Translate was not suitable for practical use, stemming from various issues, particularly in the quality of the translation from the MT engine itself. This issue was compounded by the poor “chunking” or sentence boundary detection by the ASR system, which was also previously demonstrated to have an adverse effect on the translation quality. (Fügen & Kolss, 2007) The research by Wu and Wang (2018) also showed that the automatic interpretation system still suffered from poor logic and cultural awareness.

### **2.3 Interpretation Quality Assessment**

One of the studies in assessing the quality of an interpreter that stressed the importance of linguistic quality and translation accuracy is by Bühler (1986), where the professional interpreters have overwhelmingly rated fluency of delivery and consistency with the original message being two of the most important factors in sponsoring an interpreter into the AIIC.

Regarding the quality comparison between interpretation techniques, there are a few studies in this area, one of these being the comparison between the conventional consecutive and simultaneous consecutive (Sim-Consec) techniques by Hamidi and Pöchhacker (2007). In this work, the researchers assess the quality of the translations between the conventional method and the Sim-Consec method by self-evaluation by the interpreters themselves, by the audience reception, and by analyzing the transcription of the translation. In this study, the audience was selected randomly for the evaluation. For each experimental session, the audience was separated into three groups. The eight-page questionnaire was then distributed to listeners during the experimental sessions, which revealed their overall impressions of the interpreter’s performances. The researchers also used output analysis to allow “triangulation of data” and adopt “a multi-method approach to interpreting quality assessment.”

As described above, Hamidi and Pöchhacker (2007) provided a very suitable framework for assessing the quality of interpretation via a novel method with solid experimental procedures. Their experiment design and feedback survey questionnaires for interpreters and audience were adapted for use in this study.

After extensive research, no studies were found on the subject of using automatic subtitling in aiding human interpreters when performing simultaneous interpretation. Some studies used the ASR technology in tandem with MT technology to create a simultaneous machine interpreting system that requires as little human intervention as possible. Others used an entirely different method in aiding human interpreters. Thus, this research was conducted to determine how automatic subtitling and machine-generated translation could aid interpreters to achieve more accuracy and fluency in simultaneous interpretation.

### 3 Objectives

This research aims to provide a method for reducing a load of interpreters during the interpretation process while increasing the quality of the translation, particularly in accuracy. This research also aims to compare the difference in the quality of English to Thai SI between automatic subtitling-aided interpretation, machine-translated subtitles-aided interpretation, and conventional interpretation method.

Thus, the research questions could be summarized as follows:

1. To what degree do the automatic subtitling and machine-translated subtitles help an interpreter in performing interpretation, when compared to the traditional method?
2. To what degree does the audience benefit from the methods as described?

### 4 Methodology

#### 4.1 Design

To answer the research questions, a comparison between the three interpreting methods is needed. To achieve this, an experiment where an interpreter interprets three different but comparable speeches in simultaneous mode was conducted, with the interpreter interpreting with the automatic subtitling, machine-translated subtitles, and no subtitling given.

Note that this research was conducted in the pilot study environment by the researcher before the experiment on the participants was conducted.

#### 4.2 Participants and Material

Four interpreting students (Int1, Int2, Int3, and Int4) participated in this experiment. The participants, three men and one woman, have had over a year of experience with interpretation and a number of interpretation methods, with some of them having had experience in using subtitles as means to aid SI. Note that Int2 has English as A language, and Thai as B language; the rest of the participants have Thai as A language, and English as B language. This is summarized in Table 1 below.

<b>Participants</b>	<b>Gender</b>	<b>A language</b>	<b>B language</b>
<b>Int1</b>	Male	Thai	English
<b>Int2</b>	Male	English	Thai
<b>Int3</b>	Male	Thai	English
<b>Int4</b>	Female	Thai	English

*Table 1: Participants' gender and A/B languages*

The three source videos for the experiment were excerpts from the same parent video, which contained an English speech delivered by an English native speaker and dealt with computer algorithms concerning science communication. The three excerpts were comparable in terms of content and language use. They were also comparable in terms of articulation speed, proper names, and figures. Each excerpt has a relatively comparable length (7:43, 7:54, and 7:54 minutes long, respectively).

To provide a glossary for the participants, a selection of keywords and phrases was compiled from the transcripts of those three excerpts combined, with the help of term extraction software (Term Extraction 1.1 by FiveFilters.org).

For the interpretation, the source videos were played back from the YouTube website from a computer. To provide the automatic subtitling and machine-translated subtitles, the option for English auto-generated subtitle and Thai auto-translation of the said subtitle were turned on respectively, generating a block of text at the bottom of the video of reasonable font size.

Regarding the recording device for each rendition of interpretation, each participant was free to use their recording device that they were familiar with (for this experiment, the participants used some different models of iPhone, ranging from iPhone 11 to iPhone 13). To provide the recording for the experiment, each participant was required to start and stop recording on their respective device on their own; each participant was given a brief period to test their recording device. Also, for additional data, Int3 and Int4 were asked to provide a video recording of their faces during their interpretation sessions. These video recordings provided data for analysis of interpreters' body language, facial expression, intonation and hesitation, which could shed light on their cognitive efforts during interpretation with and without the aid of subtitling.

The participating audience was a sample of eight persons aged from 23 to 51, with every participating audience member having graduated, or is expected to within a year of the date of the survey, with a bachelor's degree at minimum. All of the said audience had virtually neither experience nor knowledge in interpretation. The audience was randomly selected to listen to a number of participants' voice recording, ranging from one to four participants', taking audience members' time constraints into consideration.

The questionnaire for the audience consisted of four sections, with three of those sections including three questions to assess comprehension (one short-answer recall and two multiple choice questions); six seven-point Likert scales (7 labeled as "strongly agree," and 1 labeled as "strongly disagree") to grade the following qualities: comprehensibility, fluency, coherence, quality of expression, emphasis and intonation, and confidence and professionalism; and one seven-point Likert scale (7 labeled by labeled as "strongly desirable," and 1 labeled as "strongly undesirable") to rate the desirability of the participants' interpretation renditions. The other section asked the audience to provide a ranking of the three renditions of the interpretation from their most favorite to least favorite, and a short reason for such ranking. The questionnaire used in this study is shown in Appendix A.

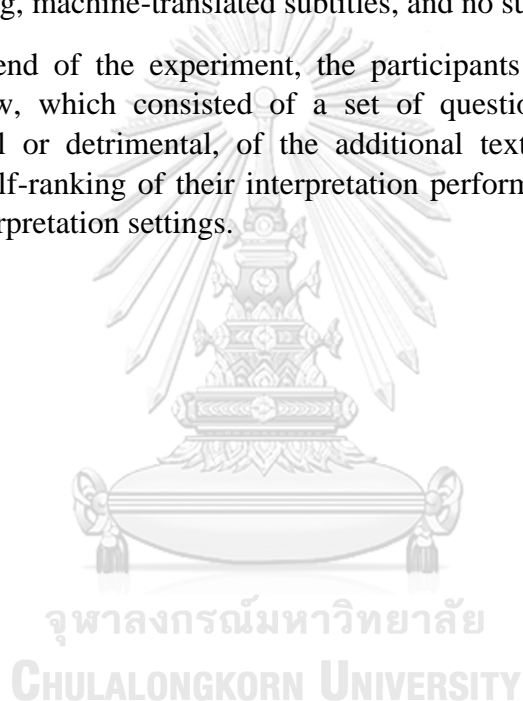
### **4.3 Procedure**

For the experiment, each participant interpreted three different speeches. At the beginning of the experiment, all four participants were given the glossary of the speeches to study for 5 – 6 minutes before the first interpretation session, in which the

participants were asked to interpret the speech from English to Thai, with the video (with or without subtitling) in full view. The end of the first interpretation session was followed by a 5-minute break before the beginning of the second interpretation session, which in turn was followed by another 5-minute break before the beginning of the third interpretation session.

Thus, to decrease the effect of familiarity with the content of the speech itself which could affect the results, the video with no subtitling was given last. In addition, to accurately compare the effects of automatic subtitling against the machine-translated subtitles, two schemes of interpretation orders were devised. Int1 and Int3 were given the videos in this order: machine-translated subtitles, automatic subtitling, and no subtitling. On the other hand, Int2 and Int4 were given the videos in this order: automatic subtitling, machine-translated subtitles, and no subtitling.

After the end of the experiment, the participants were asked to provide a feedback interview, which consisted of a set of questions regarding the effects, whether beneficial or detrimental, of the additional text below each video, self-assessment and self-ranking of their interpretation performance, and suggestions for use in the real interpretation settings.



## **5 Results**

### **5.1 Interpreters' Feedback**

All four participants were able to adapt to the automatic subtitling in the video for their interpretation without any major problems, regardless of whether they actually used the subtitling in their interpretation. When asked to rank their performance, Int3 and Int4 have ranked the session with English subtitling as their best, followed by the session with no subtitling, with the session with Thai subtitling as the last. On the other hand, Int1 and Int2 have ranked the session with Thai subtitling as their best, followed by the session with English subtitling, with the session with no subtitling as the last. It should be noted, however, that Int2 has also stated that the reason for such ranking is because the first session provided a “warm-up” session, resulting in a perceived superior performance in the second session.

The participants, in general, agreed that additional automatic subtitling in the source language (English) proved beneficial in assisting SI, thanks to the subtitling being a relatively accurate transcript of what the speaker has said. They agreed that the automatic subtitling allowed them to have a reference of the speech, with one participant stating that the subtitling provides a “lifeline” for proper names and numbers, while another has stated that the subtitling provided a method to “cross-check” what they have been thinking and what they were about to speak.

On the other hand, the participants have all stated that machine-translated subtitles (Thai subtitling) were ineffective, and in some cases even counterproductive, in the interpretation process. This was due to the relatively low quality and progress of current MT technology, which the participants have also pointed out. Two participants have also stated that machine-translated subtitles were considered a distraction when performing the interpretation, as the translation provided on the screen and what they have been contemplating often did not match and they have to actively choose which translation to articulate.

Barring minor issues, such as the subtitling having a small delay (less than one second) between the speaker and the subtitling itself, all of the participants agreed that they could imagine using automatic subtitling to aid their SI if provided in the real setting. This is exemplified by all participants who have stated that they can perceive that “something is missing” in the interpretation session without subtitling, especially when compared to the interpretation session with automatic subtitling. Other suggestions for using this method in real life setting include improving the ASR technology and transcribing numbers as digits as opposed to text (e.g., 200,000 instead of two hundred thousand).

### **5.2 Audience Response**

#### **5.2.1 Audience's Comprehension Quizzes Responses**

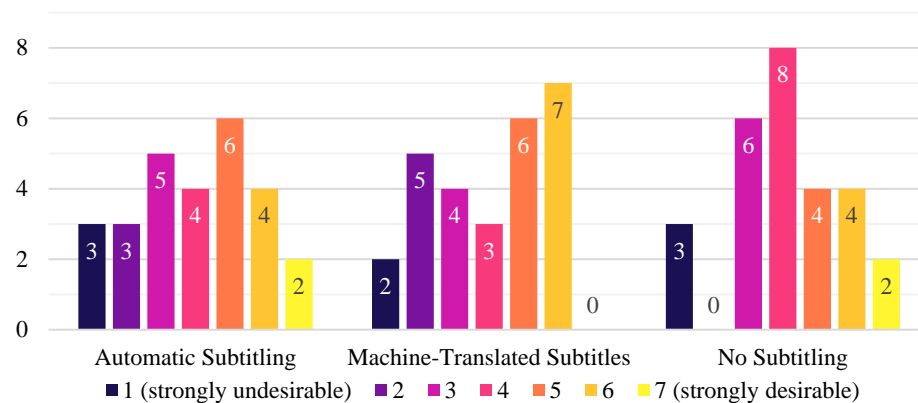
The responses to the comprehension quizzes gave a basic metric for measuring comprehension of the speech materials. The renditions using automatic subtitling allowed for a slightly improved comprehension, as opposed to those using no



subtitling (59.26% and 55.56% of correct answers given for automatic subtitling and no subtitling, respectively). In contrast, machine-translated subtitles could be described as being detrimental to the audience's comprehension (40.74% and 55.56% of correct answers given for machine-translated subtitles and no subtitling, respectively).

### 5.2.2 Audience's Ratings

The audience rated the interpretation renditions of the sessions with automatic subtitling and machine-translated subtitles to be slightly more favorable in contrast to the interpretation rendition of the session with no subtitling. As shown in Figure 1 below, the majority of the audience members (14 and 16 out of 27 responses for automatic subtitling and machine-translated subtitles respectively) rated the two former renditions with the rating of 4 – 6, as opposed to the rating of 3 – 4 for the renditions of the sessions with no subtitling (14 out of 27 responses).

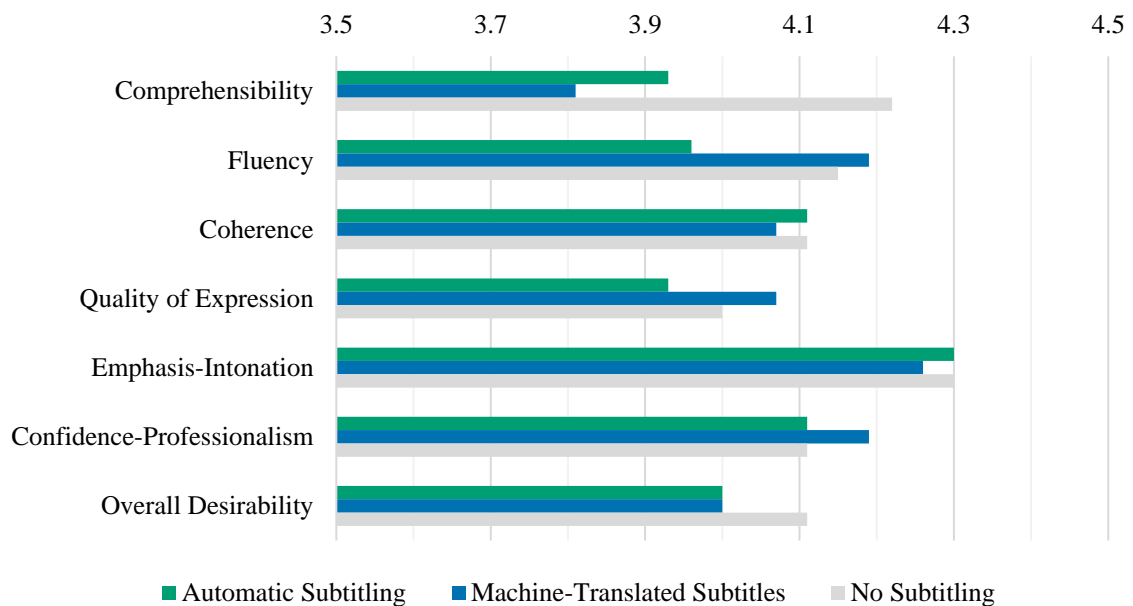


**Figure 1:** Number of responses for overall desirability

However, the average rating for each method is 4.00, 4.00, and 4.11, for automatic subtitling, machine-translated subtitles, and no subtitling respectively. This is illustrated in Table 2 and Figure 2 below, the former of which also illustrates that the median and the mode of the ratings have virtually no differences, and thus the slightly higher preference could be described as statistically insignificant.

		Comprehensibility	Fluency	Coherence	Quality of Expression	Emphasis-Intonation	Confidence-Professionalism	Overall Impression
Mean	Automatic Subtitling	3.93	3.96	4.11	3.93	4.30	4.11	<b>4.00</b>
	Machine-Translated Subtitles	3.81	4.19	4.07	4.07	4.26	4.19	<b>4.00</b>
	No Subtitling	4.22	4.15	4.11	4.00	4.30	4.11	<b>4.11</b>
Median	Automatic Subtitling	4	4	4	4	5	4	<b>4</b>
	Machine-Translated Subtitles	4	4	4	4	4	4	<b>4</b>
	No Subtitling	4	4	4	4	4	4	<b>4</b>
Mode	Automatic Subtitling	5	5	5	3	5	5	<b>5</b>
	Machine-Translated Subtitles	4	4	3	5	4	6	<b>6</b>
	No Subtitling	5	3	4	4	4	6	<b>4</b>

**Table 2:** Number of responses for overall desirability



**Figure 2:** Average value of audience's rating for each metric

Similar to the desirability ratings, the majority of the audience members also rated the quality of the sessions with automatic subtitling and machine-translated subtitles to be slightly more fluent, more coherent, easier to follow, and seemed more confident and professional. The renditions with automatic subtitling are also rated slightly higher by the majority of the audience regarding emphasis and intonation. These ratings are illustrated by the heat map in Figure 3. It should still be emphasized, however, that when considered jointly with the statistical values as shown in Table 1, this slight benefit could be described as statistically insignificant.



## 6 Conclusion

The results of the experiments show that, while there are some improvements to be made regarding the ASR technology, all four participants considered the automatic subtitling (generated from the ASR) method easily adoptable and have significant benefits, such as allowing the interpreters to have a method to cross-check the correctness of their translation and providing a live transcript for the interpreter to decrease the cognitive effort. With minor adaptations, such as transcribing numbers as digits and reducing the time delay between the speaker and the appearing subtitling, the participants found this method to be viable in a real-life setting. In contrast, the method utilizing machine-translated subtitles (Thai subtitling) was found to be ineffective and detrimental in the interpretation process, as to be expected due to the comparably inferior quality and progress of current MT technology. It should be noted, however, that these results might be true for Thai language only; varying results are possible for languages with different advancement in MT technology.

On the other hand, the audience members have rated the renditions of the speeches with subtitling, whether it was the automatic subtitling or machine-translated subtitles to be slightly more favorable than that without so. The audience has also found the rendition of the speeches with automatic subtitling to have a slight positive effect on the delivery, specifically regarding fluency, coherence, and quality of expression. This is consistent with participants' feedbacks which state that automatic subtitling has aided their interpretation process, as opposed to machine-translated subtitles. In spite of this, as shown in previous section, this slight advantage could be considered statistically insignificant, and further studies could be done to investigate this.

To the researcher's knowledge, this research provides one of the first pieces of information regarding using these tools in aiding SI for this language pair. However, it is clear that further research is needed for a more thorough analysis of the interpretation quality. Different input speed, language combinations, subtitle display options, and some other variables are the areas that this research that could potentially be expanded. Further analysis regarding accuracy and interpretation qualities, such as prosody, fluency, quality of expression, and psychoanalysis of body language during the interpretation session could also be studied.

It should also be mentioned that the implications of this method need to be assessed, despite this technological development in aiding the interpreters. Considering that as of 2022, the automatic live MT subtitling system is neither widely commercially available nor freely accessible, especially for Thai language, there might not be readily available tools or technology to employ this method to use in real-life setting.

## Appendix A: Audience's Questionnaire Sample

Note that this questionnaire sample is translated from Thai to English.

### Comprehension Quiz

Please answer questions (1) – (3) based *ONLY* on your understanding from the clip you have listened from the latest interpreter.

- (1) What does this clip talk about in general?

You may write your answer of any length you wish.

.....  
 .....  
 .....  
 .....

- (2) What imaginary dilemma did the speaker mention in the beginning?

- Algorithm's disagreement with you regarding your fundamental values  
 UK withdrawal from the European Union (Brexit) due to algorithm  
 The use of algorithms to communicate complex scientific concepts to the masses  
 Tech companies' exploitation, using algorithm as a tool

- (3) What video topic that the speaker does NOT usually make?

- Education  
 Interesting things around the world  
 Computer science  
 Linguistics

### Interpreter's Satisfaction Survey

Please truthfully assess your satisfaction with the interpreter *IN THIS CLIP* by the following criteria.

The satisfaction levels of numbers 1 – 7 are as follows:

1. Strongly disagree / Strongly dissatisfied
2. Moderately disagree / Moderately dissatisfied
3. Slightly disagree / Slightly dissatisfied
4. Neither agree nor disagree / neither satisfied nor dissatisfied (neutral)
5. Slightly agree / Slightly satisfied
6. Moderately agree / Moderately satisfied
7. Strongly agree / Strongly satisfied

Note that the satisfaction level according to numbers 1 – 7 has no relations to question numbers (1) – (7).



## REFERENCES

- Alves, L. M. (2021). *Microsoft Translator et le jargon vidéoludique: Evaluation de l'impact de la traduction automatique sur une conversation entre joueurs de jeux vidéo* [Master's Thesis, Université de Genève]. <https://archive-ouverte.unige.ch/unige:155488>
- Bansal, S., Kamper, H., Lopez, A., & Goldwater, S. (2017). *Towards speech-to-text translation without speech recognition*. arXiv. Retrieved 17 July 2022 from <https://arxiv.org/abs/1702.03856>
- Berard, A., Pietquin, O., Servan, C., & Besacier, L. (2016). *Listen and Translate: A Proof of Concept for End-to-End Speech-to-Text Translation*. arXiv. Retrieved 17 July 2022 from <https://arxiv.org/abs/1612.01744>
- Bühler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua*, 5(4), 231-236. <https://doi.org/10.1515/mult.1986.5.4.231>
- Defrancq, B., & Fantinuoli, C. (2020). Automatic speech recognition in the booth Assessment of system performance, interpreters' performances and interactions in the context of numbers. *Target*, 33(1), 73-102. <https://doi.org/10.1075/target.19166.def>
- Fügen, C., & Kolss, M. (2007). The influence of utterance chunking on machine translation performance. Interspeech 2007, 8th Annual Conference of the International Speech Communication Association, Antwerp, Belgium.
- Gaber, M., & Corpas-Pastor, G. (2021). *Automatic speech recognition systems for interpreters: Spoken corpora exploitation by interpreter trainers and trainees*. University of Malaga. Retrieved 2 June 2022 from <https://riuma.uma.es/xmlui/handle/10630/23270>
- Gaber, M., Corpas-Pastor, G., & Omer, A. (2020). La tecnología habla-texto como herramienta de documentación para intérpretes: Nuevo método para compilar un corpus ad hoc y extraer terminología a partir de discursos orales en vídeo. *TRANS – Revista De Traductología*(24), 263-281. <https://doi.org/10.24310/TRANS.2020.v0i24.7876>
- Hamidi, M., & Pöschhacker, F. (2007). Simultaneous Consecutive Interpreting: A New Technique Put to the Test. *Meta*, 52(2), 276-289. <https://doi.org/10.7202/016070ar>
- Jameson, J., & Pengelly, G. (2019). *Subtitling videos accurately and easily with CaptionHub and AWS*. Amazon Web Services, Inc. Retrieved 2 June 2022 from <https://aws.amazon.com/blogs/machine-learning/subtitling-videos-accurately-and-easily-with-captionhub-and-aws/>
- Laksana, N., & Putri, S. (2018). An Error Types Analysis on YouTube Indonesian-English Auto-Translation in Kok Bisa? Channel. *Journal of Language and Literature*, 18(1), 76-81. <https://doi.org/10.24071/joll.v18i1.1058>
- Manurung, B. M. (2021). *YouTube closed captions auto-translate performance in the Indonesian translation of what if's video entitled "what if we dumped our trash into volcanoes?"* [Undergraduate Thesis, Sanata Dharma University]. <https://repository.usd.ac.id/39597/>
- Microsoft Corporation. (n.d.). *Live conversations - Education - Microsoft Translator*. Microsoft Corporation. Retrieved 2 June 2022 from

- <https://www.microsoft.com/en-us/translator/education/live-conversations/>
- Pewnim, P. (2017). *Assessing the quality of English-to-Thai machine interpreting: translating Google Translate* [Master's Special Research, Chulalongkorn University]. <https://doi.org/10.14457/CU.the.2017.964>
- Rodríguez, S., Gretter, R., Matassoni, M., Alonso, Á., Corcho, O., Rico, M., & Daniele, F. (2021). SmarTerp: A CAI System to Support Simultaneous Interpreters in Real-Time. Proceedings of the Translation and Interpreting Technology Online Conference, Antwerp, Belgium.
- Sawaf, H. (2012). Automatic Speech Recognition & Hybrid MT for HQ Closed-Captioning & Subtitling for Video Broadcast. 10th Conference of the Association for Machine Translation in the Americas: Commercial MT User Program, San Diego, CA.
- Wu, Y., & Wang, Y. (2018). An Exploration of English-Chinese AI Interpreting Difficulties from a Perspective of Practical Application. *DEStech Transactions on Social Science, Education and Human Science*. <https://doi.org/10.12783/dtssehs/ichae2018/25626>
- Zhao, Y., Zhang, X., Hu, R.-S., Xue, J., Li, X., Che, L., Hu, R., & Schopp, L. (2006). An Automatic Captioning System for Telemedicine. 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, France.





จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

## **VITA**

**NAME** NUTDANAI SORNCHAI

**DATE OF BIRTH** 28 October 1997

**PLACE OF BIRTH** Bangkok, Thailand

**INSTITUTIONS  
ATTENDED** Chulalongkorn University (B.Eng.)

**HOME ADDRESS** 1192 Nakhon Chai Si Road  
Dusit, Bangkok 10300  
THAILAND



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**