International Journal of Multidisciplinary Research and Growth Evaluation.

On the application of speech recognition technology in interpretation

Ji Lingzhu
Taiyuan Normal University, City of Jinzhong, 030619, Shanxi, P. R. China

* Corresponding Author: **Ji Lingzhu**

Article Info

ISSN (online): 2582-7138
Volume: 04
Issue: 01
January-February 2023
Received: 01-01-2023;
Accepted: 17-01-2023
Page No: 317-320

Abstract
The rapid development of information science and technology makes it possible for interpretation from one language to another to expand in both time and space. The artificial neural network (ANN) based on HMM model works well in speech recognition. The technology can assist interpreters in preparation, recording and searching and information organization respectively before, during and after interpretation stage. Together with the advent of cutting-edge technologies such as big data, virtual reality and artificial intelligence, the application of information technology (IT) to interpreting has undergone a shift from the mode of external assistance to that of internal penetration. In the future development of interpreting techniques will be more advanced and the relationship between speech recognition technology and interpreters will be inseparable.

Keywords: speech recognition technology, interpreting technique, interpreters

1. Introduction

Since it is both time and energy consuming for human beings to translate or interpret from one language to another, computer-assisted translation or interpretation has been more and more widely used. Machine translation has been used for over 80 years mainly in the forms of machine translation and computer aided translation. As speech recognition is introduced, the assisting function of machines is more intensified and diversified, then computer-aided interpretation comes into being. The development of artificial intelligence has brought natural language processing technologies such as speech recognition, text conversion, semantic analysis, emotion recognition and speech synthesis into the family of machine translation. IT giants Google and Microsoft developed machine translation software based on the neural networks. *Mr. Tencent*, the software production of China's Tencent in 2018, provided interpretation services for Boao Forum. Although the performances of the production was not perfect at that time, it proved that the degree of participation of technology in interpreting practice was getting higher and higher, and predicted for sure that in the future technology would always play a role either implicitly or explicitly in interpreting practice. The present machine interpretation is devided into three steps: speech recognition, machine translation and speech synthesis. However, the development of the three major steps is unbalanced and not as expected. This paper attempts to use this technology as an auxiliary tool for interpretation to discuss its role before, during and after the interpreting task, so as to further study the relationship between the human interpreters and technology.

2. Working principle of speech recognition technology

Speech recognition came into being about twenty years ago. It is now widely used in many areas in our everyday life, such as communication, navigation, various kinds of service and so on. The working mode of speech recognition is as follows: computers or other devices with apps of speech recognition function first “listen to” human beings, then they extract the meanings of the sound signals, and finally the voice signals are converted to relevant text versions or commands. Speech recognition consists of feature extraction, pattern matching and model training. Automatic Speech Recognition (ASR) is the first step in the sequence of speech recognition, machine translation and speech synthesis, and it is also the best developed step. This cross discipline has been developed for more than half a century.

Hidden Markov Model (HMM), as the core technology, has kept upgrading itself. It is a prediction model divided by time points. Based on the given conditions at a certain time point t, the calculation of some parameters and mathematical models is introduced to predict the probability at the time point of t+1. Since language is linear, components such as the attributive clause and objective clause in English have their fixed positions in the sentences. The fixed collocations and structures in a language can also be used as set parameters to carry out the association and calculation in the Hidden Marcov Model in order to improve the efficiency of sound recognition. In the 1980s, the artificial neural network (ANN) algorithm based on the HMM model played an important role. Machines with mathematical models were used to simulate working procedures of human nerves. Consequently they were able to learn independently through repeated training. Apple and IBM took the lead to invest in the research and development of speech recognition technology. Although China started quite early in the research of machine translation (1950s), there was no significant progress in the first 20 years. After awakening in the middle of 1970s, it has made quite a lot of contributions. The speech recognition technology is doing well presently in machine interpretation, however, it is still in its infancy stage. For interpreters, speech recognition technology can play a more important role as a kind of computer-aided technology in all the three stages in interpretation. The present pattern of human and computer interaction will also shed light on the relationship between human interpreters and technology in the future.

3. The function of speech recognition technology

Interpretation technology mainly provides basic technical support for human interpreting activities; it aims to reduce the workload of human interpreters in the process of pre-interpretation, while-interpretation and post-interpretation by technical means, and improve their work efficiency and quality; Machine interpretation technology is used to complete the interpretation task in a certain context through the comprehensive application of speech recognition, machine translation, speech synthesis and other technologies, on the basis of inherent corpus of human interpreters.

3.1 Pre-interpretation

For many interpreters, the portable electronic devices equipped with speech recognition technology are of great help, which is also the development tendency of the technology. The popularly used two by the Chinese interpreters are Apple's Siri system from the U.S.A. and IFLYTEK, which specializes in Chinese. Siri is more like an artificial intelligence system, which can not only understand various languages, but also react based on the actual situation. To realize human –computer interaction, Siri first extracts acoustic features through the speaker and earpiece of the mobile phone, and then uses a powerful online search engine to meet the needs of users. IFLYTEK specializes in speech technology, which mainly includes speech synthesis and speech recognition. Besides, siri and I FLYEK both include voice coding, voice conversion, spoken language evaluation, voice denoising and enhancement. ® The most fundamental need of interpreters is background knowledge of the target speech. Interpretation, especially in scientific and technological meetings, poses a special challenge for interpreters in language and background knowledge. Interpreters must acquire the knowledge before the meeting and familiarize themselves with it until the degree of automatic extracting during interpreting. In meetings, interpreters serve professionals in source and target languages, while few experts in this field have no relevant background knowledge.

Therefore, there is a knowledge gap between interpreters and the professional participants of the meetings. This gap involves both language knowledge (terminology) and specific expertise (speaker information, context, etc.). To eliminate this gap, interpreters need to make preparations in advance. Most of this knowledge can be acquired by translators while they are translating, but interpreters need to acquire it before the meeting because they have to work under the time and occasion pressures when interpreting. Pre-preparation is essential to overcome many difficulties in the process of interpretation, and insufficient preparation may be the cause of many errors. Therefore, many recent studies focus on the pre-conference preparation stage, especially on the way to define and acquire the knowledge required for good performance

With the help of speech recognition technology and artificial intelligence, interpreters can quickly find the relevant materials they need for reference, convert the conference speech materials into text, or the texts into speech materials, so that interpreters' attention splitting will be more reasonable. They can be liberated from the time-consuming and tedious investigation and research work in the preparatory stage with improved accuracy and quality of information and language.

3.2 While-interpreting

During the while-interpreting stage, speech recognition generates text expression through ways of speech feature extraction, acoustic model calculation, language model calculation etc. which helps to solve the problem of information receiving failure or inefficiency caused by noise interference, fast speech speed, speech variants and other factors. Moreover, after speech recognition technology converts the speaker's words into text, the difficulty of interpretation is greatly reduced from consecutive interpretation and simultaneous interpreting to visual translation. Or simply just correct the speech recognition mistakes. For example, I said "She is a well-known paper-cutting folk artist in the southern part of Shanxi province, good at cutting opera characters" in Mandarin Chinese to an IFLYTEK speech recognizing translating App. After two seconds, the result "She is a well-known paper-cut folk artist in the southern part of Shanxi province, good at cooking opera characters" was both on the screen in the forms of text and audio, but the word "cutting" was replaced by "cooking". The App did not know that opera characters could not be cooked. However, that spared most of the mental work from the interpreters.

In November 2012, Microsoft demonstrated their significant progress in computer speech recognition. In the demonstration video, he said eight sentences in English to the newly developed speech recognition translation system. The system imitated his pronunciation and intonation feature, and two seconds later "spoke" the translation in Mandarin. The translated speech in the target language is high in quality as well as with the pronouncing features of the source language speaker. The automatic transcription system is able to recognize the speaker's voice, more effectively generate text and audio records of discussions, interviews and meetings, etc. Then the manual correction of recognized texts and the review of the transcripts will be more complete and reliable. It opened up a wide range of possible applications and areas

for further development since it put forward a method of translating with lexicalized information. Phrase based translation system can determine the starting point of translation in a sentence, with shorter delay. It will be effective for reducing the start time and processing time of machine translation, and can be used in various ways in the future.

However, it should also be noted that, especially in simultaneous interpreting, the words, phrases, paragraphs and even text recognized by speech recognition technology may affect the interpreters' attention allocation. When interpreting, the interpreters should focus on both listening for the information in the source language and monitoring the output quality in the target language. Speech recognition technology can greatly reduce the pressure on short-term memory for interpreters who are used to relying on notes. However, for interpreters who focus on listening, it may disrupt his/her attention distribution and affect the information accuracy and the output quality in the target language.

Meanwhile, we have noticed that speech recognition technology is not equally valuable in Chinese to English interpretation (translation) and English to Chinese interpretation (translation). Interpreters found that in C-E translation it functions much more effectively than in E-C translation. Therefore, speech recognition technology performs better when the mother tongue is interpreted into the target language. Maybe the fundamental reason for this is that interpreters process their own mother tongue much faster than the target language. However the technology can at least reduce the interpreters' burden of note-taking and note-decoding. In the future, the interpreters should strengthen their cooperation with the computers in order to have high quality interpretation service.

3.3 Post interpreting

After the interpretation, speech recognition technology can directly convert both the source and the target language recordings into text and store them into the corpus, which can not only be used as backup resources for the preparation of the similar topics during the pre-interpreting stage, but also liberate interpreters from repetitive labor, so as to provide more efficient interpretation services.

4. Summary and implications

Despite its present advantages and application, speech recognition technology is still far from perfection. First, the self-adaption ability is in need of improvement. Before using the speech recognition device, users need to read a lot of provided sentences or phrases in order to make the device get used to the speaker's voice. The training process is very consuming and often needs to be reconducted with different speakers. All speech recognition devices need this process to get used to the features of the speaker's voice. Even when the same speaker speaking under different circumstances, the system needs to be trained for re-adaption. Also the background noise affects its performance. So the developers and researcher of the speech recognition technology still have a lot of work to do to improve its self-adaption capability so that it can perform well in different environments. Secondly, speakers' accents vary due to different birth places or educational backgrounds etc. In China, people from different places speak in very different dialects. Dealing with dialects of the same language is very difficult for human beings, let alone the speech recognition technology. Besides, with the improvement of globalization, many speakers have the experience of overseas education, they are used to speaking in two languages. Then the system might be confused and the recognition output will not be satisfactory.

Since the 20th century, interpretation technology has been playing a role in interpretation more or less invisibly or dominantly. The wide application of artificial intelligence has given interpretation new vitality and spawned a variety of new professional forms of interpretation, such as telephone interpretation, television interpretation, video conference interpretation and artificial intelligence interpretation, and also brought about professional technical tools such as smart pens, terminologies, video conference operating systems, etc., to the interpretation profession. It has also triggered many changes in the working methods of the interpretation profession (such as remote working mode, online working mode), work content (such as computer-assisted human translation based on speech recognition).

The speech recognition technology home and abroad upgrades from day to day Mary Meeker's Internet Trend Report 2017 shows that as of May 2017, the accuracy rate of speech recognition (English) based on Google Machine Learning System has reached 95%, which is equivalent to the accuracy rate of human beings. The accuracy rate of Nuance Dragon Naturally Speaking developed by Dragon Systems of Newton, Massachusetts reached over 95%. Strong accents and fast speed are major challenges for speech recognition. In China, Sogou, Baidu and I FLYTEK respectively held press conferences in 2016 to announce their speech recognition accuracy. On November 21, Sogou's voice team announced that the accuracy rate of their voice recognition had reached 97%, supporting dictation of speech speed up to 400 words per second. Baidu announced on November 22 that it had developed four speech recognition technologies, including emotion synthesis, far-field scheme, wake-up phase II and long speech scheme. Its recognition accuracy of speech in "quiet environment" reached 97%. IFLYTEK quoted Luo Yonghao's demonstration data at the Hammer Conference in September at the conference on November 23. Its success rate of speech input recognition reached 97%, and even the accuracy rate of offline recognition reached 95%. Since 2013, as the research on deep learning has moved a big step forward, machine translation (NMT) based on artificial neural networks has begun to rise. Its technical core is a deep neural network with a large number of nodes (neurons), which can automatically learn from the corpus. After the sentence from one language is vectored, it is transmitted layer by layer in the network and transformed into a form of expression that can be "understood" by the computer. Then, it is generated into a translation of another language through multi-level complex conduction operations. The development of automatic speech recognition and machine translation will bring about possiblity for a revolution in the interpreting profession.

The outbreak of the covid-19 epidemic in late 2019 and early 2020, the interpreting demand has mainly become remote video interpretation. Remote video interpretation platforms have also emerged one after the other, allowing interpreters to safely provide interpretation services at home and familiar with the advantages and disadvantages of various remote interpretation platforms. The interpreters will be able to introduce the most suitable platform to the customers. It can be seen that although people depend on the development of

technology, they are still masters deciding the development tendency and the application. Intelligent machine interpretation is an integrated application of human interpretation skills.

With the development of interpretation technology, interpreters will inevitably come into contact with technical products, which will change the cognitive and operational process of interpretation, the model of interpreter's attention allocation and even the form of interpretation. In the case of machine assistance, the interpreter changes from merely listening and outputting to reading visual information in the process of listening and outputting. On one hand, it alleviates the memory pressure, on the other hand, it poses new challenges to the coordination ability. According to Jill's mental power distribution model of consecutive interpretation, during the first stage of consecutive interpretation, the interpreters focus on listening and analysis +notes+short-term memory+coordination, and during the second stage, memory+notes reading+communication. After the interpreters acquired the key information from the machine aided recording, their attention distribution mode may change to:

Stage I =listening and analysis+note taking process, reducing the notes of the parts done by the machine and hearing monitoring the coordination among the notes, the parts done by machine and the short term memory,

Stage II =memory+note reading and machine recording + target language output.

In this way, the form of interpretation may be more digital, forming human-computer cooperation. Remote interpretation may become wide –spread due to technological advance and finally change the definition and form of interpretation.

Now speech recognition systems operate at various levels: words, phrases, sentences etc., and even multimodal texts. However, in interpreting and translating practice, there are still mistakes to be corrected by human beings. For instance, it is found that for speeches intensive with numbers, proper nouns and terms, and those spoken at a fast speed, computer-aided translation can effectively improve the interpretation quality by relieving the memory burden of interpreters, especially the accuracy of the above-mentioned types of information. The auxiliary effect is proportional to the intensity of such information and the speed of the source speech. For speeches with low intensity of such information, the role of technological assistance is limited, and the quality of the target language output still depends on the interpreter's effective listening ability in source language, logical reasoning ability and knowledge reserve in the corresponding fields. Human beings still play a necessary role in high-level thinking, emotional cognition, cultural recognition, and machine monitoring. Therefore, even in the machine dominated intelligent interpretation mode, human intelligence is unmatched and human value is irreplaceable.

References

1. Deng Juntao, Zhong Weihe. Integration of Information Technology and Interpretation Teaching: Levels, Mechanisms and Trends [J] Chinese Translation. 2019;40(6):88-95+192.
2. Hu Kaibao, Tian Xujun. MTI talent training in the context of language intelligence: challenges, strategies and prospects [J]. Foreign Language World. 2020;(2):59-64.
3. Kong Ying. The use of speech recognition technology in interpreting practice [J]. Science and Technology Vision; c2020.
4. Li Hanyi. On the Assisting Role of Speech Recognition Technology in Consecutive Interpreting, a Case Study of Nuance Dragon Naturally Speaking [D]. Sichuan International Studies University; c2016.
5. Shen Dan. Simulation Research on Speech Recognition Assisted Simultaneous Interpretation [D]. Xiamen University; c2014.
6. Sun Maosong, Zhou Jianshe. The Development Strategy for Natural Language Processing Research Inspired from a Historical View on Machine Translation [J]. Language Strategy Research; c2016;6.
7. Sun Yuxin. The Supporting Role of Speech Recognition and Machine Translation in Interpretation [D]. Shandong Normal University; c2013.
8. Tian Xinyu, Li Junhui. Analysis of the effects of speech recognition errors on translation performance. [J]. Journal of Xiamen University (Natural Science Edition); c2022.
9. Wang Huashu, Liu Shijie. Research on the Translation Technology shift in the Era of Artificial intelligence. [J]. Foreign Language Education; 2021;42(5).
10. Wang Huashu, Yang Chengshu. The Development of Interpretation Technology in the Age of Artificial Intelligence: Concepts, Impacts and Trends [J]. Chinese Translators Journal. 2019;40(6):69-79+191-192.
11. Zheng Ze. An Empirical Study of Computer-aided Consecutive Interpretation [D]. Beijing Foreign Studies University; c2018.