Apply Robot Recognize the Motivation of Human to Solve Problems Encountered by Translation Robots During the Translation Process

Xiaonan Qiao[1,*] and Yihan Zhou[2]
[1]School of Science and Technology, Hong Kong Metropolitan University,
Kowloon City District, Hong Kong, 999077, China
[2]School of Information and Intelligent Engineering, Zhejiang Wanli University, Ningbo, Zhejiang, 315100, China

Keywords: Simultaneous Interpretation, Human-Computer Interaction, Body Language.

Abstract: During the development of science and technology, interacting between different countries become the basis for development, due to the culture and language difference simultaneous interpretation becomes more and more popular and AI began to be applied to simultaneous interpretation. However, because the difference between the culture and rapid iteration of language, there are errors during the process of simultaneous interpretation including polysemy, emotional misjudgment and new words are incorrect and imperfect. So it is very important to improve translation accuracy using human-computer interaction. In this essay, the results of some studies are presented by collecting a large number of languages to expand the richness of language content, issues with slang and new buzzwords solved. Gathering the voice of dialogue to improve the flexibility and adaptability of translation, the machine performs advance voice adaptation training instead of humans to avoid word recognition errors and applying the body language during dialogue to improve the translation efficiency and translation accuracy be explained in detail.

1 INTRODUCTION

In today's world, the connections between people and countries are closer than ever, making clear communication across different languages and cultures very important. Simultaneous interpretation is a complex and demanding task that plays a key role in helping people communicate in international settings like diplomatic meetings, global business negotiations, and conferences. It is very important that reflecting on how social workers play multiple roles and explain in the context of language interpretation (Doering-White et al., 2019). These multiple roles can facilitate communication, However, despite advances in technology and teaching methods, achieving high accuracy in simultaneous interpretation still has some distance to achieve. Although the way of interacting is direct and accessible with the robotic platform, human language is still an important and one of the most natural ways to communicate because of its expressiveness and flexibility, the ability of a robot to correctly interpret users' commands is a key element in human-computer interaction (Andrea et al., 2022). During develop simultaneous interpretation human-computer interaction also be developed, AI is more widely used in this field, machines assist humans and make them more efficient.

One major issue is the natural differences in how people communicate. These differences go beyond just language and include cultural nuances, regional accents, and non-verbal cues like gestures and facial expressions. There are results suggest that language control may be more complex than previously thought, because different mechanisms used for different languages (Babcock & Vallesi, 2015). There is a research called SIMTXT which is considered that it is more difficult than SI of improvised speech. It is helpful for interpreters to have textual content when interpreting read discourse normally (Seeber et al., 2020). A study provides a set of systems based on various existing simultaneous interpretation systems. It combines frequency modules with other modules and peripheral circuits to complete systems used in all kinds of meetings. During speech recognition technology and the system's functional module structure. The results of the experiments show that the system has great recognition efficiency and accuracy in the simultaneous interpretation (Liu, 2021). The difficulty of simultaneous interpretation is further

Qiao, X. and Zhou, Y.
Apply Robot Recognize the Motivation of Human to Solve Problems Encountered by Translation Robots During the Translation Process.
DOI: 10.5220/0014290600004718

In *Proceedings of the 2nd International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2025)*, pages 5-9
ISBN: 978-989-758-792-4

increased because it happens in real time. Interpreters must listen, understand, and translate spoken words almost instantly, often with very little delay. This requires not only strong language skills but also a deep understanding of the cultural and situational context of the conversation, meaning sensitivity to cultural and situational factors. Through expanding the range of information that machines can collect, the quality of interpretation services can be improved significantly. On one hand, this helps improve our own quality of life. On the other hand, given the current global situation, it will help promote better understanding and cooperation in our increasingly connected world. Even though most of international leaders are command of English, there are still existing many ones who fail to immerse in communicating naturally. In the circumstances, the status of such a device is enhanced one step further.

In this essay, three methods are collected and integrated, they are collecting data about what people talk, collecting data about how is the voice, and collecting data about how people act. It analyzes how these three methods can improve efficiency and accuracy, how to use the robot to make these methods become faster and more practical. And this essay also mentions the disadvantages of these methods and give the way to solve these problems.

2 DATA COLLECTION

Figure 1 shows how simultaneous interpretation works (ResearchGate, 2021).

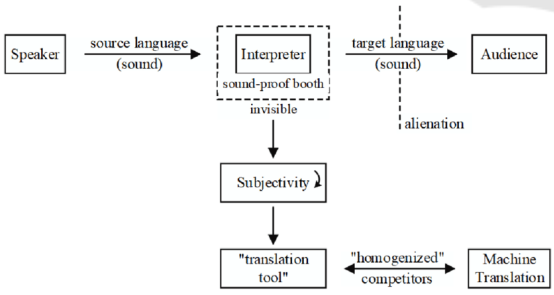

Figure 1: Simultaneous interpretation workflow diagram (ResearchGate, 2021).

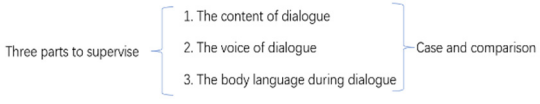

Figure 2: Three types of data stored (Picture credit: Original).

All the data collect below are stored in the device in the dotted frame of the flow chat above. And Figure 2 illustrates three kinds of data that ought to be supervised.

2.1 Collect Data About What People Talk (The Content of Dialogue)

The diversity of language expression habits poses a significant challenge in cross-language communication. Language users from different countries and regions often have unique ways of expression, slang, and culturally specific terms. There are experimental results showing a potential cause about simultaneous interpreting errors that is misled which come from the machine (Qin & Wang, 2020). To build an efficient multilingual speech recognition system, it is first necessary to comprehensively collect and analyze these language expression habits. This includes, but is not limited to, everyday language, industry terminology, local dialects, and popular slang. By constructing a large-scale corpus and applying deep learning algorithms, these linguistic features can be effectively captured and organized. Even though using a common language for scholarly communication offers some advantages, it might limit research diversity. Although artificial intelligence techniques have been used by machine translation in recent years, it still has some shortage If used by non-professionals, they need to improve their skills to make translation perfect and accuracy (Bowker, 2021). An accepted approach is to crowdsource platforms to collect language samples from different regions, combined with natural language processing technology for automatic classification and labeling. At the same time, linguistic experts should be introduced to manually verify and supplement the data to ensure its accuracy and comprehensiveness. Additionally, it is necessary to consider the dynamic changes in language expression habits and establish a continuous update mechanism to adapt to the evolution of language and the emergence of new vocabulary.

2.2 Collect Data About How Is the Voice (The Voice of Dialogue)

Accent diversity is another major challenge faced by speech recognition systems. Speakers from different regions and educational backgrounds may use the same vocabulary but with vastly different pronunciations. Imaging someone is watching an English movie without subtitles, it might be hard to catch what characters want to convey. One of the

factors could be their personalized speaking styles, with accents and pronunciations, Traditional speech recognition systems often struggle to accurately recognize various accents, leading to a decline in recognition rates. When people listen to accented speech, it will be difficult to process fluency in information processing (Sally & Tombs, 2020). To address this issue, an innovative accent adaptation technology can be put into practice. When the machine hears the first two sentences spoken by a person, it can record them and then listen to the subsequent speech of that person according to that accent, thereby improving accuracy. The core of this technology is to convert accent features from analog signals to digital format and store them in the system. Specifically, it first requires building a large-scale speech database containing multiple accents. Then, using deep neural network technology, the feature vectors of different accents are extracted and encoded into digital format. These digitized accent features can be stored in the system for subsequent speech recognition tasks. In practical applications, when the system encounters a new speaker, it can quickly match their accent features with the stored accent models, thereby adjusting recognition parameters to improve accuracy. Additionally, the system has self-learning capabilities, enabling it to continuously update and expand the accent database to adapt to more diverse accent features.

2.3 Collect Data About How People Act (The Body Language During Dialogue)

In human communication, body language plays an important role, as it can complement, emphasize, or even replace verbal information, it helps people in expressing themselves and understanding and explaining others (Turaev et al., 2023). Therefore, combining body language recognition with speech recognition can significantly enhance the accuracy of communication understanding. There is an innovative multimodal recognition method that can elevate overall recognition effectiveness by simultaneously analyzing speech and body language information. In specific implementation, the system needs to be equipped with visual sensors to capture different body movements, facial expressions, and gestures. Using computer vision technology and deep learning algorithms, the system can recognize and interpret these non-verbal cues. For example, a nod may indicate agreement, a frown may indicate confusion, and specific gestures may emphasize a certain point. While different impressions of personal traits are affected by behavior without voice, robots can study body languages such as regular movements, and diverse gestures to get them. These human–robot interactions architectures are starting to be used in various fields, Human emotions are an important part of this (Romeo et al., 2021). This information can be fused with speech recognition results to improve the accuracy of understanding the speaker's intent.

2.4 Case and Relative Advantages

Interpreters used an AI tool providing real-time terminology suggestions and context-aware prompts during sessions. Accuracy increased by 15% for technical debates (e.g., climate policy), as per 2021 internal evaluation. Errors in numbers and proper nouns dropped by 22%. That is just one of the great number of examples. In practical applications, multi-modal recognition system is particularly suitable for complex scenarios, such as multi-person meetings or noisy environments. By integrating speech and body language information, the system can more accurately identify the speaker, understand their intent, and even predict their subsequent actions. This not only improves the accuracy of speech recognition but also lays the foundation for more natural human-computer interaction.

3 DISCUSSION

Although the three practical methods mentioned earlier show effectiveness, limitations may still emerge during actual monitoring processes. To successfully advance the project, continuous attention to conversational content, tone variations, and physical gestures remains essential, along with consistent application of appropriate technological tools. Essentially, this represents an effort to enhance existing technical systems through their own capabilities. In such context, the improvement of real-time translation quality largely depends on the maturity of current sensing devices. As a consequence, before initiating large-scale data collection, the priority should be refining the accuracy of these detection instruments. This foundational upgrade will simultaneously improve both translation efficiency and result reliability. Based on existing devices, some senior sensors below can be put into practice to improve the sensation of the three parts. Initially, high-directionality microphones are able to focus on speakers' voice while suppressing background noise, relying on beamforming. Secondly, Optical Character Camera, as an advances visual sensor, can be used to

scan real time text from slides or speech timestamps. Finally, Utilizing Multi-Angle HD Cameras or Thermal/Infrared Sensors to capture gestures, facial expressions e.g. are beneficial for sensing body language. All in all, optical sensors track the position of speakers using infrared signals or changes in ambient light, activating voice recording devices. Visual sensors capture lip movements, facial expressions, and scene dynamics, using deep learning to analyze non-verbal cues and improve speech recognition accuracy in noisy environments. Audio sensors collect voice waveforms, apply noise-reduction algorithms to extract clear sound, and sync with optical and visual data. Together, they enhance performance: optics boost spatial awareness, visuals add contextual details, and audio provides core voice signals. Multimodal algorithms can improve real-time translation and noise resistance, especially in complex acoustic settings.

4 CONCLUSION

The paper summarizes how the accuracy of simultaneous interpretation can be improved through collect data from there parts. They are content, voice and action. And it reminds us that it is efficient to put our attention onto the sensors which are essential as input devices. Therefore, the paper comes to the discussion about that. The growing sophistication of artificial intelligence has fundamentally transformed global communication practices, with powerful translation technologies playing an increasingly vital role in bridging linguistic divides. However, the risk of meaning distortion in automated translation systems raises serious concerns, especially in such a politically charged international environment where diplomatic relations remain fragile. Research indicates that minor inaccuracies in conveying cultural context could inadvertently escalate tensions through misunderstandings in critical diplomatic or commercial exchanges. This evolving technological reality requires modern university students to develop practical integration of hardware design principles and software development skills. Current studies highlight promising opportunities in combining sensor technologies with language processing systems to advance real-time translation devices. For instance, developing sensors that detect vocal patterns, facial expressions, and physiological signals could significantly improve translation accuracy by better interpreting situational context. Moving forward, researchers should focus on creating collaborative systems between humans and artificial intelligence that balance advanced algorithms with human oversight. This strategy not only improves existing translation technologies but also creates new career directions for professionals working at the intersection of language technology and intelligent hardware development. Interdisciplinary efforts to refine adaptive different models and enhance sensor response speeds could help minimize communication errors while strengthening cross-cultural understanding in international exchanges.

AUTHORS CONTRIBUTION

All the authors contributed equally and their names were listed in alphabetical order.

REFERENCES

Andrea, V., Danilo, C., Emanuele, B., et al.: 'Grounded language interpretation of robotic commands through structured learning', Artificial Intelligence, 278(103181), 2022

Babcock, L., Vallesi, A.: ' Language control is not a one-size-fits-all languages process: evidence from simultaneous interpretation students and the n-2 repetition cost', Frontiers in psychology, 6, 1622, 2015

Bowker, L.: 'Promoting Linguistic Diversity and Inclusion: Incorporating Machine Translation Literacy into Information Literacy Instruction for Undergraduate Students', The International Journal of Information, Diversity, & Inclusion, 5(3), 127–151, 2021

Doering-White, J., Pinto, R. M., Bramble, R. M., Ibarra-Frayre, M.: 'Teaching Note—Critical Issues for Language Interpretation in Social Work Practice', Journal of Social Work Education, 56(2), 401–408, 2019

Liu, F.: 'Design of Chinese-English Wireless Simultaneous Interpretation System Based on Speech Recognition Technology', International Journal of Antennas and Propagation, 2021

Qin, Y., Wang, C.: 'Can Machine Translation Assist to Prepare for Simultaneous Interpretation?' International Journal of Emerging Technologies in Learning (Online), 15(16), 230-237, 2020

Romeo, M., Hernández García, D., et al.: 'Predicting apparent personality from body language: benchmarking deep learning architectures for adaptive social human–robot interaction', Advanced Robotics, 35(19), 1167–1179, 2021

Sally, R. H., Tombs, A. G.: 'When does service employee's accent matter? Examining the moderating effect of service type, service criticality and accent service congruence', European Journal of Marketing, 56(7), 1985-2013, 2020

Seeber, K. G., Keller, L., Hervais-Adelman, A.: 'When the ear leads the eye – the use of text during simultaneous interpretation', Language, Cognition and Neuroscience, 35(10), 1480–1494, 2020

The Influence of Speech Translation Technology on Interpreter's Career Prospects in the Era of Artificial Intelligence - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/The-process-and-feature-of-simultaneous-interpretation_fig2_349939923 [accessed 29 Mar 2025]

Turaev, S., et al.: 'Review and Analysis of Patients' Body Language from an Artificial Intelligence Perspective', in IEEE Access, vol. 11, pp. 62140-62173, 2023