CSCanada
Higher Education of Social Science
Vol. 23, No.2, 2022, pp. 93-96
DOI:10.3968/12822

ISSN 1927-0232 [Print]
ISSN 1927-0240 [Online]
www.cscanada.net
www.cscanada.org

Evaluation of Student Interpreters Using Voice Recognition and Automatic Grammar Correction

LIU Bin[a]; LIU Hui[a],*

[a] School of Foreign Languages, North China Electric Power University, Beijing, China.
*Corresponding author.

Received 21 September 2021; accepted 29 October 2022
Published online 26 December 2022

Abstract
The evaluation and assessment of student interpreters have long been an issue for interpreting programs. The balance between student practice throughput, the time and human cost of assessment, and the quality of feedback is notoriously difficult to achieve. Here we demonstrate a way to rapidly assess student Chinese-to-English interpreting performance using automatic speech recognition and grammar correction software. The assessment results are compared with human graders against a set of criteria for grammar, fidelity, register, and enunciation. The results show that the semiautomatic assessment process is less time-consuming, and can give adequate feedback for enunciation, grammar, and register. Student volunteers were able to maintain engagement over a three-month period with minimal intervention from the instructor, however, interest began to drop over the long term.

Key words: Interpreting Studies; Student Interpreting; Interpreting Analysis

Liu, B., & Liu, H. (2022). Evaluation of Student Interpreters Using Voice Recognition and Automatic Grammar Correction. *Higher Education of Social Science*, *23*(2), 93-96. Available from: URL: http://www.cscanada.net/index.php/hess/article/view/12822
DOI: http://dx.doi.org/10.3968/12822

1. INTRODUCTION
Interpreting assessment and evaluation are relatively new topics compared with the critique of written translations. Assessment for interpreting is conducted through qualitative and quantitative methods, and different systems of evaluation focus on different angles depending on the research and purpose.

Buhler in his seminal work, (1986) developed an evaluation framework for semantic and pragmatic criteria for interpreters, based on questionnaires sent to AIIC members. Two content-related quality criteria sense consistency with the original and logical cohesion of utterance were the two top-rated criteria in the survey, followed by correct grammatical usage, the competence of interpretation, and fluency of delivery. This result was reaffirmed by Chiaro and Nocella (2004) using a wider internet-based study.

His idea may be added on by recent scholarship. However, other research has placed contradicting weights on interpreting standards. Sun (2017) for instance believes that evaluation should be assessed through fidelity and communicative properties, and benchmark performance using accuracy, fluency, expression of ideas, and expression of emotion.

The continuous development of Internet and mobile communication technology has accelerated the wide application of mobile APP, which means that speech recognition technology has gradually entered mobile APP, providing technical support for the new simultaneous interpretation (hereinafter referred to as “simultaneous interpretation”) teaching course. The new teaching mode should make use of modern information technology so that English teaching and learning can break away from the limitation of time and place, and develop towards individualized and autonomous learning. Previous studies on interpretation were in the empirical stage of interpretation practice and teaching practice, and the discussion focused on oral features and principles of interpretation (Liu HP., 2005). Because interpreting is an activity involving skills and abilities, interpreters must have strong professional knowledge and solid basic skills, and simultaneous interpretation has the

highest requirement among all interpreting categories. Therefore, the professionalization of translation talent training and the practicality of translation teaching are urgent problems to be solved. With the development of vehicle-mounted speech systems and various speech recognition software, speech recognition has come into people's lives, and it has been recognized by users for its practicality and accuracy.

This proof-of-concept study attempts to construct a teaching evaluation model of reverse manual interpretation evaluation through speech recognition technology and analyzes the application effect of speech recognition in simultaneous interpretation teaching by analyzing and comparing the interpretation process and translation quality. Student engagement in low instructor interference self-guided study is also measured to better assess the appeal of such a platform for students of interpreting.

2. QUALITY ASSESSMENT IN INTERPRETING

Since the 19th century, the quality standards of interpretation have been put forward by researchers and interpreters. Seleskovitch pointed out that the standard of interpretation is "expressiveness and fluency". Kurz (1993) conducted research on user expectations of several groups of interpreters and put forward the following interpretation quality standards: semantic consistency with the source language, logicality of interpreters, accuracy of terms, completeness of information, fluency of expression, and accurate grammar. Bao Gang (2005:178) believes that interpreting standards can be summarized as "comprehensive, standard and smooth". Gao Liang and Lin Yuru (1996) put forward an interpretation rating scale, which divided the quality standards into information completeness, accuracy, language expression, fluency, clarity, and adaptability. Yang Chengshu (2005) described a "Scoring Table for Interpreting Professional Examination" that divides interpreting standards into four items, namely faithfulness (accuracy and completeness), fluency and clarity, language (grammar, word selection), and time control. All kinds of scoring tables reflect the objectification of interpreting quality standards.

In 1995, the Association of International Edes Interpreters de Conference (AIIC) published a large-scale survey report on the expectations of conference interpreters (Moser, 1995). The report points out that translation completeness, accuracy, and meaning fidelity of terms in "content match" and synchronicity of simultaneous interpretation, oral expression skills, and pleasant voice in "formal match" are the most critical indicators.

3. THE APPLICATION OF SPEECH RECOGNITION IN COMPUTER-AIDED INTERPRETER TRAINING

The function of speech recognition is to convert a speech signal into the corresponding text information, and its system is composed of acoustic features, acoustic models, and decoders. The process of speech recognition is to extract acoustic features from the original speech data, and then get the acoustic model after training. This forms a speech network with dictionary pronunciation and language model, then extracts its features from new speech, and finally obtains the recognition result by Viterbi decoding through an acoustic model. The hidden Markov model (Rabiner, 1989) is adopted in the acoustic modeling of the speech recognition system, which aims to show the conversion relationship between the inherent hidden state of speech and the temporal sequence.

The Construction of Speech Recognition Technology in Interpretation Teaching; The research of the interpreting process has always been one of the key issues in interpreting theory research. Different situations will occur in the actual work of simultaneous interpretation. Translators should not only have extensive knowledge and strong language skills but also need strong adaptability and stress resistance. However, students who are not deeply involved in simultaneous interpretation are far from such a requirement, so it is difficult to carry out classroom exercises smoothly, and students feel strongly frustrated. Therefore, this study aims at such a traditional interpretation classroom, using speech recognition technology as an auxiliary tool for interpretation training, and discusses its application effect in traditional interpretation teaching.

In the process of Chinese-English translation, if students have insufficient knowledge of spoken English and grammar, they can quickly locate it through speech recognition tools, so that students can efficiently complete self-feedback.

4. MATERIALS AND METHODS

The speech used in this research comes from the European Union Speech Repository (EUSR).

From the content point of view, the corpus of EUSR includes two types: a). The speech materials selected by professional interpreters based on real conference interpreting scenes (such as European Commission meetings, parliamentary debates, press conferences, hearings, and interviews, etc.); b). Interpreting training materials suitable for different levels selected by senior professional interpreters and college interpreters. All the corpus videos are stored in the online database in ITU-

TH.264 format and are classified according to language, proficiency level, interpretation form, field, discourse type, etc. Meanwhile, professional interpreters and college interpreters regularly evaluate and audit the quality of the corpus in the library.

On the principle of corpus selection, authenticity is the typical feature of this corpus. The principle of the authenticity of teaching materials is a universally accepted standard, and its specific connotations include authenticity of materials, colloquialism of style, diversity of accents, and situational presentation.

In principle, the authenticity of EUSR is embodied in three aspects: material content, language features, and presentation forms: a). As far as material content is concerned, EUSR's material comes from the speeches and recordings of conferences of various European Union institutions, various international organizations and national institutions, which realizes the seamless connection between the material content and the real interpretation task, and the authenticity of the material content is practically guaranteed; b). As far as language features are concerned, the above-mentioned corpora are in line with the colloquial and on-the-spot features of interpreting teaching corpora, so they are in line with reality in style.

Language requirements of interpreting the training in context; c). As far as the presentation form is concerned, all the corpora are presented in the form of video transcription, which provides an intuitive and visual interpretation training scene for interpreters, and ensures the authenticity, immersion, and challenge of their task participation.

Sixty students enrolled in Translation and Interpreting master-level programs were given self-learning assessments. The source of the speech comes from the Speech Repository maintained by the EU. The speech is around 5 minutes in length and is conducted through consecutive C-E interpreting. The students were asked to join in a voluntary extracurricular exercise for one year. Volunteer participants have given their informed consent, and are free to leave the practice session at any time during the exercise. Student self-guided practices are issued six times a week and were asked to use text-to-speech and grammar editing software after each session, and write a short progress report (~200 words) at the end of each month. Student participation is recorded as the number of times they participate each week, with students finishing three practices or more per week regarded as high participation, and students with at least one practice per week counted as having low participation. Students who quit midway were also counted, and an exit survey was issued.

The speech-to-text software iFlyrec is used on the student recording, and the output is fed through grammar correction software. High-quality utterance without grammatical errors is not flagged by the grammar correction software, while poor utterances produce the wrong word being recorded by the speech-to-text transcript, which are be flagged by the grammar correction module.

Manual assessment is conducted in tandem to check for machine assessment accuracy.

5. RESULTS

This proof-of-concept research used the recording of 10 students across 5 C-E interpreting assignments. On average, this machine-aided assessment can detect 30-100 mistakes in a 6 minute speech:

5.1 Grammatical Mistakes

Grammatical mistakes were easily recognized by the machine assessment process. Oral delivery is different from written texts, especially under the pressure of time constraints. Errors that would otherwise not have manifested under written translation contexts regularly appear, and are easily detected by the semi-automatic system.

Ex1

Student: Last year, October 17th in Shanxi province, Changchun county, the residents there can take the bus for Free that it marks that Changchun county have an area when the bus is Free.

Correction: Last year, on October 17th in Shanxi province, Changchun county, the residents there can take the bus for Free that it marks that Changchun county has an area where the bus is Free.

5.2 Pronunciation Mistakes

Similarly, many pronunciation mistakes were easily caught by the machine assessment system.

Ex2:

Student: If you want to established an enduring project of robust, you have to do the following following steps.

Correction If you want to established an enduring project of free bus, you have to do the following following steps.

It should be pointed out that, ironically, as speech recognition software becomes more advanced, error-proofing has become integrated into many user-end software, to the extent that some minor pronunciation errors are not caught. In the example above, the delivery of "enduring" was pronounced as "induring", but the software auto-corrected the mistake. Therefore, only mispronunciations of a certain magnitude can be caught.

5.3 Knowledge Gaps and Understanding

Theoretically speaking, mistakes that deal with the understanding of the source text could not be detected by the platform. However, due to the fact that the Chinese speech in the European Speech Repository is all relatively simple to understand from a native speaker's perspective, with little to no technical jargon, this aspect cannot be tested.

5.4 Paralinguistic mistakes

Paralinguistic mistakes are also easily detected by the machine assessment pipeline. In Ex 2, the student interpreter's repeat of the phrase "following" is detected, as are involuntary utterances such as uh (Ex 3).

Ex3:

There are 16 youth uh-huh， youth peacekeepers that have gave their lives to the UN peacekeeping mission uh.

5.5 Recognition errors:

It should be noted that the speech recognition software also produced "false positives" due to misidentified phrases. These are especially common for proper nouns and less commonly used phrases. In the case of proper nouns, the misidentified term may also "fly under the radar" of the grammar software, since capitalization changes may affect error detection (Ex4).

Ex4:

Student:

And in 2005, in again, PRO proceed, there were also Free buses covering the main street around the city. It operated only 1 year.

Correction:

And in 2005, in Fujian Province, there were also Free buses covering the main street around the city. It operated only 1 year.

In Ex4, the student pronounced Fujian Province accurately but was transcribed by the speech-to-text software. Ironically, the mistakes were not caught by the grammar software, since it did not violate any grammatical rules.

5.6 Student Feedback

Of the 60 master-level students who participated in this initial study, 8 were able to sustain high levels (more than four days per week over a period of one year) of self-guided interpreting practice with machine feedback with minimal prompting from the instructor. 17 were able to maintain semi-self-guided interpreting practice with occasional prompting from the instructor. Students who had high levels of participation cited the incorporation of grammar corrections as a reason for their engagement, while students who had lower levels of participation cited the difficulty to juggle multiple platforms and the writing of monthly reports as a major hindrance behind their loss of interest.

It should be noted that 52 students (90%) were able to maintain high levels of participation for three months without much prompting from the researcher before drop rates increased. The issue of maintaining student interests over the long term, therefore still requires further breakthroughs.

CONCLUSIONS

In the age of the internet, the automatic correction system plays a great role in the role for the assistance of interpreting trainers and students.

Compared with the traditional interpretation teaching mode, speech recognition Machine assisted interpreting learning can be conducted in a quick and self-guided manner. (2) Machine-assisted assessment has good quality, and the removal of an "authoritative grader" can effectively overcome the students' fear and make the practice go smoothly. (3) Compared with the traditional interpretation teaching mode, the speech recognition assisted mode can effectively improve training efficiency, as students can conduct more practice with semi-feedback (4) The assisted mode of speech recognition has high coverage, which can effectively shorten the ability gap between students and ensure similar interpreting efficiency. (5) Students were able to maintain voluntary participation with minimal intervention from the instructors for three months before interests drop, and a better way of maintaining student interest in a self-guided study with minimal instructor intervention is still required.

This proof-of-concept research was able to show moderate success over the short term. For future research, an all-in-one platform dedicated to interpreter training can be established, with virtual incentives to maintain higher student interests.

REFERENCES

Bao, G. (2005). *Introduction to interpreting theories*. China Translation & Publishing Corporation.

Bühler, H. (1986). Linguistic(semantic)and extra-linguistic(pragmatic)criteria for the evaluation of conference inter-pretation and interpreters. *Multilingua, 5*(4), 231-235.

Chiaro, D., & Nocella, G. (2004). Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web. *Meta, 49*(2), 278-293. Érudit. https://doi.org/10.7202/009351ar

Gao, L., & Lin, Y.T. (1996). *E-C and C-E interpreting textbook.* Fujian Education Press.

Kurz, I. (1993). Conference interpretation: Expectations of different user groups. LINT.

Liu, H. P. (2014). Theorizing Interpretation: Advances and Trends. *Chinese Translators Journal, 4*, 71-74

Moser, P. (1996). Expectations of users of conference interpretation. *Interpreting, 1*(2), 145-178. https://doi.org/10.1075/intp.1.2.01mos.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257–286. https://doi.org/10.1109/5.18626

Sun, Y. Q., & Zhou, M. K. (2017). Study on the Comprehensive Evaluation Method of Machine Translation Quality. *Chinese Science & Technology Translators Journal, 2*, 20-24 https://doi.org/10.16024/j.cnki.issn1002-0489.2017.02.007

Yang, C. S. (2005). *Interpreting teaching research: Theories and practice.* Beijing: China Translation and PublishingCorporation.