

Retraction

Retracted: Design of Intelligent Speech Translation System Based on Deep Learning

Mobile Information Systems

Received 22 August 2023; Accepted 22 August 2023; Published 23 August 2023

Copyright © 2023 Mobile Information Systems. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.


The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Y. Tan, "Design of Intelligent Speech Translation System Based on Deep Learning," *Mobile Information Systems*, vol. 2022, Article ID 2463812, 7 pages, 2022.

Research Article

Design of Intelligent Speech Translation System Based on Deep Learning

Ying Tan 

Jiangsu Maritime Institute School of International Education, Nanjing, Jiangsu 211170, China

Correspondence should be addressed to Ying Tan; 20148441@stu.sicau.edu.cn

Received 21 June 2022; Revised 29 July 2022; Accepted 17 August 2022; Published 6 September 2022

Academic Editor: Jiafu Su

Copyright © 2022 Ying Tan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to solve the problem of low translation accuracy caused by complex sentence parameters in traditional machine translation systems, a method based on deep learning was proposed. First, MCU SPCE061A is used to study the problem of complex digital signal. The training data in the synchronous translation server support the translation services of a large number of users, and the translation results were displayed through the session interface of the user terminal. The PMDL model is used to detect the keyword signal, record the PCM audio data, and slice the collected pulse code modulation signal, so as to wake up the artificial intelligence voice service. Then, this study establishes a speech recognition process that accurately outputs the speech-related semantics. In this paper, a manual interactive synchronous translation program is designed with the input text as the search criterion, and the set is trimmed to obtain the best translation effect. The experimental results show that the sentence translation accuracy of the system is 0.9 ~ 1.0. It is proved that the method based on deep learning solves the problem of low accuracy of the traditional translation system.

1. Introduction

Governments, businesses, academic organizations, humanitarian organizations, and other organizations have recently faced unprecedented internationalization and globalization. The effectiveness, market size and scope of competition of security, and trade and commerce all depend on global information awareness and global interaction and communication ability. Strengthening world integration requires natural and effective international cross language communication, and the language gap is a huge obstacle to world integration [1]. With the increasing popularity of real-time translation, it is urgent to develop a system supporting synchronous translation. Speech to text translation (s2t) refers to the process that allows the machine to automatically translate the text of the target language from the speech signal of the source language.

A traditional speech translation system usually consists of two parts: speech recognition module and machine translation module. Cascading the two modules can form a speech translation system [2]. This cascading approach can

improve the overall performance from the improvement of a single component. For example, in recent years, neural machine translation has generally replaced statistical machine translation, which not only improves the quality of text translation but also improves the quality of speech translation. For this model which separates speech recognition and translation models, two models with good performance may be generated, respectively. However, when the two models are cascaded together, the inherent error propagation problem in this way will affect the performance of the whole system.

Cascading speech recognition and machine translation is still the mainstream method of speech translation [3]. However, the result of speech recognition is spoken sentences, which contain a large number of nonstandard languages, such as repetition, ellipsis, inversion, unclear semantic logic, broken sentences, and so on. In addition, influenced by the accent of the speaker, environmental noise, homonyms, and confusing words in the language, there will be errors in the results of speech recognition. These problems bring great challenges to the follow-up machine

translation. Therefore, a speech recognition postprocessing module is usually added between speech recognition and machine translation to reduce the impact of verbalization and recognition errors as much as possible by regularizing, breaking sentences, smoothing, punctuation prediction, and even error correction on the recognition results. Due to the complexity of oral English, these problems have not been completely solved. Combined with artificial intelligence technology, a machine synchronous translation system with a speech recognition device is designed. The translator configuration system helps to improve the synchronous translation service, to improve the user experience effect, and to promote the healthy development of the online translation market.

2. Literature Review

Liu and others proposed the famous dynamic time warping (DTW) algorithm, which can effectively compare two time series with different lengths and calculate the similarity through dynamic programming. Because the algorithm is easy to implement, it once dominated the research of speech recognition technology. Subsequently, the statistical model-based method gradually stepped onto the historical stage of speech recognition, among which the well-known algorithm is the hidden Markov model (HMM). After the statistical model, artificial neural network (ANN) has opened a new door for research in the field of speech recognition and has gradually become the mainstream method of speech recognition [4]. Luo and others proposed a new time-frequency convolution neural network (TFCNN) framework, which convolutes the feature space in both time and frequency scales. Under all test conditions, for all feature sets, the performance of the framework is always better than the convolution neural network, which can significantly reduce the word error rate (WER) [5]. Zhang and others proposed an end-to-end automatic speech recognition model for mono-multispeaker on ICASSP (International Conference on Acoustics, Speech and Signal Processing). Compared with other researches on mono-multispeaker speech recognition, this model can improve the performance of end-to-end model in separating overlapping speech and recognizing separated streams and bring about a relative performance improvement of about 10.0% in character error rate (CER) and WER, respectively [6]. Das and others proposed a new acoustic model for far-field speech recognition tasks. The long short term memory recurrent neural network (LSTM) based on attention mechanism and multitask learning framework reduced the absolute word error rate by 1.5% [7]. Xue and others proposed a digital Chinese continuous recognition system. Aiming at the characteristics of Chinese speech recognition, they adjusted the acoustic model of Sphinx speech recognition system so that the recognition rate of the model for speech digital string can reach 98% [8]. HORII and others proposed a speech recognition framework—deep full convolutional neural network (DFCNN). This framework has achieved 15% improvement over the two-way RNN-CTC, a widely used speech recognition model in the industry, and provides the

recognition of Cantonese, Henan, Sichuan, and other dialects in the dictation test of Chinese voice messages within iFLYTEK [9].

The translated sentences contain complex sentence parameters, which are not eliminated immediately, resulting in poor translation results. Therefore, combined with artificial intelligence technology, a machine synchronous translation system with a speech recognition device is designed. The translator configuration system helps to improve the synchronous translation service, improve the user experience effect, and promote the healthy development of the online translation market.

3. Research Method

3.1. System Hardware Structure Design. In order to realize the reasonable connection between translation services and the needs of users at all levels and also effectively allocate translator resources, this paper designs a machine synchronous translation system based on artificial intelligence technology and speech recognition [10]. The system is mainly composed of user terminal, server, and translator terminal. The feedback module of the user terminal is responsible for communication between internal modules or between the interpreter end and the outside, displaying the input and output content. The sound input and output part is synchronously translated by the translation module, and the feedback module finally gives feedback to the input interpreter user. The modules of the server are, respectively, used for the communication between the internal modules of the server or between the server and the external modules and to store the generated or received external information. The translation scoring module scores the information received by the computer after machine translation, and the order allocation module assigns orders to specific interpreters. The corpus learning module mainly transforms the translated content into corpus through machine learning technology to improve the level of machine translation. The translator terminal includes a communication module, which is responsible for the communication between internal modules or between the interpreter terminal and the outside. The display module is responsible for displaying the input and output contents; voice module inputs and outputs voice to the module; and translation module is responsible for synchronous translation.

3.1.1. SPCE061A Single-Chip Microcomputer. 32 KB flash is embedded in the SPCE061A chip. After high-speed processing, it can easily and quickly process complex digital signals. The structure of SPCE061A single-chip microcomputer is shown in Figure 1.

An intelligent speech recognition module based on SPCE061A is designed, and the corresponding control program was written. After when different voice command signals are received, the MCU pin sends the predetermined high and low levels [11, 12]. After when the microprocessor is connected to the power level, it will immediately and intelligently identify the sentences to be translated, send the

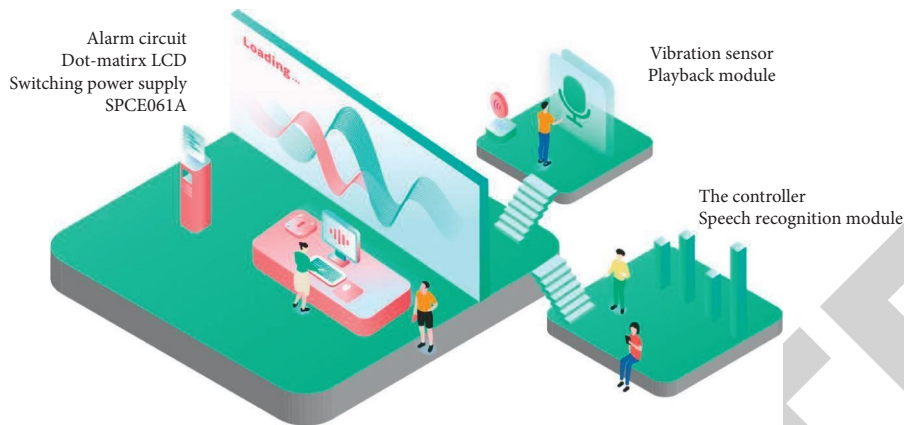


FIGURE 1: SPCE061A single-chip microcomputer structure.

letter composition instructions to the system in combination with the actual language conditions, and then combine the letters with idioms in combination with the artificial intelligence technology. The single-chip microcomputer with its own loudspeaker is used to play the intelligent recognition results synchronously in real time.

3.1.2. Machine Synchronous Translation Server. One or more translation servers and network servers running the decoder constitute a machine synchronous translation system, and the translation services between different languages are usually handled by different translation servers. Therefore, a distributed server system was built to support the translation services of a large number of users. The system was equipped with a decoding server and a network server. The online translation system runs on the decoding server. Users can query online synchronously through the HTTP network server interface.

The establishment of the machine translation server architecture is divided into two stages: first, learning the step size from the large-scale parallel corpus to obtain the maximum probability, and then second, the maximum probability solution is obtained by training. This method obtains the training data by counting the sentence pairs in the parallel corpus, then prunes them to remove the redundant data, and finds the most likely translation results according to the training data.

3.1.3. User Terminal. The user terminal provides a user interface, i.e., a session interface, which includes a session module and a translation module. This is the user interface. When the session button on the main interface is pressed, the program will jump to the session interface. After clicking the dialog button, the program will enter a dialog list page. When you click an entry in the dictionary, the program jumps to the dictionary page.

The user terminal session interface frame displays the chat list, the language information of each message, the time when the message is translated, etc. There is a language input button at the bottom of the dialog box to input Chinese and English buttons, respectively [13, 14].

3.1.4. Microphone Array. Due to various noise interference, the speech signal is inaccurate, and even the speech is submerged. Therefore, installing a microphone array in the system can convert sound signals into digital signals. Using this array can not only improve the resolution of sound but also extract accurate pure speech from noisy speech signals. This method extracts the time and space information of the sound source through the microphone array and suppresses the noise so as to accurately recognize the voice instructions in the noisy environment.

3.2. System Software Design

3.2.1. Design of Artificial Intelligence Voice Wake-Up Function. In order to realize the voice wake-up function, it is necessary to set up a process in the background regularly and real time monitoring of the surrounding environment of the device and detect whether there is a signal of the keyword required by the device in the signal. The specific process is shown in Figure 2.

Figure 2 shows that after the keyword signal detection was completed, the detection callback or static detection was performed and the action is recorded. During the static detection of the recorded action, the process returns the recorded PCM audio data to the main process, and then continues to the next step. On this basis, the PMDL model was used to slice the collected pulse code modulated signal [15, 16].

3.2.2. Speech Recognition Function Design. With the continuous development of the Internet of things and artificial intelligence, the demand for free interaction between people and computers is more and more frequent, and people pay more and more attention to it. Now, it is generally believed that speech recognition is a part of language technology, which also includes speech synthesis and natural language processing. The main purpose of speech recognition technology is to complete the conversion from speech to text, so speech recognition is also the basic condition of natural language processing. The speech signal is generated by the vibration of the vocal organ, so the lip radiation and glottic

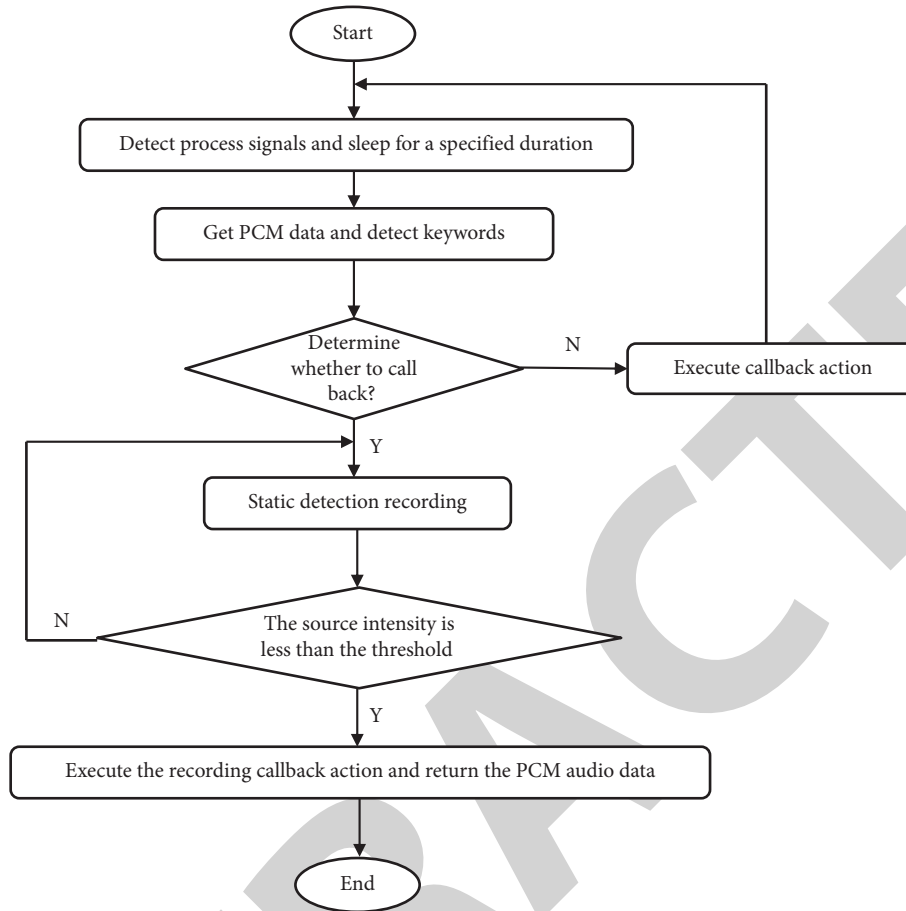


FIGURE 2: Artificial intelligence voice wake-up process.

excitation will have a certain impact on it, resulting in the reduction of the high-frequency band of the speech by 6 db/octave band above 800 Hz. Therefore, in the process of speech transmission, the high-frequency part is easy to be lost. Therefore, the preemphasis step should be introduced before the speech signal processing. The main function of speech signal preemphasis is to reduce the influence of lip radiation and improve the resolution of high-frequency part. Speech signal is a typical nonstationary time-varying signal; that is, its distribution law will change with time. Since most signal processing systems can only process stable signals at present, it is necessary to process speech signals first. After research, the speech signal has the characteristic of “short-term stability” because the occurrence of speech is closely related to the movement of the vocal organ. When speaking, the vocal organ will produce a certain inertial movement. Within a period of time (generally 10 ms ~ 30 ms), the speech signal can be regarded as approximately unchanged, so this speech signal can be treated as a stable signal [17, 18]. The speech recognition process is presented in Figure 3.

Figure 3 demonstrates that the speech recognition system can complete speech recognition based on the following four working principles. First, the speech recognition library and detection terminal are selected directionally and then combined with antialiasing bandpass filtering technology, which can effectively eliminate individual speech differences,

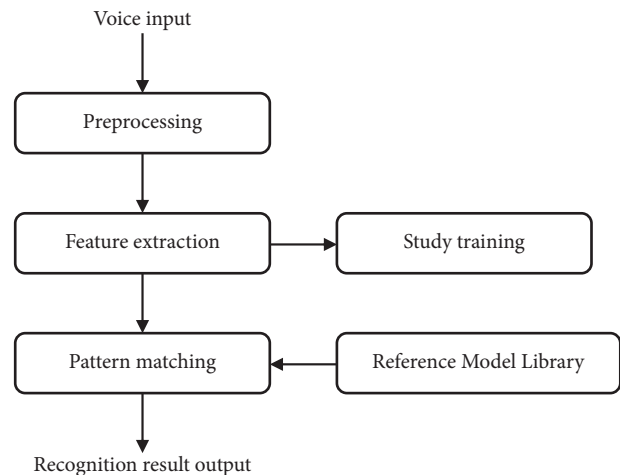


FIGURE 3: Speech recognition process.

sampling equipment, and sampling environment noise. Then the speech acoustic parameters such as average energy, vibration peak, and average zero crossing rate are extracted, which can quickly and accurately reflect the sound quality characteristics of speech; After that, a speech pattern database is established. The main link of language repetition training is to let the speaker repeat his pronunciation, directly delete the redundant speech information from the

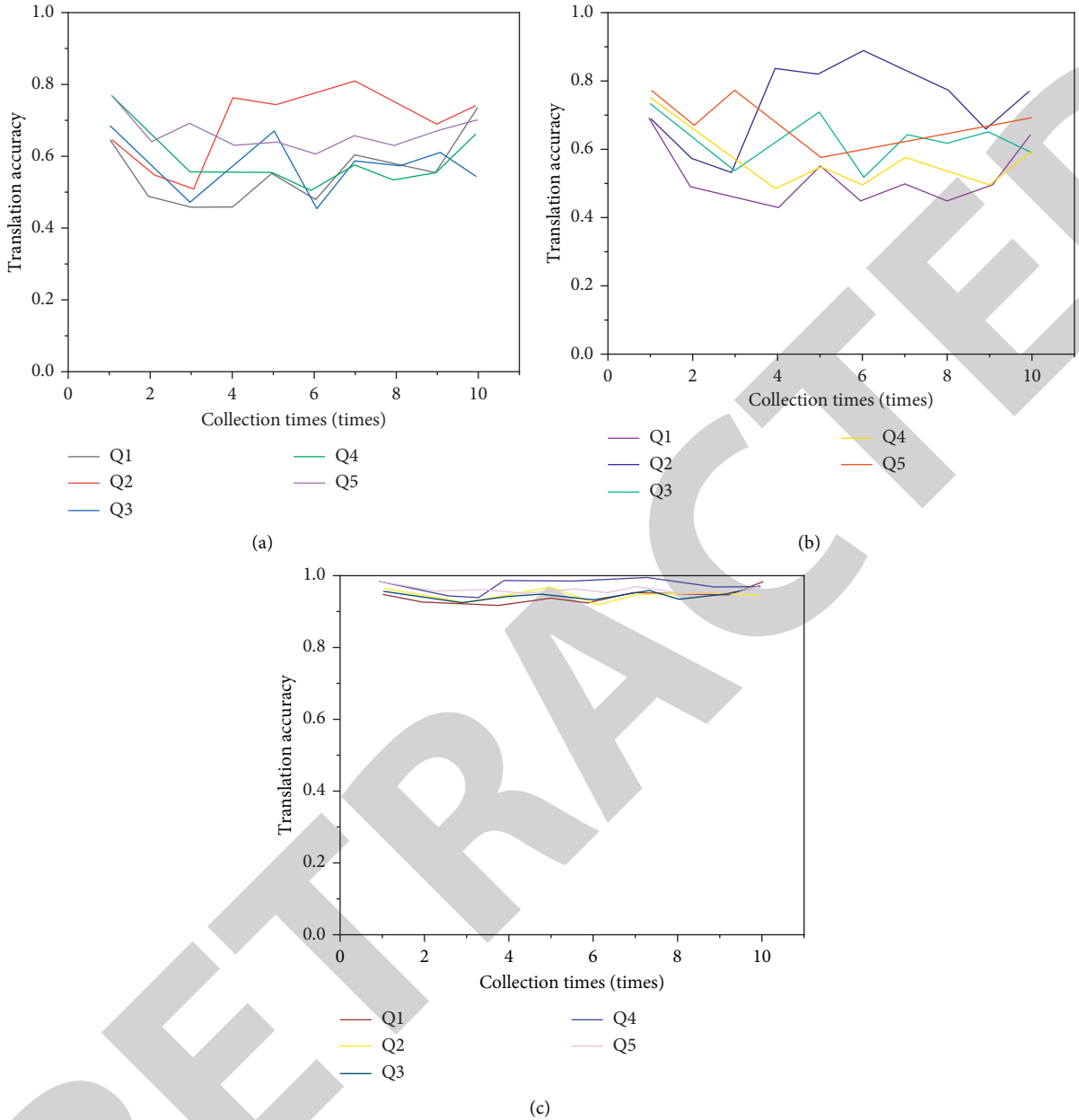


FIGURE 4: Comparative analysis of translation accuracy of the two systems. (a) Machine synchronous translation system based on the Moses statistical machine translation framework. (b) Corpus based machine translation system. (c) Machine synchronous translation system based on artificial intelligence and speech recognition.

original speech samples one by one, only retain some key speech data, and scientifically classify the key speech data according to the relevant schemes. Finally, the relevant semantics of speech are accurately output according to the speech similarity [19, 20].

3.2.3. Manual Interactive Synchronous Translation. The machine translation method of human-computer interaction function includes the following steps:

Step 1. Read the machine translation model and select the corresponding prestored domain according to the user of the translation domain.

Step 2. After reading the text, divide the text into a series of sentences to facilitate subsequent modules.

Step 3. By inputting text as search criteria, after receiving the user’s input text, search the matching translation on the search network, and different users can get different translation results corresponding to the input [21]. Since the number of states in the W_n set increases exponentially with the increase of n , it will take a lot of time if the set is not pruned. Therefore, pruning is required. The pruning process is as follows: for the determined source language sentences R_1, R_2, \dots, R_n , there is a phrase model (f, g, s, d) , and the four elements in the model represent phrase library, grammar

model, distortion limit, and distortion parameter, respectively.

Suppose W_n translates n words into m sets. If an element in the M statement in W_2 consists of two words, it means that only two words are translated into idioms. For W_i , each state has a transition state, all possible states will be added to the corresponding set, and finally the state with the highest score will be returned [22].

Let β be the search parameter and p be the transfer parameter, and the resulting syntax model is as follows (1):

$$g = \text{next}(p, d). \quad (1)$$

After determining the search parameters, remove all the parameters in the set that do not meet (2) to achieve the purpose of pruning:

$$a(g) > a(d) - \beta. \quad (2)$$

Step 4. When translating sentences in the source language, first read the translation options of the sentences in the source language, and then expand the translation assumptions from a small container to a large container. At each transition stage, if the difference between a score and the highest score in the container is greater than the threshold, then the state will decline. If the state remains unchanged, all available transition options expand. If the old and new assumptions are the same, the score increases. The best translation result is to find the translation statement with the highest score in the largest container.

4. The Experimental Results

The purpose of the experiment was to test the functions of the machine synchronous translation system based on artificial intelligence technology and speech recognition so as to ensure that the system can meet the purpose of synchronous translation.

4.1. Experimental Data Acquisition. 123425 English sentences were selected from the translation database, out of which, 1000 were derived from translation materials in the field of news. Five of the 1000 sentences were randomly selected and translated by five teachers. The translated results are as follows: Q1 is Japan is interested in China's new round of strategic technology; Q2 is that the United States exerts pressure on South Korea; Q3 is that the United States and China hold negotiations on the issue of border deployment; Q4 is that the US side attaches great importance to China US relations; and Q5 is that China will not pose a threat to other countries. In the experimental results, the Moses statistical machine translation framework translation system, corpus translation system, artificial intelligence technology, and speech recognition translation system are used to translate the above five sentences. The accuracy of the translation results is shown in Figure 4.

It can be seen from Figure 4(a)) that using the machine synchronous translation system based on Moses statistical machine translation framework, the translation accuracy of

five sentences was between 0.4 and 0.8. It can be seen from Figure 4(b)) that using the corpus based machine synchronous translation system, the translation accuracy of the five sentences was between 0.4 and 0.9. It can be seen from Figure 4(c)) that the machine synchronous translation system using artificial intelligence technology and speech recognition had a translation accuracy of $0.9 \sim 1.0$ for five sentences [23]. According to the above analysis results, the machine synchronous translation system using artificial intelligence technology and speech recognition has high translation accuracy.

5. Conclusion

A single-chip microcomputer has the characteristics of convenient and rapid processing of complex digital signals in speech recognition hardware, which has become the main advantage of machine synchronous translation system to realize artificial intelligence technology and speech recognition. The system combines the artificial intelligence technology and speech recognition technology. It has the advantages of high recognition efficiency and high accuracy and thus meets the needs of people for different speech communication. However, the system still has many shortcomings, which need to be further improved in the design process: in the next development stage, a more efficient decoding algorithm can be used to improve the decoding speed. A more robust caching mechanism and common word search mechanism can be adopted to speed up offline translation and meet the requirements of efficient translation. The application of deep learning technology can ensure the organic unity of high accuracy and high efficiency of scenic spot text translation and speech recognition. It provides a convenient and simple way for tourists to obtain scenic spot information. At the same time, it provides a good help for tourists who are unfamiliar with the local language of the tourist destination or have certain visual impairment so that they can effectively understand the scenic spot information and improve their tourism experience.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Z. Z. Gafarova, M. A. Bozorova, S. S. Jumayeva, L. I. Idiyeva, and L. U. Radjabova, "Synchronous translation – a complex set of cognitive processes," *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 1, pp. 403–407, 2020.
- [2] S. M. Shavkatovna, N. N. Soliyevna, S. M. Ahmadovna, and K. S. Shaydullovna, "Strategies for mastering synchronous translation art," *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 03, pp. 404–415, 2020.

- [3] V. Galina M, "Axiology of Faust: a tragedy translation in the context of Russian-european intercultural relations," *Humanitarian Vector*, vol. 15, no. 5, pp. 8–15, 2020.
- [4] Y. Liu, J. Zhang, H. Xiong et al., "Synchronous speech recognition and speech-to-text translation with interactive decoding," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, pp. 8417–8424, 2020.
- [5] R. C. Luo and Y. W. Perng, "Adaptive skew force free model-based synchronous control and tool center point calibration for a hybrid 6-dof gantry-robot machine," *IEEE*, vol. 25, no. 2, pp. 964–976, 2020.
- [6] P. Zhang, C. Wang, N. Kumar et al., "Artificial intelligence technologies for covid-19-like epidemics: methods and challenges," *IEEE Network*, vol. 35, no. 3, pp. 27–33, 2021.
- [7] D. K. Das, "Artificial intelligence technologies for breast cancer screening," *Oncology Times*, vol. 43, no. 5, pp. 20–21, 2021.
- [8] Y. Xue, C. Fang, and Y. Dong, "The impact of new relationship learning on artificial intelligence technology innovation," *International Journal of Innovation Studies*, vol. 5, no. 1, pp. 2–8, 2021.
- [9] H. Horii, "Advancement of vehicle occupant restraint system design by integration of artificial intelligence technologies," *International Journal of Transport Development and Integration*, vol. 5, no. 3, pp. 242–253, 2021.
- [10] A. V. Rakšnys, D. Gudelis, and A. Guogis, "The analysis of opportunities of the application of big data and artificial intelligence technologies in public governance and social policy," *Socialinė Teorija Empirija Politika ir Praktika*, vol. 22, no. 6, pp. 88–100, 2021.
- [11] L. Cui, "A preliminary study on the management strategy of university personnel files based on artificial intelligence technology," *Journal of Electronic Research and Application*, vol. 5, no. 2, pp. 1–4, 2021.
- [12] X. Kun, Z. Wang, Z. Zhou, and W. Qi, "Design of industrial internet of things system based on machine learning and artificial intelligence technology," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2601–2611, 2021.
- [13] P. Folkeard, M. V. Eeckhoutte, S. Levy et al., "Detection, speech recognition, loudness, and preference outcomes with a direct drive hearing aid: effects of bandwidth," *Trends in Hearing*, vol. 25, no. 4, p. 233121652199913, 2021.
- [14] K. A. Berg, J. H. Noble, B. M. Dawant, R. T. Dwyer, R. F. Labadie, and R. H. Gifford, "Speech recognition as a function of the number of channels for an array with large inter-electrode distances," *Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. 2752–2763, 2021.
- [15] J. S. Kempfle, A. Panda, M. Hottin et al., "Effect of powered air-purifying respirators on speech recognition among health care workers," *Otolaryngology-Head and Neck Surgery*, vol. 164, no. 1, pp. 87–90, 2021.
- [16] M. H. Alhumsy and S. Belhassen, "The challenges of developing a living Arabic phonetic dictionary for speech recognition system: a literature review," *Advanced Journal of Social Science*, vol. 8, no. 1, pp. 164–170, 2021.
- [17] J. Smith and Q. Chen, "Enabling low-dose liquid-phase TEM with advanced signal processing, machine learning, and molecular simulation," *Microscopy and Microanalysis*, vol. 27, no. S1, pp. 1314–1315, 2021.
- [18] J. Edwards, "Wearables-fashion with a purpose: a new generation of wearable devices uses signal processing to make life easier, healthier, and more secure [special reports]," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 15–136, 2021.
- [19] C. Liu, M. Lin, H. L. Rauf, and S. S. Shareef, "Parameter simulation of multidimensional urban landscape design based on nonlinear theory," *Nonlinear Engineering*, vol. 10, no. 1, pp. 583–591, 2021.
- [20] R. Huang, P. Yan, and X. Yang, "Knowledge map visualization of technology hotspots and development trends in China's textile manufacturing industry," *IET Collaborative Intelligent Manufacturing*, vol. 3, no. 3, pp. 243–251, 2021.
- [21] X. Liu, J. Liu, J. Chen, F. Zhong, and C. Ma, "Study on treatment of printing and dyeing waste gas in the atmosphere with Ce-Mn/GF catalyst," *Arabian Journal of Geosciences*, vol. 14, no. 8, Article ID 737, 2021.
- [22] M. Bradha, N. Balakrishnan, S. Suvi et al., "Experimental, Computational Analysis of Butein and Lanceoletin for Natural Dye-Sensitized Solar Cells and Stabilizing Efficiency by IoT," *Environment, Development and Sustainability*, vol. 24, 2021.
- [23] N. Yuvaraj, K. Srihari, G. Dhiman et al., "Nature-inspired-based approach for automated cyberbullying classification on multimedia social networking," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–12, Article ID 6644652, 2021.