Comput Syst Sci & Eng (2020) 5: 377–383

International Journal of
Computer Systems
Science & Engineering

Implementation of Embedded Technology-Based English Speech Identification and Translation System

Zheng Zeng*

Chengdu Technological University, Chengdu, Sichuan 611730, China

Due to the increase in globalization, communication between different countries has become more and more frequent. Language barriers are the most important issues in communication. Machine translation is limited to texts, and cannot be an adequate substitute for oral communication. In this study, a speech recognition and translation system based on embedded technology was developed for the purpose of English speech recognition and translation. The system adopted the Hidden Markov Model (HMM) and Windows CE operating system. Experiments involving English speech recognition and English-Chinese translation found that the accuracy of the system in identifying English speech was about 88%, and the accuracy rate of the system in translating English to Chinese was over 85%. The embedded technology-based English speech recognition and translation system demonstrated a level of high accuracy in speech identification and translation, demonstrating its value as a practical application. Therefore, it merits further research and development.

Keywords: Embedded technology; speech identification; HMM algorithm, translation system

1. INTRODUCTION

With the development of globalization, communication between people has gradually transcended the boundaries between countries, and interactions between people who speak different languages are becoming more frequent. In order to break through language barriers, language translation technology have emerged. Machine translation is capable of text translation, but it cannot substitute for oral communication and its prosodic nuances. In order to solve this problem, the design of speech identification and translation system has gradually become a new research direction. A speech identification and translation system can effectively eliminate some of the obstacles in cross-language communication [1].

A speech translation system which is used to translate one speech into another speech consists of speech recognition, machine translation and speech synthesis. The first speech translation experiment system in the world was the Speech Trans developed in the US in 1989. Subsequently, more and more countries have researched such systems. Bangalore et al. [2] designed a statistical model which included a speech-to-speech (S2S) system and SIP architecture to facilitate real-time, two-way, trans-language dialogue. Éva Székely et al. [3] developed a facial expression-based affective speech translation system to classify the emotional states of users and output in appropriate sound style. Sangeetha et al. [4] studied the translation from English to Dravidian, proposed a hybrid machine translation system based on Hidden Markov model (HMM), and found that it had strong speech translation performance. In terms of speech

*Address: Room 407, Unit 1, Building 4, Teachers' apartment, Chengdu Technological University, No.1, Section 2, Zhongxin Avenue, Pidu district, Chengdu, Sichuan 611730, China
Email: zhengz_zeng@outlook.com

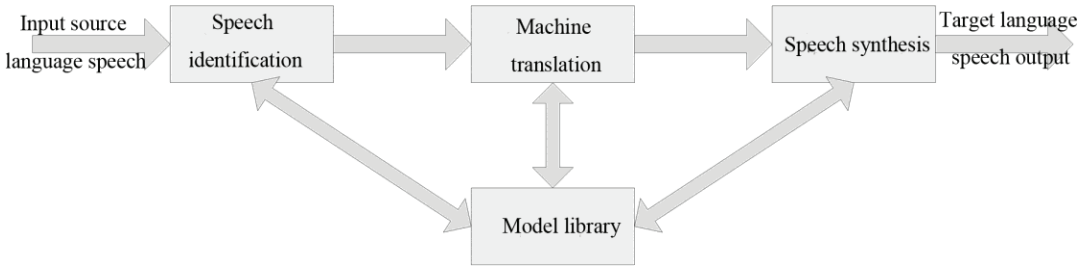

Figure 1 The structure of the speech identification and translation system.

feature extraction, Kim et al. [5] proposed an algorithm called the Power Normalized Cepstrum Coefficient (PNCC), which could provide a high level of recognition accuracy in a noisy environment and required very little calculation. Popovic et al. [6] studied the recognition of the Serbian language, designed a system based on deep neural network, trained the algorithm using used 90 hours of speech corpus, and modified the algorithm according to the linguistic features of the Serbian language. Liu et al. [7] studied the application of a support vector machine (SVM) in speech recognition, proposed a logistic kernel function, and verified the performance of the method in speech recognition through experiments. Pham et al. [8] compared the phrase-based and neural-based Vietnamese machine translation methods, and found that the accuracy of the phrase-based method was 97.32% and that of the neural-based method was 96.15%; however, the neural-based method required a high operation speed and a large development space. Embedded technology is a kind of technology with low cost and small volume, which has a wide range of applications in many fields. Cesarini et al. [9] designed an embedded system to convert the pressure value of a swimmer's injury into sound and used it as a communication channel between coaches and athletes. Lee et al. [10] studied the embedded system based on GPU and designed a novel scheduling framework to improve the system's flexibility. Al-Odat et al. [11] designed an embedded healthcare system for diabetic patients and found, through experiments, that the method had a 99.3% accuracy. Yi et al. [12] designed a video feature location system based on embedded technology for video monitoring, which had good performance, low cost and high positioning accuracy.

In this study, a technology-based speech identification and translation system was realized on the embedded operation system using the HMM algorithm. Compared to the traditional translation system, the system proposed here had a higher identification rate and better accuracy, suggesting its strong potential as a practical application.

2. OVERVIEW OF THE SYSTEM

2.1 Speech Translation System

Speech translation technology has emerged as a means of facilitating communication by circumventing language barriers. A speech translation system can translate different languages through a computer system that offers many features such as linguistics, speech recognition and speech synthesis. Hence, it has the potential to greatly influence people's lives and induce social changes. As this technology began to attract more in-depth research, the vocabulary of the speech translation system began to expand gradually, and multilingual translation and two-way translation began to appear. It has been applied in a small number of fields, such as the public transport sector that uses it for ticket sales and train timetable inquiries.

2.2 Embedded Technology

Embedded technology involves the embedding of a chip, which is written with device control programs, into a device in order to control some of its operations. With the development of science and technology, embedded technology is being applied in an increasing number of fields. In a speech recognition and translation system, embedded technology offers several advantages: high level of stability, good adaptability, and simple and convenient operation. It can improve the recognition rate and accuracy of the system. Moreover, embedded technology has high reliability, low power consumption, and a lengthy life cycle, which can save system design cost. Therefore, the embedded speech recognition and translation system has great market potential.

2.3 The Working Principles of the System

The speech identification and translation system consists of speech identification, machine translation and speech synthesis [13] (Figure 1).

Speech recognition refers to the transformation of speech into the corresponding text or instruction by means of a machine. This technology has been widely used in phonetic dial-up and intelligent toys, and can greatly improve the quality of life. With the development of science and technology, speech recognition technology is constantly improving, but the accuracy of its recognition is still an important research challenge. Context, the speaker's emotion and a change of environment will affect the recognition rate [14].

Machine translation refers to the conversion of one language to another language based on the computer's programming and calculation capabilities. It is closely related to computer technology and linguistics and can foster political, economic and cultural exchanges, offering strong scientific and practical value.

Speech synthesis refers to the conversion of text to speech, i.e., the transformation to sound of the received text information

in real time. It involves acoustics and linguistics. It consists of text analysis, prosodic control and speech synthesis. Firstly, the text is normalized; then, the pitch and length are synthesized. This is followed by speech as the output.

3. DESIGN AND IMPLEMENTATION OF THE SYSTEM

3.1 Embedded Operation System

In this study, the Embedded Windows CE operation system was used as it has favorable transportability, good real-time performance and strong functions.

The hardware of the embedded system comprised two main components: a master control core and a speech identification component. The main controller was a STM32F103C8T6 chip (STMicroelectronics Group, France) which was installed with a 64KB high-speed storage and an enhanced I/O port. An LD3320 chip (ICRoute Company) which was integrated with the optimized speech identification algorithm was used in the speech identification part. The English speech identification algorithm achieved high accuracy.

An embedded database is a light database which can operate without starting on the server side. In the speech translation system, both speech identification and machine translation need to frequently visit the data on embedded devices. The database can help manage the data. In this study, the SQL Server Mobile 2005 database was used, and the mobile databases were synchronized using a replica technique.

3.2 Speech Identification Algorithm

The Hidden Markov Model (HMM) was used to realize the speech identification function. HMM was set as H, i.e.

$$H = \{V, N, M, \pi, A, B\}$$

where V stands for the number of linguistic units included in the model, N stands for the number of states, Mstands for the number of observation symbols which might be output by the states, and π stands for the set of probability of initial state, i.e.

$$\pi = \{\pi_{wi} \mid 1 \le w \le V, 1 \le i \le N\}$$

where π_{wi} refers to the probability that the initial state of word w was state i.

A stands for the set of probability of state transition, i.e.

$$A = \left\{a_{wxij} \mid 1 \le w, x \le V; 1 \le i, j \le N\right\}$$

where a_{wxij} refers to the probability of state i of word w transiting to state j of word x.

B stands for the set of probability of output, i.e.

$$B = \left\{b_{wjk} \mid 1 \le w \le V, 1 \le j \le N, 1 \le k \le M\right\}$$

where b_{wjk} refers to the probability of word w in state ^{j} outputting k-th number symbol in VQ codebook.

HMM was denoted as $\lambda = (A, B, \pi)$.

The observation sequence obtained after vector quantization of speech signals was

$$O = o_1 o_2 \cdots o_r$$

Three problems must be solved before applying the algorithm in speech identification.

The first problem was how to effectively calculate the probability of the observation sequence $P(O|\lambda)$ when $O = o_1 o_2 \cdots o_r$ and $\lambda = (A, B, \pi)$ were given.

The second problem was how to determine the optimal state sequence $S = \{s_1, s_2, \cdots, s_N\}$

The last problem was how to adjust $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$ when the observation sequence was given.

The procedures for applying HMM in speech identification are as follows.

(1) The set of sound class of model L was defined as $V = \{v_1, v_2, \cdots, v_n\}$.

(2) Sets which included a certain number of labeled voice were accumulated for every voice class V_i.

(3) Every voice class acquired an optimal model λ_i from the training set.

(4) In the process of training, for each unknown sequence O, its probability was $\Pr(O|\lambda_i)(i = 1, 2, \ldots, L)$. Moreover, the speech corresponding to an unknown sequence O was determined for each class V_i.

3.3 Design of the Translation System

Rules-based direct translation whose basic unit was phrase was used. The system divided an English sentence into multiple connected word strings, then translated every English phase into Chinese, adjust the order, and finally output the translation text. The segmentation of phrase is shown below.

Original text:

I	will	go	shopping	in	the	mall

Division of phases:

I will	go shopping	in the mall

Translation:

我将	购物	在商场

Adjust the order:

我将	在商场	购物

The translation code was:

```
class CHTranslate:public CHObject
{ public:
      Cstring Match(Cstring&);//character matching
      void QF(Cstring &source);//sentence division
      //translate the sentence after analysis according to
      translation rules
      void Rule1();
```

```
        void Rule2();
        void Rule3();
        void Rule4();
        Cstring Translate(Cstring & source);// translate words
        segmented
        Cstring ZuHe();//integrate words

}

Calling class object with translation conversion module:
Cstring Ctranslate::Translate(Cstring & source)
{ QF(source);//segment sentence
        //Conversion rules
        Rule1();
        Rule2();
        Rule3();
        Rule4();
        return ZuHe();//combine translation texts

}
```

Firstly, the segmented words were stored in a linked list Cword to form a vocabulary linked list. Then every object was traversed using the linked list. Finally, a complete sentence was created after a string of words in the target language was obtained.

3.4 Speech Synthesis Technology

A Chinese monosyllable parameter library was established after analyzing and editing all Chinese character pronunciations as the basis of synthetic speech. Moreover, relevant rules have be formulated in regard to tone, tone modification, emphasis and pauses. These rules were used in speech synthesis to render output speech that was more fluent and natural. A simplified speech parameter library, a rule library, and linear predictive coding were used. The technology was relatively mature and could produce high-quality synthetic speech, which could reduce cost and improve synthesis speed.

4. RESULTS OF SYSTEM APPLICATION

4.1 Speech Identification

First, spoken words were recorded. Then the words which needed to be recognised were selected on the system interface. After preprocessing and feature extraction, the word recognition results were obtained (Figure 2).

Speech was recorded in a quiet environment. The words 0∼9, 'good morning' and 'hello' were spoken into a microphone and recorded. Every word was read ten times. If the output text included the word to be recognised, then the recognition was considered to be successful.

4.2 Results of English Speech Recognition

0∼9, 'good morning' and 'hello' were read ten times and identified using the embedded speech identification and translation system and the traditional translation system [15]. The recognition results are shown in Figure 3.

The results demonstrated that the recognition rate of the embedded speech recognition and translation system was obviously higher than that of the traditional translation system (88% vs. 68%). This indicated that the speech recognition function of the traditional translation system was poor, with a strong probability that the input speech would not be able to be identified. In the embedded speech recognition and translation system, the speech recognition module based on HMM had a significant advantage in terms of speech recognition.

4.3 Results of English-Chinese Translation

0∼9, 'good morning' and 'hello' were read ten times and identified using both the embedded speech identification and translation system and the traditional translation system. The results are shown in Figure 4.

The results showed that the accuracy rate of the embedded translation system was much higher than that of the traditional translation system. The accuracy rate of the embedded system was above 85%, while the accuracy of the traditional translation system was not more than 70%. Because of the high recognition rate, the embedded translation system also demonstrated high accuracy in speech translation; it could identify nearly all the input speeches and translate them accurately. Affected by the poor recognition rate, the traditional translation system was not as accurate as the embedded system when translating.

In order to further verify the effectiveness of the proposed system, it was compared with the system designed in a previous experiment [16] involving the recognition and translation of 100 complex English words such as 'phenomenon', 'ethnicity', 'remuneration', 'philosophical', etc. The level of speech recognition and the translation accuracy of the two methods are shown in Table 1.

As evident in Table 1, the speech recognition rate and translation accuracy of the system designed in this study decreased when handling complex words, but it is still over 80%; the speech recognition rate of the system designed in this study is 83.6%, which is 6.23% higher than that of the previous system designed by Fu [16]; the translation accuracy is 81.4%, which is 6.82 % higher than that obtained by the traditional system [16]. The results verified that the system designed in this study was reliable.

5. DISCUSSION

Speech is the most natural and convenient way to communicate, and also the most direct way of engaging in human-machine interaction [17]. Speech translation technology enables communication between people speaking different languages, using a computer system as the intermediary. The three components of speech translation -speech recognition, machine translation and speech synthesis- all need the support of corresponding technologies. With the development of technology, automatic speech recognition and machine translation have made significant progress [18].

HMM has gradually become an important technology for speech recognition. Compared to other technologies, HMM

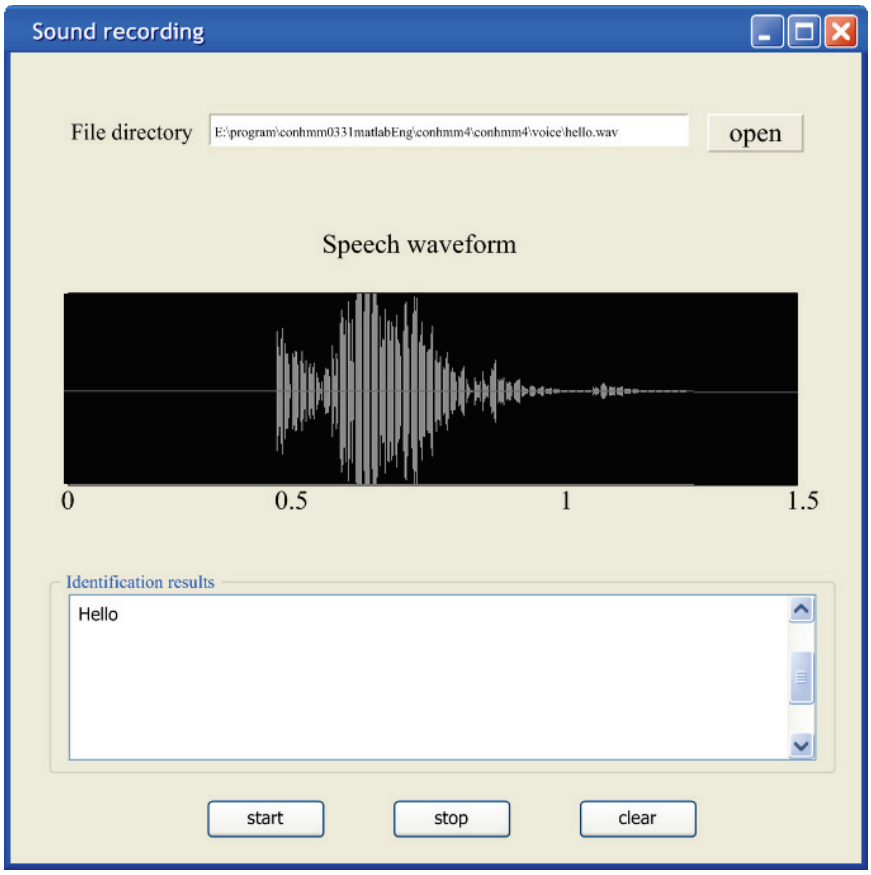

Figure 2 The speech recognition process.

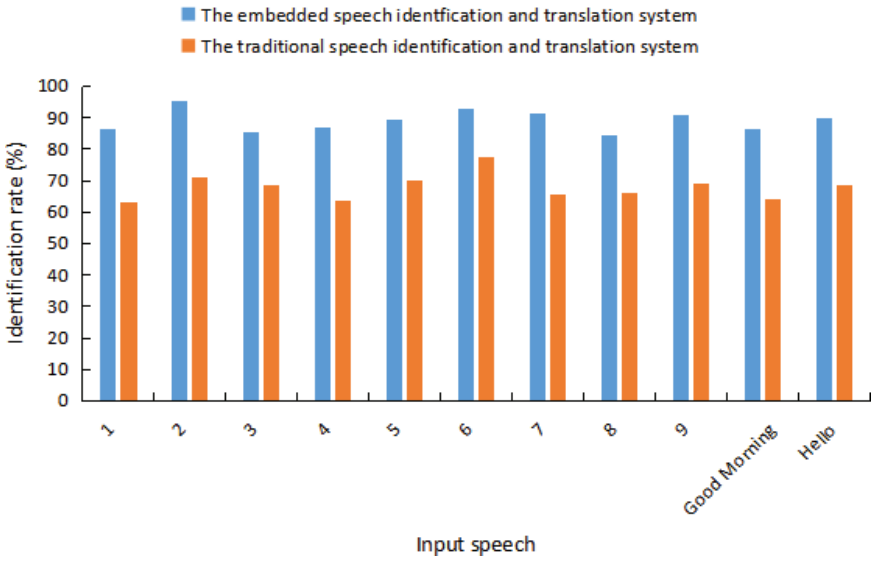

Figure 3 Comparison of recognition results for English speech.

Table 1 Comparison of speech recognition and translation systems.

	The system designed in this study	Literature [16]
Speech recognition rate/%	83.6	78.7
Translation accuracy/%	81.4	76.2

has a high recognition rate [19]. In addition, embedded technology offers good advantages in a speech recognition translation system, and can significantly improve the levels of recognition speed and recognition accuracy. Therefore, this study designed an embedded speech recognition translation system by combining HMM with embedded technology.

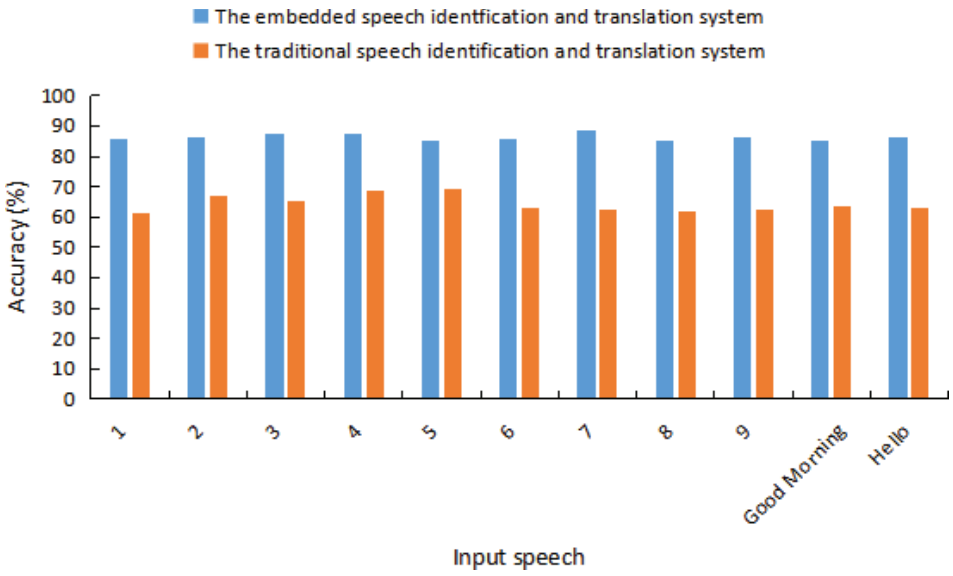

Figure 4 Comparison of speech translation results.

The experimental results showed that the system designed in this study was superior to the traditional system in terms of both the amount of speech recognition and the level of translation accuracy in simple speech. The system designed in this study achieved an average of 88% for speech recognition, 29.41% higher than the traditional system; in regard to the translation of speech, the system designed in this study achieved an accuracy of over 85%, whereas the traditional system had an accuracy of only 60%-70%. Then, for complex speech recognition, it was found through comparison with literature [16] that the level of speech recognition and the translation accuracy of the system designed in this study were higher, which highlights the advantages of the system proposed in this paper.

The combination of embedded technology and speech recognition technology is of great significance to speech translation. It plays an important role in the promotion of speech translation technology and the realization of speech translation in embedded devices. The embedding of a speech recognition algorithm in systems can improve the efficiency of speech recognition, thereby providing a significant benefit to the system user.

6. CONCLUSION

The speech recognition and translation system which combined the embedded technology and the HMM algorithm achieved a high recognition rate and accuracy and performed better in terms of English speech recognition and translation than did the traditional speech translation system. However, speech translation in a noisy environment and the translation of continuous speech were not explored in this study, and provide fertile ground for future research.

REFERENCES

1. Do Q. T., Toda T., Neubig G., et al. Preserving Word-Level Emphasis in Speech-to-Speech Translation. IEEE/ACM Transactions on Audio Speech & Language Processing, 2017, 25(3): 544–556.
2. Bangalore S., Sridhar V. K.R., Kolan P., et al. Real-time incremental speech-to-speech translation of dialogs. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2013: 437–445.
3. Éva S., Steiner I., Ahmed Z., et al. Facial expression-based affective speech translation. Journal on Multimodal User Interfaces, 2014, 8(1):87–96.
4. Sangeetha J., Jothilakshmi S. Speech translation system for English to Dravidian languages. Applied Intelligence, 2016, 46(3):1–17.
5. Kim C., Stern R. M. Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(7):1315–1329.
6. Popovic. B. M., Ostrogonac S., Pakoci E., et al. Deep Neural Network Based Continuous Speech Recognition for Serbian Using the Kaldi Toolkit. Statistical Analysis & Data Mining, 2015, 5(3):205–217.
7. Liu X. F., Zhang X. Y., Wang Z. J. Logistic kernel function and its application to speech recognition. Huanan Ligong Daxue Xuebao/Journal of South China University of Technology (Natural Science), 2015, 43(5):100–106.
8. Pham H. T., Pham K. X., Phuong L. On the Use of Machine Translation-Based Approaches for Vietnamese Diacritic Restoration. The 21st International Conference on Asian Language Processing. IEEE, 2017.
9. Cesarini D., Calvaresi D., Farnesi C., et al. MEDIATION : An eMbEddeD System for Auditory Feedback of Hand-water InterAcTION while Swimming. Procedia Engineering, 2016, 147:324–329.
10. Lee H., Faruque M. A. A. Run-Time Scheduling Framework for Event-Driven Applications on a GPU-Based Embedded System. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2016, 35(12):1956–1967.
11. Al-Odat Z. A., Srinivasan S. K., Al-Qtiemat E. M., et al. A Reliable IoT-Based Embedded Health Care System for Diabetic Patients. International Journal on Advances in Internet Technology, 2019, 12(2019):50–60.
12. Yi Y. Design of video feature location system based on embedded technology. Journal of Interdisciplinary Mathematics, 2018, 21(5):1227–1231.

13. Hashimoto K., Yamagishi J., Byrne W. et al. An analysis of machine translation and speech synthesis in speech-to-speech translation system. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2011:5108–5111.
14. Anusuya M. A., Katti S. K. Speech Recognition by Machine: A Review. International Journal of Computer Science & Information Security, 2010, 64(4):501–531.
15. Wang Y. S. Research on part-of-speech tagging using decision trees in English-Chinese machine translation system. Computer Engineering & Applications, 2010, 46(20):99–102.
16. Fu X., Lu W., Zhu L., et al. Study of the Establishment of a Reliable English-Chinese Machine Translation System Based on Artificial Intelligence. 2016.
17. Dong H. Control System of Home Service Robot Based on Embedded Speech Recognition. Microcomputer Applications, 2017, 33, 4:15–19.
18. He X., Deng L. Speech Recognition, Machine Translation, and Speech Translation—A Unified Discriminative Learning Paradigm [Lecture Notes]. Signal Processing Magazine IEEE, 2011, 28(5):126–133.
19. Nasereddin H. H. O., Omari A. A. R. Classification techniques for automatic speech recognition (ASR) algorithms used with real-time speech translation. Computing Conference. 2017:200–207.