

## Application of speech technology in the translation system

Run Zhou<sup>1, a</sup>, Wei Xiang<sup>1, b,\*</sup>

<sup>1</sup>College of Electrical and Information Engineering, Southwest University for Nationalities, Chengdu, Sichuan, China 610041

<sup>a</sup>2872994057@qq.com, <sup>b</sup>3730544@qq.com, \*corresponding author

**Keywords:** Speech Translation; SR ; Machine Translation ; TTS ; Application;

**Abstract.** The speech is always the most convenient, most natural way of communication. In this paper, the application of speech technology in the translation system is studied. First, brief the developments and basic theories of speech technology. Then make a deep research for application of Speech Recognition(SR), Text To Speech(TTS) in translation system. Give structure diagrams of SR, TTS technologies and introduce the key modules. Finally, simply analysis the application prospects, needs and challenges of speech technology in various areas.

### Introduction

Speech Technology has gradually expanded in the field of telecommunications such as voice information service application, and quickly spread. Speech Technology has two branches, one is the technology that transform human speech into text, known as speech recognition ( SR ). And the another is to switch text into natural speech, on the contrary with SR, called Text-To-Speech ( TTS ). There is interdisciplinary area, including acoustics, phonetics, linguistics, computer science theory, information theory, digital signal processing and many other disciplines; From an implementation perspective, fast getting start with SR, because SR requirements is not as strict as TTS for speech corpus. But the SR is often influenced by environment, language, pronunciation, enunciation, which led to the recognition rate is not high, in the latter part of the study process more difficult, so that SR is not able to generally applicable, also needs to be improved. TTS, requirements for voice library is relatively high, and needs to annotation to corpus .It not only needs a lot of manpower and resources but also have higher demand on basics of language and speech [1]. In recent years, the rapid development of information processing technology in the computer language, in areas such as Speech Recognition, machine translation, Text-To-Speech have made much progresses. Since the 90, China has carried out basic research and development work related to speech translation system [2]. Using translation software for carrier, can give full play to advantages of voice technology, making the people who are with different countries, different nationalities and different languages easy exchanging the culture and technology by using translation software, and not need to complex text input or expensive third-party translation. Which provides wide application prospects and development directions for speech technology, such as SR and TTS, but at the same time, it also faces many challenges.

### Application of voice technology in the translation system

Speech Translation has always been the goal pursued for many years. Just born from the computer, people expect mutual translation between different languages using computers [3]. Speech Translation System's basic function is to translate one language into another language. In order to achieve this, the speech translation system should contain three basic parts : Speech Recognition, Machine Translation, Text-To-Speech. Speech recognition is to identify the source language input, get the text description and formal description of the source language . Machine Translation is the translation of the speech recognition result into the target language text. Text-To-Speech is based on the results of machine translation, producing the speech of target language [4], as in Fig 1 :

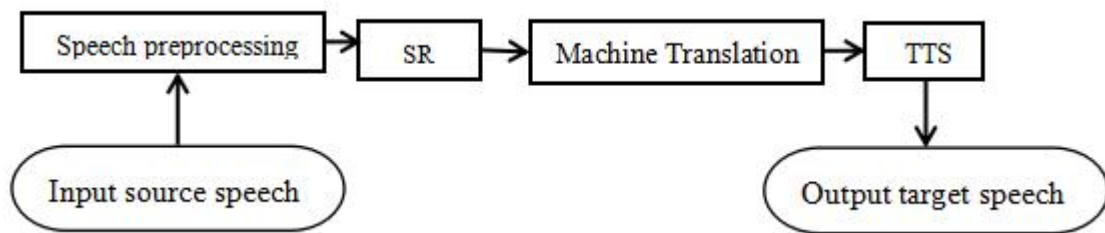


Fig 1 the system diagram of speech technology in the translation system

Increase speech processing, for the complexity of the surroundings, to filter out some of the interference or noise in the environment, improving the degree of clarity.

### Principle and analysis of key technologies

The three technologies: SR, Machine Translation, TTS are core technologies of translation system. To achieve a speech input and get another speech output, which is essential to any one of these technologies. At present, the three technologies all have been obtained great progresses and remarkable results, and have brought immeasurable convenience for human learning, travel, work and life. Below, the principles and key modules are analyzed respectively of SR, Machine Translation, and TTS in speech system.

#### SR

SR module is the foundation of the entire Speech Translation System, which is responsible for receiving speech signal after preprocessing, and makes voice signals into a speech primitive ( such as syllables, phonemes, etc ), then based on grammar and constitutive rules of speech to transform primitives into source language text stream. SR is very important for whole translation system, how the recognition accuracy directly affects the correct rate of translation of the entire system. For years, speech recognition technology has concerned by many scientists and made a significant progress in recent years at home and abroad. Such as enable achieve isolated words recognition with a large vocabulary, and the recognition rate is higher; But for continuous speech recognition, great progress has been made, but the recognition rate is far below to the recognition of isolated words. Main reasons are:

- (1) User diversity: for the reasons of gender, age, social background, dialect, channel length, different users have much different on the pronunciation of the same word.
- (2) Variability of the same users: due to the different of emotional, physical, tone, state of health, leading to the same user different pronunciation at different times.
- (3) Vagueness of Speech: paronyms, homonym, with the environment and different locations in the language stream, their voice characteristics makes a big difference, includes: volume, speed, accent, pronunciation, and ways of read.
- (4) Noise interference: although make preprocessing before speech recognition, it does not completely eliminate noise interference in speech.

Aiming at these problems, the current continuous speech recognition processing methods are proposed:

- (1) Segmentation Method: For isolated - word recognition technology has matured. Therefore, cut the speech signal flow into basic unit, then the technology of isolated words recognition can be applied to continuous speech recognition, Fig 2:

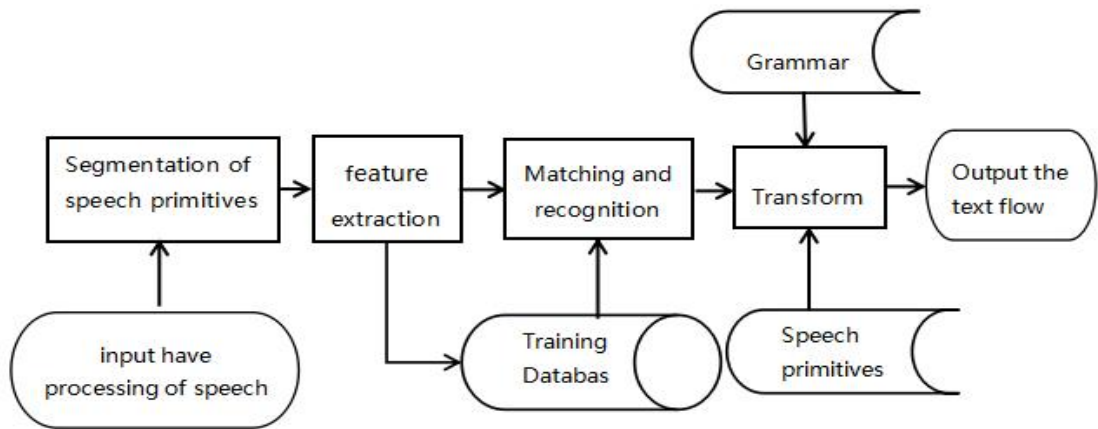


Fig 2 cutting method of SR

This approach applicable to the Chinese monosyllabic, for Western languages such as English, segmentation error rate is high, because more linking phenomena in English and other western languages, which cannot be cut down only by voice.

(2) The way of keywords : first, to choose some sentence patterns, set a thesaurus. Identification of keywords in the whole sentence, once the keywords recognition is successful, will no longer consider other components, to achieve the objective of the whole sentence recognition. The method has the advantage that allowing the input sentences are not standard, but the drawback is the recognition result may not be accurate. Because of a large number of non - standard sentences in oral, the keywords method is more appropriate used in current speech Translation System .

### Machine Translation

Machine Translation module is an important part of speech translation systems. It received the results from the SR module, through understanding and analysis to sentences , combining with the knowledge of speech, making some necessary treatment. Then, use the syntax and semantic rules for parsing, forming a parse tree, eventually generates the target language, as in fig 3 :

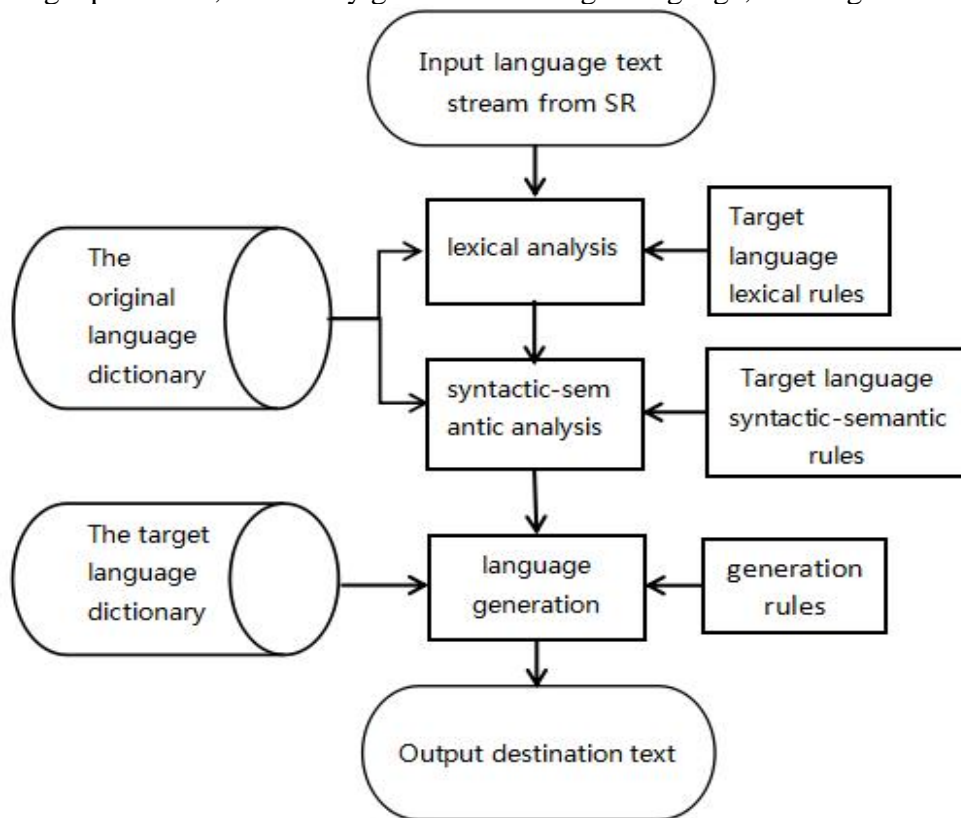


Fig 3 the schematic diagram of machine translation

## TTS

TTS module is to transform the the target language from machine translation module into the corresponding speech. According to the main design thought,the methods include synthesis rules - driven ( rule-based ) method and data - driven ( data-based ) method [5] ,As in fig 4 :

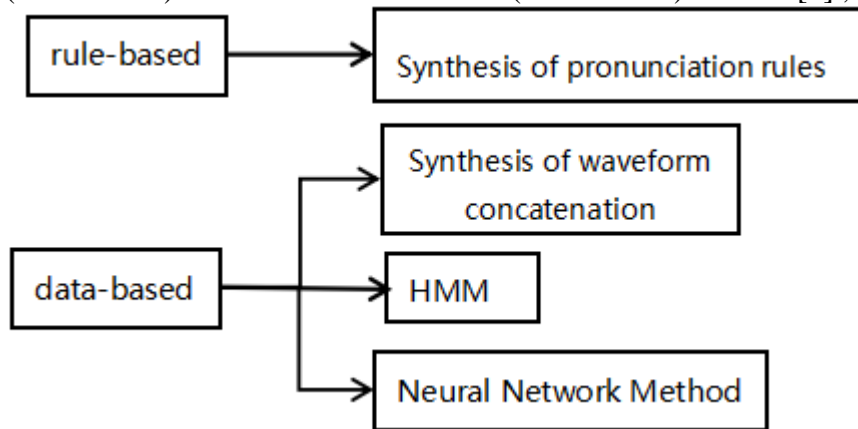


Fig 4 TTS method

The former is based on the parameters of the vocal organs and physical processes, direct simulation the pronunciation of people [6]. This method is complex, and the naturalness of synthesized speech is low, now rarely used. Waveform technology is based on text analysis information , selecting appropriate unit from prerecorded voice library ,then splicing ,getting the final synthesized speech.As the final units are copied directly from the recording sound library, the method can maintain quality of the original pronunciation of people, achieve high natural speech synthesis,as in fig 5 :

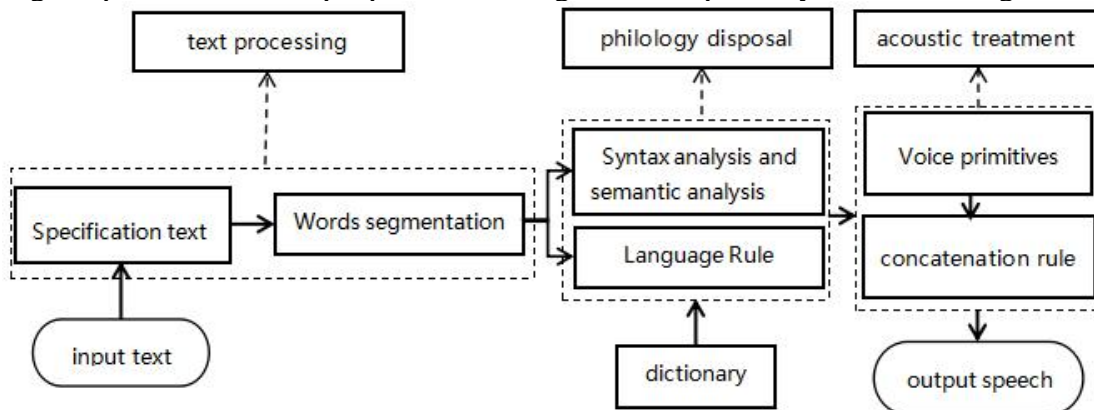


Fig 5 Waveform splicing method of TTS

The synthesis process is divided into 3 steps :

- (1) Text processing : processing the input text, remove the characters of error or unable to pronounce ;
- (2) philology disposal: depending on the text structure, composition and punctuation marks on different locations, make sure the transformation of tone and different severity ways.
- (3) Acoustic treatment : From the processing results of first two parts, select voice from the voice library primitives based on connection rules to output speech, that is synthesized speech.The advantages of this approach is synthetic speech quality high, timbre similarity well, better to learn the natural rhythm, natural high, but the disadvantage is the need for larger speech database support. Therefore, the way of waveform concatenation is more dependent on the quality ,size, and the smallest unit of speech corpora,and so on. With the development of Data Mining Technology in the computer field, many statistical methods and artificial neural network technology in the field of computer data processing application is successful. In this context, there have been based on data - driven text analysis methods at home and abroad [7]. For example : using hidden Markov Model ( HMM:Hidden Markov Model ) and neural networks method ( Neural Network Method ) [8].

## Summary and prospect

Speech technology has made great progresses in financial, education, commerce, tourism as well as other fields, and possesses broad prospects for development. Its application in translation system is more to promote mutual learning and communication between users with different languages. Especially in recent years, there are a large number of translation software with speech function, providing a language learning platform about listening and speaking. Greatly improves the efficiency of people learning a new language. Of course, speech technology is still facing many problems and challenges, such as user adaptation for speech recognition, the precision rate of consecutive translation, tone of voice process and naturalness for synthesis, etc. But as speech technology combined with other technologies, it will be gradually improved, obtain promotion and application in broader fields, certainly.

## References

- [1] Zhang Bin , Quan changqin , REN Fuji. Speech synthesis method and development summary[J]. Journal of Chinese Computer Systems, 2016, (1): 186-192.
- [2] Du Limin, Hou Ziqiang. Some scientific problems facing the Chinese speech recognition. Chinese Journal of Electronics, 1995, 23 ( 10 ).
- [3] Wei Maocheng, Zhang Sen, Zhang Fenghou. Analysis of key techniques of English - Chinese speech Translation System [ J ]. Journal of Shandong Institute of Building Materials, 1998, 12, ( 3 ) : 226-230.
- [4] Li Jing Jiao, Yao Tianshun. Medium vocabulary English - Chinese speech Translation System [ J ]. Mini - Micro Systems, 1998, 19, ( 8 ) : 44-48.
- [5] Youcef tabet , Mohamed boughazi . Speech synthesis techniques . A survey[ C ]. 7th International Workshop on Systems , Signal Processing and their Applications ( WOSSPA ), 2011: 67-70 .
- [6] Huang Nanchuan, Deng Zhenjie, Wang, Zhang haojian. Research and Development of Speech Synthesis techniques [J]. Journal of North China Institute of Astronautic Engineering, 2002, 12(3): 36-39.
- [7] SH IN YA NAKAJMIA. Automatic Synthesis Unit Generation for English Speech Synthesis Based on Multi - Layered Context Oriented Clustering [ J ] . Elsevier Science Publishers B V , 1994, 14 (4): 313-324.
- [8] ANDREJ LJOLJE, JULIA HIRSCHBERG, JAN P H VAN SANTEN. Automatic Speech Segmentation for Concatenative Inventory Selection [C] // Progress in Speech Synthesis. [ S. 1 ] : Springer Verlag, 1997: 315-321.