# Overview of the Speech Recognition Technology

Jianliang Meng, Junwei Zhang,Haoquan Zhao
School of Control and Computer Engineering
North China Electric Power University
Baoding, China
e-mail: zhangjunwei0502@163.com

***Abstract*—As a cross-disciplinary, speech recognition is based on the voice as the research object. Speech recognition allows the machine to turn the speech signal into text or commands through the process of identification and understanding, and also makes the function of natural voice communication. Speech recognition involves many fields of physiology, psychology, linguistics, computer science and signal processing, and is even related to the person's body language, and its ultimate goal is to achieve natural language communication between man and machine. The speech recognition technology is gradually becoming the key technology of the IT man-machine interface[1].The paper describes the development of speech recognition technology and its basic principles, methods, reviewed the classification of speech recognition systems and voice recognition technology, analyzed the problems faced by the speech recognition.**

***Keywords-basic principles;method; speech recognition; application***

## I. INTRODUCTION

Speech recognition is the machine on the statement or command of human speech to identify and understand and react accordingly. It is based on the voice as the research object, it allows the machine to automatically identify and understand human spoken language through speech signal processing and pattern recognition. The speech recognition technology is the high-tech that allows the machine to turn the voice signal into the appropriate text or command through the process of identification and understanding. Speech recognition is a cross-disciplinary and involves a wide range. It has a very close relationship with acoustics, phonetics, linguistics, information theory, pattern recognition theory and neurobiology disciplines. With the rapid development of computer hardware and software and information technology, speech recognition technology is gradually becoming a key technology in the computer information processing technology. Products to develop speech recognition technology is also widely used in voice-activated telephone exchange query information networks, medical services, banking services, industrial control every aspect of society and people's lives. Many experts believe that speech recognition is one of the 2000-2010 IT field ten scientific and technological developments.

## *II.* THE DEVELOPMENT PROCESS AND CURRENT SITUATION OF THE SPEECH RECOGNITION TECHNOLOGY

Speech recognition research work began in the 50's, Bell Labs speech recognition system-Audrey system first identifies the ten English digits. But it really made substantial progress, and as an important issue in conducting research in the late 60's the early 1970s. Further speech recognition in the 1980s, the HMM model and artificial neural network (ANN) are successfully used in speech recognition. 1988，FULEE Kai and others use the VQ/I—IMM method to achieve speaker-independent continuous speech recognition system-SPHINX, including 997 vocabulary. This is the first of the world speech recognition system, it is a high-performance, non-specific, large vocabulary continuous speech recognition system. People finally breakthrough of the three major obstacles, including a large vocabulary, continuous speech and non-specific. And it identified the mainstream of statistical methods and models in speech recognition and language processing.

Speech recognition system has already begun from the laboratory to practical; there have been more mature market products. Many developed countries such as the United States, Japan, South Korea, as well as IBM, Apple, Microsoft, AT&T and other well-known companies to invest heavily in research and development of practical speech recognition system [2].

## III. BASIC PRINCIPLES AND METHODS OF SPEECH RECOGNITION TECHNOLOGY

The speech recognition system is essentially a pattern recognition system, including feature extraction, pattern matching, the reference model library. Its basic structure is shown in Figure 1:

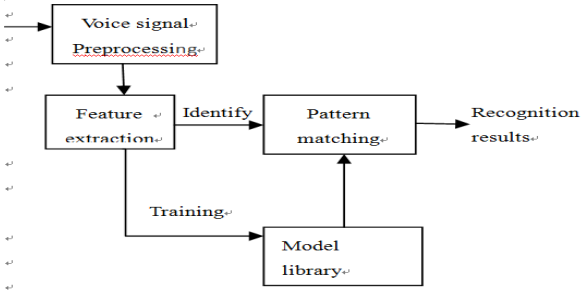


Figure 1 The basic principles of speech recognition system

978-0-7695-4789-3/12 $26.00 © 2012 IEEE
DOI 10.1109/ICCIS.2012.202

The unknown voice through the microphone is transformed into an electrical signal on the input of the identification system, the first after the pretreatment. The system establishes a voice model according to the human voice characteristics, analyzes the input voice signal and extracts the required features on this basis, it establishes the required template of the speech recognition.

Computer is used in the recognition process according to the model of the speech recognition to compare the voice template stored in the computer and the characteristics of the input voice signal. Search and matching strategies to identify the optimal range of the input voice matches the template. According to the definition of this template through the look-up table can be given the recognition results of the computer.

Representative speech recognition methods include dynamic time warping (DTW), hidden Markov model (HMM), vector quantization (VQ), artificial neural network (ANN), support vector machine (SVM) and so on.The article focuses on two methods of hidden Markov model (HMM) and artificial neural network (ANN).

### *A. Hidden Markov Model (HMM)*

As a statistical model,Hidden Markov Models Hidden Markov Model (HMM) analysis founded in the 1970s and 1980s has been the dissemination and development and successfully applied to the modeling of the acoustic signal.To the 1990s, HMM has also been the introduction of the computer word recognition and mobile communication core technology of multi-user detection. So far, it is still considered to be the most successful approach to achieve fast and accurate speech recognition system.

The HMM model parameters represent the time-varying characteristics of the voice signal. It consists of two interrelated stochastic processes common to describe the statistical characteristics of the signal. One of which is hidden (unobserved) finite-state Markov chain, and the other is the observation vector associated with each state of the Markov chain stochastic process (observable). Reveal characteristics of the hidden Markov chain depends on the signal characteristics can be observed. In this way, a certain period of time varying signals such as voice characteristics described by the random process corresponding to the symbols of state observation. Signal described by the hidden Markov chain transition probability changes with time.

HMM model in a state j under the corresponding observed values by a set of probability $b_{ik}$, k = 1,2,…, M, to describe, it is one of the M discrete countable observations, and thus known as the discrete the HMM. When the observed value of a continuous random variable X, its corresponding observed values in the state j observed by a probability density function $b_j$ (X), which became continuous HMM. Continuous HMM using the Baum-Welch algorithm to estimate model parameters applied in the estimation of $\pi$, A parameter, but the description in the estimation of $b_j$ (X) parameter must be a certain limit can be established. Current most widely used is the Gaussian $b_j$ (X) it can be represented using the following formula [3]:

$$b_j(x) = \sum_{k=1}^{k} c_{jk} b_{jk}(x) = \sum_{k=1}^{k} c_{jk} N\left(x_{\mu_{jk}}, \sum_{jk}\right)$$

$$1 \le j \le N$$

Among them, the $N(X, \mu_{jk}, \Sigma_{jk})$ for multi-dimensional Gaussian probability function, $\mu_{jk}$ mean vector, $\Sigma_{jk}$ side difference matrix, k is the $b_j(X)$ the number of mixed probability, $c_j(X)$is the combination coefficient, and

$$\sum_{k=1}^{k} c_{jk} = 1$$

HMM is a more complete expression of acoustic model of the voice,and it uses statistical methods of training the underlying acoustic model and the upper voice model into the unified voice recognition search algorithm can obtain better recognition results, and can be used for continuous speech recognition, but the drawback is the need to be very sophisticated calculations and a longer training sequence[4].

### *B. Artificial Neural Network (ANN)*

Artificial neural network ANN (based Artificial Neural Networks), analogous to the way biological nervous systems process information, using a large number of simple processing units connected in parallel to form a complex information processing system. This system has the training, highly parallel, rapid judgment, fault tolerance features applies voice signal processing. Speech recognition neural networks are usually divided into two categories, a class of neural networks or neural networks with the traditional HMM, the DP combination of hybrid network, the other is the establishment of the auditory neural network model based on human auditory physiology, psychology research.

Neural network model that more commonly used and has the potentiating of speech recognition mainly include single-layer perception model, multi-layer perception model, Kohonen self-organizing feature map model , radial basis function neural network , predictive neural network etc. In addition, in order to make the neural network reflects the dynamic of the speech signal time-varying characteristics, delay neural network, recurrent neural network and so on.

Artificial neural network technology in voice recognition applications mainly the following aspects:

*a) Reduce the modeling unit, generally in the phoneme modeling to improve the recognition rate of the entire system by improving the recognition rate of phonemes.*

*b) Depth study of the acoustic model, the auditory model, the brain operation mechanism, the introduction of context information, in order to reduce the impact of changes in voice more than the speech signal.*

*c) Extracted from the speech signal in a variety of features, a hybrid network model (HMM + NN), and apply a variety of knowledge sources (phonemes, vocabulary, syntax and meaning of the word), for voice recognition to understand the research, to improve system properties*[5].

Speech recognition using artificial neural network technology, including e-learning process and the speech recognition process, shown in Figure 2. The network

learning process is to known speech signal as a learning sample, self-learning neural network, and ultimately a set of connection weights and bias. The speech recognition process is to test the voice signal as network input, the recognition results obtained through the network of associations. The key of tthese two processes is to strike a speech characteristic parameters and neural network learning.

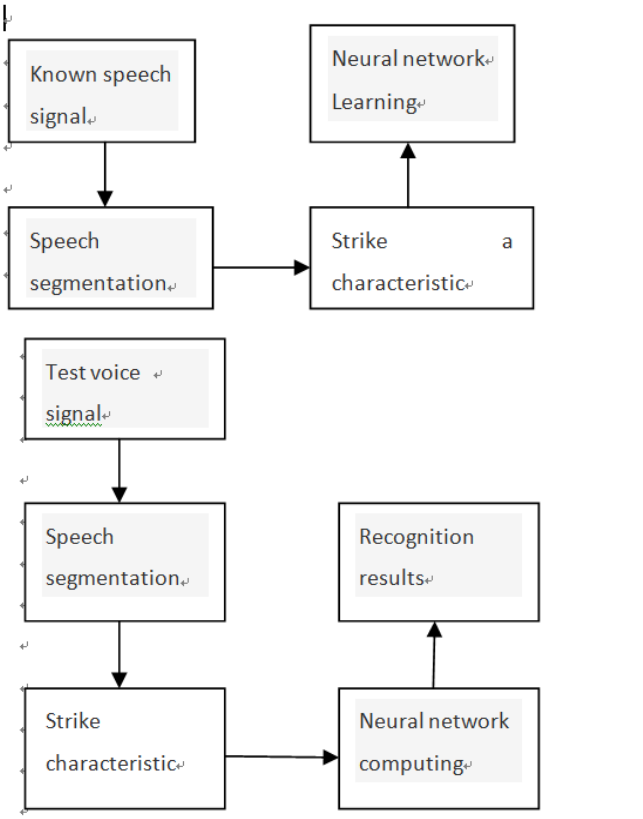


Figure 2 Artificial neural network speech recognition process

The application of artificial neural networks in the field of speech recognition has been greatly developed in recent years, artificial neural networks in speech signal processing can be divided into the following areas: firstly, improve the performance of artificial neural networks. Secondly, artificial neural network has been developed method combines a hybrid system. Thirdly, explore the use of newly emerging or widespread concern mathematical methods constitute the unique nature of the neural network, and applied to the field of speech signal processing [6].

The application of artificial neural networks in speech recognition has become a new hotspot. Artificial neural network technology has been successfully applied to solve pattern classification problems, and was shown to have enormous energy, we can predict that in the last decade, artificial neural network-based speech recognition system products will appear in the market, people will adjust their own way of speaking to accommodate a variety of recognition system.

## IV. APPLICATION OF SPEECH RECOGNITION TECHNOLOGY AND THE FACING PROBLEMS

### *A. Application of speech recognition technology*

The world to speed up research and development of speech recognition applications, there are some practical speech recognition system put into commercial operation.

The typical speech recognition system-VRCP system developed by AT&T in 1992 .The system is five words (collect, person, third number, the operator and calling card), non-specific small-vocabulary speech recognition system, has been used in AT&T Communications online, you can achieve the automatic operator-assisted call, instead of the operator completed five kinds of call type.

In September 1996, Charles Schwab launched the first large-scale commercial speech recognition application systems: the stock quotation system. The system was also the first in the financial field speech recognition system. The system is effective to improve the quality of service and customer satisfaction, and reduce call center costs. Soon, Schwab opened the speech of stock trading system.

Departments in major U.S. telecom operator Sprint PCS has the largest digital wireless network，at the same time，known for excellence and innovative customer service.The opening voice-driven systems for clients since 2000. The system provides customer service, voice dialing, check number, and change addresses and other services. In addition, China Telecom has launched a voice recognition integration of value-added services system CELL-VVAS, (VOICEVALUE-ADDED SYSTEM), the system uses a distributed excellent recognition engine ,developed a stable and efficient application. The system also perfectly integrated telecommunications switching network application to provide users with a variety of user-friendly, personalized service[7].

Another development branch of speech recognition technology is the development of the telephone voice recognition technology, Bell Labs is a pioneer in this regard, the telephone voice recognition technology will be able to telephone inquiries, automatic wiring, as well as some specialized operations, such as tourist information and other operations. After the bank use the voice query system of speech understanding technology, it can provide customers with 24-hour Phone Banking Service. Securities industry, using telephone speech recognition audio system, then, the user would like to query market could speak out the stock name or code system to confirm the user's requirements, will automatically read the latest stock price, which will greatly facilitate the user . In the 114 directory assistance artificial voice technology, you can let the computer to automatically answer the needs of users, and then playback the phone number of the query, thus saving human resources.

### *B. The facing problems*

At present, speech recognition research progress has been slow, mainly in theory has been no breakthrough. Although a variety of new amendments continue to emerge, but also the lack of general applicability. Mainly in:

Poor adaptability of the speech recognition system is mainly reflected in the dependence on the environment, If you collected speech training system in certain circumstances, the system can only be application in this environment, otherwise the system performance will be a sharp decline, anther problem is that this system does not respond correctly for the error input of users. Additionally, the progress of speech recognition in noisy environments is very difficult, because at this time people's pronounce varies greatly , like voice, slow speech rate, pitch and formant changes, which is the Lombard effect, must find a new signal analysis and processing approach.

Understanding of the human auditory comprehension, the accumulation of knowledge and learning mechanism and system of the brain control mechanism is still unclear, and secondly, the existing achievements of this aspect is used in speech recognition also remains a difficult process.

## V. CONCLUSIONS

From the problems faced by the speech recognition, speech recognition systems in order to be widely used still have a lot of areas for improvement. However, it is foreseeable in the near future that, with the voice recognition technology continues to progress, the speech recognition system will be more in-depth, the application of speech recognition systems will be more extensive[8]. A variety of speech recognition systems will appear in the market, people will adjust their speech patterns to adapt to a variety of recognition system Human beings in the short term is also impossible to create a people comparable to the speech recognition system, to build such a system is still a big challenge facing humanity, we can only forward step by step direction to improve the speech recognition system.

## REFERENCES


[1] Yu Tiecheng. The current development of speech recognition [J]. Communication World, 2005.

[2] Ren Tianping. Application of speech recognition technology [J]. Henan Science and Technology, 2005.

[3]L A Liporace.Maximum Likelihood for Multivariate Observation of MarkovSources. IEEE.Trans. IT, 1982, 28(5): 729-734

[4] Zhang Ping, Zhang Qiong. Based on HMM and BP neural network for speech recognition [J]. Cross-century, 2008.

[5] Yin Peng, Li Tao, Wang Haibing.Intelligent neural network system composed of the principle in speech recognition.Mini-Micro Systems,2000,21(8):836-839.

[6] Jiang Ming Hu, in the Yuan Baozong, Lin Biqin. Neural networks for speech recognition research and progress. Telecommunications Science,1997,13(7):1-6.

[7] Huang Shan. Voice recognition systems in the telecom prepaid business applications [J]. Information Science, 2010.

[8] Yangshang Guo, Yang Jinlong. The speech recognition technology overview [J]. Computer, 2006.