Computer Intelligent Recognition Image Speech Signal Application

Xiaowei Ma, Hongtao Deng[*]

School of Electronic Information, Qingdao University, Qingdao, Shandong Province 266000 China

Abstract: With the rapid advancement of speech recognition technology, the intelligent speech question and answer system has gained more and more attention. For the traditional question answering system, the system usually matches according to the user's input questions according to the established template rules, or searches in a large information database to get the answers to the corresponding questions that have been edited, and then the answers will be obtained. Returning to the target user, the deep question-based intelligent question answering system learns from a large number of questions and answers in the data set, extracts features, and generates corresponding answers for the user according to the questions input by the user. The intelligent voice input system is an intelligent intelligent system for automatic voice input through voice signal recognition and tone analysis. The intelligent voice input system plays an important role in voice signal analysis and speech recognition. In this paper, this paper introduces the existing speech synthesis system implementation method and speech recognition technology, then analyzes the existing methods, and proposes an improved method of server-based intelligent speech control system, and the system is implemented. The focus of inquiry and introduction has improved the shortcomings of traditional methods.

1. Introduction

With the continuous development and progress of science and technology, many new technologies have been applied in various fields, which has greatly promoted the development of modern society. Speech recognition technology, as one of the hottest advanced technologies in recent years, involves many disciplines such as signal processing, language, psychology and computer [1-4]. A large number of voice recognition products have entered the market and service fields and are widely used in smart terminals. Promote the rapid development of car voice, intelligent customer service, smart home, voice courseware and other products. All walks of life have also taken this opportunity to vigorously innovate and develop products and services combined with artificial intelligence technology to enhance user experience [5-8]. Affected by technological factors, computer-based electronic products have invaded the whole society. The development of modern society is inseparable from the information represented by information technology [9,10]. The emergence of intelligent speech technology has largely changed this way of information transmission and played a non-negligible role in many fields of society. The rational application of intelligent voice image recognition technology is conducive to the advancement and development of information technology [11]. The intelligent speech image speech signal recognition system includes speech recognition, speech interaction, natural language processing and the like. Image speech signal recognition technology often involves pattern recognition, signal processing, artificial intelligence, auditory mechanism and mechanism, probability theory and information theory. In general, the controller module, the voice recognition module, the power amplifier module, and the speaker constitute the smallest system in the voice recognition system, and the voice module receives the voice wake-up function or the voice recognition command sent from the upper computer, and the internal identification module collects the microphone. The data is converted into a corresponding recognition result and transmitted back to the controller through the communication interface.

2. Analysis of Computer Intelligent Recognition Technology

2.1. Speech Synthesis Technology

As a branch of intelligent recognition technology, speech synthesis technology includes acoustics, linguistics and multimedia technology. It is a technology that produces related syllables and sentences artificially according to presupposed procedures or instructions. The first step of speech synthesis is to deal with the analysis and understanding process of human natural language, which includes text collation, lexical segmentation, grammar and semantic analysis, so that the computer can fully understand the text information. The second step is to decode and extract prosodic features. Its function is to design and plan audio features such as length, intensity and tone for speech synthesis. The synthetic speech can correctly express semantics and make the expression more natural. Finally, speech synthesis is carried out to obtain audio output. There are a large number of speech signal processing algorithms, which are mainly composed of text analysis, prosody modeling and speech synthesis. The understanding of speech synthesis technology needs to start from two aspects. One is the application field of speech synthesis technology, which is mainly manifested in the following aspects. (1) It can be used in telephone inquiry system. Nowadays, government departments and information service industry all use unattended information inquiry system. After using voice synthesis technology, users can use telephone terminals as information inquiry tools. (2) It can be used in e-book reader. In the field of education, especially in English teaching, the emergence of e-book reader has provided many students with a new way of learning. Speech synthesis technology has prompted hand-held devices to abandon the huge display screen and convert it into e-reader. (3) Speech signal can be used in communication network. Because speech signal will occupy a lot of network resources in the transmission of network, the combination of speech signal technology can make both ends of communication use sound signal and save network resources. (4) It can be used in telecommunication value-added services and other industries to improve the applicability of the system.

2.2. Speech Recognition Technology

As another branch of intelligent recognition technology, speech recognition technology is closely related to signal processing science, linguistics, phonetics and neurobiology. The technology of conceptual understanding is to convert the received speech signal into the corresponding text or command when the machine is used for understanding and recognition. Different from speech synthesis technology, speech recognition technology is developing in two directions. One is the continuous speech recognition system applied to computer dictation system and the combination of telephone and Internet. These systems have the common characteristics of large vocabulary and need to be combined with computer operating system. Speech signal will be recognized by speech recognition system through the following links: (1) Speech signal preprocessing, including speech signal sampling, speech enhancement, endpoint detection, anti-aliasing filtering, noise reduction and so on. Endpoint detection is an important part of this part. Its purpose is to recognize speech vocabulary. The starting and ending points. (2) The purpose of feature extraction of speech signal is to extract some data information which can describe the characteristics of speech signal from the waveform of speech signal, such as resonance peak, energy mean, cepstrum coefficient, linear prediction coefficient, zero-crossing number, etc. The selection of parameters has an important relationship with the efficiency of speech recognition. (3) Training and building voice template, training is to digitize a large number of voice signal templates and voice signal databases prepared in advance, dig language information and build voice signal instruction library. At present, voice cloud data platform is used by many researchers, and it provides "cloud" in mobile terminal speech recognition and other functions. Computing is an advanced technology. (4) Pattern matching, the key point of speech recognition system is pattern matching. Its function is to analyze and calculate the distortion between the speech to be tested and the corresponding template in the pattern library according to certain standards, and output the most matched template as the recognition result.

3. Image Speech Signal Analysis Algorithm Model

Image speech signals are subject to tone changes during transmission to produce speech attenuation and distortion, resulting in poor accuracy of tone recognition. In the traditional method, the automatic voice recognition system for intelligent voice recording system mainly adopts multi-resolution fuzzy logic control method, combined with artificial intelligence control and feature extraction to improve the automatic recognition performance of intelligent voice recording system, but the traditional method is used for voice tone recognition. There are problems such as weak anti-interference ability and excessive computational overhead. Deep learning is a multi-level machine learning algorithm that performs characterization learning on very complex interrelationships between data and data. There are many ways to describe observations. For example, pixels can be described by matrices, and the ease with which algorithms can be modeled can vary from one method to another. The goal of deep learning is to be able to find better descriptions while building better models for training. Deep learning is a hierarchical structure abstracted by researching a human brain with a complex structure. The input sample data is extracted layer by layer to extract data features, so that each layer can be more abstract than the previous layer. Said. Deep learning The feature extraction of data such as text and sound is achieved by using many layers of nonlinear processing techniques. In this regard, based on wavelet multi-layer decomposition and wavelet entropy feature extraction, the intelligent voice recording system automatically recognizes the tone. The wavelet function decomposition method is used to denoise the tonal image speech signal of the intelligent speech input system, and the image speech tonal signal of the noise reduction output is characterized and adaptively filtered. In order to realize the automatic recognition of the tone of the intelligent voice recording system, the speech signal analysis method is used for feature extraction, the extracted voice tonal features are adaptively matched, the speech signal recognition and multi-scale decomposition are realized, and the multi-wavelet decomposition method is used to characterize the speech signal. Decomposition, according to the pronunciation position, vowel category and audio characteristics of the voice input, the pronunciation feature is discriminated and the tone recognition is realized. According to the above analysis, the block diagram of the realization of the automatic tone recognition model of the intelligent voice recording system designed in this paper is shown in Figure 1.
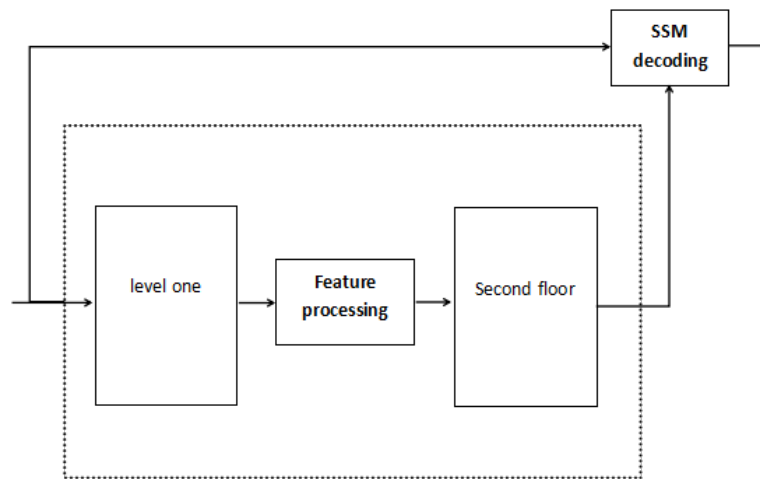

Figure 1. Schematic diagram of automatic recognition of tone in intelligent voice recording system

According to the structure diagram of the implementation principle shown in Fig. 1, firstly, the tone signal analysis of the intelligent voice input system is performed, and the original input voice signal feature sequence is x=[x(0),...,x(N-1)], where x (n) For a finite-length discrete speech signal, $0 \le n \le N-1$, then the discrete Fourier transform (DWT) of the tonal feature sequence of the intelligent speech entry system of x is defined as follows:

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(-j \frac{2\pi}{N} nk\right) \qquad (1)$$

Where $0 \le k \le N-1$, indicating the length of the speech signal, and after the signal x(n) is subjected to discrete orthogonal wavelet transform, X=DFT{x} is used to represent the finite time series x of the tonal characteristics of the intelligent speech input system. DFT. That is, the formula (2) is expressed.

$$X=[X(0),\ldots,X(N-1)] \qquad (2)$$

Based on the wavelet function decomposition method, the noise reduction processing and filtering of the tonal voice signal in the intelligent voice recording system are carried out. The automatic tone recognition algorithm design, wavelet multi-layer decomposition and wavelet entropy feature extraction are used to automatically identify the tone of the intelligent voice input system.

4. Application of Intelligent Recognition Image Speech Signal Industry

4.1. Life Area

Intelligent speech system is a new type of educational tool used in language learning in education. China's progress is universally visible, and all walks of life have undergone earth-shaking changes. However, it is indisputable that we are still slightly lacking in technology compared to Western countries. If you want to constantly improve yourself, you need to master the world-wide technical communication language - English. The intelligent voice system has a standard reading function that allows the user to master the most standard pronunciation, and the learning process seems to be in a real language environment. medical field. As we all know, the issue of drug safety is related to the life and property of citizens, and it is an important guarantee for the normal life of the residents and the long-term stability of the society. Advances in science and technology have increased the efficiency and precision of the medical field. In terms of drug safety, the intelligent voice system monitors the identification and packaging of drugs, and uses voice modules to provide voice broadcast feedback, supervise drug safety, and provide protection for citizens' health. It is worth mentioning that the voice system is applied in medical services. In today's society, the relationship between doctors and patients is relatively tight, and intimate services can create a good social image for the hospital and greatly ease the relationship between doctors and patients. The artificial intelligence robot used by the medical department also utilizes the intelligent voice interaction module to provide medical practitioners with services such as appointment registration, recharge payment, and outpatient service. Intelligent voice systems play an indispensable role in the medical field.

4.2. Corporate Domain

Intelligent voice systems are closely related to artificial intelligence technology. The application market for chat bots and virtual personal assistants also showed an explosive growth trend. Due to the particularity of certain company industries, virtual private assistants provide a wide range of application space and development prospects. Taking banks as an example, it is well known that the large number of employees is one of the characteristics of banking units. At the same time, it is known for its cumbersome workload, high repetitiveness and numerous documents. The characteristics of the banking unit have created a market demand for virtual private assistants. The counter business can be represented by a virtual personal assistant. The application of virtual personal assistants will greatly improve the efficiency of the banking department and the business processing cycle, and realize the automation and digitization of the process.

4.3. Military Field

The key speech recognition and speech synthesis technology in the human-machine interface

helps the operator to open the binding of the operating keyboard as much as possible, and directly uses the voice input command to complete a series of operations, which is very convenient to retrieve from the database. Resource information required. The intelligent and anthropomorphic devices used in the communication process make the operating system more efficient and convenient, saving time and improving efficiency for human beings. The human voice is as unique and irreplaceable as our fingerprints. No two individuals will have the same sound, just as ten fingers do not have the same texture. Fingerprint recognition systems are widely used in computers, mobile phones, vehicles, and identity authentication. Voiceprint recognition technology in speech recognition systems is also applied to computer information security. In addition to our common USB encryption key, voiceprint authentication technology can more effectively ensure that secret data in the computer is not maliciously deleted, destroyed, copied, or leaked. At the same time, the vital military site applies voiceprint recognition to the access control system and enters legal instructions into the system to avoid malicious intrusion by illegal elements.

5. Conclusion

Intelligent voice control frees the user's hands and is gradually changing the user's usage habits, bringing a new revolution to the Internet market. The future value of intelligent voice lies in user data mining, as well as the content behind it and the service. The Internet of Things era with voice as the gateway will create new business models. The traditional human-computer interaction is to realize the operation and control of the machine through the sense of touch, but this operation has great drawbacks. Inconvenient operation and time consuming, especially for people with visual impairments. As the pace of society continues to accelerate, people want to use a more convenient way to replace the manipulation of the machine by touch. In this context, the emergence of intelligent voice systems reflects its advantages, and more and more smart devices use this intelligent voice technology to achieve control. The intelligent voice control system is roughly divided into two parts of a speech synthesis system and a speech recognition system. The rational application of intelligent recognition of voice technology is conducive to the advancement and development of information technology.

References

[1] Zinchenko K, Wu C Y, Song K T. A Study on Speech Recognition Control for a Surgical Robot[J]. IEEE Transactions on Industrial Informatics, 2016, PP(99):1-1.

[2] Alimuradov A K, Murtazov F S . Methods to Improve the Efficiency of Recognition of Speech Signals in Voice Control Systems[J]. Measurement Techniques, 2016, 58(10):1107-1112.

[3] Ding I J, Shi J Y. Kinect microphone array-based speech and speaker recognition for the exhibition control of humanoid robots [J]. Computers & Electrical Engineering, 2016:S0045790615004395.

[4] Alvarez A G, Evin D A, Verrastro S. Implementation of a Speech Recognition System in a DSC [J]. IEEE Latin America Transactions, 2016, 14(6):2657-2662.

[5] Ishimitsu S. Speech recognition method and speech recognition apparatus[J]. The Journal of the Acoustical Society of America, 2001, 109(1):30.

[6] Laszlo T, Ildiko H, Gabor G, et al. A Speech Recognition-based Solution for the Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech[J]. Current Alzheimer Research, 2018, 15(2).

[7] Khademian M, Homayounpour M M. Monaural multi-talker speech recognition using factorial speech processing models[J]. Speech Communication, 2018, 98:1-16.

[8] Changsu L , Youngjoong K . Spoken Language Understanding with a Novel Simultaneous Recognition Technique for Intelligent Personal Assistant Software[J]. International Journal on

Artificial Intelligence Tools, 2018, 27(03):1850009-.

[9] Simone H, Felix W, Richard K, et al. I Hear You Eat and Speak: Automatic Recognition of Eating Condition and Food Type, Use-Cases, and Impact on ASR Performance[J]. PLOS ONE, 2016, 11(5):e0154486-.

[10] Humes L E, Kidd G R. Speech recognition for multiple bands: Implications for the Speech Intelligibility Index[J]. The Journal of the Acoustical Society of America, 2016, 140(3):2019-2026.

[11] Qi B, Liu Z, Gu X , et al. Speech recognition outcomes in Mandarin-speaking cochlear implant users with fine structure processing[J]. Acta Oto-Laryngologica, 2017, 137(3):7.