Chinese Language Information Processing Considering Efficient Decoding Algorithm for Phonetic Conversion

Fengying Wang*
Weinan Normal University, Weinan City, Shaanxi, China, 714000
*E-mail: wanqiu0122@163.com

Abstract. For Chinese language information processing system, it is a very typical 3C system, which integrates computer and communication into one. Connect smart devices through the application of various technologies, mainly including network transmission technology, audio and video technology, information processing technology, etc. and use a highly efficient management system to uniformly manage and control all devices. According to the actual situation of human communication, the most direct way is language communication. Dedicated voice recognition chips are mainly used to control and control smart devices in traditional smart home systems. The currently recognized voice recognition products usually rely on the Internet as an important means to improve their efficiency.

Keywords: Phonetic Conversion, Efficient Decoding, Chinese Language

1. Introduction

The advent of AI technology has brought great progress and development to the modern society and economy. At this stage, speech recognition technology needs to be further improved in accuracy and speed and deep learning still stays at the theoretical level of research and has not been applied to actual products. In addition, deep learning technology can integrate cloud computing and data resources to improve the recognition speed and accuracy of data resources. In order to promote the development of speech recognition technology to a higher level, it needs to be carried out from the aspects of speech collection and feature selection, simulation training and speech recognition. In this way to achieve efficient audio conversion and voice information processing.

2. Analysis of phonetic conversion technology

2.1. Classification of phonetic conversion system

The phonetic conversion technology has a variety of classification methods, the common ones are
classification according to the recognition object, whether it is for a specific speaker and the vocabulary size method of the speech to be recognized. According to the targeted speaker, the phonetic conversion technology can be divided into specific human phonetic conversion and non-specific human phonetic conversion[3]. A specific person's phonetic conversion can only recognize the voice of one or several people, while a non-specific person's phonetic conversion can recognize any person's voice. The non-specific human voice conversion system is more in line with actual needs, but it is more difficult to achieve. According to the vocabulary size of the speech to be recognized, it can be divided into a small vocabulary, a medium vocabulary and a large vocabulary phonetic conversion system. The recognition range of the small vocabulary phonetic conversion system is dozens of words, the recognition range of the medium vocabulary phonetic conversion system is several hundred words to a thousand words and the recognition range of the large vocabulary phonetic conversion system is a few thousand to several Ten thousand words, its future market prospects are shown in the figure below.

![Figure 1. Future market prospects](image)

2.2. The realization principle of the phonetic conversion system

The specific implementation details of different phonetic conversion systems are different, but the basic technical principles are similar. A basic phonetic conversion system mainly includes four parts: phonetic conversion unit selection, feature extraction technology, pattern matching criterion and model training technology.

Selection of phonetic conversion unit. The phonetic conversion unit has three types: word sentence, syllable and phoneme[3]. Feature parameter extraction technology. The purpose of feature parameter extraction is to extract useful information for phonetic conversion from speech signals, that is, to remove redundant information that is not related to recognition and to retain information related to recognition[4]. Hidden Markov model describes the statistical characteristics of a signal by two random processes related to each other, one of which is a hidden (unobservable) chain with finite states and the other is an observation vector associated with each state of the chain Random process (observable). The development of artificial neural network provides a new solution to the realization of phonetic conversion. Artificial neural network simulates the principle of human neuron activity and has the ability of self-learning, association, comparison, reasoning and generalization. These capabilities are not available in hidden Markov models, but artificial neural network does not have the dynamic time
warping performance of hidden Markov models. Therefore, if the advantages of the two are combined, the robustness of the entire model can be improved[5].

3. Decoding algorithm overview

PCM (Pulse Code Modulation), pulse code modulation. The human ear hears analog signals and PCM is a technology that converts sound from analog signals to digital signals. The principle is to sample the analog signal with a fixed frequency. The sampled signal looks like a series of continuous pulses with different amplitudes on the waveform (pulse-like electrical fluctuations) and the amplitude of these pulses Quantization is performed with a certain precision and these quantized values are continuously output, transmitted, processed, or recorded in a storage medium, all of which constitute the production process of digital audio (three processes of sampling, quantization and encoding). Voice coding is to encode analog voice signals and convert the analog signals into digital signals, thereby reducing the transmission bit rate and performing digital transmission. The basic methods of voice coding can be divided into waveform coding and parametric coding[6]. Parametric coding is based on the pronunciation mechanism of human language, finds the characteristic parameters that characterize speech and encodes the characteristic parameters.

4. Chinese language information processing based on efficient decoding algorithm for phonetic conversion

4.1. Voice collection and feature selection

In the process of collecting voice information resources, the system needs to go through sampling and endpoint testing. First of all, the sampling link usually relies on the sound card to achieve the induction and integration of voice information resources and the received information is processed accordingly to make it digital. Secondly, the end-point testing process usually verifies the initial and final stages of the appearance of voice information and is an important preparation before information resource processing. Under normal circumstances, the realization of the endpoint detection function is based on the algorithmic method of linking energy and zero-crossing rate. In addition, to ensure that the speech recognition system can work normally, it is also necessary to solve the problem of information feature screening. The accuracy and speed of the information feature screening results directly affect the accuracy and efficiency of speech recognition. This is because only the information features are realized. Accurate screening can distinguish the huge voice information, extract useful information and exclude incomplete and useless information at the same time, thus ensuring the orderly operation of the collection of voice information resources. The recognition success rate is shown in the following figure.
4.2. Simulation training and speech recognition

This part focuses on two aspects, namely training and recognition. The first training is to select an appropriate recognition path according to the requirements of the system and obtain relevant information of these Chinese characters through training and clustering measures. At the same time, these information resources are stored in the system as initial application values. When discriminating data information, the Chinese character resource to be discriminated must be stored in the system file to establish a reference template resource library. Secondly, the recognition mainly compares the text to be recognized with the text information in the reference template database and determines the content of the recognized text according to the similarity of the comparison. There are currently three main recognition methods: the language obtained in the simulation training phase. The information is based on and stored in the system. When text recognition is to be performed, the text with the smallest score and distance is selected; the method of scoring is used to calculate the text information to be recognized through the state graph model to obtain the output probability; The differences in language features are appropriately divided and then the average value in each regional feature is calculated.

5. Conclusion

In summary, the application of AI technology has an important role in promoting the rapid development of China's social economy. It is necessary to strengthen the research and development of intelligent robot technology and promote it to the research of speech recognition and other fields, to promote the innovation and progress of speech recognition technology and ultimately drive innovation in all walks of life in society.

References

[1] Nakagome Sho,Luu Trieu Phat,He Yongtian, Ravindran Akshay Sujatha, Contreras-Vidal Jose L. An empirical comparison of neural networks and machine learning algorithms for EEG gait decoding. [J]. Scientific reports, 2020, 10(1).
[2] Tsofun Algorithm Ltd.; Patent Issued for Device, System And Method Of Implementing Product Error Correction Codes For Fast Encoding And Decoding (USPTO 10,547,332)[J]. Computers, Networks & Communications, 2020.

[3] Abbaspour Sara, Lindén Maria, Gholamhosseini Hamid, Naber Autumn, Ortiz-Catalan Max. Evaluation of surface EMG-based recognition algorithms for decoding hand movements.[J]. Medical & biological engineering & computing, 2020, 58(1).

[4] Sara Abbaspour, Maria Lindén, Hamid Gholamhosseini, Autumn Naber, Max Ortiz-Catalan. Evaluation of surface EMG-based recognition algorithms for decoding hand movements.[J]. Medical & Biological Engineering & Computing, 2020, 58(4).

[5] Andrew Allman, Wentao Tang, Prodromos Daoutidis. DeCODe: a community-based algorithm for generating high-quality decompositions of optimization problems[J]. Optimization and Engineering, 2019, 20(4).

[6] Lucky Galvez, Jon-Lark Kim. Projection Decoding of Some Binary Optimal Linear Codes of Lengths 36 and 40[J]. Mathematics, 2019, 8(1).