Design and Implementation of Speech Recognition System Based on Artificial Intelligence

Yu Li

917586724@qq.com

School of International Relations, No. 12 Poshang Village, Haidian District, Beijing, 100094, China

Abstract. Speech recognition technology based on artificial intelligence AI technology has gradually been widely used under the background of the continuous development of current data processing technology, and belongs to the branch of biometrics technology. AI-based speech recognition technology can be applied to biometric recognition and matching in different fields, providing technical support for the intelligent development of various industries and fields. In car intelligent voice recognition system, for example, for the car voice recognition system speech recognition real-time and accuracy requires high characteristics, the article adopts multiple pickup as car voice system sound acquisition hardware system, the voice data pretreatment after feature extraction, in artificial intelligence AI technology and sound biometric matching model and alarm sensitive words. In addition, the sensitive word recognition transmission system is connected to the cloud platform of the public security alarm system to realize intelligent voice recognition alarm. In addition, the dialect speech library of different regions was established, 60 experimental subjects from different regions were selected, and 5 alarm sensitive words were spoken in dialect for the experiment. The results showed that the speech recognition accuracy of the artificial intelligence speech recognition alarm system designed in this article was more than 95%, which was consistent with the real alarm intention of the experimental subjects.

Keywords: artificial intelligence AI; voice recognition; on-board alarm system; system framework

1 Introduction

With the development of artificial intelligence technology, the accuracy speed of speech recognition has been significantly improved, and it has been applied to the development of various production industries in the society to promote the intelligent process of various industries. At the same time, with the intelligent development of China's automobile industry, in order to improve the safety and comfort of automobile driving, artificial intelligence AI technology is used to optimize the function of artificial intelligence identification and alarm system of automobiles[1]. At present, artificial intelligence speech recognition technology has been widely used in medical treatment, transportation, intelligent customer service, tourism and other industries, and automobile speech recognition technology has also been favored by relevant researchers, and its speech recognition function has reached quite accurate technical difficulty[2]. And car speech recognition and alarm system research is lacking, the article adopts AI algorithm optimization sound acquisition and feature extraction function, realize the intelligent free, context understanding, full duplex speech recognition emotion interaction

MSIEID 2023, December 08-10, Guangzhou, People's Republic of China

DOI 10.4108/eai.8-12-2023.2344688

function, make human-machine voice interaction closer to human communication mode, improve the alarm sensitive word recognition, strengthen the safety performance of the car.

2 Design framework of vehicle voice recognition alarm system

The framework structure of the speech recognition alarm system based on AI AI is mainly composed of audio acquisition unit, gain amplification circuit, signal filtering and conditioning, AD acquisition, speech recognition, anti-disassembly alarm and voice intercom programs. The specific framework is shown in Figure 1.

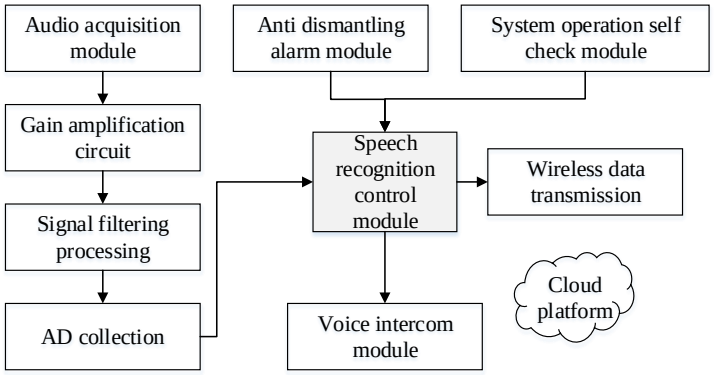

Figure 1 Voice recognition and alarm system based on AI

The main working process of the system is to use the on-board multi-channel microphone pickup to collect the sound data in real time. After the collected sound data is preprocessed through the amplification and filtering, and then the sound features are identified through the sound feature extraction system. This link is mainly through the acoustic model to establish identification library, using the way of neural network matching acoustic dictionary and decoder decoding contrast, will get the text information and the sound features of alarm sensitive word matching, if the algorithm translation and the alarm sensitive word sound features, through the signal transmission system will send alarm signal to the platform regulators and public security alarm system cloud platform, realize the alarm function[3]. In addition, the supervision platform in the background of the system can communicate with the on-board host through voice communication to ensure the accuracy of the alarm signal triggering.

3 Algorithm design

3.1 Audio signal analysis

Sound acquisition input module design is mainly distributed in the car monitoring device in front, back, left, right four positions, realize the comprehensive three-dimensional audio data acquisition, one of the sound source as the input source of audio signal analysis, the remaining three as a reference for environmental noise, its main purpose is to filter out environmental noise

to improve the signal to noise ratio of audio data acquisition, obtain more pure audio signal. Subsequently, the collected audio is analyzed by the audio signal intelligent analysis module.

1) Audio signal detection

Assuming that the collected audio data sequence is r (n), the audio signal is analyzed by formula (1).

$$R(e^{j\omega}) = \sum_{0}^{N-1} r(n) e^{j\omega n} \tag{1}$$

among $R(e^{j\omega})$ Is the discrete Fourier transform of r (n). When the distributed energy field of the spectrum is close to zero value, it is regarded as silent signal, but instead appears as sound signal.

2) Audio signal preprocessing

The detected audio signal is expressed with the formula data as

$$r(n) = \sum_{0}^{i} S(o) + \sum_{i+1}^{n} V(b) \tag{2}$$

Where S (o) represents the audio mute signal, and V (b) represents a voice signal segment where the collected audio signal removes the mute signal. This segment is used as the collected sound analysis data signal, and then the minimum mean square (Least mean square, LMS) is calculated by using the adaptive filtering algorithm[4]. In addition, the audio filtering of the ambient noise audio from the sound collector of another road, that is, the background environmental noise is defined as N (b), and the filtered pure speech segment can be obtained by formula (3) calculation, namely C (b).

$$\text{C (b)=V (b)-N (b)} \tag{3}$$

3) Feature extraction algorithm

The preprocessed audio signal is processed into frame data. Assuming that the length of each frame is S and the length of the frame movement is s, the data processing will be conducted by alternately overlapping according to S-s. Where S> 2s, the processed frame data is extracted according to the feature extraction method. The steps of the feature extraction algorithm are follows:

A) Make a Fourier transform of the audio signal in each frame;

B) the voice signals collected by the MER inversion spectrum coefficient (such as equation (4), (5));

$$\log\left|S\left(e^{j\omega}\right)\right| = \sum_{-\infty}^{\infty} C_m e^{-jwm} \tag{4}$$

$$S(w) = \lim_{N\to\infty} \frac{1}{N} \left|X(e^{j\omega})\right|^2 = \lim_{N\to\infty} \frac{1}{N} \left|\sum_{n=0}^{N-1} r(n) e^{-jwn}\right|^2 \tag{5}$$

C) Convolutional neural network is used to recognize the acquired speech signals. The structure of the convolutional neural network for processing audio is shown in Figure 2[5].

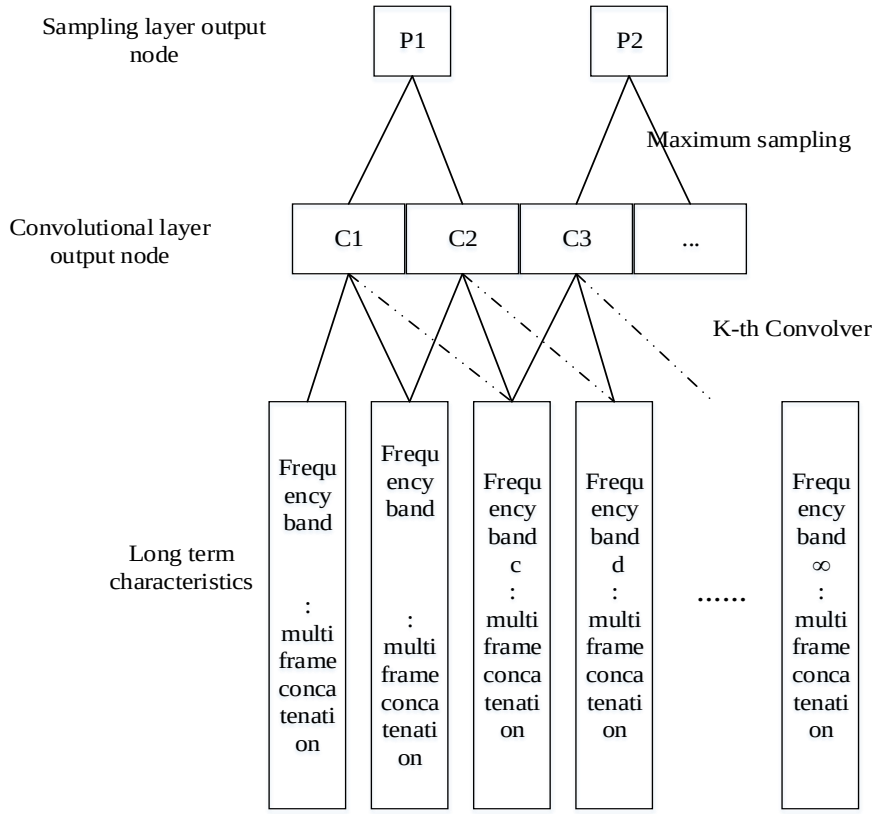

Figure 2 Structure of the audio signal processing algorithm of the convolutional neural network

In Figure 2, the structure of the convolutional layer and the speech sampling layer of the convolutional neural networks (Convolutioal Neural Networks, CNNs) are used in speech recognition. After the collected sound audio signal enters the feature extraction system, the collected audio information is taken as a two-dimensional feature input. The first bit is the time domain dimension, the second dimension is the frequency domain dimension, and the two of the physical meanings are completely different. Traditional deep neural network algorithm has better performance for multi-frame series sound attempt feature extraction in sound feature recognition system. Therefore, on this basis, the paper uses the convolutional neural network for the features to extract the extracted sound information, and connects the current frames in the collected audio and several frames to extract and identify the sound features. The physical significance of the convolutional layer in Figure 2 is to extract features and observe the local frequency domain through the convolutor, and extract local meaningful information for local convolution[6]. And the convolutor is applied on different filter bands to analyze the filter band coefficient of the current frame by the filter bands. Thus, the sound feature extraction output of the convolutional neural network can be extracted through formula (6).

$$C_{i,k} = \chi\left(\sum_{b=1}^{s-1} w_{b,k} v^{T}{}_{b+i} + a_k\right) \tag{6}$$

Among them, k represents the k-th convolutional layer; v^{T}_{b+i} represents the i-th group of sound input feature vectors; $w_{b,k}$ represents the weight parameter of the k-th convolutional layer; S represents the width of the convolver; Ak represents network bias. That is, by weighted averaging the input features of the i-th group with the k-th convolutional layer, combined with a nonlinear function β Obtain the value of the convolutional layer output node. (Note: Nonlinear function β Generally, the arctangent function or sigmoid function is chosen.)

3.2 Biometric identification model matching algorithm

The biological model covers the speech data of different regions, including Mandarin Chinese and regional dialects, and the sound database in the biological model stores N different sound model data. The sound feature recognition results in the feature extraction algorithm are matched with the sound signal features in the biological model sound database, and the specific decoding process is shown in Figure 3.

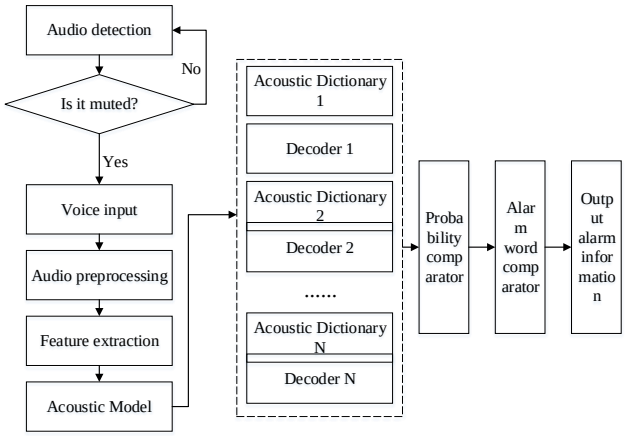

Figure 3 Design framework diagram of matching alarm system for biological model features

The sound feature matching in the biological model matches the output text characters according to the theological dictionary and decoder. If there is no consistent word in the output result feedback database, it is necessary to select the word with the largest matching probability after careful feature extraction. The alarm system will match the matching output text information with the alarm sensitive words. If the detection result of multiple times is the alarm sensitive words, the alarm system will be triggered, and then the vehicle information, location and alarm information detected by the alarm system will be feedback to the supervision platform and the cloud platform of the public security alarm system.

4 Practice results and analysis

In order to verify the effectiveness and reliability of the on-board speech recognition intelligent alarm system designed by the article, 50 volunteers from different regions were invited, and the local dialect was used as the test language of the test system. And set up five alarm sensitive

words, namely "help", "accident", "fire", "danger" and "alarm". The alarm system will trigger the alarm system after extracting the collected speech and matching the alarm sensitive words in the sound data of the biological model, or with the continuous sharp sound "ah". The 50 test objects invited in the article were tested with alarm sensitive words, and the recognition results of the on-board voice recognition alarm system were compared (as shown in Table 1).

Table 1 Sensitive word recognition of on-board voice recognition alarm system based on artificial intelligence

sensitive word	System recognition / times	precision /%
save sb.'s life	50	100
traffic accident	50	100
The fire	48	96
danger	49	98
report to the police	50	100

As can be seen from Table 1, the accuracy of the speech recognition system for alarm sensitive words is more than 96%. In addition, the alarm information recognized and matched by the system is consistent with the true alarm intention of the test object, which has high reliability.

5 Conclusion

To sum up, with the continuous promotion of the intelligent development trend of various industries, the intelligent development of the automobile industry has gradually become the trend of current research. Using artificial intelligence technology to optimize the car speech recognition system, effectively improve the safety and comfort of the car. Based on the artificial intelligence convolutional neural network algorithm of the audio data collected in the vehicle speech recognition alarm system, the purified sound audio is obtained. After using the feature extraction algorithm and compare with the sound features of the constructed biological model sound database, if the sound features of the alarm sensitive words are satisfied, the alarm system will be triggered. In this way, the driving safety of the car drivers is improved, and with the technical support of the artificial intelligence algorithm, the convolutional neural network algorithm strengthens the contrast effect of the dialect in the system, and improves the applicability of the car in different regions.

References

[1] Choudhary A ,Kshirsagar R .Process Speech Recognition System using Artificial Intelligence Technique[J].International Journal of Soft Computing and Engineering (IJSCE),2022,2(5):132-137.

[2] Lu H ,Xueqin D ,Li Y , et al.An Effective Artificial Intelligence-Enabled Error Detection and Accuracy Estimation Technique for English Speech Recognition System[J].Wireless Communications and Mobile Computing,2022,12(8):48-56.

[3] Rafael L,R.A G,Paloma V,et al.SIMO:An Automatic Speech Recognition System for Paperless Manufactures[J].Advances in Science and Technology,2023,6948 129-139.

[4] Lin Jinhong, Wu Guopei, Cai Di, et al. Development of speech recognition and voice playback system for imitation gecko wall climbing robot [J]. Mechanical Design and Manufacturing, 2023,384 (2): 219-222,227.
[5] Mahesh N T K,Ganesh K K,T.K D,et al.Group Attack Dingo Optimizer for enhancing speech recognition in noisy environments[J].The European Physical Journal Plus,2023,138(12):
[6] Widodo B.Low-Cost Cat Robot Using Speech Recognition Systems for Promotion[J]. Bulletin on Innovative Computing, Information and Control - B: Applications.,2023,14(02):117-119.