

An Improved Machine Learning Model for Pig Abnormal Voice Recognition

Ying-Ming Shi*

School of Automation Engineering, Tangshan Polytechnic College, Tangshan City 063600, Hebei Province, China
yingming5230@163.com

Received 1 November 2022; Revised 30 November 2022; Accepted 1 December 2022

Abstract. The animal's cry largely reflects its physical state, while the pig's different sound signals reflect its current physiological health and emotional state. In this paper, the pig cough is taken as the recognition object. First, the time domain and frequency domain of the pig call signal are analyzed, and the hardware system of pig audio acquisition is built. Then, the collected pig audio information is denoised. The live pig voice endpoint detection adopts the double threshold endpoint detection method, and then uses Mel frequency cepstrum coefficient to extract the features of the live pig audio signal. In the recognition of the pig cough, the hidden Markov model is used to improve the recognition accuracy and efficiency through machine learning. The experimental results show that the recognition method described in this paper can accurately identify the pig cough sound.

Keywords: hidden Markov model, feature extraction, endpoint detection, machine learning

1 Introduction

China is a big country in pig breeding and pork consumption. With the rapid development of intelligent and intensive breeding industry, precision pig breeding supported by information technology and artificial intelligence is an inevitable requirement for the sustainable development of modern pig industry. Disease prevention and control is the first priority in the process of pig breeding. The change of pig's cry is the most direct response to the change of its physiological health and growth environment. At present, in most pig farms, intelligent, automatic feeding and environmental control equipment have been basically popularized. However, in judging whether pigs' physiological health is abnormal, it mainly depends on the intuition and experience of the keepers. For large farms with high feeding density, this method not only consumes a lot of time and energy, but also often causes pig disease aggravation or even death due to human factors. Therefore, the realization of pig abnormal voice recognition monitoring can improve the breeding efficiency and reduce economic losses to a certain extent. Aiming at the problem that the current recognition of pig body abnormalities mainly depends on human experience, this paper establishes an automatic cough sound recognition system based on machine learning, and judges whether the pig is sick according to the recognized cough sound, and completes the following work:

(1) In this paper, the collected audio information is preprocessed by the methods of pre emphasis, framing, windowing combination and spectral subtraction, in order to separate the pig's voice, reduce the congestion of the voice channel, and shield the ambient sound.

(2) A double threshold endpoint detection method is proposed, which enables the start and end points of pig voice to be intercepted from the mixed ambient sound to accurately distinguish pig audio and non pig audio.

(3) An improved Hidden Markov Model is established. by adding the improved algorithm, the learning efficiency of the model and the recognition accuracy of the target sound are improved.

This paper is arranged as follows: Chapter 2 mainly introduces the relevant research achievements and shortcomings of relevant scholars, the third chapter discusses the data quantitative description method of abnormal pig voice and the recognition method of cough sound, chapter 4 establishes the simulation model of sound acquisition and recognition, and completes the simulation experiment of recognition, the fifth chapter makes a summary of this paper, and carries out the future prospects and further research arrangements.

2 Related Work

Alexandra studied the relationship between the degree of pain in pigs and the pig's voice. The results showed that the pain in piglets could be judged by sound. At the same time, for piglets, the pain degree of castration and

* Corresponding Author

tail trimming was similar. The sound judgment method was worth learning [1]. Linhart compared the changes of central frequency and sound quality of the two kinds of sounds by arousing squeals and grunts of pigs with three levels of stimulus. The change of sound characteristics of squeals is greater than that of grunts. The classification model of sounds is of reference value for this paper [2]. Shuo Peng took HMM as the recognition basis for pig cough and further identified the pig cough sound through the application of model and hidden state, but increased the recognition workload to deal with the sound [3]. Based on the idea of objective optimization, Yani Xu trained the characteristics of the sow cough sound signal and proposed a recognition algorithm for the sow cough sound. The recognition effect was good, but the sound data model was not described too much [4]. Yongjie Gong proposed a feature parameter extraction method based on the combination of static and dynamic characteristics of pig cough sound, and combined with SVM, VQ and HMM to explore the best identification search path respectively, and the search model was confusing [5]. Xuan Li learned the characteristics of pig voice signals through BLSTM network, and designed a pig voice recognition system through CTC. The average pig cough recognition rate in the experiment reached a high level, but the recognition accuracy depends on the size of the data set established [6]. Yawen Wu has established a spectral subtraction model for audio processing, which reduces the efficiency of sound processing, improves the efficiency of sound processing, and solves the problem of sound distortion to a certain extent [7]. Yan Cang believed that abnormal pig behavior was related to abnormal pig voice. By establishing an optimal recognition model for pig voice classification, the recognition accuracy of abnormal voice was improved [8].

3 Description and Identification of Abnormal Sound Data of Pigs

3.1 Voice Recognition Hardware System Structure

The structure of pig voice recognition system is described as follows:

- (1) By setting up a pickup in the pig farm, the control module is used to transmit, transfer and receive multi node audio signals;
- (2) Pre process and denoise the acquired sound;
- (3) Select appropriate sound signal features and extract them;
- (4) Train some features, establish a template library, estimate and recognize some samples, compare their similarity, and calculate the recognition results.

The structure schematic diagram of pig audio recognition system is shown in Fig. 1:

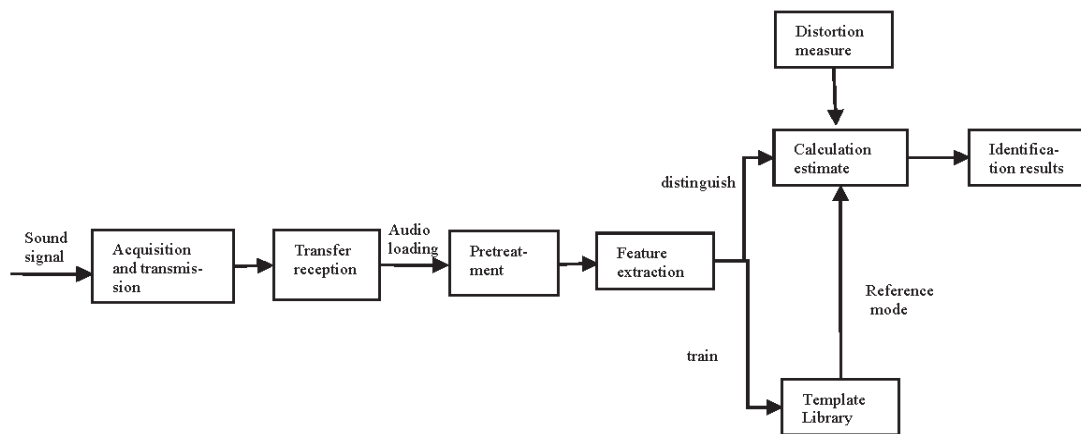


Fig. 1. Structural schematic diagram of the system

3.2 Analysis of Abnormal Sound in Pigs

The pig's voice is closely related to its own physical condition. This paper mainly studies the cough sound, which is most closely related to the pig's health, as the research object of pig's abnormal voice. Pigs may cough when suffering from respiratory diseases, bacterial infections and influenza [9]. The above diseases are often accompanied by infectivity. For intensive aquaculture enterprises, if they do not control these diseases in time, they may

cause huge economic losses. In the early stage of disease, pigs are usually accompanied by cough sounds, which are particularly obvious. Therefore, early recognition and judgment of cough sounds are more important for disease warning and prevention and control. Sound signals are generally described by time domain characteristics and frequency domain characteristics. Fig. 2(a) shows the time domain diagram of pig cough sound, and Fig. 2(b) shows the frequency domain diagram.

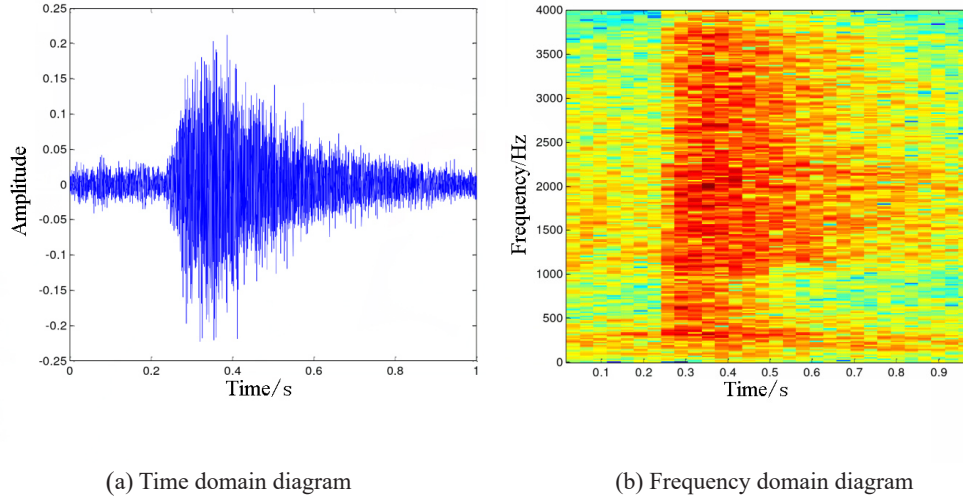


Fig. 2. Pig cough sound pattern

3.3 Pig Audio Signal Acquisition

In view of the layout of the pig farm and the range of activities of pigs, the microphone adopts the iTalk-02 produced by Tyless Company, and is fixed on the column in the middle of the pig house, with a height of about 2 meters, facing the pig eating and activity area, so that the pig voice can be collected more clearly. If you cannot provide your figures electronically, paste originals into the manuscript and center them between the margins. For halftone figures (photos), please forward high-contrast glossy prints and mark the space in the text as well as the back of the photos clearly, so that there can be no doubt about where or which way up they should be placed.

3.4 Preprocessing of Sound Signal

In order to eliminate the influence of useless audio, such as environmental noise, the collected audio is preprocessed, including pre emphasis, framing, and windowing operations, so as to obtain a stable and smooth signal.

(1) Preemphasis

The high-frequency end of audio signal $r(n)$ will be attenuated, and the high-frequency part will be reduced, so the high-frequency signal part should be added. In this paper, digital filters are used to pre emphasize the pig audio signal:

$$b(n) = r(n) - ar(n-1). \quad (1)$$

Wherein, $b(n)$ is the audio signal after the emphasis; a is the pre weighting coefficient and $a = 0.9375$.

(2) Framing

The pig audio signal can be regarded as a stable process in a short time, so the audio signal is first divided into frames, and then the audio signal is divided into frames according to the number of cycles re contained in each frame as follows:

$$f_n = (N - F) / I + 1. \quad (2)$$

Where, f_n is the number of frames, N is the signal length, F is the frame length, and I is the frame shift.

(3) Windowing

Due to the interruption in the first place of the audio signal that may be caused by framing, windowing is to transform the fragments after framing and connect the interrupted parts. The convolution form of the output time series $x(n)$ of pig audio signal generated by a finite unit impulse response filter is as follows:

$$x(n) = \sum_{n=-\infty}^{\infty} b(n)w(n-m). \quad (3)$$

Wherein, $b(n)$ is the pre weighted signal and $w(m)$ is a unit impact response. Window function $w(n)$ is expressed as follows:

$$w(n) = \begin{cases} 0.54 - 0.46 \cos[2\pi n / (L-1)] & 0 \leq n \leq L-1 \\ 0 & \text{other} \end{cases}. \quad (4)$$

In the formula, L is the window length.

(4) Noise reduction by spectral subtraction filtering

Spectrum subtraction [10] is to use the short-term stationarity of the audio signal to remove the noise signal of the short-term spectrum of the noisy audio signal, so as to extract the clean audio spectrum. The schematic diagram is as the Fig. 3:

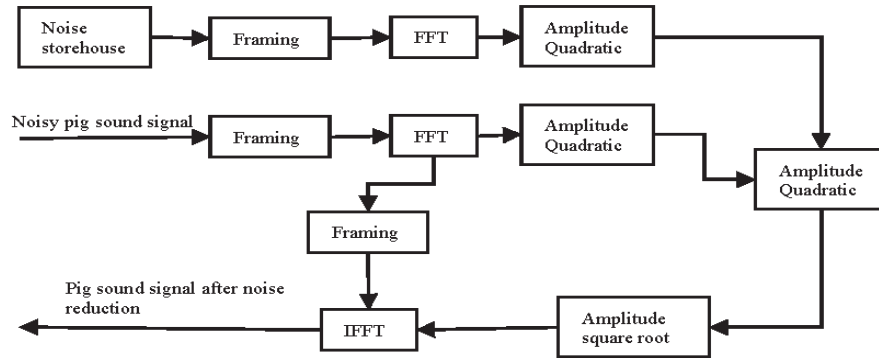


Fig. 3. Noise reduction schematic diagram

The time sequence of pig audio signal is $x(n)$, and the signal of Fourier transform after framing is:

$$X_i(k) = \sum_{m=0}^{N_1-1} x_i(m) \exp\left(j \frac{2\pi mk}{N_1}\right) \quad k = 0, 1, 2, \dots, N_1 - 1. \quad (5)$$

Where, $X_i(k)$ is the amplitude in the frequency domain after the frame transformation, $x_i(m)$ is the pre-processed i th frame signal, N_1 is the frame length, and the absolute value of the amplitude is expressed as $|X_i(k)|$, and the calculation formula of the phase angle $X_a^i(k)$ is as follows:

$$X_a^i(k) = \arctan \left[\frac{\text{Im}(X_i(k))}{\text{Re}(X_i(k))} \right]. \quad (6)$$

Where $\text{Im}(X_i(k))$ is the imaginary part of $X_i(k)$ and $\text{Re}(X_i(k))$ is the real part of $X_i(k)$. The calculation formula of average energy $E(k)$ is:

$$E(k) = \frac{1}{N} \sum_{i=1}^{N_2} |X_i(k)|^2. \quad (7)$$

Where, I_1 is the length of no session, and N_2 is the number of frames corresponding to I_1 . Each frame signal obtained after preprocessing is used as a short-time stationary state signal. All noise signals meet the requirements of spectral subtraction. The audio enhancement effect is good and the operation is simple. The spectral subtraction formula is as follows:

$$MMD_{Y_{t+1}}^2 \equiv \| \mu X_{t+1} | S_X, S_Y - \mu X_{t+1} | S_X \|_{H_x}^2. \quad (8)$$

Where, $|X_i(k)|$ is the amplitude after spectrum subtraction, c is the subtraction factor, so that $c = 4$ and d are the spectrum threshold parameters, and $d = 0.001$, the amplitude after spectrum subtraction $|X_i(k)|$ is combined with the phase angle $X_a^i(k)$ for inverse FFT to obtain the audio sequence $\tilde{x}(k)$.

3.5 Pig Voice Endpoint Detection

The collected signal contains other sounds, and only by extracting the features of pig sounds can the accuracy of subsequent recognition be improved, so it is necessary to distinguish the segments of pig sounds. This paper uses the method of double threshold endpoint detection to determine the starting point and end point of pig voice.

The short-time energy calculation method of pig sound signal is as follows:

$$E_i = \sum_{n=0}^{F-1} y_i^2(n), 1 \leq i \leq f_n. \quad (9)$$

Wherein, F represents the frame length, f_n represents the number of frames divided, and $y_i(n)$ represents the pig signal sound sequence in frame i . The representation method is as follows:

$$y_i(n) = w(n) * x((i-1) * I + n). \quad (10)$$

The average zero crossing rate of pig signal is expressed as:

$$Z_i = \frac{1}{2} \sum_{n=0}^{N-1} |\text{sgn}[y_i(n)] - \text{sgn}[y_i(n-1)]|, 1 \leq i \leq f_n. \quad (11)$$

Where $\text{sgn}[]$ is a symbolic function:

$$\text{sgn}[y(n)] = \begin{cases} 1 & y(n) \geq 0 \\ -1 & y(n) < 0 \end{cases} \quad (12)$$

Therefore, the double threshold endpoint detection method is used to detect the starting point and end point of pig cough, as shown in Fig. 4:

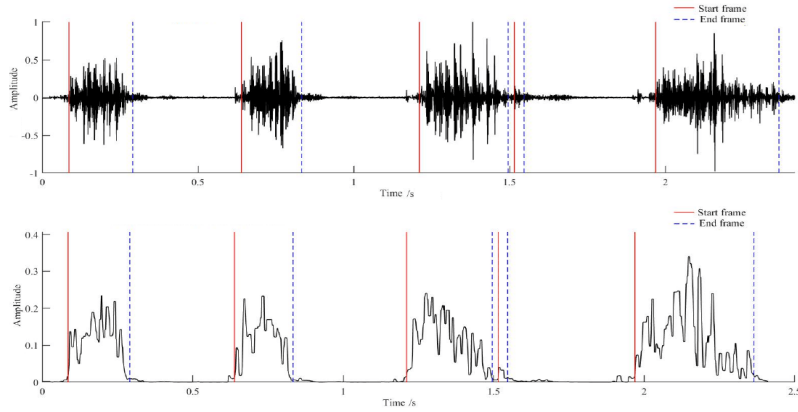


Fig. 4. Endpoint detection results

3.6 Feature Extraction

The voice feature parameters represent the pig voice information. In this paper, Mel-Frequency Cepstrum Coefficients (MFCC) is used to extract the pig voice feature. Mel-Frequency Cepstrum Coefficients can simulate the auditory characteristics of the human ear, and represent the sound signal nonlinearly. It has good anti noise characteristics. The bass of the auditory masking effect is easy to hide the high tone, but the high tone is not easy to hide the low tone. This paper sets a group of band-pass filters to highlight the low-frequency signal, so as to achieve filtering. The conversion relationship of Mel frequency is expressed as:

$$F = 2595 \lg(1 + f / 700). \quad (13)$$

Wherein, represents Mel frequency scale, and f represents the actual frequency. In this paper, the sampling frequency of the collected samples is set to, the frame length is set to, and a Mel filter is given. The representation of the filter bank is shown in Fig. 5.

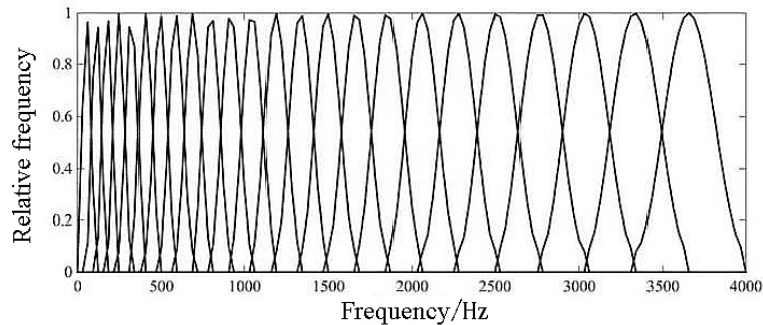


Fig. 5. The response curve of Mel filter bank

The flow schematic diagram of pig audio signal extraction MFCC parameters is shown in Fig. 6. After cepstrum processing, MFCC feature parameters are extracted.

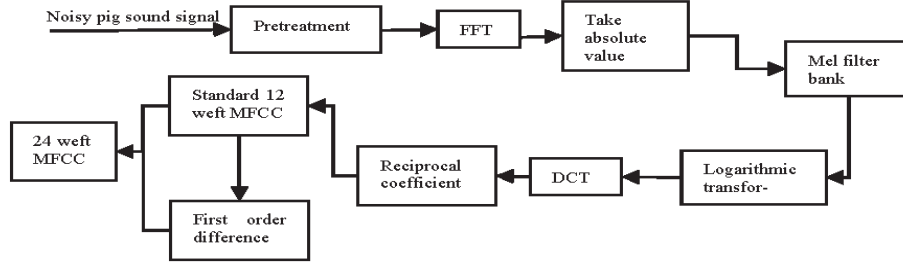


Fig. 6. MFCC extraction process of live pig sound signal

The interference information of the acquired pig audio signal is removed, and several frames of pig audio signal are obtained, and each frame of pig audio signal is Fourier transformed:

$$X(i, k) = FFT[x_i(m)]. \quad (14)$$

Wherein, $x_i(m)$ is the pig audio signal of frame i , $X(i, k)$ is the sound signal spectrum, k is the k spectral line, and then calculate the power spectrum $E(i, k)$ of pig audio signal of each frame.

$$E(i, k) = |X(i, k)|^2. \quad (15)$$

Calculate the power spectrum of the signal for each frame passing through the Mel filter bank. The energy of the power spectrum is shown as follows:

$$S(i, m) = \sum_{k=0}^{L-1} E(i, k) H_m(k) \quad 0 \leq m \leq M. \quad (16)$$

Where, $S(i, m)$ represents Mel filter energy, m represents m Mel filters, and $H_m(k)$ represents Mel filters. The calculation formula is as follows:

$$H_m[k] = \begin{cases} 0 & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \leq k < f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases}. \quad (17)$$

The logarithmic energy output by each Mel filter is as follows:

$$C(i, m) = \ln[S(i, m)] \quad 0 \leq m < M. \quad (18)$$

To sum up, MFCC parameters are shown as follows:

$$mfcc(i, n) = \sqrt{\frac{2}{m} \sum_{m=0}^{M-1} C(i, m) \cos \left[\frac{\pi n(2m-1)}{2M} \right]}. \quad (19)$$

In the formula, $mfcc(i, n)$ is the MFCC parameter, i is the number of frames, n is the spectral line, and m is the order number of Mel filter. The characteristics of pig sound signals are often both static and dynamic, and most of them are dynamic. Therefore, the differential expression of dynamic MFCC parameters is as follows:

$$\Delta mfcc_n = \sum_{i=-1}^T i \cdot mfcc_{n+1} / \sum_{i=-1}^T i^2. \quad (20)$$

In the formula, $\Delta mfcc_n$ is the first-order difference of MFCC characteristic parameters, and I is taken as 2. The characteristic parameters of pig cough sound are as Fig. 7:

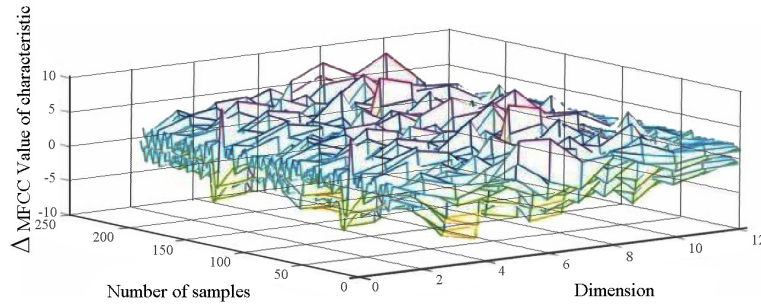


Fig. 7. Characteristic parameters of pig cough sound

3.7 Voice Recognition

Hidden Markov Model (HMM) is a typical machine learning algorithm. A complete HMM model defines the hidden state, observed state, initial state, hidden state transition probability matrix, and hidden state to observable state output probability matrix [11], as shown in Table 1.

Table 1. The content of five factors in hidden Markov model

Factors	Meaning
N	Number of states
M	Number of observables corresponding to each state
A	State transition probability matrix
B	Probability distribution of observations
π	Probability distribution in initial state

An observation sequence $O = \{o_1, o_2, \dots, o_T\}$ can be sorted out by using the above table. The HMM process is as follows:

- (1) Select an initial state $q_1 = i$;
- (2) Set observation time $t = 1$;
- (3) According to the probability matrix B of the current state observation value, obtain $O_t = v_k$;
- (4) From the state transition probability matrix A , the current state $q_t = i$ enters the next state $q_{t+1} = j$;

(5) Set $t = t + 1$, repeat (3) if $t < T$, otherwise end.

Therefore, a complete HMM model can be expressed as:

$$\lambda = (A, B, \pi). \quad (21)$$

Let the algorithm continuously learn to adjust the model parameters. At the beginning, use the known training samples to initialize the model parameters, then train the model, and then continue to optimize the parameters through continuous learning until the model parameters achieve the project approval effect. In this paper, Baum Welch algorithm is used to solve the learning problem of HMM. According to the maximum likelihood criterion, the forward backward algorithm is iterated repeatedly to continuously adjust the optimal model parameters. Observe sequence $O = \{o_1, o_2, \dots, o_T\}$, initialize model $\lambda = (A, B, \pi)$, re estimate model parameter $\bar{\lambda} = (\bar{\pi}, \bar{A}, \bar{B})$, and satisfy $P(O|\bar{\lambda}) > P(O|\lambda)$. The probability that the state of auxiliary variable $\sigma_t(i, j)$ changes from i to j from time t to time $t + 1$ is expressed as:

$$\sigma_t(i, j) = P(q_t = i, q_{t+1} = j | O, \lambda) = \frac{P(q_t = i, q_{t+1} = j | O, \lambda)}{P(O | \lambda)}. \quad (22)$$

In the formula, $\zeta_t(i)$ represents the state probability of i at time t under the conditions of $O = \{o_1, o_2, \dots, o_T\}$ and $\lambda = (A, B, \pi)$, as shown below:

$$\zeta_t(i) = \sum_{j=1}^L \sigma_t(i, j) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)}. \quad (23)$$

In the formula, $\alpha_t(i)$ and $\beta_t(i)$ are forward and backward variables respectively. The revaluation formula updates the model parameters after iteration. The updated formula is:

$$\bar{\pi}_i = \zeta_1(i). \quad (24)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \sigma_t(i, j)}{\sum_{t=1}^{T-1} \zeta_t(i)}. \quad (25)$$

$$\bar{b}_j(k) = \frac{\sum_{t=1, O_t=V_t}^{T-1} \zeta_t(i)}{\sum_{t=1}^{T-1} \zeta_t(i)}. \quad (26)$$

After obtaining the required parameter $\lambda = (A, B, \pi)$, the training stops.

4 Simulation Experiment and Result Analysis

4.1 Results of Pig Noise Reduction

In this paper, the pig sound noise reduction method is run on Dell workstation, the CPU is Inter Xeon E5-2603 v4, the memory is 16GB, and the program running platform is Python 3.6. In this paper, the noise reduction method described in Section 3.4 is used to add $10dB$ noise to the pure pig cough sound, and then the noise is reduced through the algorithm. The waveform of the noise reduced pig cough sound is consistent with that of the pure pig cough sound, which proves the effectiveness of the noise reduction algorithm. The waveform of the pig cough sound before and after the noise reduction is shown in Fig. 8:

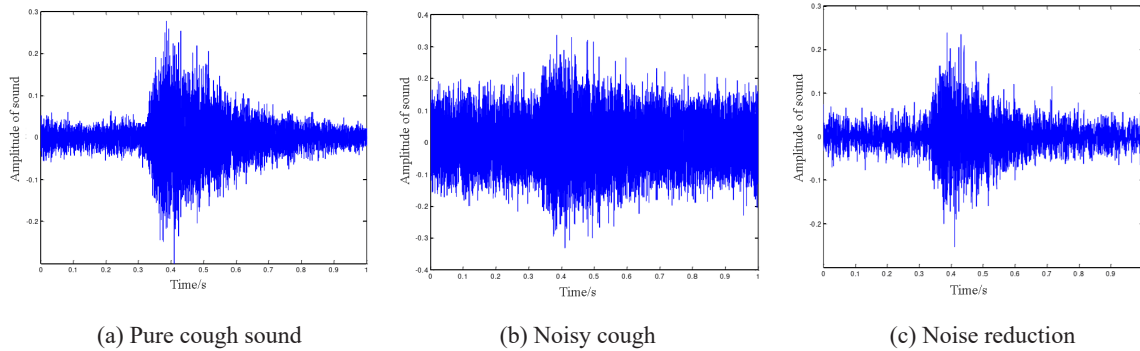
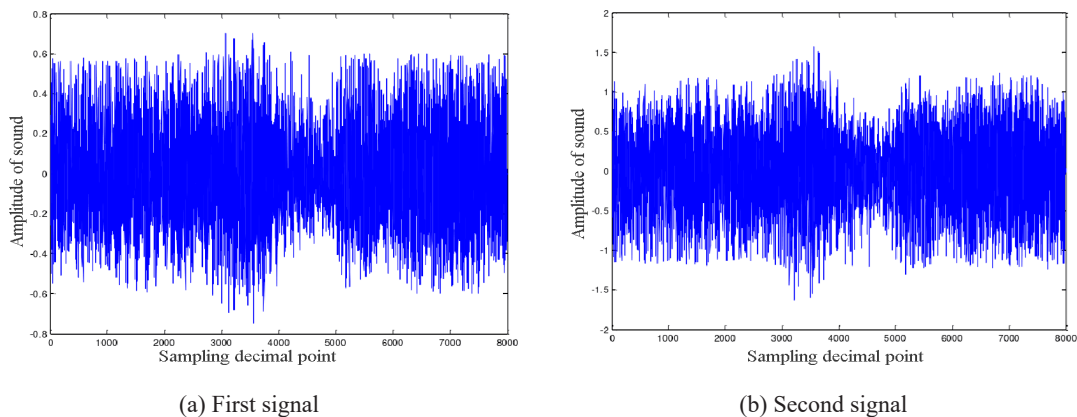


Fig. 8. Waveforms of porcine cough before and after denoising

4.2 Improved HMM Voice Recognition

The professional recording pen of Philips was used to sample at Jidong Pig Farm in Tangshan. The format used for sample collection was high frame. wav format. The sampling frequency of the recording pen was 44.1 kHz, the accuracy was 16 bits, and the sampling space was selected separately. Then use MATLAB for simulation test.

30 groups of data of pig cough were recorded, of which the first 25 groups were used as training sets, and the last 5 groups were used as test sets. Then MFCC features of pig cough audio are extracted, HMM model is trained by training samples, and then recognized by algorithm. Initialize the number of sound hiding states to, the length of the initial probability is equal to the number of states, the first element is 1, the other elements are 0, the initial transition matrix is 4×4 , and the total value is 1. Then repeat iterative calculation through Baum Welch algorithm, accumulate the output probability of observation sequence, and calculate the required probability value. When the probability is less than a certain set threshold, judge the convergence and end the convergence, otherwise return to continue the iteration, set the threshold value as 5×10^{-6} , and the number of iterations is 40.



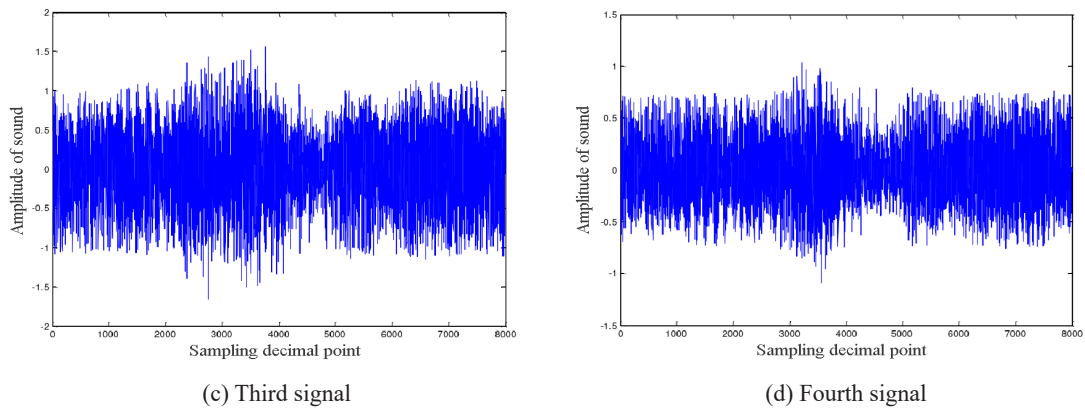


Fig. 9. Sound sampling sample

Fig. 9 shows the four channels of pig sound samples with cough sound collected from the pig farm, input the sound samples into MATLAB, and then obtain the cough sound through noise reduction and separation for comparison.

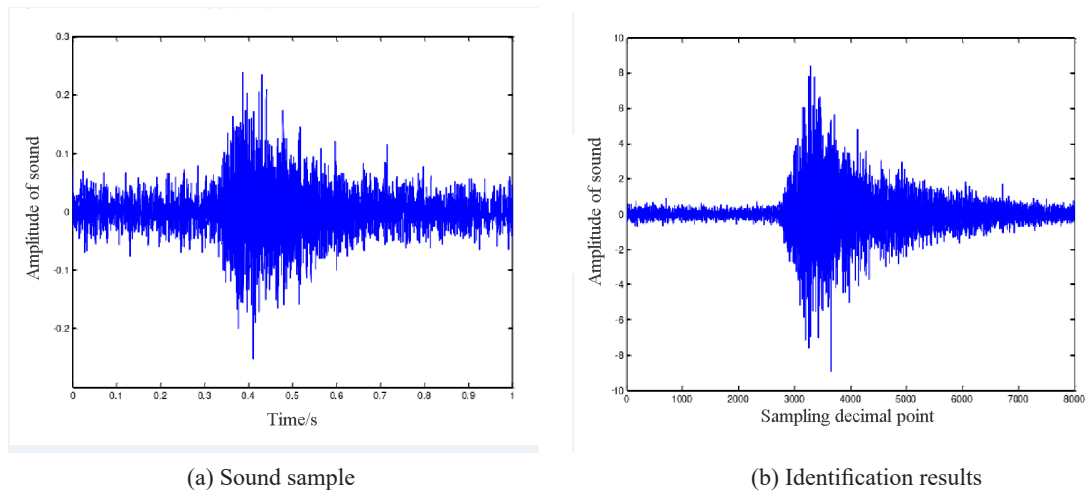


Fig. 10. Comparison chart of identification results

Fig. 10(a) shows the identified pig cough sound, and Fig. 10(b) shows the standard figure. After similarity comparison, it is considered as pig cough sound.

5 Conclusion

This paper discusses a method of pig audio recognition, which can judge whether the pig's body state has changed by identifying whether the pig's voice contains cough, so as to help large-scale farms improve breeding efficiency and reduce the incidence of pigs. In this paper, an improved sound noise reduction method is designed and a machine learning recognition algorithm is introduced to improve the efficiency of sound recognition. The simulation results show that the method provided in this paper is reliable and efficient.

In the future, according to the living habits of pigs, we will identify and study the laws of abnormal behavior and diet of pigs, so as to help farmers carry out scientific breeding more deeply and bring higher social benefits.

6 Acknowledgement

This work has been supported by University Level Projects: Research on Monitoring and Forecasting Service System of Pig Breeding Based on Cloud Computing Technology -- Taking Xihuang Village Pig Breeding Base in Leting County as an Example (YJKT202204).

References

- [1] A.-F. Cordeiro, I.-A. Nääs, M.-S. Baracho, F.-G. Jacob, D.-J. Moura, The use of vocalization signals to estimate the level of pain in piglets, *Engenharia Agrícola* 38(4)(2018) 486-490.
- [2] L. Pavel, R. Victoria, R. David, Š. Marek, Expression of emotional arousal in two different piglet call types, *Plos ONE* 10(8)(2015) 1-13.
- [3] S.-Peng, D.-Y. Liu, G.-L. Shi, G.-B. Li, J.-S. Mu, L.-C. Gu, J.-Jiao, Pig state audio recognition based on deep neural network and hidden Markov model, *Journal of China Agricultural University* 27(6)(2022) 172-181.
- [4] Y.-N. Xu, M.-X. Shen, L. Yan, L.-S. Liu, C.-R. Chen, P.-Q. Xu, Research of predelivery Meishan sow cough recognition algorithm, *Journal of Nanjing Agricultural University* 39(4)(2016) 681-687.
- [5] Y.-J. Gong, X. Li, Y. Gao, M.-G. Lei, W.-H. Liu, Z. Yang, Pig cough recognition based on vector quantization *Journal of Huazhong Agricultural University* 36(3)(2017) 119-124.
- [6] X. Li, J. Zhao, Y. Gao, M.-G. Lei, W.-H. Liu, Y.-J. Gong, Recognition of pig cough sound based on deep belief Nets *Transactions of the Chinese Society for Agricultural Machinery* 49(3)(2018) 179-186.
- [7] Y.-W. Wu, R. Shao, M. Li, F. Zhang, H.-B. Tao, L.-C. Gu, J. Jiao, Endpoint detection of live pig audio signal based on improved EMD-TEO cepstrum distance, *Journal of China Agricultural University* 26(4)(2021) 104-116.
- [8] Y. Cang, S.-Y. Luo, Y.-L. Qiao, Classification of pig sounds based on deep neural network, *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)* 36(9)(2020) 195-204.
- [9] X.-W. Chen, Clinical symptoms and preventive measures of swine asthma, *Modern Animal Husbandry Science & Technolog* (11)(2021) 114-115.
- [10] W.-X. Fan, M.-N. Jia, W. Chen, X. Liu, H.-Y. Han, Cable local discharge signal denoising method based on spectral subtraction and wavelet analysis, *Modern Electronics Technique* 44(19)(2021) 71-75.
- [11] L.-P. Ye, T. He, HMM speech recognition study of an Improved Particle Swarm Optimization based on Self-Adaptive Escape (AEPSO), *IOP Conference Series: Earth and Environmental Science* 634(1)(2021).