Personalized Power Consumption Control Method for Intelligent Buildings That Perceiving Speech Recognition Identity

Fang Jianjun 1,2, Liu Penghui 3

1China Energy Investment Corporation Center for Green Energy and Architecture, Beijing102211, China;
2National Institute of Clean-and-Low-Carbon Energy, Beijing 102211, China;
3Inner Mongolia Key Laboratory of Perceptual Technology and Intelligent Systems, Huhehaote, 010062, China
E-mail: 13520185312@163.com

Abstract: With the development of IoT technology, the meaning of the concept of energy conservation and environmental protection in people's minds is also constantly changing, and it is moving towards the direction of intelligence and personalization. At present, China's economy has entered the era of artificial intelligence. Electricity control such as building lighting has gradually changed from a single manual control to a multi-modal control. This paper designs an embedded control method based on speech recognition of personnel to achieve personalized power consumption, and improves the existing defects of intelligent power consumption in buildings such as unconfirmed operators and single decision-making factors.

The method is to extract the mel frequency cepstrum coefficient feature vector after preprocessing the speech signal; Aiming at the problem of singularity matrix in the parameter estimation of Gaussian mixture model, the modified EM algorithm was used to optimize the parameter estimation of the Gaussian mixture model to complete the identification of personnel. The experimental test results show that the system can control the electricity facilities in combination with the environmental information and realize the personalized lighting control function while confirming the identity of the personnel.

1. Introduction
Under the current rapid development of intelligent buildings, it is required that the lighting control system be intelligent. As China's environmental protection efforts continue to increase, higher energy conservation and emission reduction requirements are imposed on energy users, especially buildings. In addition to ensuring the basic needs of building lighting, the building intelligent lighting control system also needs to reduce the power consumption as much as possible. With the development of power electronics technology and intelligent control technology, intelligent control of building lighting can be realized. With the development of people's living standards, the individualized demand for lighting has gradually become richer, and building lighting has become increasingly demanding to meet comfort levels.
Personalized electricity use in buildings is to integrate the building environment and facilities with voice recognition technology, wireless sensing, and control networks to build a convenient and efficient control system, thereby improving the convenience and comfort of electricity use in buildings [1]. With the development of technology, embedded controllers with voice processing capabilities have gradually become voice interaction objects. Electricity control for building lighting has gradually changed from a single manual control to a multi-modal control. This is also a new era product combining speech recognition technology, automatic control technology and communication technology [2].

At present, the smart power industry has a promising market, but its development still has shortcomings, such as the lack of a unified industry standard, potential safety hazards, the inability to identify operators, and single decision-making factors [3].

Aiming at the problem of the inability to identify operators and single decision-making factors, the direct and convenient features of voice can be used, while the building environment information can be integrated, and the identity of the operator can be confirmed by voice to improve the convenience and comfort of intelligent electricity consumption, which is conducive to system application promotion. The embedded system has the characteristics of high performance and low price. Combining it with speech recognition technology and wireless sensor control network for intelligent power consumption has considerable research and practical significance. Intelligent power consumption based on voice perception is an inevitable trend for people to pursue high-quality, convenient and safe lives. Voice interaction is the foundation. Identity recognition is accomplished through voice perception. The system completes decision control appropriate for the member according to environmental information, and realizes personalized power consumption services.

2. Person identification algorithm based on Gaussian mixture model
The Gaussian Mixture Model (GMM) is the most common method for speaker recognition. The main idea is to use the superposition and combination of different Gaussian probability density functions to express the distribution of feature vectors in the probability space [4]. Considering the limitation of personnel identification in specific application scenarios, this will not only ensure the security of the entire system, but also improve the self-adaptation and self-learning ability of the subsequent system control to a certain extent. GMM personnel identification process is shown in Figure 1.

![GMM validation process schematic diagram](image)

2.1. Model parameter description
The system builds the corresponding identity feature probability distribution model for members in the personnel set. The parameter values involved in the model are determined by the distribution of the characteristic parameters of the members themselves. In order to facilitate processing, it is assumed that the probability density function of each member has the same form, and the linear weighted combination of M single Gaussian distributions can be used to describe the distribution of frame characteristics in the feature space. The Gaussian mixture model is expressed as follows:
Among them, $M$ is the number of components in the GMM; $X$ is a D-dimensional MFCC feature vector; $w_i$ is a weighting coefficient. $b_i(X)$ satisfies the following formula:

$$b_i(X) = w_i \frac{1}{\sqrt{(2\pi)^D \left| \Sigma_i \right|}} \exp \left\{ -\frac{1}{2} (X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i) \right\}$$

Among them, $\mu_i$ is the mean vector and $\Sigma_i$ is the covariance matrix. The mixed weight value satisfies the following equation:

$$\sum_{i=1}^{M} w_i = 1$$

Each component's mixing weight, mean vector, and covariance matrix form a complete Gaussian mixture model [5]. In the model, $\lambda$ is used as a parameter description, and $\lambda = \{w_i, \mu_i, \Sigma_i\}, i = 1, ..., M$.

2.2. Model parameter estimation

In speech perception recognition system, each member is characterized by its own GMM model. After getting the voice of a member, it can build his model through training. The process of building a model is to estimate the parameter values of the GMM model.

Suppose the sequence of speech feature vectors is $X = \{X_1, X_2, ..., X_T\}$, and the likelihood function is expressed as:

$$P(X \mid \lambda) = \prod_{t=1}^{T} P(X_t \mid \lambda)$$

In order to facilitate the calculation of the above formula, logarithms are taken on both sides of the above equation, and its log-likelihood can be expressed as follows:

$$L(X \mid \lambda) = \log P(X \mid \lambda) = \sum_{t=1}^{T} \log P(X_t \mid \lambda)$$

To find the best model $\lambda$ of $\hat{\lambda}$, we need to get the optimal model parameter $\hat{\lambda}$ such that $\hat{\lambda} = \arg \max \ L(X \mid \lambda)$. According to the maximum likelihood estimation principle, when the parameter $\lambda$ is adjusted to make the average probability of the $T$ series reach the maximum, $L$ also reaches the maximum value. The non-linearity of parameter $\lambda$ directly makes it difficult to find the maximum value, so it is implemented by the Expectation Maximization (EM) iteration algorithm [6]. Iterate according to formulas (6) to (8):

$$w_i = \frac{1}{T} \sum_{t=1}^{T} P(i \mid X_t, \lambda)$$

$$\mu_i = \frac{\sum_{t=1}^{T} P(i \mid X_t, \lambda) X_t}{\sum_{t=1}^{T} P(i \mid X_t, \lambda)}$$
\[
\sigma_i^2 = \frac{\sum_{t=1}^{r} P(i \mid X_t, \lambda) (X_t - \mu_i)^2}{\sum_{t=1}^{r} P(i \mid X_t, \lambda)}
\]

(8)

Among them, \(P(i \mid X_t, \lambda)\) is the posterior probability of the i-th Gaussian mixture component.

\[
P(i \mid X_t, \lambda) = \frac{w_i b_j(X_t)}{\sum_{k=1}^{m} w_k b_k(X_t)}
\]

(9)

Starting from setting the initial value of the model parameter \(\lambda\), the EM algorithm is used to re-estimate \(\hat{\lambda}\) so that the new model parameter satisfies the following conditions: \(P(X \mid \hat{\lambda}) \geq P(X \mid \lambda)\). The new model is used as the initial value and continues to be recursive until the model converges or reaches the maximum number of iterations.

According to the literature \([7]\) in the estimation of the GMM model parameter \(\lambda = \{\mu_i, \Sigma_i\}\), it is easy to generate a singular matrix using the above method. Therefore, this paper uses maximum likelihood estimation (ML) as the initial model, and then uses \(\alpha\) to control the correction ratio to modify each step model of the EM algorithm. The specific formula is as follows:

\[
w_j = \frac{1}{T} \sum_{t=1}^{r} P(i \mid X_t, \lambda)
\]

(10)

The modified revaluation formula for mean \(\mu_i\) is as follows:

\[
\mu_i = \alpha \mu_{i-1} + (1 - \alpha) \frac{\sum_{t=1}^{r} P(i \mid X_t, \lambda) X_t}{\sum_{t=1}^{r} P(i \mid X_t, \lambda)}
\]

(11)

The modified revaluation formula for mean \(\sigma_i^2\) is as follows:

\[
\sigma_i^2 = \alpha \sigma_{i-1}^2 + (1 - \alpha) \frac{\sum_{t=1}^{r} P(i \mid X_t, \lambda) (X_t - \mu_i)^2}{\sum_{t=1}^{r} P(i \mid X_t, \lambda)}
\]

(12)

Among them, the posterior probability of the component \(i\) is as follows:

\[
P(i \mid X_t, \lambda) = \frac{w_i b_j(X_t)}{\sum_{k=1}^{m} w_k b_k(X_t)}
\]

(13)

It is not difficult to see from the above formula: when \(\alpha = 0\), the optimized EM algorithm is the traditional EM algorithm.

2.3. Identification of personnel

After the system completes the GMM training, it is to confirm the identity of the person, that is, the system determines whether it is the model \(\lambda_i = \{\lambda_{i1}, \lambda_{i2}, \lambda_{i3}, \lambda_{i4}, \lambda_{i5}\}\) corresponding to the predetermined person identity \(i^* = \{i_1, i_2, i_3, i_4, i_5\}\), so that it has the largest posterior probability
According to the Bayesian formula, the following formula is known:

\[ P(\lambda_i | X) = \frac{P(X | \lambda_i)P(\lambda_i)}{P(X)} \]  

(14)

Where \( P(X) \) and \( P(\lambda_i) \) values are determined, so \( P(\lambda_i | X) \) is determined by \( P(X | \lambda_i) \), that is:

\[ i^* = \arg \max \ P(X | \lambda_i) \]  

(15)

2.4. Experimental test

According to the experimental scheme provided in Reference [8], the experimental data uses the voices of 5 members in the voice database. One sentence is randomly selected as the training voice sample under the folder of any recording of each member, and the remaining 13 sentences are used as test files. Based on the above theoretical knowledge, the system establishes a GMM model for confirming the identity of a person, and strictly extracts feature parameters according to the pre-emphasis coefficient, frame length, and frame offset conditions selected in this paper. Feature parameters were extracted from the original MFCC to form a 13-step MFCC. The training convergence condition of the system is when the relative deviation of the two log-likelihood function values before and after iteration is less than 0.001. The simulation program design adopts adjustable GMM mixing numbers, and the optional Gaussian mixing numbers are 4, 8, 16, 24, and 32.

Figure 2 shows the effect of the system using \( x \) to control the correction rate on model training time, recognition time, and recognition rate under different Gaussian mixture numbers. It is found from Figure 2 that the modified EM makes the data sufficiently trained, but with the increase of the mixing order, the recognition rate decreases somewhat; the recognition rate is the highest when the Gaussian mixture of order 24 is used.

![Figure 2. Comparison of time and recognition rate correction of different mixed orders](image)

The system trains the identity characteristics of five personnel. The process of training and confirmation is shown in Figure 3. Then the system tests again and calculates their likelihoods.

Figure 4 shows the MFCC characteristic information of the person identity \( i_2 \) and the distribution of the coefficients in the 13-dimensional MFCC.

In Figure 4, the MFCC of member \( i_2 \) is plotted on the left. As can be seen from the figure, the first 13 steps of information are relatively obvious, and the characteristics of the speaker can be basically described. The figure on the right is the distribution profiled according to the 13-dimensional MFCC feature values, and the thick outer lines represent the information after adding the distribution probabilities.
3. Conclusion

This paper mainly studies the basic method of identifying people by speech perception. The Mel frequency cepstrum coefficient is extracted as the feature vector of the speech. Then, the Gaussian distribution probability density function with different weights is superimposed to express the distribution of the feature vector in the probability space. Finally, the improved EM algorithm is used to estimate the parameters of the Gaussian mixture Model and complete personnel identification.

References

[1] Guo Tanna, Jiao Yanbing. Analysis on the Impact of Internet of Things Technology on the development of smart home [J]. Digital Users, 2013, 22 (3): p77-77
[2] Li Ming, Guo Liang, Jia Guangcheng, et al. Talking about the influence of intelligentization of electrical products on the development of detection technology [J]. Electric Measurement & Instrumentation, 2009, 46 (S1): p 6-8+86
[3] Wang Jinshuai. Design of smart home control system [D]. Zhengzhou University. 2017
[4] Zhou Lei. Research on speaker identification method based on voicereprint recognition [D]. Shanghai Normal University, 2016
[5] Pang Cheng, Wang Xiuling, Zhang Jie, et al. GMM mandarin chinese accent recognition based on multi-feature fusion [J]. Journal of Huazhong University of Science and Technology (Natural
[6] Camilla Damian, Zehra Eksi, Rüdiger Frey. EM algorithm for Markov chains observed via Gaussian noise and point process information: Theory and case studies[J]. Statistics & Risk Modeling, 2018, 35(1-2): p 51-72

[7] Zheng Liwu. Research and implementation of speaker recognition system based on embedded platform [D]. Southeast University, 2016

[8] Liu Penghui. Research on smart home embedded control system based on voice perception [D]. Inner Mongolia University of Technology, 2018