English Speech Feature Recognition Based On Digital Means

Yuji miao (Yujimiao34@gmail.com)
North China University of Science and Technology
Yanan Huang
North China University of Science and Technology
Zhenjing Da
North China University of Science and Technology

Research Article

Keywords: Digitization, English speech, feature recognition, fuzzy recognition

Posted Date: October 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-941510/v1

English speech feature recognition based on digital means

Yuji Miao,Yanan Huang,Zhenjing Da
Faculty of International Languages in North China University of Science and Technology, Qinggong College , Tangshan, Hebei , 063000, China
Email:mmww122021@163.com

Abstract. In order to improve the effect of English speech recognition, based on digital means, this paper combines the actual needs of English speech feature recognition to improve the digital algorithm. Moreover, this paper combines fuzzy recognition algorithm to analyze English speech features, and analyzes the shortcomings of traditional algorithms, and proposes the fuzzy digitized English speech recognition algorithm, and builds an English speech feature recognition model on this basis. In addition, this paper conducts time-frequency analysis on chaotic signals and speech signals, eliminates noise in English speech features, improves the recognition effect of English speech features, and builds an English speech feature recognition system based on digital means. Finally, this paper conducts grouping experiments by inputting students' English pronunciation forms, and counts the results of the experiments to test the performance of the system. The research results show that the method proposed in this paper has a certain effect.

Keywords: Digitization; English speech; feature recognition; fuzzy recognition

1. Introduction

English speech feature recognition plays an important role in supporting English learning. From a practical point of view, improving English speech feature recognition through digital means can effectively improve the effect of English learning. Therefore, it is necessary to study the application of digital technology in English speech feature recognition [1]

Language is the most basic, most natural, most important, most effective, most commonly used, most convenient and longest-used communication method for human beings. Language is also the foundation of human society. But after human beings entered the computer age, computers have become a daily tool in human life. Language, the most basic and most natural way of communication, has lost its usefulness. Human hands freed from walking upright are once again confined by the keyboard. This is obviously unacceptable to human beings who have created a glorious civilization with both hands [2]. Therefore, almost at the same time that humans use computers, humans are trying to communicate with computers through voice so that computers can understand human languages, but this has always been impossible to be as effective as keyboards and mice. With the advent of the multimedia and network era, mankind has entered the post-PC era. At this time, a variety of intelligent equipment has been widely used in human production and life. The natural, fast, stable, and reliable way of interaction between humans and intelligent terminals makes voice once again an urgent requirement for human-machine communication. This gave rise to the research topic of speech recognition. The voice signal as the carrier of voice recognition has also become an important research object [3]. Moreover, as an important research field, speech signal processing has a long research history. It is a comprehensive subject formed on the basis of speech linguistics and digital signal processing, and it is closely related to disciplines such as psychology, physiology, computer science, communication and information science, pattern recognition and artificial intelligence. At the same time, the research on speech signal processing has always been an important driving force for the development of digital signal processing technology. Many new methods of processing are first achieved in speech processing and then extended to other fields. For example, the birth and development of many high-speed signal processors are inseparable from the research and development of speech signal processing. The so-called speech signal processing is the use of digital signal processing technology to analyze and process the speech signal, which includes speech communication, synthesis, recognition, and speech enhancement. One of its purposes is to obtain some speech parameters reflecting the important characteristics of the speech signal through processing so as to transmit or store the speech signal information with high effect. The second is to process certain calculations to meet the requirements of a certain purpose, such as artificially synthesizing speech, identifying the speaker, and identifying the content of the speech. These two purposes represent the two aspects of the theory and research of speech signal processing that are closely integrated. On the one hand, it is researched from the perspective of speech production and perception. This research is inseparable from subjects such as speech, linguistics, cognitive science, psychology, and physiology. On the other hand, speech is processed as a signal, including traditional digital signal processing technology and some new technologies applied to speech signal processing methods. In addition, voice signal processing is one of the core technologies used in emerging fields such as information superhighway, multimedia technology, office automation, modern communications and intelligent systems.

This paper uses digital means to study English speech feature recognition algorithms, and analyzes actual cases to improve the effect of English speech feature recognition.

2. Related work

The speech signal is a typical non-stationary random signal. Due to the wide variety of voice environments and the complexity of the voice signal itself, people will inevitably be interfered by noise introduced from the surrounding environment and transmission media, electrical noise inside the communication device, and even other speakers during voice communication. These interferences will eventually make the voice received by the receiver no longer pure original voice, but noisy voice polluted by noise. This causes the deterioration of the performance of the speech processing system, affects the recognition rate, and even causes the system to be completely unable to process speech. Moreover, these interference signals cause great interference to the effective information carried by the voice signal in the voice communication. As a result, the denoising processing of the speech signal is produced [4].

The literature [5] carried out research on the safety communication problems in air traffic control based on the voice of the pilot and the identity of the aircraft, and achieved outstanding research results. In order to establish the JP telephone network intrusion detection system, the literature [6] implanted watermark information in the telephone voice, and verified the feasibility of the method through simulation experiments. The literature [7] studied the related problems of the anti-interference ability of watermark in digital-to-analog conversion, and used ICA algorithm to separate two different sounds from a piece of a cappella recording. The literature [8] successfully realized the interactive modification of music recording by virtue of the good separation effect of ICA algorithm. The literature [9], the extraction of small target signal under clutter background and the investigation of system security performance are discussed.The literature [10] proposed a recursive algorithm based on neural simulation structure. However, this algorithm only uses second-order statistical properties, so its convergence needs to be further optimized. Aiming at the problem of blind source separation, the literature [11] gave a relatively complete research framework, and by analyzing the separability and uncertainty in the blind source separation algorithm, it proposed a joint diagonalization method. The BP neural network algorithm proposed in the literature [12] has become one of the most widely used neural network models. The literature [13] established an objective function based on information maximization, and constructed a unified framework of ICA algorithm on the basis of information theory.

The literature [14] proposed a disguised voice hidden telephone system. Moreover, due to the consideration of the real-time nature of the system, it combines the classic hiding method to successfully realize the secure transmission of real-time voice. The literature [15] used audio as a hidden carrier to mask the secret voice signal, so as to enhance the signal's anti-low-pass filtering ability and anti-interference ability. The literature [16] derived an adaptive blind source separation switching algorithm based on kurtosis, and used this algorithm to achieve blind separation of speech signals. The simulation experiment verifies that the algorithm has good separation performance. The literature [17] established a dual-channel blind source separation model on the basis of the 4th order cumulant, and used the feature matrix joint approximate diagonalization algorithm to further realize the separation of the frequency hopping signal and the interference. Aiming at the problem of noise interference in multiple received signals, the literature [18] clearly pointed out that the signal can be de-noised by wavelet first, and then blindly separated. This method has a significant improvement effect on the bit error rate performance of frequency hopping communication against strong uncorrelated interference. The literature [19] systematically summarized the blind source separation algorithm, and further explained the future development direction of the blind source separation algorithm. The literature [20] used the feature matrix joint approximate diagonalization algorithm to blind source separation of the signal. The previous blind source separation anti-interference methods require a large number of antennas at the receiving end. In response to this problem, the literature [21] proposed a semi-blind separation anti-interference algorithm for direct spread communication based on the periodicity of the spreading code under a single channel.

3. Digital English speech processing algorithm

(1) Data fuzzification

Fuzzification is the process of transforming numerical attributes into fuzzy attributes. Nowadays, many methods have been proposed to realize this transformation

Generally speaking, it can be studied by experts in the field of specific discretization algorithm, or by any fuzzy clustering algorithm. Clustering is a popular data mining technology, which allocates data instances to several groups called clusters, so that the instances belonging to different clusters are as different as possible. The similarity between clusters and instances can be measured by distance, connectivity and strength. Non fuzzy clustering algorithms, such as DBSCAN or K-means, assign each instance to a cluster. In the case of fuzzy clustering, each instance can be associated with more clusters with membership degree. Membership indicates the degree to which the instance belongs to the cluster. One of the most commonly used fuzzy clustering algorithms is fuzzy C-means (FCM).

The purpose of this algorithm is to minimize the following criteria[22].

$$\min imize \sum_{j=1}^{m_i} \sum_{i=1}^{k} \left(u_{i,j}\right)^2 d\left(a_i, c_j\right)^2 \qquad (1)$$

Among them, $u_{i,j}$ represents the membership degree of the i-th instance to the j-th cluster, f is the ambiguity, and $d\left(a_i, c_j\right)$ is the distance metric.

$$d\left(a_i, c_j\right) = \sqrt{\left(a_i - c_j\right)^2} = \left|a_i - c_j\right| \qquad (2)$$

This algorithm defines the constraint of the membership degree $u_{i,j}$ of each level, and the sum of the membership degree of each cluster of a_i must be equal to l. This constraint starts directly from the use of fuzzy C-Means, but it is also necessary to achieve FDT induction. The constraint can be expressed as the following formula:

$$\begin{aligned}&\textstyle\sum_{j=1}^{m_i} u_{i,j}=1, j=1,2,\mathrm{L}\ ,k\\&0\le u_{i,j}\le 1, i=1,2,\mathrm{L}\ ,k; j=1,2,\mathrm{L}\ ,m_i\\&0\le \textstyle\sum_{j=1}^{m_i} u_{i,j}\le k, i=1,2,\mathrm{L}\ ,k\end{aligned} \tag{3}$$

FCM assigns a membership degree $u_{i,j}$ to each cluster's instance of each cluster. The degree of membership is calculated based on the distance between the instance and the cluster. The membership degree of the i-th instance to the j-th cluster is as follows:

$$u_{i,j}=\frac{\dfrac{1}{d\left(a_i,c_j\right)}}{\sum_{t=1}^{m_i}\dfrac{1}{d\left(a_i,c_t\right)}} \tag{4}$$

When all the values of $a_i\in a$ are assigned to each cluster, the new center can be calculated by the following formula:

$$c_j=\frac{\sum_{j=2}^{k}\left(u_{i,j}\right)^2 a_i}{\sum_{j=1}^{k}\left(u_{i,j}\right)^2} \tag{5}$$

The scalar value can be converted into a degree of membership, and the continuous value can be converted into a fuzzy attribute mapping.

(2) Fuzzy decision tree

DTS is a data mining tool for classification and prediction. The main purpose of classification is to map a new unknown instance to one of the predefined classes.

At present, there are many known DT sensing algorithms, such as ID3, C4.5, CHAID, cart or cmie based fuzzy decision tree. ID3 has good performance in linguistic attributes, but it can't deal with numerical attributes. The next problem with ID3 is that it tends to prefer attributes with more values. These problems are improved in C4.5 algorithm. C4.5 deals with numerical attributes by defining split points, which divide numerical data into intervals. Unfortunately, this can have a negative impact on the performance of classification, especially when some values are close to the margin. In addition, the identification of interval boundary may not be completely correct. The introduction of fuzzy logic can improve this aspect. At present, various algorithms of fuzzy decision tree have been introduced. Many of them are based on traditional algorithms, such as ID3 to fuzzy correction. Other FDT guidance algorithms are based on cmie proposed and applied in.

The associated attribute of each internal node selected by CMIE is defined as $I\left(B;A_{i_1j_1},A_{i_2j_2},\mathrm{L}\ ,A_{i_{q-1}j_{q-1}},A_{i_qj_q}\right)$. According to $U_{q-1}=A_{i_1j_1},A_{i_2j_2},\mathrm{L}\ ,A_{i_{q-1}j_{q-1}},A_{i_qj_q}$, the sequence of attribute values is defined from the root of the tree to the node at the q-th level. The sequence U_{q-1} defines the path from the root to the node under investigation. The attribute of the maximum value of CMIE is associated with the investigated node in the obtained selection criterion, and the CMIE is divided by the entropy of the attribute, thereby preventing a larger number of attribute preferences. The standard has the following form:

$$\arg\max\left(\frac{I\left(B;A_{i_1j_1},A_{i_2j_2},\mathrm{L}\ ,A_{i_{q-1}j_{q-1}},A_{i_qj_q}\right)}{H\left(A_{i_q}\right)}\right) \tag{6}$$

The entropy of attribute A_{i_q} has the following definition:

$$H\left(A_{i_q}\right)=\textstyle\sum_{=1}^{m_i} M\left(A_{i_qj_q}\right)*\left(\log_2 k-\log_2 M\left(A_{i_qj_q}\right)\right) \tag{7}$$

DT is sensitive to overfitting. When the classifier overfits the data it senses, overfitting occurs. Therefore, DT will have lower classification performance when the new data is unknown. In the decision tree, overfitting means that the tree that is induced fits all the instances in the data set. In this case, it classifies training examples perfectly, but the classification of new samples is usually inaccurate. Therefore, the pruning technique is applied, which replaces the subtrees of the tree with leaves. Tree pruning reduces the computational complexity of classification. This paper establishes leaves in two thresholds A and B, and determines the leaf nodes during tree derivation according to the pruning of the leaves. Threshold A represents the minimum frequency of occurrence in a given branch. Frequency represents the percentage of instances that are covered by the path to which the analysis node belongs. When the following conditions are true, the node is converted to a leaf.

$$\alpha\ge\frac{M\left(A_{i_1j_1}\times\mathrm{L}\ \times A_{i_qj_q}\right)}{k} \tag{8}$$

Each node contains a confidence level that reflects the credibility of the output class. If the confidence is greater than the threshold parameter B, the node is transformed on the leaf. The mathematical form of this pruning criterion is as follows:

$$\beta\le 2^{-I\left(B_j\left|A_{i_1j_1},\mathrm{L}\ ,A_{i_zj_z}\right.\right)},\ j=1,\mathrm{L}\ ,m_b \tag{9}$$

$I\left(B_j\left|A_{i_1j_1},\mathrm{L}\ ,A_{i_zj_z}\right.\right)$ can be calculated by the following formula:

$$I\left(A_{i_1j_1}\left|A_{i_zj_z}\right.\right)=H\left(A_{i_q}\right)=\log_2 M\left(A_{i_1j_1}\right)-\log_2 M\left(A_{i_zj_z}\times A_{i_1j_1}\right) \tag{10}$$

These threshold parameters affect the size of the final decision tree branch. The greater the α value, the smaller the branch depth of the tree, while the greater the β value, the greater the branch depth of the tree. The setting of α and β has a great influence on the classification performance of the decision tree. Therefore, in order to achieve the best classification performance,

the most appropriate threshold must be selected. After many repeated steps, the appropriate threshold is finally obtained, and the decision tree is determined by trying different values of α and β.

4. English speech feature recognition system based on digital means

From the perspective of service design, English speech feature recognition system needs to consider the service value of the system in addition to users, products, interactive behaviors and scenarios. Therefore, English speech feature recognition is a system composed of five basic elements: person, process, object, product use scene, and service value. The English speech feature recognition system is designed around the above five basic elements. Its purpose is to coordinate the relationship between them and analyze their functions and properties to optimize the user experience. The system components are shown in Figure 1.

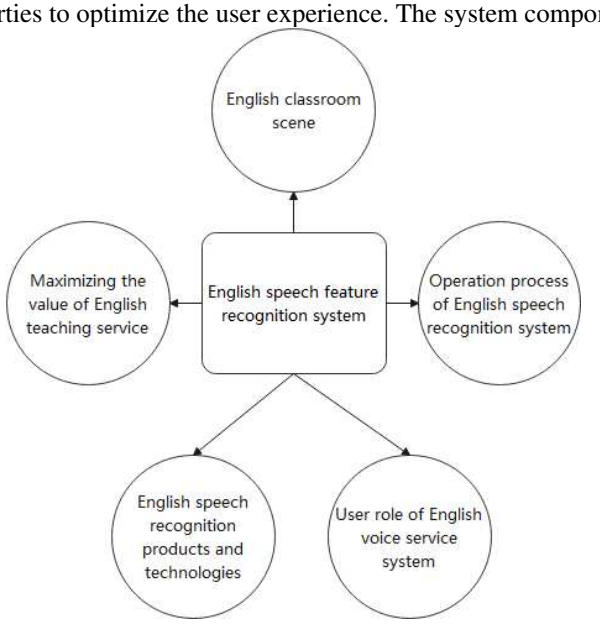
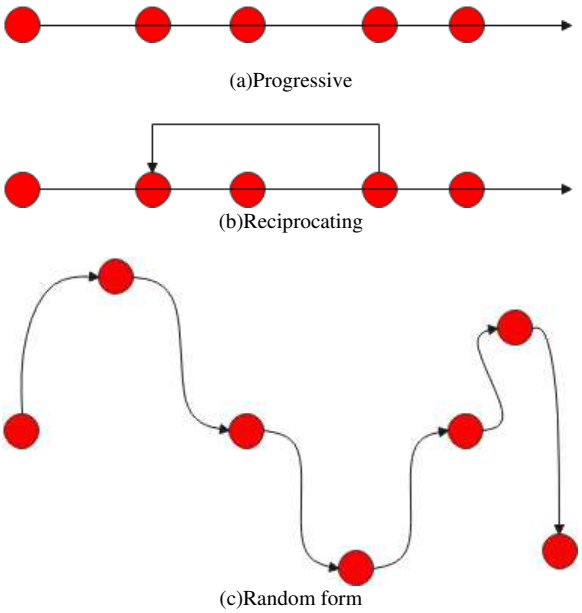


Figure 1 Elements of the system

The process mainly refers to the activities brought about by the user using the product in the service system, and the feedback behavior of the product to the user. It also includes the behavior between users and different stakeholders. Good behavior can make users interested, and make it more convenient, fast and comfortable to use, reduce mis-operations, and evoke better emotional experience. Figure 2 shows the interactive behavior path of the English speech feature recognition system based on digital means.



(a)Progressive

(b)Reciprocating

(c)Random form

Figure 2 Interactive behavior path

The English speech feature recognition system based on digital means is applied to the Internet of Things technology. The three main characteristics of the Internet of Things are comprehensive perception, reliable transmission and intelligent processing. From the perspective of the entire architecture of the Internet of Things, the Internet of Things consists of three major parts, from the low-end to the top-level are the perception layer, the network layer, and the application layer. The functional attributes carried by each layer are different. The perception layer uses various sensors to collect environmental data. The network layer is composed of mobile communication and the Internet, forming the bottom-end data transmission channel, and is responsible for transmitting the information collected by the sensors to the application layer, as shown in Figure 3.

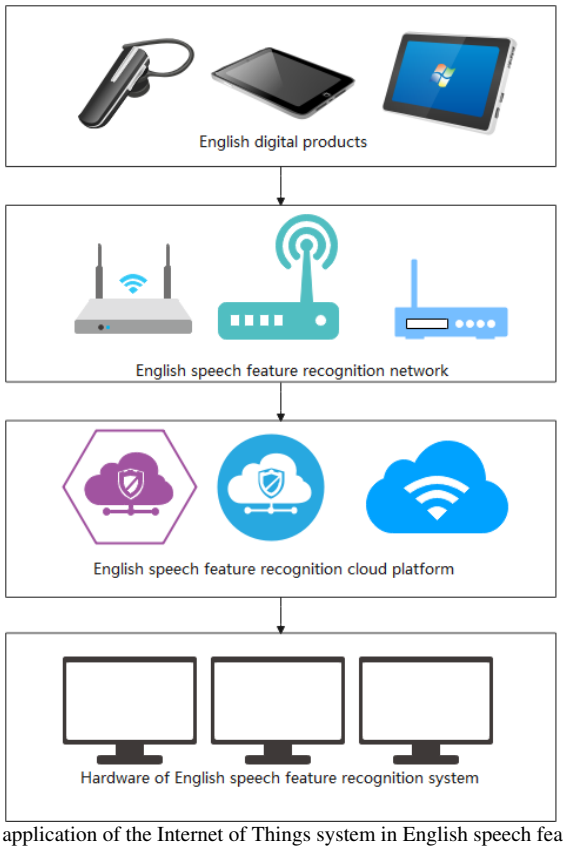


Figure 3 The application of the Internet of Things system in English speech feature recognition

With the development and maturity of technologies such as cloud computing and big data, artificial intelligence-related technologies have entered a period of explosive growth. This not only changes the innovation and transformation of traditional industries, but also penetrates into people's daily life and brings new behavioral experiences to human-computer interaction. Intelligent speech technology involves knowledge of multiple disciplines, such as acoustics, cognition, pattern recognition, artificial intelligence technology, etc. The system framework of intelligent voice contains 5 modules. The voice recognition module is responsible for accepting the user's voice input and converting it into text to the natural language understanding module. After understanding the semantics of user input, the natural language understanding module inputs specific expressions into the dialogue management module. The dialogue management module is responsible for coordinating the calls of various modules and maintaining the current dialogue status, and handing over the specific reply method to the natural language generation module for processing. The natural language generation module generates a specific reply text and inputs it into the speech synthesis module. The speech synthesis module is responsible for outputting the text to the user in the form of speech. The system framework of intelligent voice is shown in Figure 4.

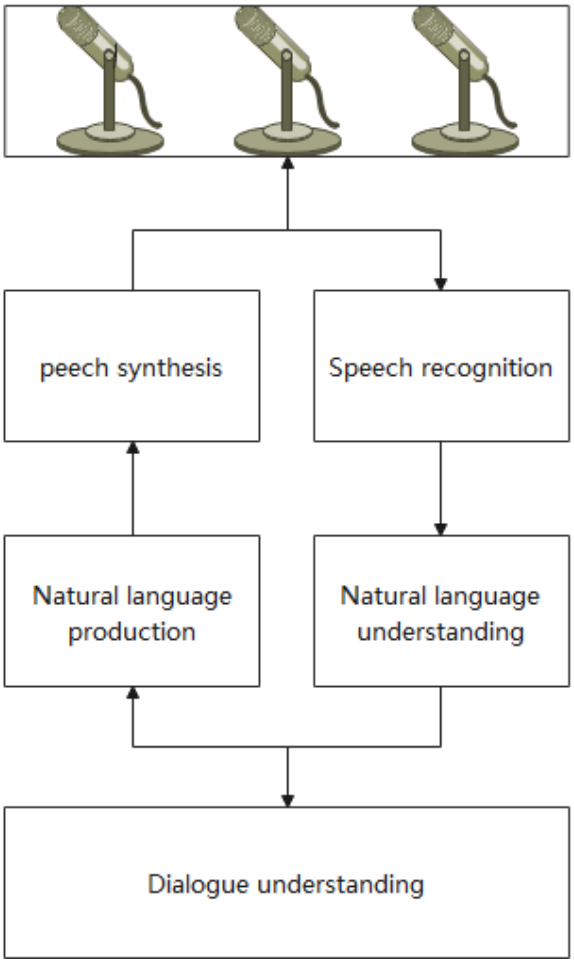


Figure 4 Intelligent voice related technologies

Intelligent voice brings users a new way of interaction, which is more flexible, natural and efficient than graphical interaction. After decades of development, the design process and specifications of graphical interactive design are very detailed, covering almost all graphical interactive scenes. From graphic interaction to voice interaction, the way of human-computer interaction has changed from visual to auditory and from tangible to intangible. The difference in the way of information transmission makes the elements of graphic interaction design and voice interaction extremely different. Graphic interaction design mainly takes six elements of color, font, motion, material, layout and shape as the main design objects, while the voice interaction design needs to use the timbre of intelligent voice as the object, as shown in Figure 5.

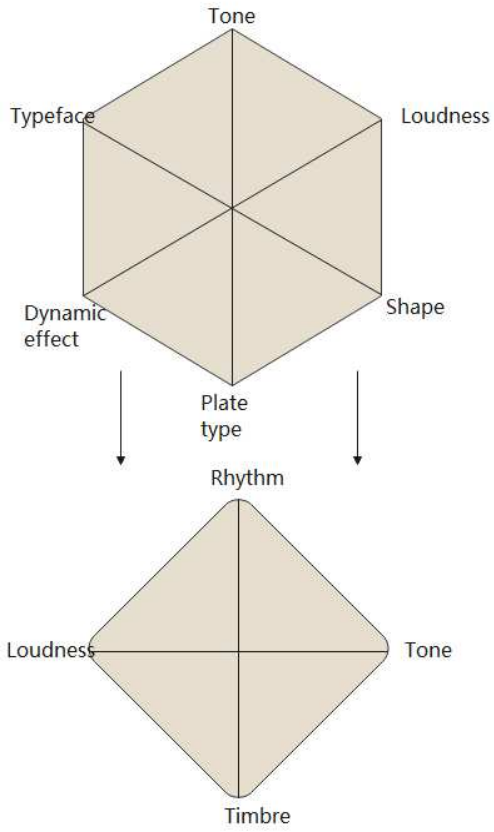


Figure 5 The difference between graphic interaction design and voice interaction design elements

The key figure map refers to the systematic representation of each key figure and the relationship between the key figures in the form of a chart or map. It can help designers quickly figure out the relationship between each role in the project, so as to discover the problems that exist in their interactive scenes, and provide help for the subsequent optimization of the experience. In the traditional sense, after users purchase goods, producers and consumers no longer have an interest relationship. However, with the transformation of product innovation models, producers have become service providers, and intangible services and tangible products are integrated and interdependent.

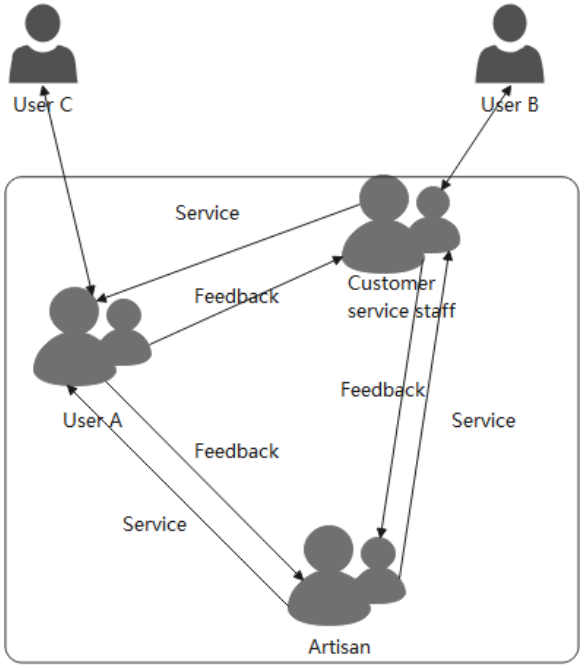


Figure 6 Map of key figures

In the process of service design, the systematic innovation method, from the overall perspective, comprehensively considers the elements of human, service and environment as well as the relationship between the elements, and then reasonably plans the combination order and cooperation degree of the elements in the system, so as to maximize the performance of the whole service system. Based on the system innovation, this paper analyzes the scattered problem points in the intelligent voice service, as well as many stakeholders that may be involved in the service system, and reestablishes the task relationship model. On the one hand, we need to pay attention to the pain points of users' experience in the whole process of using intelligent voice service, which is the key element of service optimization; On the other hand, people, products and environment elements in the intelligent voice service

system need to be regarded as an interactive, interdependent and comprehensive system with specific goals, so as to maximize the service of home English speech recognition system.

The whole system consists of five parts: PC, PCI adapter card, main switch, branch switch, and student terminal. The block diagram of the whole system is shown in 7:

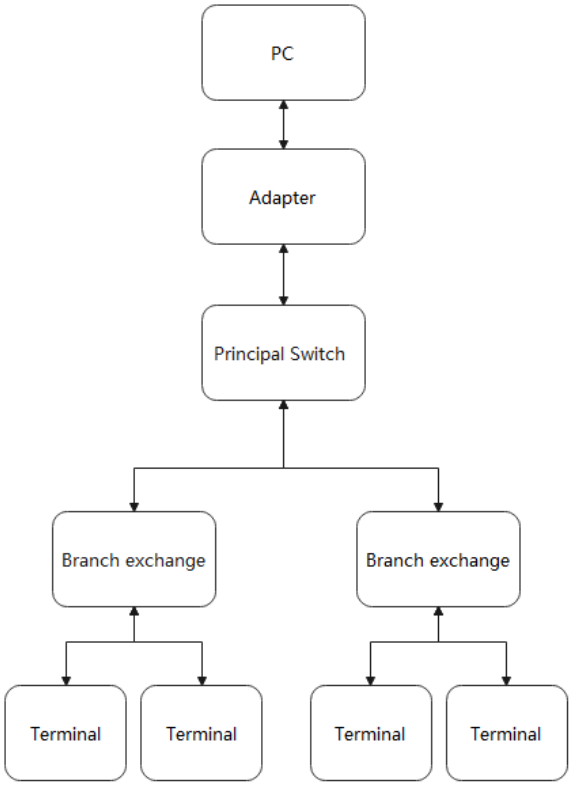


Figure 7 Block diagram of English speech feature recognition system

5. Performance test of English speech feature recognition system based on digital means

The performance of the English speech feature recognition system based on digital means constructed in this paper is tested through experimental research. The system of this paper mainly converts English speech into digitized form and then recognizes it. Therefore, in the experimental research, this paper mainly counts the effect of digitization conversion of English speech and the recognition effect of English speech features of this system. First of all, this paper carries out the statistics of the digital conversion effect of English speech of the system constructed in this paper. We use the system to identify 80 groups of student conversations, and use the system to convert the conversations into digital form. The results are shown in Table 1 and Figure 8.

Table 1 Statistical table of the effect of digital conversion of English speech

NO.	Digital effect	NO.	Digital effect	NO.	Digital effect
1	79.5	28	87.6	55	91.5
2	81.4	29	85.1	56	79.6
3	86.8	30	89.7	57	84.3
4	80.2	31	80.8	58	81.4
5	87.0	32	81.0	59	85.9
6	90.0	33	80.7	60	79.5
7	81.0	34	79.6	61	91.6
8	81.4	35	86.3	62	89.6
9	91.0	36	82.7	63	85.8
10	87.3	37	88.5	64	88.3
11	79.3	38	84.0	65	85.4
12	84.4	39	83.0	66	79.7
13	80.6	40	91.9	67	80.9
14	87.1	41	89.1	68	90.8
15	86.2	42	86.9	69	84.6
16	91.0	43	89.8	70	88.5
17	87.8	44	91.3	71	83.4
18	82.0	45	83.4	72	79.5
19	84.4	46	86.4	73	91.0

20	81.6	47	91.1	74	83.3
21	84.3	48	91.1	75	79.6
22	89.5	49	84.5	76	85.3
23	86.6	50	82.3	77	88.9
24	88.5	51	86.2	78	79.5
25	83.2	52	80.9	79	83.0
26	81.7	53	79.3	80	89.0
27	83.1	54	86.5		

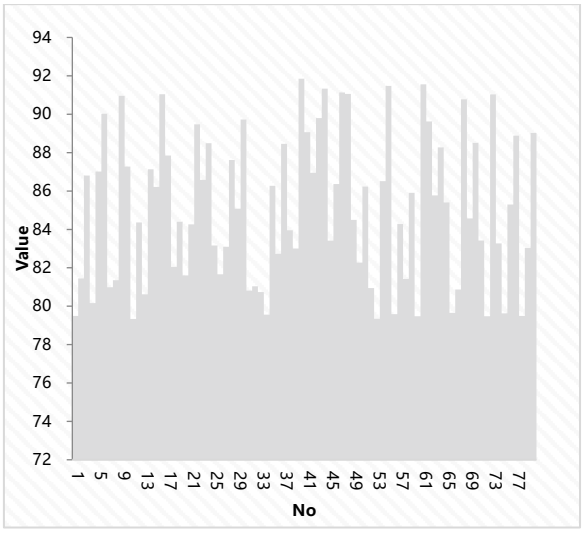


Figure 8 Statistical diagram of the effect of digital conversion of English speech

Through the above analysis, it can be seen that the English speech feature recognition system based on digital means constructed in this paper has a certain effect on the digital conversion of English speech. After that, this paper analyzes the accuracy of the English feature recognition of the system in this paper. The results obtained are shown in Table 2 and Figure 9.

Table 2 Statistical table of the accuracy of English feature recognition

NO.	Feature recognition effect	NO.	Feature recognition effect	NO.	Feature recognition effect
1	83.5	28	90.7	55	88.0
2	85.5	29	77.1	56	93.2
3	93.8	30	81.5	57	91.6
4	89.5	31	86.3	58	81.6
5	91.0	32	80.7	59	77.9
6	82.9	33	83.9	60	92.3
7	81.5	34	77.6	61	93.0
8	93.0	35	84.8	62	83.6
9	92.2	36	78.8	63	89.7
10	79.5	37	90.7	64	77.8
11	79.8	38	92.1	65	88.4
12	81.0	39	78.4	66	81.7
13	89.5	40	92.8	67	93.7
14	89.4	41	77.3	68	76.3
15	91.9	42	82.3	69	79.6
16	79.5	43	79.1	70	79.2
17	77.8	44	91.6	71	81.0
18	88.3	45	78.9	72	80.3
19	80.5	46	88.2	73	91.6
20	93.3	47	78.2	74	91.8
21	90.7	48	83.6	75	91.6
22	93.3	49	91.9	76	93.1
23	86.9	50	92.2	77	89.0
24	82.0	51	80.2	78	80.0
25	80.1	52	84.2	79	81.0
26	92.2	53	79.2	80	78.3
27	90.3	54	91.2		

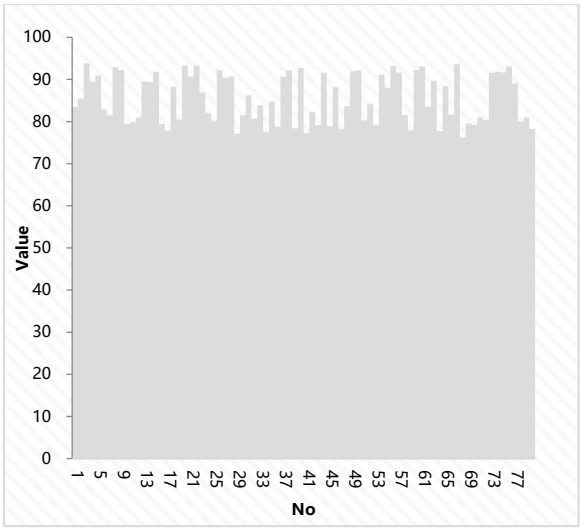


Figure 9 Statistical table of the accuracy of English feature recognition

From the above research, it can be seen that the English speech feature recognition system based on digitized segments constructed in this paper has certain effects and can play a certain role in intelligent English teaching.

6. Conclusion

English speech feature recognition plays an important role in supporting English learning. From a practical point of view, the improvement of English speech feature recognition through digital means can effectively improve the effect of English learning. This article combines the actual needs of English speech feature recognition to apply digital means to English speech recognition. In addition to the advantages of the high sensitivity to initial conditions, high-capacity dynamic storage, and low observability of digital chaotic signals, the reliability brought by digitization can effectively improve the ability of the chaotic system to resist channel interference and channel distortion, especially in its anti-attack and interception capabilities. In this paper, time-frequency analysis of chaotic signals and speech signals is carried out to eliminate noise in English speech features and improve the recognition effect of English speech features. Meanwhile, this paper constructs an English speech feature recognition system based on digital means to test the system's performance. The research results show that the method proposed in this paper has a certain effect.

DECLARATIONS

Funding : Not applicable

Conflicts of interest : The author has no conflicts of interest

Availability of data and material: Not applicable

Code availability: Not applicable

References

[1]. Rhodes, Richard. Aging effects on voice features used in forensic speaker comparison[J], international journal of speech language & the law, 2017, 24(2):177-199.
[2]. Ngoc Q. K. Duong, HienThanh Duong. A Review of Audio Features and Statistical Models Exploited for Voice Pattern Design[J], computer science, 2015, 03(2):36-39.
[3]. Sarria-Paja M , Senoussaoui M , Falk T H . The effects of whispered speech on state-of-the-art voice based biometrics systems[J], Canadian Conference on Electrical and Computer Engineering, 2015, 2015(1):1254-1259.
[4]. Leeman A , Mixdorff H , O'Reilly M , et al. Speaker-individuality in Fujisaki model f0 features: Implications for forensic voice comparison[J], International Journal of Speech Language and the Law, 2015, 21(2):343-370.
[5]. Hill A K , Rodrigo A. Cárdenas, Wheatley J R , et al. Are there vocal cues to human developmental stability? Relationships between facial fluctuating asymmetry and voice attractiveness[J], Evolution & Human Behavior, 2017, 38(2):249-258.
[6]. Marcin Woźniak, Dawid Połap. Voice recognition through the use of Gabor transform and heuristic algorithm[J], Nephron Clinical Practice, 2017, 63(2):159-164.
[7]. Haderlein T , Michael Döllinger, Václav Matoušek, et al. Objective voice and speech analysis of persons with chronic hoarseness by prosodic analysis of speech samples[J], Logopedics Phoniatrics Vocology, 2015, 41(3):106-116.
[8]. Nidhyananthan S S , Muthugeetha K , Vallimayil V . Human Recognition using Voice Print in LabVIEW[J], International Journal of Applied Engineering Research, 2018, 13(10):8126-8130.

[9]. Malallah F L , Saeed K N Y M G , Abdulameer S D , et al. Vision-Based Control By Hand-Directional Gestures Converting To Voice[J], International Journal of Scientific & Technology Research, 2018, 7(7):185-190.
[10]. Morgan Sleeper. Contact effects on voice-onset time in Patagonian Welsh[J], acoustical society of america journal, 2016, 140(4):3111-3111.
[11]. Mohan G , Hamilton K , Grasberger A , et al. Realtime voice activity and pitch modulation for laryngectomy transducers using head and facial gestures[J], Journal of the Acoustical Society of America, 2015, 137(4):2302-2302.
[12]. Kang T G , Kim N S . DNN-Based Voice Activity Detection with Multi-Task Learning[J], Ieice Transactions on Information & Systems, 2016, E99.D(2):550-553.
[13]. Choi, HaNa, Byun, SungWoo, Lee, SeokPil. Discriminative Feature Vector Selection for Emotion Classification Based on Speech[J], Transactions of the Korean Institute of Electrical Engineers, 2015, 64(9):1363-1368.
[14]. Herbst C T , Hertegard S , Zangger-Borch D , et al. Freddie Mercury—acoustic analysis of speaking fundamental frequency, vibrato, and subharmonics[J], Logopedics Phoniatrics Vocology, 2016, 42(1):1-10.
[15]. Al-Tamimi J . Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations[J], Laboratory Phonology, 2017, 8(1):1-40.
[16]. Abdel-Hamid O, Mohamed A, Jiang H, et al. Convolutional neural networks for speech recognition[J], IEEE/ACM Transactions on audio, speech, and language processing, 2014, 22(10): 1533-1545.
[17]. Kim C, Stern R M. Power-normalized cepstral coefficients (PNCC) for robust speech recognition[J], IEEE/ACM Transactions on audio, speech, and language processing, 2016, 24(7): 1315-1329.
[18]. Noda K, Yamaguchi Y, Nakadai K, et al. Audio-visual speech recognition using deep learning[J], Applied Intelligence, 2015, 42(4): 722-737.
[19]. Qian Y, Bi M, Tan T, et al. Very deep convolutional neural networks for noise robust speech recognition[J], IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(12): 2263-2276.
[20]. Li J, Deng L, Gong Y, et al. An overview of noise-robust automatic speech recognition[J], IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 22(4): 745-777.
[21]. Besacier L, Barnard E, Karpov A, et al. Automatic speech recognition for under-resourced languages: A survey[J], Speech Communication, 2014, 56(3): 85-100.
[22]. Watanabe S, Hori T, Kim S, et al. Hybrid CTC/attention architecture for end-to-end speech recognition[J], IEEE Journal of Selected Topics in Signal Processing, 2017, 11(8): 1240-1253.

Yuji Miao works on Faculty of International Languages in North China University of Science and Technology, Qinggong College, China. Her research interests include Applied Linguistics, Second Language Acquisition, Pedagogy . More than 5 papers published.

Yanan Huang works on Faculty of International Languages in North China University of Science and Technology, Qinggong College, China. Her research interest is foreign linguistics and applied linguistics. More than 5 papers published.

Zhen-

jing Da works on Faculty of International Languages in Qinggong College, North China University of Science and Technol ogy. His research interests include syntax, translation skills. More than 8 papers published and 1 book published.