

# Design of Intelligent Voice Acquisition System Based on Cloud Resource Scheduling Model

Zhiwu Cui<sup>1\*</sup>, Ke Zhou<sup>1</sup>, Jian Chen<sup>2</sup>

<sup>1</sup>Book Information Center, Shandong Vocational College of Industry, Zibo 256414, China

<sup>2</sup>Educational Information Center, Qingdao Vocational and Technical College of Hotel Management, Qingdao 266100, China

\*Corresponding author: czw78@163.com

Received: June 21, 2021. Revised: December 15, 2021. Accepted: January 11, 2022. Published: January 12, 2022.

**Abstract**—The existing acquisition system has the problem of imperfect communication link, which leads to the weak signal receiving strength of the system. This paper designs an intelligent voice acquisition system based on cloud resource scheduling model. **Hardware:** select S3C6410 as hardware platform, optimize audio access port, connect IIS serial bus and other components; **Software part:** extract the frequency agility characteristics of intelligent voice signal, predict the future sample value, establish the communication link with cloud resource scheduling model, obtain the communication rate information, code and generate digital voice data, set the transmission function of intelligent acquisition system with overlay algorithm. **Experimental results:** the average signal receiving strength of the designed system and the other two intelligent voice intelligent acquisition systems is 106.40 dBm, 91.33 dBm and 90.23 dBm, which proves that the intelligent acquisition system integrated with cloud resource scheduling model has higher use value.

**Keywords**—Cloud resource scheduling model; Intelligent voice; Intelligent acquisition system; Voice signal; Communication link; Overlay algorithm.

## I. INTRODUCTIONS

Voice is a kind of voice that human body sends out through a series of vocal organs, which has information value and can be used for communication. People can easily and naturally transfer information through direct, accurate and efficient voice communication. Although the sources of human access to information are increasing, voice, text and image are still the most basic, important and accurate sources. In information communication, the natural advantage of voice is irreplaceable by other ways. On the one hand, voice

contains not only language content, but also emotion, which can make the audience more accurately grasp the real intention of the speaker. On the other hand, voice is closely related to the improvement of language and communication ability of human society. In the accurate performance of various situations, it fully embodies the intuitive and efficient characteristics. At present, with the continuous development of intelligent AI technology and information network technology, intelligent voice, as a necessary direct form of input and output, has become a hot topic of research from all walks of life [1], [2]. As early as the 1960s, the intelligent voice acquisition system has been studied by scientists all over the world. In the 1970s, speech acquisition has formed some basic research results on the basis of the continuous development of digital signal processing and signal processing theory. After continuous development, it has gradually formed an independent scientific research and technology field [3]. In the 1980s, it entered the practical stage of voice acquisition system. In the 1990s, intelligent speech acquisition has attracted the attention of scientists, and has become one of the hottest research directions. Although the domestic research on intelligent voice acquisition system started late, in recent years, a relatively perfect system has been formed in the theoretical research of its system, and some research results have been produced.

[4] presented a hidden Markov model optimization method based on a genetic algorithm (GAs) that can collect and transmit speech with a Gaussian mixed density. When assigning probability values to HMM, a GAs-based method is introduced to find out the appropriate probability values so that the case ratio of probability values for HMM has been used to assign probability values. The proposed method enables automatic speech acquisition, but the collected speech signal intensity is weak. [5] based on the characteristics of FPGA efficiency, flexibility and high integration, a high-performance audio processor platform is built through the interconnection between localbus bus and

DSP processor. The digital design method and table tennis operation logic are used to realize the real-time collection, exchange and mixing processing of multi-user voice signals to meet the needs of the system design. The speech signal intensity collected by this method is insufficient.

Based on the shortcomings of the above research results and further improve the intensity of signal acquisition, this paper designed an intelligent voice acquisition system based on the cloud resource scheduling model. The hardware of the system selects S3C6410 as the hardware platform to optimize the audio access port to lay a physical foundation for improving the strength of audio acquisition; the software design of the system is the innovation point of this paper, predict future sampling value based on the frequency agility characteristics of voice signals; use cloud resource scheduling model to establish communication links, encode and generate digital voice data, and optimize the transmission function of intelligent acquisition system with superposition algorithm. The experimental results show that the present system enhances the strength of the speech signal and improves the comfort of the human-computer experience.

## II. HARDWARE DESIGN OF INTELLIGENT VOICE ACQUISITION SYSTEM BASED ON CLOUD RESOURCE SCHEDULING MODEL

According to the needs of the system, S3C6410 is selected as the hardware platform. The hardware support provided by the platform for the development board is: ARM11 microprocessor, recording device, flash memory, RS232 serial port, A/D and D/A converter, HDMI, SDRAM memory decoder and touch screen. The specific hardware configuration is shown in Figure 1.

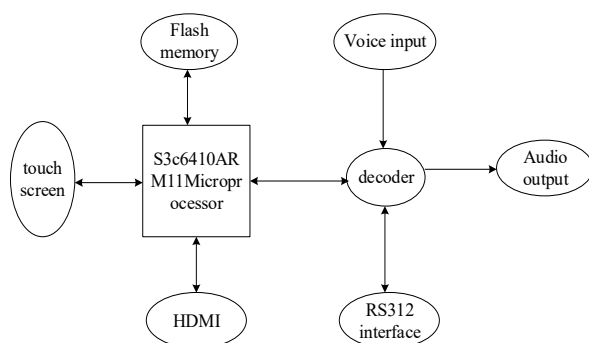


Figure 1. System hardware architecture

As can be seen from Figure 1, the microprocessor (MPU) adopts S3C6410, the architecture is ARM 11, and the main frequency reaches 800 MHz. Compared with S3C2440 of ARM9 architecture, the operation speed has been greatly improved. Flash uses NAND Flash memory developed by Samsung, which has 256 MB storage space. NAND Flash memory is an ideal solution for high data storage density. SDRAM memory adopts Hynix Company's, using two pieces, each 128 MB. The interface equipment required by this platform: sound output interface and sound input

interface, an adjustable resistor, can be used for A/D conversion test. Among them, arm 11 microprocessor manages the whole system and provides various operations [6], [7]. Recording equipment is to obtain voice signals from the outside world. A/D and D/A converters are used for the conversion of voice signals between digital signals and analog signals to prepare for the later voice processing. Flash is mainly used to solidify the boot program, embedded operating system, application program and various data of the system. RS232 serial port is mainly used to connect the development board and PC to provide convenience for development [8], [9]. In order to minimize the output pins of the chip, this paper optimizes the audio access port. The connection of IIS serial bus is: data line (providing time-sharing recovery function), clock signal line and field selection line (sound channel selection). Now the mainstream audio codec and embedded microprocessor support IIS bus standard [10]. Using the built-in IIS standard audio interface of MPU, it can connect with most audio decoders and has good compatibility. At the same time, S3C6410 also provides DMA data transmission mode, so that the embedded system can directly transmit data without the microprocessor and realize the simultaneous sending and receiving of data, which is very practical in the case of limited system resources. UDA1341 supports the data format of IIS (inter IC sound bus) bus, the connection mode with S3C6410 is very simple, and it has the playback mode with DSP characteristics, so it is very suitable for the development of voice acquisition system. Using IIS bus to solve the data interaction between S3C6410 and digital audio codec. Based on this, the system hardware design steps are completed.

## III. SOFTWARE DESIGN OF INTELLIGENT VOICE ACQUISITION SYSTEM BASED ON CLOUD RESOURCE SCHEDULING MODEL

### A. Frequency agility feature extraction of intelligent speech signal

In the whole process of feature extraction, because the features and feature parameters of speech are time variables, it is a non-stationary process and cannot be analyzed and processed by the existing methods of processing digital signals. However, in a short period of time, its characteristic is basically a fixed value, that is, a stable state. Voice sensor nodes can not guarantee the stability of communication channel quality in the process of data communication. When the communication channel quality decreases due to interference, it will cause the increase of voice packet loss rate in the process of communication. If effective measures can not be taken, it is likely to cause the decline of voice communication quality or even the interruption of communication. Voice signal windowing can detect whether the node is subject to continuous interference, and if it is continuous interference, change the node communication channel [11]-[13]. The expression formula of voice signal windowing is as follows:

$$F_q(h) = F(h) \times q(h) \quad (1)$$

In Formula (1),  $F$  is any window function,  $q$  is amplitude frequency value, and  $h$  is frame length. If the interference signal comes from the frequency hopping interference source, the node communication channel will not be changed, and the selected new communication channel has the least interference. It can be seen that the sensor node will choose a new communication channel to resume communication only under continuous interference. According to the protocol, the communication between the sending and receiving nodes adopts the mechanism of waiting for retransmission, that is, each time the sending node sends a data frame, and the receiving node is required to respond [14], [15]. Therefore, the sending node can detect the current communication channel through the received response frame. There is a strong correlation between speech signal samples, which can be used to approximate or predict the future sample values by using the linear combination of sample values at some time in the past, the expression formula is as follows:

$$Q(h) = \sum_{i=1}^e \delta_i F(h-i) \quad (2)$$

In Formula (2),  $\delta$  is the first-order statistic,  $F$  is the arbitrary window function,  $h$  is the frame length,  $i$  is the time series, and  $e$  is the main lobe value of hamming window. If the current channel is in good condition, the sending node can receive all the response frames correctly, otherwise the response frames are lost. By calculating the number of response frames lost continuously, we can judge whether the current channel is continuously interfered. The receiving node can judge the status of the current channel by counting the number of received voice data frames. When the packet loss rate of received data exceeds a certain threshold, the quality of the current communication channel can be considered to be degraded. When the communication channel is subject to continuous interference, the transmitting node reselects the communication channel through the idle channel selection method, and the communication channel selected by this method is the channel with the least interference. After the channel selection is completed, the sending node continues to send voice data frames on the channel until the response frame is received or the maximum waiting time is exceeded. It is assumed that speech signal sequence is stationary and random in a certain period of time, in the autocorrelation method, it is defined as:

$$w = \sum_{k=t}^k F(h) \times (k-t) \quad (3)$$

In Formula (3),  $F$  is any window function,  $h$  is frame length,  $k$  is autocorrelation sequence, and  $t$  is random variable. The receiving node reestablishes communication with the sending node by means of channel scanning and

monitoring. The receiving node first sets the communication channel as 0 B channel, and monitors the intelligent voice data frame on this channel. When receiving the intelligent voice data frame from the sending node, the communication channel is set as the current channel. Otherwise, if the maximum waiting time is exceeded, the channel switches to the next channel to monitor again until the communication is reestablished with the sending node. The adaptive frequency agility characteristics of the sending node will change according to the actual situation. When the response frame lost by the sending node exceeds the preset threshold value, the voice signal collection will be stopped and the smallest channel will be re selected as the new communication channel. Then the sending node will send the empty intelligent voice data frame on the new communication channel and wait for the response frame to be received. If the response frame is received, the answer number will be lost and cleared, and then the intelligent voice signal acquisition will be turned on again, otherwise a new communication channel will be selected again. Based on this, the frequency agility feature extraction of intelligent speech signal is completed.

#### B. Establishing communication link based on cloud resource scheduling model

The essence of resource scheduling is to allocate multiple independent tasks to the available resources of multiple organizations, so as to minimize the total task completion time [16]. The cloud computing scheduling model produces a large number of tasks after receiving user demands, and reasonably assigns tasks to different computational nodes to maximize task execution efficiency. Cluster resource scheduling model can also be used as an alternative to scheduling resources, but in practice, the cluster resource scheduling model has a low positioning degree to users, the resource scale that can be scheduled is small, while the cloud resources have a high positioning degree to users, and the resource scale is large. Because the essence of cloud resource scheduling is resource management, such as dynamically adjusting the voltage according to the load of the task, reducing the frequency of CPU, and achieving the purpose of energy saving. Therefore, this paper adopts the cloud resource scheduling model for resource scheduling. In a deterministic environment, it is often assumed that the time virtual computing nodes perform tasks in a cloud resource scheduling model is a value. The calculation formula of the corresponding relationship matrix between task and virtual machine is as follows:

$$S = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & & \vdots \\ S_{m1} & S_{m2} & \cdots & S_{mn} \end{bmatrix} \quad (4)$$

In Formula (4),  $s$  is the scheduling task,  $m$  is the behavior virtual machine queue, and  $n$  is the task queue.

According to people's interest points, dynamically close or open the corresponding process. According to the principle that a task can only be executed on one virtual machine, establishing a communication link is regarded as a scheduling scheme, the calculation formula of the time spent is as follows:

$$T = \sum_{p=1}^r S \times g \quad (5)$$

In Formula (5),  $S$  is the correspondence matrix,  $g$  is the scheduling target, and  $r$  is the link communication channel,  $p$  is the communication rate. According to the calculation results, the communication protocol, communication rate and modulation mode are determined. There are many communication protocols recommended by ITU, and the same communication protocol includes many communication rates and modulation modes. In order to receive and demodulate the data correctly, the receiver should know what communication protocol, what rate and modulation mode the sender uses to transmit the data. Test whether the channel can complete the communication at the set communication rate. The state of the channel determines whether the information can be transmitted normally during communication. If there is more interference on the channel, the correctness of the information can not be guaranteed when the data is transmitted at a higher communication rate. There are many errors after the information is transmitted, which seriously affects the intelligibility of the information. At this time, it is necessary to reduce the rate to ensure the demand of low error rate [17], [18]. During the construction of the link, both sides of the communication evaluate the channel condition by sending the training sequence to judge whether it meets the BER requirements. If the bit error rate does not meet the requirements, the modem will reduce the rate and train again until the link can be built at the rate allowed by the bit error rate. The establishment process of communication link can be divided into two stages. The first stage is to set modem register to start the process of establishing link. The second stage is to handshake and send training sequence according to the register setting. At this time, the program queries the status of modem chip, judges whether the establishment of link is successful, and obtains the communication rate information. In addition, when the voice/fax module requests to establish a connection, it also needs to consider giving the connect instruction, and when the voice/fax module actively establishes a connection, it needs to provide the number of the other party for the voice/fax module to establish a connection. So the program design does not need too much algorithm principle, just need simple logic operation, implementation is relatively simple. The voice/fax module can modulate/demodulate the voice information. The voice information collection is realized by microphone. When playing, the voice can be played out by a voice player. The module also reserves a voice device interface, which can be used directly. Based on the above

description, the steps of establishing communication link are completed.

### C. Overlay algorithm to set the transmission function of intelligent acquisition system

Overlay algorithm, which is the intersection of different data layers in the target area, creates a new cut data set and quantitatively analyzes the spatial range and characteristics of the interaction between different types of spatial objects. In order to improve the transmission accuracy of the compressed voice transmission system without increasing the spectral overhead of the system, the collected voice compression is used to transmit the superposition algorithm. Characteristic information of speech signal; extracted feature information is superimposed on the compressed voice signal in superposition sequence for transmission. Compared with the traditional compressed voice transmission method, the superposition algorithm can still improve the noise with a high signal-to-noise ratio, without increasing the spectral cost of the transmission system. The superposition algorithm improves the efficiency of voice transmission by simplifying the data package and possibly improving the proportion of voice data in each data packet. In addition, by turning off the automatic response and retransmission function, the system transmission delay is reduced, the code operation efficiency is improved, so as to improve the actual transmission rate of voice and ensure the voice quality. Because the voice transmission is not very sensitive to the packet loss rate, it is not necessary to ensure that every data packet must be sent successfully, but it is very sensitive to the software transmission efficiency and delay. Therefore, in the voice transmission part, the sender is only responsible for sending, while the receiver is only responsible for receiving. The data reading and writing in the process of sending voice is realized by SPI port. First, the wireless module enters the transmission mode and waits for transmission. Then the voice code is activated, and the voice module starts to collect the voice signal, encodes and generates the digital voice data, which is stored in the cache port. After that, the voice data is read out through the main controller and written into the memory. Because the wireless module sends 32 bytes of voice data each time, when the cache is full of 32 bytes of voice data, it will be sent out immediately. In addition, in this process, LED1 light will flash once every 8000 bytes of voice data is read and written to indicate that the system is working in the transmission state. In the process of speech transmission, the far-field narrow-band signal processing model is the foundation, so it can be regarded as a plane wave. When there is a uniform linear array composed of one element, the far-field narrow-band signal can be expressed as:

$$\eta = \tilde{f}(\varphi) d^{\frac{1}{2}} \quad (6)$$

In Formula (6),  $f$  is carrier angular frequency,  $\varphi$  is signal complex envelope,  $d$  is signal phase. It is known that the traditional array usually processes the far-field

signal, but because of the particularity of acoustic transmission, the system often processes the near-field signal. The incident wave of far-field signal can be approximately plane wave, and that of near-field signal can be approximately spherical wave, then the expression formula of the received signal of the array element is as follows:

$$y = l(u - \lambda) \quad (7)$$

In Formula (7),  $l$  is the distance between adjacent array elements,  $u$  is the wavelength difference of array element signal, and  $\lambda$  is the number of array elements. Because the frequency switch of the transmitter is controlled by the key, the wireless module will terminate the voice transmission of the current frequency every time the send key is pressed in the process of voice transmission. After that, 8 communication interrupt packets are continuously sent to the receiver at the current frequency. Then change the transmission frequency in order, jump to the next frequency point and wait for the establishment of communication. After the receiver enters the receiving mode for the first time, it will automatically search the frequency and select a receiving frequency. The implementation method is to let the receiver wait on each frequency point in turn, that is, wait for the received data on the broadcast frequency, two multicast frequencies and each unicast frequency in turn. Once the data is received on a certain frequency point, the receiver will establish communication with the sender on this frequency point, and then carry out voice and data transmission [19]-[21]. In the process of voice transmission, the receiver will automatically judge whether it has received the communication interrupt data. If it has received the interrupt data, it will jump out of the frequency point and return to the frequency scanning cycle to select the frequency point again, the calculation formula of spatial phase of frequency point is as follows:

$$\sigma = \frac{2\pi z \cdot \cos\theta}{z} \quad (8)$$

In Formula (8),  $z$  is the carrier wavelength and  $\theta$  is the signal incidence angle. According to the calculation results, the operation types of spatial data can be divided into intersection, union, inversion and combination modes. In the merge mode, the outer loop of the new sequence selects the largest outer loop set of the boundary lines of the target and the input two sequences, while the inner loop retains all the attributes of the two input feature sets. The intersection mode only preserves all the data of the common part of the target sequence and the input sequence in the topological relationship. The inversion mode retains the remaining data of the target sequence after removing the intersection part of the input sequence. After the communication link is established, both sides of the communication begin to transmit data. The data transmission stage is divided into four parts: bit synchronization, transmission command code, transmission

session key and transmission voice data. In the stage of in place synchronization and session key transmission, the voice collector and the local adapter send data to each other. After that, the local adapter sends the command code to the voice collector according to the control command of the host computer, and the voice collector completes different types of voice transmission operations according to different command codes. In the stage of voice transmission, the voice collector receives, sends or simultaneously sends voice data according to the command code sent by the local adapter. Data transmission is divided into four states: bit synchronization, command code transmission, session key transmission and voice transmission. When the system queries the success mark of chain building, it starts the data transmission and enters the bit synchronization state. After the synchronization is completed, the command code is transmitted. According to the command code, the session key transmission mode is determined. After the session key transmission is completed, it enters the voice transmission state. When the voice is transmitted, the link state is detected at the same time. When the link off flag is detected, the data transmission is ended and the reset operation is performed. Based on the above description, the link of setting system transmission function is completed.

#### IV. EXPERIMENT AND RESULTS ANALYSIS

##### A. Setting up experimental environment

In order to verify the effectiveness of the design of intelligent acquisition system, the simulation test is carried out, and the experimental results are obtained. Using the serial debugging tool comtools, the frame data of mg2455 can be transmitted to the display interface of comtools through the serial port through programming, which can be used to verify whether the node can send and receive data correctly in the communication process. Set the serial port parameter format as “BBBB, P, D, s”, where BBBB represents baud rate, P represents parity, D represents data bits, and s represents stop bits. Here, the baud rate is set to 115200, the parity check is set to no check state, the number of data bits is set to 8, the stop bit is set to 1, and the data display mode is set to hexadecimal. WM8974 provides Loopback function. When loopback function is enabled, the data output by ADC audio interface will be fed back to DAC data input directly. After the voice sensor node completes the connection, the sending node will continuously collect voice signals and transmit them to the receiving node. On the basis of the above experimental environment, the experimental test is carried out.

##### B. Experimental result

In the experiment, the method proposed by Benmachiche and Makhlof [4] and the method proposed by Ge [5] are selected and compared with the designed system. Under different communication distance conditions, the signal receiving strength of the three systems is tested respectively.

The larger the value is, the better the system performance is. The experimental results are shown in Figures 2-4.

As can be seen from Figure 2, the average signal receiving strength of the designed system and the other two intelligent voice acquisition systems is 132.25 dBm, 121.67 dBm and 118.52 dBm; As can be seen from Figure 3, the average signal receiving strength of the designed system and the other two intelligent voice acquisition systems is 121.46 dBm, 106.31 dBm and 103.83 dBm; As can be seen from Figure 4, the average signal receiving strength of the designed system and the other two intelligent voice acquisition systems is 65.51 dBm, 46.02 dBm and 48.36 dBm, which indicates that the practical application effect of the designed system is better. This is because S3C6410 is selected as the hardware platform to optimizes the audio access port, improves the strength of the audio signal from the physical basis; the software design uses the superposition algorithm to optimize the transmission function of the intelligent acquisition system, the receiver automatically determines whether the communication interruption data is received, and reselects the frequency point when the signal interruption to ensure the reception strength of the signal.

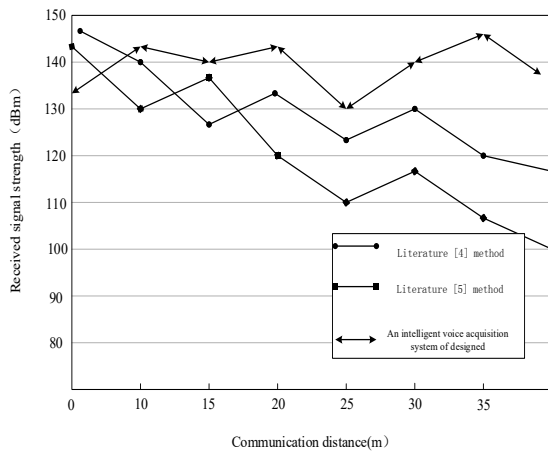


Figure 2. Signal receiving strength of 35 m communication distance

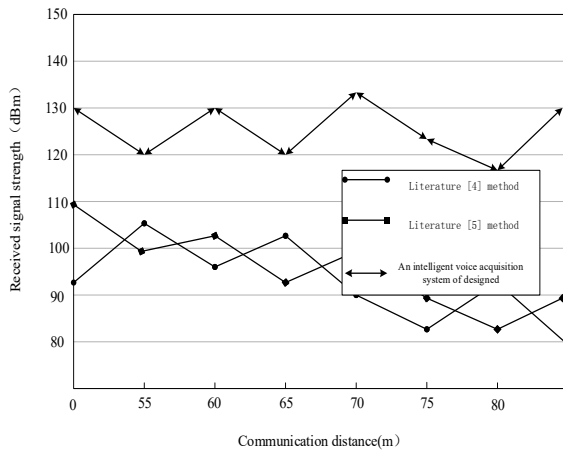


Figure 3. Signal receiving strength of communication distance 80 m

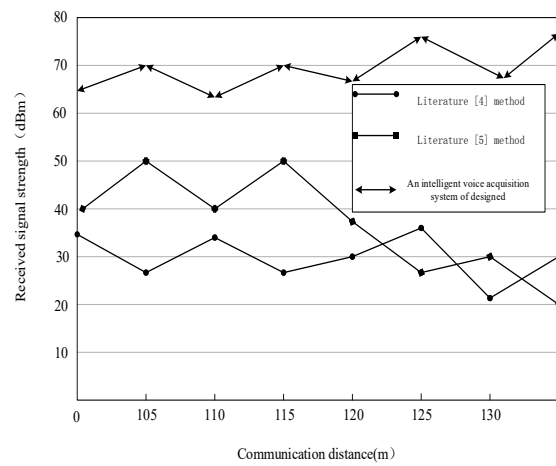


Figure 4. Signal receiving strength of communication distance 130 m

## V. CONCLUSIONS

In order to improve the voice intensity of the existing voice signal acquisition system, the intelligent voice acquisition system based on the cloud resource scheduling model. The system selects S3C6410 as the hardware platform to optimize the audio access port; the software design of the system to predict the future sampling value based on the frequency agility characteristics of the cloud resource scheduling model to establish communication links, generate and encode digital voice data; and optimize the transmission function of the intelligent acquisition system. This optimization method is the innovation point of this paper. The optimization method can still improve the speech with a high signal-to-noise ratio without increasing the spectral cost of the transmission system. The experimental results show that the present system significantly enhances the strength of the speech signal. The design of the intelligent acquisition system has laid a theoretical and practical foundation for the academia to carry out related research, and broadened the application field of cloud resource scheduling model. At the same time, it enriches the literature in the field of intelligent voice acquisition, and provides a feasible method for system optimization. Due to the limited research conditions, this paper has not been tested in the level of network false data identification. Future studies can identify false data on the basis of collecting audio data and continue to improve the performance of the system.

## REFERENCES

- [1] C. E. Chelloug and A. Farrouki, "Real time implementation of voice activity detection based on false acceptance regulation", *International Journal on Electrical Engineering and Informatics*, vol. 12, no. 3, pp. 654-666, 2020.
- [2] R. M. Garg, "Intelligent voice assistant for desktop using NLP and AI", *International Journal for Modern Trends in Science and Technology*, vol. 6, no. 12, pp. 328-331, 2020.

- [3] G. Stamatescu, A. Sachenko and P. Dan, "Convergence of intelligent data acquisition and advanced computing systems". *Sensors*, vol. 21, no. 7, pp. 2262, 2021.
- [4] A. Benmachiche, A. Makhlof, "Optimization of hidden Markov model with Gaussian mixture densities for Arabic speech recognition", *WSEAS Transactions on Signal Processing*, vol. 15, pp. 85-95, 2019.
- [5] M. L. Ge, "The process and application of audio data exchange control based on FPGA", *Electronic Design Engineering*, vol. 27, no. 12, pp. 189-193, 2019.
- [6] Z. Lin, J. Lu and X. Qiu, "An effective hybrid low delay packet loss concealment algorithm for MDCT-based audio codec". *Applied Acoustics*, vol. 154, pp. 170-175, 2019.
- [7] S. Raj, P. Prakasam, and S. Gupta, "Multilayered convolutional neural network-based auto-codec for audio signal denoising using MEL-frequency cepstral coefficients". *Neural Computing and Applications*, vol. 33, pp. 10199-10209, 2021.
- [8] R. Hinrichs, T. Gajecski, J. Ostermann and W. Nogueira, "A subjective and objective evaluation of a codec for the electrical stimulation patterns of cochlear implants". *The Journal of the Acoustical Society of America*, vol. 149, no. 2, pp. 1324-1337, 2020.
- [9] C. Y. Yeh and H. Y. Lai. "High-efficiency vector quantization codebook search algorithms for extended adaptive multi-rate-wideband audio coder". *Sensors & Materials*, vol. 31, no. 1, pp. 153-163, 2019.
- [10] S. Seshadri and O. Rasanen. "Synnet: An adaptable end-to-end syllable count estimator for speech". *IEEE Signal Processing Letters*, vol. 26, no. 9, pp. 1359-1363, 2019.
- [11] S. Kaul and A. Jain. Opus and session initiation protocol security in voice over IP (VOIP). *European Journal of Engineering Research and Science*, vol. 4, no. 12, pp. 27-37, 2019.
- [12] S. N. Kirillov, D. Y. Mamushev, A. N. Kuchurkin and E. V. Ban', "Noninvasive simultaneous diagnosis of multiple upper respiratory tract diseases using speech signal parameters". *Biomedical Engineering*, vol. 54, no. 3, pp. 216-219, 2020.
- [13] H. Relao-Iborra, J. Zaar and T. Dau, "A speech-based computational auditory signal processing and perception model". *The Journal of the Acoustical Society of America*, vol. 146, no. 5, pp. 3306-3306, 2019.
- [14] H. S. Das and P. Roy. "Optimal prosodic feature extraction and classification in parametric excitation source information for Indian language identification using neural network based q-learning algorithm". *International Journal of Speech Technology*, vol. 22, no. 1, pp. 67-77, 2019.
- [15] Y. A. Alotaibi, S. A. Selouani, M. S. Yakoub, Y. M. Seddiq and A. Meftah, "A canonicalization of distinctive phonetic features to improve Arabic speech recognition". *Acta Acustica united with Acustica*, vol. 105, no. 6, 1269-1277, 2019.
- [16] M. A. Isah and B. S. Kim. "Integrating schedule risk analysis with multi-skilled resource scheduling to improve resource-constrained project scheduling problems". *Applied Sciences*, vol. 11, no. 2, 650, 2019.
- [17] B. Kwaha, E. P. Ogherowo and J. T. Zhimwang, "Transmission evaluation of mobile cellular voice and data services under rain condition in Jos. Nigerian Journal of Physics", vol. 28, no. 1, pp. 155-162, 2019.
- [18] H. Huang, X. Wang and M. Hu, "Applied to mobile multimedia intelligent speech system interactive topic guiding model". *IEEE Access*, vol. 7, pp. 182348 - 182356, 2019.
- [19] P. Byahatti and M. S. Shettar, "Fusion strategies for multimodal biometric system using face and voice cues." *IOP Conference Series: Materials Science and Engineering*, vol. 925, no. 1, pp. 12031, 2020.
- [20] H. B. Zhang, "Simulation of depth spectrum feature extraction method for speech signal under Mel frequency". *Computer Simulation*, vol. 37, no. 5, pp. 197-200, 267, 2020.
- [21] S. Yin, H. Fang and X. Hou, "Athlete's respiratory frequency and physical energy consumption model based on speech recognition technology". *International Journal of Speech Technology*, vol. 23, no. 4, pp. 389-397, 2020.

**Creative Commons Attribution License 4.0  
(Attribution 4.0 International, CC BY 4.0)**

This article is published under the terms of the Creative Commons Attribution License 4.0

[https://creativecommons.org/licenses/by/4.0/deed.en\\_US](https://creativecommons.org/licenses/by/4.0/deed.en_US)