

Research Article

Borderless Fusion Financial Management Innovation Based on Speech Recognition Technology

Lin Yang ^{1,2}, Zehao Fan ³, and Jiaqi Zhou¹

¹College of Economics and Management, Beijing Institute of Petrochemical Technology, Beijing 102617, China

²Enterprise Development Research Center of Beijing Institute of Petrochemical Technology, Beijing 102617, China

³Beijing Haidian Foreign Language Shiyuan School

Correspondence should be addressed to Zehao Fan; zehaofan@bjfiles.com

Received 4 July 2022; Revised 30 July 2022; Accepted 17 August 2022; Published 14 September 2022

Academic Editor: Juan Vicente Capella Hernandez

Copyright © 2022 Lin Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, the rapid development of information technology has affected the operation of the world economy, and the emergence of e-commerce has decreased significantly. The development of modern information technology, especially the development of Internet technology, provides a solid foundation for e-commerce. The time and space gap between economic operators is shortened, and resources are shared as widely as possible. Based on a comprehensive and in-depth study of language recognition technology. However, the current e-commerce financial management still uses the traditional accounting method, which is inefficient and cannot be integrated with new technologies. This paper proposes a design scheme for a nonspecific person and small vocabulary language recognition system based on SPCE061A processor. This integrated system can be combined with data acquisition and language input technology and constitute a data acquisition and input system. It is used for data collection and input of instruments and meters in the workplace to improve work efficiency. This paper proposes algorithms based on time-domain energy and zero-crossing rate detection, detection based on entropy function method, and detection based on the double-threshold method. The speech recognition technology is analyzed in detail, and a new model of financial management is constructed. Accounting information disclosure is a virtuous circle of “promoting capital management with appropriate decentralization.” The experimental results in this paper show that through model construction, the calculated WER is 0 dm, 7 dm, 15 dm, and 11 dm, respectively, and it is concluded that the system exhibits good antinoise performance.

1. Introduction

1.1. Background. Language is one of the important ways for people to communicate, and it is the result of useful information that people retain in the long-term evolutionary process. With the advent of networked technology, in addition to communicating with people using language, it has become a dream for people to allow machines to understand human language. This problem that was previously considered impossible to solve has now become possible. With the advent of the information age, the use of voice instead of computer keyboard input has become one of the future social development trends, and voice recognition technology has emerged at the historic moment and has become one of the hotspots of research by many researchers. With the

continuous development of computer and artificial intelligence technology, smart machines participate in human production activities and social activities. How to improve the relationship between people of these machines and make it easier for people to manage machines. More importantly, language is the best way for humans and machines to communicate. Speech processing is a cross-discipline between language and digital signal processing technology. It is an important method of computer intelligence and human-computer interaction. The main technology of the interaction between humans and computers is speech recognition. Therefore, the research of speech recognition technology has become the research focus of countries all over the world. The development of computers and the importance of social life are becoming more and more prominent. Various

applications of voice coding technology have been developed by the product, including voice telephone exchange. Voice dialing systems, industrial control systems, etc., have penetrated into all sectors of society and have good application prospects. Boundless management was first aimed at the overall management of an enterprise. According to its meaning, the borderless theory is applied to the field of financial management and combined with the characteristics of integrated finance, a borderless integrated financial management model is proposed.

1.2. Significance. Speech recognition is a comprehensive subject, including sound, computer science, signal processing, mathematical statistics, human intelligence, and engineering. The ultimate goal of speech recognition is to make a technology product that can communicate with people normally. This product can convert the received sound information into corresponding text and execute the corresponding action, thereby realizing human-computer interaction (Human Computer Interface, HCx). At present, the speech recognition technology in a quiet environment is quite satisfactory, and the speech recognition rate is also very high. However, in real life, the speaker's voice is easily affected by the surrounding environment and various superimposed noises and cannot achieve the ideal recognition effect. Boundless integrated financial management is guided by corporate strategy, emphasizing that finance breaks through the existing work framework and mode with a borderless active management awareness, communicates and transmits financial concepts in all aspects of the value chain, and forms the integration of finance and other aspects. The department implements the financial management model to promote the continuous growth of the overall value of the enterprise. Therefore, improving speech recognition technology under noisy conditions and extracting accurate sound information has become the focus of researchers. In a society where the scope of information technology is gradually expanding, the development of research on speech recognition technology in noisy environments is of great significance to improving people's living standards and quality of life.

1.3. Related Work. Human society increasingly shows the characteristics of an information society. Not only between people but also between people and machines, a large amount of information needs to be exchanged at any time. Voice is the most direct, most convenient, and effective tool for information exchange between people, and it is also an important means of communication between people and machines. Use the concept of big data to optimize the industrial structure and improve work efficiency. First, it outlines the main content and development changes of financial management. The latest research progress of digital construction financial budget management and cost management is analyzed. Then combined with actual work, the main content and workflow of the financial management system are summarized and improved. Finally, an intelligent framework for modern hospital financial management is proposed, including budget management, cost management, and

performance management. [1] Dagnew D K uses a parallel hybrid research method. Collect data through questionnaire surveys and interviews. Quantitative data is analyzed by descriptive and econometrics, and interview data is analyzed by narrative. In addition, the lack of guidance and consulting services, the high cost of training and consulting services, the complexity of most standards, and the lack of time for regulatory agencies are the most important challenges. The international comparability, transparency, and quality of the financial reporting system are one of the prospects for the implementation of IFRS. However, his data is not clear enough ([2] Maria Ruiz-Jimenez J). There is an ethical debate about the impact of gender diversity in the top management team (TMT) on the organization. The research aims to contribute to this debate by analyzing the influence of gender diversity of TMT on the relationship between knowledge combination ability and organizational innovation performance. A sample of 205 small and medium enterprises (SMEs) belonging to the Spanish Technology Company (TBF) sector was used. The research results show that gender diversity positively regulates the relationship between knowledge combination ability and innovation performance. They discussed the impact on theory and practice among them, how to promote more equal gender distribution, and the benefits of gender diversity in senior management positions [3]. Although the analysis is in place, some theories do not have practical value. Through their research, it can be found that for financial management, most scholars still improve the efficiency of traditional methods and do not propose essential innovations in combination with current technology.

1.4. Innovation. The innovation of this article is as follows: (1) First of all, the innovation of the topic selection angle. This article is a new perspective from the perspective of topic selection. At present, there is not much research on the integration of speech recognition, boundless fusion, and financial management. It is of exploratory significance. (2) The second is the innovation of research methods. The proposed detection based on time domain energy and zero-crossing rate, detection based on entropy function method, etc., has theoretical value. (3) The other is the innovation of project practice. Boundless integrated financial management penetrates the financial concept into all aspects of production and operation, so that information communication can break the barriers of departments and professions, improve the information transmission, diffusion, and penetration capabilities of the entire organization, and realize the symmetrical distribution of information, experience and skills and sharing, thereby stimulating innovation and improving work efficiency, and realizing the optimal allocation of enterprise resources and the maximization of value creation. The project is of great significance to improving people's living standards and quality of life.

2. Commonly Used Voice Endpoint Detection Algorithms

2.1. Based on Time Domain Energy and Zero-Crossing Rate Detection. Speech endpoint detection is generally used to identify the presence and absence of speech in an audio

signal. Short-term energy analysis and zero-crossing rate analysis are the most basic methods in time domain analysis of speech signals, and they are widely used, especially in the detection of speech signal endpoints. Since these two methods are usually used independently in speech signal endpoint detection, it is easy to miss important information during endpoint detection. Judging from the instability of the speech signal [4], its energy may change over time, and the energy of nonhuman voice and human voice is also very different. Therefore, by analyzing the short-term energy and the short-term energy zero pass rate, it is possible to distinguish the speech segment with and without the speech segment. The short-term energy of speech sampling sequence $g(i)$ is defined as follows:

$$W_i = \sum_{-\infty}^{+\infty} [f(e)h(i-d)^2] = \sum_{-\infty}^{+\infty} f(e)^2 h(i-d) = g(i)^2 * h(i). \quad (1)$$

Among them, W_i represents the short-term energy of the i speech frame, $h(i) = g(i)^2$. Set the windowed speech signal to $f_w(i)$, N represents the window length, and the short-term energy [5] is expressed as follows:

$$E_i = \sum_{e=-\infty}^{\infty} |x(r)|w(i-r) = \sum_{m=e}^{i+A-1} |x_w(r)|. \quad (2)$$

The endpoint of the speech signal is obtained by setting and comprehensively using two threshold levels, and the principle is simple, and the real-time performance and precision are relatively high, so it has a wide range of applications. The zero pass rate of the speech signal should be the number of times the signal crosses the branch in one second. When the speech signal is a continuous signal, the zero pass rate can be calculated as the number of times it passes the segmented signal. If it is an obvious voice signal, the number of times the symbols of two adjacent sampling points change in a unit time can be used to represent the short-term zero-crossing rate [6]. It is defined as follows:

$$\text{WDR}_i = \frac{1}{2} \left[\sum_{k=0}^{A-1} |\text{dsn}(h(i))| - \text{dsn}(h_i(h-1)) \right]. \quad (3)$$

In formula (3), the frame length is L , and $\text{dsn}(x)$ 1 is the sign function.

$$\text{dsn}(x) = \begin{cases} 2, & x \geq 0, \\ -2, & x < 0. \end{cases} \quad (4)$$

In real life applications, in order to prevent low-frequency interference [7], a threshold value T is usually set, and the number of times the threshold value T is crossed is used instead of the number of zero-crossings. Therefore, the improved zero-crossing rate is shown in the following:

$$\text{WDR}_i = \frac{1}{2} \sum_{m=n}^{L-1} |\text{dsn}[x(m) - T]| + |\text{dsn}[x(m) + T]| - \text{dsn}[x(m-2) + T]. \quad (5)$$

2.2. Detection Based on Entropy Function Method. Entropy comes from statistical thermodynamics to physics [8]. It characterizes the degree of disturbance. With the cross-development of science, entropy has important applications in the fields of government, probability theory, information theory, and life sciences. In speech signal processing, the concept of entropy is widely used for endpoint detection [9], such as spectral entropy and cepstral distance entropy.

Assuming that the noisy speech signal $h_i(m)$ in the i th frame after the windowing process, after Fourier transform (FFT), the spectral probability density of the k th spectral line of each frequency component is $h_p(k)$, then the spectral entropy of each frame is defined as follows:

$$X_i = - \sum_{k=0}^{H/3} L_i(G) \log L_i(G), \quad (6)$$

where N is the FFT length, and the probability space of the discrete speech signal X can be expressed as follows:

$$\begin{bmatrix} S \\ P(g) \end{bmatrix} = \begin{bmatrix} x_1, x_2, \dots, x_c \\ j_1, j_2, \dots, j_c \end{bmatrix}. \quad (7)$$

Then, the entropy function of the speech signal X is as follows:

$$F_i = \sum_{i=1}^q P \log P_i. \quad (8)$$

From the definition of spectral entropy, it can be seen that speech entropy reflects the “disorder” of the amplitude distribution of the source in the frequency domain.

The idea of offspring spectral entropy is as follows: first divide a frame of speech signal into several offspring, and then find the spectral entropy of each offspring. Suppose the energy probability of the m th offspring of the i th frame is $P(h, i)$, and the offspring energy is $E(h, i)$; then,

$$P(h, i) = \frac{E(h, i)}{\sum_{k=1}^{g_v} E(h, i)}. \quad (9)$$

The offspring power spectrum is as follows:

$$S_i = \sum_{a=1}^{g_v} P(h, i) \log P(h, i). \quad (10)$$

The difference between the detection method based on the offspring spectrum entropy method and the original spectrum entropy method is that the latter is to find the probability density function of each power point and then the spectrum entropy, while the former is to find the power density function of a child of the speech frame [10], then calculate the spectral entropy, because, in this way, the amplitude of a single spectral line is protected from the interference of background noise.

2.3. Detection Based on Double Threshold Method. The dual-threshold method was originally proposed based on short-term energy and short-term average zero-crossing rate [11].

Its principle is because the energy of the finals is very large, short-term energy can be used to find the finals, and the frequency of initials is also very high and can be used. Find the short-term average zero pass rate of the agreement. The dual-threshold method needs to use two levels of judgment to realize voice endpoint detection. Set Q_F as the high threshold and Q_H as the low threshold.

- (1) First-level judgment: according to the difference in short-term average energy;
- (2) After the judgment of (1), according to the difference in the short-term average zero-crossing rate, the second-level judgment is carried out so that the starting point and ending point of the speech can be determined, and the effect we want to achieve by the endpoint detection can be obtained.

The specific process is as follows:

- (1) Starting from the silent segment, if the short-term energy [12] and the short-term zero-crossing rate of the speech frame are both greater than Q_F , we can think that the frame may be the starting point of the speech, and the system starts to detect it, but it is not sure that it is the starting point. At this stage is defined as a transitional section.
- (2) Continue to track the short-term energy and the short-term zero-crossing rate of the next frame. If both of these parameters are less than Q_F , it is definitely not a voice endpoint, and the silent segment is restored.
- (3) If one of the parameters is greater than the high threshold Q_H , and the duration of such a state exceeds the minimum speech segment length, it can be marked as the starting point of entering the speech segment.
- (4) After entering the speech segment, if the short-term energy or the short-term zero-crossing rate of the detected speech frame is less than Q_F , it enters the end point mark, but it is still a transitional state at this time. If these two characteristic parameters are both higher than Q_F , it is considered that it is still in the speech segment. If the detected two parameter values are less than Q_H and last for a period of time, it is marked as the end point of the speech signal [13]. The overall flow chart of dual-threshold endpoint detection is shown in Figure 1.

The above endpoint detection method is only suitable for speech in a pure environment and is not applicable when noise is added [14] because the zero-crossing rate of noise is much larger than initials and finals, so the above method needs to be improved. In fact, the short-term zero-crossing rate of the silent section in a noisy environment is less than that of the voiced section [15], which is just the opposite of the situation in a pure environment. Therefore, when looking for the starting point, you only need to find the original zero-crossing rate greater than Q_H . If it is changed to less than Q_H , the situation of finding the end point is also

the opposite, and the situation of short-term energy remains unchanged. Using the above-mentioned improved double-threshold method to detect the starting point and ending point of a noisy segment of Chinese speech, the detection results are shown in Figures 2 and 3:

Although the improved dual-threshold method has some antinoise ability [16], the false positive rate is still very high in a noisy environment. This is because the improved method cannot completely screen the noise of the speech, and the noise of different frequencies will lead to a false alarm rate. In the past ten years, in the analysis and research of voice signals, a feature value based on the entropy function has emerged as it is used in voice endpoint detection, which has a certain degree of robustness in complex environments.

3. Construction of a New Model of Financial Management

3.1. The Establishment of "CIO" Model. This paper conducts a standardized analysis by reading the literature and sorting out related concepts and theoretical foundations [17]. Comprehensively analyze the operating conditions and operating models of each enterprise and compare their comprehensive scores. The analysis results show that most of my country's SMEs' existing financial management models have serious problems, mainly: weak capital management capabilities lead to low corporate returns and poor performance. The company's business performance has led to unrealistic and unrealistic accounting information disclosure, and the high concentration of financial power allocation has led to unscientific capital management decision-making. The capital management [18], accounting information disclosure, and financial organization of small and medium-sized enterprises have been severely divided [19]. The current financial management model does not adapt to the new e-commerce environment and cannot survive and grow in the new economic competition environment. It is necessary to discuss how to regulate the accounting information disclosure of listed companies under the current market economy conditions, propose effective strategies to solve the problem, and expect the relevant laws and regulations of accounting information disclosure to be more complete.

Starting from the problem, this article selects three typical submodes according to the standard of financial management module-Capital Management Mode, Accounting Information Disclosure Mode, and Financial Organization Mode (Financial Organization Mode), as the research object and content, as shown in Figure 4.

This model focuses on the study of the relationship between corporate capital management, accounting information disclosure, and financial organization and aims to combine the three organically to restrict, promote and coordinate work with each other so as to achieve "capital management to promote accounting information disclosure, the virtuous circle of "promoting capital management with appropriate decentralization," thereby improving the level of

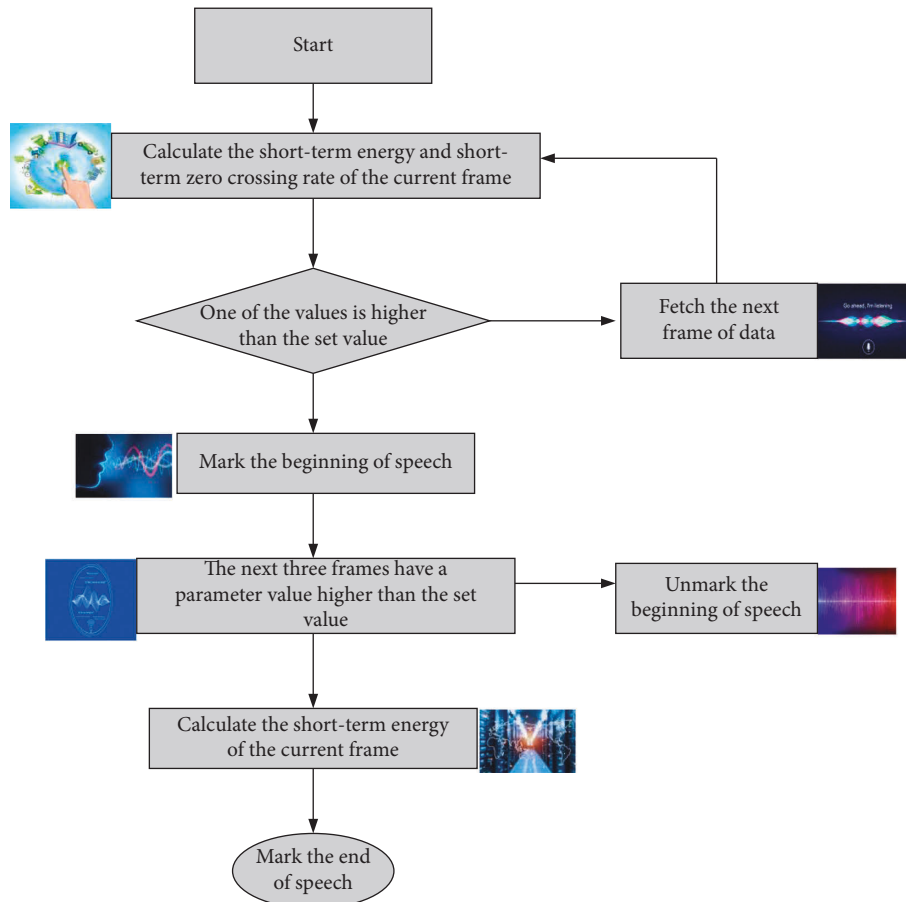


FIGURE 1: Flow chart of the dual-threshold endpoint detection program.

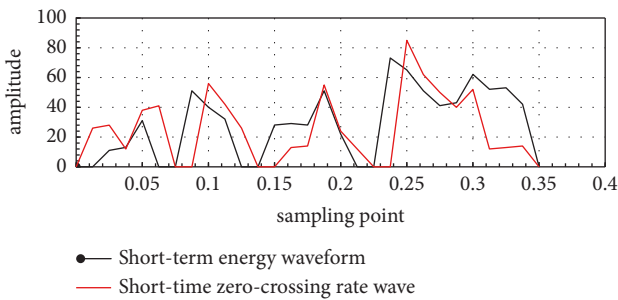


FIGURE 2: Short-term energy of noisy speech signals.

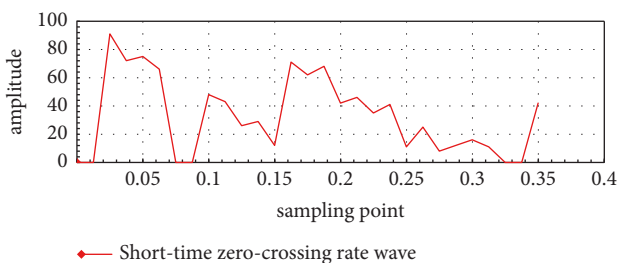


FIGURE 3: Short-term zero-crossing rate of noisy speech signals.

corporate financial management [20]. Accounting information disclosure means that an enterprise provides important accounting information that directly or indirectly affects the decision-making of users to the information users in the form of public reports fairness between objects. The financial model is to classify, organize, and link various information of the enterprise according to the main line of value creation, so as to complete the functions of analyzing, predicting, and evaluating the financial performance of the enterprise. The following will use empirical analysis to study the interaction between these three and propose the “Capital-Information-Organization” integration model, referred to as the “CIO” model. All in all, through the reorganization of the vertical, horizontal, external, and geographical boundaries of financial management, the purpose of financial management has changed from the traditional management model with strict upper and lower levels, a horizontal division of left and right, and obvious internal and external boundaries to a flat organizational structure, top and bottom. Communication is convenient, departments are integrated with each other, and there is no boundary between internal and external management models.

3.2. Chip Selection. The speech recognition system is built on a certain hardware platform. Nowadays, the types of commonly used speech recognition chips are a special IC for speech recognition composed of a single-chip

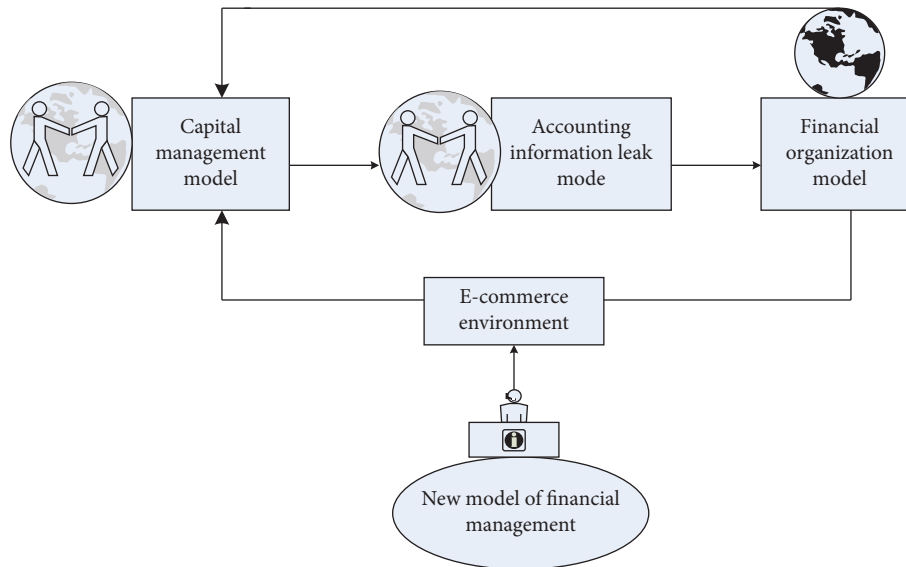


FIGURE 4: Components of the financial management model.

microcomputer (MCU) [21], a speech recognition system composed of a digital signal processor DSP, and a speech recognition system-level chip SOC (System on Circuit). The SPCE061A chip was selected as the core component of this system because it was determined after the following comparisons of these types of chips.

- (1) An application specific integrated circuit composed of multiple band-pass filters and linear adaptive circuits [22]. It is a product in the early 1980s and the first integrated circuit dedicated to speech recognition. It consists of a set of band-pass filters to form a feature extraction circuit and then uses a linear matching circuit for pattern matching. This circuit has poor speech recognition performance.
- (2) The on-chip system integrates circuits such as a single-chip microcomputer or DSP [23], A/D, D/A, RAM, ROM, host computer [24], power amplifier, and so on. As long as peripheral circuits such as power supply are added, functions such as speech recognition, speech synthesis, and speech playback can be realized. This is the most advanced speech recognition chip that has emerged in the past two years, with high performance and low power consumption, but the price is relatively high.
- (3) SPCE061A is a highly integrated single-chip. It integrates single-chip, A/D, D/A, RAM, and ROM on a chip. It also has 16 bit \times 16 bit inner product operation and multiplication functions, and the highest CPU clock can reach 49 MHz. Therefore, it is equivalent to DSP in terms of complex digital signal processing, but the price is lower than dedicated DSP chips and has a strong interrupt handling capability. The system supports 10 interrupt vectors and more than 10 interrupt sources. It is suitable for real-time voice processing. It has a dual-channel 10-bit DAC audio output function and is configured with

dynamic white gain. The microphone input method provides great convenience for voice processing. The system hardware circuit is composed of SPCE061A, SPR module, LED keyboard module, and power module. The hardware structure block diagram of the system is shown in Figure 5.

3.2.1. MIC Input Module. Figure 5 shows the entry circuit diagram. As mentioned above, the D D converter of SPCE0611A has eight channels, one of which is the MIC-IN input, which is used to sample specific audio signals. The noise signal is converted to MIC signal and then enters the amplifier in the gain control circuit of SPCE0611A. The options that can be detected, such as the AGC circuit, when the input signal increases at any time. The purpose of the AGC (Automatic Generation Control) automatic gain control circuit is to realize the amplification gain control of the detection object whose signal amplitude changes greatly. For example, in the audio design, in order to ensure that the speaker outputs the sound of the appropriate volume, of course, the volume is independent of the input (Otherwise, it is not necessary to constantly adjust the volume knob). AGC control needs to be added, and AGC can also be used in the image acquisition design to achieve stable image signal detection under strong and weak light conditions. When the input signal is reduced, the AGC circuit will automatically add Hamiltonian to an optimal level, which will help reduce corrective actions. Therefore, the input circuit of the resistance and capacitance system of the DC filter is very simple as shown in Figure 6.

3.2.2. Communication Module. (See Figure 7) shows the communication module circuit. The data in the single-chip microcomputer in the picture is converted to RS-232 level through the serial port through MAX232 level and

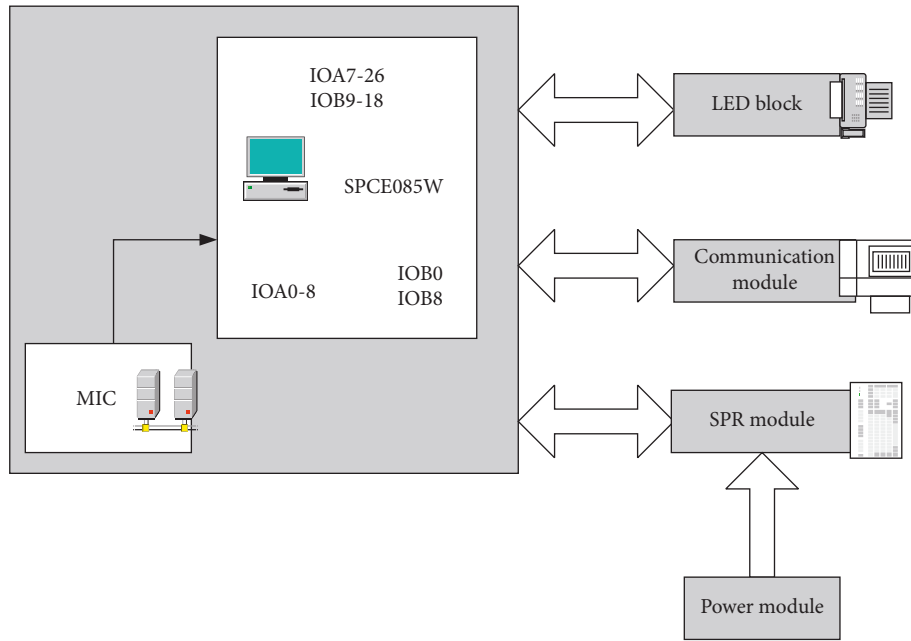


FIGURE 5: System hardware circuit.

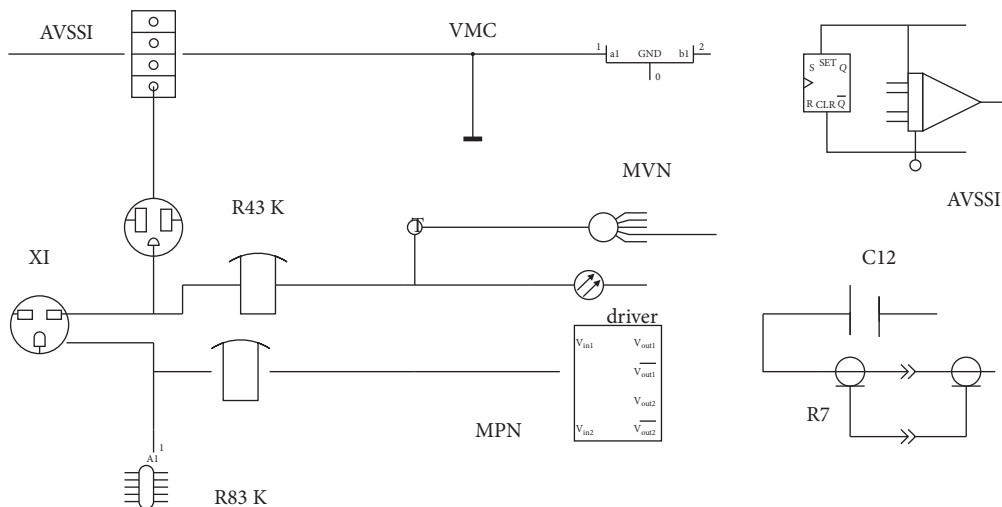


FIGURE 6: Audio input module.

transmitted to the upper level. The receiving pin RXD and the sending pin TXD of Sunplus SPCE061A MCU serial communication are shared with IOB7 (pin 79) and IOB 10 (pin 76), respectively. The serial interface of SPCE061A is a full-duplex interface with a receiving buffer. The processing of the communication module is optimized with the built-in chip of the system, which can reduce the delay of communication.

The most widely used standard in asynchronous serial communication is the standard RS-232C serial interface used in the circuit. Since SPCE061A serial ports are all TTL levels, they are not compatible with RS-232C levels, so level conversion must be performed on both interfaces. Using the MAX232 chip to connect the external power supply SV and external capacitors, a 10 V power supply

can be generated to form a 232C transponder. The communication circuit in this system uploads a large amount of voice data to the computer for processing, which is completed by the computer. For example, calculating noise energy and zero flow, designing digital filters, training model libraries, etc.

3.3. System Implementation

3.3.1. *Software and Hardware Environment.* The DELL computer used in this system, Core i5 processor, 4G memory, 2.4GHZ frequency, and Philips headset for recording. The computer system is Win7, and the development platform is Matlab.

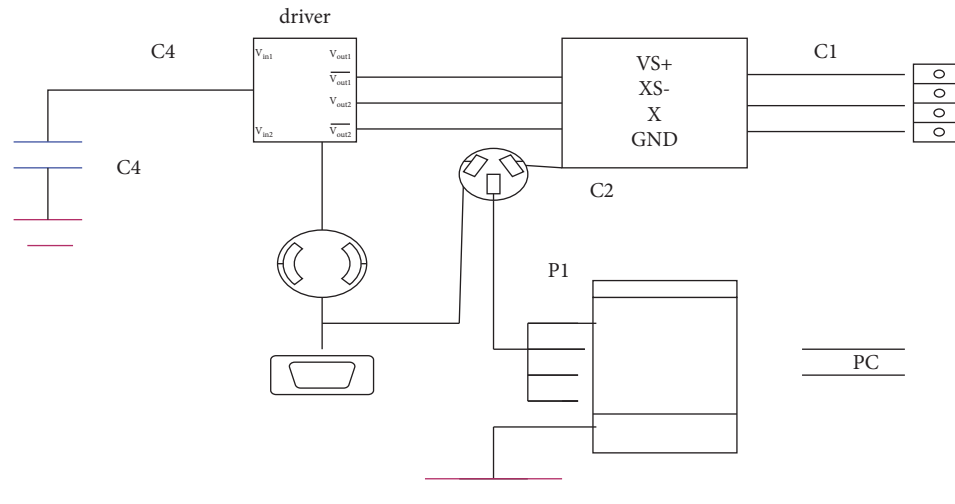


FIGURE 7: Serial port communication circuit.

3.3.2. The Establishment of the Corpus. Corpus refers to a large-scale electronic text library that has been scientifically sampled and processed, which stores the language materials that have actually appeared in the actual use of the language. There are two types of corpora, one is keyword corpus, and the other is experimental test corpus.

The keyword corpus is a total of 64 commonly used keywords in the three languages of Chinese, English, and Japanese recorded by 6 people as standard as possible. Due to gender differences, each keyword is recorded using 1 male and 1 female; for different regions, pronunciation and dialects are temporarily not included in the corpus due to a large amount of data. In the future, the collected key sounds of various common dialects of speech can be cut and added to the keyword corpus in practical applications.

The experimental test corpus is divided into telephone speech, actual scene recording, and reading-aloud recording. Limited to the environment, the telephone speech and actual scene recording only include Mandarin Chinese, and the reading-aloud speech includes three languages: Chinese, English, and Japanese. The phone voice uses 50 minutes of recording files of ordinary calls; the actual scene recording is the sound recorded by placing a voice recorder in the school laboratory, home, and office for a long time, and each of them intercepts the more concentrated voice for 1 hour. The reading speech includes news speech in 3 languages for 20 minutes each, and the reading material recorded by 3 people is 20 minutes each. Convert all voice data into 8K sampling rate and WAV format, and store them in two folders, respectively. Name the keywords with the content of the keywords, and name the test voices with the recording method and the label.

3.3.3. Acoustic Model. Usually, in speech recognition, the phonemes and syllables of a certain language are commonly used as primitives for acoustic modeling. This will be a big problem for training data because sufficient training data is required to obtain adequate training. Robust model. This article is aimed at different languages and it is difficult to

fully train, so each keyword as a whole is used as a primitive. The problem with words as primitive is that the recognition process is limited by the content of the vocabulary, and words that are not in the vocabulary cannot be recognized and will reduce the recognition rate. However, the situation described in this article is that the content of the keywords that need to be recognized is known before the recognition, and the keywords to be recognized have been added to the keyword corpus, so using keywords as primitives best meets the needs. After determining the primitives, extract the average short-term amplitude, peak size, and peak spacing as time-domain feature parameters. 12-dimensional MFCC coefficients, 12-dimensional MFCC coefficients first-order difference, and 12-dimensional MFCC coefficients second-order difference are used as the frequency corresponding to the speech signal. Domain characteristic parameter: After further analysis of 3 groups of 12-dimensional parameters in the frequency domain, 27-dimensional frequency domain characteristic parameters are obtained after weighted reconstruction. Without repeated training, the time domain and frequency domain parameters of a keyword are directly used as its time domain and frequency domain acoustic models, that is, the matching template in the recognition process. After the one-step recognition is completed, it is generally necessary to confirm the keywords through the language model. When choosing a better language model, you can check whether the identified keywords meet the requirements from the perspective of language habits and prosodic features, which can reduce the false alarm rate of the system. However, since this article does not distinguish the language and recognizes the keywords as a whole, it does not establish a language model. Instead, it uses the time domain and frequency-domain joint search method to confirm each other, which can also better reduce the false alarm rate.

3.3.4. Performance Test. For the existing corpus, a 20-minute reading material recorded in the Chinese reading-aloud speech was selected for the recognition test. There are

TABLE 1: Voice system performance test results.

Key words	Number of false alarms	Number of recognitions	False alarm rate	Missed recognition rate (%)	Recognition rate (%)
Management	3	16	13%	12.5	97.5
Academy	6	36	12.7%	19.6	76.5
The university	5	25	8.6%	11.4	79.6
Education	4	14	43.3% 100%	8	100

4 keywords: management, college, university, and education, which appear 16, 22, 39, and 12 times respectively. The sampling frequency is 8000 Hz, and the frame length is 128. The recognition results obtained are shown in Table 1.

The system also separately shows the time when each keyword appears, and saves the results, so that you can easily query the results.

4. Analysis of Experimental Results

4.1. Selection of the Number of Transformation Matrix Clusters. When generating the transformation matrix of the feature space, how to share training data and choose the number of clusters is a problem that requires comprehensive consideration. The degree of detail in the division of acoustic modeling units and the amount of training data will directly affect the choice of the number of clusters. In the Aurora2 experiment, since there are only 180 emission states and the noisy training data has only 1760 sentences, it is not appropriate to use too many classifications. When modeling the parameter trajectory of the feature space transformation matrix separately, we tried to use 24816 matrix classifications respectively, and the result of the recognition system is shown in Figure 8.

When the GVP-HMM method is used to cascade the transformation matrix and model parameters for modeling, the selection of the number of transformation matrix clusters needs to be reconsidered. At the training end, the transformation matrix parameter trajectory is generated on the polynomial trajectory of the Gaussian model parameters. At this time, the polynomial coefficient model is obtained through various trainings and already contains the information of the noisy speech at the training end; at the decoding end, the transformation matrix also needs to be loaded on the acoustic model. The acoustic model at this time is more consistent with the test background noise. At this time, if too many clusters are used to refine the impact of noise on each modeling unit, it will have a training effect. As can be seen from the figure, in the GVP-HMM cascade system containing variable Gaussian model and feature space transformation matrix, the one that obtains the lowest word error rate is the two-category "GVP-mv-fran" system (WER9.79%), where the pronunciation words and silence sil and sp are divided into two classification matrices. The same result also appeared in the "GVP-mv-tran" system containing a variable Gaussian model and mean space transformation matrix. The lowest is 0 and the highest is 16.

4.2. Recognition Results of Various Parametric Trajectory Modeling Systems. In the experiment, we used every 1001 sentences of 0 dB, 5 dB, 10 dB, and 20 dB under 4 signal-to-

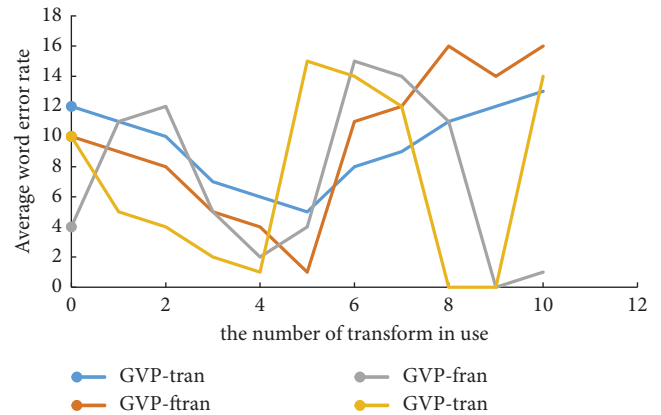


FIGURE 8: Selection and comparison of the number of transformation matrix clusters.

noise ratio conditions to decode the initial pure model, the mcond reference model, and the variable parameter system under the GVP-HMM framework.

- (1) When using the GVP-HMM method to model the parameter trajectory of the Gaussian model, we chose the mean value and the mean value simultaneous parameter trajectory modeling system as a reference. In the experiment, it is simplified and marked as "mean" and "mv."
- (2) When modeling the trajectory of the mean transformation matrix in the model space, we selected 2 and 8 categories of mean transformation matrices "tran2" and "tran8", respectively. At the same time, the cascade system "mv-tran2" with variable Gaussian model parameters is used as the reference system.
- (3) When modeling the trajectory of the feature space transformation matrix, we also use "ftran2" and "ftran8" to represent the transformation matrix of the corresponding number of categories and use "mv-ftran" to represent the cascade system. The variable parameter modeling types represented by the various systems of GVP-HMM are shown in Table 2.

At the decoding end, the signal-to-noise ratio information corresponding to the test environment is loaded onto the polynomial coefficient trajectory, and corresponding parameters are generated for decoding. The recognition results of the GVP-HMM system based on the above-mentioned traditional methods are given in Table 3.

TABLE 2: Various types of variable parameter trajectory modeling under the GVP-HMM framework.

GVP Syetem	Parameter polynomials				Poly Coef
	Mean	van	Tran	Fran	
Mean	√	×	√	×	3.66 m
Mb	√	×	√	×	7.88 m
Tran3	×	√	×	√	12.5 m
Tran9	√	√	×	×	15.6 m
Fran3	×	√	√	×	11 m
Fran8	×	×	√	√	18.23 m

TABLE 3: Recognition results of variable model spatial parameters under traditional GVP-HMM method Aurora 2 experiment.

WER (%)	0 dm	7 dm	15 dm	11 dm	Ave	25 dm
Clean baseline	88.26	24.63	8.5	8.6	14.5	25.63
Mecond baseline	23.56	5.26	7.48	8.69	9.54	4.25
Mean	23.56	9.87	9.56	9.47	9.63	9.87
mv	24.36	8.56	7.45	8.36	4.23	1.23
Tran2	22.86	8.56	7.56	9.36	4.56	15.23
Tran4	11.23	15.23	5.36	5.45	8.69	12.25
Nv-tran12	12.35	11.25	25.65	4.59	75.6	88.56

TABLE 4: Recognition results of variable parameter feature transformation matrix system Aurora 2 experiment.

WER (%)	0 dm	7 dm	15 dm	11 dm	12 dm	Ave	Fall
Mecond baseline	22.25	33.56	21.56	23.25	15.23	11.23	10.23
Mv	12.56	15.36	23.63	24.56	26.98	25.36	25.36
ftran2	5.23	6.52	9.36	8.24	2.56	6.52	11.23
Ftran8	22.56	33.36	45.56	11.25	5.6	2.3	11.25

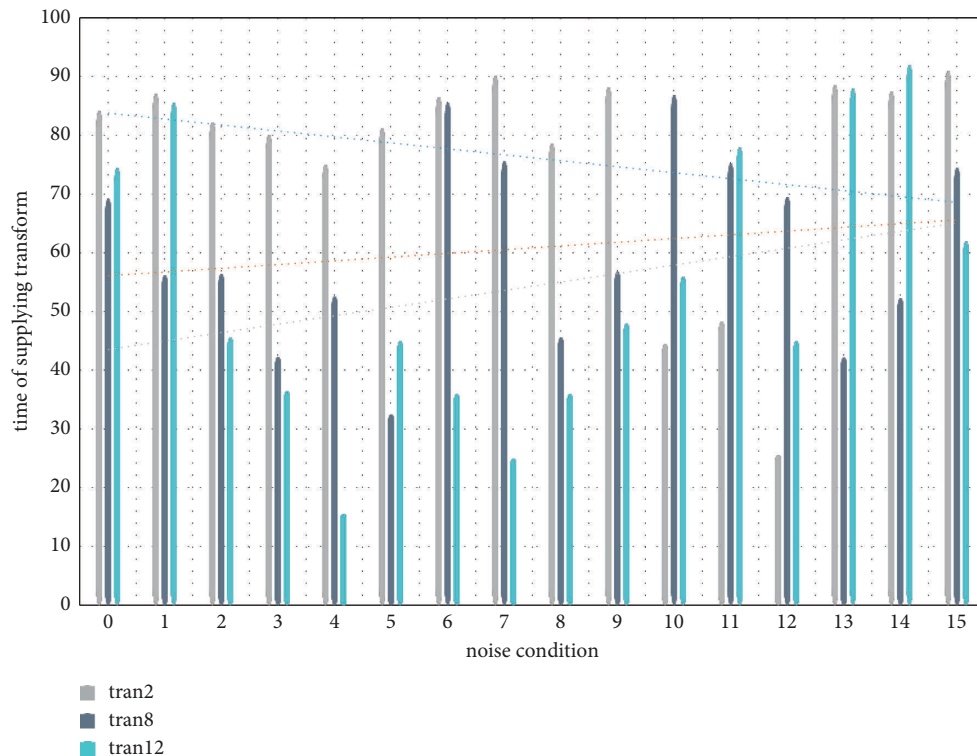


FIGURE 9: Comparison of loading speed of two types of transformation matrices.

In the third chapter of the main content of this article, the application of the GVP-HMM method extended to feature space antinoise recognition is emphatically discussed. Therefore, in this experiment, the antinoise ability of the variable parameter feature transformation matrix, the variable parameter feature transformation matrix cascade Gaussian model parameter trajectory system, and the variable parameter feature transformation matrix cascade mean transformation matrix are used to verify GVP. The antinoise ability of the HMM system in the characteristic domain. The characteristic equation of the matrix is that the matrix can actually be regarded as a transformation. The left side of the equation is to change the vector x to another position; the right side is to stretch the vector x , and the stretching amount is λ . Then its meaning is obvious, expressing a characteristic of matrix A that this matrix can lengthen (or shorten) the vector x by λ times.

4.2.1. Independent Modeling of Variable Parameter Feature Transformation Matrix. On the pure model, we also use a variety of training methods, using four different signal-to-noise ratio training data to generate a transformation matrix trajectory in the form of polynomial coefficients in the feature domain instead of a static transformation matrix. At the decoding end, read in the signal-to-noise ratio information of the test environment to generate the corresponding feature transformation matrix. Compared with the reference system, the recognition results of the variable feature space transformation matrix based on the GVP-HMM method are shown in Table 4.

A series of experiments based on the Aurora 2 data set effectively supports the method of GVP-HMM extended to the feature space in this paper. It is verified through experiments that whether it is using the variable feature space transformation matrix alone, cascading with the variable Gaussian model parameter system, and cascading with the variable model domain transformation matrix, it shows good antinoise performance. At the same time, relying on its unique compact storage space and strong portability, it can become an effective means to solve robust speech recognition in noisy environments. The comparison is shown in Figure 9.

5. Conclusions

In today's era of big data, all departments of an enterprise must make adjustments or even reforms according to the changes in the new environment. Whether an enterprise chooses strategic, lean, or information-based finance, it all reflects multidepartment and multidimensional financial management without exception. The characteristics of the field and multidisciplinary integration. Speech recognition is an important artificial intelligence technology that has played an important role in computer applications, office automation, computer networks, and many other aspects, such as communication, robots, and intelligent human-machine interfaces. Speech recognition technology has advanced by leaps and bounds, and effective recognition

technologies such as standard matching and HMM have been formed. Some successful speech recognition systems have emerged. For now, speech recognition technology has gradually moved from the laboratory to commercialization, and for this reason, there have been many large-scale speech recognition products. In daily work and people's lives, what people need most is cheap voice recognition equipment installed on specific products. In the past two years, this small, integrated voice detector has become an important direction for technological development. After a lot of research, this article uses a combination of regulatory analysis and empirical analysis to analyze the current situation of financial management of my country's small and medium-sized enterprises; my country's business from three aspects: financial organization and management, accounting, and capital information disclosure under the new situation of e-commerce, finance. The problems and deficiencies of the management model have established a new model of e-commerce-based financial management for small and medium-sized enterprises-CIO environment interface model and conducted a basic analysis of structural equations and structural equations.

Data Availability

No data were used to support this study.

Conflicts of Interest

There are no potential conflicts of interest in this study.

Acknowledgments

This work was supported by the Social Science Project of Beijing Education Committee (project no. SM202010017002) and National College Student Innovation Entrepreneurship Training Program (project no. 2019J00015).

References

- [1] Y. . Gong, "Construction of financial management and decision system based on big data," *PPTA: Quarterly Journal of Indian Pulp and Paper Technical*, vol. 30, no. 8, pp. 135–143, 2018.
- [2] D. K. . Dagneu, "Implementation of international financial reporting standards: its practices, challenges and prospects: the case of commercial banks in Ethiopia," *Asia-Pacific Journal of Management Research and Innovation*, vol. 16, no. 4, pp. 259–271, 2020.
- [3] J. M. Ruiz-Jiménez, M. d M. Fuentes-Fuentes, and M. Ruiz-Arroyo, "Knowledge combination capability and innovation: the effects of gender diversity on top management teams in technology-based firms," *Journal of Business Ethics*, vol. 135, no. 3, pp. 503–515, 2016.
- [4] Z. Wang, M. Ying, Q. Wu, R. Wang, and Y Li, *Gene*, vol. 593, no. 1, pp. 100–109, 2016.
- [5] P. E. N. G. Lan, "Teachers' teaching innovation abilities in financial management courses," *Journal of Modern Accounting & Auditing*, vol. 5, no. 1, pp. 20–26, 2017.

- [6] W. . Liang, "Research on innovation of financial management mode in high-tech marine engineering enterprises," *Journal of Coastal Research*, vol. 115, no. sp1, p. 99, 2020.
- [7] D. Xiaocun, S. Xuan, and Z. Huan, "Innovation of internal control system of financial management in public institutions," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 17, pp. 263–268, 2017.
- [8] R. C. Mayer, R. S. Warr, and J. Zhao, "Do pro-diversity policies improve corporate innovation," *Financial Management*, vol. 47, no. 3, pp. 617–650, 2018.
- [9] M. Dziura and T. Rojek, "Management of the company's innovation development: the case for polish enterprises," *Journal of Risk and Financial Management*, vol. 14, no. 4, p. 156, 2021.
- [10] Y. Rahmawati, F. G. Dewi, and Y. Yuliansyah, "REGIONAL: regional management information system and training for regional financial accountability," *International Journal for Innovation Education and Research*, vol. 8, no. 11, pp. 213–223, 2020.
- [11] M. S. Peiris, N. J. Dewasiri, and Y. Banda, "Book review: I," Edited by M. Pandey, Ed., vol. 16, no. 2, pp. 167-168, Asia-Pacific Journal of Management Research and Innovation, 2020.
- [12] S. Sheikh, "The impact of market competition on the relation between CEO power and firm innovation," *Journal of Multinational Financial Management*, vol. 44, no. MAR, pp. 36–50, 2018.
- [13] S. Harith and R. H. Samujh, "Small family businesses: innovation, Risk and value," *Journal of Risk and Financial Management*, vol. 13, no. 10, p. 240, 2020.
- [14] A. Irjevskis, "Exploring the link of real options theory with dynamic capabilities framework in open innovation-type merger and acquisition deals[J]," *Journal of Risk and Financial Management*, vol. 14, no. 4, p. 168, 2021.
- [15] V. P. Shestak, E. I. Moreva, and I. G. Tyutyunnik, "Financial management of innovative activity," *Financial Theory and Practice*, vol. 23, no. 6, pp. 63–75, 2019.
- [16] T. Overmans, "Managing slack: how city managers create slack for innovation in times of fiscal stress," *Journal of Public Budgeting*, vol. 30, no. 3, p. 00, 2018.
- [17] Mizrahi, K. Sarit, A whole new meaning to having our head in the clouds: voice recognition technology, the transmission of our oral communications to the cloud and the ability of Canadian law to protect us from the dangers it presents," *Canadian Journal of Law and Technology*, vol. 15, no. 1, p. 10, 2017.
- [18] D. Poap and M. Wozniak, "Voice recognition by neuro-heuristic method," *Tsinghua Science and Technology*, vol. 24, no. 01, pp. 11–19, 2019.
- [19] J. Jiang and H. H. Wang, "Application intelligent search and recommendation system based on speech recognition technology," *International Journal of Speech Technology*, vol. 24, no. 1, pp. 23–30, 2021.
- [20] D. Cao and Y. Guo, "Algorithm research of spoken English assessment based on fuzzy measure and speech recognition technology," *International Journal of Biometrics*, vol. 12, no. 1, p. 120, 2020.
- [21] D. Purbohadi, S. Afriani, N. Rachmanio, and A Dewi, "Developing medical virtual teaching assistant based on speech recognition technology," *International Journal of Online and Biomedical Engineering (ijOE)*, vol. 17, no. 04, p. 107, 2021.
- [22] M. Liu, X. Chen, Y. Mo et al., "Improving English pronunciation via automatic speech recognition technology," *International Journal of Innovation and Learning*, vol. 25, no. 2, p. 126, 2019.
- [23] L. Mao and J. Miao, "Comprehensive evaluation and analysis on spoken English based on computer speech recognition technology," vol. 42, no. 2, pp. 778–782, 2017.
- [24] Z. Yun, "Research on spoken English speech recognition technology in computer network environment," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 16, pp. 445–449, 2017.