

Research Article

Adaptive Pronunciation Proofreading of Spoken English in a Wireless Sensor Network Environment

Shufang Yang ¹ and Junying Lv ²

¹Department of Foreign Languages, Xinyang University, Xinyang, Henan 464000, China

²Further Education College, Xinyang Normal University, Xinyang, Henan 464000, China

Correspondence should be addressed to Junying Lv; ljy@xynu.edu.cn

Received 3 September 2021; Revised 15 September 2021; Accepted 17 September 2021; Published 22 October 2021

Academic Editor: Guolong Shi

Copyright © 2021 Shufang Yang and Junying Lv. 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.

This paper presents an in-depth study and analysis of adaptive proofreading of spoken English pronunciation in a wireless sensor network environment. This paper addresses the above problem by combining two common methods for controlling the transmission rate of sensing nodes maximizing network utility algorithm and congestion control mechanism. Firstly, the transmission rate of one-hop nodes at a distance from the aggregation node is dynamically adjusted by the increasing exploration algorithm under the premise of unknown link transmission capacity, while the transmission rate of one-hop nodes is proportionally allocated to multihop nodes in multihop nodes by the congestion control mechanism based on the average reception success rate of the link. A design framework for a speech recognition system with complementary offline recognition and online recognition based on the C/S model is proposed, and a speech recognition system in swarm intelligence awareness is implemented based on the Sphinx engine. The client side implements the speech recognition of decoder in the offline state, and the server side provides the functions of recognition consistency detection, model adjustment training, monitoring, and recommendation in the online state as well as the interface for external access. The scene adaptation module effectively improves the speech recognition system's speech recognition correct rate under different scenes, and the discourse topic recognition module verifies the recognition effectiveness of the speech recognition system under different discourse topics, which can meet the requirements of users' personalized speech input.

1. Introduction

Wireless sensor network technology quantifies the information of the physical world into digital information through the rich variety of sensors carried by itself and connects it with the digital world through wireless communication devices, which is one of the important basic technologies of industrial informatization. Information collection is the basic task of wireless sensor networks, in terms of the current development trend, although the current sensor devices continue to enrich, able to obtain more useful information from the physical world, but to meet the large-scale use of scenarios sensor nodes mostly use small low-cost devices, which leads to the sensor nodes in the function of the general computing power, storage capacity, and communication capacity is limited [1]. The initial needs of informatization,

still can not meet the collection of information in the physical world; heterogeneous systems can collaborative interaction, to achieve a deep integration of people, things, and service networks. Therefore, this paper argues that it is necessary to conduct further research on the information collection mechanism of wireless sensor networks and designs efficient information collection mechanisms to meet the performance requirements of different application scenarios under the current limited sensing device capability, to lay a solid foundation for breaking the bottleneck of IoT to seize the high ground of the fourth industrial revolution [2]. Information transmission and information exchange play an indispensable role in the production life of human beings. Connecting it with the digital world through wireless communication devices is one of the important basic technologies of industrial informatization. Information collection is

the basic task of wireless sensor networks. In the process of sending and receiving information, as the information itself carries many physical properties such as sound and light, and does not have the ability to diffuse outward and information processing, it can only be used by people through a certain way to convert information into knowable signals [3]. This way of acquiring information is called data acquisition, and the process of data transmission is data transfer or data forwarding [4]. In recent years, there has been an explosive growth in the rate of information updates and the ways of dissemination are becoming increasingly diverse. The need for people to have end products that can provide information anytime and anywhere in their lives is also becoming increasingly urgent, and this urgent need has prompted the wave of development of wireless communication technology to come early. The scope of short-range wireless communication technology is quite broad, usually refers to the communication radius within a few dozen meters to a hundred meters; if the communication parties take the radio wave communication mode to receive and send information, this can be called short-range wireless communication.

Small-volume multifunctional embedded devices have begun to dominate the market and with increasingly powerful computing and perception capabilities are bound to lead group intelligence perception computing on the core stage of mobile computing, and the application of voice recognition technology in group intelligence perception on embedded has become the most attractive function at present. Human-computer communication will be rapidly applied and popularized under the background of is Internet of Things, which will provide support for the electronicization of various fields such as high-grade intelligent toys voice conversation, car navigation, online meeting, business management, medicine and health, education, and training [5]. At the same time, many the masses of mobile phones or tablets and other intelligent terminal devices consciously or unconsciously collect various aspects of sensor data as well as voice data as the input data set for group intelligence perception applications, which can do analysis and prediction of complex scenes or behaviors, with far-reaching scientific research significance and application value. It can only meet the initial needs of industrial informatization, but still cannot meet the collaborative interaction of heterogeneous systems when collecting information in the physical world, and realize the deep integration of people, things, and service networks. The era of big data has arrived, and big data technology embodies and predicts the awareness and behavior of the crowd with data. The real meaning of big data is to use machines to analyze future trends quickly, efficiently, and intelligently. It is to reduce the analysis time that would take days or even weeks to an instant based on huge amount of data to realize its application value. The technology of swarm intelligence perception requires the support of big data, the source of which is participatory perception. Group intelligence perception has been proposed in foreign countries, and there has gradually started forward-looking and technical research on group intelligence-aware speech recognition systems in China, such as unconscious group intelligence perception [6]. The significance of unconscious

swarm intelligence perception is that the popular smart mobile devices constitute a powerful Internet of Things, which will be the source of big data sets and use cloud technology to accomplish the task of collaborative perception.

Before the rise of deep learning, hybrid Gaussian and hidden Markov models have been widely used as very effective acoustic models, but traditional speech recognition systems consist of several modules together, which are cumbersome to operate and at the same time do not facilitate the unified optimization of the whole model. Only in a certain way can information be transformed into knowable signals to be used by people. This way of obtaining information is called data collection, and the process of data transmission is data transfer or data forwarding. In today's big data eras, these traditional speech technologies are no longer sufficient to support the development of more efficient speech recognition systems, and with the development of deep learning, end-to-end models based on deep neural networks are gradually becoming a new research trend. To address the difficult problems that language models cannot be integrated into the training process of acoustic models and cannot effectively integrate language models for joint optimization, this paper proposes a new end-to-end speech recognition algorithm that incorporates language models, which uses the output of CTC as input for training language models after matrix transformation, so that language models can participate in the training and testing stages of acoustic models, and to a certain extent corrects the CTC model's wrong output, and from the experimental results the method does reduce the word error rate and improve the recognition performance. To simplify the subsequent processing work as well as to facilitate the use by others, the proposed end-to-end language recognition algorithm with fused language models is implemented in a process, and a website based on the Django framework is built; the website can complete the offline and online recognition of speech documents and test the practicality of the improved algorithm in this paper.

2. Current Status of Research

The application of swarm intelligence perception technology to speech recognition systems, where the complexity of the environment in which the speech signal is captured, is greatly increased, and necessarily requires preprocessing of the captured speech signal, such as denoising and compression. Compared with the traditional speech recognition system, the speech recognition system with swarm intelligence perception needs better robustness and compression ratio [7]. The complexity of the environment causes the unpredictability of background noise, and the superposition of various noises and low signal-to-noise ratio brought by different environments test the robustness and adaptability of the speech recognition system model. Also, the speech recognition system in group intelligence perception needs to improve and adapt to the diversity of personalized input. The smaller storage capacity and valuable traffic of embedded devices to cope with the large data volume brought by swarm intelligence awareness require speech recognition systems to adopt processing strategies with smaller

computation and storage volumes [8]. The essence of speech recognition technology is to convert human speech content into computer-readable input information; unlike speaker recognition whose main purpose is to identify the person who emits the speech, speech recognition focuses more on the semantic understanding and computer reproduction of the speech signal itself [9]. The purpose of speech recognition includes two kinds: one is to let the computer “understand” human speech without converting it into written text, but to respond according to the instructions or requests contained in the speech and the other is to generate written text from the valid content of speech information word by word and sentence by sentence according to the grammar rules to achieve the correct understanding of vocabulary.

Infrared data communication is mainly used in point-to-point control signal transmission; most of the current remote-control design are using infrared transmission, which has its own low cost, small size, low power consumption, and easy to connect and other characteristics make its application in the field of low-cost communication very wide. But at the same time, infrared data communication requirements in the angle between the signal transmitter and receiver is not only more demanding at the same time but also requires the existence of obstacles between the two ends; these reasons greatly constrained the development of infrared wireless communication [10]. In a wide variety of wireless communication methods, each communication method has its own advantages and has its own field of application; we can not expect a communication method can solve all the wireless problems. In practical applications, wireless communication is still a complex field, in the design of which issues such as power consumption, cost, effective distance, operability, and effectiveness in the face of interference need to be considered, and in the case of high requirements for data confidentiality, the security of the communication also needs to be considered [12]. The mutual constraints among these factors lead to the coexistence of multiple communication methods, so it is very important to choose a reasonable wireless communication method according to the actual situation [13]. Gul et al. proposed NLP analysis of the test documents of signal vendors to improve the testing process [14]. Fu et al. applied speech recognition to the train ticket inquiry system, which provided convenience for special groups and ticket purchase in special environments [15]. But the recognition effect for new Chen et al. designed a railway speech-guided ticketing system, which realized the design of mobile terminal, but the system collected fewer samples and did not reflect the real application environment, and the interference of noise and dialect recognition also needed to be solved [16]. Latif et al. designed a train on-board speech recognition system for the “train-machine joint control” of the railway engine department [17]. However, the former two tend to be more open domain recognition, while the latter does not really realize the intermodulation with LCK, and its accuracy rate needs to be further tested.

This paper studies postprocessing techniques for car service speech text, unlike language recognition preprocessing which focuses on the extraction of speech signals; postpro-

cessing mainly completes the conversion of speech to text, i.e., recognition into computer-understandable information. Postprocessing techniques involve text processing, natural language understanding, and artificial intelligence and differ in different fields; the same point is that postprocessing all use linguistics to correct the results of preprocessing. Through preprocessing, the system obtains a string of pinyin; firstly, the string is sliced according to natural language processing techniques, and the text information transformed into pinyin is sliced into words, and then, the sliced results are checked and corrected using text proofreading methods; commonly used methods include rule-based methods, statistical methods, machine learning-based methods, and a mixture of these methods. The common methods include rule-based methods, statistical methods, machine-learning-based methods, and a mixture of these methods.

3. Adaptive Proofreading Analysis of English Spoken Pronunciation in a Wireless Sensor Network Environment

3.1. Design of Wireless Sensor Network Environment Construction. In the wireless sensor network as a distributed sensor network, the end is mainly probing and sensing external sensor devices; wireless sensor networks can achieve three functions of data collection, processing and transmission; and communication technology and computer technology have to become the three pillars of information technology. The energy problem always serves as the primary factor in the design of wireless sensor networks, and the energy consumption model determines to a large extent the merit of the design of the clustering algorithm for wireless networks. The analysis time that originally took days or even weeks was shortened to an instant to realize its application value, the group intelligence perception technology needs the support of big data, and the data source is participatory perception. The first-order RF energy model is a more comprehensive way to reflect the energy consumption model of WSN nodes, as shown in Figure 1.

In the above figure sensor node, as the basic unit of wireless sensor network, mainly contains a sensing unit, processing module, wireless communication module, and power supply module. Among them, the sensing module is mainly responsible for collecting the data sensed by the sensors and transmitting the data. And the aggregation node is a special class of nodes, as the mainstay of the wireless sensor network, which has more powerful storage space and longer communication radius than the common sensor nodes. Moreover, to complete the data transmission process, there are two protocols within the aggregation node, which are both a sending protocol that can issue monitoring tasks to the sensor nodes in the sensing area and a receiving protocol that can transmit the data collected by the wireless sensor nodes to the network or to the satellite. While in the communication process, there is a rapid depletion of the remaining energy due to the limited energy carried by the aggregation nodes and the high transmission burden [18]. It can realize the three functions of data collection,

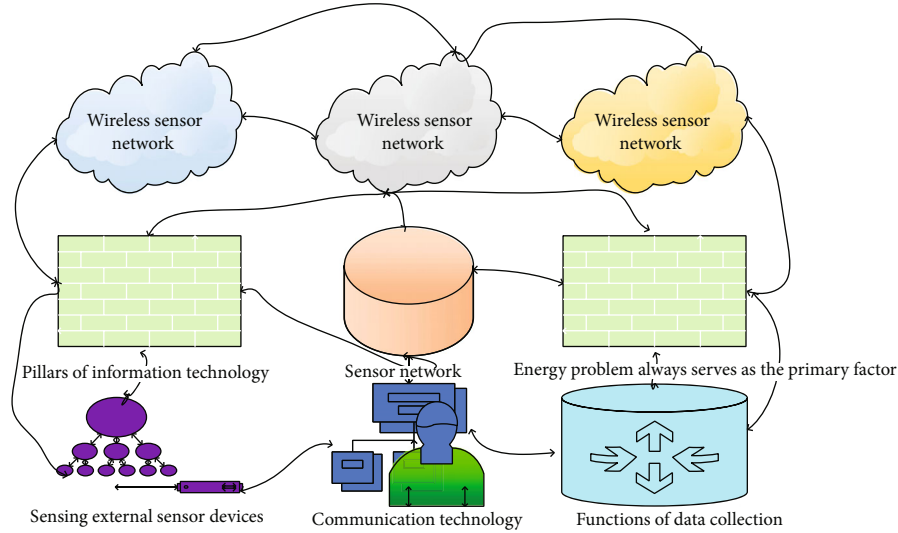


FIGURE 1: Wireless sensor network.

processing and transmission, and communication technology and computer technology have become the three pillars of information technology. The energy problem has always been the primary factor in the design of wireless sensor networks. The energy consumption model determines the wireless sensor network to a large extent. The pros and cons of network clustering algorithm design are found. Once many sensor nodes are randomly dropped around the area to be detected, the nodes are free to search and connect to the sensor nodes around them and the data probed by each sensor node is transmitted along the other sensor nodes. When passing through other nodes, the transmitted sensed data is disposed by other sensor nodes, and after multihop transmission, the information is reunited at the rendezvous node and then transmitted to the processing node via satellite or internet; this multihop routing information transmission is also the main way of network energy consumption. In addition, in the communication process, because in WSNs in the sensor nodes after the placement of the node, there was almost no possibility of replacement, and thus, enhancing WSNs in the limited energy in the network transmission efficiency of the subject has been the focus of scientific and technical workers.

For the path problem in the transmission process, wireless sensor network routing protocols can be divided into single-path and multipath routing protocols, and based on the timing of route establishment, the relationship between data sending can be divided into active routing protocols and passive routing protocols. Based on whether geographic identifiers are reflected in the routing information, whether geographic environment factors are considered in the routing calculation can be divided into environment-based routing protocols and nongeographic factor-determining routing protocols. It mainly includes a sensing unit, a processing module, a wireless communication module, and a power supply module. The sensor module is mainly responsible for collecting the data sensed by the sensor and transmitting the data. According to the above different division, no mat-

ter what factors affect the routing protocol, the final reflection is the survival period of the node and the information throughput, and whether it is the survival period or the information throughput will eventually be reflected in the node energy, so this paper is first based on the difference of the initial energy carried by the node, the wireless sensor network into homogeneous and heterogeneous two aspects, and then according to the difference in network topology the communication strategy of nodes into two modes, planar routing protocol, and layer routing protocol.

Since sensor nodes are mostly battery powered and the nodes have limited energy, how to efficiently use the limited energy to extend the life cycle of sensor nodes has been the focus of research in wireless sensor networks. The energy consumption of sensor nodes is mainly composed of three parts, sensing energy consumption, computing energy consumption, and transmission energy consumption. Among them, transmission energy consumption accounts for 80% of the total energy consumption, so how to reduce network transmission energy consumption is considered an important performance indicator when designing network transmission protocols. In application scenarios such as environmental monitoring, the packet rate of wireless sensor network nodes is low, typically at the level of a few packets per minute, and when the wireless transceiver module of a wireless sensor node is on, the energy consumption of the node is 1000 times higher than when the node is in a dormant state. In the wireless sensor network transmission protocol, it is proposed to let the node periodically enter a dormant state to reduce the node's energy consumption when there is no data transmission energy consumed by idle listening.

$$P(S_{t-1}|S_t) = P(S_{t-1}|S_t, S_{t-1}, \dots, S_1). \quad (1)$$

The transfer matrix consisting of each state transfer in the above equation represents the probability of each node moving to the next node, and the sum of its transfer

probabilities sums to one. A complete Markov process contains the full number of node states and the probability of each state moving to each other.

$$P_{ss} = P[S_{t-1} = S]. \quad (2)$$

And the incentive process from a Markov process to a Markov process with the addition of $R(i, a)$ is considered an incentive. It corresponds to the expectation value of each state transfer process from one state to another. For example, in random, a state starts through constant state changes and finally reaches the final position smoothly, as long as by moving then there must eventually be a path of state transfer, each state change will have a $R(i, a)$, and thus, eventually $V_t(i, a)$ denotes the sum of all $R(i, a)$ from the initial start to the end.

$$V_t(i, a) = \min \left\{ R(i, a) - \gamma \sum_{j \in I} T(j|i, a) V_t(j) \right\}. \quad (3)$$

Existing approaches have studied this problem on deterministic routing, while the problem has not been well addressed in opportunity routing. Therefore, in this paper, delay-preserving low-power information collection mechanisms in opportunity routing are studied as the first part. For application scenarios with large data volumes, the limited transmission capacity of wireless sensor networks becomes a key factor affecting the transmission delay, and existing methods have drawbacks such as the need for known link transmission capacity and low network utilization, so the second part of this paper works on high-throughput transmission mechanisms in wireless sensor networks [19]. Also, considering that wireless rechargeable sensor networks mostly use RFID-based communication protocols, the many-to-one communication mechanism between nodes and readers makes the time complexity of the information collection mechanism proportional to the number of nodes, which cannot meet the demand of high time efficiency of the system.

$$\text{Square}(\Delta f * t) = \sum_{n=1,3,5,\dots}^{\infty} \frac{1}{n} \sin(2\pi * n * \Delta f * t). \quad (4)$$

In the Markov ant colony-based decision process, the Markov decision model is an optimal decision model for stochastic dynamic systems based on the Markov process theory, which can be used to select the paths satisfying the requirements as the allowed decision set for data forwarding by using its decision set at each node and does not depend on the historical moments of the system, as the allowed decision set may contain multiple paths, subject to further optimization process to select the optimal paths. Firstly, the Markov nodes are filtered and then based on the remaining nodes set, and this is used as the starting point to evolve through the ant colony algorithm generation by generation until the set optimization goal is satisfied then terminated, as shown in Figure 2.

When the source node sends the data to the destination node, the communication link establishment process between the source node and the destination node is considered an ant colony searching for food process taking the source node as the anthill and the destination node as the food, the nodes calculate the transfer probability of the allowed decision set within the communication range and the optimal path is selected by the transfer probability. Since direct transmission of data packets causes an increase in network latency and energy consumption, forward ants are proposed to be used to perform the optimal path finding before data transmission.

$$P\{R_i = r\} = \prod_{j \in \text{for}_i} (1 + p_{ij}), \quad (5)$$

$$E(S_i) = \frac{1}{T} \sum_{s=0}^T s P\{S_i = s\}. \quad (6)$$

Feature extraction is used to extract the information that best reflects the essence of a segment of speech. Feature extraction starts with windowing and framing the speech signal, treating each frame as a smooth signal, followed by FFT to obtain the frequency domain signal and extract acoustic features. The commonly used acoustic features are Mel frequency cepstrum coefficients, filter bank features, and perceptual linear prediction coefficients. Finally, the extracted speech features are downsampled and regularized to ignore the insignificant features and filter out noise interference.

3.2. Adaptive Proofreading Design for Spoken English Pronunciation. In the design of the speech recognition system, the continuous analog speech signal obtained from the speech acquisition circuit should first be digitized so that the system can perform feature extraction, template matching, and other operations on it in the later speech recognition process. In the speech signal acquisition, to avoid the phenomenon of signal overlapping and the occurrence of industrial frequency interference, the prefiltering operation should be completed before sampling the speech signal to filter out the interference including power supply noise. The interference due to power supply noise can be eliminated by adding a high-pass filter at 50 Hz during signal sampling. After the filtering of the power supply noise is completed, a suitable filtering frequency should be designed to consider the superposition of frequencies that occurs during the discretization of continuous signals.

After sampling the speech signal, a discrete signal in the time domain is obtained, but its amplitude is still continuous, so the signal needs to be quantized and coded to discrete the signal amplitude. The essence of quantization is to divide the discrete sampled values into several intervals after the discretization is completed, and the sampled values within the interval are identified as the same value, which is called the quantized value [20]. The data collected by the wireless sensor node can be transmitted to the network or satellite receiving protocol. In the communication process,

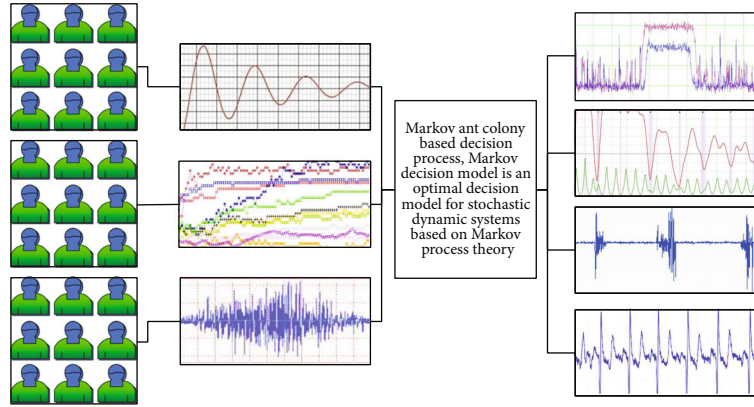


FIGURE 2: Signal separation process.

the convergent node carries limited energy and the transmission burden is large, resulting in rapid consumption of the remaining energy. As a result of rounding several values with slight differences and expressing them as the same quantization value, this gives rise to quantization error which we call quantization noise, and the way to reduce quantization noise is to increase the number of quantization bits and expand the range of amplitude variation of the speech signal as much as possible, so that the fewer values in a single quantization interval the better. Figure 3 shows the amplitude and phase frequency characteristics of the first-order high-pass filter used in the preemphasis process.

The speech signal is a signal that varies with time and mainly contains a clear tone part and a turbid tone part. The fundamental period of the turbid tone as well as the amplitude of the clear tone and the vocal tract parameters vary slowly with time, but due to the motion habits of the vocal organs, the speech signal can be approximately constant over a short period of time (generally, 10-30 ms), i.e., the speech signal possesses short-time smoothness. In the analysis of speech signals, the speech signal is divided into several analysis frames for processing. This is done by using a finite-length movable window function for weighting.

The overzero rate of effective speech under the same conditions is much lower than the overzero rate of sudden noise; for example, in the home use environment preset by the speech recognition control system designed in this paper, the overzero rate of effective speech commands is often less than the overzero rate of noise generated by actions such as opening and closing doors. This is because the energy of normal speech is generally concentrated in the low to medium band range, while the energy band range of noise can often be very high, which also results in an unusually high short-time overzero rate. Therefore, the definition of the short-time transient zero rate is improved in the system design. This way the calculated short-time zero rate has a certain anti-interference capability, even if there is random noise if it does not exceed the threshold band composed of positive and negative thresholds will not produce a false zero. In this paper, the dual threshold detection algorithm is designed to set the threshold values for the short-time amplitude and the improved short-time overzero rate,

respectively. When judging the speech onset, it is required to be able to avoid sudden noise caused by, for example, door and window opening and closing sounds and object collisions. This situation generally causes a sharp increase in the overzero rate within a very short period, but due to its short duration, short-time sudden noise can be excluded by setting the minimum duration of valid speech, and the current speech signal exceeds one or all the two thresholds, and the duration exceeds the minimum duration threshold. Then, the earliest time point that exceeds one or all of the two thresholds is returned and recorded as the starting point of speech; the intermediate silent segment of the word link cannot be discarded when discerning the end point of speech, and this situation is avoided by setting the longest duration threshold of the silent segment, which is lower than the two thresholds at the same time and exceeds the longest duration of the silent segment; Then, the time point that is most initially lower than the threshold is marked as the end point of valid speech, as shown in Figure 4.

Then, it is transmitted to the management node via satellite or the Internet. This multihop routing information transmission is also the main method of network energy consumption. In the communication process, after the WSNs are put on the sensor node, there is almost no possibility of node replacement. The advantage of using mutual information to determine the degree of association between single words is that it can identify word pairings that do not go together, which is very suitable for domain-specific noun pairings, but its reliance on the respective frequencies of two words will assign a larger mutual information value to chance cases once the frequencies are low, which is obviously unrealistic. Mutual information-based text error checking targets more than 2 consecutive scattered strings that appear after word separation, and constructing a mutual information model and combining it with scattered strings can find word errors in the text. If there is an error in a sentence, the position where the error is located will be cut into loose strings. In mutual information-based text error checking, the text is first divided into words, and then, the place where it is cut into loose strings is found, which is divided into single-word, double-word, and multiword loose strings.

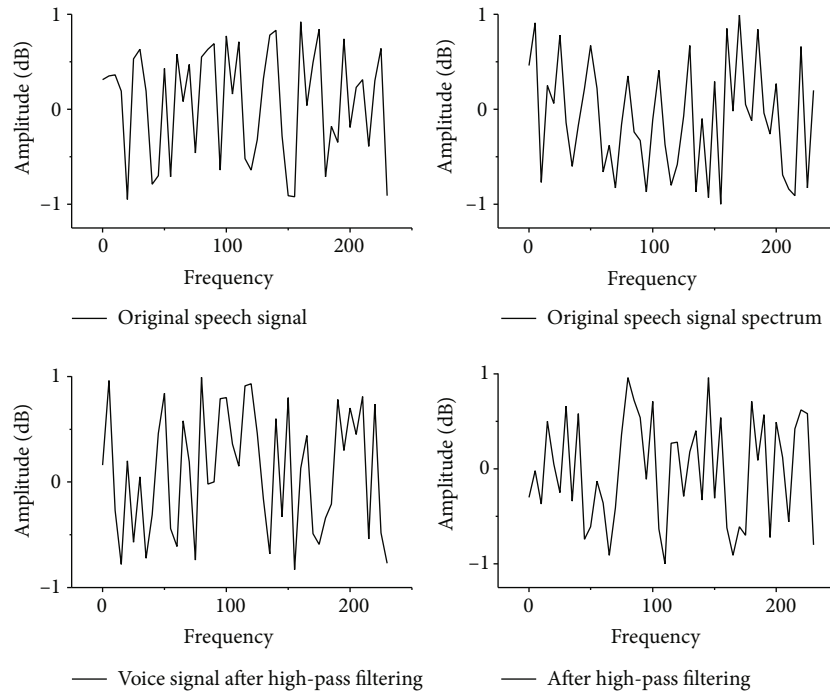


FIGURE 3: Comparison of speech signal and spectrum before and after preemphasis.

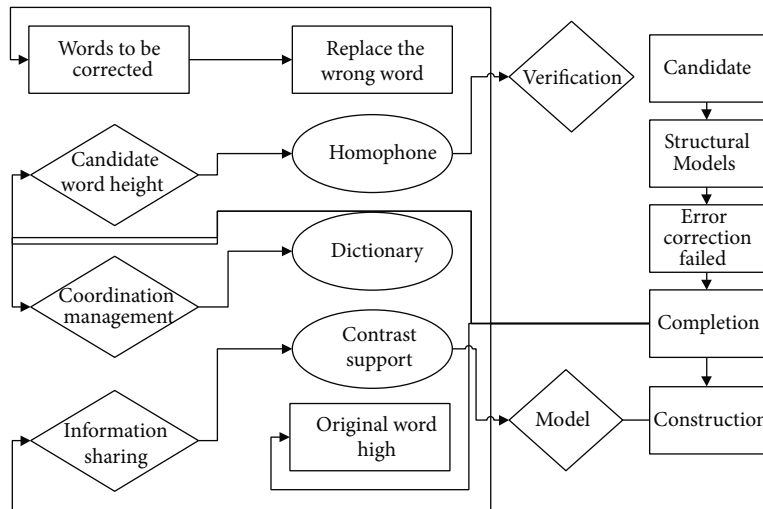


FIGURE 4: Error correction process of homophone dictionary.

The decoder module consists of a trained acoustic model, a dictionary and language model, and a speech decoding search algorithm. Both client side and server side have their own decoders and are not identical. The client side decoder requires small computing and storage volume and fast recognition speed, so the acoustic model, dictionary and language model, and even the decoding search algorithm need to be simplified, and the recognition accuracy is relatively low, while the server side decoder can use cloud technologies such as distributed computing and distributed storage to meet the big data processing requirements of speech intelligence recognition and group intelligence per-

ception applications, so the acoustic model, dictionary and language model, and decoding search algorithms can all be used in the most robust and robust way, and the recognition speed is faster when the network is smooth. In addition, the client does not have the training capability of the acoustic model and dictionary and language model but can do model tuning, and all model trainings are undertaken by the server.

The result of speech recognition is output by the client in the text format (i.e., word string or sentence) corresponding to the speech. The offline state directly outputs the search results from the client side decoder. The online state, on the other hand, requires the client to decode and then feed

the text result back to the server and compare it with the text result decoded by the server to see if it is consistent, and if it is inconsistent, the client will make model adjustments. At the same time, considering that RFID-based communication protocols are mostly used in wireless rechargeable sensor networks, and the many-to-one communication mechanism between nodes and readers, the time complexity of the information collection mechanism is proportional to the number of nodes, which cannot meet the high-efficiency requirements of the system. The online state speech recognition system output to the server side of the decoding results prevails, regardless of the client recognition results. The server records the client profile statistics recognition error information to train the personalized model. The dictionary and language model are targeted for discourse topic recognition. Since the discourse topics of speech input in group wisdom perception are diverse and limited by the knowledge domain of a particular person, and it is difficult and unnecessary for the dictionary and language model to include all words and semantics, and the performance of the embedded device will also limit the storage size and computing power, the client's dictionary and language model need to be gradually strengthened to meet the specific user's Generic and personalized needs are sufficient. In the online state, the server dynamically detects whether a new special topic is encountered in the speech input, and if the special topic vocabulary exceeds a self-defined threshold, the server will download the trained dictionary and language model of the special topic to the client, and the client will perform model mixing to achieve correct topic recognition. Of course, if a new special discourse topic vocabulary is encountered in the offline state, the client is unable to recognize it accurately.

4. Analysis of Results

4.1. Wireless Sensor Network Environmental Performance Results. This paper considers a multicategory RFI system, where the number of node categories will obviously have a critical impact on the performance of the proposed protocol. To investigate the effect of the parameters on the performance of the protocol, this paper keeps the other parameters constant and varies the value of L from 50 to 250. The experimental results in Figure 5 show that the execution time of the FDP protocol proposed in this paper remains essentially stable as the number of node categories increases. This is because this paper keeps the value of L constant, and only the category in the TOP- k set requires an accurate number estimate, while that node category in the non-TOP- k set can be eliminated quickly. Thus, even increasing the value of L does not make the execution time of FDP increase significantly. Unlike the protocols in this paper, this paper finds that the execution between other protocols increases significantly with the number of node categories. This indicates that the FDP protocol in this paper has good scalability with respect to L . The ZDE protocol needs to be executed separately on each node category, so the overall execution time should be proportional to the number of categories. The EDFSA protocol requires scanning to read all nodes in the current system, as shown in Figure 5.

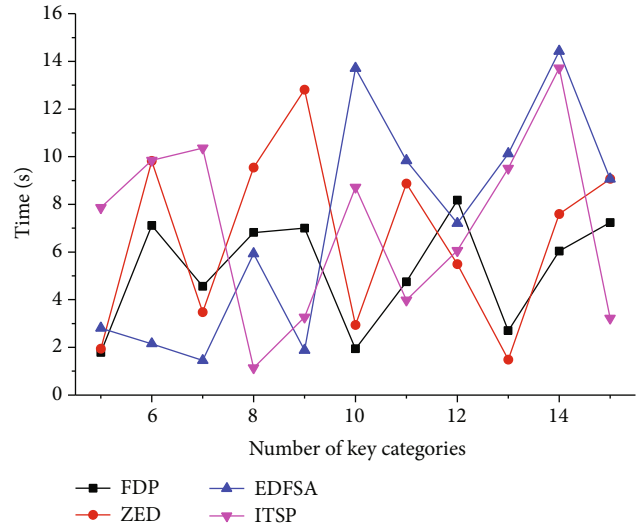


FIGURE 5: Impact of the number of key categories on the execution time of each protocol.

In this paper, the number of nodes in the current system is therefore proportional to L . Adding a 50 Hz high-pass filter can eliminate the interference signal caused by power supply noise. After the power supply interference noise is filtered, the overlap phenomenon that occurs during the discretization of the continuous signal must be considered, so it is necessary to design a suitable filter frequency. In the home environment preset by the voice recognition control system, the zero-crossing rate of effective voice commands is often smaller than the zero-crossing rate of noise generated by actions such as opening and closing doors. This is because the energy of normal speech is generally concentrated in the middle- and low-frequency bands. The ITSP protocol uses the interactive Bloom filter technique to identify anomalous nodes, and the length of the forward filter is proportional to UA , while the length of the reverse filter is proportional to UB . The size of the two sets above and hence the communication overhead increase proportionally. Moreover, when the value of L is large, the time efficiency of the FDP protocol in this paper is significantly higher than other protocols. For example, when $L=250$, the execution time of ZDE protocol is 621.9 seconds, the execution time of EDFSA protocol is 929.7 seconds, and the execution time of ITSP protocol is 316.6 seconds, while the execution time of the FDP protocol proposed in this paper is only 62.2 seconds, which is 90%, 93.3%, and 80.4% higher in terms of time efficiency, respectively.

In large-scale RFID systems, multiple readers usually need to be deployed. In this paper, we study the query algorithm when a single reader is used. The overall execution time does not increase if multiple readers are parallelized to execute the algorithm in this paper and thus has a relatively good scalability, as shown in Figure 6.

In this section, this paper will study the estimation accuracy of the key classes of anomalous nodes returned by the FDP protocol. The frequency of node category 18 appearing in the TOP- k set K is 95%, which meets the default query

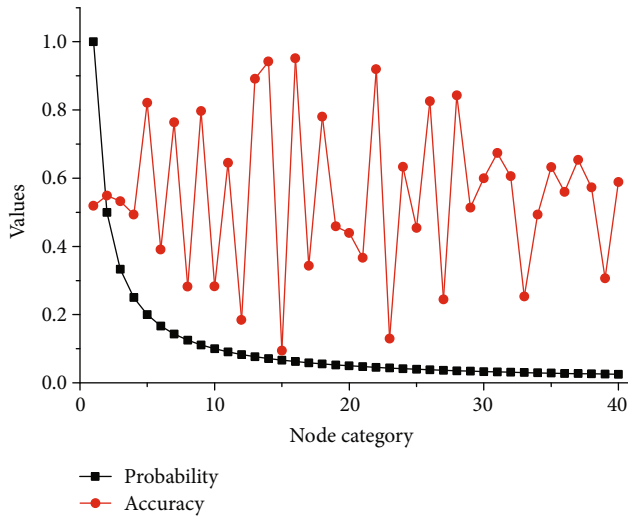


FIGURE 6: Estimated accuracy.

accuracy of this article, while the 9th and 10th types of nodes appear in the TOP- k set at a frequency lower than expected. The estimation accuracy contains two levels: firstly, whether the probability of the nodes in class k with the highest number of anomalous nodes appearing in set K is satisfied and secondly whether the number of anomalous nodes in these k classes is estimated accurately. Assuming that the nodes with the highest number of anomalous nodes are in the top 10 categories, respectively, the experimental results in Figure 6 show that node category 18 appears in the TOP- k set K with 95% frequency, which satisfies the default query accuracy of this paper, while nodes in categories 9 and 10 appear in the TOP- k set with a frequency that seems to be lower than the expected value. However, the number of anomalous nodes in categories 9 and 10 are 543 and 541, respectively; both of which are less than the threshold value just mentioned. Therefore, it is also normal that the frequency of nodes in these two categories appearing in the TOP- k query results is lower than β . The probability frequency of nodes in categories 9 and 10 appearing in TOP-10 is still much higher than the other node categories that follow. Because the first 10 categories of nodes appear in the TOP- k set with the greatest frequency, this paper counts whether their estimates of the number of anomalous nodes are accurate when they appear in the TOP- k set. For example, in 500 experiments, category 10 appears in the TOP- k set 315 times, and a total of 303 of these estimates satisfy the counting accuracy of $(\alpha, 3)$. Then, in this paper, 0.9619 is recorded as the estimation accuracy of the number of anomalous nodes in category 10.

4.2. English Spoken Pronunciation Proofreading Results. The verification block diagram of the speech recognition module is shown in Figure 7 and includes a speech input acquisition module, a result display module, and an input control module. Among them, the input control module as well as the display module is implemented on the TFTLCD display that serves as the GUI interface. When setting the GUI display needle, the screen backlight is adjusted for different situa-

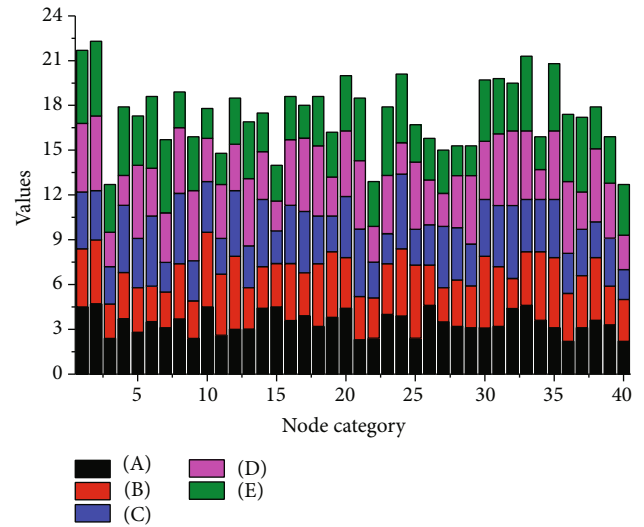


FIGURE 7: Phase A and B waveforms.

tions to achieve the reminder function. The main task of the stepper motor in the design of this paper is to control the movement of household items including smart curtains in the horizontal direction, which requires the motor to have the functions of forward rotation, reverse rotation, stop, acceleration, and deceleration. For example, when the voice command prompts the need to pull open the curtain or close, the STM32 control core board through the timer output pulse through the GPIO to the TB6560 as the core of the motor drive circuit CLK pin to control the motor to start movement preset the initial situation; the definition of the pull open curtain command corresponds to the positive rotation of the motor, corresponding to the CW direction control pin being low or suspended. When the curtain pulling action is to be performed, the CW pin needs to be set high, so that the motor reverses to complete the opposite control action, as shown in Figure 7.

Through the introduction of the previous part of this section, the various functional modules of the speech recognition control terminal can achieve the predefined functions and finally the overall evaluation of the speech recognition-based control terminal is carried out to test the control range of the speech recognition-based control terminal. Again, many tests are conducted for the commands in the previous speech recognition test, and the testers are trained with feature templates separately, and then, the control tests are conducted at different distances according to the effective range of the speech acquisition module as well as the wireless communication transmission module. The test results for different distances are shown in Figure 8. The data in the table shows that the control range of the control terminal is within 15 meters, beyond which the wireless communication success rate plummets, in addition to the differences in control success rate due to different rates of correct speech recognition in different environments. The control terminal designed in this paper has an effective control range of about 15 m within the light noise generated by normal homes.

The final implementation of the control terminal was determined through analysis and research, and the overall

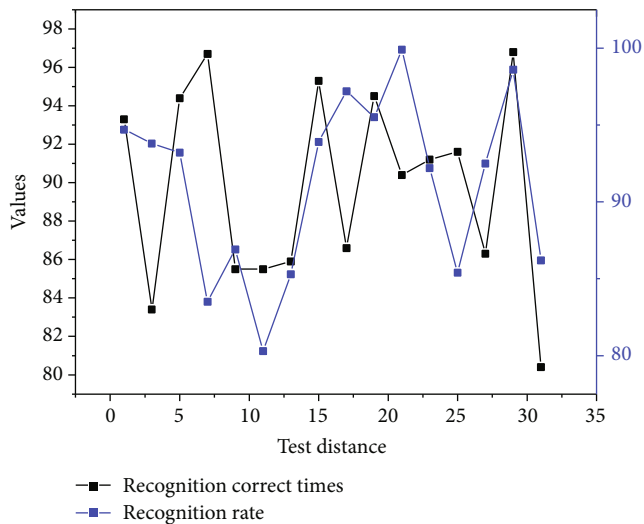


FIGURE 8: Calibration results at different distances.

structure of the terminal was determined to include a voice recognition processor as the main device and a control output processor as the control part. The information is transmitted between the two by wireless communication. The study discusses the key technical aspects of the control terminal design, such as the selection of the speech recognition matching algorithm and the determination of the wireless communication method. It is shown that although both their recognition accuracy is about the same, there is less corpus that is not recognized and more stable recognition by KDDI, which is the reason the recognition result of KDDI is selected as the system input in this paper. From Figure 8, although the traditional simulation system incorporating speech recognition has improved accuracy, with a concentrated distribution of 60% to 70%, it has the same problem as Baidu speech recognition, that is, loose distribution and more distribution in smaller intervals, causing more corpus to be unrecognized. It will not increase the execution time of FDP significantly. For the system studied in this paper, the accuracy has improved greatly and the distribution is more concentrated, mainly in the interval of 90% to 95%, which is due to the inclusion of the domain lexicon, which reduces the number of unrecognized corpus.

5. Conclusion

This paper focuses on high-throughput transmission mechanisms under high-load application scenarios in wireless sensor networks. With the continuous development of wireless sensor networks, sensors have a rich means of acquiring information about the physical world, which imposes high-throughput requirements on information collection mechanisms. Since link transmission capacity is very difficult to predict online, the problem of maximizing network utility cannot be directly used to improve the throughput of the network by dynamically adjusting the transmission rate of the nodes. To improve the training process where the language model cannot be integrated into the acoustic model, this paper proposes an end-to-end model based on spell cor-

rection model which can correct the errors generated by CTC-based speech recognition systems; this fusion method uses the output of CTC as input to train the language model after doing some matrix operation, which truly achieves end-to-end and alleviates the condition between the output of the CTC model. The unreasonable assumption of independence and the drawback of not being able to fuse the language models are cons; the feasibility of this improved algorithm is proved by the comparative analysis of experimental results, and the model shows better recognition results than other models. The end-to-end speech recognition algorithm and model based on the fused language model are flowed to construct a web version of the speech recognition system, which can complete offline and online recognition of speech files, verifying the applicability of the algorithm in this paper.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] S. Z. Sweidan and K. A. Darabkh, "VREG: a virtual reality educational game with Arabic content using android smart phone," *Journal of Software Engineering and Applications*, vol. 11, no. 10, pp. 500–520, 2018.
- [2] H. Y. Wang, V. Lin, G. J. Hwang, and G. Z. Liu, "Context-aware language-learning application in the green technology building: which group can benefit the most?," *Journal of Computer Assisted Learning*, vol. 35, no. 3, pp. 359–377, 2019.
- [3] G. Z. Liu, J. Y. Chen, and G. J. Hwang, "Mobile-based collaborative learning in the fitness center: a case study on the development of English listening comprehension with a context-aware application," *British Journal of Educational Technology*, vol. 49, no. 2, pp. 305–320, 2018.
- [4] L. Dell'Angela, A. Zaharia, A. Lobel, O. Vico Begara, D. Sander, and A. C. Samson, "Board games on emotional competences for school-age children," *Games for Health Journal*, vol. 9, no. 3, pp. 187–196, 2020.
- [5] M. Vacher, F. Aman, S. Rossato, F. Portet, and B. Lecouteux, "Making emergency calls more accessible to older adults through a hands-free speech interface in the house," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 12, no. 2, pp. 1–25, 2019.
- [6] D. J. Clink and H. Klinck, "Unsupervised acoustic classification of individual gibbon females and the implications for passive acoustic monitoring," *Methods in Ecology and Evolution*, vol. 12, no. 2, pp. 328–341, 2021.
- [7] L. Gurney and A. Diaz, "Coloniality, neoliberalism and the language textbook," *Language, Culture and Society*, vol. 2, no. 2, pp. 149–173, 2020.
- [8] F. Kodama, "An empirical study of contrasting IoT with IT: evidences of differences drawn from Japanese experiences,"

- American Journal of Industrial and Business Management*, vol. 8, no. 1, pp. 27–58, 2018.
- [9] A. Jangizehi, F. Schmid, P. Besenius, K. Kremer, and S. Seiffert, “Defects and defect engineering in soft matter,” *Soft Matter*, vol. 16, no. 48, pp. 10809–10859, 2020.
- [10] A. K. Alpotte, M. Zivkovic, I. Branovic, and R. Popovic, “Multilingual virtual environment for wireless sensor networks,” *Computer Applications in Engineering Education*, vol. 25, no. 2, pp. 200–213, 2017.
- [11] I. Fathi, Q. Ibrahim, and J. M. Abdul-Jabbar, “Real-time voice transmission over wireless sensor network (VoWSN) based automatic speech recognition (ASR) technique,” *Al-Rafidain Engineering Journal (AREJ)*, vol. 24, no. 2, pp. 23–34, 2019.
- [12] S. Chen, “Design of internet of things online oral English teaching platform based on long-term and short-term memory network,” *International Journal of Continuing Engineering Education and Life Long Learning*, vol. 31, no. 1, pp. 104–118, 2021.
- [13] D. Zhou, L. Liu, T. Tang, Y. Huang, Y. Lee, and J. Hong, “Design on intelligence music system in the cultural center based on IoT,” *Personal and Ubiquitous Computing*, vol. 24, no. 3, pp. 319–332, 2020.
- [14] S. Gul, M. Asif, S. Ahmad et al., “A survey on role of internet of things in education,” *International Journal of Computer Science and Network Security*, vol. 17, no. 5, pp. 159–165, 2017.
- [15] Y. Fu, K. Mechitov, T. Hoang, J. R. Kim, D. H. Lee, and B. F. Spencer Jr., “Development and full-scale validation of high-fidelity data acquisition on a next-generation wireless smart sensor platform,” *Advances in Structural Engineering*, vol. 22, no. 16, pp. 3512–3533, 2019.
- [16] C. Chen, “A study on hybrid course design of the society and culture of major English-speaking countries based on CBI approach,” *International Journal of Social Science and Education Research*, vol. 4, no. 6, pp. 136–143, 2021.
- [17] A. I. Latif, A. M. Daher, A. Suliman, O. A. Mahdi, and M. Othman, “Feasibility of internet of things application for real-time healthcare for Malaysian pilgrims,” *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 3, pp. 1169–1181, 2019.
- [18] P. T. Daely, H. T. Reda, G. B. Satrya, J. W. Kim, and S. Y. Shin, “Design of smart LED streetlight system for smart city with web-based management system,” *IEEE Sensors Journal*, vol. 17, no. 18, pp. 6100–6110, 2017.
- [19] M. Yu, M. Bambacus, G. Cervone et al., “Spatiotemporal event detection: a review,” *International Journal of Digital Earth*, vol. 13, no. 12, pp. 1339–1365, 2020.
- [20] K. L. Hsiao, T. C. Huang, M. Y. Chen, and N. T. Chiang, “Understanding the behavioral intention to play Austronesian learning games: from the perspectives of learning outcome, service quality, and hedonic value,” *Interactive Learning Environments*, vol. 26, no. 3, pp. 372–385, 2018.