

## Research Article

# The Technique of English Word Syllable Division in Speech Synthesis Based on Neural Network

Hongjuan Ma 

*Applied Foreign Languages Department, Xinyang Vocational and Technical College, Xinyang 464000, Henan, China*

Correspondence should be addressed to Hongjuan Ma; [xyzrscwl@xyvtc.edu.cn](mailto:xyzrscwl@xyvtc.edu.cn)

Received 12 July 2021; Revised 14 August 2021; Accepted 23 August 2021; Published 21 September 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Hongjuan Ma. 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.

With the increasing maturity of speech synthesis technology, on the one hand, it has been more and more widely used in people's lives; on the other hand, it also brings more and more convenience to people. The requirements for speech synthesis systems are getting higher and higher. Therefore, advanced technology is used to improve and update the accent recognition system. This paper mainly introduces the word stress annotation technology combined with neural network speech synthesis technology. In Chinese speech synthesis, prosodic structure prediction has a great influence on naturalness. The purpose of this paper is to accurately predict the prosodic structure, which has become an important problem to be solved in speech synthesis. Experimental data show that the average error of samples in the network training process is  $1e/85$ , and the minimum value of the training error after 500 steps is 0.00013127, so the final sample average error is  $1e/85 * 0.00013127 = 0.112 < 0.5$ , and use the deep neural network (DNN) to train different parameters to obtain the conversion model, and then synthesize these conversion models, and finally achieve the effect of improving the synthesized sound quality.

## 1. Introduction

The Internet is changing the teaching methods of teachers. Nowadays, one teacher alone can teach thousands of students. Using the interactivity of the Internet, teachers and students can interact in a timely manner, so as to achieve a similar effect to small class teaching. At the same time, people can enjoy high-quality online courses without leaving home. As long as there is a computer, anyone can receive a university education anywhere [1].

The basic unit of foreign English pronunciation is phoneme; it is composed of 41 phonemes. An English word contains multiple phonemes, and each phoneme corresponds to a syllable, so the pronunciation of an English word is composed of multiple syllables. The process of converting the characters of English words into the phonemes of words has been relatively mature and will not be discussed here [2, 3]. The division of syllables in English words is one of the most important parts in the technical requirements of machine synthesized sound, which seriously affects the accuracy and naturalness of pronunciation [4, 5]. Dictionary

lookup table method was compared with traditional methods [6]. At the same time, there are still many applications in the research. These applications are very attractive, which shows the great potential of neural network applications. Neural network technology is especially good at processing which has a large number of the sample applications [7]; the problem of pronunciation of English words is solved by using complete learning samples. The problem involves phonemes and phonetic symbols. The entire learning sample involves tens of thousands of English words. The following is the use of neural network technology to solve the problem [8, 9].

Geoffrey Hinton obtained some surprising results on MNIST and showed that the acoustic model of a frequently used commercial system can be significantly improved by distilling the knowledge in the model set [10, 11]. Yang Shen's program has been trained to extract trunk. The program is generally very robust and yields more than 90% of information about the residual and side-chain 1 rotator isomers with a torsion angle over 90%. In addition to reliably predicting the secondary structure, deep neural networks

can be applied to study end-to-end speech synthesis methods, so that the text features can be directly mapped to the speech spectrum to get better effect of speech synthesis but requires a larger corpus and longer training time [12–15].

This paper firstly introduces the development history and current situation of speech recognition technology at home and abroad and then conducts theoretical research and analysis on each link, including speech acquisition, preprocessing, endpoint detection, feature parameter extraction, time sorting, and other aspects. In each stage of network and the theory and algorithm of speech recognition model, a complete design scheme of speech recognition system is presented by selecting MFCC as the speech characteristic parameter. This paper focuses on the selection of identification models. By comparing various recognition algorithms, the BP neural network is selected as the basic unit of the recognition model. Aiming at the accuracy of speech recognition and the deficiencies of the BP neural network algorithm, this article introduces the neural network integration theory, using individual differences and combining k-means clustering method to improve the integration network, the improved neural network will be able to effectively integrate multiple BP network construction cost.

## 2. Proposed Method

### 2.1. Classification and Basic Composition of Speech Recognition System

2.1.1. *Classification of Speech Recognition System.* Speech recognition system is generally classified based on two classification methods.

(1) Based on Recognition Objects. According to different objects of recognition, speech recognition tasks can be roughly divided into two categories, namely, isolated word recognition and continuous speech recognition. Among them, the task of isolated words is to recognize speech.

(2) Based on the Size of Vocabulary. Some speech recognition systems can only recognize the terms in the preset vocabulary of the system, and the terms outside the vocabulary cannot be recognized. Other systems are based on the recognition of phonemes, which often have no restrictions on the recognition of terms. In short, according to the size of the vocabulary, it can be divided into small vocabulary (less than 100), medium vocabulary (100–500), and large vocabulary (more than 500). The larger the vocabulary, the greater the confusion between terms and the lower the recognition rate of the system.

2.1.2. *Basic Structure of Speech Recognition System.* The speech recognition system is mainly composed of four parts: signal processing and feature extraction, acoustic model (AM), language model (LM), and decoding search part. The typical implementation scheme of the speech recognition system is shown in Figure 1.

Then enter the training stage, process the acquired feature parameters, extract the data that can best reflect the

voice features, and save it to the template library. Finally, in the recognition stage, voice parameters are obtained through the same channel to form the test template, which is then matched with the reference template in the template library. Then, the similarity is calculated according to a certain quasi-selection (such as distance measurement), and the similarity with the highest matching score is output as the result. At the same time, some prior knowledge or fuzzy theory can be added to improve the accuracy of the recognition system.

### 2.2. Basic Principles of Speech Recognition

#### 2.2.1. Mathematical Model of Speech Signal

(1) *Acoustic Model of Speech Signal.* After selecting a series of system parameters related to the vocalization process, speech synthesis technology is to process the input text type signal sequence through a specific synthesizer to produce high naturalness, high sound quality, and rich expressive voice output, so that the computer or related system can be equipped with a technology that can emit natural and smooth sound like a “human.” Speech synthesis technology acts as the output part of the machine in the human-machine voice interaction and plays an important role; some desired speech feature sequences can be obtained. In general research and application of speech coding and speech recognition, the discrete time domain model shown in Figure 2 is usually selected.

Among them, the excitation model is divided into voiced excitation and unvoiced excitation based on voiced and soft sounds. The vocal tract model can build three practical models based on the pronunciation of various phonemes, which are cascade, parallel, and hybrid models. The airflow formed by the radiation model in the articulation cavity radiates through the lips and reaches the listener’s ears. The sound signal will be attenuated and has the characteristics of high-pass filtering. In the excitation model, the contents of the excitation source package do not sound and sound. The silent signal is the output of the linear system excited by the white noise sequence, while the sound signal is generated by the periodic pulse generator. When voiced, the transfer function of the system is

$$H_v(z) = G(z)V(z)R(z). \quad (1)$$

The function of  $Av$  is to adjust the amplitude or energy of the voiced signal, where  $z$  represents the amplitude of the voiced signal.

In order to make the sound excitation signal have the actual waveform of glottic pulse, it should pass the glottic pulse model filter and convert its  $z$ -field to

$$G(z) = \frac{1}{(1 - g_1z^{-1})(1 - g_2z^{-2})}. \quad (2)$$

Both  $g_1$  and  $g_2$  are close to 1, so the spectrum of turbidity excitation is close to that of glottic flow pulse. When a voiceless sound is produced, whether it is a blocking sound

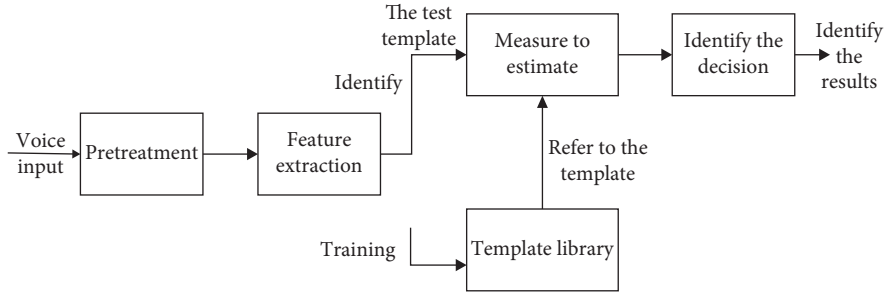


FIGURE 1: Speech recognition system structure.

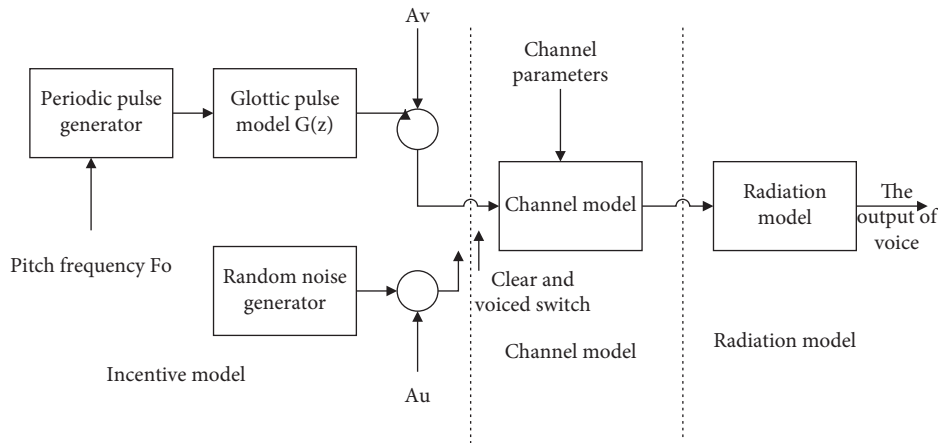


FIGURE 2: The discrete time domain model generated by voice signal.

or a frictional sound, the sound channel will be blocked to form turbulence. Therefore, the silent signal can be simulated as random white noise with average value of 0 and variance of 1, which is randomly distributed at any time or amplitude.  $V(z)$  in the sound channel model is the sound channel transfer function, and its expression is

$$V(z) = \frac{1}{1 - \sum_{k=1}^N \alpha_k z^{-k}} \quad (3)$$

$R(z)$  in the radiation model is a first-order high-pass filter, whose expression is

$$R(z) = R_0(1 - z^{-1}). \quad (4)$$

(2) *Human Auditory Model.* The auditory system of the human ear is a very precise and complex system. Speech is transmitted from the vocal organ through the medium to the auditory organ of the human ear. The parameters in the acoustic model will more or less change with energy loss. After a long period of research, it is found that better results can be obtained by modeling speech recognition system from the perspective of human ear hearing. On the one hand, it explores human hearing, physiology, and psychology, which makes up for the deficiency of voice characteristic parameters in the voice channel model. On the other hand, it analyzes the sound principle of voice signal

from the aspect of frequency domain. It is found that there is a pairwise relationship between high and low sound and treble sensation and frequency, and people's perception of complex sound patterns is a complex auditory process driven by knowledge and data.

### 2.2.2. Speech Signal Preprocessing

(1) *Speech Signal Preaggravation.* Prew weighting actually refers to the pre weighting of the high-frequency part of the speech signal. Because the average power of the speech signal is stimulated by the radiation of the mouth and nose and the glottis, it has been attenuated by more than 800 Hz. In the high-frequency part-of-speech spectrum, because the high-frequency components are lower than the low-frequency components, it is difficult to find an image in the frequency part of the spectrum. For places where the brightness or gray level changes drastically, high-frequency components, such as edges, should be drawn; and for places where the changes are not large, a histogram of low-frequency components, such as large color blocks, should be drawn. Large areas are low frequencies, small areas or discrete areas are high frequencies. The image is regarded as a two-dimensional function, and the places with drastic changes correspond to high frequencies, and vice versa, coupled with the low-frequency end mixed with 50 Hz or 60 Hz interference

information, so it is necessary to improve the high frequency part of the speech recognition spectrum to remove the interference part to facilitate subsequent speech parameter analysis, as follows:

$$H(z) = 1 - aZ^{-1}, \quad 0.9 < a < 1. \quad (5)$$

Generally,  $a = 0.9375$  should be used to calculate the short-term energy of each frame of speech, which can eliminate dc drift, suppress random noise, and improve the clearing energy.

(2) *Framing and Windowing of Voice Signal.* In order to use digital signal processing technology to analyze and process speech signals changing with time, speech signals are divided into several frames to analyze their characteristic parameters. The length of each frame is 10–30 ms, and each frame has short-term stability. Framing can be continuous or overlapping. In order to smooth the transition between frames and maintain the continuity, the overlapping segmentation method is generally used. There are two window functions: ① The width of the main lobe of the window spectrum should be narrow to obtain a steeper transition zone. ② Compared with the amplitude of the main lobe, the side lobes should be as small as possible so that the energy is concentrated in the main lobe as much as possible, so that the small shoulder peak and after vibration can be reduced, and the stop band attenuation and pass band stability can be improved. The ratio of frame move to frame length is usually 0-1/2, as shown in Figure 3.

The commonly used window functions in speech signal processing are rectangular window and Hamming window, whose expression is shown in equation (6) ( $N$  is the frame length):

Rectangular window:

$$W(n) = \begin{cases} 1, & 0 \leq n \leq (N-1), \\ 0, & \text{others,} \end{cases} \quad (6)$$

Hamming window:

$$W(n) = \begin{cases} 0.54 - 0.46 \cos\left[\frac{2\pi n}{N-1}\right], & 0 \leq n \leq (N-1), \\ 0, & \text{others.} \end{cases} \quad (7)$$

(3) *Endpoint Detection of Speech Signals.* The starting and ending points of speech signals are not given in advance, and speech signals are often mixed with some background noise. Therefore, in the practical application of speech recognition, it is necessary to accurately find out the starting point and ending point of speech and remove the background noise unrelated to the signal, which not only reduces the amount of data, computation, and processing time, but also helps to improve the recognition rate of the system. Endpoint detection is a key part-of-speech recognition, which directly affects the performance of the whole system.

The function of short-term average energy is to distinguish between voiced and unvoiced sounds in speech signals. After

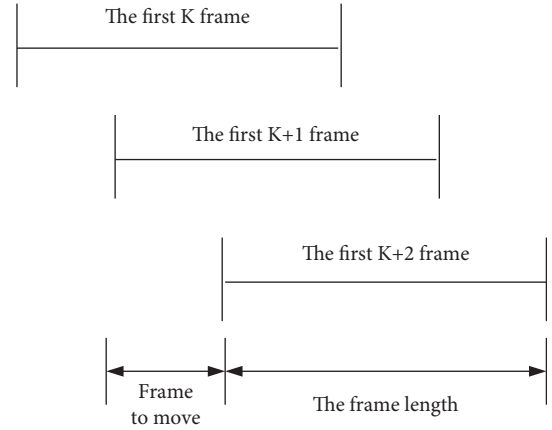


FIGURE 3: The frame move and the frame length.

the speech signal is divided into several frames and windowed, the energy of each frame (the amplitude of each frame) is calculated, and the calculated energy is called the short-term average energy. Therefore, after calculating the energy value of each frame, a threshold is set to theoretically separate the silent segment from the sound segment. The short-term average energy  $E(I)$  of the speech signal in frame  $I$  can be obtained by one of the following three algorithms, in which  $N$  is the frame length and  $X_i(N)$  is the amplitude energy of the speech message at the  $N$ th point. The simple method is to directly calculate the volume of the sound to find the part whose volume is greater than a certain threshold, and consider this part as the required speech signal, the intersection of this part and the threshold as the endpoint, and the rest as a nonspeech frame.

$$\begin{aligned} E(i) &= \sum_{n=1}^N |X_i(n)|, \\ E(i) &= \sum_{n=1}^N X_i^2(n), \\ E(i) &= \sum_{n=1}^N \log X_i^2(n). \end{aligned} \quad (8)$$

Voiced signals are issued by the vocal cords, and the corresponding voice signals have a high amplitude, but generally there is no change between higher positive and negative values, while voiceless signals do not have the vibration of the vocal cords but are produced by the friction, impact, or blasting of the air in the oral cavity, with a lower amplitude and a higher zero point sound repeated back and forth. Therefore, the short time average energy can be used to get the voiced signals better, but the zero crossing rate is needed to solve the problem of unvoiced signals segment extraction.

The short time average zero crossing rate is a method to estimate the sinusoidal frequency in time domain analysis. The calculation formula is shown in the following equation:

$$Z_n = \frac{1}{2} \sum_{m=-\infty}^{\infty} |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| w(n-m), \quad (9)$$



The two output values of the neural network are inversely mapped to two binary numbers, and the same phoneme position must be a syllable. The syllable identifier can be added to the corresponding input position to realize the syllable partition phoneme.

**3.2. Experimental Design.** Neural network structure design has always been a very meaningful but difficult to solve problem. This article reviews and summarizes the research status of neural network structure design in recent years. First, the four criteria that should be considered in the design of neural network structure are analyzed, namely, the function approximation error of the neural network, the complexity of the network structure, the generalization ability of the network, and the fault tolerance of the network. Fortunately, with the MATLAB neural network design platform, the purpose of modifying the neural network structure can be easily achieved, and only one sentence needs to be completed. Therefore, it is very convenient to use the trial and error method to design the structure of the neural network [16, 17]. The sample input starts with the smallest number and then enters 5000 samples in turn to find a suitable neural network structure and then takes the final 10000 samples. After a lot of simulation experiments, a neural network with a four-layer structure of  $16 \times 30 \times 20 \times 2$  was found. The activation function of each layer is expressed as  $\tanh(x)$ ,  $\tanh(x)$ , and  $\text{sigmoid}(x)$ , that is, the first two layers of a function are the tangent S after a logarithmic sigmoid function. The neural network structure is shown in Figure 5.

## 4. Discussion

**4.1. The MATLAB Platform.** The platform data come from the data set of Spoken Arab Digit, the UCI machine learning repository, which contains a total of 8800 voices and data, of which 6600 pronunciations are used as training samples and 2,200 pronunciations are used as test samples. Using the bootstrap method, each time 5000 from the training samples are selected as the number of training samples for the weak classifier, and the training is repeated 200 times [18, 19].

The error outputs of the single BP network model, the traditional BP-AdaBoost model, and the model in this paper were compared under 2,200 test samples. The error accuracy of each BP network was set as 0.03. The number of base classifiers in the latter two integrated models was 25. The result is shown in Figure 6.

When the number of base classifiers is different, the correct recognition rate of the two models is shown in Table 1 (where the error accuracy of each base classifier is 0.008).

As shown in Table 1, when the number of base classifiers is 50, the difference of base classifiers in BP-AdaBoost model is not obvious, and the recognition rate is lower than that when the number of base classifiers is 25, but the recognition rate in this model can continue to increase [20]. 5 experiments are conducted under three models when the error precision of each base classifier is 0.06, 0.03, 0.01, 0.008, and 0.003, respectively. The results are shown in Figure 7. The X-

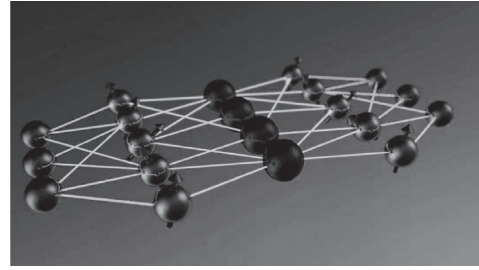


FIGURE 5: Experimental design of neural network structure.

axis is the experiment sequence number, and the Y-axis is the recognition rate under 2,200 test samples.

The model in this paper and the decision tree model were tested in the same size data set many times. The comparison results of the recognition rate are shown in Table 2, which shows the effectiveness of the proposed method.

**4.2. Analysis of Experimental Results.** While trying to design the neural network structure, we are also looking for appropriate training methods. The results show that with the increase of sample size, BP method, BP method with momentum term, and self-adaptive method are more and more effective [21]. The training results of the acquisition rate BP method, the adaptive learning rate and momentum term method, and the G method are not ideal; although the Levenberg–Marquardt and quasi-Newton methods have good training effects, the memory of these two methods is very large. When the number of samples increases to more than 10,000, the computer memory is not enough [22, 23].

**4.2.1. Prediction of Speech Structure.** The labeled corpus used in speech synthesis was completed by the Institute of Human-Computer Interaction and Media Integration (TH\_CoSS) of Tsinghua University, with a total of 5,406 sentences, of which 5,000 were training sentences. The file was TH\_CoSS.txt, and the remaining 406 sentences were test sets. Prediction of rhythmic structure is mainly divided into the following three steps [24].

The first step is to segment the news data of the whole network (Sogou CA, the size is 2.1 G), deal with the problems in the data files, and obtain clean full-text data. Use the Chinese word segmentation open source tool `jiaba` run `jieba_seg.py` to perform word segmentation and merge the word segmentation text corpus and the labeled corpus (TH\_CoSS.txt) with good word segmentation as the word vector training input data: (1) perform word segmentation and part-of-speech tagging on the target sentence; (2) use word segmentation and part-of-speech information to implement prosodic word formation; (3) perform syntactic segmentation and syntactic labeling on the target sentence; (4) build a prosodic structure prediction tree; (5) determine the location of the prosodic boundary. [25].

The second step is to use `word2vec` for word vector training. When training the word vector, the word vector we

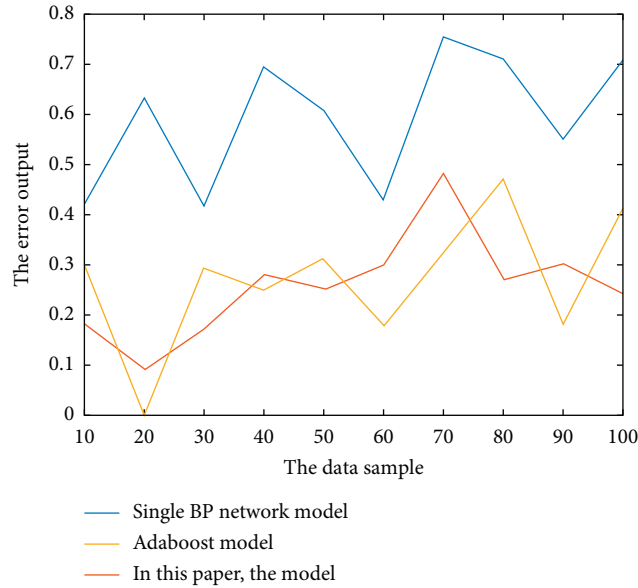


FIGURE 6: Comparison of error output under different models.

TABLE 1: Comparison of the AdaBoost model and the KASWT model in training time.

Number of base classifiers	BP-AdaBoost recognition rate (%)	The model recognition rate in this paper (%)
5	88.36	88.50
15	88.47	89.66
25	89.50	91.03
50	89.37	92.68

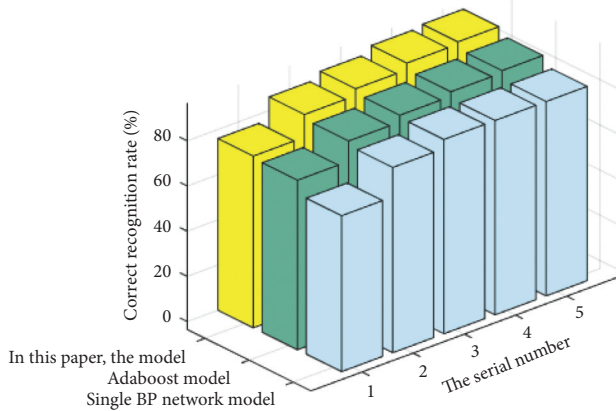


FIGURE 7: Recognition rates of different models with varying error accuracy.

TABLE 2: Comparison of recognition rate between the decision tree model and the method in this paper.

Training data size	Tree model (%)	KASWT model (%)
400	72.25	72.60
800	87.45	93.75
1200	87.40	88.65
1600	87.75	89.65
2000	90.35	91.80

trained will eventually be used to train the network model; that is to say, the words in the annotated corpus used to train the network model need to be found in the word vector scale.

In the third step, the training corpus is used to find the word vector, and a small program prepared by C is used to mark prosodic words as 0/1 according to the word segmentation results to form the training data.

The predicted results obtained are shown in Table 3.

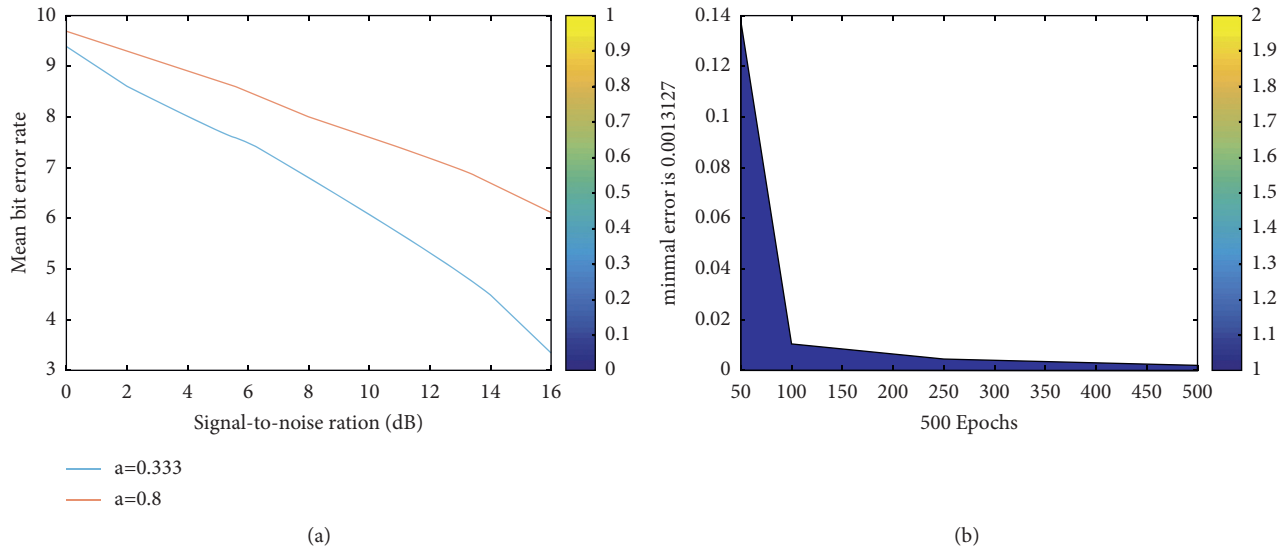
It can be seen from Table 3 that as the dimension of the word vector increases, the corresponding training time of the model also increases. The error rate of prosodic word prediction decreases with the increase of vector dimension. After the minimum 20 dimensions, the error rate will not decrease further as the dimension of the word vector increases, but will increase slightly. Therefore, when the dimension is too large, the prosodic error rate will not increase, but will increase the training time of the model.

**4.2.2. Speech Structure Preprocessing.** After a lot of simulation training, it is found that 35,161 sample points can be fully fitted within the required accuracy range. Figure 8 shows the average error of samples in the training process of neural network (1e1/85). The minimum training error after 500 steps is 0.00013127.

After the neural network training in MATLAB software, the weight matrix  $W$  and the threshold vector  $b$  written by the C language training module of the neural network

TABLE 3: Relation between input feature dimension and error rate.

Word vector dimension	Error rate (%)	Time
5	14.6	About 5 hours
10	10.0	About 7 hours
20	9.82	About 9 hours
60	9.95	About 12 hours

FIGURE 8: Mean error ( $lel/85$ ) of samples during neural network training.

calculation module are transplanted and removed, and the output processing is rounded to obtain two positive integers, and then put them into a 16-bit binary string with 0 and 1, and use the binary string information to annotate the X type to get the final result. After running the actual C language program and comparing the results with the actual samples, it is found that the recognition accuracy can reach 94.7%, which indicates that the performance of this method is very high.

## 5. Conclusions

Speech recognition technology, Android platform, and neural network integration theory are not only hot issues in academic research, but also a direction of market application promotion. Based on the analysis and summary of the existing speech recognition theoretical models and recognition algorithms, in view of the low accuracy of traditional BP neural network theory in speech recognition, the introduction of speech recognition technology combines neural network theory with it and improves the separate k-means clustering method of the traditional integrated network generation part in the network. Finally, voice synthesis function is added on Android2.3 platform, and simple application of speech recognition is carried out in combination with relevant theories of neural network. An effective way to improve the accuracy of speech recognition in the mode of local recognition and cloud storage is presented.

At the same time, there are still many applications in the research. These applications are very attractive, which shows the great potential of neural network applications. Neural network technology is particularly good at dealing with a large number of sample application problems. Problems can be solved by fully learning samples. Even tens of thousands of English words from phonemes to phonetic symbols can be analyzed and resolved well.

In this paper, a new method of processing inter-phonemic syllable labeling in English words by neural network is proposed. In fact, this method is instructive and can be applied to the main stress and secondary stress of each syllable in the word in the same way. When the deep neural network matures and is closely integrated with speech synthesis technology, just like how people have learned to speak, the development of speech technology will rise to a new level and will be human. The development of computer interaction and voice-related fields provides important forces and makes important contributions to the advancement of human science.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The author states that this article has no conflicts of interest.



## References

- [1] H. He, F. Gang, and C. Jinde, "Robust state estimation for uncertain neural networks with time-varying delay," *Journal of Jishou University (Natural Sciences Edition)*, vol. 19, no. 8, pp. 1329–1339, 2019.
- [2] Y. Dong and H. Wang, "Robust output feedback stabilization for uncertain discrete-time stochastic neural networks with time-varying delay," *Neural Processing Letters*, vol. 51, no. 1, pp. 83–103, 2020.
- [3] C. Hartmann, D. Opritescu, and W. Volk, "An artificial neural network approach for tool path generation in incremental sheet metal free-forming," *Journal of Intelligent Manufacturing*, vol. 30, no. 2, pp. 757–770, 2019.
- [4] S. Chan and A. H. Elsheikh, "Parametric generation of conditional geological realizations using generative neural networks," *Computational Geosciences*, vol. 23, no. 5, pp. 925–952, 2019.
- [5] A. N. Gorban, E. M. Mirkes, and I. Y. Tukin, "How deep should be the depth of convolutional neural networks: a backyard dog case study," *Cognitive Computation*, vol. 12, no. 1, pp. 388–397, 2020.
- [6] Y. Zhang, T. S. Lee, M. Li, F. Liu, and S. Tang, "Convolutional neural network models of V1 responses to complex patterns," *Journal of Computational Neuroscience*, vol. 46, no. 1, pp. 33–54, 2019.
- [7] S. Guo, B. Zhang, T. Yang, D. Lyu, and W. Gao, "Multitask convolutional neural network with information fusion for bearing fault diagnosis and localization," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 8005–8015, 2020.
- [8] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Integration of accelerated deep neural network into power transformer differential protection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 865–876, 2020.
- [9] D. Xie, Y. Jiang, and M. Han, "Global exponential synchronization of complex-valued neural networks with time delays via matrix measure method," *Neural Processing Letters*, vol. 49, no. 1, pp. 187–201, 2019.
- [10] Y. Zhang, J. Hou, V. Towhidlou, and M. R. Shikh-Bahaei, "A neural network prediction-based adaptive mode selection scheme in full-duplex cognitive networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 540–553, 2019.
- [11] J. Alvarez and A. Leon, "1D convolutional neural network for detecting ventricular heartbeats," *IEEE Latin America Transactions*, vol. 17, no. 12, pp. 1961–1968, 2019.
- [12] N. Ståhl, G. Mathiason, G. Falkman, and A. Karlsson, "Using recurrent neural networks with attention for detecting problematic slab shapes in steel rolling," *Applied Mathematical Modelling*, vol. 70, no. JUN, pp. 365–377, 2019.
- [13] L. Ren, Z. Liu, and Z. Zhou, "Determination for 235 U enrichments of same geometry and mass uranium components by BP neural network," *Hedongli Gongcheng/Nuclear Power Engineering*, vol. 40, no. 1, pp. 48–50, 2019.
- [14] S.-c. Ou, "The role of lexical stress in spoken English word recognition by listeners of English and Taiwan Mandarin," *Language and Linguistics/語言暨語言學*, vol. 20, no. 4, pp. 569–601, 2019.
- [15] A. Gottardo, N. Amin, A. Amin, R. Al-Janaideh, X. Chen, and J. Paradis, "Word reading in English and Arabic in children who are Syrian refugees," *Applied PsychoLinguistics*, vol. 41, no. 6, pp. 1305–1328, 2020.
- [16] N. Nway Han and A. Thida, "Annotated guidelines and building reference corpus for Myanmar-English word alignment," *International Journal on Natural Language Computing*, vol. 8, no. 4, pp. 25–38, 2019.
- [17] S. Lindstromberg, "Surplus interword phonological similarity in English multiword units," *Corpus Linguistics and Linguistic Theory*, vol. 16, no. 1, pp. 95–118, 2020.
- [18] A.-Z. Yen, H.-H. Huang, and H.-H. Chen, "Learning English-Chinese bilingual word representations from sentence-aligned parallel corpus," *Computer Speech & Language*, vol. 56, no. JUL, pp. 52–72, 2019.
- [19] H. Afzal and T. Mukhtar, "Semantically enhanced concept search of the Holy Quran: Qur'anic English WordNet," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 3953–3966, 2019.
- [20] J. Song, Jinsoo and K. H. Kyu-Hong Hwang, "Discourse-pragmatic functions of non-cannonical word order in English and their recognition by English native speakers," *The New Korean Journal of English Language & Literature*, vol. 61, no. 1, pp. 267–288, 2019.
- [21] K. Krenca, E. Segers, X. Chen, S. Shakory, J. Steele, and L. Verhoeven, "Phonological specificity relates to phonological awareness and reading ability in English-French bilingual children," *Reading and Writing*, vol. 33, no. 2, pp. 267–291, 2020.
- [22] M. Abdulmalik Ali, "Investigation of vocabulary learning strategies to identify word meanings for Saudi EFL students in reading context," *Arab World English Journal*, vol. 11, no. 3, pp. 149–169, 2020.
- [23] D. A. S. El-Dakhs and J. Altarriba, "How do emotion word type and valence influence language processing? The case of Arabic-English bilinguals," *Journal of Psycholinguistic Research*, vol. 48, no. 5, pp. 1063–1085, 2019.
- [24] P. ten Hacken, "Elisa Mattioli: analogy in word-formation: a study of English neologisms and occasionalisms," *Folia Linguistica*, vol. 53, no. 1, pp. 287–292, 2019.
- [25] K. Hesseltine and J. Davis, "The communicative function of adjective-noun order in English," *Word*, vol. 66, no. 3, pp. 166–193, 2020.