

Simulation of English Linguistic Phonetic and Lexical Variation Based on Sociology Perspective

Na Wang, M.Sc

Na Wang, Lecturer, Foreign Language Department, Zhoukou Normal University, Henan, Zhoukou 466001, China

Correspondence LecWang;wxunb@163.com

Objectives: From the perspective of sociology, the speech and vocabulary variation simulation of English linguistics is discussed in depth. Firstly, the background of the research object and the significance of the research are elaborated, and then the research theory related to the English phonetic and lexical variation simulation is analyzed. **Methods:** Through the design of the English phonetic intonation network teaching system, the design ideas that conform to the development of each function of the platform are proposed. **Results:** Furthermore, the English linguistic speech and lexical variation simulation model algorithm based on sociological perspective is used to design and verify the function of the teaching system, and the effectiveness of the algorithm is verified by empirical analysis. **Conclusion:** The final results of the experiment show that by using the Internet of Things (IoT) technology to develop a system tool that conforms to the teaching method and put it into specific teaching work can improve students' English linguistics pronunciation and vocabulary learning ability.

Keywords: IoT; English phonetic intonation; network teaching system; function; development

Tob Regul Sci.™ 2021;7(5-2): 4752-4762

DOI: doi.org/10.18001/TRS.7.5.2.40

In the process of teaching pronunciation in modern English, English teachers pay more attention to the explanation of pronunciation methods, and usually also demonstrate the pronunciation method on the spot, and ask students to practice together with the demonstration to improve students' speaking ability. In recent years, many educators have carried out the teaching mode of distance learning. In the process of distance learning, students can learn by themselves according to their own time and their own English pronunciation. This teaching mode breaks the traditional face-to-face teaching method, which can improve the pronunciation of English without leaving home, solve the difficulties of listening, speaking and reading in English learning, so that improve the overall learning level of students by making students comm

unicate more fluently with foreign friends. However, students who use distance learning can only objectively imitate, but cannot form their own characteristics to accurately grasp their own pronunciation. Abstract and difficult to think English pronunciation and tone of learning methods will make the atmosphere of learning English very boring, so it will make more students lose the pleasure of learning English, which leads to the decline of academic performance.

From the perspective of sociology, after collecting relevant experience and analyzing the social phenomena brought about by English linguistics speech and vocabulary variation simulation, a very effective structured method is explored. This method aims to study the influence of network teaching system on the simulation of speech and vocabulary variation, and summarizes the methods suitable for students to learn English. It uses experimental phonetics, phonology and

modern multimedia and network technology related theories to construct a network teaching system suitable for English majors' speech intonation teaching, so that the pronunciation and intonation can be embodied, visualized and dynamic, thus overcoming the blindness of speech imitation. Therefore, an improved particle filter algorithm is proposed for the English speech rationality recognition model, and the function development of the English phonetic intonation network teaching system under the background of IoT is discussed.

In order to meet the above-mentioned teaching needs of English phonetic intonation, the designed network teaching system has certain innovations, including front-end function modules and back-end function modules, with basic video learning functions, voice uploading functions and message feedback functions. At the same time, in order to stimulate students' interest in learning, the network teaching system is required to have a beautiful and simple human-machine interface, good query function, easy to maintain and operate. Through the online teaching system, students can express their own voices and feedback. Teachers can publish innovative video tutorials, lab guides, and feedback on speech tones.

The research is mainly divided into three main parts. The first part is the previous research summary, which lays a theoretical foundation for the research. The second part is the function design and algorithm of the network teaching system of English phonetic intonation. It mainly analyzes the multi-objective genetic algorithm based on embedded multimedia music player and the application of ACC audio file decoding in speech recognition system construction are analyzed. Finally, the method proposed is tested.

In recent years, the research on speech and vocabulary variation simulation is increasing. More scholars have done a lot of research on the connection between speech and lexical variation simulation by using sociological methods such as observation, empirical verification and statistical research. For example, Liu S of East China Normal University held a school on the English intonation competition as a research object, the method of empirical research proves that the

competition can improve the oral English results, thus improving the positive influence of English pronunciation teaching, and put forward some positive and useful suggestions for the teaching methods of oral English and pronunciation¹. Huang L et al. edited the English film as the basis for the training of English teaching, and compared the speech materials of the English major in the freshman year. Through more than three months of training and learning, it is concluded that the editing of English film can improve students' learning effects on the tone and phoneme of spoken English, improve students' skills in all aspects of spoken English, and provide a basis for students to choose learning English pronunciation and vocabulary learning materials². Yuan S et al. scholars believed that the human-machine voice dialogue technology was currently recognized as a highly technical field of technology. This technology included three functions: natural language understanding, speech recognition, and synthesis. The speech synthesis system could be used as a computer language output method to improve the educational outcomes of English spoken speech³. Wang F studied the English learning achievements of sophomores in various colleges and universities. It is found that students other than English majors lack knowledge and skills in English pronunciation and speaking, and cannot express their thoughts in English. Therefore, it is necessary to infiltrate voice teaching in English teaching with the support of policies, teachers and students⁴. Zhang J used the theory of language transfer to study the influence of Cantonese on English pronunciation and intonation. It was found that Cantonese and Cantonese tones had a positive impact on English communication and English intonation⁵. Gooch R et al. explored the problems in oral English teaching from the perspective of phonetics, and proposed the teaching strategies of oral English teaching in higher vocational colleges, which provided a reference for oral English teaching⁶. Huang CN et al. used poetry reading to train English pronunciation in the context of language context, so that learners can finally cross-cultural communication with correct pronunciation and intonation⁷. Li X research believed that accurate English phonetic intonation plays an important

role in improving learners' English listening and expression ability⁸.

In summary, with the development of information technology, designing an online teaching system to realize the teaching of English phonetic intonation can effectively help students improve their oral communication skills.

METHODS

Application of ACC Audio File Decoding in the Construction of Speech Recognition System

After the AAC audio data structure is parsed, the entire bit stream information and the configuration information required by each function module are obtained, so that the decoding of each of the following function modules can be performed. M/S coding uses matrix operation, which uses M channel (middle channel) and S channel (side channel) instead of the original L channel (left channel) and R channel (right channel). The M/S coding formula is as shown in (1):

$$\begin{bmatrix} M \\ S \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} L \\ R \end{bmatrix} \quad (1)$$

In the M/S matrix stereo mode, the normalized side channel value is transmitted. At the decoding end, the corresponding frequency domain data in the left and right channels needs to be added and subtracted to restore the original left and right channel data, so that the reconstruction of the left and right channels is obtained by matrix vector operation, and the M/S decoding formula is as shown in (2):

$$\begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} M \\ S \end{bmatrix} \quad (2)$$

L, R represent the left and right channel values after conversion, M represents the intermediate channel value, and S represents the side channel value. The matrix formula for ordinary channel pairs is shown in (3):

$$\begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} L \\ R \end{bmatrix} \quad (3)$$

When the frequency domain correlation of the left and right channel signals is high, the use of M/S coding can reduce the frequency domain energy of one channel, thereby reducing the number

of coded bits required. When the L and R channels are similar, the energy of the M channel is similar to that of the L channel, and the energy of the S channel is much smaller than the energy of the R channel; similarly, when the L and R channels are inversely similar, the energy of the S channel is similar to that of the R channel, and the energy of the M channel is much smaller than the energy of the L channel, so that the number of bits required can be reduced.

Multi-objective Genetic Algorithm Based on Embedded Multimedia Music Player

At the time of research, a very practical Linux system has been transplanted, and the experimental environment in which the Linux system is running is built. At the same time, various functions of the AAC decoder are activated by the corresponding software, and the preparation for analysis and research is completed. A variety of fully functional optimization utilities are available on Linux systems, and optimization utilities are open source. In order to divide the software and hardware, in the research topic, the complexity of the AAC software model needs to be analyzed. The method is to use an optimization utility called gprof in the Linux system. The biggest function of this optimization tool is to perform GNU performance measurement. The function of performance measurement is mainly to measure the time value of the code function segment in the running of the software, observe the result of the program in the running process, and based on this result, distinguishes the parts of the code that can be used as optimization hotspots, so performance measurement is also an indispensable part. The gprof tool works by including a mcount function in each application's function. In the process of running the program, the mcount function will be called continuously. In this way, a function call relationship can be formed by the function. The function of the gprof tool is to display the function call relationship, which includes the time of all functions and the number of calls, and the relationship between the functions. This clearly shows the flow of the program and finds the most time-consuming functions or operations in a program. The most time-consuming functions or operations in the system can be considered in

hardware for real-time purposes. Therefore, here, gprof is used to analyze the complexity of the AAC software model. The AAC test audio is ADTS format, the sampling rate is 48000 Hz, the bit rate is 128 kbps, and the audio duration is

17.387 seconds. The software decoding process is analyzed. The calculation results of each decoding module are shown in Table 1:

Table.1
AAC Software Model Complexity Analysis Results

Modular	Cycle number	Percentage (%)
IMDCT	70848589	65.9
Huffman Decode	22431734	20.9
M/S	1224026	1.1
PNS	168978	0.2
Intensity stereo	81387	0.1
Other	12751453	11.9
Total	107506167	100.0

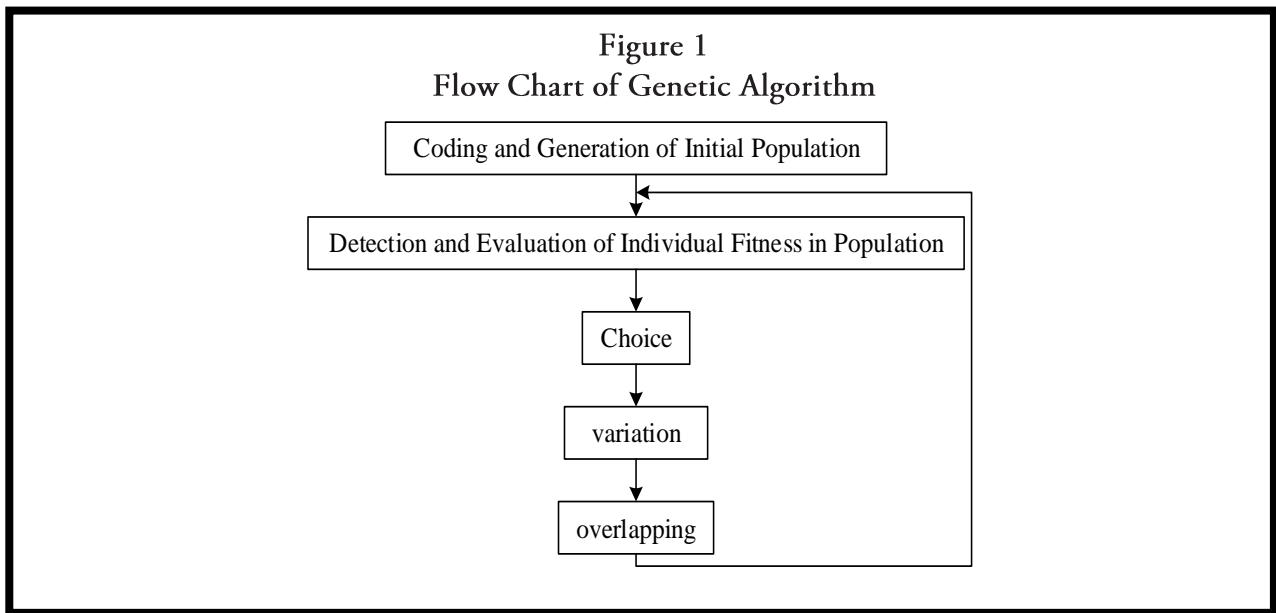
Using the gprof tool to output the measured results, we can see from the above results that the IMDCT module and the Huffman module are the highest in the various types of operations, and the values are 65.9% and 20.9%, respectively. It also reflects that the algorithms involved in these two modules are very complicated and very difficult to calculate. The calculation results are obtained by using a large number of data operations, which will occupy most of the decoder's running time, and the result will cause the AAC decoder to be unstable. With the very fast running speed of IP core hardware, these very special modules can be replaced with hardware. At the same time, these IP core modules can be reused. If similar calculations occur in the future design process, these IP core modules can be directly called, so that subsequent design development can be carried out more smoothly, which is conducive to saving time and improving development level. However, the software and hardware partitioning schemes obtained by analyzing the software complexity have certain limitations. This method considers the performance index of the system's execution speed. In fact, there are many factors that affect the system hardware and software partitioning. For example, the hardware implementation needs

to occupy the chip area, which will affect the system cost and power consumption. Therefore, in this topic, the application of multi-objective optimization genetic algorithm in hardware and software partitioning is further studied. The traditional multi-objective optimization algorithm has also been popularized in the past. It needs multiple optimizations in the running process to get the optimal solution set, but the optimization process is not related, and needs to be modified several times, so that various results of geology appear. Decision makers can't make optimal decisions from so many results, so the drawbacks of using traditional multi-objective optimization algorithms to calculate complex problems are more prominent; therefore, a new multi-objective optimization method based on genetic algorithm has emerged. The space of this method is relatively large, and several effective solutions can be obtained in the optimization process. It has a very good advantage in solving complex optimization problems, so this method is preferred when dealing with multi-objective problems. The multi-objective optimization genetic algorithm is suitable for solving the combinatorial optimization problem of software and hardware partitioning of this SOPC system. In the optimization process, multiple targets such as running time, hardware area, power

consumption and cost are optimized, and finally the software and hardware partitioning scheme of the SOPC system can be obtained.

In the specific SOPC system design, the main parameters include system running time, chip area, power consumption and cost. Obviously, the optimal requirements for system design are short run time, small hardware area, low power consumption, and low cost. In this topic, the optimization goals are set as: running time, hardware area, power consumption, and constraints are set to meet the predetermined system cost. The flow chart of the basic genetic algorithm is shown in Figure 1. The genetic algorithm must first be executed from a population that represents the possible potential solution set of the problem. When generating this initial population, it is necessary to realize

the mapping from phenotype to genotype, that is, the coding work. The application usually adopts binary coding. After the initial population is produced, according to the principle of survival of the fittest and the principle of survival of the fittest, evolution from generation to generation produces an increasingly better approximate solution. Each generation selects individuals according to the individual fitness size in the problem domain, and uses genetic operators to select, combine crosses and mutations, and continuously generate new populations. Through the whole process, the new population will be more adapted to the environment than the previous generation, and the optimal individual in the final last-generation population will be decoded as the approximate optimal solution to the problem.

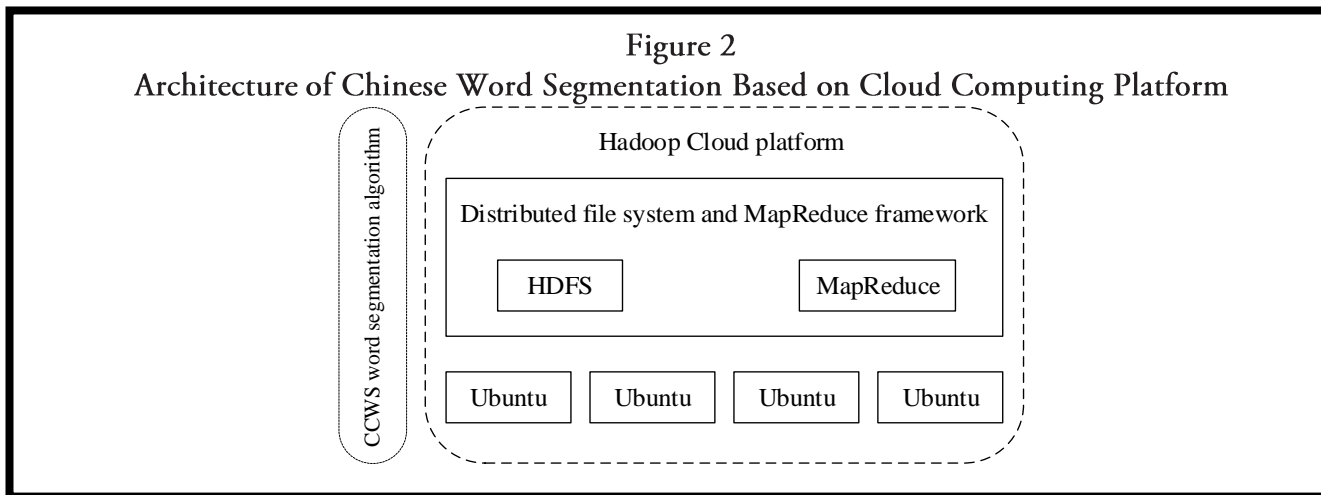


In this topic, the hardware and software implementation is divided. Therefore, there is only two possibilities for a node to be implemented. Therefore, the binary string is directly selected for encoding, and the chromosome is defined as $(k_1, k_2, k_3, \dots, k_n)$, $k_i \in \{1, 0\}$, $i = \{1, 2, \dots, n\}$. n is the number of system nodes, $k_i=1$ means that the node is implemented by hardware, and $k_i=0$ means that the node is implemented by software. The initial population is randomly generated, but

in the process of evolution, the continuously generated ones need to more and more satisfy the constraints of the system, and continuously remove the individuals who do not meet the constraints of the system, so as to consider the diversity of the algorithm while satisfying the requirements of the efficiency of the algorithm. The reasonable condition for the individual to generate randomly is that the hardware implementation cost is greater than the software implementation cost, and at the same time, the hardware implementation time is less than the

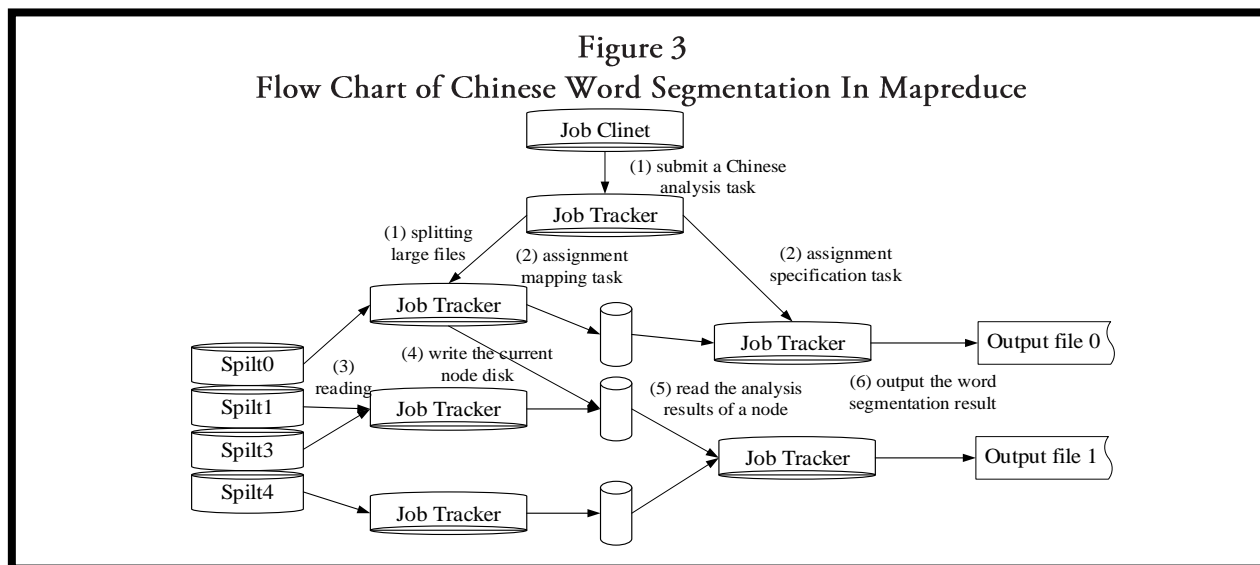
software implementation time. The population size can be set to twice the number n of nodes, and it is assumed that the population size remains constant during the process of evolution.

The CCWS English word segmentation algorithm is deployed on the Hadoop IoT platform. The system architecture diagram is shown in Figure 2.



In the CCWS Map Reduce process, the Map Reduce task is controlled by a Job Tracker and multiple Task Tracks: two types of nodes. The main job of Job Tracker is to schedule and manage the Task Tracker. It generally runs on the Name Node. After the Job Tracker assigns the Map task and the Reduce task to the free Task Tracker, the Task Tracker executes the

Tasks in parallel. At the same time, Job Tracker also supervises the operation of these Tasks. If a Task Tracker is abnormal, Job Tracker will assign this task to other suitable Task Trackers. The flow chart of the English word segmentation Map Based on the IoT is shown in Figure 3. The specific steps are as follows.



Submitting the word segmentation code to the Job Tracker in the Name Node, Job Tracker is in charge of the Map Reduce work, decides which files

are processed, and divides the large files into several blocks according to the actual situation; Job Tracker assigns different task tasks to the blocks in the first step, and assigns compute nodes

to each task. The specific tasks are independently executed by Task Tracker. Task Tracker reads file data from HDFS according to the Map function; in the Map stage, Task Tracker writes the output directly to the local disk and then tells Job Tracker where the results are. Job Tracker then informs the Task Tracker's Reduce task to go to the corresponding node to take these intermediate results; During the Reduce phase, Task Tracker: according to the location information of the Job Tracker: the corresponding node reads the Data and executes the reduce task; finally, the Task Tracker: writes the result of the:educate phase to the distributed file system.

RESULTS

Experimental Data Source and Processing

The entity, the video information, the audio information, the feedback information body, the experimental unit name, and the experimental guide are entities as the entity objects of the database system. The database has 6 entities corresponding to the table, the content of which contains user information, video tutorials, voice experiments, feedback information, experimental unit names, and experimental guides. The user information table is used to store user information; the feedback table is used to store information such as the results of the teacher's teaching; the experimental unit name table is used to store the name information corresponding to each experimental unit; the experiment guide table is used to store various storage paths; the video tutorial table is used to store related information of the video. The voice experiment table structure is used to store information about the audio file. Experiments are needed to demonstrate the effectiveness of particle filter for English linguistic speech and lexical variation simulation models. Under the basis of running MATLAB7.0, a reasonable simulation platform model related to English linguistics speech and vocabulary variation simulation should be built, and the pronunciation data collected from 10 college students and 3 English teachers should be used.

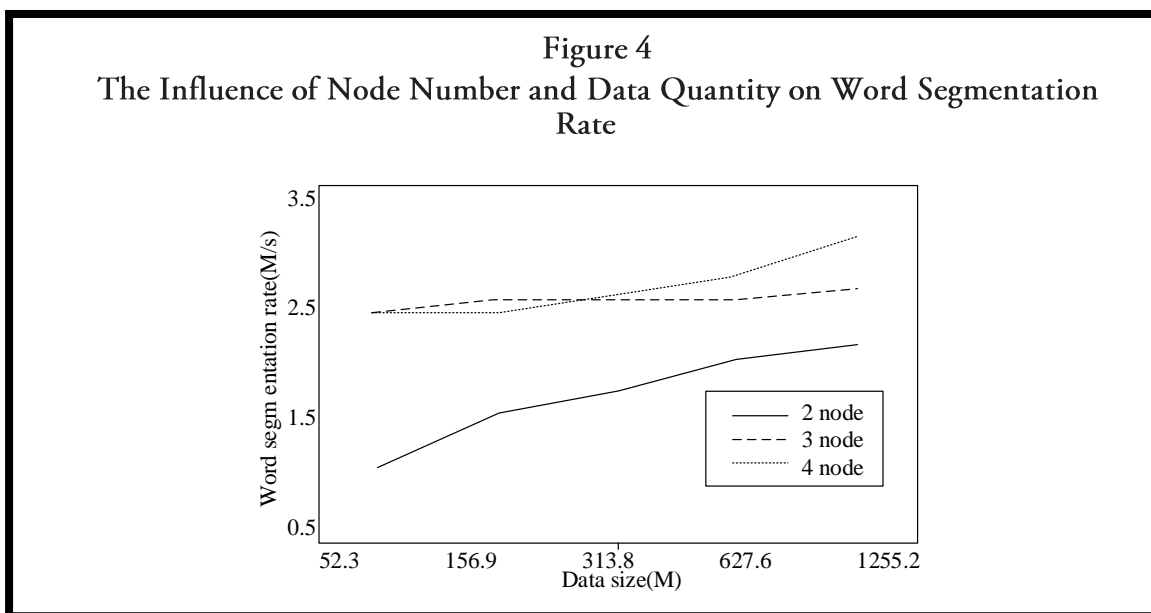
Performance Experiment and Result Analysis of English Speech Recognition Algorithm Based on Network Teaching System

For the original corpus experiment, we mainly focus on the selection of large primitive words. We select some long sentences with more speech samples from the corpus as test texts. Experiments have shown that in most cases, words as real words can generally be selected to the most ideal unit, while the chances such as the articles a and the and some high-frequency prepositions are selected to the most ideal unit are not very high, but the selected unit is similar to the ideal unit from the text environment and prosodic features, so it is very similar to the original speech in terms of auditory effect. Take the 12th sentence in the original corpus as an example: Following a problem solving formula used teams throughout federal express, members team monitored tomorrow routine, carefully noting where and when the work group's resources are used and where they are idle, waiting others upstream, the process to send packages their way. This sentence consists of a total of 47 words. There are 22 words in the sentence that are high-frequency words, and there are more than 10 samples of most of the words. In the sentence, 34 words are selected into very suitable units, and the other 13 words have no corresponding units with the original corpus, but from the perspective of synthetic effects, the synthesized speech has a similar synthetic effect to the original speech, and the tone of the pronunciation is reproduced, and the synthesis effect is also natural and close, thereby obtaining a higher unit selection efficiency. The platform implemented in this experiment is Hadoop's IoT platform, and the test experiments are carried out in strict standards of two, three and four nodes. The test data is divided into five types of continuous Chinese paragraphs, and the data sizes are 51.2M, 157.1M, 314.2M, 630.7M, and 1249.9M, respectively. Taking into account the impact of accidental factors will lead to errors in the experimental results, in order to make the experimental results closer to the true value, each test data was expanded to perform four tests, and then the average of the four results was used as the final result of the experiment. The following table is the time value of each data node running

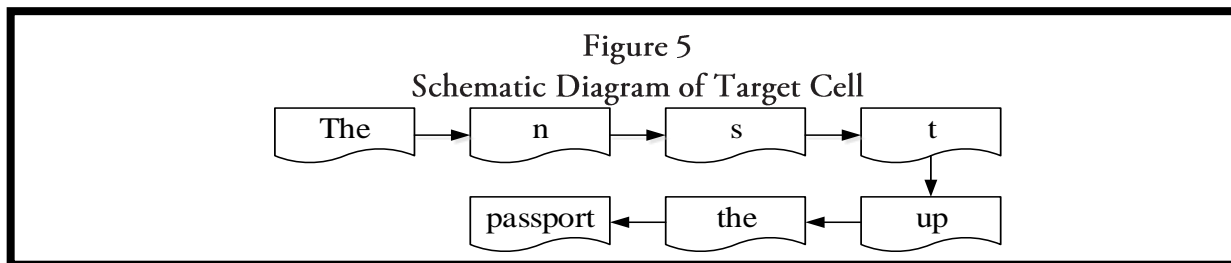
Table.2
The Word Segmentation Time on 2 DataNode

File size (M)	51.2	157.1	314.2	630.7	1249.9
Time 1 (s)	63	120	209	380	678
Time 2 (s)	69	121	231	352	673
Time 3 (s)	65	119	228	344	688
Time 4 (s)	64	125	210	341	684
Mean time (s)	65.25	121.25	219.50	354.25	680.75
Average rate (M/s)	0.78	1.30	1.43	1.78	1.84

In order to make the results more intuitive, some data is represented by the graph, and the results are as follows:



From the above experimental results, it can be found that many of the test texts extracted from the China Daily website can be spliced using word primitives. The excerpted text accounts for 90% or more of the total number of words. From this industry, the coverage of the corpus is relatively high. According to SPTE-O1: The luggage glanced up the black passport. The text analysis module is divided according to the unit sequence shown in Figure 5.



In particular, the splicing of the two words "luggage" and "glanced" requires the use of

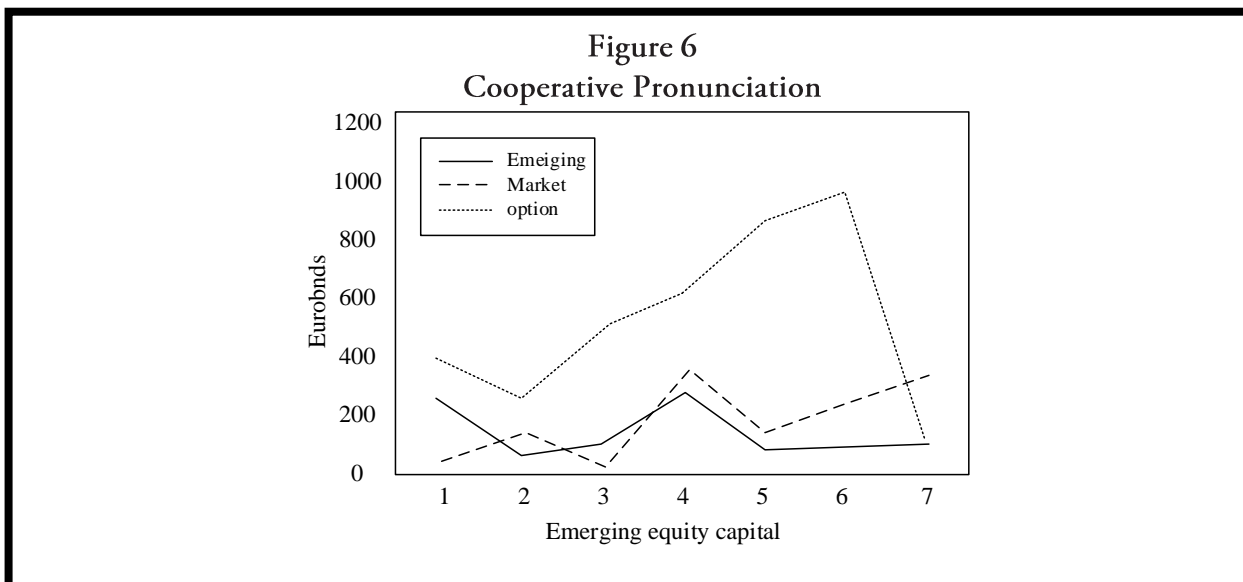
smaller primitives, while taking the units of Table 3 to select the units.

Table 3
Test Sample SPTE-O1 Small Unit Splicing Experimental Results

Target unit	Frequency	Synthesis Unit
Li	39	With luxury
gi	2	Lover
gl	11	Full of
n	5127	Plan to
s	2841	Will the items
t	2690	notice

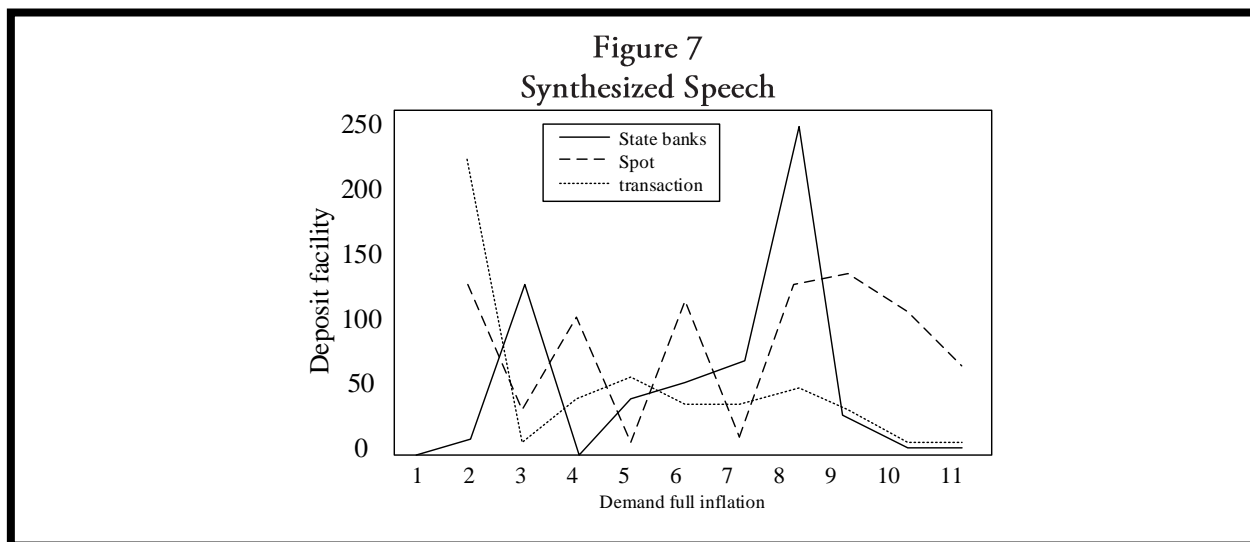
Based on the perspective of auditory effects, in this study, although small units are used in the splicing process, most of the cooperative pronunciation phenomenon will be covered inside the unit, and the sound quality loss at the splicing point is almost inaudible to the human ear, and the synthesized speech effect is relatively natural and close. At the same time, each splicing tone corresponds to a sample of a very large number of candidate units, and the selection speed is also very fast for the whole unit. If the

test text needs to use a small primitive in the splicing process, the synthesis effect is not very good, and there will be inconsistency and instability. The main reasons for these unhealthy phenomena are: after the text is divided, the principle of giving priority to large primitives is adopted in the process of text synthesis, and the number of candidate samples is relatively small, which causes the dissonance of synthesized speech and reduces the naturalness.



This situation is particularly prominent when using words to synthesize; when the primitives in the corpus cannot cover the strong cooperative pronunciation phenomenon and have to be spliced with smaller primitives, that is, when the strong co-sounding phenomenon occurs at the splicing point, the synthesized speech may have problems such as jumping and noise. For example, in the test sample TECE-O1 (see

Appendix A), the word foul, because there is no //faul/ in the syllable and the pre-syllable, it has to be stitched with /fau/ and //I/, and there is a jump in the synthesized speech. In the case of change, the sense of hearing is not natural enough; although the selected splicing primitives cover most of the strong cooperative pronunciation, the coordinated pronunciation between syllables is not controlled.



DISCUSSION

English linguistics speech and vocabulary variation simulation extends the method of English teaching. It is a very useful tool to assist teachi

ng. This method will not neglect or replace the traditional education teaching mode, but promote the development of teaching work. In order to better the teaching results, we must make full use of the good resources of the network teaching

system and excellent human-computer interaction performance. At the same time, two mutually connected and relatively independent teachers and student self-control systems are needed. When these systems actively correct the output information, the teaching dynamics can be balanced. Therefore, from the perspective of sociology, the speech and vocabulary variation simulation of English linguistics is discussed in depth. When the traditional recognition model of the pitch recognition is performed, the tone feature is used for recognition, and the speech that is affected by the accidental factor cannot be eliminated, and the modeling error is very large. A method of English linguistic speech and vocabulary variation simulation based on sociological perspective is proposed. The simulation results show that the recognition accuracy of the English speech rationality recognition is higher when using the optimization model, which provides a theoretical basis for the difference teaching of spoken English pronunciation. The development and design of the English phonetic intonation network teaching system has a very good effect. It can effectively feedback the learning outcome information, pays great attention to the development of students' ability in all aspects, and will improve the implementation of English pronunciation and intonation. However, there are still some shortcomings in the research results. The next step should be strengthened in terms of the comprehensiveness and robustness of the system functions.

Human Subjects Approval Statement

This paper did not include human subjects.

Conflict of Interest Disclosure Statement

None declared.

References

1. Gooch R, Saito K, Lyster R. Effects of recasts and prompts on L2 pronunciation development: Teaching English /r/ to Korean adult EFL learners. *System*, 2016,8(6):117-127.
2. Huang C N, Department F L. The Application of Poems Recitation in English Pronunciation Teaching in the Higher Colleges. *Journal of Hubei Correspondence University*, 2016,56(9):19-28.

3. Huang L, Liu Y. An Experimental Study of the Effect of Using Film Clips and Pronunciation Textbook Exercises in English Pronunciation Teaching. *Journal of Guangdong University of Foreign Studies*, 2016,70(89):12-34.
4. Li X. The Study of English Oral English Teaching in the Perspective of Phonetics. *Journal of Jiamusi Vocational Institute*, 2018,90(56):68-80.
5. Liu S. A practical study on the English pronunciation and intonation contest in college English pronunciation teaching. *Foreign Language Learning Theory & Practice*, 2016,8(12):90-111.
6. Wang F. On Penetration of Phonetic Teaching in College English Teaching for Non-English Majors. *Journal of Lianyungang Technical College*, 2016,34(10):35-54.
7. Yuan S, Yuhua J I. A case study on the dynamic development of college students' learning motivation in the English classroom. *Foreign Language World*, 2017,6(12):78-85.
8. Zhang J. Research on Influences of Cantonese on English Pronunciation and Intonation Study and Countermeasures from the Perspective of Language Transfer Theory. *Journal of Changsha University*, 2017,4(1):102-110.