*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

# Research of MCU Bilingual Teaching Based on Speech Recognition and English Sound Change

**Dongyuan Ge[1,*], Zhongliang Luo[2], Jian Li[1], Tuo Zhou[3], Qin Wang[1], Zhenfei Zhu[1], Wenjiang Xiang[4], Xifan Yao[5,*]**

[1] School of mechanical and Transportation Engineering, Guangxi University of Science and Technology, Liuzhou, China

[2] School of Information Engineering, Shaoguan University, Shaoguan, China

[3] Library, Guangxi University of Science and Technology, Liuzhou, China

[4] School of Mechanical and Energy Engineering, Shaoyang University, Shaoyang, China

[5] School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou, China

*Corresponding Authors.

***Abstract:***

Long short term memory LSTM) and VGGish are introduced for continuous English speech recognition system, which can recognize liaison, incomplete explosion, and voiceless consonant voicing in English oral. And in the process of constructing the model, the self-attention is introduced in alignment with the model. Then some phenomena of sound changes in English pronunciation were explored and studied, from which some enlightenments are obtained, that is, according to the mechanism of liaison, incomplete explosion, voiceless consonant voicing and the cognitive law of learning, and the teaching scheme of MCU course reforming is achieved and improved. According to the sound changes of English pronunciation, a novel online MCU hybrid bilingual teaching system is designed, which can not only meet the learning needs of students, but also enhance students' enthusiasm for exploring professional knowledge, and make the MCU course teaching emerge a new look.

***Keywords****: Bi-LSTM; VGGish; Speech Recognition; MCU Bilingual Teaching; pronunciation mechanism.*

## I. INTRODUCTION

The form of traditional face-to-face classroom teaching is dull, and prone to being limited by time and space. With the development of computer network technology, Internet has been integrated in education gradually, and various online live classes have begun to enter in teaching. For example, according to the actual teaching situation of the MCU course in China universities, a flipped classroom teaching mode

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

based on virtual simulation and spiral progress is explored [1]. The development of machine learning technology has promoted continuous progress in many fields, and in particular big data processing ability is constantly improving. The study of human's thinking and language's evolution may be complicated, but there are also certain rules to be explored, and from which we can draw some lessons about the development progress and formation mechanism of human languages[2]. For example, some enlightenments can be obtained from the phonetic change phenomenon of English pronunciation, so as to reform the teaching of some professional courses. We can judge a person's behavior patterns from his learning and expression methods, and then by modeling his behavior patterns, so as to improve and modify related models in a targeted manner, which is one of an important basis and content for the reform of educational teaching. On the other hand, in the process of conducting, with the user's existing learning data, we can use convolutional neural networks and long short-term memory as well as continuous speech recognition technology to develop a new type of bilingual online teaching system for MCU, so the users' learning contents of interesting or weak parts are pushed to the users, and the system can also intelligently plan the user's learning process, conduct students to learn step by step, so as to further improve students' learning efficiency.

## II. RELATED WORK ON SPEECH RECOGNITION AND BILINGUAL TEACHING

With the development of research on speech recognition technology being more and more in-depth, many scientists developed a lot of significant approaches for speech signal recognition with Markov model, among which the recognition of continuous speech has become the current research hotpot[3,4]. A novel Markov model with a state space being linear(not exponential) in the number of sources is designed, which can recognize speech from recordings of a priori known speakers' simultaneously speaking[5]. Petridis et al. presented an end-to-end speech recognition system based on Long-Short Term Memory (LSTM) networks consisted of two streams, which extract features directly from the mouth and different images respectively, and the model simultaneously learns to extract features directly from the pixels and performs classification as well as achieves state-of-the-art performance in visual speech classification[6]. Han, Kang and Mao et al proposed a load-balance-aware pruning method that can compress the LSTM model size by 20 times with negligible loss of prediction accuracy. At the same time a scheduler that encodes and partitions the compressed model to multiple PEs for parallelism and a hardware architecture named ESE that works directly on the sparse LSTM model are designed, and the proposed technology speeds up prediction and makes its energy efficient[7]. Many laboratories and companies are also studying the methodology of the Markov model. For example, Carnegie Mellon University has developed a continuous speech recognition system, which is also the world's first speech recognition system. The research on speech recognition system in China only started in the last few decades, but its development is very fast, and its applications in engineering practices and daily life is very extensive and remarkable[8, 9]. Graves, Mohamed and Hinton investigated deep recurrent neural networks in 2013, which combine the multiple levels of representation that have proved effective in deep networks with the flexible use of

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

long range context that empowers RNNs. Through trained end-to-end and suitable regularization, the designed deep recurrent neural networks achieve a test set error of 17.7% on the TIMIT phoneme recognition benchmark[10].

The speech communication channel is considered as one of the most important modality to benefit the blind and low vision persons, a robust blind digital speech watermarking technique has been proposed for online speaker recognition systems based on discrete Wavelet Packet Transform, which is a significant work in the state-of-the-art the accessibility field[11, 12]. In order to explore the extent to which different instantaneous frequencies due to the presence of formants and harmonics in the speech signal may predict a speaker's identity, a novel parameterization of speech that is based on the AM-FM representation of the speech signal is presented, which can assess the utility of these features in the context of speaker identification[13]. Researchers have applied machine learning in the field of speech recognition, which can promote the development of speech recognition technology[14, 15]. By jointing optimization of the deep learning models (deep neural networks and recurrent neural networks) with an extra masking layer, the designed system can enforce a reconstruction constraint, and adopted discriminative training criterion can further enhance the separation performance[16]. Based on cloud computing and artificial intelligence (AI), the MOOC platform for college English cross-cultural teaching is carried out, where the teaching content in the classroom be compressed, but the overall teaching content don't be cut down[17]. With MOOCs students can sort out their learning knowledge in the outside of class, which can not only play the leading role of teachers, but also improve students' learning efficiency and teaching quality of class. Based on perception psychology, the bilingual teaching model is introduced in the class of MCU course, which is focused on the internal structure, instruction, comprehensive engineering design set and so on [18]. On the other hand, there are researches on contributions of oral language in young bilingual students' English reading outcomes being investigated, which is an under-explored topic[19]. Applications of big data and AI in education have made significant headway which demonstrates a novel trend in leading-edge educational research. The convenience and embeddedness of data collection for educational technologies, paired with computational techniques all made the analyses of big data possible [20]. Although the scope of Internet and AI has become more and more extensive, but their application in the education is still relatively primary or even weak, and the integration of teaching resource with advanced technology falls behind compared with the AI in engineering. During our exploration process of teaching practice, by adopting Internet and machine learning as well as integrating various accumulation teaching resources of curriculum, an online hybrid MCU bilingual teaching platform is built, which not only promotes the diversification of education models, but also improves the smartness of education and teaching, and enhances the effect of learning; not only improve students' self-study ability, but also give full play to students' subjective initiative and improve their learning efficiency. According to the obtained data of students behavioral and learning results and so on, the system is trained with big data technology. The relevant decision-making algorithm is used to model the data, and then the established model is further judged and analyzed. This model can

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

contribute to optimize and improve students' learning behaviors and learning outcomes, and also provides a basis for subsequent further teaching research.

## Ⅲ. NOVEL ENGLISH CONTINUOUS SPEECH RECOGNITION FOR TEACHING SYSTEM DESIGN

### 3.1 Structure of Bi-direction Long Short Term Memory

The structure of bi-direction long short term memory is shown as Figure 1, which is with two separate hidden layers by processing the data in both directions, which are then fed forwards to the same output layer. As illustrated in Fig. 1, the networks computes the forward hidden sequence $\vec{\mathbf{h}}$, the backward hidden sequence $\overleftarrow{\mathbf{h}}$, the output sequence $\boldsymbol{y}$ by iterating the backward layer from $t = T$ to 1, and the forward layer from $t = 1$ to $T$, and then updating the output layer.

$$\vec{\mathbf{h}}_t = \mathrm{H}\left(\mathbf{W}_{\vec{x}\vec{h}}\mathbf{x}_t + \mathbf{W}_{\vec{h}\vec{h}}\vec{\mathbf{h}}_{t-1} + \mathbf{b}_{\vec{h}}\right) \tag{1}$$

$$\overleftarrow{\mathbf{h}}_t = H\left(\mathbf{W}_{\overleftarrow{x}\overleftarrow{h}}\mathbf{x}_t + \mathbf{W}_{\overleftarrow{h}\overleftarrow{h}}\mathbf{h}_{t-1} + \mathbf{b}_{\overleftarrow{h}}\right) \tag{2}$$

$$\mathbf{y}_t = \mathbf{W}_{\vec{h}y}\vec{\mathbf{h}}_t + \mathbf{W}_{\overleftarrow{h}y}\overleftarrow{\mathbf{h}}_t + \mathbf{b}_o \tag{3}$$

At the same time we concatenate the forward and backward states to obtain the annotations ($h_1$, $h_2$, $\cdots$, $h_{Tx}$) as follows,

$$\mathbf{h}_i = [\vec{\mathbf{h}}_t, \quad \overleftarrow{\mathbf{h}}_t]^{\mathrm{T}} \tag{4}$$

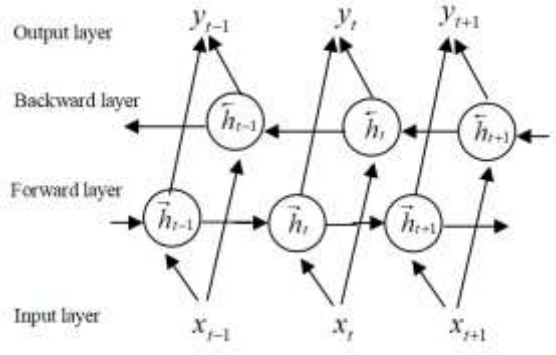



Figure 1. Bidirectional Recurrent Neural Network

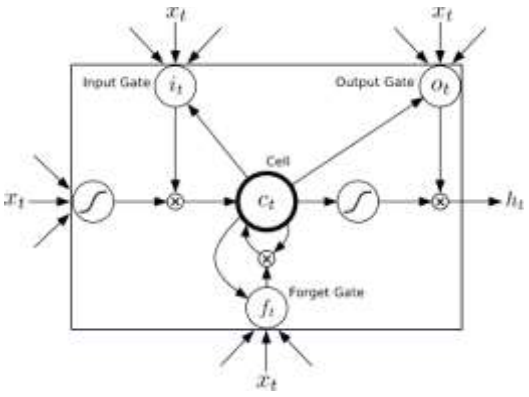



Figure 2. Long Short-term Memory Cell

The LSTM is used to purpose-built memory cells and store information [21], for better finding and

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

exploiting long range context, whose structure is shown in. Fig.2, a version used in this paper.

$\boldsymbol{h_t}$ is implemented by the following composite function[22],

$$\mathbf{i}_t = \sigma(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{W}_{ci}\mathbf{C}_{t-1} + \mathbf{b}_i) \quad (5)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{W}_{cf}\mathbf{c}_{t-1} + \mathbf{b}_f) \quad (6)$$

$$\mathbf{c}_t = \mathbf{f}_t\mathbf{c}_{t-1} + \mathbf{i}_t.\tanh(\mathbf{W}_{xc}\mathbf{x}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c) \quad (7)$$

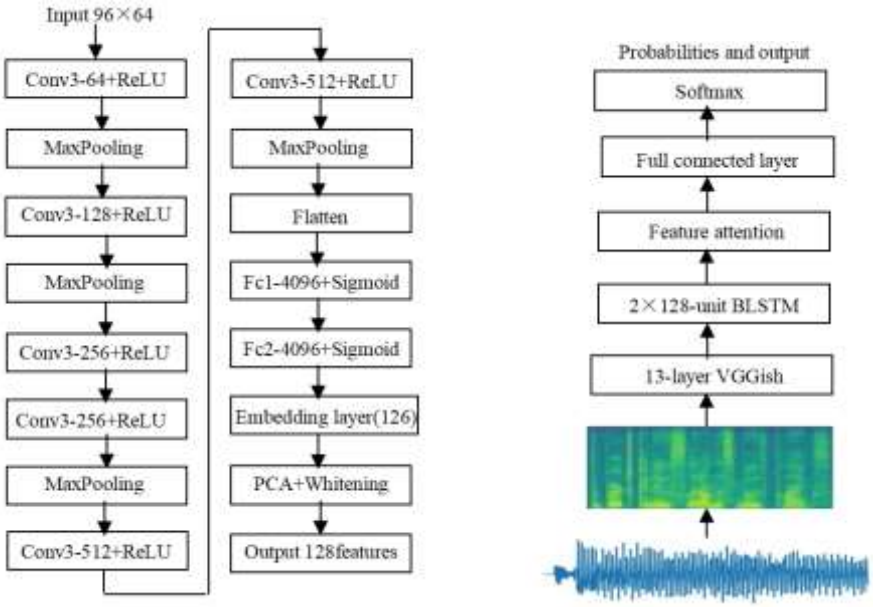
$$\mathbf{o}_t = \sigma(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{W}_{co}\mathbf{C}_t + \mathbf{b}_o) \quad (8)$$

$$\mathbf{h}_t = \mathbf{o}_t.\tanh(\mathbf{c}_t) \quad (9)$$

where ***i, f, o*** and ***c*** are respectively the *input gate*, *forget gate*, *output gate* and *cell* activation vectors, all of which are the same size as the hidden vector *h*. The weight matrix $\mathbf{W}_{xi}$, $\mathbf{W}_{hi}$,…, $\mathbf{W}_{xo}$, are the input-input gate, the hidden-input gate,…, the input-output gate matrix etc. The weight matrices from the cell to gate vectors (e.g. $W_{ci}$) are diagonal, so element *m* in each gate vector only receives input from element *m* of the cell vector. The $\boldsymbol{b}_i$, $\boldsymbol{b}_f$, $\boldsymbol{b}_c$ and $\boldsymbol{b}_o$ are the bias terms. $\sigma(.)$ is the logistic sigmoid function.

3.2 VGGish model and speech recognition model

VGGish model is designed as Fig.3[23] and Speech recognition model is designed as Fig.4.



*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

Figure 3 VGGish Model and its later processing

Figure 4 Speech recognition model based on Bi-LTSM and VGGish

As can be seen in the Figure 3, Relu function is written as follows,

$$\mathbf{s} = \max(\mathbf{0}, \mathbf{r}) \tag{10}$$

Sigmoid function is written as follows,

$$f(\mathbf{x}) = \frac{1}{1+e^{-\mathbf{x}}} \tag{11}$$

During the experiment, VGGish model is used to extract embedding features from speech data, and carries out cluster visualization.

At the same times, attention mechanism is introduced. The essence of this mechanism is to introduce the context information and location information of the input corresponding to the current prediction.

The context vector $c_i$ is computed as a weighted sum of these annotations $h_i$,

$$c_i = \sum_{j=1}^{T_x} a_{ij} h_j \tag{12}$$

The weight $\alpha_{ij}$ of each annotation $h_j$ is computed according to the softmax function as follows [24, 25],

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})} \tag{13}$$

where $e_{ij} = a(s_{i-1}, h_j) = v_a^T \tanh(w_a s_{i-1} + U_a h_j)$, is an alignment model which scores how well the inputs around position $j$ and the output at position $i$ match. The score is based on the LSTM's hidden state $s_{i-1}$ and the $j^{\text{th}}$ annotation $h_j$ of the input sentence.

3.3 Experiment of speech recognition

While the system is trained for continuous English speech recognition, the TIMIT corpus is adopted. First, we extract the some feature signals according to the corresponding method, and then integrate them into a matrix as training data. We divide all the data into binary for avoiding errors caused by too large differences in the data magnitude. The adopted technology based on BiLSTM-VGGish is used to carry out simulation experiments. From the experimental results, we can see that the accuracy of using neural network algorithm for classification can reach about 95%, and the proposed algorithm can also performs well in the field of signal classification.

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

## IV. IMPROVEMENT OF MCU BILINGUAL TEACHING BY ENLIGHTENED FROM SPEECH RECOGNITION

### 4.1 Analysis of some phenomena in English pronunciation

The phenomenon of sound change of English speaking originates from an economy principle, which can make our lips and tongue in a relaxed state when we speak. By exploring the mechanism of liaison, incomplete explosion and voiceless consonant voicing and so on, so as to obtain some enlightenments for bilingual teaching of MCU. First we explore the mechanism of liaison, incomplete explosion and voiceless consonant voicing in English speaking.

In the same sense group, if the first word ends with a consonant and the second word begins with a vowel, when we speak or read the sentence, it is natural and customary to spell the two phonemes together. This kind of phonetic phenomenon is called liaison. In generally, a sentence with liaison means does not need to be stressed. For example, the sentences “Not at all” and “Nice to meet you”, if we take liaison means it would sound smoother.

The incomplete explosion is a phonetic phenomena too. If the end letter of the previous word and the beginning of the following word are two or the same explosions (such as/ p/ / b/ / t / / d / / k / / g /), then the pronunciation of first phonetic symbol only retains mouth shape and time, but does not pronounce and with a slight pause, then the following phonetic symbol pronounces later. In fact the speech recognition system or listeners can feel the sound. For example, in the sentence “You make me wanna call you in the middle of the night”, the phonetic symbol /k/ in the word “make”, doesn’t pronounce by adopting the means of incomplete explosion.

Voiceless consonant voicing is another kind of pronunciation phenomenon, but does not have a pronunciation rule. For example, if the phonetic symbols after [s] are voiceless consonant [t], [k] or [p], in general, they would pronounce the voiced consonant [d], [g] and [b] respectively.

The above phonetic changes of English pronunciation are the evolution in the process of language developing for convenience, which are pronunciation phenomenon formed naturally in a long time. People who speak English as their mother tongue or as secondly official language all can recognize the above phonetic changes in communication; at the same time, with the artificial intelligence, the existing speech recognition systems can basically distinguish the phonetic changes of English speaking, such as liaison, incomplete explosion and voiceless consonant voicing. There are many research works done in the fields, for example, by exploiting the influence of different emotions on the prosody parameters, and emotion conversion methods are employed to generate the word level non-uniform prosody modified speech, where the modification factors for prosodic components such as pitch, duration and energy are used, so as to improve the performance of automatic speech recognition systems [26]. By analyzing and

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

studying this phonetic change phenomenon, we find that students' learning of professional knowledge has a similar mechanism with it[27, 28]. By introduction of the phonetic change mechanism of the English pronunciation into the teaching system, the teaching of the MCU course takes on a new look.

## 4.2 Pronunciation phenomenon and MCU course Teaching

**English liaison and teaching planning**

SCM is a widely used component in engineering. With the rapid development of MCU technology, a chip of MCU is equivalent to a neuron, a device made of which can run some simple neural network models, and are applied in engineering practices. By adopting MCU (Renesas RX65N) raw data are transferred into images with short-time Fourier transform, and then feature extracting is carried for classifying objects with CNN algorithm[29]. There are many researches on intelligent teaching, for example, according to the AI advance technology, neural network is adopted for teaching quality evaluation of MCU course teaching, which provided a novel programme for smart teaching[30]. The knowledge points of each chapter of MCU are not independent, and have strong coupling. For example, the timers/counters, interrupt technology, serial port communication and so on, are closely related to the hardware structure of MCU, and the realization of their functions needs the MCU's instruction system. So the teaching arrangement for this knowledge points has many forms. In general teaching plans have the following several schemes: 1. Computer hardware structure, timers/counters, interrupt, serial communication, instruction system, and system expansion (including AD conversion, etc.). 2. Computer hardware structure, instruction system, timers/counters, interrupt, serial communication, and system expansion (including AD conversion, etc.). 3. Computer hardware structure, instruction system, interrupt, timers/counters, serial communication, and system expansion (including AD conversion, etc.). These teaching plans and arrangements have their own grounds. The timers/counters, interrupt, serial communication, including the following AD conversion and other knowledge points, are related to interrupt technology, which is also the difficulty and key content of MCU course. On the other hand, the interrupt technology is also the clue of the course and can string up the main knowledge points of MCU course. If the interrupt technology is taken as a linking clue for each knowledge point among the timers/counters, interrupt operation, serial communication, and system expansion (including AD conversion, etc.), such doing will help students to master the relevant part of the content better. According to many years of teaching practice and research on MCU, and drawing a salutary lesson from the phenomenon of liaison in English pronunciation, we adopt the second scheme, that is, the chapter of timers/counters is arranged before the interrupt technology, the serial communication is arranged after the interrupt technology, and the system expansion is scheduled at the end. This arrangement is more consistent with the mechanism of liaison in English pronunciation, which can make the teaching more smoothly among timers/counters, interrupt technology, serial communication, and system expansion. By the scheme, classroom teaching of MCU is more fluent and in line with characteristic of economy, so the students also feel more facilitating and economy. At the same time, we found that we can save one

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

class-hour for the novel teaching scheme. According to our teaching practices and English liaison pronunciation phenomenon and mechanism, the teaching arrangement of MCU course is shown in Figure 5.

In Figure 5, the "Anti-incomplete explosion" mean that the parts of interrupt technology in serial ports, ADC and keypad are overlooked in class after the content of interrupt handing are taught, and the [.1] in "activity layer1 " and [.2] in "Activity layer 2" mean their activity functions respectively. In fact, liaison of English pronunciation is an instinctive evolution of the language's development, and students' learning of corresponding professional knowledge has a similar mechanism with language acquisition. Because the chapters or classes of MCU are coherent to form a strong coupling knowledge chain, seen from the cognitive aspect, the arrangement in Figure 5, makes the course's knowledge points among the four chapters more similar to the structure of liaison between consonant and vowel, so that the teaching of each knowledge point is smooth and economical, and students can efficiently learn, digest, review and consolidate the corresponding knowledge points through this teaching mode.

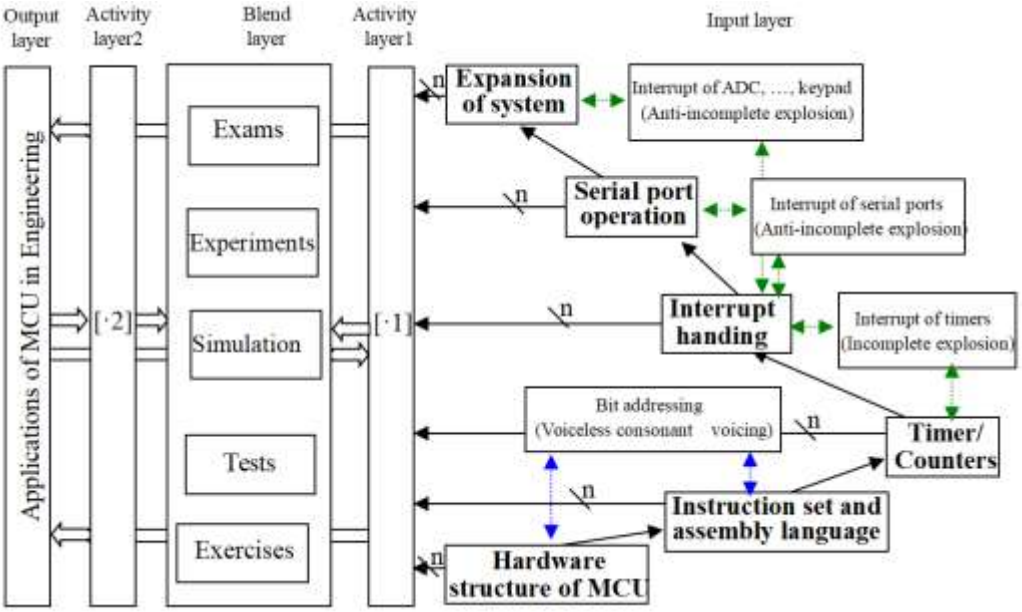



Figure 5 The Teaching Scheme of MCU Bilingual Teaching

4.3 Shortcomings and countermeasures of the designed system for bilingual teaching of MCU

4.3.1 Insufficient practical application

The application of machine learning such as VGGish and LSTM can promote the development of continuous speech recognition technology. It also has many shortcomings. It is easy for us to use some modern advanced methods for the teaching process to make everyone pay more and more attention to the teaching process and despise learning itself. At present, most of the education application researchers are

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

computer professionals. These researchers may not be rigorous enough in their grasp of pedagogy and psychology, and ignoring the laws of teaching. Besides teaching professional knowledge in English environment, another important factor of the bilingual teaching is that the teaching process should be consistent with the law of human cognitive development, or the development history and mechanism of professional knowledge.

4.3.2 Difficult knowledge / key knowledge and incompletely explosion

MCU interrupt technology is the difficulty and key point in this course. The interrupt technology is involved in the timers/counters, serial port communication, ADC, keyboard response and other knowledge points. When the basic principle of timers/counters is taught in class, because of the interrupt technology involved, according to the incomplete explosive voice in English pronunciation, just make a brief introduction for it, i.e. no detail explain it for students, then turn to the chapter of interruption technology, which is similar with the phonetic phenomenon of incomplete explosion in English pronunciation. After the main content of the next chapter i.e. interrupt technology knowledge point is completed, conversely the relating interrupt technology is added as complement of the timers/counters, so that the two knowledge points are embedded into each other's content. As for the following chapters such as serial port communication and A/D conversion, besides their basic principles and conventional query methods, interrupt technology is also involved. In the process of teaching, we also learn from the mechanism of incomplete explosion in English pronunciation, which is not taught in detail, but a little time is reserved in the classroom for students to learn and digest by themselves, and teachers will answer students' questions when they encounter difficulties during the studying. This method is called anti-incomplete exposition in our novel teaching practice. We found that this method can promote students' interest in the learning the key content of MCU, and mobilize students' learning initiative. With this scheme students' knowledge gained through their own exploration is more thorough and comprehensive. In fact the students' participation degree is more adequate. At the same time, 1.5 class hours are saved compared with the previous teaching mode. In addition, it also makes the teaching process relaxed and lively, and the classes can void the dilemma of full classroom cramming teaching.

4.3.3. Bit addressing and voicing

In MCS-51, the bit addressing looks similar to byte addressing, and its associated bit operation is not the key content or difficult content, so students are prone to confuse the knowledge point because of students' thinking habits of before or unfamiliar with the new characteristics of MCU. Thus students are easily ignored this convenience advantage of MCU, which gives rise to the students often fail to flexibly use the function of the bit operation in engineering practice. Therefore, while the knowledge points of the bit operation are taught, in the light of the mechanism of voiceless consonant voicing in English pronunciation, for example, the consonants / p /, /k/, /t/, /t/ / in the words "speak", "sky", "stuff", "water", are voiced as /b/, /g/, /d/, /d/ respectively, and general speech recognition system or English communicators can distinguish the speech phenomenon. According to the teaching practice and research,

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

in the teaching scheme, the internal structure of the MCU the course only gives a rough introduction for the part i.e. the bit operation, which is taken as a cushion, and in the part of the addressing module, which is overlooked in the class, that is as the transition. Till the sections of the bit operation instructions, according to the mechanism of voicing consonant voicing, this part is taken as the key content so as to help students understand the difference between bit addressing and byte addressing. This method is equivalent to the phenomenon of voiceless consonant voicing in English pronunciation. By the new scheme it takes more 0.5 teaching hour than former teaching programme in the class, but the teaching effect is better.

4.3.4 Implementation Strategy

In the process of bilingual teaching of MCU, we should be able to simulate the real teaching environment so that students and machines can interact. At the same time, the system should also pay attention to teachers' guiding ability and cultivate students' self-learning ability.

The teachers of bilingual teaching should combine the knowledge of education and teaching with the development of new technologies to improve their own teaching methods continuously, and to improve themselves professional vision. At the same time, the class's teaching schemes should also be improved and adjusted according to the cognitive law, teaching law and actual conditions self-adaptively.

The established online MCU bilingual teaching platform should keep openness and developers and professional teachers keep co-operating, so that the functions of the platform can be changed according to user's needs correspondingly, and become more and more in line with the actual application requirements. This research is mainly aimed at MCU bilingual teaching, and come to a good teaching effect, but parts link and content, such as the mechanism of English liaison phonetic phenomenon is not suitable for those courses with weak coupling of chapters, so it is difficult to transplant or extend to this kind of courses.

## V. CONCLUSIONS

This paper designs and implements a continuous English speech recognition system based on Bi-LSTM and convolutional neural networks. At the same time, according to the mechanism of sound change of English speaking and the relationship and characteristics between professional knowledge of class, by combining internet technology and multimedia technology a novel MCU bilingual teaching program is structured for smart online teaching platform. By this novel teaching method, we can not only save 2 class hours for the course, but also complete the scheduled teaching task with good teaching effect. What's more, it is particularly important for the current teaching situation of professional courses with heavy teaching tasks and less teaching-hours.

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

## ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China under grant Nos. 81960332, 51765007, 51675186, and the Undergraduate Teaching Reform Project of Guangxi Higher Education under grant No 2018JGA204.

## REFERENCES

[1] Ge D. Y., Yao X. F., Liang M. A., Liu E. C., Xie G. J. (2018). Research on Flipped Classroom for MCU Based on Proteus and Spiral Progress, Journal of Electrical & Electronic Education, 40(3): 57-60.
[2] Wang H., Guo B., Hao S. Y, et al. (2020). Personalized dialogue content generation based on deep learning, Journal of Graphics, 41(2), pp. 210-216.
[3] Lee K. S. (2008). EMG-based speech recognition using hidden markov models with global control variables. IEEE Transactions on Biomedical Engineering, Vol. 55, No. 3. pp.930-40.
[4] Yang W. X., Tang S. Y., Li M. Q., Zhou B. B., Jiang Y. J. (2018). Markov bidirectional transfer matrix for detecting LSB speech steganography with low embedding rates. Multimedia Tools and Applications, Vol.77, No.14. pp. 17937-17952.
[5] Smaragdis P., Raj B. (2012). The Markov selection model for concurrent speech recognition, Neurocomputing, Vol. 80, pp 64-72.
[6] Petridis S., Li Z., & Pantic M. (2017). End-To-End Visual Speech Recognition With LSTMs. ICASSP 2017 - 2017 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE. pp.2592-2596.
[7] Han S., Kang J. L., Mao H. Z., et al. (2017). ESE: Efficient Speech Recognition Engine with Sparse LSTM on FPGA. the 2017 ACM/SIGDA International Symposium. Pp.1-10.
[8] Yuan L. C. (2008). A speech recognition method based on improved hidden Markov model. J. Cent. South Univ. (Science and Technology), Vol.39 No.6, pp.1303-1308.
[9] Xu C. D., Xia R. S., Ying D. W., Li J. F. (2014). Time-frequency speech presence probability estimation based on sequential hidden Markov model for speech enhancement. Acta Acustica, 39(05), pp. 647-654.
[10] Graves A., Mohamed A. R., and Hinton G.. (2013). Speech recognition with deep recurrent neural networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6645-6649.
[11] Freitas, D., Kouroupetroglou, G. (2008). Speech technologies for blind and low vision persons. Technology and Disability, 20(2), 135–156.
[12] Nematollahi M. A., Rosales H. G., Akhaee M. A., & Al-Haddad S. A. R. (2015). Robust digital speech watermarking for online speaker recognition. Mathematical Problems in Engineering. https://doi-org.proxy2.cl.msu.edu/10.1155/2015/372398.
[13] Grimaldi M., Cummins F. (2008). Speaker identification using instantaneous frequencies. IEEE Transactions on audio, Speech, and Language Processing, 16(6), 1097–1111.
[14] Saifan R. R., Dweik W., Abdel-Majeed M. (2018).A machine learning based deaf assistance digital system. Computer Applications in Engineering Education, Vol.26, No. 4, pp. 1008-1019.
[15] Fayek H. M., Cavedon L., Wu H. R. (2020).Progressive learning: A deep learning framework for continual learning. Neural Networks, Vol. 128, pp.345-357.

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*September-October 2021 Page No. 814-826*
*Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021*

[16] Huang P. S., Kim M., Hasegawa-Johnson, M., & Smaragdis, P. (2014). Deep learning for monaural speech separation. In 2014 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 1562–1566). IEEE.
[17] Xie H. Y., Mai Q., Anand P., et al. (2021).College English cross-cultural teaching based on cloud computing MOOC platform and artificial intelligence, Journal of Intelligent & Fuzzy Systems, Vol. 40, No.4. pp. 7335-7345.
[18] Chen J., Jiang M., Lang L., Wang G. L.. (2009). The Teaching Research of Bilingual Education and Comprehensive Engineering Design of MCU, Journal of Electrical & Electronic Education, 31(2), pp.110-111.
[19] Huang B. H., Bedore L. M., Niu L. P., Wang Y. T., et. al. (2021). The contributions of oral language to English reading outcomes among young bilingual students in the United States, International Journal of BilingualismVol. 25, No.1, 40-57.
[20] Luan H., Geczy P., Lai H., et al. (2020) Challenges and Future Directions of Big Data and Artificial Intelligence in Education. Frontiers in Psychology, Vol 11. pp.580820-580820.
[21] Hochreiter S, Schmidhuber J. (1997) Long Short-Term Memory. Neural Computation, 9(8): 1735-1780.
[22] Graves A., Jaitly N. (2014). Towards end-to-end speech recognition with recurrent neural networks. pp.1-9.
[23] Li S , Tang M , Guo Q , et al. (2017). Deep neural network with attention model for scene text recognition. IET Computer Vision, 11(7):605-612.
[24] Bahdanau D., Cho K., Bengio Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. Computer Science, pp.1-15.
[25] Chorowski J., Bahdanau D., Cho, K., et al. (2014). End-to-end continuous speech recognition using attention-based recurrent nn: first results. Eprint Arxiv. pp.1-10.
[26] Vegesna V. V. , Gurugubelli K., Vuppala A. K. (2018). Prosody modification for speech recognition in emotionally mismatched conditions. International Journal of Speech Technology. Vol.21, No.3, pp.521-532.
[27] Ma T. H., Yang H. M., Tian Q., et al. (2021). A Hybrid Chinese Conversation model based on retrieval and generation. Future Generation Computer Systems, Vol. 114, No. 2021, pp. 481-490.
[28] Yang H. M., and Ma T. H. (2021). Compound Conversation Model Combining Retrieval and Generation, Computer Science, doi: 10.11896/jsjkx.200700162. (Online)
[29] Hung C. W., Zeng S. X., Lee C. H., Li W. T.(2021). End-to-End Deep Learning by MCU Implementation: An Intelligent Gripper for Shape Identification, Sensors, Vol.21, No.3, pp.891-891.
[30] Ge D. Y., Yao X. F. (2008). Application of Neural Network in Teaching Quality Evaluation of SCM Course. Proceedings of 2008 International Seminar on Education Management and Engineering. pp.638-642.