www.cadanda.com
Computer-Aided Design AND APPLICATIONS

Augmented Reality-based Oral Chinese Teaching System under Human-Computer Interaction

Yingying Zhang* and Huiyu Guo

International Education College, Gannan Medical University, Ganzhou, 341000, China

Corresponding author: Yingying Zhang, stellazhang@gmu.edu.cn

Abstract. There has been a spike in research papers studying the use of Augmented Reality (AR) in academic settings over the previous decade. AR's capacity to extend the classroom atmosphere with augmented 3D teaching materials with greater features has been shown in various researches. With the structuring of learning and practice, an increasing number of teachers have used information-based instructional techniques when constructing classrooms; stringent laws use of images and audio-visual equipment to boost the number of data input, while dynamic stationary instructional content keeps the school wealthy and intriguing. Static school books have evolved into vibrant textbooks due to multi-faceted vocabulary instruction, impactful characters (message, images, sounds), diverse exercise topic establishing, and better situations focusing on student features, mental acceptance, and requirements. Learners can easily grasp and execute the Chinese languages acquired in personal communication training. An oral Chinese teaching auxiliary system (OCTAS) is proposed in this research. A great spoken textbook emphasizes what to study and how to gain knowledge. Using the information-based instructional methods, the usage and teaching impact on Chinese as a second language is considerably improved.

Keywords: Chinese teaching, oral teaching system, human-computer interaction, foreign language teaching, AR.

DOI: https://doi.org/10.14733/cadaps.2023.S9.40-60

1 INTRODUCTION TO LANGUAGE TEACHING SYSTEMS

People conducted a study on speech feelings in the 1980s; this included several aspects connected to emotion. Human contact has become a prominent study topic, with the fast growth of artificial learning and the refinement of associated research approaches [25]. The research objective is for computers to identify and show emotion as people do. Conventional emotion detection research focuses on the following facets: the conceptual foundations of emotion detection and database choice, voice signal pre-processing, edge detection of vocal emotion signals, and identification model design [1].

New developing technologies such as augmented reality (AR) and virtual reality (VR) are expanding educational opportunities beyond the restrictions of human help, location, time, and resources [5]. Many of these apps have grown in importance as a result of the urgent drive in academia towards digital and remote learning. However, one aspect that is sometimes missed is the capacity to give individualised learning, which will be described in this article.

The efficiency of spoken Chinese lessons has been a focus of interest for overseas Chinese employees as instructional paradigms and methods for teaching Chinese as an overseas language continue to evolve [24]. The instructors of Chinese as a second language face determine the most appropriate spoken texts and teach techniques to express themselves fully and actively. There is a specific teaching approach in every guidebook. The author's analyses and research of the existing spoken courses discover that many of the school books it has been using are far removed from everyday life and have insufficient teaching materials [21]. Having a strict plan for the instructor in the school and a specific training approach for the class teacher, which frequently requires less work, is tremendously damaging to the students' progress.

With the fast growth of digital, an increasing number of instructors are incorporating information-based learning approaches into their classroom designs [6]. It could not only compensate for the deficiencies of the above issues if it uses digital technologies in spoken language education, and it can also tear through the blackness and tier constraints of curriculums and achieve maximum use of all the huge distinctive amount of information, intuitive and quickly adaptable and diverse benefits [26]. It is shown to learners in a more fascinating and adaptable form by enhancing different data and linguistic elements in books to study in a calm and pleasant setting and have fun instructing [13].

The pronunciation, shape, and meaning of Chinese characters are all intertwined. On the other hand, several Chinese characters are words in and of themselves. As a result, international students who understand Chinese symbol sound, meaning, and handwriting establish a solid basis for studying Chinese pronunciations, vocabulary, and syntax [14]. The value of teaching Chinese letters cannot be overstated in this context. The progress of Chinese verbal skills such as hearing, speaking, writing, and typing, particularly at the basic level, is primarily dependent on the learner's understanding of Chinese symbols [9]. It is hard for overseas students if the study of Chinese letters is not given sufficient emphasis from the start. The more Chinese they study, the more difficult this becomes.

The features of the availability and value of expertise regulation rely on the educators' expertise processing technologies. The efficient use of picture instructional practices greatly helps promote the sensible organization of students' experiential learning and optimizes the procreation of understanding, as per the hypothesis [16]. In reality, knowledge education has successfully realized the coupling of teaching microstructures and teacher preparation, which is difficult to do with conventional teaching techniques.

The main contributions of this article are as follows

- The teacher-student model is proposed in this article to teach the Chinese language.
- The smart evaluation system helps to identify the efficiency of the suggested model.
- The knowledge database is used to improve the learning outcomes of the system.
- The oral speech output is analyzed using the human-computer interaction model.

The remainder of this article is as follows: section 2 describes the background to the Chinese language teaching models. The proposed oral Chinese teaching auxiliary system (OCTAS) is designed and evaluated in section 3. Section 4 discusses the software outcomes of the proposed system. The conclusion and future scope of the proposed system are depicted in section 5.

2 BACKGROUND TO THE CHINESE LANGUAGE TEACHING MODELS

"Design thinking" has gained prominence in recent years, and it is today seen as an amazing new model for effectively addressing a wide range of issues in fields such as information technology, commerce, school, and medical [22]. The phrase "product design" was its title in 1989, and it had since been ingrained in the collective psyche of design academics. Ever since a slew of creative thinking frameworks had arisen, each based on a unique perspective on the design issue and drawing on theories and concepts from design method, sociology, and schooling, among other fields suggested by [3]. These lines of inquiry had resulted in a vast and diverse knowledge of human existence.

A comprehensive literature review was expected to understand the comprehensive effort to develop interactive multimedia within the educational establishment using AR, in order to see how AR learning implementations could be augmented by a program implemented Handsfree interplay in moderating effect higher education classrooms by autonomous algorithms [17].

Design theory solved complicated issues and created creative solutions based on a consumer perspective and multi-disciplinary groups [18]. There was yet to be a widely recognized description of creative thinking, as well as the name itself was a source of debate among professionals and proponents. However, one aspect unites them all in terms of ideas: a focus on individuals and the creation of the next. From new products to the assessment of the interaction with people and goods, then to the analyzing the relationship among people, the journey from designing to creative thinking was truly from new products to the analyzing the relationship among people suggested by [19]. As a result, design thinking was a human-centered creative process to create solutions that incorporated genuine consumer demands rather than required functionality.

Design thinking was a method for developing new products and fostering creativity. Furthermore, design thinking was frequently employed to concentrate on problem-solving [2]. Consequently, this study employed design theory to comprehend the existing situation and identified the problem with Thai learners' Chinese language acquisition. While innovation was employed, academics also provided specific techniques, patterns, or procedures on functioning or executing it. From the user's standpoint, risk management was an iterative process that started with creating user knowledge and ideas and began testing and implementation suggested by [27]. Although there were other versions of the design phase, this article concentrated on the test methodology developed. By speaking with specialists or performing research, the developer's expertise was increased and deepened.

The issue and its environment were defined to fix it and developed relevant ideas [8]. This step translated empathized user intents into more in-depth user demands and observations. The idea was to come up with a definition of the problem. Finally, this explanation directed what adjustments to improve the user experience.

The use of augmented reality to generate digital information for studying science topics has yielded promising results, whether using marker-based, indicator, or tracking devices technologies. AR Interaction approaches using virtual items in real-world situations increase interactivity and improve learning engagement. Global positioning system (GPS), gyroscopes, inertial measurement units, compass, wearable network, touch identification, natural language processing, action recognition, monitoring markers, hand monitoring, and eye movements are some of the sensors and input methods that are utilised in augmented reality applications for engagement [20].

Come up with some ideas - Ideate entails expanding one's mind, being innovative, and coming up with several solutions to an issue. At the idea-generating stage, all essential people's thoughts were gathered suggested by Mahalingappa et al. [15]. The key notion was that the ideas created were not evaluated or limited in any way. This phase fosters collaboration among ideas while encouraging the number of ideas. Its major features were curiosity and originality, represented in lateral thinking to investigate a larger optimum solution [10].

Create a prototype - Developing a prototype Users had a quick and flexible approach to prototypes during the initial design suggested by [12]. The earlier identified issues, the more beneficial the overall design phase was. Simultaneously, the prototype process determined if a remedy was too sophisticated or simple, allowing for quick learning and consideration of other options.

Put it to the test - Testing puts workable solutions developed throughout the design phase into action and takes feedback on these suggested by [28]. The screening was used to improve and develop the answer, place it in an actual operating context, and link the embryonic prototypes to the patient's reality through a battery of experiments to arrive at a good method.

Due to the issues above, this study looked at the automated scoring system for English-Chinese oral translating problems [23],[7]. In the vast bulk of speaking Communication translation examinations, applicants have very little time listening to issues and responding to them. Some Chinese-English speaking translations gave the contenders the language to interpret in live time. In contrast, others included listening to a tape of a topic replayed once or twice, and the applicants were answering questions following the sounds of prompting [4]. As a result, while hand scoring oral English-Chinese interpreters, the professor's scoring point firstly analyzed the fullness of crucial data in the replies and the coherence of the paragraph's primary concept before continuing with the thorough grading based on the participant's talking proficiency [11].

As per the research, the major information elements, the primary concept of the phrase, and the flow of language were chosen as important factors of the Chinese translation score system. The points system provided comments on the person's response based on the results of the three variables and the specialist skillset after evaluating the candidate's response recording. The outcome was used as a benchmark for scoring with the help of a teacher.

3 PROPOSED ORAL CHINESE TEACHING AUXILIARY SYSTEM

The instructor model, the learner model, and the teacher-student tutorial mode are all discussed in depth in this section. While the instructor and learner models have the same network topology as the benchmark baseline, their decoding techniques are different. In practice, it first trains a conventional instructor model in the instructor forcing method for end-to-end teaching systems, referred to as the training set. The instructor model is meant to replicate the genuine levels of natural voice signal as it develops in the instructor pushing mode.

Then, in a free-roaming state, it educates another prototype system. The learner model is taught by simultaneously understanding the ground-truth series and the instructor model's concealed variables.

3.1 System Model

The learner model efficiently learns the real distributions of the real voice signal by understanding from the concealed states of the instructor model via levels of learning. Because the learner classifier is developed in free-roaming mode with the anticipated voice fragments as the decoder's feed, it should adapt to the run-time interpretation scenario.

3.1.1 Teacher model

It uses the teaching pushing mode for the decoding in the instructor model, which forecasts a voice frame using the preceding speech framing in the series as input. Let $P(q|p,\theta)$ be the instructor model of these θ is the parameter estimates, given a data defined clearly $p = \{p_1, p_2, \cdots, p_T\}$ and its goal mel-spectrogram characteristics $q = \{q_1, q_2, \cdots, q_T\}$. The preceding frames $q_1, q_2, \cdots, q_{t-1}$ from the goal, speech patterns are used as input by the instructor models with

instructor pushing mode to forecast the characteristic frame q_t at time step t, as defined in Equation (1).

$$P(\hat{q}|p,\theta) = \prod_{t=0}^{T} P(\hat{q}_t|q_{<t},p,\theta) \quad (1)$$

where $\hat{q}$ is the expected value, and q comes from the everyday speech of the subject. The instructor model is predicted to learn the real probabilistic model (θ) from a real voice signal in this decoder mode, which would be highly useful for the learner model.

3.1.2 Student model

The core network of the learner model is identical to that of the professor model, except for an entirely new decoder mode: free-roaming mode. The student model is incorporated with AR technology. The decoder anticipates a voice panel in this manner by using the sequence's prior anticipated speech pixels as feed. The learner model's decoder method is denoted in Equation (2).

$$P(\hat{q}|p,\theta) = \prod_{t=0}^{T} P(\hat{q}_t|\hat{q}_{<t},p,\theta) \quad (2)$$

where $\hat{q}$denotes the expected value. The deviation value is denoted θ.

3.2 Knowledge Distillation

Information distillation is often a procedure in which a tiny model is taught to replicate a bigger model that has already been trained. This study uses a cognitive level to execute the teacher-student workout program. The goal is to employ a teacher model that has been trained in teacher pushing mode to lead the development of a learner model that is freely operating. It anticipates the instructor model's outputs probabilistic model to represent the genuine probability of the real voice signal since it is taught using speech sounds phrases as the decoder's feed. The free-roaming mode is used to train the prototype system. As a result, it is more closely related to the current inference situation. At the same moment, through the cognitive level, the learner model's concealed states are tuned to be similar to those of the instructor model.

The workflow of the proposed OCTAS model is denoted in Figure 1. It has a character embedding layer, convolutional layer, long short-term memory (LSTM) layer, and decoder for analyzing and processing the voice signal. The final reconstructed signal is compared with the computer-trained samples.

It creates one goal function for the instructor model, the characteristic loss. It creates two optimization methods for the learner model: one for characteristic loss, the same in the instructor model, and the other for the level of learning or evaporation loss. The AR technology helps the students to learn easily and increase their learning outcomes.

Following that, it outlines the complete procedure. The encoder turns the one-hot vectors to continuous top-level characteristics description f: from the incoming letter sequence $p = \{p_1, p_2, \cdots, p_T\}$ from the provided text. The characteristics description is denoted in Equation (3).

$$f_t = En(f_{t-1}, p_t) \quad (3)$$

The characteristic feature is denoted f_t, and the incoming letter sequence is denoted p_t. At every step t, the instructor Decoder D produces a hidden layer h_t and it is expressed in Equation (4).

$$h_t = D(h_{t-1}, \hat{q}_{t-1}, \vartheta(f_t)) \quad (4)$$

where $\vartheta()$ is a method that takes the location-sensitive attentiveness method for calculating the vector representation. The past outcome of the sequence is denoted $\hat{q}_{t-1}$. The features are denoted f_t. The hidden layer feature is denoted h_t. Likewise, the undergraduate decoding Decoder S analyses the identical input data and simultaneously construct the hidden layer $\hat{h}_t$ and it is expressed in Equation (5).

$$\hat{h}_t = D\left(\hat{h}_{t-1}, \hat{q}_{t-1}, \vartheta(f_t)\right) \quad (5)$$

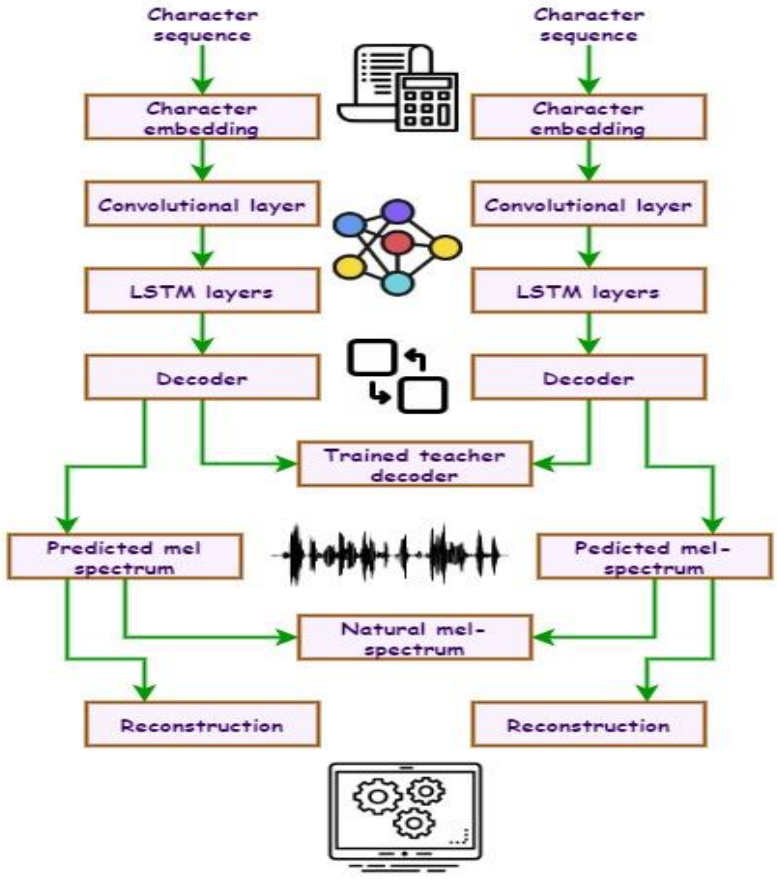

Figure 1: The workflow of the proposed OCTAS model.

The past outcome of the sequence is denoted $\hat{q}_{t-1}$. The features are denoted f_t. The predicted hidden layer feature is denoted $\hat{h}_t$. The characteristic loss function L_f guarantees that the produced speech is near the goal language in both the instructor and learner models. The loss function is denoted in Equation (6).

$$L_f = \sum_{t=0}^{T} L_g(\hat{q}_t, q_t) \tag{6}$$

The predicted outcome of the sequence is denoted $\hat{q}_t$, and the actual outcome is denoted q_t. It incorporates the distilling penalty L_d in the learner model to reduce the disagreement among the concealed states h and $\hat{h}$ of the instructor model and the learner model. The distilling loss is denoted in Equation (7)

$$L_d = \frac{1}{T}\sum_{t=0}^{T}\left|h - \hat{h}\right|^2 \tag{7}$$

The hidden state is denoted h, and the expected state is denoted $\hat{h}$.

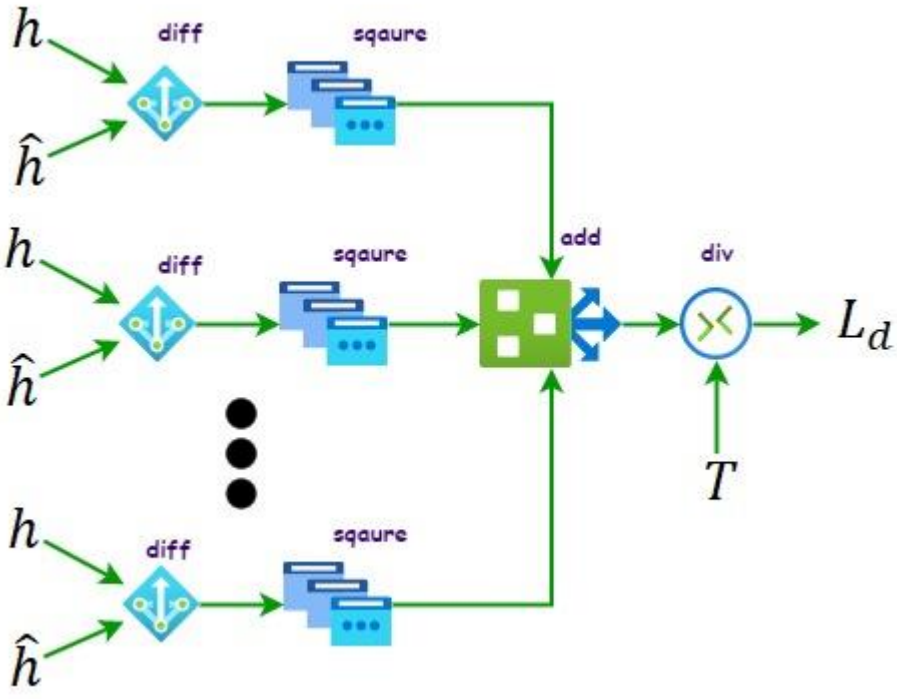

Figure 2: The pictorial view of the function L_d.

The pictorial view of the function is denoted in Figure 2. It uses hidden layer states and the expected hidden layer states for the computation of $\mathrm{L_d}$. As a result, the complete loss functional for the school model is denoted in Equation (8)

$$\mathrm{L_{tot}} = \mathrm{L_f} + \gamma \mathrm{L_d} \tag{8}$$

The loss factor is denoted as $\mathrm{L_f}$, and the distilling loss is denoted $\mathrm{L_d}$. When γ is a trade-off variable for the two damage factors, the suggested 2-step instructor development learning method enables a more compacted End-to-End network than the others, like the generative model, thanks to the level of learning. In the instructor pushing mode, the optimization problem $\mathrm{L_f}$ is used to teach the instructor model; while in the free-roaming mode, a mixture of multiple loss measures L_{tot} teaches the prototype system.

3.3 Voice Processing System Design

First, the input audio emotional signal is pre-weighted, and the vocal emotion information is pre-processed. The goal is to accentuate the high-frequency portion of the voice, reduce the impact of lip emission, and improve the verbal signal's better definition. The voice signal is separated into small segments to process based on its short-term reliability. The frame is achieved by utilizing a moveable finite-length frame to weigh; a rupture in the waveforms diagram of the voice signal terminus recognition is employed to fill it in.

The identification is based on the zero-crossing frequency and the short-time spectrum sensing approach to increase the shows that a strong of the extracting features limit the influence of unrelated data. The voice signal's mel frequency correlation coefficient (MFCC) and energetic parameters are then retrieved and fed into the recurrent neural network (RNN) for retraining. The

RNN was created to better represent the correlation time in the temporal series' output voltage and the signals across the instances.

3.3.1 Feature extraction

To acquire the spectrum properties of the spoken emotion signals, the MFCC combines basic arithmetic to mimic the auditory neural system of the auditory system and a collection of triangle filters to replicate the sensory nerves of an inner ear. On the other hand, the Mel spectrum is non-linear, and triangular filter dispersion is dense in the low-frequency region but scarce in the high-frequency and mid-frequency regions. The MFCC computation precision and sampling rate are good in the lower frequencies section; however, the MFCC computation precision and sampling rate are inadequate in the middle to upper frequencies, and some impulses are lost. As a result, it enhances the existing MFCC extraction method.

1) Pre-aggravation is defined as the transformation of the original audio emotional input $s(n)$ through a high-pass filter $H(s)$, with the high-pass filter represented in Equation (9)

$$H(s) = 1 - bs^{-1}; \text{and } 0 < b \leq 1 \tag{9}$$

The benefit of s in the preceding formula is that the output doesn't vary at a lower frequency; nevertheless, the signal grows as the frequency increases. The temporary variable is denoted b.

2) Windowing utilizes the voice signal's short-term normality, and the signal's properties are considered to remain stable throughout a short period. An interleaving approach divides the speech into parts, referred to as frames. Hanning window and Hamming window are two often used window algorithms. This work uses the Hanning window operator, as shown in Equation (10).

$$wf(k) = \begin{cases} 0.51 - 0.42\cos\left(\frac{2\pi k}{N-1}\right) & k = 0,1,2,\cdots,N-1 \\ 0 & \text{else} \end{cases} \tag{10}$$

The filter size is denoted N. The window function is used to compute errorless voice signals.

3) The Discrete Fourier Transform (DFT) is a discontinuous transfer function in both the temporal and frequencies domains, translating the signal's time-frequency patterns into spatial frequency patterns. Every filter bank is the total of its spectral filtering characteristics, using a Mel-scale-based bandpass filter.

4) The Mel filter bank works the same way as the human ear recognizes sounds, and it normally brings 24 filter banks. Because the normal hearing system is unaware of the stage, the maximum value is used, and the small difference is removed. The log was chosen because human feelings are relative, and the log is a unique mathematical connection.

5) It returns to the timeframe after receiving the inverse DFT (IDFT). Lastly, the divergence is calculated twice to yield the characteristic vector's change over time. Every time the derivatives acquire a 12-dimensional characteristic, it also gets a 36-dimensional characteristic. The MFCC's convolutional is reduced using the differential function. The number of channels is represented by m, and the first and second-order distinctions are written in Equations (11) and (12).

$$\Delta o_x(q) = \frac{\sum_{l=0}^{N} n \times o_{x-l}(q)}{\sum_{l=0}^{N} l^2} \tag{11}$$

$$\Delta^2 o_x(q) = \frac{\sum_{l=0}^{N} n \times \Delta o_{x-l}(q)}{\sum_{l=0}^{N} l^2} \tag{12}$$

The outcome of the layer is denoted $o_{x-l}(q)$, the size of the layer is denoted l, and the number of elements in the layer is denoted n.

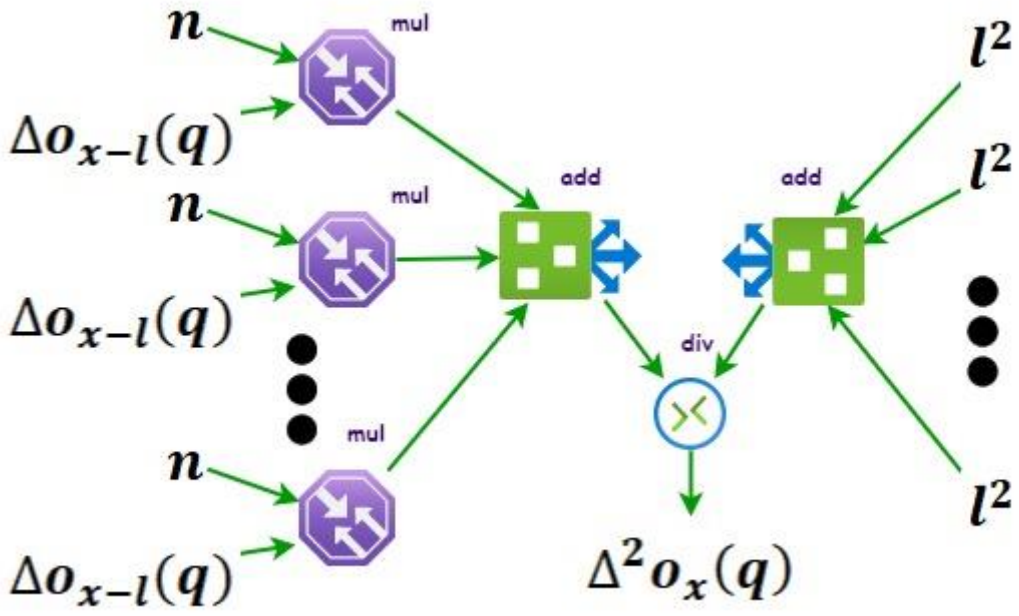

Figure 3: The pictorial representation of the function $\Delta^2 o_x(q)$.

The pictorial representation of the function $\Delta^2 o_x(q)$ is denoted in Figure 3. It uses the outcomes of each layer, the number of elements, and the filter's size to compute the final results.

6) The signal's power is separated twice to provide a 3-dimensional power characteristic that more properly represents the speech's emotional aspects. Eventually, the 38-dimensional MFCC characteristics are acquired.

3.3.2 RNN structural design

The nodes of the RNN structure's concealed layer are linked. The concealed layer's power supply comprises the present incoming signal and the output current from the prior time step at every timestep.

The workflow of the speech analyzing model is depicted in Figure 4. The parameters are set, and the voice input is collected from the students. The collected information is processed with layers like hidden layer, calculation layer. The calculated error is removed by adam optimizer, and the trained system is considered to analyze testing samples. As shown in the diagram above, the recurrent neural channel's input is provided an outcome regarding the changing condition at all moments in history. The input data and a condition describing the overall time point are fed into the primary model. At every particular time, the RNN receives an input, produces an output based on the present, and then passes it on to the next particular time. The outcome is compared with the actual AR environment with the training and testing results.

The fundamental RNN unit architecture is used in this article, which employs the TensorFlow computational model. The experiment shows that the microstructures are 20 intake nodes, six output vectors, and 25 concealed layer neurons nodes, with a 0.02 and 500 loop cycles detection rate. The learning date is the first 50 examples of every emotion, and the final exam is the final 50 samples.

Figure 4: The workflow of the speech analyzing model.

The following are the network's primary characteristics: the Adam optimizer is designed to decrease the difference between the estimated and true values during learning. Adam optimizer is the best option to reduce loss. Second, the output categorization evaluation method classification report() is being used to examine the reliability, several responses, and F1 value of separate classes to enable the precision and recall rate per the classification. The confusion matrix() is a scenario analysis desk that summarises the categorization model's deep learning forecast outcomes. The data is presented in a matrix format based on the real categorization and the categorization parameters suggested by the SVM classifier.

The Python language retrieves the signal's characteristics using the TensorFlow learning algorithm. In this study, 38-dimensional MFCC elements and 50-dimensional MFCC characteristics are chosen to see how various features affect the outcome of vocal emotion identification. Short-time power relates to the voice signal's energy content over a short period, connected to the strength of sound waves, implying that the energetic characteristic allows speech signals to communicate emotional aspects more precisely.

3.4 Oral Teaching Model

Learners must improve their academic effectiveness and continue growing their vocabularies while introducing new Chinese terms to outsiders, particularly at the sound pedagogical stage at the basic level. How to get learners to speak new language accurately, swiftly, and efficiently and increase their capacity to retain and utilize new words is an issue we need to investigate more.

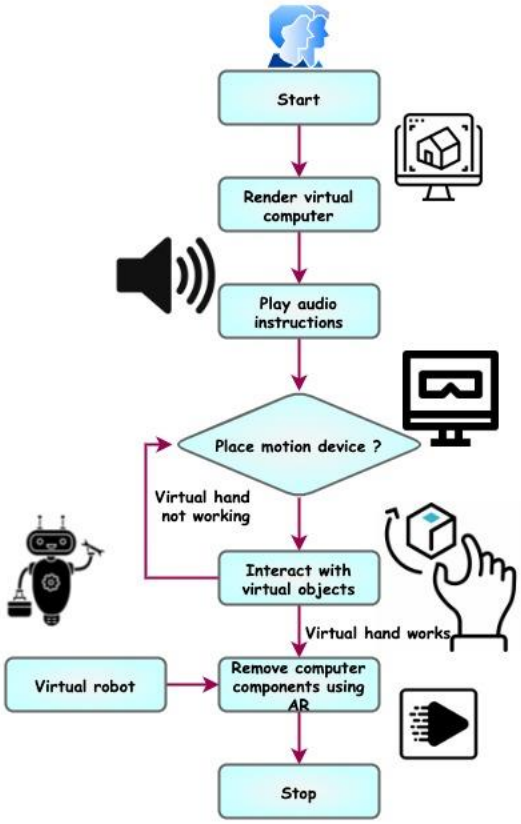

Figure 5: The AR-based learning model in the proposed OCTAS model.

The AR-based learning model in the proposed OCTAS model is depicted in Figure 5. The necessary features are gathered and rendered on the computer. The instructions are sent via audio signals and the motion of the students are continuously monitored and the respective AR motions are adjusted. The virtual robots help Chinese oral teaching methods.

1) Learning to pronounce new words: Because of the unique characteristics of the Chinese phonological systems and the personal variances among students, some thousands of students have abnormalities in their pronouncing after completing the intense speech supervised learning. It is still obligatory for schools to teach it.

Teachers often read conventional language orthographic techniques and pronunciation, but the issue of tongue placement can only be conveyed by sketching or motions. If it employs information systems, it uses lip charts, animation, and other tools at any moment to examine and feel the sound, form, and audio again and again in the human-computer interaction system. It videotapes and compares the recorded version. Teachers utilize students' assignment recordings to do speech tailoring, identify phrases with pronouncing issues, summarise them, and "consult" instructors and learners in the classroom to swiftly and efficiently fix students' pronouncing issues.

2) New word use: Many instructors' absorption of new terms, as per the author, is generally confined to studying and evaluating translation. Some phrases are forgotten quickly using this single approach of memorization. The use of computer technology improves students' comprehensibility of words learned in various formats, boosts memory activation about other parts, and leaves a lasting impact on their brains.

It watches, chats, waits, and if the topic is "eating," it shows a movie of someone eating or a photograph of someone eating; if the topic is "air conditioning," it can display a photograph of air conditioners, etc. When the phrase "change" is uttered, a change is shown; whenever the word "anger" is being spoken, an image of a smiling face on the skull is presented. This method of studying connects the new vocabulary with the existing ones.

The statement "These clothes are quite fat" is another instance. It can provide a photograph of a set of fat and slim jeans. Based on the photo, learners can pronounce "fat" and "thin," and afterward, "this pair of trousers is fat." Learn to articulate what they've learned about what they've learned. Study phrases and grammar with the use of rapid and simple computer technology. Train pupils' verbal or phrase response and oral expression abilities with brief examples. Instructors could further analyze textbooks, construct different information-based practice forms, completely activate all recollection organs, and employ them audio-visually to improve students' capacity to use language effectively. For instance, look at photos to talk, look at photos to speak, look at images to complete in the blanks, listen to words to pick images, etc.

The classroom engagement is heightened in this practice mode, and the learners' interest in learning new words is substantially raised, as is the terminology mastered in a unit period. With the rise in the quantity of information, current discovery is successfully input, and the effectiveness of evaluating ancient knowledge is increased, particularly in a short time. The development of information education has aided in the overall acquisition of new words by optimizing and improving the learning process.

3.4.1 Teacher-student training

The instructor model, the learner model, and the teacher-student tutorial mode are all discussed in depth in this section. At the same time, the instructor and pupil models have the same network topology as the standard baseline. In practice, it first educates a conventional teaching model in instructor pushing mode for end-to-end systems, referred to as the instructor model. The instructor model is meant to replicate the genuine levels of natural voice signal as it develops in the instructor pushing mode. The AR model helps to develop a better teacher-student environment in the virtual condition.

Then, in free-roaming mode, it educates another learner model. The student model is taught by simultaneously understanding the ground-truth series and the instructor model's concealed states. The student model efficiently learns the real distributions of the actual voice signal by understanding from the concealed states of the instructor model via the level of cognitive. Because the learner model has been trained in free-roaming mode with the anticipated speech phrases as the decoder's inputs, it should adapt to the run-time inferences scenario.

3.5 RNN Model Optimization

The early recollection is gone if the RNN is trained over a lengthy period. In reality, when the input passes via the RNN, little information is removed at each step. In the RNN state, there is no sign of first inputs immediately after. This research presents an optimization method for Long Short-Term Memory networks (LSTM) to address the issues above. The fundamental difference between an LSTM network and a standard RNN is the addition of three gate components in the LSTM network. The LSTM unit's unusual gate format enables it to store and retrieve long-period audio signals. Improving the stable complex of the RNN model is vital to get a greater correct

identification rate in less time. The LSTM's novel gate structure eliminates the issue of gradient vanishing or slope inflation in the typical RNN time axis, allows for more speech signal reserve and transfer, and increases the experiment's reliability.

3.5.1 Process

The 5-step Empathizes-Defines-Ideates-Prototypes-Tests technique is utilized to verify the whole study in this paper. Due to the strong relationship between define and ideate. This section explains what happens in each phase by following these processes.

Step1: This article attempts to locate the issue within the company in the first phase, based on empathy and from the standpoint of the leader. While instructing, it collects the advisor's responsibilities for Chinese instructors. The writer starts with no pre-determined research topics other than to investigate the existing state of the organization and attitude toward which aspects of the issues they genuinely desired to fix.

Step 2: The second phase concentrates on the most pressing problem to be resolved. This section reduces the emphasis and demonstrates the relevance of the chosen problem after studying the manager's diverse demands.

Step 3: The prototypes for fixing the issue are suggested in the third step. This part imagines how to uncover the underlying cause of the specific problems from learners and lecturers, depending on the most significant concerns. The interview protocol inquiries were created to analyze the issues from the viewpoint of Chinese instructors. In addition, the tool is used to assess Thai students' ability to read Chinese letters.

Step 4: The prototype's outcome is implemented in the fourth step. Prototype survey questionnaires were employed to obtain insight and specific details from Chinese instructors. Exam boards would be used to assess the pupils' Chinese language skills. In the meantime, after collecting exam results, an in-depth session with Chinese instructors to delve deeper and thorough information.

The proposed OCTAS model is designed in this section with the student-teacher model, training, and evaluation of speech signals. The simulation outcomes of the proposed OCTAS model are computed and analyzed in the next section.

4 SIMULATION OUTCOME ANALYSIS

The content was written using the TensorFlow1.12.0 machine learning application's GPU edition. This study compares RNN to support vector machine (SVM) and classical classifiers. The Libsvm toolbox is used to construct the SVM classifiers in MATLAB2012b. It's matched to an existing vocal emotion detection algorithm using the public talk dataset to check the networking model's legitimacy.

The recognition rate analysis of the existing support vector machine (SVM) model and the proposed OCTAS model are shown in Figures 6 and 7. The simulation is carried out by considering the database consists of some words, and the teacher trains the students in Chinese oral speaking. The spoken Chinese words are compared with the trained signals in the human-computer interface. The proposed OCTAS model is evaluated with the help of recurrent neural network (RNN), human-computer interface models, and evaluation models. The proposed OCTAS model exhibits a higher recognition rate of the speech signal from students than traditional methods.

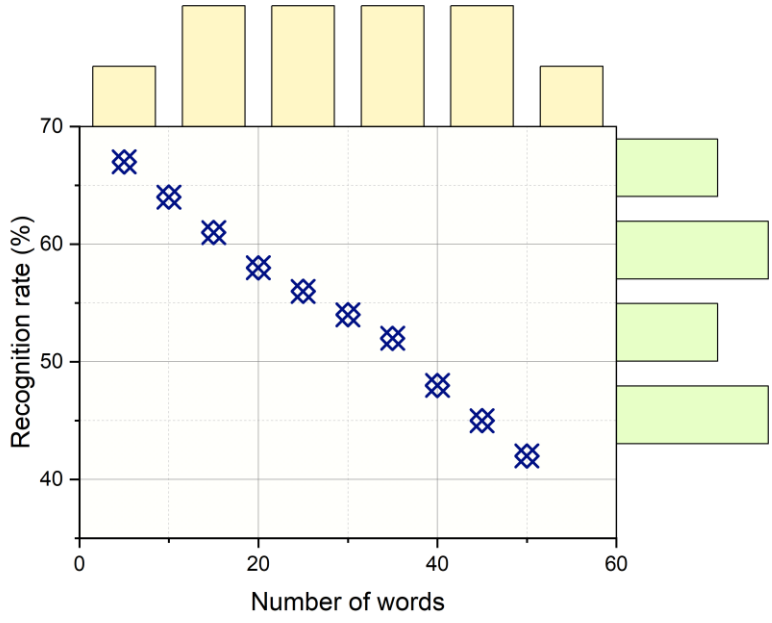

Figure 6: The recognition rate of the existing SVM model.

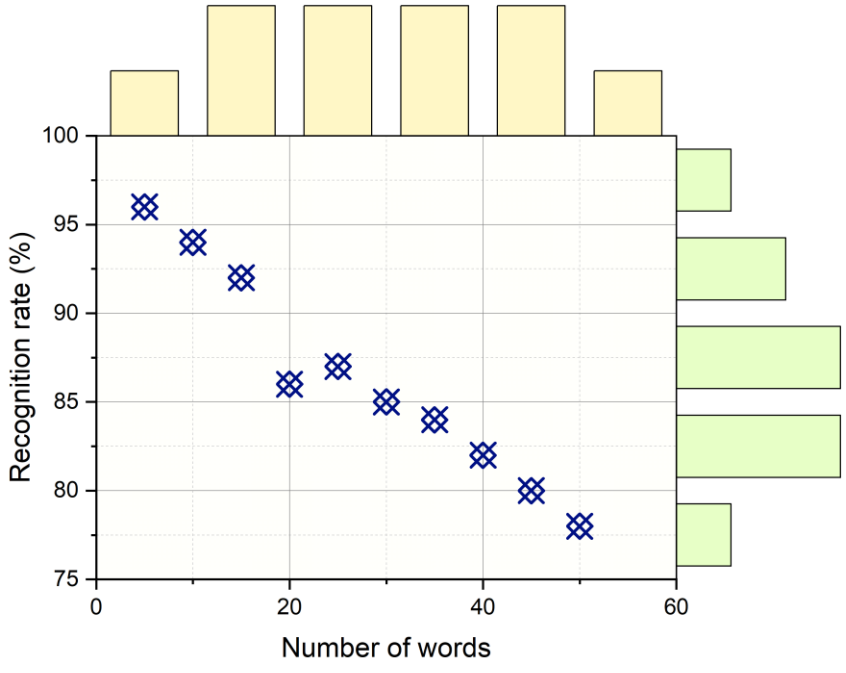

Figure 7: Recognition rate analysis of the proposed OCTAS model.

Number of words	*SVM (%)*	*OCTAS (%)*
5	67	96
10	64	94
15	61	92
20	58	86
25	56	87
30	54	85
35	52	84
40	48	82
45	45	80
50	42	78

Table 1: Recognition rate analysis of the proposed OCTAS model.

Table 1 shows the recognition rate analysis of the existing SVM models and the proposed OCTAS model. The simulation outcomes of the proposed OCTAS model are evaluated under the given simulation environment and simulation tools. The recognition rate of both existing and the proposed OCTAS models are monitored by varying the number of words from a minimum of 5 words to a maximum of 50 words. The oral Chinese sound is analyzed with the help of the human-computer interface, and the proposed OCTAS model produces higher outcomes with the help of mathematical models.

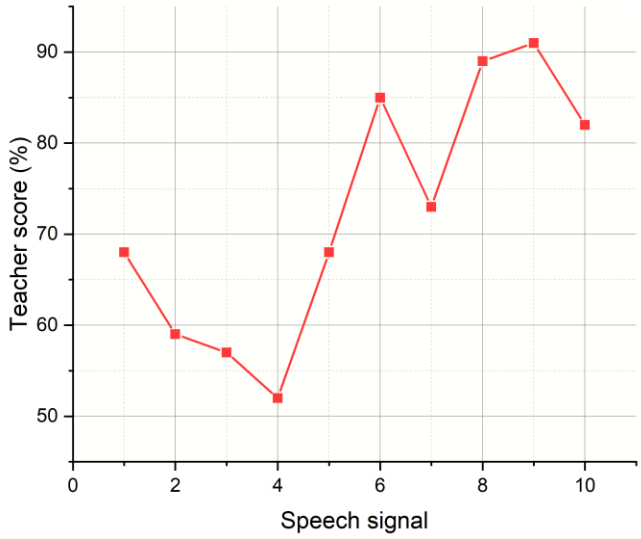

Figure 8: The teacher score analysis of the proposed OCTAS model.

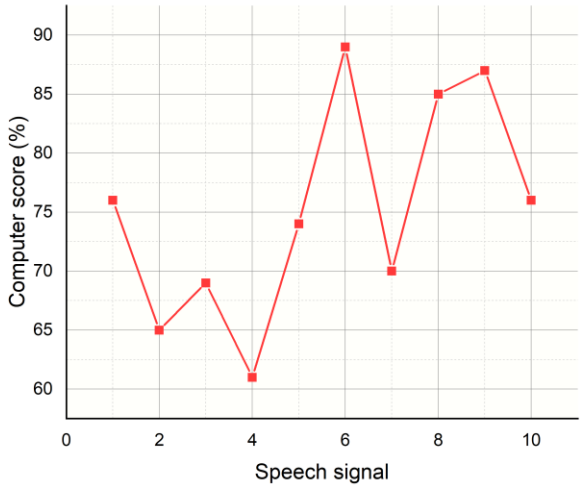

Figure 9: The computer score analysis of the proposed OCTAS model.

The proposed OCTAS model's teacher score and computer score analysis are shown in Figures 8 and 9. The simulation outcomes of the proposed OCTAS model are evaluated using the given simulation environment by analyzing the voice signal of the students. The students are taught Chinese with oral speaking and listening. The teacher and also the computer evaluate the simulation outcomes of the students' Chinese speaking ability. The proposed OCTAS model exhibits higher evaluation results in analyzing the voice signal of the students.

Iteration	*Accuracy (%)*	*Precision (%)*
10	68	65
20	72	69
30	75	73
40	78	74
50	80	78
60	83	81
70	85	83
80	87	85

Table 2. Simulation outcome analysis of the proposed OCTAS model.

Table 2 shows the simulation outcome analysis of the proposed OCTAS model. The simulation outcomes of the proposed OCTAS model are analyzed and evaluated, and the simulation outcomes of the proposed OCTAS model in terms of accuracy and precision are monitored continuously. The simulation findings are evaluated by considering different iteration sizes by varying from 10 to a maximum of 80 iterations with an increment level of 10 iterations. As the iteration increases, the simulation findings of the proposed OCTAS model also increase with the help of the human-computer interface and mathematical model.

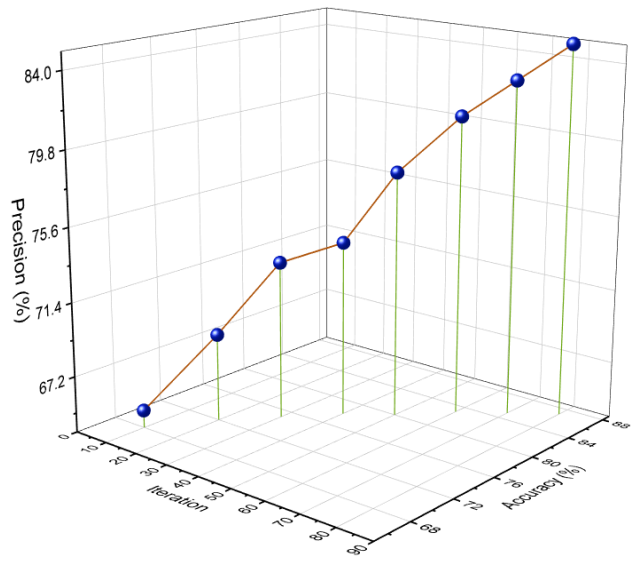

Figure 10: Precision and accuracy analysis of the proposed OCTAS model.

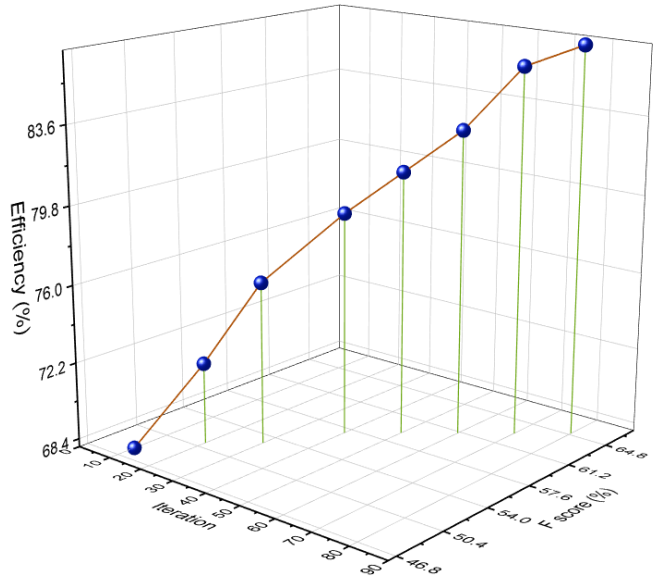

Figure 11: Efficiency and F score analysis of the proposed OCTAS model.

Figures 10 and 11 show the proposed OCTAS model's precision and accuracy, F score, and efficiency analysis. The simulation analysis of the proposed OCTAS model is evaluated, and the outcomes are monitored in terms of accuracy, precision, efficiency, and F score. The software outcomes of the proposed OCTAS model are monitored by varying the iteration size from a minimum size to a maximum size with a step size of 10 iterations. The proposed OCTAS model produces higher simulation outcomes with the help of human-computer interface modules and mathematical models.

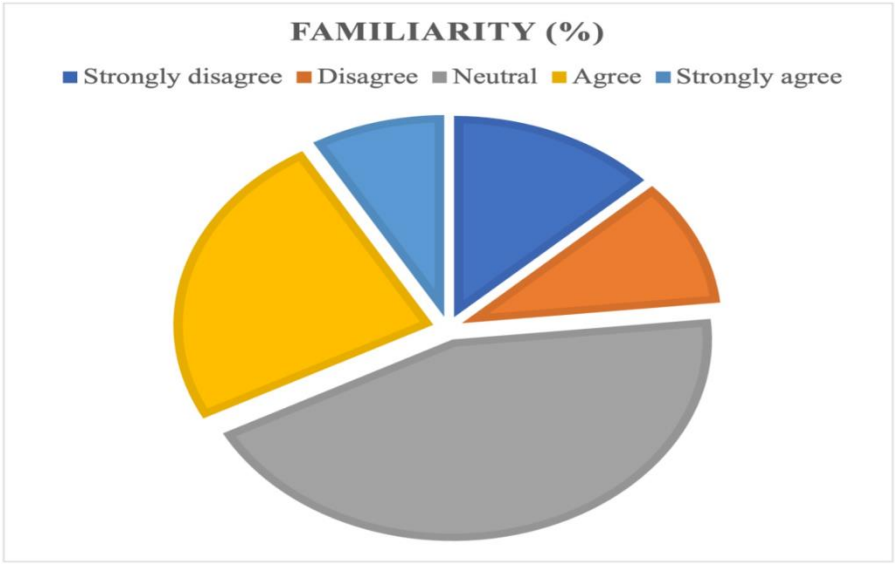

Figure 12: Familiarity analysis of the OCTAS model.

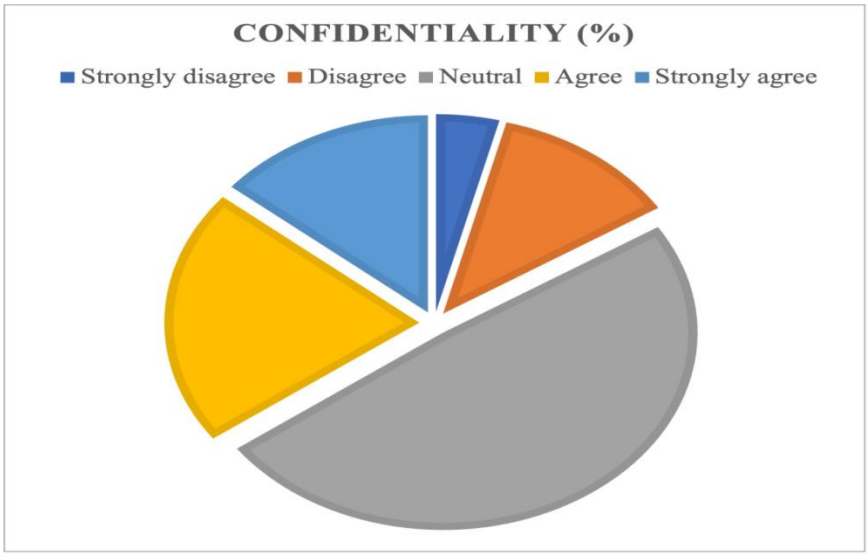

Figure 13: Confidentiality analysis of the OCTAS model.

The familiarity and confidentiality analysis of the OCTAS model are shown in Figures 12 and13. The OCTAS model collects feedback from the participants about the interaction between students and the AR environment. The majority of the students voted against the familiarity and confidentiality of the OCTAS model. The OCTAS model with AR module is easy to use and the system helps to increase the learning effectiveness of the students to learn the Chinese language. Very few people faced issues interacting with the AR environment.

The proposed OCTAS model is evaluated, and the outcomes are continuously monitored in this section. The software outcomes of the proposed OCTAS model exhibit higher simulation outcomes with the help of the human-computer interface and recurrent neural network.

5 CONCLUSION AND FUTURE WORK

Oral Chinese teaching is a physical training device for "stimulus-response" and fits the cognitive demands of learners learning Chinese as a second language. It is used as a cognition tool to aid students in activities. Using an informatics role in the instruction of spoken Chinese as a different tongue enhances teaching techniques and brings the learning materials closer to the learners' cognition law, promoting the major construction of spoken Chinese as a second language. An oral Chinese teaching auxiliary system (OCTAS) is proposed in this research. It can permit learners to practice skills widely, accomplish the schoolbook objectives to the maximum extent possible, and truly enhance the students' verbal learning outcomes if it merges abilities and averts weak points, naturally incorporates information-based instruction with spoken curriculums fairly incorporate teacher education. The system's performance will be increased in the future by incorporating artificial intelligence models. While the number of learners has good expectations for these technological advances, the findings indicate that there are still a considerable number of misunderstandings among them, particularly around the impacts of AR/VR. This is likely owing to the fact that only a small percentage of pupils are acquainted with virtual reality and augmented reality technology or how to use them in the classrooms.

Yingying zhang, https://orcid.org/0000-0003-0568-1899
Huiyu Guo, https://orcid.org/0000-0003-0614-0904

REFERENCES

[1] Bao, R.: Oral corrective feedback in L2 Chinese classes: Teachers' beliefs versus their practices, The system, 82, 2019, 140-150. https://doi.org/10.1016/j.system.2019.04.004
[2] Bao, R.: Collaborative dialogue between complete beginners of Chinese as a foreign language: implications it has for Chinese language teaching and learning, The Language Learning Journal, 48(4), 2020,414-426. https://doi.org/10.1080/09571736.2017.1422136
[3] Dalim, C.-S.-C.; Sunar, M.-S.; Dey, A.; Billinghurst, M.: Using augmented reality with speech input for non-native children's language learning, International Journal of Human-Computer Studies, 134, 2020, 44-64. https://doi.org/10.1016/j.ijhcs.2019.10.002
[4] Hulme, C.; Zhou, L.; Tong, X.; Lervåg, A.; Burgoyne, K.: Learning to read in Chinese: Evidence for reciprocal relationships between word reading and oral language skills, Developmental Science, 22(1), 2019, e12745. https://doi.org/10.1111/desc.12745
[5] Jang, J.; Ko, Y.; Shin; W.-S.; Han, I.: Augmented reality and virtual reality for learning: An examination using an extended technology acceptance model, IEEE Access, 9, 2021, 6798-6809. https://doi.org/10.1109/ACCESS.2020.3048708
[6] Jianling, L.: The impact of face-to-face oral discussion and online text-chat on L2 Chinese writing, Journal of Second Language Writing, 41, 2018, 27-40. https://doi.org/10.1016/j.jslw.2018.06.005
[7] Jin, T.; Liu, X.; Lei, J.: Developing an effective three-stage teaching method for collaborative academic reading: evidence from Chinese first-year college students, Journal of English for Academic Purposes, 45, 2020, 100853. https://doi.org/10.1016/j.jeap.2020.100853
[8] Ju, Z.; Zhou, Y.; DelMas, R.: The contributions of separate pinyin skills and oral vocabulary to Chinese word reading of US Mandarin immersion third graders, Reading and Writing, 34(9), 2021, 2439-2459. https://doi.org/10.1007/s11145-021-10150-9

[9] Jüngling, S.; Lutz, J.; Korkut, S.; Jäger, J.: Innovation Potential for Human Computer Interaction Domains in the Digital Enterprise, In Business Information Systems and Technology 4.0 Springer, Cham, 2018, 243-256. https://doi.org/10.1007/978-3-319-74322-6_16
[10] Li, L.; Wang, H. C.; Castles, A.; Hsieh, M.-L.; Marinus, E.: Phonetic radicals, not phonological coding systems, support orthographic learning via self-teaching in Chinese, Cognition, 176, 2018, 184-194. https://doi.org/10.1016/j.cognition.2018.02.025
[11] Li, W.; Liu, C.; Liu, S.; Zhang, X.; Shi, R.-G.; Jiang, H.; Sun, H.: Perceptions of education quality and influence of language barrier: graduation survey of international medical students at four universities in China, BMC medical education, 20(1), 2020, 1-13. https://doi.org/10.1186/s12909-020-02340-w
[12] Li, X.; Chu, S.-K.: Using design-based research methodology to develop a pedagogy for teaching and learning of Chinese writing with wiki among Chinese upper primary school students, Computers & Education, 126, 2018, 359-375. https://doi.org/10.1016/j.compedu.2018.06.009
[13] Lü, C.: Teaching and learning Chinese through immersion: A case study from the North American context, Frontiers of Education in China, 15(1), 2020, 99-141. https://doi.org/10.1007/s11516-020-0005-9
[14] Luo, H.; Yang, C.: Twenty years of telecollaborative practice: implications for teaching Chinese as a foreign language, Computer-assisted language learning, 31(5-6), 2018, 546-571. https://doi.org/10.1080/09588221.2017.1420083
[15] Mahalingappa, L.; Polat, N.; Wang, R.: A cross-cultural comparison in pedagogical beliefs about oral corrective feedback: The case of English language teachers in China versus the US, Language Awareness, 2021, 1-21. https://doi.org/10.1080/09658416.2021.1900211
[16] Peng, X.; Chen, H.; Wang, L.; Tian, F.; Wang, H.: Talking head-based L2 pronunciation training: Impact on achievement emotions, cognitive load, and their relationships with learning performance, International Journal of Human-Computer Interaction, 36(16), 2020, 1487-1502. https://doi.org/10.1080/10447318.2020.1752476
[17] Qiu, X. Y.; Chiu, C.-K.; Zhao, L.-L.; Sun, C.-F.; Chen, S.-J.: Trends in VR/AR technology-supporting language learning from 2008 to 2019: a research perspective, Interactive Learning Environments, 2021, 1-24. https://doi.org/10.1080/10494820.2021.1874999
[18] Sandnes, F.-E.; Eika, E.: Hostage of the software: experiences in teaching inferential statistics to undergraduate human-computer interaction students and a literature survey, Research on e-Learning and ICT in Education, 2018, 167-183. https://doi.org/10.1007/978-3-319-95059-4_10
[19] Tian, L.; Li, L.: Chinese EFL learners perceive peer oral and written feedback as providers, receivers, and observers, Language awareness, 27(4), 2018, 312-330. https://doi.org/10.1080/09658416.2018.1535602
[20] Tseng, M.-F.; Geng, Z.: Implementing Virtual Reality–Enhanced Tasks in Chinese Language Teaching, In Contextual Language Learning, Springer, Singapore, 2021, 91-118. https://doi.org/10.1007/978-981-16-3416-1_5
[21] Van Ha, X.; Nguyen, L.-T.; Hung, B.-P.: Oral corrective feedback in English as a foreign language classroom: A teaching and learning perspective, Heliyon, 7(7), 2021, e07550. https://doi.org/10.1016/j.heliyon.2021.e07550
[22] Vuletic, T.; Duffy, A.; Hay, L.; McTeague, C.; Campbell, G.; Grealy, M.: Systematic literature review of hand gestures used in human-computer interaction interfaces, International Journal of Human-Computer Studies, 129, 2019, 74-94. https://doi.org/10.1016/j.ijhcs.2019.03.011
[23] Wasko, C.-V.: Can We Use an Outside-Expert, Conversational ESP Stimulus with Adjustable Oral Communication Anxiety to Motivate Chinese ELL, First Year, Undergraduate Students During the Final Semester Quarter? In Asian Research on English for Specific Purposes, Springer, Singapore, 2020, 11-21. https://doi.org/10.1007/978-981-15-1037-3_2

[24] Xie, Y.; Chen, Y.; Ryder, L.-H.: Effects of using mobile-based virtual reality on Chinese L2 students' oral proficiency, Computer-assisted language learning, 34(3), 2021, 225-245. https://doi.org/10.1080/09588221.2019.1604551

[25] Yan, J.; Goh, H.-H.; Zhou, H.-X.: Improving the Teaching of Chinese Speaking of Young Students from English-Speaking Families: Teacher's Professional Development, In Teaching the Chinese Language in Singapore Springer, Singapore, 2018, 65-82. https://doi.org/10.1007/978-981-10-8860-5_5

[26] Yang, J.: Understanding Chinese language teachers' beliefs about themselves and their students in an English context, The system, 80, 2019, 73-82. https://doi.org/10.1016/j.system.2018.10.014

[27] Yang, L.: Pragmatics learning and teaching in L2 Chinese, The Routledge handbook of Chinese second language acquisition, 2018, 261-278. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315670706-11/pragmatics-learning-teaching-l2-chinese-li-yang

[28] Zhu, Y.; Wang, B.: Investigating English language learners' beliefs about oral corrective feedback at Chinese universities: A large-scale survey, Language awareness, 28(2), 2019, 139-161. https://doi.org/10.1080/09658416.2019.1620755