



# A VR Interactive 3D Mandarin Pronunciation Teaching Model

Yujia Jin<sup>1</sup>, Yanlu Xie<sup>1</sup>, Jinsong Zhang<sup>1</sup>

<sup>1</sup>Beijing Language and Culture University, Beijing, China

blcu16wenyijyj@163.com, xieyanlu@blcu.edu.cn, jinsong.zhang@blcu.edu.cn

## Abstract

This paper presents a Virtual Reality interactive three-dimensional Mandarin pronunciation teaching model with higher clarity, more friendly interaction and more comfortable user experience. Our system provides four learning modes: initial, final, word and confusion. According to personal needs, learners can switch the demo views, customize visualization of 3D anatomical model's physiological structure, control play speed of 3D animation, and change position to observe movements and deformations of articulators. Moreover, the system provides assistant pronunciation guidance, such as the pronunciation method, 2D oral section animation, target zone of critical articulators, and contrastive models of confusion minimal pairs to help learners understand the essentials of Mandarin pronunciation.

**Index Terms:** 3D articulatory modeling, Mandarin pronunciation visualization, articulatory animation, Intelligent aided language learning, Virtual Reality

## 1. Introduction

In the context of global outbreak of COVID-19, artificial intelligence is developing rapidly in the field of education, online teaching is gradually becoming a popular learning method. In the absence of teaching aids and teachers' face-to-face demonstration and guidance, how to make learners quickly understand the essentials of pronunciation and imitate effectively becomes a question in this new situation. With the development of articulatory motion observation technology, some use modern phonetics instruments, such as EMA, EPG, X-ray, MRI and etc. to record 3D articulatory movements and construct 3D articulatory models. There are works which present 3D model in the form of video or animation in the pronunciation teaching system. Rathinavelu [1] constructed a 3D vocal tract articulatory model of Indian Tamil language and developed a Tamil pronunciation teaching system for hearing-impaired children. Wang [2] used a 3D virtual tutor in the pronunciation training system for ASD children. Li [3] combined the visual articulatory model with the training system of the IPA and confirmed the effectiveness of the 3D model through experiments.

3D oral modeling technology is greatly developed recently, however, due to the limitations of data collection and processing, the application of this technology in pronunciation teaching feedback is still less. In particular, there is still a lack of 3D articulatory modeling research for Mandarin pronunciation teaching feedback. Inspired by the Metaverse and wisdom education of VR field, this paper further explores 3D model based on pronunciation attributes and anatomy [4], we build a VR version of an interactive 3D Mandarin pronunciation teaching model with higher clarity, more friendly interaction and more comfortable user experience.

This model can be applied to C2L learners' Mandarin pronunciation teaching, language impairment treatment, children's pronunciation teaching and etc.

In this paper, the construction method of the model is described in section 2. Functions of the system are demonstrated in section 3. Section 4 shows the user guide of our system. Section 5 is the conclusion.

## 2. Method of model building

### 2.1. Mandarin EMA data collection and processing

The collected corpus is selected from Chinese Proficiency Grading Standards for International Chinese Language Education. There are 9 sensors in total participated in data collection. Articulatory features are extracted after the process of screening, annotating, filtering and smoothing.

### 2.2. Synthesis of model animation

Due to physiological similarity, a linear fitting method is used to approximate the mapping relationship between EMA data and model data.

In 3Ds MAX [5], an oral anatomical model which includes tuning organ, resonating cavity and sound source is constructed based on 3dbody model.

There are two kinds of animation, deformation of soft tissue and open-close movement of rigid-body. The positions of lattices and control points are adjusted to correspond with EMA data sampling points, and the matched model tracks are used to synthesize key-frame interpolation animations.

### 2.3. Implementation of model interaction

We design the interactive system in Unity [6], import SteamVR and VRTK plug-ins to realize the human-computer interaction in VR state, and develop a VR learning system.

## 3. Functions of the system

The learning page of the system is shown in Figure 1. The main functions of the system are as follows.

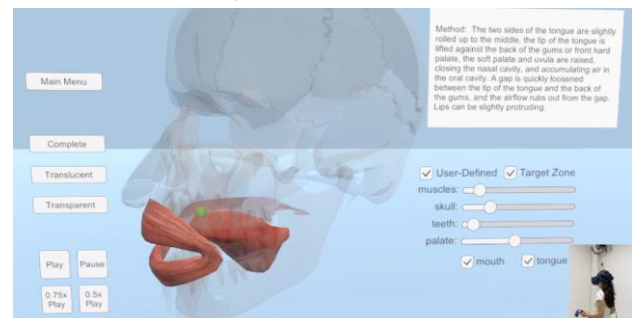


Figure 1: The learning page of the system.

### 3.1. Learning modes

There are four learning modes designed: initial, final, word and confusion. Word mode shows the pronunciation of pinyin which takes co-articulation into account. Confusion mode helps learners to distinguish the pronunciation of confusing minimal pair, such as l/r, s/sh by showing two 3D models and pronunciation methods at the same time.

### 3.2. Demo views

There are three kinds of default view in the system, which are complete, translucent and transparent. In the complete view, there is a complete head model including the skull, 14 facial muscles, teeth, lips and tongue. In the translucent view, the skeleton, muscles and other organs are hidden or been translucent. Users who use this view can clearly observe the relative position of tongue in the oral cavity, the mutual contact and speed change between active and passive articulators. Transparent view shows only the lips, tongue, hard palate and soft palate, allowing learners to further view the better articulation movements.

### 3.3. User-defined

Besides the three default demo views, learners can customize the visualization of 3D anatomical model's physiological structure by themselves. They can adjust the transparency of muscles, skull, teeth, hard palate and soft palate, and hide the view of lips and tongue according to their needs.

### 3.4. Assistant target zone

Spatial target zones corresponding to position of articulation can provide meaningful feedback [7]. The target zone is a red sphere area which turns green if the critical articulator comes into the area. It provides enhanced visual feedback to help learners quickly understand the essentials of pronunciation and get the target pronunciation position.

### 3.5. Animation player

Learners can control the play and pause of animation and audio. The system also provides "0.75x Play" and "0.5x Play" functions to slow down the animation, which help learners to observe the articulatory movements more clearly.

### 3.6. Movement

The 3D model can be scaled, moved, and rotated arbitrarily. Learners can adjust the position and direction of the model according to their demands, and dynamically observe the pronunciation process from multiple angles. If learners want to observe the articulatory movements more closely, the system also supports teleportation.

### 3.7. Other functions

A method panel with the pronunciation method and a 2D oral section animation is provided as guidance. After finishing the learning process, the system provides the learning notes to help learners review the content.

## 4. User guide of the system

The system is exported as an EXE executable file. Learners need to wear the HTC VIVE device and use left and right controllers to interact with the learning system.

- Press and hold the touchpad on the right controller to show a pointer. When the pointer touches the button on the UI, the color of the button changes. Then press the grip button on the right controller to click the button. The action can select the learning mode, switch the demo view, control the play of animation, and return to the main menu, etc.
- Touch the upper part of the left controller touchpad to move the model. Touch the lower part to rotate the model. Press trigger on the left handle to scale the model.
- Hold the pointer from right controller and press the grip button to check "user-defined" option. A custom panel with sliders and options appears. Use the pointer from right controller to slide the sliders and uncheck the tongue and mouth option.
- Use the pointer from right controller to check "target zone" option to show the red target zone.
- Aim the pointer from right controller at the ground to realize teleportation.
- Press the menu button on the left controller to show learning notes.

## 5. Conclusions

In this paper, a VR 3D Mandarin pronunciation teaching model is provided. It can be used as an innovative teaching tool in visual pronunciation training and feedback system. In the future, we will evaluate the effect of the model in Mandarin pronunciation teaching.

## 6. Acknowledgements

This work is supported by National social Science foundation of China (18BYY124), Center for Language Education and Cooperation (YHJC21YB-128), Science Foundation of Beijing Language and Culture University (the Fundamental Research Funds for the Central Universities) (19PT04), the Fundamental Research Funds for the Central Universities, and the Research Funds of Beijing Language and Culture University (22YCX099). The corresponding author of the paper is Yanlu Xie.

## 7. References

- [1] A. Rathinavelu, H. Thiagarajan, and A. Rajkumar, "Three Dimensional Articulator Model for Speech Acquisition by Children with Hearing Loss," in *4th International Conference on Universal Access in Human-Computer Interaction*, Beijing, China, Jul. 2007, pp. 786–794.
- [2] F. Chen, L. Wang, G. Peng, N. Yan and X. Pan, "Development and evaluation of a 3-D virtual pronunciation tutor for children with autism spectrum disorders," *PLOS ONE*, vol. 14, no. 1, pp. 1–22, Jan. 2019.
- [3] N. Zhi and A. Li, "Phonetic Training Based on Visualized Articulatory Model," *Journal of Foreign Languages*, vol. 43, no. 1, pp. 59–74, Jan. 2020.
- [4] X. Feng, Y. Xie, Y. Deng and B. Li, "A Dynamic 3D Pronunciation Teaching Model based on Pronunciation Attributes and Anatomy," in *INTERSPEECH 2020*, Shanghai, Beijing, Oct. 2020, pp. 1023–1024.
- [5] <https://www.autodesk.com/products/3ds-max/>
- [6] <https://unity.cn/>
- [7] W. Katz, T. Campbell, J. Wang, et al, "Opti-Speech: A real-time, 3D visual feedback system for speech training," in *INTERSPEECH 2014*, Singapore, September. 2014, pp. 1174–1178.