INTERSPEECH 2020: Show & Tell Contribution
October 25–29, 2020, Shanghai, China

A Dynamic 3D Pronunciation Teaching Model based on Pronunciation Attributes and Anatomy

Xiaoli Feng [1,2], *Yanlu Xie*[1] , *Yayue Deng*[1], *Boxue Li*[2]

[1] Beijing Language and Culture University, Beijing, China
[2] Yunfan Hailiang (Beijing) Technology co., LTD, Beijing, China
fengxiaoli314@163.com, xieyanlu@blcu.edu.cn,1420048330@qq.com, 1240146049@qq.com

Abstract

In this paper, a dynamic three dimensional (3D) head model is introduced which is built based on knowledge of (the human) anatomy and the theory of distinctive features. The model is used to help Chinese learners understand the exact location and method of the phoneme articulation intuitively. You can access the phonetic learning system, choose the target sound you want to learn and then watch the 3D dynamic animations of the phonemes. You can look at the lips, tongue, soft palate, uvula, and other dynamic vocal organs as well as teeth, gums, hard jaw, and other passive vocal organs from different angles. In this process, you can make the skin and some of the muscles semi-transparent, or zoom in or out the model to see the dynamic changes of articulators clearly. By looking at the 3D model, learners can find the exact location of each sound and imitate the pronunciation actions.

Index Terms: dynamic 3D pronunciation model, speech learning system, pronunciation attributes, anatomy

1. Introduction

Pronunciation has always been difficult for the learning and teaching of Chinese as a second language(C2L). In traditional classroom, experienced Chinese teachers often adopt phonological attribute-based approaches for pronunciation-teaching, such as using the "blowing paper method" for the teaching of aspirated consonants. With the application of new techniques in research, physiological phonetics continues to develop. Two-dimensional physiological data of articulators' movements were obtained by X-ray or MIR. [1,2] The movement of the lips and the tongue can be tracked when speakers produce the particular sound through the acquisition of EMA data. [3] The research results of physiological phonetics provide another more intuitive way of thinking for pronunciation teaching, namely 2D/3D articulators imaging method. This method is more suitable for an online learning environment without direct teacher involvement or feedback in computer-aided pronunciation teaching.

It is still difficult to obtain physiological data and apply it to pronunciation teaching directly. Based on the 3D-body's head and neck model [4], we consider developing a dynamic 3D pronunciation teaching system for Chinese teachers and Chinese as a second language (C2L) learners by combining 3D modeling technique, head and neck anatomy data and phonology knowledge together. The system contains the oral 3D animations of 21 initials and 38 finals in Chinese. Chinese teachers and C2L learners can see the vocal cords and articulators in the mouth from different angles. And they can choose to hide the facial skin and parts of muscles or cause them to be semi-transparent, or enlarge or shrink the model to make the specific articulators more visible.

In the field of computer assisted pronunciation teaching (CAPT), the Goodness of Pronunciation (GOP) is a common feedback method. It is difficult for learners to understand their pronunciation problems simply with a GOP score.[5] After the pronunciation error recognition, the 3D animation feedback can help the learners to correct the pronunciation more intuitively. In addition, the model can also be applied to children's pronunciation teaching and language rehabilitation.

In the paper, knowledge background and technical details of model development will be present in section 2. The function application of the model is described in section 3. Section 4 is the conclusion.

2. Method of model building

2.1. Chinese initial and final system

21 initials and 38 finals are contained in Mandarin Chinese system. [6]

- According to articulation methods, Chinese initials can be divided into six categories: stop, affricate, fricative, nasal, lateral, and approximate. Among them, stop and affricate consonants have the opposite of aspirated and unaspirated, such as /p/ans/ph/. According to the place of articulation, Chinese initials can be divided into bilabial, dentil-labials, apical, retroflex, lingua-palatal, and velar consonants.
- According to the characteristics of the composition of finals, Chinese vowels can be divided into single vowels, compound vowels, and nasal vowels. The articulation of a single sound is mainly described by three dimensions: the height of the tongue position, the front or back of the tongue position, and the round or spread of the lip shape. Compound and nasal vowels should consider the main vowel sounds and the role of co-articulation between phonemes.

From the above classification of Chinese initials and finals, we can see that it is possible to describe the phonemes through oral physiological characteristics. What, then, is the physiological basis for phonological diversity?

2.2. The physiological basis of pronunciation

Air flow from the lungs is the motor of speech production. As a whole, the speech system includes two aspects: phonation and articulation. [7]

Anatomically speaking, the phonation organs mainly refer to the larynx and vocal cords, including thyroid cartilage, arytenoid cartilage, cricoid cartilage, epiglottis cartilage, and vocal cords. The muscles of the larynx mainly include the

thyroarytenoid muscle, cricothyroid muscle, lateral cricoarytenoid muscle, posterior cricoarytenoid muscle, interarytenoid muscle, and aryepiglottic muscle.

As shown in Figure 1, the articulation organs mainly include the upper lip, lower lip, upper dental, lower dental, gingiva, hard palate, soft palate, uvula, the tip of the tongue, blade of the tongue, lingual surface, lingual root, oral cavity, and nasal cavity, etc. The movement of the tuning organ is the basis for the production of phonetic diversity.

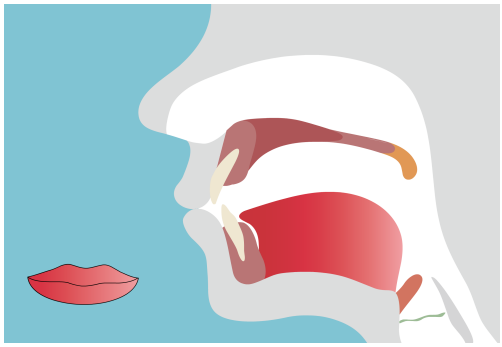

Figure 1: *Schematic diagram of articulatory organs*

2.3. Technical details

Users can download the toolkit "Mega-Fires 3.48 Toolkit",[8] then add the “Modifiers” attribute to the facial muscles, select FFD3*3*3 or FFD4*4*4, and then select the “Scripts” attribute and Mega FFD Animate. After that, add the corresponding animation script to the muscles, create an animation in the animation property panel, and then add the property of “add property”. If you select FFD3*3*3, you will add a total of 27 points from p00 to p26. and if you choose FFD4*4*4, you will add a total of 64 points from p00 to p63. At this time, the muscle is wrapped by a cube adjustment frame. Each added point corresponds to a point of the adjustment frame. By adjusting the x, y, and z coordinates of each point, the deformation of the muscle can be controlled.

For the process of optimizing voice animation and interaction, considering the situation that the articulatory organs in the oral cavity are not completely visible, or even completely invisible, we try to make an interactive design of the model from the perspective of the pronunciation organ visualization in order to allow the users to observe the movement of the pronunciation organs more clearly. For dental, hard palate, and mandible, which only produce slight deformations during pronunciation, or even do not produce deformation at all, we consider them as rigid bodies and perform motion simulation. For tongues and soft palate, which produce a large amount of deformation during pronunciation, we try to simulate its deformation effect and use the collected Electromagnetic Articulography (EMA) data to drive and control its movement.

3. The using of the 3D learning system

21 initials and 38 finals of Chinese are contained in this 3D pronunciation learning system. The main functions of the 3D model are as follows:

1. The model can be rotated to view the articulatory organs from different angles;

2. The model can be zoomed in and out; and,

3. The facial skin and some muscles can be made translucent.

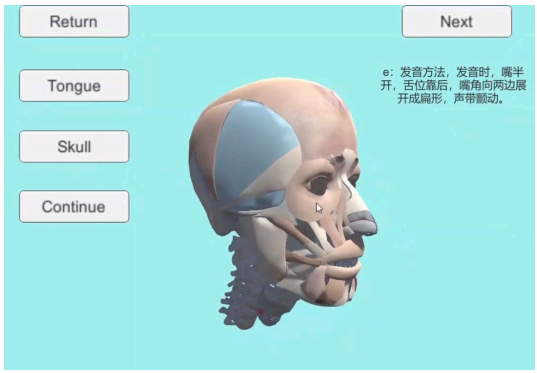

Figure 2: *The interface of the 3D learning system*

Specific instructions will be given in reference to Figure 2. Choose a learning object first.

- Click "Pause" to stop sound playback.
- Click “Continue" to restart the animation.
- Slide the mouse wheel to zoom in or out the model.
- Press and hold the left mouse button to pan the model.
- Hold the right mouse button to rotate the model.
- Click "Skull", "Tongue" and "Head" to show the 3D animation of the three states of the model respectively.
- Click"Return" will return to the previous menu.

4. Conclusions

In this paper, a dynamic 3D articulation model is explained. The model can be used as a tool to help the Chinese teachers and C2L learners understand the pronunciation of each Chinese initial and final intuitively.

5. Acknowledgments

This work is supported by the National Social Science Foundation of China (18BYY124), Advanced Innovation Center for Language Resource and Intelligence (KYR17005), Discipline Team Support Program of Beijing Language and Culture University (GF201906), Education and teaching reform project of Beijing Language and Culture University (XJGZ201904) and Wutong Innovation Platform of Beijing Language and Culture University (19PT04). The corresponding author of the paper is Yanlu Xie.

6. References

[1] Zhou Dianfu, Wu Zongji, Mandarin Pronunciation map, 1963.
[2] Wang Gaowu, Kong. JP, Bao, HQ, “Formant Prediction from Vocal Tract of Mandarin “, Chinese Journal of Phonetics, The Commercial Press，2008.
[3] Hu Fang, On the lingual mechanism in vowel production in Ningbo Chinese, Chinese Journal of Phonetics, The Commercial Press，2006.
[4] www.3dbody.cn
[5] S. M. Witt, and S. J. Young. "Phone-level pronunciation scoring and assessment for interactive language learning." Speech Communication, vol. 30, no. 2, p. 95-108, 2000.
[6] Lin Tao, Wang Lijia, YuyinxueJjiaocheng(revised edition), Peking University Press, p,105-120,2013.
[7] Kong Jiangping, Shiyan Yuyinxue Jichu Jiaocheng, Peking University Press, p8-15.2015.
[8] https://assetstore.unity.com/packages/tools/modeling/mega-fiers-644.