A study on multi-speech models of Mandarin

and multi-media learning system

KONG Jiangping

Abstract

This paper introduces the multi-speech modality research of Mandarin in the department of Chinese language and literature, Peking University and discusses the role and possibility in establishing a multi-media learning system of Mandarin. In the study on speech production of Mandarin, five basic models which are 1) the model of vocal fold vibration established by high-speed digital imaging; 2) the model of dynamic vocal tract establishes through X-ray and MRI, 3) the model of lip motion set through video and motion capture, 4) the model of palatal contact studied through electropalatography and 5) the model of speech aspiration studied through an respiration belt are introduced. The advantages and possibility of these models used for Mandarin learning are discussed including the teaching and learning for people with hearing and pronouncing problems. Finally the application prospects in multi-media teaching and learning system of Mandarin are talked.

Keywords:

Mandarin, multi models of speech production, multi-media learning system

1. Introduction

With the development of China, the activities of economics and cultures have been increasing rapidly. So people of the world have been paying more and more attention to the learning of Mandarin which is also an official language used in the United Nations. Mandarin belonging to the Sino-Tibetan language family is a typical tonal language and has many speech characteristics in speech physiology, acoustics and psychology. As a standard Chinese spoken language, studying on the Mandarin speech production physiologically, acoustically and psychologically is very important for its teaching and learning.

In the traditional education, people tends to regard the education as a kind of "art" but not "science" or a kind of scientific methods. At present, since the speech technology and internet develop fast, many techniques and methods are used in establishing e-teaching system which leads language teaching and learning into a new field. As to the spoken language teaching and learning, the study on multi models in speech production scientifically and technologically can improve the way of teaching and learning and set up new educational system.

The spoken language acquisition of second language has its own property which has close relationship with speech production and the modeling is well benefit from the new techniques used in recent ten years. In Peking University, the multi models of Mandarin had been studied for almost ten years and the models of vocal fold vibration, vocal tract, lip motion, palatal contact and speech aspiration were established. In this paper, the multi-speech models of Mandarin are briefly introduced and the application prospects in Mandarin teaching and learning system are discussed.

2. Model of dynamic vocal tract

X-ray was used in studying on vocal tract when it was just invented. Through X-ray, phoneticians got to know the activities of human’s speech organs and defined vowels by mouth openings and tongue positions. In China, only one set of X-ray materials including 250 single syllables and disyllables of 3 persons, 1 male subject and 2 female subjects, was captured by Professor Bao
Huaiqiao in 1970s. Since X-ray is invasive, these materials are very valuable in studying the movement of articulation in Mandarin.

With the development of image signal processing, it is relative easy to process these videos and detect the edges of vocal tracts and tongues of different vowels. A database of Mandarin vocal tract was established by 32000 frames of images by which a 2D model of dynamic vocal tract in Mandarin was set up. In this model, the speech organs were divided into 6 parts which are: 1) lips; 2) mouth; 3) soft palate; 4) tongue, 5) tongue tip and 6) vocal folds and 12 parameters were used to drive the model. See Figure 1. There are 4 plots in figure 1. which displays the vocal tract of initial /b/, the vocal tract of vowel /a/, the 3D speech organs by MRI and the separated 3D speech organs.

![Figure 1: Vocal tracts of Mandarin captured by X-ray and MRI](image)

In the teaching of Mandarin pronunciation, pictures of speech organs are often used in class and the disadvantage is that they are not able to show the movement of articulations. Based on the model of dynamic vocal tract, a teaching system for Mandarin pronunciation was established, in which the Mandarin articulation was carefully described by images of real vocal tracts and videos. It is a very useful and convenient tool and has plenty of materials for both teachers and second language learners. In addition, the system is also useful for Chinese children who have partially hearing or hearing problems to learn the pronunciation of Mandarin in the first language acquisition.

3. Model of vocal fold vibration

Since the vocal folds are in the throat and can't be easily observed directly, the phonetic study on phonation types developed much late than that of articulation. Since the technique of high-speed imaging was used in studying on phonation types and the vibration procedure of vocal folds can be observed by eyes, phonation types of different languages were greatly understood. The fundamental frequency (F0) and its contours were usually regarded as the only distinguish feature in Mandarin tones, and now people found that the phonations, especially the phonation in the third tone (shang sheng) was a kind of creaky voice which was significant to intelligibility and naturalness of Mandarin.

The high-speed image samples of 4 basic Mandarin tones of 3 males were captured by fiberscope and the high-speed image samples of 8 persons, 4 males and 4 females, were captured by endoscope, while the speech sound and electroglottography signal were sampled simultaneously. After signal processing, three kinds of parameters can be extracted from these 3 signals. Based on these parameters, phonation models of Mandarin tones can be established. See figure 2. There are 2 plots for each tone in figure 2 in which the upper display the parameters of F0 and speed quotient (SQ) which is defined as opening phase over closing phase and the lower displays the parameter of open quotient (OQ) which is defined as open phase over period.
Figure 2: Phonation parameters of 4 basic tones in Mandarin.

From the phonation patterns of tones, we found that OQ and SQ changed along with F0. In the learning of Mandarin tones, all attention had been paid to the change of pitch and the phonation had been disregardful. Now we know that phonations are very important in learning Mandarin tones, especially the third tone in which creaky voice is the most important thing that a learner has to imitate. By taking advantage of modern techniques, visual phonation feedback should be designed in a Mandarin learning system to help the learners. With the visual feedback of F0, OQ and SQ, phonations, it is not only helpful for the normal Mandarin learner to learn tones, but also very useful and helpful for the partially hearing people to learn tones. In addition, phonation type visual feedback in a learning system will also be helpful for singers to imitate different phonation types in different operas or original folk songs.

4. Model of lip motion

From the viewpoint of speech communication, lip reading or speech reading is very important in learning of spoken language and lip reading is obliged to the language teaching of deaf mute child. From the viewpoint of linguistics, viseme is defined as a unit which is significant in distinguish meanings. The results of present research on viseme show that static viseme was not very useful in spoken language learning and the dynamic viseme was more useful and significant in language learning. So dynamic lip reading should be designed in the spoken language teaching system which needs multi-media technique.

Mandarin lip motion has been studied through video materials and image signal processing. In this study, video of 4000 basic Chinese words were captured by video camera and motion capture and the contours of inner and outer lips were detected for setting up a database. With this parameter database, a model of lip motion was established which could be used in a Mandarin teaching and learning system. See Figure 3. There are 3 plots in figure 3, in which the first plot displays a lip image with detected contours, the second plot displays the definition of the model, and the third plot displays the 3D parameters sampled by motion capture.

Figure 3: Lip parameters and definition of Mandarin.

According to our study, the dynamic viseme produced by the model of lip contours would lose information of speech in communication. So a 3D model of lips should be studied and established for a multi-media Mandarin teaching and learning system which is not only necessary for normal language learners, but also for the learners who need lip reading in their work.

5. Palatal contact by EPG

One of the methods in the study of Mandarin articulation is to capture the signal of palatal contact area by EPG, an instrument often used in the research of physiology. According our research, we found that it was a very good method to study on the articulation of Mandarin, especially the co-articulation between consonants and vowels in one syllable and the co-articulations in running speech. However, this instrument was originally designed for the training of child with cleft palate after medical operation and an artificial palate whose cost was not low should be made for each user beforehand. All this impedes the application of EGP in normal language training.
A database of EPG in Mandarin was established in the phonetic lab in Peking university and the co-articulations within syllables and between syllables were studied. According to the research results, the language training on cleft child is necessary after medical operations and the method of EPG can also be used in the training of normal learners of Mandarin, if the language training system with EPG is low.

6. Model of speech respiration

Respiration is the power of speech and was rarely studied phonetically because there was no appropriate instrument. In recent 10 years, an respiration belt which was an instrument in the study of human physiology, was used in the phonetic study on reading aloud, sutra chanting and oral cultures in China. Some results show: 1) people usually would use both costal respiration and abdominal respiration in different kinds of speech communication and oral culture performance; the abdominal respiration would mainly provide power of speech and the costal respiration had more close relationship with articulation; 3) different patterns of respiration were used in different languages because of their different syntax structures.

In the study of respiration in Mandarin, two respiration belts were used to capture the signals of costal respiration and abdominal respiration. We found that there were at least 3 kinds of respiration resets in reading Chinese seven quatrains (七言绝句, 七言律诗) and at least 4 respiration resets in broadcasting news. See figure 5.

7. Application prospects of multi-media teaching system

It is a new task to put basic phonetic results and models into use of language teaching and learning system. From the viewpoint of basic phonetic study, there are two aspects which we should consider in a multi-media spoken Mandarin system, one is demonstration and the other is visual feedback. As to the first aspect, acoustical and physiological parameters or models can display speech sound in different forms. The acoustical parameters of tones and diatones can be displayed by F0, dynamic phonation types by SQ and OQ, vowels by F1 and F2 in an acoustical chart, consonants by the parameters of palate contact area and respiration rhythm by signals of respiration. The physiological demonstrations can be implemented by the models of vocal tract driven by parameters of X-ray and MRI videos, by the model of vocal folds driven by parameters of high-speed images, by the model of dynamic lip driven by parameters of videos and the by the model of respiration driven by parameters of the signals of respiration. As to the second aspect, the visual feedback of speech sound, especially the visual feedback of speech physiology is very indispensable in a multi-media language system. Physiological visual feedback of a learner's speech sound can be regarded a real-time test by which learners can judge himself in speech sound learning. At present, the implementation of speech sound visual feedback still has many technical problems in speech signal processing and physiological signal processing. However, we can display...
the acoustic speech feedback of tones by F0 or the feedback of vowels by F1 and F2 in acoustic vowel chart. But it is still difficult to display dynamic vocal tract or vibration of vocal folds of speech sounds pronounced by Mandarin learners.

8. References


