AN ARTICULATORY MODEL OF
STANDARD CHINESE USING MRI AND X-RAY MOVIE

Gaowu Wang*  
Peking University  
Beijing Normal University

Jiangping Kong†  
Peking University

ABSTRACT
To better understand speech production from the phonological inputs to articulatory movements and then to acoustic outputs, it is important to establish an elaborate articulatory model of the vocal tract. This paper has explored the articulatory mechanism of speech production in Standard Chinese and developed a geometric articulatory model in both the visual and acoustic modalities. This model was based on the data of MRI images and X-ray movie, with the former providing detailed volumetric information of the vocal tract, and the latter the dynamic information of articulation. In this model, the seven articulators have been studied and modeled, including the hard palate, pharynx, jaw, lips, velum, tongue, and larynx. The tongue is modeled as two parts: tongue tip and tongue body, thus reducing the necessary number of parameters. The relation between larynx height and fundamental frequency in regard to the four tones is also modeled. These two improvements on tongue and larynx modeling have contributed new ideas to the articulatory modeling of Standard Chinese. This model can serve as a research tool for linguists, phoneticians, and speech engineers, and can be used in parameter speech synthesis, virtual speaker, and visual assistant speech training of Standard Chinese.

* Gaowu Wang 汪高武 conducts research at The Department of Chinese Language and Literature, Peking University and completed this paper at Beijing Normal University. He also teaches at Beijing Normal University.
† Jiangping Kong 孔江平 is a professor of Peking University at Center for Chinese Linguistics, Department of Chinese Language and Literature, and Joint Research Center for Language and Human Complexity. He is the corresponding author of the article: Dept. of Chinese Language and Literature, Peking University, Beijing 100871, China; [jpkong@pku.edu.cn]
In this special issue, 10 papers on physiological aspects of phonetics are presented. There are four papers on speech models, including three papers for geometrical models of Mandarin, and one paper for the physiological articulatory model. “An Articulatory Model of Standard Chinese Using MRI and X-ray” has explored the articulatory mechanism of speech production in Standard Chinese and developed a geometrical articulatory model, both in visual and acoustic modalities, based on the data of MRI images and an X-ray movie. “A Two-Dimension Lip Model for Mandarin Chinese”, established a two dimensions lip model with inner and outer lip contours which has well defined the linguistic term 'lip rounding' and was used to generate audio-visual stimuli for the speech perception experiment of the McGurk Effect. “A Dynamic Glottal Model through High-speed Imaging” has introduced a dynamic glottal model based on high-speed imaging and the model controlled by dynamic glottal widths, lengths, F0, Open Quotient (OQ) and Speed Quotient (SQ) which can produce speech sources with different phonation types. “A Control Strategy of a Physiological Articulatory Model for Speech Production” has constructed a full three-dimensional physiological-articulatory model, including the tongue, jaw, hyoid bone and vocal tract wall, based on the continuum finite element method.

There are three papers on physiological phonetic studies that used electropalatography (EPG), instruments for air-pressure and flow, and aspiration. “Prosodic Boundaries Effect on Segment Articulation in Standard Chinese: An Articulatory and Acoustic Study” has investigated the EPG and the acoustic data of the prosodic boundaries effect on the domain-initial segments in Standard Chinese, aiming to examine the domain-initial strengthening in both spatial and temporal dimensions. “A Study on the Features of Chest and Abdominal Breathing between Reciting and Chanting Chinese Poetry” has studied the features of chest and abdominal breathing between reciting and chanting Chinese poems of different styles. “An Aerodynamic Study on Articulation of Mandarin Initials” has studied the aerodynamic features of Mandarin initial consonants with different articulatory places and manners and found that the parameters of expiratory airflow duration (EAD), peak air-pressure (PAP), peak expiratory airflow (PEA) and expiratory volume (EV) can all

1. INTRODUCTION

Speech is the important tool for humankind's thinking and communication, therefore the study of speech sounds, known as phonetics, has fundamentally theoretical and practical value. With the development of advanced instruments and scientific methods, traditional phonetics can overcome its limitations of "by mouth and ear" (Luo 1956) to develop into modern experimental phonetics, thus bridging the gap between the traditional language culture and modern scientific study. In modern phonetics, speech production has become the frontier, since it involves interdisciplinary cooperation, including linguistics, phonetics, psychology, acoustics, computer engineering, and other fields.

To better understand speech production from the phonological inputs to articulatory movements and then to acoustic outputs, it is important to establish an elaborate articulatory model of the vocal tract. At present, this topic has attracted broad international interests, based on the speech materials of many languages, including English, French, German, Japanese, Swedish, and others, but few from Chinese (Wang et al. 2008). And there are few advanced studies to establish an articulatory model embracing the specific properties of Chinese.

This paper serves as exploratory research in the field of speech production modeling in Standard Chinese, as it has established an articulatory physiological geometric model of Standard Chinese. In order to establish the model, a database of articulatory movements of Standard Chinese was established, including MRI images and X-ray movie, for which the former provides detailed volumetric information of the vocal tract and the latter the dynamic information of articulation.

2. MRI AND X-RAY DATABASE
2.1. MRI Experiment

Magnetic Resonance Imaging (MRI) allows a tomographic view
In this special issue, 10 papers on physiological aspects of phonetics are presented. There are four papers on speech models, including three papers for geometrical models of Mandarin, and one paper for the physiological articulatory model. "An Articulatory Model of Standard Chinese Using MRI and X-ray" has explored the articulatory mechanism of speech production in Standard Chinese and developed a geometrical articulatory model, both in visual and acoustic modalities, based on the data of MRI images and an X-ray movie. "A Two-Dimension Lip Model for Mandarin Chinese", established a two dimensions lip model with inner and outer lip contours which has well defined the linguistic term 'lip rounding' and was used to generate audio-visual stimuli for the speech perception experiment of the McGurk Effect. "A Dynamic Glottal Model through High-speed Imaging" has introduced a dynamic glottal model based on high-speed imaging and the model controlled by dynamic glottal widths, lengths, F0, Open Quotient (OQ) and Speed Quotient (SQ) which can produce speech sources with different phonation types. "A Control Strategy of a Physiological Articulatory Model for Speech Production" has constructed a full three -dimensional physiological-articulatory model, including the tongue, jaw, hyoid bone and vocal tract wall, based on the continuum finite element method.

There are three papers on physiological phonetic studies that used electropalatography (EPG), instruments for air-pressure and flow, and aspiration. "Prosodic Boundaries Effect on Segment Articulation in Standard Chinese: An Articulatory and Acoustic Study" has investigated the EPG and the acoustic data of the prosodic boundaries effect on the domain-initial segments in Standard Chinese, aiming to examine the domain-initial strengthening in both spatial and temporal dimensions. "A Study on the Features of Chest and Abdominal Breathing between Reciting and Chanting Chinese Poetry" has studied the features of chest and abdominal breathing between reciting and chanting Chinese poems of different styles. "An Aerodynamic Study on Articulation of Mandarin Initials" has studied the aerodynamic features of Mandarin initial consonants with different articulatory places and manners and found that the parameters of expiratory airflow duration (EAD), peak air-pressure (PAP), peak expiratory airflow (PEA) and expiratory volume (EV) can all

---


用磁共振成像和 X 光聲道資料建立漢語普通話調音模型

汪高武* 孔江平†

北京大學 北京大學

北京師範大學

提要

爲了更好地理解言語的產生過程，即如何從音位輸入到調音器官的動作、再到聲學輸出，需要建立詳盡的聲道調音模型。本文是對漢語普通話言語產生中調音機制的探索與研究，建立了一個具有視覺和聲學輸出的幾何調音模型。該調音模型的數據源自於聲道的磁共振圖像和 X 光錄影，前者主要提供聲道的立體形狀，後者提供調音的動態過程。在這個模型中，聲道被分解為七個調音部位進行研究：硬齶、喉腔後壁、下顎、雙唇、軟齶、舌頭和喉管。其中創新性地，舌頭又被分為舌體（相對簡單）和舌尖（更為靈活）兩部份，從而簡化所需要的參數；另外，普通話四聲中基頻高低與喉管上下高度的關係也加入到模型中。該模型可以作為研究工具服務於語言學、語音學和語言工程，並可用於語音參數合成、虛擬說話人、普通話輔助教學等領域。

主題詞

言語產生 調音模型 声道 磁共振成像 X 光

* 汪高武 教授於北京師範大學並治學於北京大學中文系。
† 孔江平 教授於北京大學教授執教研究於中文系，中國語言學研究中心，語言與人類複雜系統聯合研究中心，並為本文聯繫作者：中國北京，郵編 100871，北京大學中文系；[jpkong@pku.edu.cn]