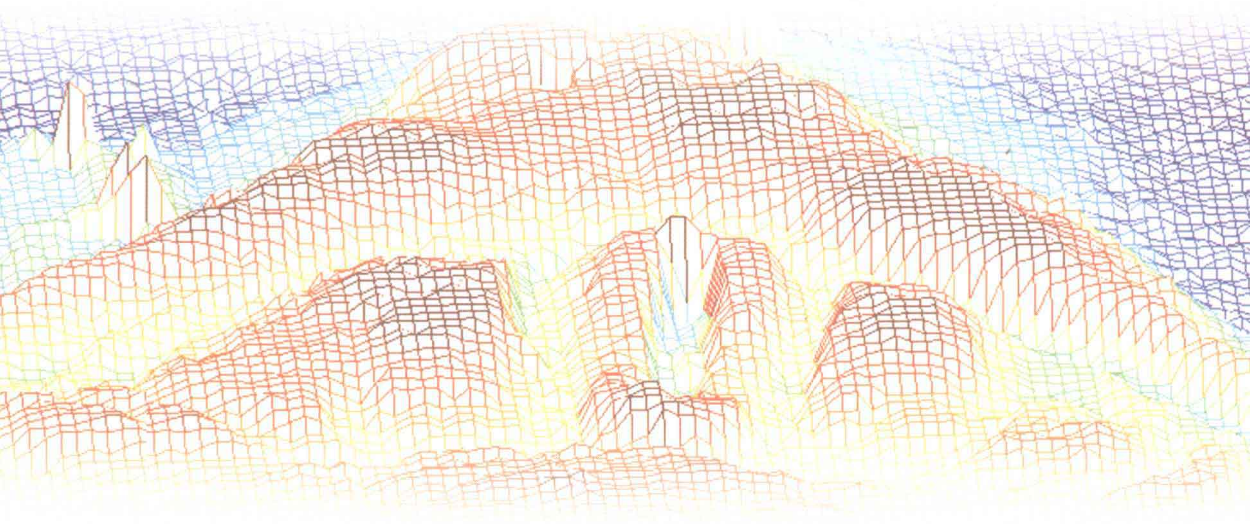
LARYNGEAL DYNAMICS AND PHYSIOLOGICAL MODELS

—— High Speed Imaging and Acoustical Techniques

动态声门与生理模型

KONG JIANGPING



北京大学出版社
PEKING UNIVERSITY PRESS

Laryngeal Dynamics and Physiological Models

——High Speed Imaging and Acoustical Techniques

动态声门与生理模型

Kong Jiangping

北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

Laryngeal Dynamics and Physiological Models: High Speed Imaging and Acoustical Techniques＝动态声门与生理模型/Kong Jiangping 著. —北京:北京大学出版社,2007.11

ISBN 978-7-301-13672-0

Ⅰ.动… Ⅱ.K… Ⅲ.发声法-研究-英文 Ⅳ.H0818.1

中国版本图书馆 CIP 数据核字(2008)第 053541 号

书　　名: **Laryngeal Dynamics and Physiological Models: High Speed Imaging and Acoustical Techniques**
动态声门与生理模型

著作责任者: Kong Jiangping 著

责 任 编 辑: 白 雪 杜若明

标 准 书 号: ISBN 978-7-301-13672-0/H・1968

出 版 发 行: 北京大学出版社

地　　址: 北京市海淀区成府路 205 号 100871

网　　址: http://www.pup.cn

电 子 邮 箱: zpup@pup.pku.edu.cn

电　　话: 邮购部 62752015 发行部 62750672 编辑部 62752028
出版部 62754962

印 刷 者: 北京宏伟双华印刷有限公司

经 销 者: 新华书店

730 毫米×980 毫米 16 开本 18.5 印张 285 千字

2007 年 11 月第 1 版 2007 年 11 月第 1 次印刷

定　　价: 38.00 元

世人以竹木牙骨之類爲叫子置人喉中吹之能作人言謂之顙叫子嘗有病瘖者爲人所苦煩冤無以自言聽訟者試取叫子令顙之作聲如傀儡子粗能辨其一二其冤獲伸此亦可記也

Devices made by people from materials, such as bamboo, wood, ivory and bone, are called sound generators. A sound generator, which can be put into the throat and produce speech sound by whistling, is called a voice generator. A dumb person, who suffered from injustice, could not argue in court for himself. The judge let people put a voice generator in his throat, and asked him to speak. The speech articulated like puppet talk, but could roughly make sense. His injustice was finally redressed. The case is worthy of being documented.

From volume 13.1 of an ancient Chinese book titled "*Mengxi Bi Tan*" (writing in the Mengxi Garden) by Shen Kuo (1031—1095). The story above is translated by Kong Jiangping, 2000.

To Professor William S-Y Wang, I can not enter this new area without his encouragement and help.

Jiangping

Foreword

Phonetic sciences have progressed remarkably over recent decades, dating to the introduction of the sound spectrograph and the source-filter theory of speech production in mid-20th century. However, the great bulk of the new knowledge is focussed on the movements of the tongue, jaw and palate, that is, on the segmental aspects of speech. While no one would deny the importance of what the larynx contributes to the speech process, the source of the source-filter partnership, our knowledge here remains sketchy and imprecise.

This relative neglect of the voice is no doubt due to the great difficulty in observing phonation directly. The vocal folds are well hidden at the base of the throat, and they vibrate at frequencies which must be slowed down considerably before they can be seen, measured, and analyzed. The research needs to build upon the technology of high speed photography via endoscopy, and of the extensive digital processing which ensues, which is a daunting challenge.

Professor KONG Jiangping took up this challenge as his PhD project, directly after he came to Hong Kong to study with me. His work took him to Tokyo University and to the Royal Institute of Technology in Stockholm, to gather physiological data on phonation, and to learn new methods of analyzing these data. He has also frequently travelled to all-but-inaccessible villages in China's rugged southwest to do linguistic fieldwork, recording the priceless speech of many languages on the verge on extinction. This volume is the fruit of much of this dedicated labor.

The volume is a major contribution to our knowledge of phonation-its various modes of vibration and detailed mechanisms for their production in selected languages. As such it provides valuable knowledge to areas of medicine concerned with the voice and its disorders. It is also important to areas of the arts concerned with how the voice is harnessed for a diversity of astonishing effects in various musical cultures.

There are of course many other applications of the new knowledge contained in this volume that one can think of. From the perspective of language evolution, which is my own area of interest, the new knowledge promises a deeper understanding of how tone languages arise. Some 40 years ago, in discussing the phonological features of tone, I wrote:

"They have a particularly close relation, synchronically and diachronically, with features which are controlled primarily at the larynx, e. g., voicing, aspiration, glottalization, length, breathiness, etc. This relation is easily understandable since the primary determinants of tone is the rate of laryngeal vibration."

Now Professor Kong has not only deepened our understanding of laryngeal vibration by physiological research, he has also collected much data on Tibeto-Burman languages in China's southwest. These languages range over the entire spectrum of being non-tonal to more or less tonal to fully tonal, such as Chinese. To satisfactorily answer the question of how languages become tonal, we need to precisely plot the evolutionary trajectory of these languages, bearing in mind all the relevant laryngeal activities as they interact with the segmental systems. The same methodology can be applied, of course, to other areas of the world where tone languages are spoken. I believe that this volume opens up an important new avenue of linguistic research, and recommend it to the field with enthusiasm.

William S -Y. Wang
Academia Sinica, Taibei
Chinese University of Hong Kong
International Institute of Advanced Studies, Kyoto

Preface

The development of science depends on two aspects, one is the discovery of the nature of the world, and the other is the application of new technology. In the research of voice science, many techniques have been developed, such as: 1) fiberoptic observation, 2) laryngeal electromyography, 3) photoglottography, 4) electrog-lottography, 5) high-speed digital imaging, and 6) inverse filtering in acoustical analysis. Among these techniques, high-speed digital imaging is the one with the most potential in the recent decade.

The technique of high-speed filming appeared in the 1930s, which had been first used in the study of vibration of vocal folds in Bell Telephone Laboratories in 1937. Since it is extremely costly and complex in operation, it has not been widely used in the research of vibration of vocal folds and the diagnosis of pathological voice in hospitals. With the development of video technology and computer, high-speed digital imaging has replaced the position of high-speed filming in the area of voice science. At present, most of the studies have been done in hospitals by medical doctors who observe and explain the cases of pathological voice, and only very few parameters have been extracted and used for systematic research.

The questions that can be raised are: 'Can high-speed digital imaging be used to study the vibration of vocal folds quantitatively?' and 'Can this technique be used to establish a physiological model based on the parameters of dynamic glottis?' To answer these questions is the aim of this research. The research has two parts: 1) to study and quantify the dynamic glottis of different phonation types by developing an effective platform of image processing and parameter extraction, and 2) to study the relationships between the dynamic glottis and the glottal source by establishing a model to simulate the vibration of vocal folds.

The objectives we are pursuing are 1) processing the image and extracting the parameters automatically, 2) describing the properties of different phonation types quantitatively, and 3) modeling the dynamic glottis mathematically. In accordance with these objectives, the book is divided into 4 parts and has 10 chapters. The 4 parts are: 1) image processing and parameter extraction, 2) analysis of different phonation types, 3) the nature of phonation and quantification and 4) model construction and glottal synthesis.

A high-speed digital image system called 'VoiceLab_Image' has been established for the study of vibration of vocal folds through the dynamic glottis. After the analysis and study on different phonation types, it has been confirmed that the system is an effective and useful tool. It has functions to process the images automatically, such as image rotation, contrast adjustment, shift adjustment and area detection. It can extract parameters automatically and calculate them statistically. The parameters extracted by this system cannot be obtained by other techniques, such as the methods used in speech acoustics and physiology. There are 24 basic parameters, such as F0, the left and right glottal area functions, the speed and open quotients of the glottis, and the direct current component. Of these parameters, ten parameters are calculated statistically, such as the index of left and right glottal area balance, the index of symmetry of the dynamic glottis, and average ratio of glottal shape.

Based on these parameters, different phonation types have been analyzed and explained. In this book, three types of materials of the high-speed digital images are analyzed and studied. They are 1) the modal voice with different pitches, 2) the special phonations including falsetto, vocal fry, breathy voice, inspiration and diplophonia, and 3) the phonations with gliding pitch and in the basic tones in Mandarin. Many properties and features have been found and specified through the analysis. Then, 3 modes of dynamic glottis, 7 types of glottal pulse and 8 types of glottal period, such as the leakage and open modes, the double and triple glottal pulses and the supra-glottal period, are defined. Based on these, the definition of fundamental frequency and the method of decomposing the asymmetrical glottal area are discussed.

The quantification of phonations through the high-speed digital imaging is part of the research objective. In this study, three kinds of quantification have been discussed. The first is the quantification of phonation types. The second is the quantification of parameters. The third is the quantification of the relationships among the parameters. The phonation types have been quantified through the averaged statistical parameters obtained by using the method of correlation analysis and classification. The result shows that the phonation types can be specified and defined by parameters of high-speed images quantitatively. The different types of parameters are also quantified by the methods of correlation analysis and classification. The result indicates that the parameters can be considered as two types, the basic parameters and the individual parameters. The basic parameters are defined as those, which have relationship with each other, and the individual parameters are defined as those, which have no relationship with other parameters. The quantification of relationships among parameters has been done by establishing relation functions through polynomial curve fitting. Fourteen relation functions of parameter pairs are finally established, which reflect the natures of different phonation types and form the basis of the physiological model.

A physiological model of dynamic glottis has been established on the model of static glottis in three modes and the model of the dynamic parameter control. The model of the static glottis has been established by the basic parameters of the glottis through the basic mathematical formulas of ellipse, square, rectangle, triangle and trapezoid. The models of the parameter control function are constructed by the parabola function and the function approximated by sinusoid.

Based on the physiological model, a synthesis system of dynamic glottis has been established to study the relationship between the dynamic glottis and the glottal source. Using the system, three kinds of glottal sources are synthesized for different purposes. Firstly, five different types of glottal pulses have been synthesized. Secondly, the properties of phonations with different F0, different SQ, different OQ, and both different F0 and OQ have been studied. In addition, the properties of the balance and unbalance of the left and right glottal areas, the symmetry and asymmetry of the left-right glottal periods, and the shift of the contour of the control function have also been investigated. Thirdly, six phonation types are synthesized with actual formants[①] for engineering purpose.

Finally, a general conclusion has been drawn and further researches on vibration of vocal folds through high-speed digital imaging in the areas of voice pathology, speech acoustics and phonetics are suggested.

---

① See chapter 9.

Acknowledgement

I would like first to express my deepest gratitude to my supervisor Professor William S-Y Wang for encouraging me and supporting me to enter this new area, and for his valuable guidance during the years in which I have completed this research. His valuable supervision together with his kindness has greatly benefited me both academically and personally.

I am grateful to Professor Seiji Niimi for his supervision, contributions and kindness. He is the supervisor of mine when I registered as a special graduate student at the Department of Speech Physiology, Faculty of Medicine, the University of Tokyo. With his help, the samples of high-speed digital images used in this research are obtained.

I would like to give my thanks to Professor Hirose for helping me to be accepted as a special student in the University of Tokyo to start my study in this area. Thanks are due to all the staff and fellow students, Imagawa, Murano, Fujimoto, Ishige, Konishi and Yan Qun, for their help in capturing the samples and their friendship when I was in Japan. I also would give my thanks to all the subjects. There will be no this research without their contributions. Here I would like to extend my thanks to the Department for the usage of the advanced high-speed digital imaging system and the financial support of the experimental cost.

I greatly appreciate Professor Gunnar Fant for his guidance on the acoustic models of glottis, especially his explanation on the development of the LF-model and his kindness when I stayed in Stockholm with him. I would also like to thank Professor Gunnar Fant and his colleagues at the Department of Speech, Music and Hearing, KTH for their opinions and comments on the model of dynamic glottis established in this book.

I would like to express my thanks to Professor Cheng Chin-Chuan for his suggestion on the structure of this book. I would also like to give my thanks to

Dr. Liew, Alan W C and Sum Koon Lung for their help in the image processing when I program the image processing and parameter extraction platform, and the dynamic glottis synthesis system at the Language Engineering Lab and Signal Processing Technology Center of the Department of Electronic Engineering, City University of Hong Kong. Thanks are also due to all the staff and fellow students for their help and friendship at the Center and at the Language Engineering Laboratory.

Finally, I would like to give my thanks to my wife, Chen Hua and my son, Kong Weihan for their love, support and patience. They are always there to share my joy and frustration.

本项研究的样本采集于东京大学医学院言语生理系，图像和声学信号处理系统建于香港城市大学电子工程系语言工程实验室，并得到了这两个机构的资金资助。

研究同时得到了国家自然科学基金数理学部面上项目“汉语普通话嗓音声学研究（项目批准号：10274105）”和“汉语普通话言语发声模型研究”（项目批准号：10674013）的资助。

本书的后期写作完成于中国社会科学院民族学与人类学研究所和北京大学中文系。出版得到了国家人事部留学归国人员研究启动资金的资助。

内容简介

人类言语的产生、演化和形成主要基于类人猿到人类的演化过程中发音器官的进化，它包括人类直立行走以后声道的弯曲和喉头发声器官的形成，而喉头的形成和发展奠定了现代人类言语声源的重要基础。现代言语声学理论认为，言语的产生主要包括嗓音声源和声道共鸣两个部分，语音学称为“发声”和“调音”。其中，嗓音声源在生理上对应于声带的振动方式。因此，对声带振动和动态声门类型的研究直接涉及到人类自然言语交际、言语工程、言语病理、司法语音、嗓音声乐等言语领域的基础理论问题。

在我国的宋代，人们已经对嗓音声源有了深刻的认识，沈括在他的著作《梦溪笔谈》第十三卷中就记录了当时人们对言语嗓音声源的认识、人造喉的制造和在法庭上的应用。① 然而，从语音学研究的发展历程上看，人们对语音的科学研究主要在语音的调音方面，而对语言的发声研究很少。随着人们对世界各地语言研究的深入，不同的发声类型被发现，特别是中国境内的民族语言中发现了许多不同的发声类型。但声带的隐蔽性使人们很难直接观察声带的振动方式，从而难以精确描述和研究嗓音的不同发声类型，这在很大程度上阻碍了语言发声类型的研究。计算机和高速数字成像技术的发展，使得人们直到近年来才有可能精确研究声带的振动方式和相对应的嗓音声学性质。

本书就是利用目前国际上研究语言发声中最先进和复杂的高速数字成像技术和语音信号处理技术，对语言的常用发声类型和汉语普通话声调的发声类型进行的研究。研究的样本采集于东京大学医学院，所有样本的采样频率为每秒钟 4500 帧。图像和语音信号处理系统建立于香港城市大学电子工程系。本书分为六个部分，十个章节。

第一部分由第一章“引言”组成，主要介绍了国际上高速数字成像嗓音研究的形成和发展过程以及基于高速数字成像的语言发声类型研究背景。同时，介绍了本书的目的和读本书所必备的喉部生理解剖知识。

第二部分由第二章“图像处理和参数提取”组成，主要介绍了高速数字成像系

① 见彩页。

统和图像处理及参数提取平台，它包括高速数字成像系统的系统参数、文件格式、图像处理中的旋转、选择、声门面积检测和录像帧间漂移的校正等。同时讨论了动态声门的类型、参数提取方法、统计参数的定义、24个参数的提取算法和用于建模的10项参数的算法和定义。最后讨论了声门图像处理的精度和发声类型的测定。

第三部分由第三章至第五章组成，主要分析研究了常见的语言发声类型和汉语普通话声调的发声类型及声带振动模式。第三章"正常嗓音"主要研究了标准正常嗓音、低音调嗓音和高音调嗓音声带的振动方式、振动特性和与声带振动方式相关的声学性质。第四章"特殊嗓音"研究了特殊嗓音的振动方式、振动特性和声学性质。具体包括假声、气泡音、气嗓音、双音调嗓音和吸气音声带的振动方式、振动特性和相关的声学性质。其中特别讨论了最高音调的假声和最低音调的气泡音的特殊性质以及声带不对称振动的方式、特性和声学结果。第五章"变调嗓音和汉语普通话四声的嗓音"研究和讨论了变调嗓音和汉语普通话四声的发声类型。其中特别讨论了汉语普通话四声的声带振动方式、振动特性和嗓音模式。这是目前唯一有关汉语普通话四声声带振动的高速数字成像资料和研究成果。

第四部分由第六章和第七章组成，主要介绍了动态声门特性的量化分析和特性研究。第六章"声门的性质"主要研究了声门的性质，包括新发现的七种声门脉冲类型和新定义的八种声带振动周期类型。同时还研究了声门面积和声门气流的关系，并建立了声门面积和声门气流之间的关系函数。第七章"统计分析"主要对数据进行了统计分析，包括对数据进行了相关分析和聚类分析，同时也对发声类型进行了相关分析和聚类分析。

第五部分由第八章"动态声门模型"组成，主要研究了基于高速数字成像的动态声门生理模型。其中主要研究和介绍了目前世界上最重要的八种嗓音声学模型，重点介绍了方特等人LF模型。在此基础上讨论了建立生理模型的不同层次、静态声门模型的建模和声门控制函数的建模。最终讨论和介绍了动态声门的生理模型。

第六部分由第九章和第十章组成。主要介绍了基于高速数字成像的嗓音生理参数合成和声带振动的科学研究前景。第九章"嗓音声源的合成"主要讨论嗓音声源的生理合成，其中包括不同声门脉冲的合成、不同声门生理参数的合成和不同发声类型的生理合成，并全面介绍了声门的生理合成方法。第十章"结论和进一步研究的建议"主要总结了全书研究所有的结论、讨论了这一技术对该领域的贡献并探究了基于高速数字成像研究动态声门的未来前景。

本书采用最新的科学技术，从声带生理振动和声学发声类型两个方面研究了言语声源产生的基本原理，是一本研究声带振动和嗓音发声类型基本原理的研究性专著，其成果主要可用于语音学、语言学、生理语音学、言语声学、病理语音学、嗓音声学、声乐学等相关领域。同时还可以当作以上各个领域的教学参考。

Contents