语音乐律研究报告

Status Report of Phonetic and Music Research

2007

北京大学中文系语言学实验室

Linguistics Lab
Department of Chinese Language and Literature
Peking University

说　明

本论文集是北京大学中文系语言学实验室年度研究报告的第二期，收录了实验室2007年发表和未发表的论文14篇。2007年，本实验室主要开展了如下活动：

一、科研

实验室语音学的研究主要有以下几个方面：基础理论和方法、普通语音学、生理语音学、民族语语音学、病理语音学等，详细内容参见本论文集正文。

二、学术交流

实验室与许多科研机构和大专院校建立了广泛的合作关系，如 Berkley、JAIST、台湾中央研究院、香港大学、香港中文大学、西北民族大学、广西大学等，并互派访问学生。同时，实验室开设的语音学课程，也接受了各地院校机构的学生选课或旁听。

2007年，邀请纽约眼科耳鼻喉科医院 R.J. Baken 教授、澳大利亚 Griffith 大学信息与通讯技术学院 Alan Wee-Chung Liew 博士、法国国家科学院米可研究员、台湾中央研究院语言学研究所郑秋豫研究员到实验室访问，并作了相关方面的研究报告。

实验室的大部分人员参加了2007年10月在安徽黄山举办的“第九届全国人机语音通讯会议”。

三、田野调查

2007年7月孔江平教授带领研究生在云南中甸进行藏语方言调查，2007年10月在江苏张家港、常山，11月在仙居、扬州进行汉语吴方言调查。

北京大学中文系语言学实验室
Linguistics Lab, Department of Chinese Language and Literature, Peking University
Tel: 86-10-62753016
Email: ell@pku.edu.cn
Website: http://chinese.pku.edu.cn

实验室人员		联系方式
孔江平	教授	jpkong@pku.edu.cn
李铎	副研究员	lid@pku.edu.cn
周燕	高级工程师	ell@pku.edu.cn
汪高武	04 级博士生	wanggaowu@pku.edu.cn/wanggaowu@gmail.com
尹基德	05 级博士生	yoonrealjdm@gmail.com
吉永郁代	05 级博士生	i_yoshinaga@hotmail.com
杨若晓	06 级直博生	yangruoxiao@pku.edu.cn
李英浩	07 级博士生	leeyoungho@pku.edu.cn
潘晓声	07 级博士生	itol_xs@pku.edu.cn
吴韩娜	07 级博士生	hanna.pku@gmail.com
谭晶晶	05 级硕士生	pkutjj@163.com
钱一凡	06 级硕士生	yifanqian@pku.edu.cn/yifanqian@gmail.com
叶泽华	07 级硕士生	zehuay@gmail.com

（自78年以来，实验室共培养研究生21名，其中博士16名，硕士9名。）

本期执行编辑：谭晶晶

目录

语音多模态研究和多元化语音学研究

——北大语音乐律研究的进展与理念

Research on Multi-Speech Models and Variety of Phonetic Study

孔江平

Kong Jiangping

摘要：语音学是一门既古老又现代的学科，说它古老是因为我们至今还在沿用其口耳之学的基本方法研究语音，说它现代是因为语音学已采用了大量现代科学技术，其研究也已进入了各个相关的科学领域。1925 年刘复先生在北京大学建立了第一个语音实验室—语音乐律实验室，它标志着中国语音学进入了科学的领域。经过了八十多年的发展，北大的语音乐律研究也进入了一个更广阔的领域。本文将依据北大语音学研究的现状，从“语音多模态研究”和“多元化语音学研究”两个方面和广大同仁一起探讨我国语音学研究的学科范畴和发展方向。

Abstract: Phonetics is an old and also a modern subject because of very old method, such as recording speech sound by hearing and imitation, and modern method, such as speech and neural signal processing, are all used at present. The phonetic research has also been related with many different fields. Prof. Liu Fu established the first phonetics lab, the phonetic and music lab, in Peking University in 1925, which symbolized that phonetics had become scientific in China. After more than 80 years, the lab has been coming into a more extensive research area. This paper introduces the study in the phonetic and music lab from the viewpoint of ‘multi-model phonetic study’ and ‘variety study of phonetics’ and discusses the research category and direction of phonetics in China.

关键词：语音多模态研究 多元化语音学研究 语音学 语音乐律 语音科学

Key words: Multi-models of speech, Variety of Phonetic Study, phonetics and music, phonetic science

1 引言

语音学最初作为语言学的一个分支有着悠久的历史，在早期的语言学研究中起了重要的作用。西方的传教士在传教过程中将语音学带到了世界各地，在语音学的研究普及上起了很重要的作用。随着科学技术的发展，语音学采用了大量科学的研究方法使其发展成为了一门和许多学科相关的现代科学。因此，我们说：语音学是一门既古老又现代的科学，说它古老是因为我们至今仍在用口耳之学的基本方法研究语音，说它现代是因为它已采用了大量现代科学技术，其研究已进入了各个相关的科学领域。本文将依据北大语音乐律研究的现状，从“语音多模态研究”和“多元化语音学研究”两个方面和广大同仁一起探讨我国语音学研究的学科范畴和发展方向。

1.1 北大语音乐律研究

刘复先生 1925 年在北京大学建立了第一个语音实验室—语音乐律实验室 ，它标志着中国的语音学进入了科学的研究领域。刘复先生的《四声实验录》（1924）是我国语音学领域最早的语音学研究成果，首次科学地揭示了声调的物理特性，这是中国语音学对世界语音学理论的重要贡献。经过了八十多年的发展，北大的语音乐律研究已进入了一个更广阔的领域。

1.2 语音多模态研究

在中国利用物理的方法研究语音已有八十多年的历史（刘复，1924），后来又采用了生理的方法（周殿福，吴宗济，1963）。随着科学技术的发展，许多新的科学方法用于语音学研究，特别是电子计算机和生理设备的应用，使得语音学的学科范畴有了很大变化。语音学研究最初只是为语言学服务，研究的目的比较单一，而现代语音学研究在方法和基础理论范畴方面已进入了物理学、生理学、心理学、病理学等科学领域（William J. Hardcastle et al, 1997; G. Fant et al, 2004），研究的目的和范畴已经大大超出了语言学，单纯面向语言学的语音学已不能满足当今科学研究和社会的需求。为了探讨和明确语音学的基础理论范畴，我们提出了“语音多模态研究”的理念，并根据近几年我们在“汉语语音多模态研究 ”方面的实践，希望与大家共同探讨语音学的理论范畴。

1.3 多元化语音学研究

除了面向语言学的语音学研究领域外，我国的现代语音学研究已经进入了许多相关的科学领域，例如，语音工程、语音病理、司法语音、声乐教学、播音教学等。语音学在和相关学科结合后，研究对象有了很大的扩展。语音学研究对象的扩展，一方面反映了社会发展的需求，同时，也对语音学的研究范畴提出了新的要求。为了更好地划分和明确语音学的研究范畴，我们在语音学研究的对象方面提出了“多元化语音学研究”的思路，以期能够和大家一同探讨。

2 北大语音乐律研究

北大语音乐律研究经过八十年的发展已有了一定的规模，本节主要从 1）实验室基础设备、2）语音乐律研究分析平台、3）语音多模态数据库三个方面简单介绍一下我们的研究条件和研究基础。

2.1 实验室基础设备

在北大 985 项目的支持下，从 2003 年开始逐步购买了一些先进的仪器，主要有：1）美国 KAY 公司的计算机语音工作站 3700，其选件有：语音合成分析选件 ASL、多维嗓音分析选件 MDVP、实时声门阻抗分析选件 EGG、音域分析软件、语音教学选件等；2）澳大利亚 ADI 公司的肌电脑电仪，其硬件包括肌电脑电采集器和生物电放大器；3）美国 KAY 公司的声门阻抗信号采集器；4）美国 UCLA 的气流气压计；5）英国产电子腭位仪；6）高质量的录音室、录音设备和录像设备。设备基本能够满足语音学基本的研究。

2.2 语音乐律信号分析平台

国外的语音学仪器一般只随机带有信号的采集软件和简单的分析软件，无法进行信号的参数提取和深入研究。为了使这些语音仪器能用于研究，我们逐步建立了自己的语音乐律分析平台，它主要有：1）语音声学信号分析系统，该系统具有自己的文件格式和用于语音学及言语工程目的的标注方法；2）X 光动态声道分析系统，具有图像处理、声道截面积计算和共振峰推算功能；3）动态声门分析系统，具有处理高速数字图像、声门参数提取和嗓音生理合成的功能；4）基于呼吸信号的韵律分析系统，具有提取呼吸重置、基频和嗓音参数的功能；5）电子腭位信号分析系统，能将电子腭位参数和声学参数结合起来分析；6）语音唇形分析系统，主要提取唇的外轮廓和内轮廓的参数功能，并能和语音声学参数结合起来进行分析；7）语音共振峰参数合成系统，主要是模拟声道进行语音合成；8）气流气压信号分析系统，提取气流、气压、鼻流等参数。这些分析系统的逐步建立和完善，从软件上为语音乐律的研究提供了基本的条件。

2.3 语音多模态数据库

根据现有的仪器设备，我们围绕语音多模态研究建立了汉语普通话多模态数据库和相关语音数据库，主要用于语音学的科研和教学。这些数据主要包括：1）汉语普通话单音节库；2）汉语普通话嗓音库；3）汉语普通话 EGG 信号库；4）汉语普通话词汇视频库；5）汉语普通话动态声道库 ；6）汉语普通话电子腭位数据库；7）汉语普通病理语音库，目前只有听障儿童语音库；8）声乐音库包括民族唱法头部和胸部共鸣信号库、标准元音头部共鸣信号库、呼麦声乐样本和日本“能乐”样本；9）特殊音库，主要有腹语样本和藏传佛教密宗诵经样本。

3 语音多模态研究

语音多模态研究主要是指对某种语音进行语言学、语音学、语音声学和语音生理学的全方位研究。在理论上，注重语音产生理论的研究和语音共性的研究。在方法上，尽可能采用声学、生理、心理学的研究方法采集各种声学和生理信号，如语音声学信号、生理信号和心理信号等。下面从 1）汉语普通话动态声道研究、2）动态声门研究、3）声学与电子腭位图谱研究、4）汉语普通话发声模型研究、5）基于呼吸信号的韵律研究、6）基于声门阻抗信号的嗓音模型研究几个方面介绍我们有关语音多模态研究的基本理念。

3.1 汉语普通话动态声道研究

在语音学研究中，动态声道研究是一项最有价值的研究。在早期的语音学研究中，人们就开

始了基于 X 光的声道研究，但由于 X 光对人体有一定的伤害，因此，样本采集非常慎重，X 光的动态声道研究也局限在较小的范围内。虽然研究规模小，但由于涉及到言语产生的最根本问题，其研究的成果具有很高的学术价值。随着技术的进步，CT 和核磁共振技术都有了很大的发展，声道图像信号已经从二维图像发展到了三维立体图像，大大促进了声道的语音学研究。

目前我们利用汉语普通话的 X 光动态声道资料对汉语语音产生的各个方面进行初步的研究 。研究主要分为：动态声道检测、声道截面积到共振峰的推算和动态声道的生理模型研究三个方面（汪高武等，2007）。另外，我们正在利用核磁共振采集汉语普通话基本元音的三维立体声道样本。同时，也正在和日本尖端科技大学党建武教授合作，采集动态核磁共振的声道数据，深入进行汉语普通话语音产生的生理模型研究。动态声道的研究除了可以解释语音学的许多理论问题外，其研究成果会大大促进言语工程和言语病理的研究和实际应用。另外，动态声道的基本资料可以用来制作语音多媒体教学系统，用于汉语普通话的语音教学，特别是聋哑儿童的语音康复训练。

3.2 动态声门研究

语音学将语音的产生分为调音和发声两个部分，调音的生理活动是通过动态声道来研究的，但对于语音声源的生理活动一直研究得很少。现在除了通过声学信号外，研究声带振动最先进的方法是声带高速数字成像技术。通过与东京大学言语生理系和香港大学言语听觉科学研究部的合作，我们对常见的发声类型和汉语普通话四声的声带振动方式进行了研究（孔江平，2001），在研究过程中建立了高速数字成像的分析系统。研究主要是先对声门录像进行图像处理，然后提取出动态声门的各项参数，研究这些参数和语音声学参数之间的关系，最终利用生理参数建立一个嗓音生理模型，合成出不同发声类型的声源。动态声门研究使我们能够更好地认识语音发声的生理机制、语音发声的微观运动、各种发声类型的特性以及和语音声学信号的关系，使我们更好地认识语音发声的原理和在语言交际中的作用。

3.3 声学与电子腭位图谱研究

《可视语言》(Ralph K. Potter et al, 1947）一书是在出现了语图仪以后第一本全面反映英语声学特性的研究性专著和图谱，从此，语言图谱研究就成为了语音学研究的一个重要方面。汉语普通话也出版了一本语图（吴宗济，主编，1986），由于当时的技术所限，只出了宽带和窄带语图。现在信号处理技术已经有了很大的发展，提取的参数在数量上和精度上都有了很大提高。因此，语图的研究已经和声学及生理参数数据库结合在了一起，研究的内容和目的都有了很大的变化，其重要性也越来越凸显。

现在研究声学图谱和数据库，首先要考虑研究的目的。众所周知，语音学研究现在和其它学科的交叉已经越来越大，因此，语音学的研究也要考虑其它学科的用途，在图谱和数据库的设计上不仅仅要满足语音学研究的需要，也要满足语音工程、语音病理等各个学科的需要。我们在汉语普通话的图谱制作和数据库建立方面主要考虑了一个原则，即声学数据的集合是否能够合成回语音波形。如果能较好地合成语音，则表明主要信息完全保留，没有太多的损失，这样就可以满足不同学科的研究需求。

3.4 汉语普通话发声模型研究

汉语普通话发声的声学模型研究一直是我们研究的重点，通过国家自然科学基金的多个项目支持，实验室建立起了一个由八个年龄段共 800 人的汉语普通话嗓音数据库，并对这 800 人的嗓音进行了声学研究。研究主要是提取常用的声学参数，如基频、基频抖动、振幅抖动、开商、速度商、嗓音的谱倾斜率等，然后利用这些参数建立汉语普通话嗓音的统计模型。例如，高中年龄段的平均基频要低于大学年龄段的基频；又如，幼儿园年龄段的基频男女已经开始有了微小的差别等。根据这些基本的特性，我们最近的研究主要集中在建立汉语普通话单音节声调和双音节声调的嗓音声学模型上（Kong jiangping, 1998, 2004），这些研究主要是为嗓音感知和嗓音的参数合成奠定基础。

3.5 基于呼吸信号的韵律研究

汉语是声调语言，同时又有很丰富和多变的韵律特征。目前国内大部分的研究主要是利用基频、音长和振幅等几个声学参数进行研究，并取得了很好的成果。然而，韵律是复杂的言语活动，对汉语韵律的研究采用新的生理信号和利用新的方法是非常必要的。从这个角度来考虑语音韵律研究的发展，我们购置了能够采集呼吸信号的仪器，结合呼吸、嗓音和声学三种基本信号来研究汉语普通话的韵律规律。

目前我们建立了一个基于呼吸信号的汉语普通话韵律数据库，数据库包括了不同的汉语文体，如五言七言诗、宋词、散文、小说和新闻。同步采集的信号有语音信号、呼吸信号和嗓音信号（EGG）。从这些信号中可以提取出基频、音长、振幅、开商、速度商、语音重置最大值和最小值、重置时间、重置覆盖时长等参数。研究表明汉语普通话呼吸重置有许多类型，在呼吸信号的强度上至少有三级重置，它们和韵律都有密切关系。另外，还可以确定，在言语交际过程中，呼吸和不同的文体、说话人的风格以及情感都有密切的关系。从呼吸信号本身来看，它对汉语普通话的语音教学和朗读等都有实际应用价值。

3.6 汉语普通话多维嗓音分析

嗓音的声学研究包括许多方面，多维嗓音分析主要是面向语音病理的研究。在自然科学基金的资助下，我们对汉语、藏语、彝语和蒙语做了多维嗓音分析（Kong Jiangping, 1999, 2000; 孔江平，2001b）。研究结果表明，不同语言的嗓音特性有较大差异，这说明了嗓音发声对语言的重要性，也说明了不同语言在使用自己特有的发声类型。

目前，我们正在使用 KAY 的多维嗓音分析软件，分析汉语普通话 800 人的嗓音样本。这项研究的主要目的是按年龄段和性别建立汉语普通话发音人的基本声学和生理参数数据库，多维嗓音参数共有三十二个。根据这些参数，希望最终能建立标准汉语音声学模型和嗓音生理模型。

4 多元化语音学研究

多元化语音学研究主要是指对各个不同科学领域中的语音学问题进行的研究。下面从 1）侗台语声调的声学研究、2）藏语声调起源的研究、3）语音病理研究、4）声乐语音学研究、5）腹语的语音学研究、6）诵经的发声研究几个方面介绍一下我们有关多元化语音学研究和教学的思路。希望在语音学研究对象的范畴上进行一些探讨。

4.1 侗台语声调声学研究

面向语言学的语音学研究是语音学基础理论的来源，对汉语、汉语方言和民族语言的语音进行研究是我们的基础。除了汉语语音多模态研究外，目前我们正在进行的一项研究是侗台语声调的声学研究 。研究主要是从调类出发，研究我国境内主要侗台语单音节和双音节声调的声学性质。

我们研究的侗台语包括：武鸣壮语、龙洲壮语（李洪彦等，2006）、德宏傣语、西双版纳傣语、莫语、临高语、榕江侗语、拉珈语、毛南语、仫佬语、水语、标语（谭晶晶等，2006）、仡佬语、佯黄语、加茂黎语。从这些侗台语声调声学研究的结果看，侗台语声调有很多自己的特点，如，单音节声调数量较多，调类之间有明确的对应关系，声调和声母及韵母有互补分布关系，声调分舒声和促声。侗台语的双音节声调组合数量一般都很大，但个别组合不一定就能找到实际的词。这些组合中，有的会产生变调现象。由于侗台语声调和音节结构有复杂的关系，同时具有大量的双音节组合形式，因此，侗台语声调的音位负担在汉藏语声调中有极其特殊的性质，是汉藏语声调理论研究的一个重要方面。

4.2 藏语声调起源的研究

历史语言学主要研究不同语言和方言之间的语音对应规律，包括历史文献的资料和活的语言及方言的资料，这些资料实际上也包括了语言接触对音变的影响。因此，怎样从语音学的角度来看待历史语言学的研究是语音学研究的一个非常重要的方面（孔江平，2006）。在这个方向上，我们主要开展了针对藏语声调起源的语音学研究 。

藏语是汉藏语系中一个重要的语言，也是我国一个使用人口较多和分布较广的民族语言。藏语有自己的传统文字，大约创立于公元七世纪。从藏

文上看，历史上的藏语有 200 多个声母。其中大部分是复辅音声母，从现代藏语方言的语音上看，安多藏语保留有几十个到一百多个不等的声母，没有声调，只有所谓的“习惯调”。同时，康方言和卫藏方言都有声调，但数量不同。研究表明，安多藏语的音调根据声母的清浊有一定的分化。藏语康方言中，有些单音节和双音节声调不是很稳定。另外，藏语音节和声调在整个音位系统的分布上也都有明显的特点。藏语语音方言上的这种分布，为研究藏语声调的产生和发展提供了活的资料，也使我们从语音学的角度研究藏语声调的产生和起源过程成为可能。

4.3 语音病理研究

在我国约有听力残疾者 2004 万，每年有 3 万新生听障儿童，因此，从语音学的角度研究听障儿童的语音习得和言语康复有重要的学术价值和实际应用意义。近几年我们研究了北京中国儿童康复中心部分听障儿童单音节声调和双音节声调的获得情况，研究包括佩戴助听器的听障儿童和植入人工耳蜗听障儿童的声调获得。研究结果表明，佩戴助听器儿童的声调获得要好于植入人工耳蜗儿童的声调获得（李洪彦，2007）。系统的研究不仅解释了语音获得的一些语音学问题，而且为听障儿童的语音康复提供了教学上的依据。同时，也对人工耳蜗在技术上的改进提出了新的要求。

4.4 声乐语音学研究

歌唱是人们言语交际的一种特殊方式，在我国许多民族都有对歌的习俗。它除了具有言语交际的功能外，在情感的表达上更为丰富。从声学的角度看，唱歌时声音可以很大，这样可以传得更远，延长了交际的距离。由于歌唱更注重情感的表达，因此，在歌唱时会用更为复杂的发声方式和调音方法，这为语音学研究提供了更为丰富的研究内容。

我国从声学和语音学的角度研究声乐的人比较少，而需要研究的声乐领域却十分广阔。我国有大量的戏曲，如昆曲、京剧以及各种地方戏等，同时还有大量的不同民族的民歌，这些戏曲和民歌具有丰富多变的发声方法。近几年，我们主要进行了民族唱法头部共鸣和胸部共鸣的研究（钱一凡，2007）。同时，也对标准元音头部共鸣进行了研究，找出了不同元音在头部不同部位共鸣的强弱程度（孔江平，2007）。另外，我们还正在研究蒙古族的“呼麦”和日本的“能乐”。呼麦是蒙古族一种独特的演唱形式，根据我们的研究，呼麦至少有九种基本的演唱方法，有的是运用不同的发声方式，有的是运用不同的调音方式，或者两者结合，它是一种具有奇特表现力的艺术形式。能乐是从中国传入日本的一种戏剧，在日本得到了发展，它使用一种低音调发声类型演唱，具有很独特的艺术风格。

4.5 腹语的语音学研究

腹语是一种独特的舞台言语表演艺术，表演者在操纵木偶的同时，唇型不动，同时能模仿出不同人的发音。腹语研究的意义在于，它完全采用代偿性发音来表演。代偿性发音主要是言语有残疾的儿童，如腭裂儿童，在言语学习过程中形成的发音方法。而腹语是一整套系统的代偿性发音，研究腹语不仅能使我们更好地了解这种言语艺术，而且还能充分认识人类发音器官的潜在能力，为言语儿童康复提供训练的方法。

我们对一名特殊的腹语者的语音进行了研究 。这名腹语者是闭着嘴说话，从初步的研究看，气流是从嘴角挤出，元音共振峰的模式大多能区分，整个舌的运动靠后，辅音主要靠舌尖调节。由于闭着嘴，整个音调比正常说话要高，声调模式和正常没有区别。

4.6 诵经的发声研究

在我国，佛教和藏传佛教的诵经都会使用特别的发声类型来营造虔诚的气氛和达到震撼心灵的效果。根据我们的调查，在藏传佛教中，这种诵经要从小经历严格训练才能学会。我们调查研究了一些藏传佛教的诵经，声学分析表明，诵经样本的基频并不是很低，大约在 80 赫兹左右，但在听感上却很低。从共振峰上看，诵经在高频段有很强的峰值，类似于歌唱共振峰。因此，诵经的发声类型和共振峰结构都很复杂，这些现象还需要进一步研究。另外，诵经时往往会采用两三个人和声的方式，和声能造成基频的差频，从而出现一个更低的基频，这种高能量的低频声音往往能造成人们心灵上的共鸣和震撼。近几年情

感语音的研究越来越受到重视，诵经在发声和调音方面的技巧很值得研究。

5 结语

中国的现代语音学研究从刘复先生在北大建立语音乐律实验室起已走过了 80 多年的历程，随着技术的进步和研究的深入，语音学和其它学科在研究方法和内容上越来越交叉，研究的领域也在不断融合。在这种情况下我们有必要从理论上、方法上和研究对象上对语音学的学科范畴进行讨论和界定，这种讨论和界定将会十分有利于现代语音学的发展和进步，从而推动我国的语音学理论和应用的研究。

“语音多模态研究”是随着科学的发展，在语音学采用了大量其它相关科学研究方法和基础理论的情况下发展出来的。另一方面，语音多模态研究的发展也反映出了社会对语音学提出了新的要求。虽然许多不同学科的学者都在进行语音多模态的研究，但研究目的不尽相同。语音学的学科范畴应该限定在研究语音的交际功能及内在规律方面。因此，我们认为“语音多模态研究”主要是从研究方法和基础理论上探讨语音学的学科范畴。虽然现在很难将语音学的学科范畴界定得很清楚，但大的范畴应该明确，这将有利于学科的发展。

“多元化语音学研究”是从研究对象和研究领域上来探讨和扩展语音学的研究及应用范畴。从我国具体的情况看，许多领域十分需要语音学的研究。如，聋哑儿童的语音研究和语音康复研究、民族声乐中的语音研究、特殊语音的研究等。目前，我国语音的多元化研究还十分薄弱，有许多空白，需要大力发展，这对我们的研究生教育也提出了新的要求。总之，我们希望通过讨论，能更加明确语音学的学科范畴和研究领域，推动中国语音学的发展。

参考文献

[1]孔江平，2001，《论语言发声》，民族大学出版社

[2]孔江平，2006，现代语音学研究与历史语言学，《北京大学学报》，哲学社会科学版，第 43 卷，第 2 期，北京大学出版社

[3]孔江平，2007，标准元音头部共鸣声学研究，第七届中国语音学学术会议论文集（电子版）

[4]李洪彦 蓝庆元 孔江平，壮语龙州话声调的声学分析，《民族语文》，2006 年第 6 期，商务印书馆

[5]李洪彦，2007，听障儿童普通话声调获得研究，北大研究生论文

[6]刘复，1924，《四声实验录》，上海群艺出版社

[7]钱一凡，2007，民歌男高音共鸣的实验研究，第七届中国语音学学术会议论文集（电子版）

[8]谭晶晶，孔江平，2006，广东诗洞标话双音节连字调的音系研究，《语言学论丛》，第三十四辑，商务印书馆

[9]汪高武，孔江平，鲍怀翘，2007，从声道形状推导普通话元音共振峰，第七届中国语音学学术会议论文集（电子版）

[9]周殿福，吴宗济，1963，《普通话发音图谱》，商务印书馆

[10]吴济宗，主编，1986，《汉语普通话单音节语图册》，北京，中国社会科学出版社

[11]G. Fant, H. Fujisaki, J. Cao, Y. Xu, 2004, From [12]Traditional Phonology to Modern Speech Processing, 语音学与言语处理前沿，外文教学与研究出版社

[13]Kong Jiangping, 1998, Egg Model of Diatones in Mandarin, Proceedings of 5th International Conference on Spoken Language Processing. Tu5P16, Sydney, Austrlia.

[14]Kong Jiangping, 1999, Correlation and Classification Study on EGG Parameters of Mongolian, Proceedings of 4th National Conference on Modern Phonetics, Jiancheng Publishing House.

[15]Kong Jiangping，2001a, Study on Dynamic Glottis through High-Speed Digital Imaging, PhD. Dissertation at the Department of Electronic Engineering, City University of Hong Kong

[16]Kong Jiangping, 2004, “Phonation Patterns of Tone and Diatone in Mandarin”, From Traditional Phonology to Modern Speech Processing, edited by G. Fant, H. Fujisaki, J. Cao, Y. Xu

[17]Kong Jiangping, Shen Mixia, Chen Jiayuou and Caodao Bater, 2000, Correlation and Classification Study on EGG Parameters of Tibetan, A collected Essays for the Technology of Man-Machine Speech Communication, Tsinghua university Publishing House.

[18]Ralph K. Potter, George A. Kopp, Harriet C. Green, 1947, Visible Speech, D.Van Nostrand Company, INC. New York

[19]William J. Hardcastle, John Laver, 1997, The Handbook of Phonetic Sciences，Blackwell Publishers

中国语音学的历史与现状

History and status quo of modern phonetics in China

孔江平

Kong Jiangping

摘要：中国现代语音学研究始于上个世纪初期刘复先生在北京大学国文系“语音乐律实验室”，刘复的《四声实验录》第一次阐明了基频是声调的物理基础，这是中国学者对世界语言学的重要贡献。经过 80 余年的发展，中国的语音学有了显著的发展。本文即是对中国语音学的历史和现状的简要介绍。

Abstract: The study of modern phonetics in China appeared in the early period of last century and the landmark of modern phonetics in China is the establishment of the Phonetic and Music Lab of Peking University. The book ‘Experiment on 4 tones in Mandarin’, written by Prof. Liu Fu first explained that the fundamental frequency is the physical foundation of tones, which is an important contribution of Chinese scholar to the modern phonetics. After more than 80 years, phonetics in China has greatly developed. This paper will briefly introduce the history and status quo of modern phonetics in China.

关键词：中国语音学 历史 现状

Key words： Phonetics in China, history, status quo

1 引言

中国的现代语音学研究可以认为起始于刘复先生在北京大学国文系建立“语音乐律实验室”，它标志着现代语音学在中国进入了系统科学的研究阶段。刘复先生的《四声实验录》[1]第一次阐明了基频是声调的物理基础，这是中国学者对世界现代语音学理论的重要贡献。经过了 80 多年的历程，我国语音学研究已经有了很大的发展，本文将从现代语音学在我国的发展过程、研究机构和研究现状简述语音学在中国的发展历史与现状。

2 语音学的发展历史

1921 年 11 月，刘复先生正式向蔡元培先生提交了一份《提议创设中国语音学实验室计划书》(《北京大学日刊》1921 年 11 月 16 日)，提出“鉴于研究中国语音，并解决中国语言中一切与语音有关系之问题，非纯用科学的试验方法不可”。1925 年春，在法国巴黎大学专攻语音学并获得博士学位的刘复回到北大国文系任教，同年 9 月，在其主持下，“语音乐律实验室”正式成立。实验室隶属于国文系，内有研究语音乐律及进行教学实验的仪器多种(部分仪器由刘复亲自制造)。当时的重要工作主要是调查全国方音，制成各种声调曲线及图表。1928 年，实验室改属文科研究所国学门。1934 年起又归属研究院文科研究所。语音乐律实验室的成立标志着我国现代语音学的正式开端。

2.1 三十年代的语音学研究

在 30 年代，随着新的学科制度的建立，学者的研究也趋向专业化。时为国文系研究教授兼研究院文史部主任的刘复，在这一时期将主要精力投入到对古声律的研究中，除任课、撰写论文和专著外，还亲自创制语音实验所用的仪器，并对故宫、天坛古文物陈列所收藏的古代乐器的音律进行了测定。刘复还利用暑假到开封、上海、南京、曲阜、济南等地探访测验私人所藏古今乐器，并到平绥铁路沿线调查方音。

2.2 四十年代的语音学研究

刘复以后由罗常培先生接任语音乐律实验室工作。抗日战争开始后，北大南迁到昆明，与清华、南开合并为西南联大，在条件艰苦，仪器缺乏的情况下，依然做了许多调查研究工作。1936 年罗常培先生调查临川语音，使用浪纹计和声调推断尺完成了《临川音系》这部专著[2]。西南联大时期，罗常培先生组织人力开设了与边疆语言调查相关的课程，在仪器缺乏的情况下，坚持进行对边疆语族的研究，收集了许多西南少

数民族语言的材料，做了许多实际而卓有成效的工作，为语言学的研究积累了大量珍贵的第一手材料。1946 年 10 月，北京大学复校回到北京，实验室恢复工作。

2.3 五十年代的语音学研究

1950 年 6 月中国科学院语言研究所成立时实验室大部分归入语言所，其后又归入中国社会科学院语言研究所，这时现代语音学的主要研究力量已转入中国社会科学院，主要代表人物有周殿福和吴宗济两位先生，为了配合普通话的普及开始对普通话的语音性质开展了大规模的研究。在北大只留有一小部分人员和仪器。五十年代中，为配合汉语方言课，由甘世福讲授语音学，购置了浪纹计等设备，由王福堂负责管理。50 年代后期，在下厂下乡开门办学的浪潮中，林焘、王福堂等也曾在城子煤矿用钢丝录音机等设备进行语音教学和推广普通话。1966 年文化大革命开始后，社科院和北大的语音学研究工作和教学都基本停顿。

2.4 文革以后的语音学研究

文革以后，社科院语音室对普通话开展了大规模的生理和声学研究，1978 年北大中文系决定恢复建立语音实验室，由林焘先生主持，为北京大学重点文科实验室。并从当年起设立实验语音学研究方向，招收研究生。1979 年，美国加州大学伯克利分校著名语言学家王士元先生应邀来北大讲授实验语音学，听讲者来自全国各地共一百多人。王士元先生系统介绍了实验语音学，带来了最前沿的学术知识和信息，讲座内容出版了专辑[3]。此次讲学活动激起了国内语音实验研究的热潮。同年，实验室开办语音学研究生班，由林焘、吴宗济、张家騄授课，这个研究生班为语音学方向培养了一批中坚力量。1991 年 9 月社科院语言所、民族所和北大中文系联合策划，由北大中文系语音实验室主办了第一届“现代语音学研讨会”，这是我国历史上第一次全国性的语音研讨会，与会代表近百人，这次学术研讨会引起了广泛关注，对跨世纪的中国语音学的发展起到了积极的推动作用。

纵观中国现代语音学的发展历史，有两个重要的标志，一个刘复先生在北大建立“语音乐律实验室”，它标志着现代语音学被正式引入中国。另一个是文革以后，王士元先生到北大做学术报告和在北大举办语音学研究生班，它标志着中国语音学研究现代化的开端。另外，由北大主编并即将由商务印书馆出版的《中国语音学报（创刊号）》将会成为中国现代语音学学术发展的另一个重要标志。

3 相关研究机构

语音学是一门交叉学科，无论是文科还是理工科的机构都用自然科学的方法研究语音，但研究目的不尽相同，这些理科机构很多，比如中国科学院的声学所、心理所、自动化所；高校如清华大学、北京大学的相关院系。这里主要介绍文科单位和理工科单位以外的机构。另外，这几年许多学校都在建立语音实验室，因此，本文的介绍可能会有遗漏。

中国的语音学研究机构主要有科研院所、高校和一些特定的与语音相关的研究机构。下面将从科研机构、普通高校研究机构、民族院校研究机构和其他研究机构四个方面对语音学研究机构做一简单的介绍。

3.1 科研院所

社科院语言所的语音研究室有很悠久的历史，主要研究汉语普通话及其方言，这个研究室不仅有较多的专业科研人员，而且有很好的设备。早期对汉语普通话进行了大量的声学和生理研究，出版有《普通话发音图谱》[4]，和《实验语音学概要》[5]，这几本书对中国现代语音学研究的影响都很大。语言所的语音研究室还是我国最早研究共振峰参数合成的机构。目前语言所的语音室在建立大型汉语及方言的口语语料库和生理语音学方面都做了大量的研究，取得了很好的研究成果。社科院民族所的语音实验室建立于 80 年代，主要是进行中国民族语言的语音学研究，其研究主要集中在两个方面，一是藏语、蒙语和哈萨克语的大型声学参数数据库的研究；另一个是中国民族语言嗓音发声类型的研究，主要是对民族语言的嗓音发声类型进行了系统的声学和生理研究，出版有《论语言发声》[6]。

另外，民族所和北大中文系合作正在进行中国境内汉藏语声调的声学研究。

3.2 普通高校

在高校的语音学研究是和语音学研究生培养相结合的，我国目前有语音学研究和能够培养语音学方面研究的高校大约有十几所，下面将分别作简单的介绍。北京大学中文系是我国最早建立语音实验室的大学，由刘复先生于 1925 年在北大国文系正式成立，北大的语音乐律实验室主要以中国境内的汉语、汉语方言、民族语以及民族乐律为研究对象，实验室具有许多生理和声学仪器和设备。《北京语音实验录》[7]是对汉语普通话声学研究的一个主要成果。这几年主要在进行汉语普通话语音多模态的研究和汉藏语声调的声学研究。香港城市大学语言学及翻译系建立有一个很好的语音学实验室，有许多语音设备和仪器。主要研究汉语普通话和方言的语音特性。南开大学是文革后建立语音实验室比较早的学校，有一些语音仪器和设备，有自己编制的语音分析软件，主要进行汉语方言和民族语言的语音学研究，并培养了许多语音学方面的研究生，另外，在暑期还开办语音学的学习班。北京语言大学主要是培养外国学生，所以一直对语音学研究很重视，北京语言大学的实验室有非常好的实验室仪器和设备，主要开展汉语韵律的研究以及汉语和外语的对比语音学的研究。中国传媒大学一直对语音研究比较重视，主要是用语音学的基础理论和方法研究播音语言的语音特性，其中主要集中在播音语言韵律、节奏特点和风格的研究方面。上海师范大学建立了一个较完善的语音实验室，主要进行汉语方言语音特性的描写和研究，其研究有自己的特色，即通过汉语方言语音特性的研究来解释汉语语音发展和历史音韵的问题。南京师范大学的实验室主要进行汉语及方言的语音学研究，特别是对汉语方言的声调及其理论进行比较深入的研究。广西大学语音实验室的研究主要集中在广西境内的民族语言的语音和汉语方言语音的研究方面，如标准壮语声学图谱和参数数据库等。另外，中国人民大学、复旦大学、广西师范大学、伊利师范大学和暨南大学也都建立有语音室并在做语音学方面的研究。

3.3 民族院校

在研究民族语言的语音方面主要有内蒙古大学、中央民族大学、西北民族大学等，这些学校都开展了这方面的研究。内蒙古大学建立语音实验室比较早，并对蒙古语的语音进行了大量语音声学分析和研究，建立有大型的蒙古语音数据库，培养了不少民族语言的学生和人才。中央民族大学和美国暑期语言学院合作曾建立了一个语音实验室，但主要是在教学方面，培养了一些学生。近几年民族大学的语音学研究有了较大的发展，并在汉语和不同民族语的语音对比研究方面出了许多成果。西北民族大学近几年在语音学研究方面发展很快，在学校的支持下，购置了大量的语音实验仪器和设备，其研究集中在西北的汉语方言和民族语言的语音学研究上。近几年主要进行了藏语安多方言的声学图谱和声学参数数据库的研究、藏语拉萨话的韵律研究、临夏汉腔和回腔的语音研究以及东乡族的汉语声调的研究。另外，西南民族大学和云南民族大学也在建立语音实验室，并开展了相关的民族语言的语音学研究。

3.4 其他科研机构

比较特殊的语音学研究机构主要有司法声纹鉴定机构和有关的医疗机构。司法声纹鉴定的机构有东北刑警大学、公安部第二研究所、广东刑事侦查技术中心、江苏省公安厅和中国政法大学等。东北刑警大学和公安部第二研究所是开展声纹研究比较早的单位，研究主要侧重在声纹的个人差异方面。在实际的应用方面广东省刑事物证鉴定技术中心做了大量的实际案例，有丰富的经验和数据。另外，江苏省公安厅、中国政法大学等也在开展这方面的基础和实际应用研究。在语音病理方面，主要的机构有中国聋儿康复研究中心、北大口腔医院、北京医科大学聋儿康复研究所、协和医院整形医院等。这些机构的研究和治疗主要涉及聋儿的听力康复、语训、腭裂的治疗和语音康复。

4 研究内容

我国的语音学研究已经发展到了一个比较繁荣的时期，无论从基础理论和研究方法，还是从研究领域和研究队伍都已经达到了相当的规模。特别是高校近几年纷纷建立语音实验室和培养语音学专业的研究生，扩大了语音学的研究队伍。另外，国家科研力度的加大也大大促进了语音学研究的发展。本文将对我国语音学的现状从宏观语音学论坛、面向语言学的语音学、声学语音学研究、语音韵律研究、生理语音学研究、语音工程研究、方言语音研究、对比语音学研究与语音教学、心理语音学研究、司法语音学研究和民族语语音研究 11 个方面对我国语音学的现状做一简单介绍。全面介绍我国语音学的研究是比较困难的事，因此，本文主要是根据 2006 年在北大召开的“第七届中国语音学学术会议暨语音学前沿问题国际论坛”的 110 多篇论文，简要介绍语音学的研究现状

4.1 宏观语音学论坛

语音学论坛从宏观语音学、台湾语音学及相关研究、社科院语言所的语音学研究和北大语音实验室的语音学研究情况反映我国语音学研究的概貌。在我国宏观语音学的研究比较少，大部分学者都是在比较具体的语音学研究领域，近今年来宏观语音学的研究对于指导我国语音学的发展方向、开拓新的语音学领域和语音学研究的定位都有很重要的意义[8]。虽然我国语音学的领域已经十分广泛，但距国际语音学的研究前沿还有一定的差距，如利用核磁共振（MRI）对动态声道的研究和利用高速数字成像对声带振动的研究；利用功能性核磁共振（fMRI）和脑电仪（EEG）对语音活动的大脑定位研究等。台湾语音学的研究主要集中在中央研究院，其特点是充分利用科学和工程的技术，建立大规模的语音数据库，研究汉语韵律的语音学特征，然后将研究成果应用于言语工程[9]。大陆的语音学研究有两个方面的发展，一是语音学开始出现面向言语工程应用，二是开始语音多模态研究和多元化语音学的研究。如建立大规模的语音合成和识别的语音数据库，建立汉语及其方言的口语数据库等[10]。另一个发展方向是开始了汉语语音多模态的研究和语音多元化的研究。在语音多模态研究方面，主要是从语音学的角度来研究汉语普通话的声学、唇位形状、声道运动和声带振动的内在规律和模型[11]。在语音的多元化研究方面，逐步在开展语音相关的研究领域，如聋哑儿童声调习得、腹语、声乐、宗教诵经的语音学声研究等[12]。另外，国内的语音学研究和国际上的语音学和言语工程研究一直都有广泛的联系，如在工程方面和早稻田大学的交流[13]及 ATR 的交流[14]。

4.2 面向语言学的语音学

面向语言学的语音学研究一直都是语音学基础理论研究的重点，在这个方面目前的研究主要包括：外语译词的音系研究、汉语韵母异同的声学分析、普通话基本调位研究、汉语方言与普通话推广的研究、优选论框架下的音系研究、汉语官话方言元音格局的研究、普通话声韵调音值的研究、语气词重读研究、方言语音比较研究、中文语音学术语的标准研究等。在语音学的理论方面主要有关于赵元任汉语语调思想和边界调的研究[15]、汉语方言元音类型学的研究[16]、语音学术语中文翻译及规范的研究[17]以及上海话韵律词声调的优选论分析[18]。

4.3 声学语音学研究

在语音声学方面，研究涵盖语音发音机理和声学关系的研究、普通话元音和辅音的声学研究、普通话音长和音高的综合研究、普通话轻声的声学研究、元音相似度的量化研究、普通话平翘声母区别特征的声学参数研究、北京话元音的统计分析等。目前研究趋势主要集中在声学参数数据库的研究上。信号处理技术的发展，使系统研究一种语音的语音体系已经成为可能，因此，声学参数数据库的标准成为大家讨论的热点，总的来说，元音的声学参数相对比较统一，但在辅音的声学参数的提取上，其标准却很难统一[19]，学者普遍认为声学参数的标准不能是唯一的，在现阶段主要看研究的目的是什么，根据研究的目的制定声学参数的标准是比较合理的。随着信号处理技术的深入，制定同时能用于语言学研究目的和用于言语工程目的的声学参数标准

势在必行。另外，在研究发音增强与减缩[20]、普通话元音过渡与辅音腭位的关系[21]以及北京话一级元音的统计分析[22]等方面的研究反映了语音声学研究的现状。

4.4 语音韵律研究

韵律研究一直都是我国语音学研究中的一个热点，这和目前的汉语普通话的教学及言语工程有密切关系。因为，在教学方面，外国留学生对于掌握汉语韵律特征比较困难，汉语的韵律包含声调变化和语调变化，是一种双重的音调控制。因此，这方面的研究对汉语普通话的教学有重要的意义。另外，由于在汉语的合成方面韵律对自然度有很重要的作用，因此在合成方面，韵律规则也一直是研究的重点。目前在韵律方面的研究包括语调核心单元的研究、汉语语调标记研究、汉语韵律短语的结构研究、普通话焦点重音的研究、连上变调的声学研究、汉语语调研究、普通话词重音的研究、韵律词的结构规则研究、情感句重音研究、节奏支点研究、新闻播音语言节律研究、主持人口语的韵律研究、朗读的喘息研究等。韵律研究主要集中在普通话双音节韵律词时长的研究[23]；韵律不同层级的声学研究[24]；普通话重音的研究[25]。另外，情感句重音模式的研究是韵律研究的一个新的动向[26]，这方面的研究将汉语普通话韵律的研究推向一个更新的研究领域。

4.5 生理语音学研究

语音生理的研究一直是语音学研究的一个重要方面，因为语音生理机制研究是语音学的理论基础。这些研究包含食道语发声与正常声带发声的对比研究、普通话言语测听方面的研究、聋哑儿童普通话声调获得的研究、普通话和上海话辅音发音部位 EPG 的研究、电磁发音仪的元音研究、民族唱法胸部和头部共鸣的声学研究、元音头部共鸣的声学研究、新闻朗读的呼吸节奏研究等[27]。这方面的研究可以归为 4 个重要的方面，一是利用电磁发音仪对发音机理的研究[28]，电磁发音仪是近十年才发展和开发出来的语音学研究仪器，口腔发音部位的定位和时间域上都有较好的精度，是一个有较好前景的语音学研究设备。利用呼吸信号来研究语音的韵律特性是另一个语音生理研究的新思路和新方法[29]。呼吸信号可以用不同的方法来采集，以往对呼吸的研究主要是用气流气压计研究比较微观的语音现象，如辅音、嗓音等。用于研究语音韵律的呼吸信号是利用呼吸带采集胸围或腹围的变化，从而达到研究呼吸和韵律特征的目的。利用这种方面研究语音也是首次出现，呼吸信号引入语音学的研究对韵律风格和个人风格的研究都会有很大推动。在生理研究方面，利用电子腭位仪（EPG）进行语音学的研究也是近十年来语音学研究发展的一个重要方面，目前电子腭位的研究主要是集中在汉语普通话和上海话的发音研究上[30]，电子腭位的利用不仅可以对语音的发音动作和机理学进行研究，还对腭裂术后的语言康复有重要意义。利用 X-光研究声道的形状有比较长的历史，但随着图像信号和语音信号处理技术的进步，对 X-光声道录像的研究又展现出了广阔的前景。声道形状和语音声学关系的研究在语音产生的基础理论方面有重要的意义，同时，通过这些研究可以建立声音的视觉反馈系统，因此，在语音教学，特别是聋哑儿童的语音学习方面有重要的应用前景[31]。

4.6 语音工程研究

语音学和语音工程研究的界限很难划分，本文在这里只是根据第 7 届中国语音学学术会议论文进行一点简单介绍一点。在语音工程的研究上，有时很难说是属于工程还是属于语音学。由于在研究的方法上有时是完全相同的，因此很难界定，或者说没有必要界定。但有一点是比较明确的，语音学的研究更关注语音的发音机理、声学信号的物理意义和语音的交际功能。而语音工程更关注语音应用的实现。目前面向语音工程的语音学研究主要集中在以下几个方面：人工神经网络识别的研究、基于模式识别的声调识别、基于联合源-滤波器模型优化的语音声门源模型估计、基于感知的韵律层级与基于 F0 曲线生成模型的语调研究、基于 F0 曲线生成模型的方言声调系统的分析、语音识别中的音子集设计研究、语音基频结构与情感正负性的研究、面向说话人识别的汉语情感语音库、音节之间相互制约关系

的研究、基于发音稳定段的自适应步长段模型解码及其在 LVCSR 中的应用等。总的来说，言语工程中的语音学研究除了对语音的各个方面进行建模外，正逐步开始关注到语音韵律感知层面的建模[32]。另外一个新的发展动向是语音工程由于语音合成等的需要，也逐步开始关注语音的情感计算[33]和情感的感知建模[34]，这是面向工程的语音学研究的一个新的发展点。

4.7 方言语音研究

方言的语音学研究近几年发展的很快，国内许多高校和研究方言的机构都逐步开始利用语音学的方法对方言的语音特性开展更科学的研究，近来的研究包括：汉语方言单字调音高分组统计研究[35]、方言单、双字调调长分析、香港粵語陽上陰去相混研究、方言元音的声学研究、汉语方言中轻声调值的类型、方言吞音现象的分析、普通话与方言的声学特征对比及转换、基频归一和调系规整的研究、方言复合元音的声学分析、方言入声实验研究、汉语方言元音普遍现象研究、方言声母的气流与声学分析、方言入声字时长特性分析、方言连读变调的语音分析等。一个比较新的特点是对方言基频和调系规整的研究[36]。

4.8 对比语音学研究与语音教学

由于到中国学习汉语的外国学生越来越多，在汉语普通话教学方面，各国留学生学习汉语语音的困难各有不同，这使得汉语的语音教学越来越受到重视，不少具有语音学基础老师开始了面向汉语对外教学的语音学研究，这一部分老师的数量很大，可以预见这方面的研究将会很快成为我国语音学研究的一个重要方面。目前的研究主要包括：中国学生英语朗读的语调模式研究、节律操练法研究、哈萨克族汉语声调的声学研究、维吾尔族汉语语音的声学分析、越南留学生汉语声调的声学研究、基于语料库的中国英语学生朗读介词突显性研究、维吾尔族学习汉语普通话塞音的实验研究、中国学生英语朗读的韵律特征研究、泰国学生汉语语调研究、英语与汉语节奏模式研究等。其中汉语的声调对大多数留学生来说都是一个难点，因此，这方面的研究也就显得更为重要，比较典型的研究，如越南留学生习得汉语单字调的声学考察与分析[37]，集中反映这方面的研究情况。

4.9 心理语音学研究

语音心理的研究是语音学研究的一个重要范畴，也是语音学研究的一个比较高级的阶段，因为通过语音心理的研究，使我们能够深入到语音交际活动大脑层面。近期这方面的研究主要包括：儿童言语错误与方言差异研究、时长与广州话元音的感知研究、跨语言元音的声学感知分析、维吾尔族学习者对汉语普通话塞音的范畴感知、认知语言学与语音认知研究、维吾尔族学生对英语松紧元音的感知研究、汉语的节奏及其获得研究、汉语儿童语音系统获得研究、汉语表情话语中的调型改变及其感知等。比较典型的研究主要是在语音的微观感知方面，如，方言语音特性的感知[38]和汉语表情话语中的调型改变及其感知[39]。

4.10 司法语音学研究

司法语音学（forensic phonetics）的研究是随着语音通信设备的广泛使用和相关案件的增加在近十年内开展起来。另外，我国刑法有关条文的改进，使得声纹鉴定得到一定的运用，这为司法语音学的发展鉴定了基础。目前我国声纹鉴定主要分为两个方面，一是语音样本的特征分析，另一个方面是声纹样本的自动检索。近期的研究包括：说话人自动识别系统的研究、噪声对声纹鉴定的影响研究、开口度对声纹鉴定的影响、录音剪辑检验的实验研究、利用鼻韵母共振峰特征进行声纹鉴定的研究、语音特性在鉴别双胞胎语音中的价值、语音的动态性及声纹鉴定的动态分析方法、电声伪装语音的声学研究等。目前比较典型的声纹鉴定研究还是语音的特征分析方面，如利用鼻韵母共振峰特征进行声纹鉴定的研究[40]反映了国内声纹研究与应用的水平。

4.11 民族语语音研究

我国有 55 个少数民族，已确定的语言就有 80 多种，分属汉藏、阿尔泰、南亚、南岛、印欧等语系。这些语言的语音学研究是语音学理论研究的基础。近期这方面的研究包括：金平傣语单字调的声学分析、毛南语中的鼻音对立、蒙古

语单词自然节奏模式、安顺仡佬语声调的实验研究、维吾尔语元音的声学特征分析、安多藏语语调的时长和基频的统计分析等。另外，少数民族学习汉语普通话的语音学研究也正在逐步开展。从大的方面看汉藏语的声调研究和阿尔泰语的重音及节奏研究对中国的语音学理论研究有重要意义。如，安顺仡佬语声调的实验研究[41]和蒙古语单词自然节奏模式[42]反映了民族语言语音的研究状况。

5 我国语音学的发展前景

纵观中国现代语音学的历史、研究队伍和研究现状可以预见我国现代语音学的研究在今后一个阶段将会蓬勃发展，这主要有以下几个原因：

(1)首先是我国的综合国力增强，国家在科研上的资助力度逐年增加，因此，有关课题的设置大量增加，以前无法做的课题现在都在逐步实现；

(2)教育的发展培育了大量研究人才，目前，我国具有博士学位的青年语音学研究人员大量增加，而且分布在各种不同的科研机构和高校；

(3)国家教育经费投入力度逐年加大，使得许多高校的汉语系、外语系和语言学系纷纷建立语音实验室，这使得语音学研究的硬件条件不断改善，和国际的差距也越来越接近；

(4)中国经济的发展，使得到中国学习汉语的人大量增加，汉语对外教学直接促进了语音学的研究和发展；

(5)言语技术的发展直接涉及许多语音通信领域的创新和开发，这些领域的需求反过来由对语音和言语的研究提出格更高的要求，形成相互促进的局面。

总之，可以相信中国的语音学研究会快速发展，并为语音学和言语工程的进步做出更大的贡献。

参考文献[1]

[1]刘复，1924 年，《四声实验录》，上海群益书社印行

[2]罗常培，1940 年，《临川音系》，历史语言研究所出版

[3]王士元，1983 年，《语言学论丛》，第十一辑，商务印书馆

[4]周殿福 吴宗济，1963 年，《普通话发音图谱》，商务印书馆，北京

[5]吴宗济 林茂灿，1989 年，《实验语音学概要》，高等教育出版社

[6]孔江平，2001 年，《论语言发声》，中央民族大学出版社，北京

[7]林焘 王理嘉，1985 年，《北京语音实验录》北京大学出版社，北京

[8]王士元，2006 年，宏观语音学

[9]郑秋豫，2006 年，台湾语音学及相关研究近况

[10] 李爱军，2006 年，面向言语工程应用的语音学研究

[11]汪高武 孔江平 鲍怀翘，2006 年，从声道截面积推导普通话元音共振峰

[12] 孔江平，2006 年，语音多模态研究和多元化语音学研究

[13]Sagisaka ， 2006 年 ， Towards computing phonetics

[14] 党建武，2006 年，Speech production and its modeling

[15]林茂灿，2006 年，赵元任汉语语调思想和边界调

[16] 徐云扬，2006 年，汉语方言元音的类型学研究

[17] 朱晓农，2006 年，中文语音学术语的几个问题

[18] 王嘉龄，2006 年，上海话广用式变调的优选论分析

[19]鲍怀翘，2006 年，辅音声学特征简议

[20]曹剑芬，2006 年，发音增强与减缩——语言学动因与语音学机理

[21]哈斯其木格，郑玉玲，2006 年，普通话元音过渡与辅音腭位关系解析

[22]石锋，王萍，2006 年，北京话一级元音的统计分析

[23]邓丹，石锋，吕士楠，2006 年，普通话双音节韵律词时长特性研究

[24]邝剑菁,王洪君，2006 年，连上变调在不同韵律层级上的声学表现

[25]王韫佳，初敏，2006 年，关于普通话词重音的若干问题

[26]李爱军，2006 年，情感句重音模式

[27]钱一凡，2006 年，民歌男高音共鸣的实验研究

[28]胡方，2006 年，论元音产生中的舌运动机制——以宁波方言为例

[29]谭晶晶，2006 年，新闻朗读的呼吸节奏研究

[1] 参考文献除了注明出处的其余的为第七届中国语音学学术会议的论文，并被《中国语音学报》第一辑录用，即将出版。

[30]郑玉玲 刘佳，2006 年，基于 EPG 的普通话辅音发音部位及约束研究
[31]李洪彦 黎明 孔江平，2006 年，听障儿童普通话声调获得研究
[32]顾文涛，广濑啓吉，藤崎博也，2006 年，汉语韵律结构:基于感知的韵律层级与基于 F0 曲线生成模型的语调分析之比较
[33]蔡莲红，2006 年，情感语音计算性研究的基本问题
[34]陶建华，2006 年，基于情感矢量的情感语音自动感知模型
[35]贾媛，熊子瑜，李爱军，2006 年，普通话焦点重音对语句音高的作用
[36]刘俐李，2006 年，基频归一和调系规整的方言实验
[37]关英伟,李波，2006 年，越南留学生习得汉语单字调的声学考察与分析
[38]李蕙心，2006 年，时长与广州话元音的感知
[39]朱春跃，2006 年，汉语表情话语中的调型改变及其感知
[40]王英利，2006 年，利用鼻韵母共振峰特征进行声纹鉴定的研究
[41]杨若晓，2006 年，安顺仡佬语声调的实验研究
[42]呼和，陶建华，格根塔娜，张淑芹，2006 年，蒙古语单词自然节奏模式

水语(三洞)声调的声学研究

An Acoustic Study on Tones of the Sandong Shui Language

汪锋

Wang Feng

提要 本文以贵州三洞水语为代表点,对水语声调在单音节和双音节中的表现做了详细的声学记录与分析，在大量声学数据的基础上，总结了水语单字调系统以及各声调在双音节前字和后字上的表现,结合时长材料的分析，认为前字时长短是声调简化的主要原因。文章还对个人变异以及性别差异做了比较和分析。

Abstract: This paper analyzes the acoustic data on tones of the Shui language in Sandong, the Guizhou province. Based on tonal representations among monosyllables and disyllables, the tonal system of monosyllables and the reflexes of each tone in both positions of disyllables are summarized. In connection to the data of duration, the short duration of the first syllable is thought to be the cause of simplification of tones in disyllables. Moreover, individual variation and gender difference have been analyzed.

关键词 三洞水语 声调 基频 时长 单音节 双音节

Keywords: The Sandong Shui; Tone; F0; Duration; Monosyllable; Disyllable

水语是水族的语言，一般划归壮侗语族侗水语支。水语主要分布在贵州的三都水族自治县，在贵州的榕江、荔波和广西的融水等地也有部分。本族人自称 ai3sui3。现有人口约 40 多万，而贵州境内超过 32 万。水语通常分为三个方言：三洞、阳安、潘洞。一般以三洞水语作为标准点。水语与汉语的接触比较频繁，历史也比较长久。在水语中有较多的汉语借词。

1 水语声调的相关研究

李方桂调查了贵州荔波 Ngam 和 Li 村的水语，认为舒声调有六个(Li 1948)，分别为 1 Ngam 村是低平，Li 村是低升；2 低降：中调起始，慢慢降到低调；3 中平；4 高降：高调起始，迅速降到中调或低调；5 高升：中调起始，然后升到高调；6 高平。促声调型有两个，类似于 4 调与 5 调。如 mit4‘刀’、mit5‘爱’。从声母与声调的配合关系，Li 1948 发现 1、3、5 调可以与任何声母配合，而 2、4、6 调则很受限制，根据汉语和泰语声调发展的启示，将前者判断为清声母来源，后者判定为浊声母来源。李方桂(Li 1949)对水语歌谣押韵做了详细研究，发现水语歌谣中三对声调(1 和 2；3 和 4；5 和 6)自由通押，而这一自由通押没有任何语音上的原因，同时水语中汉语老借词的声调对应表现出这三对声调分别与汉语的平、上、去对应，其中 1、3、5 调对应汉语阴调字，2、4、6 对应汉语阳调字。

张均如(1983：521)描写了三洞、阳安和潘洞，倪大白(1990:98)描写了瑶庆，如下：

表 1　水语的声调

调类			调值		调值
		三洞	阳安	潘洞	瑶庆
舒声调	1	24	11	12	13
	2	31	31	31	31
	3	33	33	22	33
	4	42	54	53	53
	5	35	35	35	35
	6	55	23	44	55
促声调	7 短	55	55	55	55
	7 长	35	35	35	35
	8 短	43	31	32	53
	8 长	43	31	31	53

张均如还提到“阳安、中和等地因吸收汉语借词增加了一个高平调，相当于汉语的上声调”。

而夏勇良、姚福祥(1994:146)根据三洞乡达便村的水语描写如下：

舒声调：			
第一调	˧¹³	dal(ta¹)肥	hal(ha¹)肩膀
第二调	˨³¹	mac(ma²)舌头	hac(ha²)药
第三调	˧³³	bas(pa³)伯母	has(ha³)杀
第四调	˥⁵³	gax(ka⁴)汉族	hax(ha⁴)酒药
第五调	˧³⁵	dav(ta⁵)中间	hav(ha⁵)呵（气）
第六调	˥⁵⁵	mah(ma⁶)价钱	fah(fa⁶)糖
促声调：			
第七调（长）	˧³⁵	lagv(la:k⁷)绳子	hagv(ha:k⁷)舂（米）
第七调（短）	˥⁵⁵	laegh(lak⁷)洗（衣）	haegh(hak⁷)拥挤
第八调（长）	˩⁴²	dagt(ta:k⁸)测量	hagt(ha:k⁸)屋基
第八调（短）	˩⁴³	daegt(tak⁸)雄性	gaegt(kak⁸)层

就三洞水语而言，这些描写的主要差别是：(1) 在 1 调上，袁、姚(1994)定的 13 比张均如(1987)的 24 要低 1 度; (2) 在 4 调上，前者的 53 要比后者的 42 高 1 度; (3) 而在 8 调上，张均如(1987)认为在长、短元音条件不同的情况下调值相同，而袁、姚(1994)认为在长元音的条件下该声调 42 的终点要比短元音的 43 终点低 1 度。(4) 袁、姚(1994)还认为，第 5 调有先降后升的趋势，实际应该标为[325]。

上述这些描写主要是基于记音人的听感。而此类描写的分歧可能源自多种语言变异，比如：人际差异、声调内部变异、记音人的不同标准、发音人的发音时候的不同风格等。而在研究中如何才能处理这些差异需要深入研究。对声调的声学分析或许能够提供一个较为客观的讨论基础，尽管上述造成差异的因素难以避免，但能够通过各种手段加以合理的处理。(孔江平 2007)

2 水语声调的声学实验

为了更准确地描写水语的声调，在编制录音词表的时候，对单音节声调和双音节连字调进行了全面的考虑。对于单音节，我们按照 10 种调类进行选词，即促声调 7 调和 8 调都按照长短元音音节再一分为二，每个调类选词 5 个。对于双音节，由于选词的困难，按照 8 个调类进行组合，得到 8×8=64 种双音节调类组合，每种组合都是 4 个。

三洞水语发音人男女各 2 名，均为 40 岁左右。每个词发音 2 遍，这样每种调类或调类组合就可以得到 24 个语音样本。采用 Cooledit 2.0 录音，左声道采集声音信号，右声道采集声门阻抗信号，采样频率为 22050Hz。

本研究用自相关的方法提取基频。以实验室编写的语音处理系统来人工标记声调的负载音段，对部分采点失误进行校正。再以Matlab程序来提取基频，舒声调采 30 个点，促声调采 15 个点[1]。最后，选取男女两组样本（各 12 个样本）的基频值分别进行平均，再将这些数值输入 Excel，就可以得到水语单音节声调和双音节连字调的基频曲线图，这些基频曲线图就是本文讨论的基础。另一方面Matlab程序同时还提取出声调的时长，以供分析之用。

3 声学实验结果及分析

3.1 单字调

根据上述步骤，我们得到的基频曲线图如下(T1 表示 1 调; L 表示长元音；S 表示短元音；以此类推)：

图 1 三洞水语的女声单字调基频

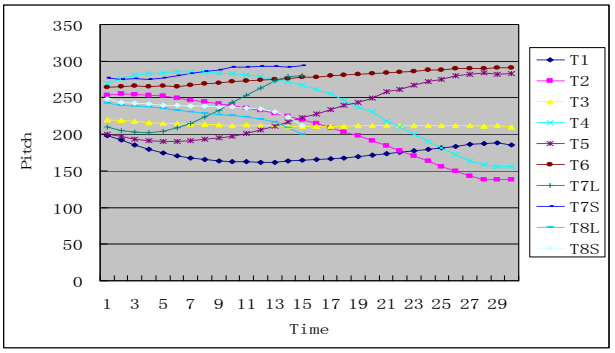
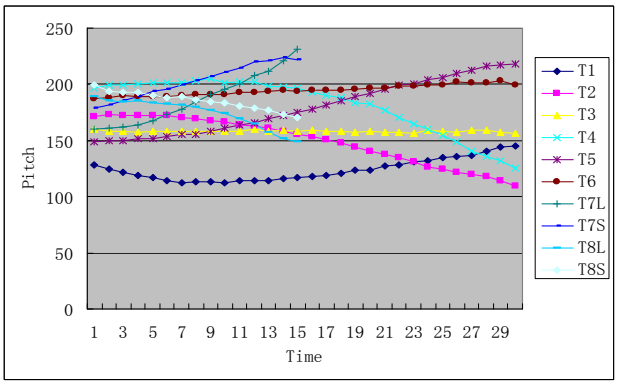


图 2 三洞水语的男声单字调基频



基频是声调的物理基础，但是只有基频值，还不能直接说明语言学的问题，因此可采用下面的公式将基频值转换成 5 度值：

$$5\text{度值} = ((\lg x - \lg b)/(\lg a - \lg b) \times 4)$$

[1]本研究用到的程序均为孔江平、吴西愉编写，特殊说明的除外

+ 1[1]

从而得到水语声调的基频情况如下：

T1： 女声数据平均后，可以看出有比较明显的先降后升的“凹”特征，在整个声调的中点附近倒到最低点。如果根据 5 度法的处理，以 323 的标调为宜，但起点要比终点略高。男声数据表现出同样的调型特征。如果依 5 度法处理，以 213 为宜，与女声不同的是，男声终点比起点要高，前半段更低。但在前人的研究中，该调的曲折都被忽略不计，因此，一个值得进一步思考的问题是，基频上表现为曲折的声调，拐点在什么情况下才能被听出来？

T2： 女声 T2 调是一个降幅较大的降调，其终点在整个声调域中最低，根据 5 度法处理，可以标为 41 调。男声调型同样，但降幅比女声略小，可以标为 31。

T3: 女声、男声数据各自平均后，该调都是一个典型的中平调，可以标为 33。

T4: 女声前半段微升，后半段下降，可以标为 452。但其起点比也标为 4 的 T2 要略高些。男声调型同样，但升幅比女声更小，或许应标为 442。

T5：女声数据平均后，前三分之一段有很小的降幅，之后一直升到调域的最高处。可以标为 35。男声数据一直保持上升的状态，没有前面的降幅，是典型的 35。

T6：女声一直处在最高处，可标为 55，但不如 T3 那么平直，表现出一个微升的趋势。男声趋势同样，但相对高度没有达到最高处，标为 44 较合适。

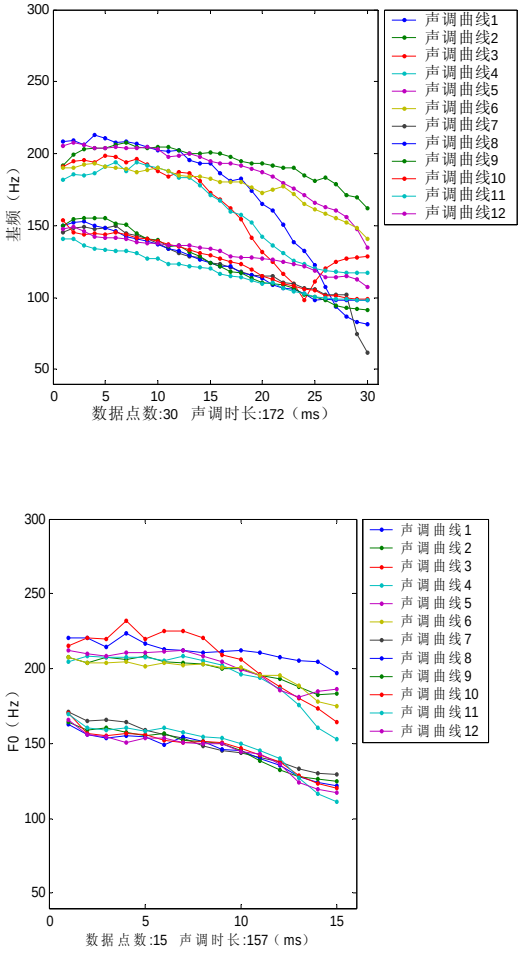
T7：女声 T7L 在长元音中，表现为一个升调，与 T5 类似，开头一小段有很小的降幅，可以标为 35。但其整个长度要比 T5 短很多。男声是典型的 35，没有开始的小降幅。有意思的是，T7 似乎与 T5 的调型总是保持同步，在女声中都有降幅，在男声中则都没有。女声， T7S 在短元音中，与 T6 的模式很相似，有微升趋势，可标为 55，在绝对值上都要比 T6 略高些，长度相对短很多。男声表现出更大的升幅，以标为 45 合适。这样看来，似乎女声中 T7 的分化要比男声中的清楚得多，后者甚至可以不再细分。

T8：女声 T8L 在长元音中，表现为一个降调，幅度不大，可以标为 43；T8S 在短元音中，表现为一个降调，幅度也不大，可以标为 43。但是，与 T8L 相比，则略高。男声表现出同样的趋势，可以标为 43，但在短元音中要略高一些。

3.2 单字调的个人变异

在男声数据中，可以看出其中一个样本明显比另外一个样本低一度左右，但调型几乎保持为完全一样，以 T2 和 T8L 为例：

图 3 男声的差异



男声 1 和男声 2 各六个样本，各自聚集成束，男声 2 基频低，与上部的男声 1 分割十分清楚。从听觉上，也能明显听出男声 2 相对要低。但在 T1 调中，二者的差别不明显。而在女声数据中，两个样本之间就没有什么差别。这些人际差异提

[1]公式中的 ***a*** 和 ***b*** 分别代表整个调域中基频数值的上限和下限，***x*** 代表的是数据点的基频值。基频值换算成 5 度值时，就是把各终点的数值分别代入 ***x*** 进行计算。

醒我们在田野调查时要根据研究目的的不同，选取合适的数据收集方法，在分析语音时要充分考虑到人际差异造成的影响(孔江平 2007)。

倪大白(1990:289)：“长短元音与声调之间，一般来说，似乎没有必然的联系，特别是舒声调，不论对长元音韵还是短元音韵几乎总是一视同仁。但在促声调里，情况就不一样了。侗台语的不少语言和方言，长元音韵的促声调和短元音的促声调，调值往往不一样。”在三洞水语中，值得注意的是，促声调中，7 调音节上呈现的是元音长短与调型不同的双重对立，从音系处理的角度看，似乎适合将元音长短处理为主要区别特征，而调型差异处理为羡余特征，因为长短元音的对立在其它调类情况下仍旧是对立的特征，因此，7 调中的两类调型是可以根据长短元音的条件预测的，可以算作 7 调这个调位的两个条件变体。而 8 调音节中，无论长短，调型都是一致的，在男声和女生的材料中都是同样的反映。

3.3 二字调

前字声调情况同样按照男/女分开考察。汇集某个特定的声调在 8 个声调中的分布(每种组合每个发音人的样本中取 6 个，共 6*2=12 个)来比较，如果没有变换调型的情况，则取平均，与单字调对比。(7 调和 8 调的长短由于材料的限制，不再细分，8 调在长短元音中差别不大，合起来考察应该没有问题，但 7 调要分高平调和升调两类，下文中根据实际情况有详细讨论。)

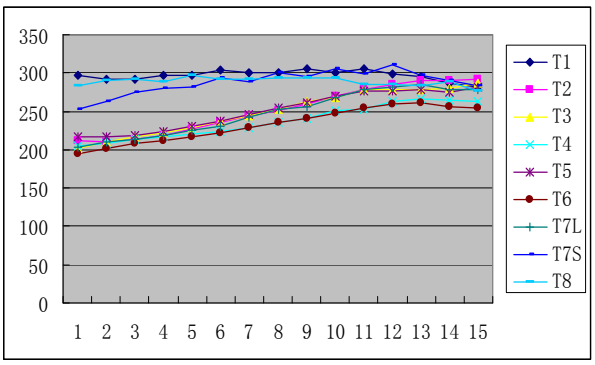
3.3.1 女声前字

T1 作为前字与单字调比，“凹”的特征不明显，而体现为先降后平，以 5 度法，大致为 322。T2 在前字上基本都保持调型，在后字是 T8 时，整体音高略有提高。除此以外，取平均之后与单字调对比，可以发现在连调情况下，T2 的收尾要高出 1 度，即可以标为 42。T3 在各类调之前的表现如一。平均后与单字调相比，除了略高一些外，调型没有什么差别。T4 在各类调之前的表现如一。平均后与单字调相比，调型没有什么大差别，但收尾要高 1 度。

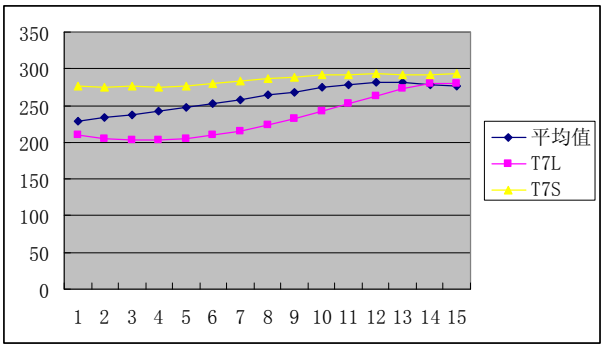
T5、T6、T7L、T8 作为前字没有什么变化，与单字调相比也相同

T7S 则出现了变调：在 T2,T3,T4,T5,T6, T7L 调前，变为升调；在 T1,T7S 和 T8 前，则保持高平调。

图 4 T7S 的变调



这样，如果取平均值的话，自然就落在单字的两调之间了。

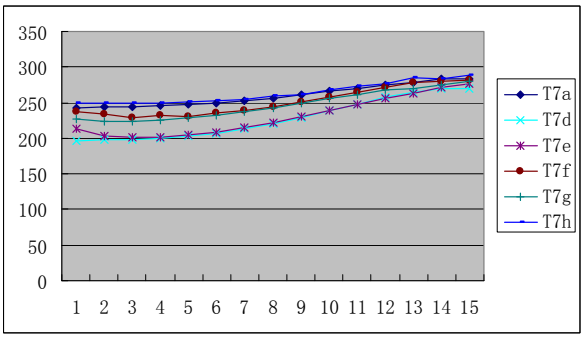


3.3.2 女声后字

T1、T2、T4、T5、T8 在后字位置上，与单字几乎完全一样。

T3 调型与单字调完全一样，但要高 1 度。T6 与单字调相比，调型也几乎没有差异，但整体平均值要低约半度。

T7 有略升的趋势，更多的是与 T7S 有相一致的倾向。



3.3.3 男声前字

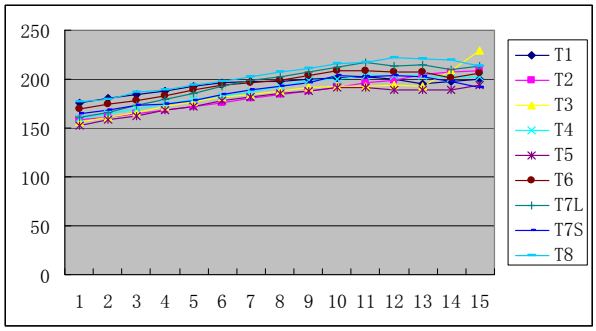
T1 在后调是 T4 时，基频明显降低。但在各调中，T1 调型没有大的变化。相比较单字调而言，在双音节前字的 T1, 趋向于微升的平调，

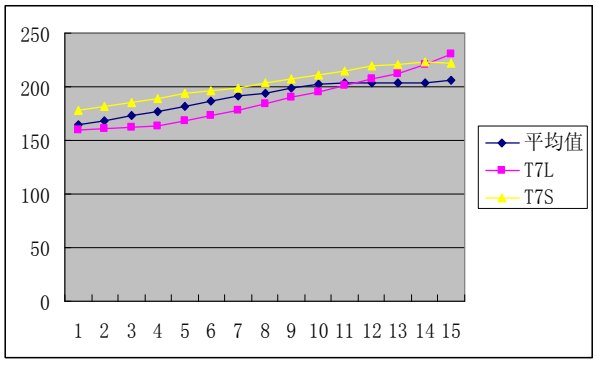
整体比单字调略高。T2 调型比较一致，呈降调。相比较单字调而言，在双音节前字的 T2 调型保持一致，但整体比单字调高出约 1 度。T3 调型比较一致，呈平调。相比较单字调而言，在双音节前字的 T3 调型保持一致，但整体比单字调略高。

T4 调型基本保持一致。相比较单字调而言，基本保持一致，但收尾比单字调高出约 1 度。

T5、T6、T8 调型与单字调基本保持一致。

T7L 在前字位置上表现基本一致，但在平均值上介于单字调的 T7L 和 T7S 之间。





而 T7S 则与单字调 T7S 完全吻合。

从这个角度说，T7L 在男声中处于不稳定状态，在双音节前字时从基频上看是两可的。再加上男声的这两种长短调的区分只是 35 与 45 的区分，本身在基频上的区分度不大。因此，在双音节状态下似有混同的倾向。在加上双音节词赖以区别的特征比单音节词多，可能又加强了这一倾向。

3.3.4 男声后字

T1 基本一致，只是比单字调在起点略高。T2、T4、T5、T8 与单字调一致。T3 基本一致，比单字调高约 1 度。T6 基本一致，只是比单字调略高。T7 在后字上基本上都保持与 45 更接近的升调趋势。

3.3.5 小结

从上述分析中可以看到，男\女差异很重要。突出表现在 T7 根据长短元音的变化上不同，女声分化相对明显，在单字调上调型明显不同，而在双音节中，女声的 T7S 还发生变调，变为 T7L，具体细节见前文；而男声在单字调上即区分不大，在双音节的前字中更是倾向于完全混同。

声调在双音节前、后字位置上的表现有差异，在前字位置上更容易简化，甚至导致变调。

3.4 时长研究

在单字调中，短元音促声韵明显比舒声韵要短。其他单字调时长的变异度比较大，每个调的长短并不恒定，比如，T1 和 T6 的音节类型上没有什么差别，但在女声中 T1(84)比 T6(344)短很多，而在男声中，T1(318)比 T6(245)要长一些。

但在双音节组合中，一个十分明显的趋势是前字比后字短。只有在前字为舒声，后字为促声的组合中，才可能出现前字比后字长的情况。抽取每一组组合中的第 12 个样本，同样显示出这样的趋势：女声数据中显示，舒声调为后字时，前字的时长一定比后字短，没有反例。都是促声时，也只有一个反例[8+8]，前字比后字略长[12ms]。在前后调相同的情况下，也只有一个反例[8+8]。男声也存在这一倾向，但例外有所增多(1+2 ; 1+5; 3+4 ; 5+4 ; 5+6)。

这一趋势是如何形成的呢？通过考察所有样本的前后字的类型及数量，可以排除音节类型的因素。因此，可以采取后字拖长的一般解释，也就是，水语发音人的发音策略是“先紧后松”。进一步说，该趋势是由位置决定的，而与其他因素无关。

在§3.3 的讨论中，我们可以看出，前字的调型都有简化的趋势，而后字却能与单字调保持一致。因此，可以认为由于水语发音人在发双音节时采取了“先紧后松”的策略，导致了前字音节时长变短，进而造成了调型和音高等的变化，而后字有充分的发音时间，因此可以保持与单字调一致。如果这种趋势不受其他新的因素干扰，我们或许可以预测，如果水语将来出现更多变调的话，应该是在前字位置先变。进一步推测，如果双音节发音策略采取先松后紧的状况，就有可能

导致音变时候的后字变化。

4 结语

通过上述数据分析，水语的单字调系统可以归纳如下：

调类		调型	调值(男)	调值(女)
舒声调	1	降升	213	323
	2	降	31	41
	3	中平	33	33
	4	升降	442	452
	5	升	35	35
	6	高平	44	55
促声调	7 短	高平	45	55
	7 长	升	35	35
	8 短	降	43	43
	8 长	降	43	43

在双音节系统中，除了 T7S，各调在前字位置上不变调。在男声前字中，T7S 在调型上与 T7L 基本混同(参见 3.3.3)，而在女声前字中，T7S(在 T2,T3,T4,T5,T6, T7L 调前)，体现为升调,与 T7L 的单字调一致；而在 1 调、T7S、T8 前，为高平调，与 T7S 的单字调一致。这种新发现的趋势似乎体现了 T7S 向 T7L 演变的趋势，但在男、女中的表现却并不相同，男声中没有后字限制条件，或者说限制条件在演变完成后已经看不出来了，而在女声中则仍有后字限制条件。在张均如(1983：521)、倪大白(1990:98)等早期研究中，并没有报道 T7S 向 T7L 的调型趋同的情况，或许可以据此推断，这是近来发生的新变化，也就是在长短元音对立的情况下，调型作为羡余特征而容易变化，而且 T8 的长短音节不分调型的模式对此变化也可能是一种促进。

后字位置上，各调基本保持与单字调一致。

水语声调系统在整个侗台语族中都显得十分稳定，少有变调。其双音节中的前字易发生简化，而后字十分稳定，如前所论，这与水语双音节词前字时长相对要短或许有密切关系。

参考文献

[1]孔江平. 2007. 语音学田野调查的一些基础理论问题. 语言学论丛 36.

[2]倪大白. 1990. 侗台语概论. 北京：中央民族大学出版社.

[3]夏勇良、姚福祥. 1994. 三洞水语的音系. 贵州民族研究 59:142-148.

[4]张均如. 1980. 水语简志.北京:民族出版社.

[5]张均如、梁敏. 1996. 侗台语族概论. 北京：中国社会科学出版社.

[6]Li, Fang-kuei. 1948. The distribution of initials and tones in the Sui language. Language 24: 162-7.

[7]Li, Fang-kuei. 1949. Tones in the rhyming system of the Sui language. Word 5. 262-7.

汉语普通话不同文体朗读时的呼吸重置研究

Breathing-Resets in Reading Aloud the Texts of Different Literature Genres in Mandarin Chinese

谭晶晶 李永宏

Tan Jingjing, Li Yonghong

摘要： 本文目的在于探讨人们在使用汉语普通话朗读不同文体时的呼吸重置特点。本文运用肌电脑电仪和呼吸带传感器测量了发音人在朗读不同文体时呼吸节奏的变化，通过自动标注提取出呼吸重置的幅度和时长，并分析了它们的频度分布，发现在朗读不同文体时，发音人的呼吸节奏有较明显的差别，朗读韵文和非韵文时，呼吸重置分别可以分成 2 级和 3 级。实验结果表明，呼吸重置的时长和幅度显著正相关，重置幅度比重置时长更能反映不同文体的差别，也更能反映不同呼吸级别的大小。

Abstract： This paper is a preliminary study on the breathing-reset characteristics in reading aloud texts of different literature genres in Mandarin Chinese, namely, Chinese ancient poetry, song lyrics, novel, prose and news. Breathing signals are recorded by EMG and respiration sensor. The parameters of duration and amplitude of breathing-resets are automatically detected. Some conclusions have been drawn by analyzing the frequency distribution of the duration and amplitude: 1) The breathing rhythm varies with the literature genres. There are two levels of breathing-resets in verse-reading and three in essay-reading. 2) The duration and amplitude are highly-correlated across genres. 3) The amplitude is more useful in indicating the literature genres and the hierarchies of breathing-resets.

关键词：语音生理学 韵律 呼吸重置 文体

Key words：physiological phonetics, prosody, breathing-reset, literature genres

0 引言

语篇的韵律特征是近年来语音学界研究的一大热点。由于语流的韵律构造是一个复杂的过程，它受到生理、心理、语法等各个方面的制约，因此研究者的角度也各不相同。目前的研究大多集中在音系学、声学和心理学方面，如文[1]、[2]、[3]，从生理角度进行研究的则并不多见。在这一方面，Liberman（1967）[4]曾提出过一个“呼吸群（breath-group）”的理论，认为人们是通过呼吸群来产生和感知语调的，在郑秋豫（2005）[5]近年的研究中，也把呼吸看成是划分韵律层级的重要线索。“呼吸群”的理论可以有效地解释语流在实现过程中受到的生理因素的制约，也能有效地解释语流中的音高下倾现象，可是目前完全从生理（呼吸）的角度对汉语语流的韵律问题进行的研究还非常少见。

文[6]利用肌电脑电仪和呼吸带传感器记录了呼吸信号，对新闻朗读中呼吸节奏的变化做了一些分析，试图寻找人们朗读时呼吸节奏的类型，探讨呼吸节奏和韵律结构的关系，发现呼吸结构可以反映人们对表达内容的整体认知规划，同时它也反映了人们生理机能的制约。由于当时实验手段的制约，还不能直接提取呼吸信号的参数，划分呼吸重置级别时，主要依靠人工判断。最近，本文编写了程序来自动提取呼吸重置的时长和重置幅度，并用它对近体诗、词、小说、散文、新闻等不同文体的朗读语料进行了分析，试图寻找人们在朗读不同文体时呼吸节奏的特点。

1 语料说明

1.1 实验语料

本实验选取了 1 名发音人朗读的近体诗 20 首、词 30 阕、小说 20 篇、散文 20 篇、新闻 40 篇，为了尽量减少发音人的断句错误，本文选取的都是发音人比较熟悉的语料。该发音人为北京大学电视台新闻播音员，女，22 岁，受过良好的播音训练。

1.2 语料录制

本实验的录音工作在北京大学中文系的隔音室内进行。录音使用的主要设备是澳大利亚 PowerLab 公司生产的肌电脑电仪，本实验用它

自带的录音软件 Chart5 采集了 4 个通道的信号：1 通道为语音信号，2 通道为嗓音（EGG）信号，3 通道为通过 MLT1132 呼吸带传感器采集的呼吸信号，4 四通道闲置，如图 1 所示。

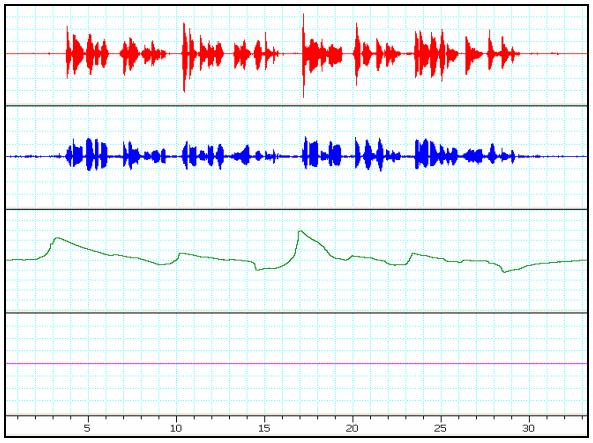


图 1 用 Chart5 采集的四通道信号

2 参数提取

2.1 参数定义

将呼吸信号的变化反映在二维图谱上，横轴是时间，纵轴是由呼吸导致的电压变化（称之为“呼吸曲线”），呼吸曲线上升表示吸气，下降表示呼气，如图 2 所示。

本实验提取的参数主要是呼吸重置的时长和重置幅度，其中重置时长指的是从开始吸气到开始呼气之间经过的时间，即图 2 中波谷和波峰之间的水平距离 T，重置幅度指的是一次吸气过程中呼吸信号数值的变化幅度，即图 2 中波谷和波峰之间的垂直距离 A。

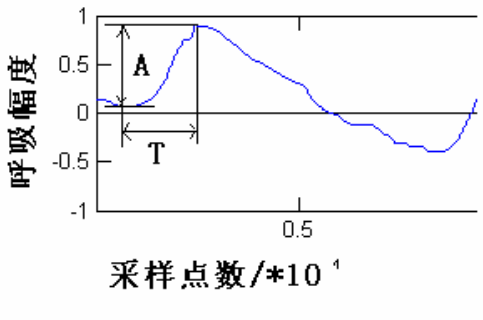


图 2 呼吸重置

2.2 程序说明

本实验采用的呼吸信号处理程序是北京大学中文系语音乐律平台子程序，在 Windows 平台下用 Matlab 编程实现。该软件功能包括：自动或手动对呼吸信号的重置进行标记，包括重置谷底标记和重置峰值标记，手动标记用不同的标记形式对不同大小的呼吸重置分别进行标记，主要的参数为重置谷底的位置和相对幅值，重置峰值的位置和相对幅值，计算每一个呼吸重置需要的时长和幅度，并保存自己设置的带参数的四通道文件，可以用带参数的四通道读取函数直接读取，标记直接显示在呼吸信号之上，便于复查，最后对标记好的 wav 文件进行批处理，自动读标记数据到 Excel 文件中。具体处理步骤如下：

1}. 考虑到朗读不同文体时呼吸信号幅度和时长的可比性，首先对第 3 通道的呼吸信号进行归一化处理。breathe_signal 为原始采集的呼吸信号。

```
breathe_signal=
breathe_signal /max(abs(breathe_signal));
```

2}. 对归一化后的呼吸信号进行平滑滤波。由于采集的原始呼吸信号带有很多细微的高频和部分干扰信号，因此对归一化后的呼吸信号进行了低通平滑滤波，滤波器采用零相位数字滤波 filtfilt(b,a,x)，通过将输入数据前向和反向处理，以完成零相位数字滤波。它先将数据按顺序滤波，然后将所得结果逆转后反向通过滤波器，这样得到的序列为精确零相位失真，并使滤波器的阶数加倍。Filtfilt 通过与初始条件相匹配。可使起始和结束阶段的暂态过渡过程最短[4]。

Filtfilt 差分方程表示为

```
y(n) = b(1)_x(n) + b(2)_x(n-1) + ... +
b(nb+1)_x(n-nb)- a(2)_y(n-1) - ... -
a(na+1)_y(n-na)
```

参数 b 序列的取值越小，滤波后的信号就越接近原始信号，但运算量越大，耗费的时间就越多，一方面考虑对高频信号的滤除，以满足呼吸的处理，另一方面要误差尽量小，方差限制在一定范围内，参数的取值为：

```
b=ones(1,400)/400
a=1
```

3}. 自动标记波峰和波谷。对平滑后的呼吸信号，利用局部最大法，综合考虑呼吸重置峰值点之间的距离和幅度差、谷值点之间的距离和幅度差，多次迭代循环，尽量消除小的类似于呼吸重置的峰和谷，最终获得能满足要求的呼吸重置的峰值和谷值。程序中还设置了手动修改功能，

如果标记位置有误，可进行手工调整。

4}．手动标记波峰和波谷。经过对呼吸信号的大量研究，大致根据呼吸重置的幅度大小和时长，确定出正常人说话，一般有 3 级呼吸，程序中可以分别对每一级别的呼吸重置和呼吸单元进行标记，也可以作为自动标注的补充手段。

5}．参数保存。标记打完后，可以按实验室自己的格式保存为带参数的四通道 wav 文件，所有的标记参数都保留在原始 wav 文件的后面，参数按国际 RIFF 标准定义。

6}．对提取呼吸参数的 wav 文件进行批处理，程序中针对手动和自动标记分别做了 2 套自动提取参数的程序，能够直接把文件夹下面的所有 wav 文件的标记数据直接读取到 xls 文件中。

3 实验结果

3.1 近体诗[1]

本实验选取了五言律诗、七言律诗、五言绝句、七言绝句各 5 首。由于近体诗的格律严格，结构简单，因此呼吸信号的类型也比较简单。其呼吸信号的特点是：无论是五言诗还是七言诗，每一联（每联包含两小句）开始处都有一个呼吸重置，七言诗则每小句开始处都有一个呼吸重置，小句开始处的呼吸重置比每一联开始处的呼吸重置小。如图 3 所示。

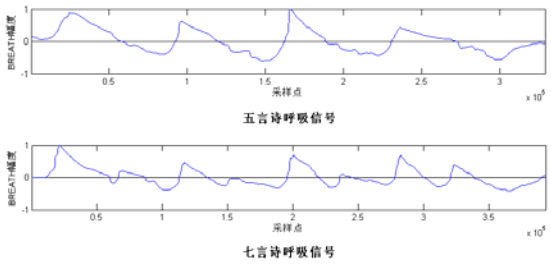


图 3 五言诗（上）和七言诗（下）呼吸信号示例

图 3 为五言律诗《天末怀李白》和七言律诗《登高》的呼吸信号，这 2 首诗都有 4 联（共 8 小句），每一联开始处都有一个较大的呼吸重置，七言律诗由于每小句字数较多，无法一口气念完一联，因此在每一联的第 2 个小句开始的时候会先吸一小口气，这样就形成了一个比较小的呼吸重置。

[1]格律诗有律诗和绝句之分，律诗分四联八小句，绝句为两联四小句，每小句 5 或 7 个字，即“五言”、“七言”。

从图 3 中可以看出呼吸信号的重置幅度有大有小，为了探讨大小不同的呼吸重置幅度的频度分布，本文自动提取了 20 首近体诗中所有的呼吸重置时长和重置幅度（共计 119 处），在 SPSS12.0 中将它们画出直方图，如图 4、5 所示。

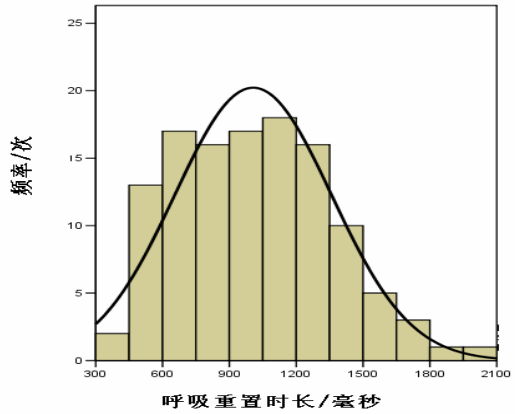


图 4 近体诗重置时长频度分布图

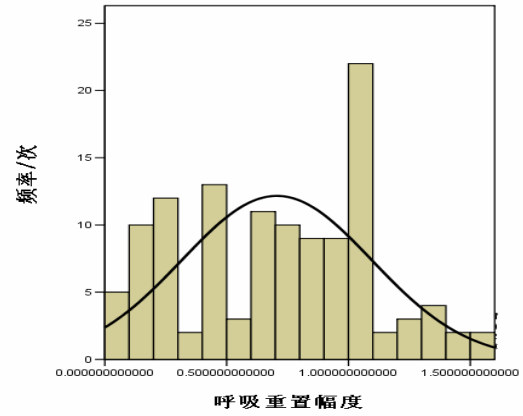


图 5 近体诗重置幅度频度分布图

从图 4、5 中可以看到，近体诗重置时长的频度分布基本符合正态分布，重置幅度的分布则在 0.50 和 1.00 处分别出现了 2 个较明显的峰值，尤其是 1.00 处频度达到 20 以上，这说明近体诗的呼吸重置大致可以分成 2 级。这种频度分布可以从近体诗的呼吸节奏特点中得到解释：每首诗都会有大呼吸重置，但是只有七言诗有小呼吸重置，因此大呼吸重置的数量就要比小呼吸重置的数量多。

3.2 词[2]

本次实验选取了 10 个不同词牌，每个词牌选取 3 首词，共 30 阕词，涵盖小令、中调和长

[2] 词按字数多少分为小令（58 字以下）、中调（59-90 字）、长调（91 字以上）。词按“阕”计，一首词为一阕，有些词分为上下两段，称为“上阕、“下阕”。

调。词的呼吸信号有如下特点：相同词牌的三阕词呼吸节奏完全相同，每阕词的上阕和下阕开始处都有一个较大的呼吸重置，除长调的呼吸重置幅度大致可以分成大、中、小 3 级之外，其余词的呼吸重置大致可以分成大、小 2 级。如图 6 所示：

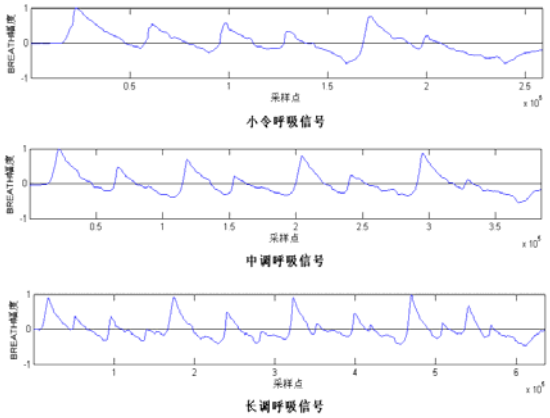


图 6 小令（上）、中调（中）、长调（下）呼吸信号示例

我们将 30 阕词中所有的呼吸重置时长和重置幅度（共 229 处）画出如图 7、8 所示的频度直方图，从图中可以看出，词中的呼吸重置时长和重置幅度都有 2 个较明显的峰，尤其是重置幅度的频度分布图中 2 个峰（分别出现在 0.3 和 1.0 处）基本对称，这说明将词中的呼吸重置分为 2 级是比较合理的。

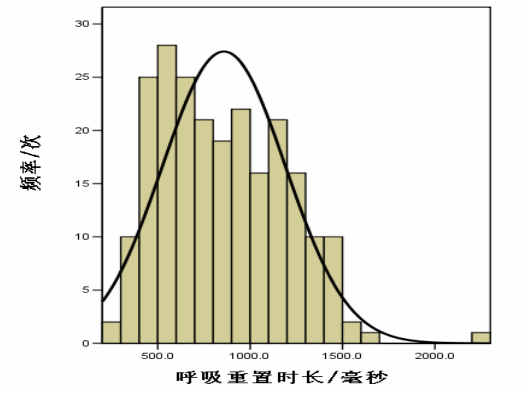


图 7 词重置时长频度分布图

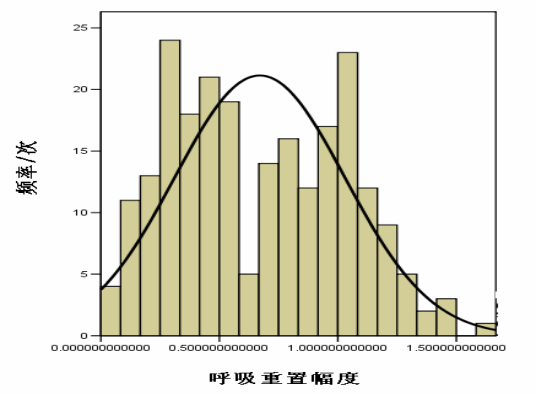


图 8 词重置幅度频度分布图

3.3 小说

本次实验从可以作为现代汉语典范的白话小说中选取了 20 个完整自然段作为实验材料，每段大约 250 字。 小说呼吸信号的特点是：每段小说中会出现 1~3 个较大的呼吸重置、若干个中等大小的呼吸重置以及很多小呼吸重置，如图 9 所示。

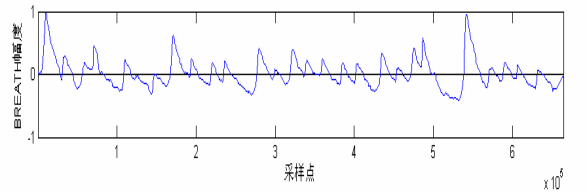


图 9 小说呼吸信号示例

从图 10、11 中可以发现：小说的呼吸重置时长和重置幅度的频度分布呈偏态分布，重置时长在 500ms 处有一个高峰、此后频度急剧下降，重置幅度在 0.20 处有一个高峰，重置幅度的频度下降的趋势与重置时长相比显得比较和缓，0.60~1.20 之间有一段比较稳定。这种频度分布也反映了小说大呼吸重置少、小呼吸重置多的特点。

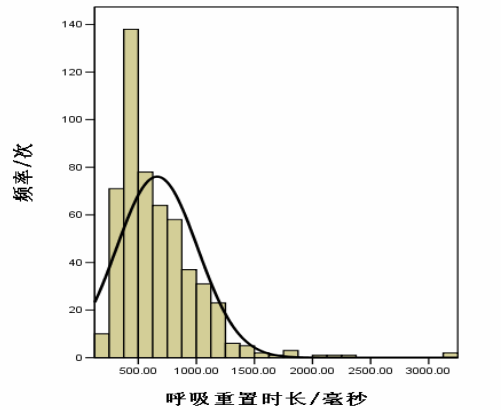


图 10 小说重置时长频度分布图

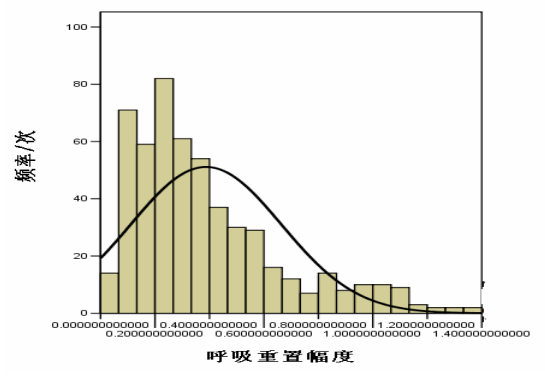


图 11 小说重置幅度频度分布图

3.4 散文

本实验选取的 20 段散文来自于可以作为现代汉语典范的白话散文，字数也都控制在 250 字左右。散文呼吸信号的特点和小说非常相似：每段散文中会出现 1~3 个较大的呼吸重置、若干个中等大小的呼吸重置以及很多小呼吸重置，如图 12 所示。

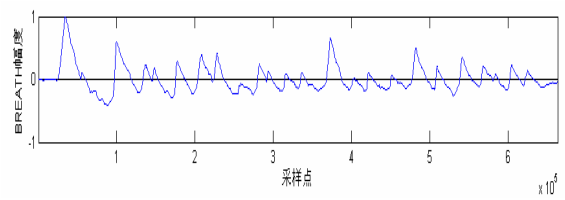


图 12 散文呼吸信号示例

散文呼吸重置时长和重置幅度的频度分布也是偏态分布（如图 13、14 所示），重置时长在 300ms 处有一个高峰，频度高达 180 余次，之后时长频度迅速下降。重置幅度则在 0.12~0.40 之间分布非常集中，之后在 1.00 处又有一个小峰。这种分布说明朗读散文时发音人的呼吸节奏也呈现出大呼吸重置少、小呼吸重置多的特点。

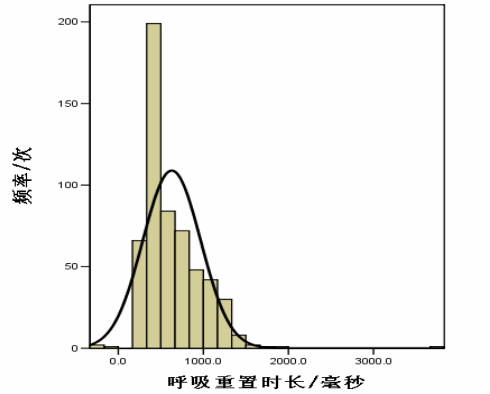


图 13 散文重置时长频度分布图

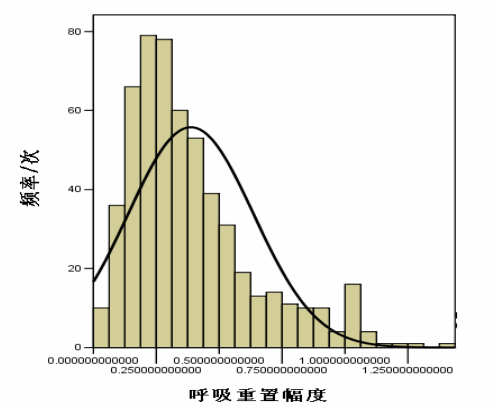


图 14 散文重置幅度频度分布图

散文和小说的呼吸信号从重置时长和重置幅度的分布频度方面来看非常相似，只是在具体的频度峰值上有细微的差别，这可能是因为散文和小说从广义的文学体裁上来说都属于和“韵文”相对应的“散文”，因此朗读的时候在节奏上会有相似之处，在朗读小说和散文时能感受到它们之间的区别，可能有其他方面因素（如音高、语速甚至语义）的影响。

3.5 新闻

新闻的文章结构有其固定的格式，发音人在朗读新闻时也容易形成相对固定的呼吸节奏模式。文[6]曾分析过该发音人朗读 40 篇新闻的呼吸节奏，认为在新闻朗读中呼吸节奏可以分为 3 级呼吸单元，分别和自然段、复句和分句相对应，相应的也有 3 级大小不同的呼吸重置，如图 15 所示：

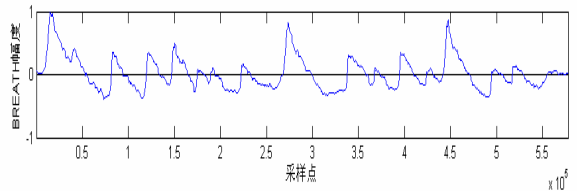


图 15 新闻呼吸信号示例

在本次实验中，本文自动提取了这 40 篇新闻中共 671 处的呼吸重置时长和重置幅度，并将其画成了频度分布直方图，如图 16、17 所示：

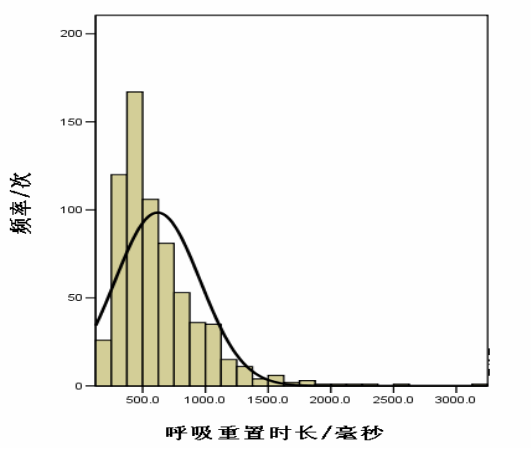


图 16 新闻重置时长频度分布图

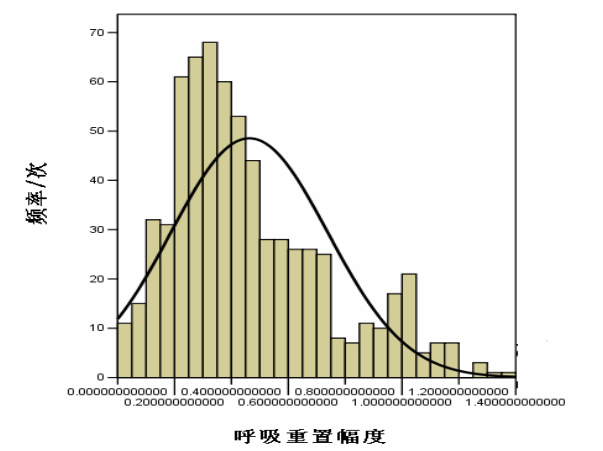


图 17 新闻重置幅度频度分布图

从图 16 中可以看到，新闻的呼吸重置时长呈明显的偏态分布，在 400ms 处有一个高峰，随后频度分布曲线急剧下降，重置时长频度的这种分布和小说、散文非常相似。

从图 17 中可以看到新闻的呼吸重置幅度在 0.30 处有 1 个很明显的高峰，在 1.00 处也有 1 个比较明显的小峰。新闻和小说、散文的不同之处在于，新闻大呼吸重置的数量比散文、小说多，另外，新闻的重置幅度在 0.50~0.80 之间有一段比较稳定的频度分布，这一段大致对应着中等的呼吸重置，这一级的呼吸重置数量比散文、小说多。这说明之前将新闻的呼吸重置分成 3 级是有道理的。

3.6 呼吸重置时长和重置幅度的关系

本文运用 SPSS12.0 对呼吸重置时长和重置幅度作了相关分析，得到如下结果：

Correlations

			duration	amplitude
Kendall's tau_b	duration	Correlation Coefficient	1.000	.517(**)
		Sig. (2-tailed)	.	.000
		N	2108	2108
	amplitude	Correlation Coefficient	.517(**)	1.000
		Sig. (2-tailed)	.000	.
		N	2108	2108
Spearman's rho	duration	Correlation Coefficient	1.000	.692(**)
		Sig. (2-tailed)	.	.000
		N	2108	2108
	amplitude	Correlation Coefficient	.692(**)	1.000
		Sig. (2-tailed)	.000	.
		N	2108	2108

** Correlation is significant at the 0.01 level (2-tailed).

表 1 重置时长和重置幅度的相关性统计结果

重置时长和重置幅度之间的 Kendall 和 Spearman 等级相关系数分别为 0.517 和 0.692，说明二者之间是正相关关系，重置时长越长，重置幅度越大。经过双尾 t 检验，发现其显著性水平为 0.01，也就是说重置时长和重置幅度之间的正相关性非常显著。

3.7 各种文体重置幅度的共同点

尽管各种文体呼吸重置幅度的频度分布各有其特点，但是仔细观察，也可以从中找到一些共性：

1）虽然频度大小不一，但各种文体的呼吸重置幅度在 1.00 处都有一个小高峰。这是因为一段语流中大呼吸重置一般有 1~3 个，从数量上来看比小呼吸重置少很多。呼吸信号一般都是从 0 开始的，接着就达到整段语流呼吸信号的最高值。而我们的呼吸信号在进行归一化处理的时候，是把整段语流的呼吸信号最大值定为 1，因此每段语流基本都会出现一个幅度为 1 的重置，这样一来 1.00 的出现频率就会比邻近的数值多一些，在统计某种文体的重置幅度频度时，就会在 1.00 处出现一个高峰。

2）重置幅度在 1.00 两侧都会有少数分布，这可能反映了一级呼吸重置的不同类型。通过观察呼吸信号，发现如果语流中有一句话很长，在说完这句话的时候发音人就会出现呼气过度的情况，这时呼吸信号就是负值，在这句话之后如果有一个大呼吸重置，其重置幅度就有可能超过 1.00。此外，一般位于语流中间位置的大呼吸的重置幅度不会大于位于语流最初的重置幅度，因此此时的重置幅度就小于 1.00。上述 2 种情况不如 1）多见，所以频度上就会小一些。

3）从二维图谱上的呼吸信号来看，可以看到比较明显的 1、2、3 级呼吸重置，但是从重置幅度的频度分布图上，2 级和 3 级重置之间的界限则不太清楚。

上述分析表明，不同文体朗读时的呼吸节奏从宏观上来看是具有一致性的。

4 结论和余论

通过分析，本文对发音人朗读不同文体的语料时的呼吸节奏变化有了初步的了解，并得到了以下结论：

1）从宏观上看不同文体的呼吸重置幅度有一定的一致性：呼吸信号的重置时长和重置幅度之间显著正相关。1 级呼吸重置的分布一致性较高，2 级呼吸重置和 3 级呼吸重置之间的界线不是很清楚。

2）不同文体的呼吸节奏有较明显的差异。宏观上，近体诗、词等韵文的呼吸节奏比较简单，其呼吸重置的大小大致可以分成两级。小说、散文、新闻等非韵文的呼吸节奏较复杂，其呼吸重置可以大致分成三级。具体到韵文、非韵文内部，不同文体朗读时的呼吸重置时长和幅度也有细微的不同。

3）从数量上看，小说、散文、新闻的大呼吸重置（1 级呼吸重置）最少，小呼吸重置（3 级呼吸重置）最多。

4）呼吸信号的重置幅度比重置时长更能反映不同文体的差别，也更能反映不同呼吸级别的大小。

目前，对语流中的呼吸节奏研究还刚刚起步，有很多问题等着我们去探索。例如，我们知道人们正常情况下只有呼气的时候才能说话，吸气的时候不能说话，因此呼吸信号在二维图谱中的上升段对应的就是语音信号的静音段。但是在对呼吸信号的波峰和波谷进行自动标注的时候我们发现了这样的情况：

图 18 是五言律诗《望岳》的语音、嗓音、呼吸信号。在自动标注呼吸信号的时候，程序按照局部最大值的办法，把波谷的标记打到了 A 的位置，但对照语音和嗓音信号可以看到一直要到 B 的位置，发音人才把“阴阳割昏晓”这句诗朗诵完。A、B 之间的呼吸信号是上升的，但这期间发音人一直在发音，这种和我们已有的知识不符的情况也许是因为这次测量的是呼吸时胸腔的变化，而此时发音人使用了腹式呼吸，腹部的力量将横膈膜往上顶，导致了胸腔扩大，反映在图谱上就成了呼吸信号上升的假相。这种发音时胸腹部呼吸的交互作用就是一个很值得研究的问题。

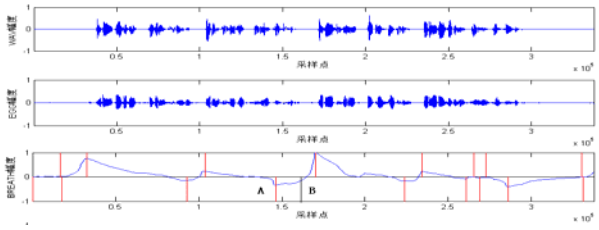


图 18 胸呼吸和腹呼吸交互作用示例

另外，呼吸信号和语音信号、嗓音信号等声学参数之间有着什么样的关系？根据呼吸信号划分出来的韵律单元和人们感知到的韵律单元是否一一对应？呼吸节奏和语法结构之间有什么样的关系？呼吸节奏是否能反映不同发音人的个人特征？能否根据呼吸信号建立出不同风格的韵律模型用于语音合成？只有对呼吸节奏作更深入的研究，才能揭示这些问题的答案。

参考文献

[1] 王洪君.普通话中节律边界与节律模式、语法、语用的关联，《语言学论丛》， 2002，26：279-300

[2] 熊子瑜韵律单元边界特征的声学语音学研究，《语言文字应用》，2003，2：116-121

[3] 王蓓、杨玉芳、吕士楠汉语韵律层级边界结构的声学相关物，《声学学报》， 2004，29（1）：29-36

[4] Liberman P. Intonation, Perception, and Language. Cambridge. Massachusettes. The MIT Press, 1967

[5] Tseng C, Pin S, Lee Y, et al. Fluent speech prosody: Framework and modeling. Speech Communication. 2005, 46: 284-309

[6] 谭晶晶.新闻朗读的呼吸节奏初探. 孔江平.第七届中国语音学学术会议论文集，电子版

[7] 陈亚勇.Matlab 信号处理详解，人民邮电出版社 2001

[8] 卢纹岱.SPSS for Windows 统计分析（第 3 版）.北京.电子工业出版社.2006

民歌男高音嗓音研究初探

A Study on the Voice Property of Chinese Folk Tenors

钱一凡

Qian Yifan

摘要： 本文以民歌男高音为研究对象，采用提取 EGG 信号中的嗓音基频和开商参数，分析音调与嗓音参数之间的关系。得出的结论主要有：1）民歌男高音的开商值在 g~g2 的调域，即 98Hz~396Hz 的基频范围之内，其变化可以明显分为低、中、高三个音区；2）开商的变化值在某一音区的范围之内呈稳定的下降趋势，即与基频的变化成反比；3）开商的变化到达音区间的临界点时会有突然的“跃高”现象；4）不同发音人之间低中音区和中高音区的临界点非常统一，分别集中在某一音调范围之内；5）不同发音人之间音区临界点开商的变化值有明显的差异，并可表现为开商变化曲线平滑度的差异；6）不同元音在频率域开商的变化形式不同，嗓音振动模式与控制控制声道形状的组织有关。

Abstract: Chinese classical tenors have special voice properties. Microphones and EGG signal collectors were used in our experiment to collect the speech and EGG signals of the a, i and u, the three vowels which were produced by 2 Chinese folk tenors ranging from 98-398 Hz. We extracted and analyzed pitch and open quotient from EGG signals at different ranges. The result shows: 1) Open quotient and speed quotient show regular changes while pitch changes from low to high and could be divided into 3 ranges; 2) The open quotient descends smoothly during one range and jumps high at the beginning of next range; 3) Open quotient varies depending on different vowels and different singers.

关键词： 嗓音；民歌；男高音

Key words: Voice property, Chinese folk singing, tenor

0 引言

嗓音发声类型的研究是语音学研究的重要课题之一，随着语音学逐步由“口耳相传”的传统语音学向追求实证的实验语音学甚至声学语音学的发展，对音乐艺术特别是声乐艺术的研究也纳入到语音学的范围之内来。

实验语音学对音乐艺术的关注其实并非新事物，早在 1925 年，刘复先生就在北京大学建立了“语音乐律实验室”，把“语音”和“乐律”放在同等的研究地位。音乐和语言本来就有密切的联系，比如音乐中的音调和语言中的声调和语调，都与声学中的基频（F0）这一概念有关。作为以语言为载体的声乐艺术，既包含语言的内在本质，又有其显著的特点。对声乐艺术的研究，既要着眼于它与语言的共性，也要关注语言之外的个性

本文的研究主要从民歌男高音这一角度切入，着重分析嗓音参数中的开商值与音调高低的关系。以往对于这方面的研究很少，主要有孔江平（1995），孔江平（1997d），Chen Jiangyou等(1998)，魏春生等（1999）和孔江平（2001）。比较具有代表性的是孔江平（2001）对嗓音和音调关系的研究。该项研究主要采用EGG信号，观察音调、速度商和开商的关系。提出的跟本文相关的结论主要有：1）男声平均开商值为55.13%；2）开商和音调高低成反比。

由于研究对象的不同，本文的研究与先前的研究相比，有如下的特点：1）嗓音样本的类型不同，如孔江平（2001）研究的样本是正常说话的嗓音样本；本文研究的样本是经过多年专业训练的民歌男高音歌唱演员的嗓音样本；2）音调对应的基频范围不同，如孔江平（2001）的研究，由于是针对正常语音的研究，音调的高低也在正常语音的范围之内，男声的基频范围稳定在129.46Hz~196.13Hz之间，跨度为66.67Hz；魏春生等(1999)的研究中男声基频范围在156Hz～278Hz之间，跨度为122Hz；而本文的研究由于是针对民歌男高音的艺术嗓音的研究，音调高低的范围更大，基频范围在98.49Hz~387.76Hz之间，跨度为289.27Hz。

1 研究方法

1.1 硬件设施

信号样本在北大中文系录音室采集，本底噪声小于等于 18dB。使用硬件设施主要有：麦克风，声门阻抗信号（EGG）采集器，调音台，Powerlab 肌电脑电仪，录音用笔记本电脑一台，定音用笔记本电脑一台。

1.2 录音及分析软件

录音用 Powerlab 肌电脑电仪配套软件 ADInstruments 公司 Powerlab Chart5 for Windows，定音高用 Cakewalk Pro Audio 9.03，后期参数提取使用基于 MATLAB 的北大语音乐律信号处理平台，以及美国 KAY 公司的“多功能语音分析系统（KAY-Multi-Speech Model 3700）”，主要使用其子系统“实时声门阻抗分析软件（Real-time EGG Analysis Program）”

1.3 发音人

发音人一：郭长雷，男，28 岁，原北京大学合唱团成员，学习民歌 7 年。

发音人二：刘洛克，男，25 岁，北京大学合唱团成员，学习民歌 11 年；

1.4 实验方法

前期使用录音软件分二通道采集声音信号，其中，第一通道采集经由麦克风得到的语音信号，第二通道采集经由喉头仪得到的嗓音信号，如图 1。信号采集参数为 16bit，20k。

实验前将麦克风固定在距发音人口部约 15cm 处；喉头仪固定在喉头。实验时发音人根据定音高软件给出的标准音高，从小字组 g 到小字二组 g2，共 25 个半音，使用民族唱法，每个音按元音[a][i][u]各唱两遍，每遍约持续 2s，即每位发音人每个元音共记录时长约 2s 的 50 个双通道信号。

实验后将记录的每个信号选取稳定段 30～40 个周期，保存为 PCM 码的 WAV 格式，使用实时声门阻抗分析软件，提取出基频（F0）、开商（Open Quotient，OQ）和速度商（Speed Quotient，SQ）等嗓音参数。

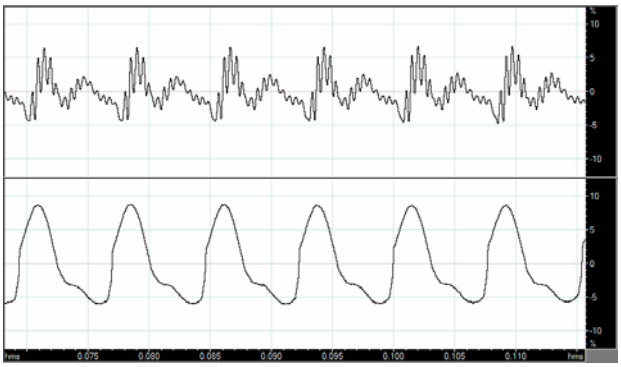


图 1 双通道录音信号举例

1.5 嗓音参数的定义

以往的研究当中对于嗓音参数中特别是开商(OQ)和速度商(SQ)的定义大多不太明确。孔江平（2001）对开商和速度商有明确的定义：

开商等于开相和周期之比（开商=开相/周期）。

速度商等于开启相和关闭相之比（速度商=开启相/关闭相）。

本文对于开商和速度商的定义基本沿用该文的定义，如图 2 所示：

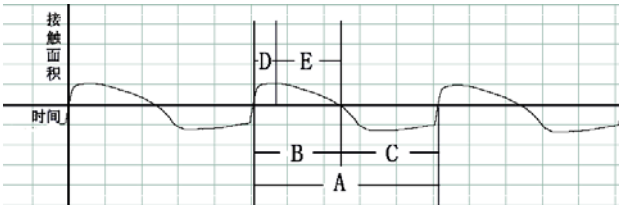


图 2 正常嗓音波形图

这是一张正常嗓音的波形图，X 轴是时间，Y 轴是声带接触面积。图中各字母表示时间段的意义如下：

A 是一个周期（Cycle）；

B 是声门面积>0 的段落，即闭相（Closed phase）；

C 是声门面积<0 的段落，即开相（Open phase）；

D 是闭相当中，声带接触面积逐渐增大，声门正在关闭的段落，即关闭相（Closing phase）；

E 是闭相当中，声带接触面积逐渐减小，声门正在打开的段落，即开启相（Opening phase）。

开商（Open Quotient）=开相/周期×100%；

速度商（Speed Quotient）=开启相/关闭相×100%。

由于篇幅的限制，本文对嗓音特性的分析主

要针对开商这一参数。

2 基音特性分析

关于基频的检测内容主要是发音人对于标准音高的把握程度，我们使用标准音高曲线分别与两位发音人实际唱出的音高曲线作对比，结果如下：

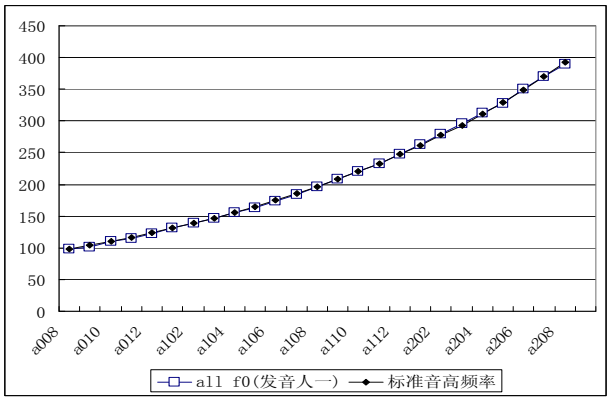


图 3 发音人一与标准音高频率曲线对比

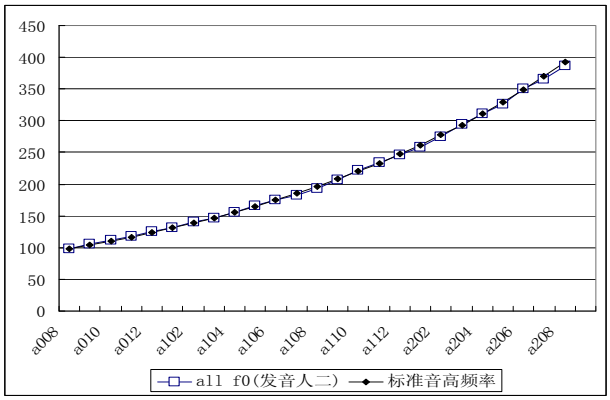


图 4 发音人二与标准音高频率曲线对比

从图 3 和图 4 中我们可以发现，无论是发音人一还是发音人二，实际所发的音高基频曲线与标准音高频率曲线几乎重合，音高掌握比较准确。

我们再使用发音人与标准音高频率的偏差值相对与标准音高频率的百分比所得的曲线，就更加一目了然：

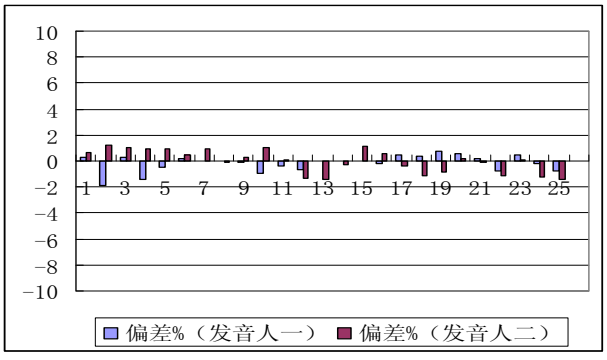


图 5 发音人与标准音高偏差比率（%）

图 5 的 X 轴是基频值不断增大的各个样本，Y 轴是偏差的比率（%）(=（实际音高对应基频-标准音高对应基频）/标准音高对应基频)。从这张图中我们可以更直观地观察到，无论是发音人一还是发音人二，在全音域中音高对应基频的偏差在-1.85%~1.38%之间。

3 嗓音特性分析

对于开商的分析主要分成以下三个部分：首先，综观两位发音人在发各个元音时的开商表现，推断开商在频率域中变化的基本特征；其次，分别比较两位发音人的开商表现，分析两位发音人不同的开商表现类型；最后，分别以两位发音人为例，对比发三个不同元音时开商的表现。

这里要说明一点的是，以往语音分析中对很多参数的分析都是采用平均数据的方法，以综合考量各个参数的总体特征，本文对于开商这个嗓音参数的分析之所以不使用平均数据，而分别对两位发音人的数据进行考量，主要是因为开商是表现个人嗓音特征的重要参数之一，每个人的开商表现都不尽相同，简单的平均可能会抹杀了个人嗓音的特点和开商在频率域中变化的特点。本文正是考虑了嗓音这个参数的特点，而采用了单独对发音人进行考量，并综合其共性的方法。

3.1 开商表现综观

图 6、图 7、图 8 分别是[a]、[i]、[u]三个元音在频率域上的变化曲线，两位发音人同一个元音的曲线画在一张图上。图中 X 轴是音高或基频递增的各个样本的序列，Y 轴是开商的比率（%）。

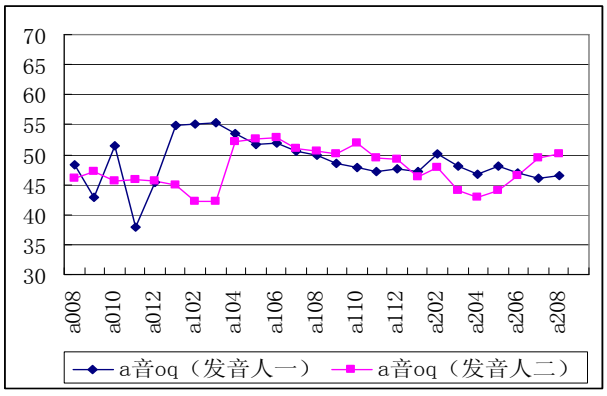


图 6 元音[a]开商变化曲线

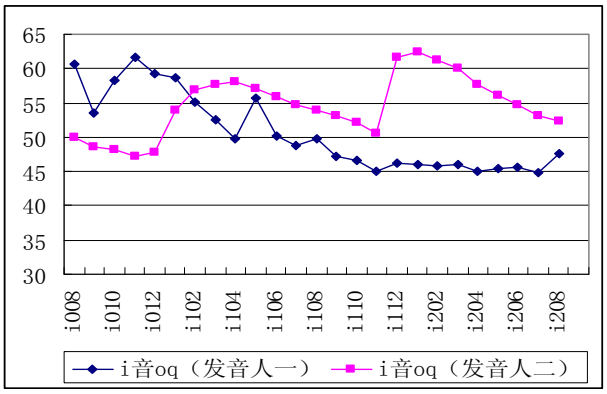


图 7 元音[i]开商变化曲线

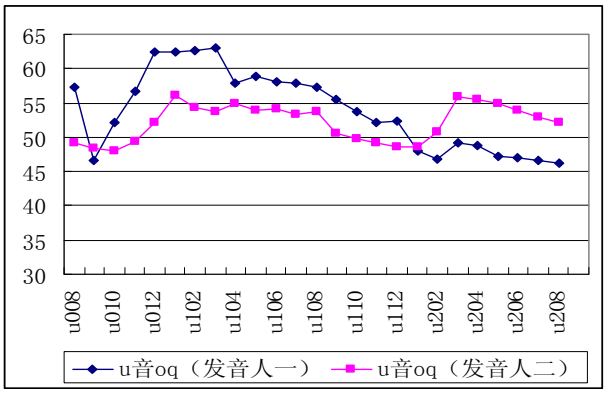


图 8 元音[u]开商变化曲线

从图 6、图 7、图 8 中我们可以看出：

无论是元音[a]、[i]还是[u]，其嗓音的开商范围都在 40%~65%之间。根据计算，发音人一开商平均值为 51.11%，发音人二开商平均值为 51.45%，总平均为 51.27%。

在本文所考察的音高基频范围，即男声小字组 g~小字二组 g2，基频 98Hz~392hz 的范围之内，开商的曲线变化可以分为低音区、中音区和高音区三段，三段有明显的临界点，且不同发音人的临界点不尽相同。由低到高的走向表现是：在某一音区的范围之内，开商值呈下降趋势；到达音高段的临界点时，开商值突然跃高，然后继续进入下一音高段的下降曲线。

发元音[a]时，发音人一低音区和中音区的临界点在样本a101，即c1（131.37Hz[1]）附近；中音区和高音区的临界点在样本a202，即#c2（282.52Hz）附近。发音人二的两个临界点都稍高，低音区和中音区的临界点在样本a104，即#d1（155.49Hz）附近；中音区和高音区的临界点在样本a208，即g2（382.86Hz）之后，实际上高音区比我们考察的频率范围更高。

发元音[i]时，发音人一低音区和中音区的临界点在样本 i011，即#a（115.51Hz）附近；中音区和高音区的临界点不明显，但是可以推断在样本 i201，即 c2（237.02Hz）附近。发音人二低音区和中音区的临界点在样本 i101，即 c1（131.19Hz）附近；中音区和高音区的临界点在样本 i201，即 c2（260.04Hz）附近。

发元音[u]时，发音人一低音区和中音区的临界点在样本 u104，即#d1（151.73Hz）附近；中音区和高音区的临界点在 u203，即 d2（294.02Hz）附近。发音人二低音区和中音区的临界点在样本 u101，即 c1（131.58Hz）附近；中音区和高音区的临界点在样本 u203，即 d2（295.75Hz）附近。

我们把所有样本当中表现出来的临界点归纳如下:

样本	低-中	中-高
a 发音人 1	c1	#c2
a 发音人 2	#d1	g2
i 发音人 1	#a	c2
i 发音人 2	c1	c2
u 发音人 1	#d1	d2
u 发音人 2	c1	d2

表 1 各临界点音高

从表 1 可以看出，虽然两位发音人低音区和中音区的临界点及中音区和高音区的临界点不尽相同，但除了发音人二的元音[a]中音区和高音区的临界点比较高，在小字二组 g2 之外，其余的临界点音高值还是可以统一为两组。低音区和中音区的临界点在小字组#a~小字一组#d1 之间，跨度为 5 个半音；中音区和高音区的临界点在小字二组 c2~d2 之间，跨度仅为两个半音。这两组之间界限分明，互不交叉。

3.2 对两位发音人开商变化模式的考量

从图 6、图 7、图 8 中我们已经可以大致看出两位发音人开商变化模式的不同。下面的这张图可以让我们更直观地进行比较。

[1] 实际发音测得的基频值

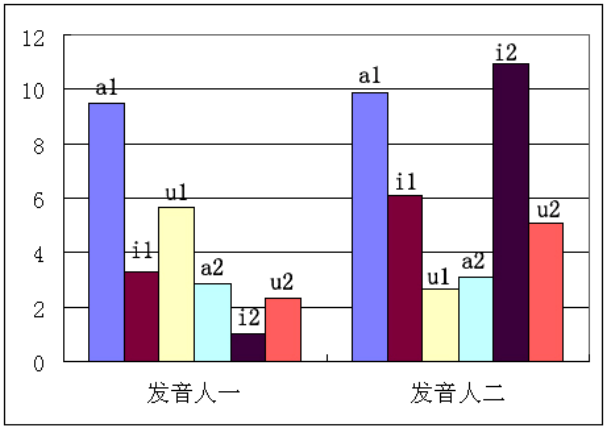


图 9 两位发音人开商临界点变化值比较

图 9 是对两位发音人开商临界点变化值的比较图，Y 轴代表临界点开商值与前一样本开商值的差值，X 轴的左边是发音人一各元音低中音区临界点（a1、i1、u1）的表现和中高音区（a2、i2、u2）的表现；X 轴的右边是发音人二各点的表现。

由图 9 可以看出，发音人一临界点左右特别是中高音区临界点的开商差值要远小于发音人二。这也是造成图 6、图 7、图 8 中发音人一曲线的变化幅度要高于发音人二曲线的原因。

3.3 对三个元音开商变化模式的考量

图 10 和图 11 分别是两位发音人 3 个元音的开商变化图：

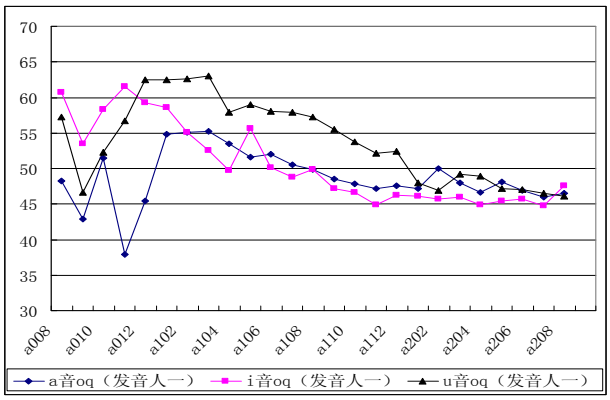


图 10 发音人一各元音开商对比

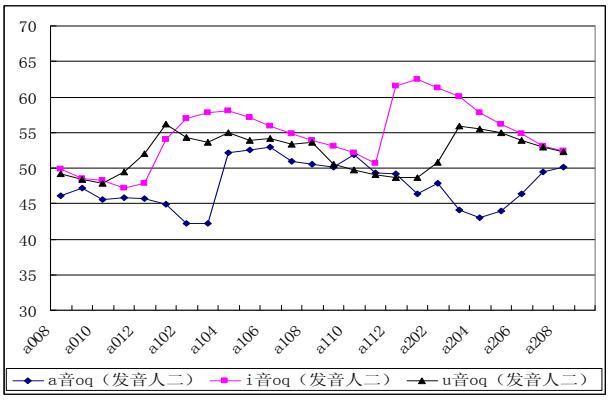


图 11 发音人二各元音开商对比

对比图 10 和图 11 我们可以发现，三个元音的开商关系在不同的发音人当中并不一致。发音人一元音[u]的开商值总体较高，元音[a]次之，元音[i]总体最低；发音人二元音[i]总体最高，元音[u]次之，元音[a]最低。开商值在不同发音人之间与元音的关系不同，是由哪些深层因素决定，还需要后续更深入的研究。

从这两张图中我们还可以看到的是，各元音的临界点在频率域上的位置有差别，如图 11 中，元音[a]的两个临界点就比[i]的要明显靠右，整体往高音域移动。元音的不同是由声道形状的不同决定的，而构成各种声道形状的肌肉神经组织的变化，也可能对嗓音振动模式产生影响。这一现象也是值得注意和挖掘的。

4 结论

综上，本文的研究可以得出一下几条结论：1）民歌男高音的开商值在 g~g2 的调域，即 98Hz~396Hz 的基频范围之内，其变化可以明显分为低、中、高三个音区；2）开商的变化值在低、中、高任一音区的范围之内呈稳定的下降趋势，即与基频的变化成反比；3）开商的变化到达音区间的临界点时会有突然的“跃高”现象；4）不同发音人之间低中音区和中高音区的临界点非常统一，分别集中在某一音调范围之内；5）不同发音人之间音区临界点开商的变化值有明显的差异，并可表现为开商变化曲线平滑度的差异；6）不同元音在频率域开商的变化表现不同，说明嗓音振动模式与控制声道形状的组织有关。

5 讨论

5.1 关于本文研究结果和之前研究结果的差异

孔江平（2001）得出的关于开商的结论是：开商与音调的高低成反比。本文的结论看似与先前的结论不尽相同，但其实并没有矛盾之处，表面矛盾的结果是由不同的研究对象和实验方法造成的。如，孔江平（2001）取样的基频范围在129.46Hz~196.13Hz 之间，恰好处于本文研究中的中音区对应的基频范围之内，而本文的研究结果中，在中音区的范围之内，开商与音调的高低的确成反比。因此，两者的结论并不矛盾，本文的研究也可以作为该研究在音域方面的一个扩展和补充。

5.2 关于临界点的问题

本文的研究得出，开商的变化曲线可以由稳定的临界点分成低、中、高三个音区，这就不禁让人联想起传统民歌教学中关于低、中、高三个音区的界定。其实，传统民歌教学当中关于三个音区的界定已经得到很多实验研究的验证，如笔者的《民歌男高音共鸣的实验研究》（未发表）当中，对民歌男高音头共鸣和胸共鸣的分析结果，就可以验证传统民歌教学中音区的划分。本文研究同样得到三个音区的结论，且低中音区的临界点和中高音区的临界点分别在#a~#d1 之间和 c2~d2 之间，与传统的界定基本一致，这样的结果进一步验证了民歌教学对音区划分的科学性。

5.3 关于两位发音人开商曲线的差异

本文的研究结果显示，不同发音人的开商变化曲线有一致的共性，也有显著的区别。开商在音调变化到临界点时跳跃性的变化，反应了嗓音振动模式的一种突变。不同发音人的这种跳跃性变化的明显程度，反应了对嗓音调节控制能力的大小和嗓音稳定的程度，嗓音变化曲线越平滑，说明从低音区到高音区嗓音的振动模式越稳定。传统民歌教学当中同样有低中高三个音区音色统一的要求：各音区音色统一，听感没有明显区别，是民歌演唱技巧中高水平的体现。考虑到嗓音声源对音色的贡献，嗓音振动模式的变化和对嗓音的控制能力与音区音色统一的问题的相关性，也是一个值得研究的课题。

参考文献

[1] 孔江平，2001，论语言发声[M]，中央民族大学出版社

[2] 孔江平，1995，汉语普通话嗓音特征相关分析[J]，中国声学学会 1995 年青年学术会议论文集，西北工业大学出版社

[3] 孔江平，1997d，汉藏语发声类型研究[A]，第 30 届国际汉藏语会议论文提要，中国北京

[4] 魏春生、王薇、陈小玲 等，1999，声带振动功能的定量检测[J]，临床耳鼻咽喉科杂志

[5] Chen Jiangyou and Kong Jiangping，1998，Acoustic study on relationships among pitch，open quotient and speed quotient extracted from EGG of Mandarin speakers[A]，Proceedings of Conference on Phonetics of the Languages in China，May 28-30，Hong Kong

藏语文-音自动规则转换及其实现

Rules for the auto-transformation of Tibetan text to IPA

李永宏

Li Yonghong

摘要：为满足语言学、音韵学和工程语音学的需要，该文根据现代藏文与 3 大方言语音之间的对应规律和藏文正字法，提出了从文字上对藏文声母和韵母拆分的“字丁分解法”，实现了藏文到各方言国际音标的自动转换。并对算法和实现过程进行了详细的阐述，建立了藏语 13 个方言点的方音数据库。方音数据库的建立为藏语方言研究和语言教学提供了科学、方便的工具，为藏语标准音的制定、推广及应用提供原始的语音材料，也能作为藏语语音识别和语音合成的标音基础。

Abstract：Modern Tibetan orthography and the sound rules relating between written text and Tibetan dialects were analyzed to develop rules for the transformation of Tibetan text to IPA for linguistics, phonology and engineering phonetics analyses. The transformation separates the Tibetan single syllable words into initial and final parts and then uses an automatic IPA transformation from the Tibetan text to the speech sounds of the dialects. A corpus of thirteen Tibetan dialects has been automatically developed. The transformation system provides a convenient tool for Tibetan dialect study an language teaching, for establishing original speech materials, for improving Standard Tibetan tongue an for engineering phonetics study.

关键词：藏文信息处理 藏语方言 国际音标 藏语文-音转换

Key words： Tibetan information processing; Tibetan dialect; International Phonetic Alphabet; Tibetan text to IPA transformation

0 引言

藏语属于汉藏语系藏缅语族藏语支，主要分布在中国西藏自治区和四川、云南、青海、甘肃等 5 省区，使用人数达 500 多万，有古老的拼音文字及浩瀚的文献典籍[1]。 在世界范围内，尼泊尔、不丹、巴基斯坦、印度等国家的境内也有一部分地区使用藏语。

在藏文初创时，以藏语口语为基础，在语音上，严格按照一字一音的原则，准确标记；在语法和词汇上，以口语为规范书写。随着语言的发展，藏文和口语失去严格对应，字母的标音功能减弱，不同地区的藏语朝着各自不同的方向发展变化，形成了各具特色的方言土语。

国内学术界将藏语主要分卫藏、安多和康 3 大方言。它们之间的差别主要表现在语音方面，如有无声调、有无清浊声母的对立、辅音韵尾的多寡等，其次是词汇和语法的差别。语音感知和科学研究的结果表明，卫藏方言与安多方言在“两端”，康方言介于中间。各方言的发展速度也不尽一致，卫藏方言最快，次为康区方言，安多方言发展最慢。安多话是藏语中保留古面貌较多的藏语方言，有很多特殊的语言现象，例如：音调不具有区别词义的功能，音节只有习惯调；有较多的复辅音等。虽然 3 大方言在口语上有较大差距，但与文字各有严格的对应规律，因此，各方言都可以拼读文字[2]。

由于藏语没有类似汉语普通话的口语标准音，所以各方言之间交际有一定的困难，这成为困扰一代又一代藏学专家的难题。虽然国内提出藏语标准语的方案很多，但都难以完全达到一致，而藏语各方言内部的语音特点和方言之间的语音差异的研究为科学、合理的制定藏语标准语提供了依据。此外，本文提出的文-音转换方案不仅能够为音韵学的研究提供方便的工具，也能作为藏语语音识别和语音合成的标音基础。

本文重点对藏语单音节词从文字到不同方言语音的转换进行了详细阐述，并在 windows 系统上进行了程序实现，建立了藏语方音数据库。藏语采用基于大字符集编码系统，国际音标采用基于 Unicode 的编码系统，藏语方言语音主要采用北大孔江平教授实地调查的藏语方言数据库（泽库和红原除外），方言点包括：拉萨、日喀则、德格、巴塘、泽库、夏河、同仁、循化、

化隆、红原、天峻、道孚、阿柔，共 13 个点。藏语词表来源于《藏汉大字典》、《安多口语字典》、《拉萨口语字典》、《格西曲扎藏文辞典》、《新编藏文字典》、《藏文同音字典》、《藏语文课本（小学 12 册、初中 6 册、高中 6 册）》6 部藏文字典和 24 册藏语文课本，共 13 万余条词汇。

根据藏语内在特点和程序处理中的实际需要，本文在已有的术语基础上[3]，约定了一些新的术语，例如“基字丁”，“带音基字丁”，“不带音基字丁”等。

1 藏语音节的基本概念

藏文是在梵文天成体的基础上发展而成的一种拼音文字，共有 30 个辅音字母，4 个元音符号，其中/a/为零位。藏文有一套严格而完整的字母组合排列规则，它的字符流是两维呈现，自左向右横向书写，传统藏文文法根据字母在音节中的结构位置，将字母分为“基字”、“上加字”、“下加字”、“前加字”、“后加字”和“再后加字”，基字为整个藏字的核心，30 个辅音字母都可以做基字，元音不能做基字，其中鈲[i]，鈴[e]，鈬[o]，加在基字丁上面，鉤[u] 加在基字丁下面，不带元音符号的默认为是元音[a]，30 个辅音字母中有 5 个可做前加字（“ག་ད་བ་མ་འ”），3 个上加字（“ར་, ་ར་, ་ལ”），4 个下加字（“ཡ་, ་ར་, ་ལ་, ་ཝ”），10 个后加字（“ག་ང་ད་ན་བ་མ་འ་ར་ལ་ས”），再后加字是后加字中的“ད”（今不用）和“ས”，当再后加字出现时，后加字和再后加字的组合在现代藏文中只有四种形式（“གས་, ངས་, བས་, མས”），否则不符合藏文的正字法[4]。

藏文大字符集编码方案在计算机中是以上下叠加的字母为一个整体进行编码的，称之为字丁，例如“སྐྱ”就是一个带有上加字（ས）和下加字(ར)和元音(ྱ)的字丁。藏文音节包含的字丁数，称之为位长，一个藏语音节位长最小值为 1，最大值为 4，例如“བསྒྲུབས”位长为 4。包含有基字的字丁，称之为“基字丁”，带元音的基字丁称为“带音基字丁”，不带元音的基字丁称为“不带音基字丁”，实际藏文中所有的基字丁都是带音基字丁(没有元音符号表示带元音[a])，不带音基字丁是在字丁分解中，为了研究方便而引入的。

2 国际音标及其拉丁转写的实现

为了能够得到较为全面的藏语词汇，经过长达一年的词典录入和校正，共收录了 6 部藏文字典和 24 册藏语文课本，总词汇达 13 万余条，查重后得到 9 万余条藏文词汇，其中单音节（不包括梵音藏文和借词）有 0.5 万余条，占到总词汇量的 5.6%；双音节词汇有 4.3 万余条，占到总词汇的一半；三音节词汇大约有 2 万余条，四音节词汇大约有 1.6 万余条，四音节以上词汇大约有 0.5 万余条，如图 1 所示。

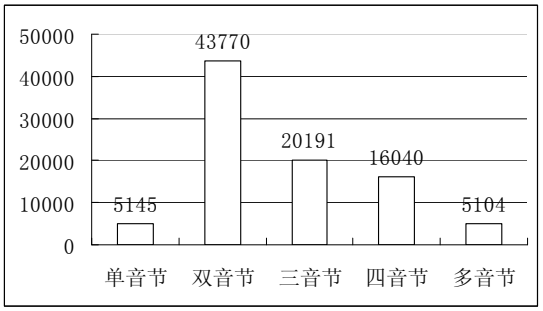


图 1 音节数量分布图

无论是词汇还是篇章，文字到国际音标转换的基础是藏文单音节的转换，转换过程如图 2 所示。核心思想为：对拟要转换的单音节藏文进行声韵母的分离，对分离的声母和韵母分别进行音标转换，然后合并生成国际音标，对有声调的方言还需要加上声调。程序中需要已经处理好的字丁分解表、声母音标转换表、韵母音标转换表的支持。

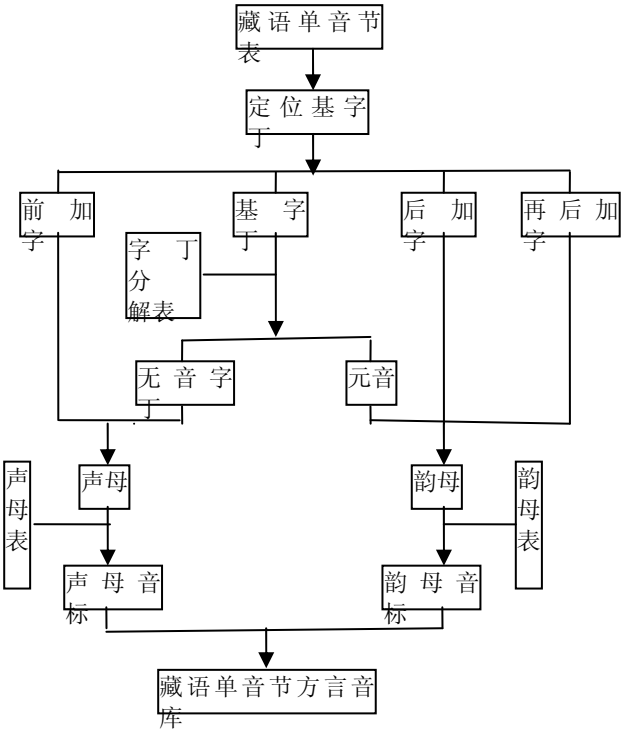


图 2　单音节音标转换框图

2.1　基字丁的定位

根据藏语的拼写方式，元音是叠加在基字丁上的，要完成藏语的声韵母分离，首先要正确的找到基字丁，这也是整个转换过程中最关键的一步。

虽然藏文有一套严格而有规律的字母组合规则，但是在一些特殊的组合形式中，并不能依靠规则直接判断，存在着歧异。例如：前加字+（'གས','ངས','དས','མས'），首字丁和中字丁都符合做基字的条件，但显然只有一种情况是正确的。因此本文按照音节位长的不同，考虑藏语单音节词的所有的字母组合类型，分别进行基字丁的判断。

类型	表现形式	基字丁	例词
单字丁	单个字丁	1	ཀ
双字丁	前加字+非后加字&非“འི་ཞུ་ཏ”	2	གཡུ
	前加字+后加字	1	གབ
	非前加字+后加字	1	དེད
	首字丁+“འི་ཞུ་ཏ”	1	ཀེའི
三字丁	前加字+非（'གས','ངས','དས','མས'）	2	བསྒྲུབ
	前加字+（'གས','ངས','དས','མས'）	2	འགས
		1	གངས
	非前加字+（'གས','ངས','དས','མས'）	1	ཡངས
	前加字+中字丁+“འི་ཞུ་ཏ”	2	གཤེའི
四字丁	前加字+基字丁+（'གས','ངས','དས','མས'）	2	གསུམ གས

表 1　藏语整字基字丁的确定

注：1 表示基字丁为第 1 个字丁；2 表示基字丁为第 2 个字丁。

2.2　基字丁分解

在藏文音节中找到基字丁后，利用藏文字丁拆分表，把基字丁包含的辅音成分和元音成分分开。基字丁分解表包括 550 多个现代藏文字丁，其中不带元音符号的字丁，本身带有[a]音，因此把字丁分解为不带音基字丁（无音字丁）和 5 个元音，ཨ[a]、ཨི[i] 、ཨུ[u] 、ཨེ[e] 、ཨོ[o]，如表 2 所示。

表 2　藏文字丁拆分表

字丁	无音字丁	元音	字丁	无音字丁	元音
ཀ	ཀ	ི	ཀེ	ཀ	ེ
ཀོ	ཀ	ོ	ཀཻ	ཀ	ཻ
ཀུ	ཀ	ུ	ཀཽ	ཀ	ཽ
ཀྲ	ཀ	ཿ	ཀཾ	ཀ	ཾ
ཀྱ	ཀ	ཾ	ཀྀ	ཀ	ྀ

2.3 音节拆分

把藏文单音节基字丁中包含的辅音成分和元音分开，基字丁前面如果有字丁就是前加字，后面有字丁就是后加字和再后加字，这便得到单音节拆分表，前加字、后加字、再后加字都可以为空，如表 3 所示。

藏文	前加字	基字丁			后加字	再后加字
		带音	不带音	元音		
བསྒྲུབ	བ	སྒྲ	ག	ུ	བ	
གཡོག	ག	ཡ	ོ	ག	ག	
བདེབས	བ	དེ	བ	ས	བ	ས

表 3　藏语单音节拆分表

2.4 藏文声韵母组合

藏语和汉语类似，也是单字单音的声韵母结构。对预转换的藏文查找藏语单音节拆分表，就很容易从文字上进行声韵母分离了，结果如表 4 所示。

藏文声母=前加字+不带音基字丁

藏文韵母=元音+后加字+再后加字

现代藏语从文字上来看，声母共有 213 个，韵母有 77 个[1]，理论上声韵母组合能产生 16401 个单音节藏文，而实际存在的藏字大约占理想组合的 1/3。

序号	藏文	汉意	藏文声母	藏文韵母
1	བསྒྲུབ	力量	བསྒྲ	ུབ
2	གཡོག་ས	靠	གཡ	ོག་ས
3	བདེ་བས	盖子	བད	ེ་བས

表 4　藏文声韵母分离表

2.5 声调的处理

由分开的藏文声韵母，分别查找声母音标对照表和韵母音标对照表，便能得到藏语声母和韵母各自对应的国际音标，然后直接组合就能得到整个音节对应的国际音标。但是，对有声调的藏语方言来讲，其音节的实际调值与声母和韵母都有直接的关系，需要在声母音标对照表和韵母音标对照表中进行标识。

以藏语拉萨话为例，声调有六个，用五度标调法记音，应记为 55，114，52，13，51 和 132。但实际上 51 调和 132 调有“ʔ”韵尾，因此藏语拉萨话的声调也可以处理成 4 个，以宽式记音可标作：55，14，53（52 和 51）和 12（12 和 132）。拉萨话的声调有长短之分，长韵母为长调 55 和 14，短韵母为短调 53 和 12；也有高低之分，清声母音节为高调 55 和 53，浊声母音节为低调 14 和 12[5] [6] [8]，如表 5 所示。

	声母		韵母	
	清	浊	短韵	长韵
形式	古藏文清声母	古藏文浊声母	vp、v、$\tilde{v}^{ʔ}$、$v^{ʔ}$、$vm^{ʔ}/vŋ^{ʔ}$	vr、vm/vŋ、v:、$\tilde{v}$:、vv:
调号	1	2	3	4
调型	高调	低调	短调	长调
调值	55,53	14,12	53,12	55,14

表 5　藏语拉萨话声调表（一）

注：调号为方便理解和程序处理而引入。

以上分析可知，藏语拉萨话单音节声韵母组合调号共有 4 种类型，其对应的实际调值，如表

6 所示。

韵 调	清声 短韵	清声 长韵	浊声 短韵	浊声 长韵
调号	13	14	23	24
调值	53˥	55˥	12˩	14˩

表 6　藏语拉萨话声调表（二）

注：藏语有些方言词汇调值不稳定用“00”来表示。

藏语其它方言的声调相对要复杂的多，黄布凡在《藏语方言声调的发生和分化条件》(1994)中提到，藏语有声调方言都经过有自然声调阶段，自然声调起源于声母和韵尾的附带特征。音位声调起源于声母和韵尾的演变导致自身辨义功能的减弱和转移。各方言声调分化并非都是清高浊低，而是条件各异，自成系统[7]。这就为声调的自动标注带来很大的困难，还需要对每一个方言点的声调做系统的分析，才能够更科学、合理的表怔藏语在不同方言中的实际的声调。

2.6　音标转换

根据得到的藏文声母和韵母，查找声韵母音标对照表（参考文献[1]），得到各自对应的音标和调号。

声母音标对照表包含现代藏文的 213 个声母，字段有声母对应的拉丁转写，包括拉萨、日喀则、德格、巴塘、泽库、夏河、同仁、循化、化隆、红原、天峻、道孚、阿柔，共 13 个点的国际音标，字段包括结构藏文声母、拉丁转写、每个方言点的音标和调号，其中安多方言内的点没有声调字段。韵母音标对照表包含 77 个藏文韵母，字段同声母音标对照表。

藏文单音节音标直接由声韵母对应的音标叠加得到，音节的实际调值由对应的声韵母的调号进行组合，然后查找表 6 就得到对应的 5 度调值。

拉丁转写=声母拉丁转写+韵母拉丁转写

音节音标=声母音标+韵母音标+声调

经过以上步骤，最后得到藏语单音节方言音库，如表 7 所示。

如果需要转换的是多音节藏语词汇或语篇，会带来更为复杂的问题。首先根据音节点，把多音节切分为单音节，对每个单音节分别转换成各自的国际音标。此时，连续变调成为我们考虑的首要问题，拉萨话的连续变调规则研究的比较多，规律也相对清楚，康方言的连续变调比较复杂，很多调值变化并不稳定，需要进一步的做深入调查和分析；安多方言牧区话和农区话差别比较大，传统说法认为安多方言没有区别意义的声调系统，但安多方言习惯调的内部规律，还在进一步的研究中。

2.7　结论

我们可以认为古藏文就是古代藏文的实际音值，字音转换是根据藏语不同方言的现代拼读法与古代藏文的对应关系建立的现代藏语方音数据库。从转换的结果来看，首先，安多方言由于其音系相对复杂，声韵母数量较多，其音节数量在 1600 左右，同音词概率为 5145/1600=3.2，有些同音词的习惯调值并不完全相同，还需要做进一步的研究，卫藏和康方言音节数量大致都在 1200 左右，但因其都有声调，所以同音词的概率要远远小于安多藏语，而且声调在不同的方言内部，分布也有很大的差异。其次，个别词汇的实际音值和利用规则转换的国际音标之间并不是完全相同，也就是说某一方言内部的同一声（韵）母在不同的词汇中其发音并不相同，其原因主要有以下几点：

（1）由于其它语种或藏语其它方言的影响，造成部分词汇发音并不符合大的系统性规律，属于语言接触的范畴。

（2）语音的演变并没有系统性的完成，有些内部词汇还保留其演变前的发音。

（3）由于藏语方言发展不平衡，藏语方言内部的部分借词，其发音并不遵从本语言点的演变规律和拼读规则。

藏语方言内部的复杂性也证明了语言发展是有内部层级性的，这就给语言工作者提出了更高的要求，必须根据研究的目的对藏语进行多角度，多层次的综合分析。

3　结束语

本文通过对藏语 3 大方言 13 个点的语音数据采集和音系整理，6 本藏文字典和 24 册藏语文课本的统计分析，结合藏文在计算机内部的处理机制，提出了“字丁分解”的藏文-音转换方法，建立了 500 多个藏语字丁的分解表，213 个

藏文声母音标对照表，77 个藏文韵母音标对照表，完成了 5145 个藏语单音节文字到语音的自动转换，建立了藏语方言语音数据库。

本文的研究成果对藏语语言学研究、语言文字教学和藏文信息处理领域的许多方面，具有重要的学术价值和广泛的应用价值。

（1）利用计算机能够识别的符号表达文字和语音信息，便于计算机进行存储、传递和数据处理，有效地保护民族语言文化。

（2）为藏语方言研究、语言教学提供了科学、方便宜用的工具。

（3）在文字和语音的标注上，提供一致的平台和符号，便于研究人员的阅读和相互交流。

（4）为语音的韵律特征分析和包括识别、合成的工程领域的研究，提供字音转换功能。

（5）为藏语标准音的制定、推广及应用研究提供原始语音材料。

一方面为了能够更好的完善藏语方音数据库，还需要进行藏语方言的调查，陆续的加入更多的方言点；另一方面，对调查的方言语料进行声学分析，建成藏语方言语音声学参数数据库，研究藏语方言的发展和内部差异，更好的为语言学，语音学研究，甚至工程语音学服务。

参考文献

[1]格桑居冕，格桑央京．藏语方言概论[M]．北京：民族出版社，2002

[2]金鹏．藏语简志[M]．北京：人民出版社，1983

[3]于洪志．计算机专用藏文文字术语探讨[J].，《术语标准化与信息技术》，1997，3：26-27

[4]周季文．藏文拼音教材[M] 北京：民族出版社，1982

[5]胡坦．藏语(拉萨话)的声调研究[J]．民族语文，1980，3：22：36

[6]谭克让，孔江平．藏语拉萨话元音、韵母的长短及其与声调的关系[J]．民族语文，1991，2：12：21

[7]黄布凡．藏语方言声调的发生和分化条件[J]．民族语文，1994，3：1：9

[8]孔江平．藏语（拉萨话）声调感知研究[J]．民族语文，1995，3：56-64

韩语呼吸节奏与语调群的关系初步研究

Correlations between Respiration Rhythm and Intonation Group in Korean: A Preliminary Study

尹基德

Yoon Kiduk

摘要： 本文以观察韩语呼吸节奏的主要表现以及呼吸与语调群的关系为目的。实验运用肌电脑电仪和呼吸带传感器记录几篇韩语新闻和散文以及简单复句的朗读风格语料，然后用语音分析软件观察呼吸节奏和音高的关系。本文得到了下面的几个结论。1）大呼吸单元对应于语义群。呼吸节奏是人对整个话语内容的认知结构的产物。2）与汉语的研究结果不同，韩语新闻朗读的呼吸充置对应于语流中停顿的地方。3）实验语料中有两个常见的呼吸结构模型，命名为呼吸下倾（respiration declination）和呼吸上升（respiration uphilling）4）呼吸群和语调群的不一致。在语调群边界上音高调节的形式决定呼吸的重值。韩语的语调群边界常见的 5 个核调形式中， 低升（lowrise）和降升（fallrise）的两个调型引起呼吸的重值。这意味着特定音高模式会影响呼吸节奏。韩语朗读语料呼吸节奏的这些表现，不但表明于在前的汉语语料的研究所提出的呼吸现象的定义相比基本上的呼吸形式很相似，而且也表示呼吸节奏与语调模式有复杂的关系。

Abstract: The main goal of this preliminary study is 1) the observation of the physiological realization of the respiration rhythm of Korean news reading corpus, and 2) the verification of the traditional presupposition that has been identified the respiration group with the intonation group. The experiment used news reading corpus for question 1) and a simple sentence for question 2) recorded with 5 types of nuclear tones which is usually spoken in the intonation group boundary. The result shows that Korean respiration rhythm is considerably identified with the break of speech, also characterized by the respiration declination and the respiration uphilling, which seem to be universal for other languages. Contrary to the traditional supposition, intonation groups are not identified with respiration groups. Except for some with lowrise and fallrise nuclear tones, the other nuclear tone types don't recharge respiration at the intonation boundary. These results imply the complexness of respiration rhythm in human speech, especially related to intonation and pitch control.

关键词： 呼吸节奏，呼吸下倾，呼吸上升，语义群，语调群，句法结构，边界语调，音高调节

Key words: respiration rhythm, respiration declination, respiration uphilling, intonation group, intonation boundary, pitch control

0 引言

到目前为止，关于语调的实验语音学研究，无论是语调语言还是声调语言，主要是依靠音高曲线上的信息来进行的，已经获得不少研究成果。可是，单单依靠音高信息的研究方法有的时候显露不少限制和问题。照此，为了进一步的语调研究我们需要另一个信息来源能够协助音高信息。一个好例子是呼吸信号的研究（如覃晶晶、孔江平（2006））。

呼吸是语音产生的第一步骤。呼吸对语调肯定有一定的影响，这个想法虽然对很多语调研究者来说不足为奇，可是有关的讨论和实验研究却很少。

在英语语调研究中，Liberman(1967)曾经提及呼吸群的概念，建议呼吸群不但是英语语调结构的基本单位，而且是一个普遍语调现象‘音高下倾（Declination）’和‘Downstep’的主要生理原因。这个看法关于普遍的语调现象提供非常直观的解释，而受到很多语调研究者的欢迎，已经成为一个无疑的前提。韩语的语调研究也按照这种看法来把呼吸群当作语调群的同一概念。韩语是个非声调语言，明显地显示语调语言的特征，因此在很多方面比较容易将英语的语调研究方法应用于其语调研究。可是，虽然在前的语调研究大部分都以音高信息的分析为主，还没有呼吸节奏有关的实验语音学研究。关于韩语呼吸节奏的研究至好也不过是诗歌文学研究或句法研究上的一个抽象的概念。

这次实验的主要目的是运用电脑电仪和呼吸带传感器记录和分析韩语新闻和散文的朗读

语料，察探呼吸节奏与音高调节之间的关系。这是以后汉韩两个语言语调的对比研究的一个初探。

0.1. 韩语语调结构简介

韩语语调的语音学研究基本上支持一个看法，就说韩语是个语调语言。Lee(1990)应用英国印象主义派的英语语调研究（如 O'Connor, J.D. and Arnold, G.F.(1973), *Intonation of Colloquial English*, 2nd edn. London: Longmans.）记述了韩语语调结构。Lee(1990)指出韩语的语调基本上有 9 个核语调（nuclear tone）,低平(low level)、中平(mid level)、低降(low fall)、升降(rise fall)、低升(low rise)、高平(high level)、高降(high fall)、全升(full rise)、降升(fall rise)。他提出韩语句子以一个以上的语调群（intonation group）来组成，而语调群（intonation group）里有一个以上的节奏单位（rhythm unit）。韩语语调的特征就在这些韵律单位的边界上出现的特定核语调来决定。其主要表现如下。

a. 单句语调

一个语调群（intonation group）组成整个句子的语调。语调群里有一个以上的节奏单位（rhythm unit），其边界上可以用 4 个核调之一，最后节奏单位最后音节是整个语调群的边界，9 个核调中可以选择一个表达整个语调群的语调。

b. 复句语调

复句的语调按照句法结构以两个以上的语调群来组成，就说‘语调群 1 + 语调群 2 等’的结构。每个语调群里节奏单位组成的语调跟单句的语调结构相似。可是语调群的单位来看，除了最后语调群，那些前面的语调群不使用全 9 个核调而仅仅用 5 个核调，就是，低平(low level)、低降(low fall)、中平(mid level)、低升(low rise)和降升(fall rise)。其中，低平(low level)主要运用于朗读、演讲和宣布等正式的话语，低降(low fall)和中平(mid level)常常表达亲密或随意的语气。低升(low rise)和降升(fall rise)的主要用处是向别人温和、亲切地说明某种事情的语调。

从另一个角度来看，随着 80 年代以来美国韵律音系学的语调标记体系 ToBI 的发展，Kang(1992), Kwak(1992), Jun(1993), Chung(1995)等运用 ToBI 研究韩语语调的特征。虽然研究者之间有若干的意见差异，K-ToBI(韩语语调标记体系)定义的韩语韵律单位有重音短语(Accentual Phrase)与语调短语(Intonation Phrase)。重音短语(Accentual Phrase)是以一个以上的音系短语(Phonological word)组成的单位，语调短语(Intonation Phrase)是以一个以上的重音短语(Accentual Phrase)组成的单位。关键是每个语调短语(Intonation Phrase)的边界上出现 6 个边界声调（这并不是说词汇层级的声调，而是说特定的语调模式），则是 L%、H%、LH%、HL%和 LHL%、HLH%。其中 LHL%、HLH%是比较复杂而不常用的，所以不算基本边界声调。

我们可以留意这两个研究看法实际上其基本概念非常相似，就是说节奏单位（rhythm unit）相当于重音短语(Accentual Phrase)，语调群（intonation group）相当于语调短语(Intonation Phrase)。这是因为 K-ToBI 基本上是一个标记系统，其研究对象语调结构与其主要表现而言，以前印象主义研究已经提出相当直观的解释，所以基本上两个研究标记方法不同研究内容相同。那么、本文章以后的讨论主要按照 Lee(1990)的整理来进行。

1 实验说明

1.1 录音语料与录音人

两篇较短的新闻报道和一篇散文，还有几个简单的复句。录音人由本人来担当，把每个语料几次练习后录了两次录音。录几个简单的复句时，为了观察呼吸和音高的关系，尽量运用所可能的语调模式来录音。其中下面主要讨论的语料是一篇新闻报道和一篇散文中的部分，与一个并列复句。

本次实验是个初步研究，所使用的语料和录音人在数量和人数方面有限制。这是以后的研究要改进的部分。

1.2 录音工作与分析方法

录音工作在北京大学中文系的隔音室进行，使用的设备是 Powerlab 公司的肌电脑电仪与

MLT1132 呼吸带传感器。录音和分析软件为 Powerlab 公司的 Chart 5，然后为了音高分析运用 praat 软件。

2 实验结果与讨论

本次实验的结果可以分成两个部分，一个是用 Powerlab 公司 Chart 5 软件的分析结果讨论韩语新闻朗读语料在大概的呼吸节奏上有什么样的独特的特征，另外一个是通过比较简单的复句运用 praat 软件提出音高曲线和呼吸节奏的比较。

2.1 韩语朗读语料的呼吸节奏基本特征

本节的讨论主要与谭晶晶、孔江平（2006）的汉语新闻朗读语料的呼吸节奏研究结果比较来进行，发现韩语朗读语料的呼吸节奏基本特征基本上与在前的研究结果相同。本节主要以韩语的语料中发现到的不同的表现为主讨论。

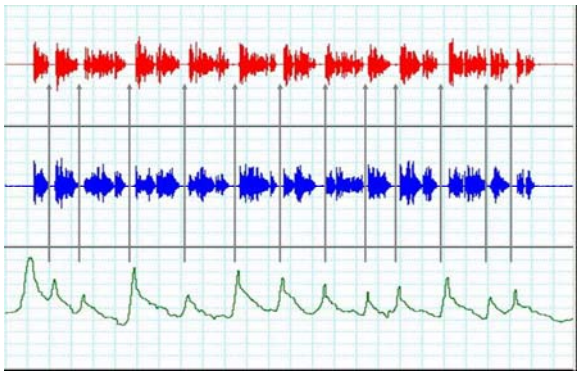


图 1 韩语新闻朗读语料的呼吸节奏

用 chart5 采取的 3 通道信号，第 1 通道是语音信号，第 2 是喉头义信号，第 3 是呼吸带信号。灰色的箭头记号是停顿的地方，也是各层级的呼吸单位，表示两个单位有互相对应的关系。

谭晶晶、孔江平（2006）的主要研究结果是如下。

- 三级不同大小单元，大呼吸对应于自然段，中呼吸对应于复句，小呼吸对应于分句或句子成分。
- 呼吸重置的地方必然有停顿，但停顿不一定有呼吸重置。两者之间的对应不是一对一的。
- 处于不同位置的呼吸单元，其曲线上达到最大值的方式不一样，说明呼吸节奏反映语篇的组织架构。
- 呼吸结构反映认知规划和生理机制，两个因素的冲突形成复杂的呼吸结构。

本次实验也能够确认这些在前的研究结果大部分可以运用于韩语的呼吸节奏研究。但是，我们要注意的是韩语语料的有些表现确实于汉语不同，表示韩语语调以及句法结构上的特点。（这表明以后需要进一步运用呼吸节奏的研究，特别能够协助汉韩语调的对比研究。）

首先，韩语朗读语料的呼吸节奏于汉语不同的地方是，与上面汉语的呼吸节奏第 3 表现不同，韩语的呼吸节奏基本上呼吸重置的地方与停顿的位置一致。图 1 的每箭头记号表示每个停顿和各个呼吸单元有一对一对应。这样的呼吸节奏表现的差异可能由两个语言韵律结构的差异产生的，也可能由句法结构上的差异产生的，也可能是两者与其他因素复合作用的结果。到目前为止，还没有适当的对比研究，当然无法决定正确的原因，以后进一步的对比研究才能够解释。可是，这次实验提出的一个结论，就是这些呼吸节奏表现的不同基本上影响到两个语言韵律结构上的差异，韩语新闻朗读的呼吸重置与停顿是一对一的关系，表示比汉语更整齐的样子。

2.2 呼吸下倾（respiration declination）

本次实验语料中出现的不同呼吸单位的组合方式当中有一个常见的现象，呼吸下倾（respiration declination）。

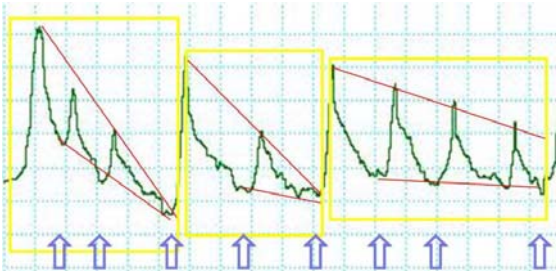


图 2 韩语新闻朗读语料的呼吸节奏主要特征

红线指示呼吸下倾(respiration declination)。黄色箱子是大呼吸群，也是语意群。蓝色的箭头记号是句法上句子边界。录音词表的翻译内容是“这里是首尔的一座邮局。正在进行分类邮件的工作。到目前为止以前的地址方式更容易分类。

/未经确认地址的邮件的确麻烦邮局职工们。这样的邮件还是直接去办公室确认地址最好。/（后略）”

我们首先从语义的角度讨论呼吸下倾（respiration declination）和语义群（semantic group）的关系。图 2 黄色的箱子非常明显地表示每个大呼吸群的边界。一个大呼吸群里几个不同层级的呼吸单位按照呼吸下倾来被排列着。每个大呼吸群里的上下红线，上线是呼吸最大值，下线是最小值，都表示呼吸下倾现象。要注意的是，照呼吸下倾现象来分类的大呼吸群（黄色的箱子）不但指出大的呼吸群，而且相当准确地符合每个语义单位（semantic group）。

呼吸节奏的表现还指出新闻朗读的句法特征。一个语义单位有时以几个简单的句子构成，有时单一个冗长的复句就是一个语义单位。当然，一个非常冗长的句子里可能存在两个以上的语义单位，可是一般新闻或散文里很少见这种非常冗长的句子。我们在实验里也可以看到韩语新闻里超过一个语义单位的极端地冗长的句子很少，主要以几个简单的句子或一个复句组成。图 2 蓝色的箭头记号是每个句子边界的标记，表示每个句子不超过语义群边界，反而几个尽短的单句构成一个语义群。尽短的单句，按照呼吸下倾曲线整齐地呼吸重置，这的确符合新闻本身的目的，就是信息的迅速而准确传达，让读者或听者容易接受所传达的信息。

那么，句法结构更复杂的呼吸节奏情况如何呢？别的语料，比如，散文朗读语料中复杂的句法结构和冗长的复句出现得比较多，会出现呼吸上升（respiration uphilling）现象。

2.3 呼吸上升（respiration uphilling）

本实验采用的呼吸信号处理程序是北京大学中文系语音乐律平台子程序 图 3 表示呼吸上升（respiration uphilling）现象。这是呼吸信号曲线的上升现象，特别是最小值曲线的上升。

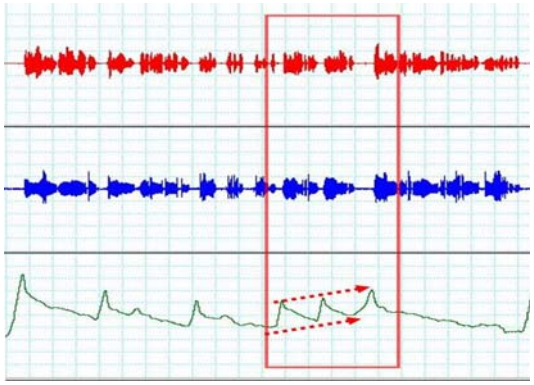


图 3 呼吸上升（respiration uphilling）现象

录音词表的翻译内容是‘并且语音学被运用于外国语的研究和教学、/方言研究、/言语障碍的诊断与治疗、/通讯工程、/发声法、播音及/**演剧的口才等，/其运用领域非常开阔。**’

这篇散文句子长度相当冗长，其句法结构是并列结构。这里出现的呼吸上升现象好像是一种呼吸调节上的错误。这样的并列结构句子虽然其句法结构很简单，但是句子的长度太长。如果没有充分的朗读练习的话，比较容易失去整个句子的节奏感，到后面就呼吸不顺出现呼吸上升。从句法结构的角度来看，该句子正确的呼吸节奏可以预测到如下。

‘并且语音学被运用于外国语的研究和教学、/方言研究、/言语障碍的诊断与治疗、/通讯工程、/发声法、/播音及**演剧的口才等，/其运用领域非常开阔。**’

事实上，经过几遍的练习后，同一个句子的呼吸信号上不再出现呼吸上升现象，就在上面预测到的地方出现呼吸重置，并且整个句子的呼吸节奏形成正常的呼吸下倾。所以，一般来讲，句法结构的复杂性和冗长的句子长度的环境提供呼吸上升（respiration uphilling）的主要原因。

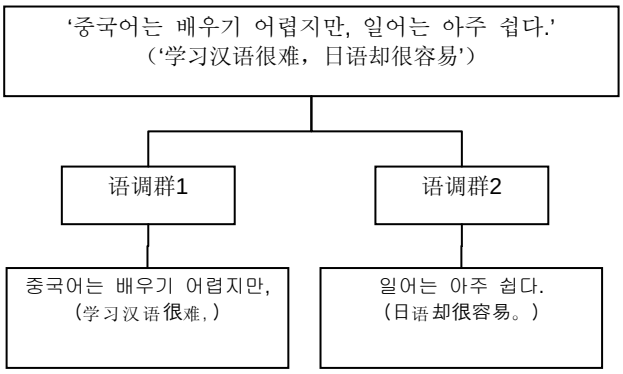
但是，这并不能说所有呼吸上升都是呼吸调节上的错误，因为除了上面提到的几个表现以外随着语料的种类和话语风格的变化，说话人可能运用更复杂的呼吸节奏模式。因此为了充分解释呼吸节奏有关的诸现象我们需要多方面的进一步研究。

2.4 语调群的音高调节对呼吸节奏的影响

Lee(1990)对呼吸群的解释如下。他提及语调群也是呼吸单位，也是语义和信息单位。换句

话说，人说一个句子说完之后一般重置一次呼吸，但说冗长的句子时自然地在句子里某个地方实行呼吸重置，这地方就成为语调群的边界。这当然有一定的规律，差不多和句法上的短语边界一致。

这是韩语语言学研究中第一次提到呼吸群的解释，而到目前为止没有适当的实验研究。至此，本节关于呼吸节奏与语调群的对应关系，通过一个并列复句来进一步讨论。为了观察语调群的音高调节，运用 praat 软件把语音通道的音高曲线提出来了。实验句子的结构如下。



这比较简单的并列复句以两个语调群组成，就是‘语调群 1+语调群 2’的结构。按照上面提到的韩语语调群的音高模式，前面的语调群 1 的边界使用 5 个核调。其中低平调（low level）是朗读和演讲等正式话语主要使用的调型，通过本次实验的新闻朗读语料也可以确认主要语调群边界上出现低降调音高模式。问题是其他音高模式与呼吸节奏的关系，就是说我们使用其他音高模式来读同样的复句时呼吸节奏上产生什么样的变化。

下面 5 个图表表示 5 个音高模式产生的呼吸节奏的不同。

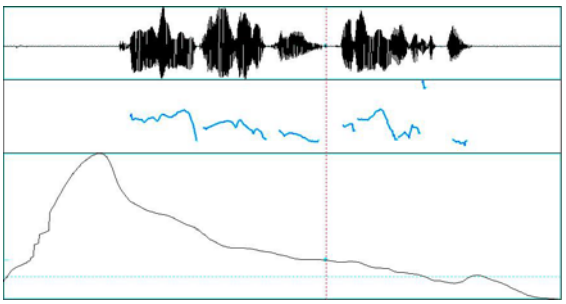


图 4 低平调（low level）

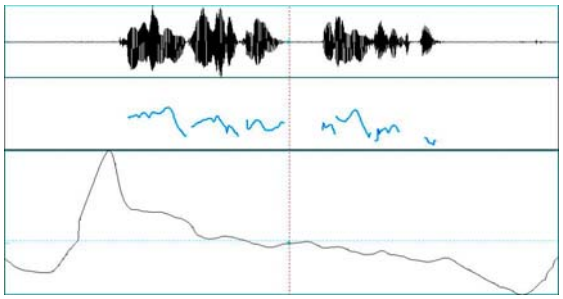
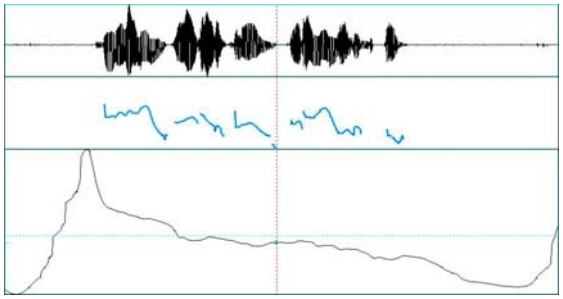
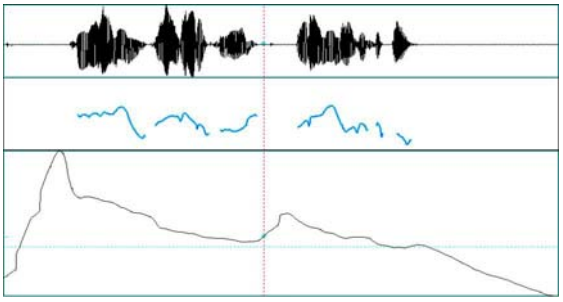


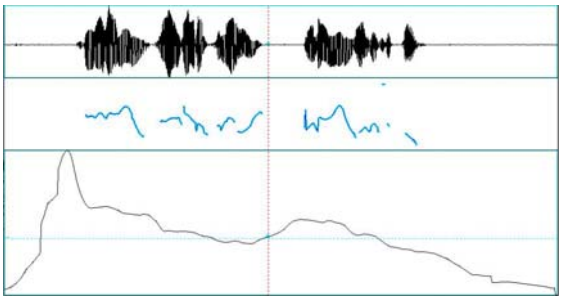
图 5 低降调（low fall）



图表 6）中平调（mid level）



图表 7）低升调（low rise）



图表 8）降升调（fallrise）

各图表上的红线表示语调群 1 的边界，中间的音高曲线助于确认语调群 1 所运用的音高模式。引起我们注意的是图表 7、8 的两个调型对呼吸节奏的影响。图表 4、5、6 表示语调群之间的边界上运用低平调（low level）、低降调（low fall）及中平调（mid level）的话，整个句子里除了开端的大呼吸以外没有任何呼吸充置，意味着两个语调群对应于一个呼吸群。但是运用低升

调（low rise）和降升调（fallrise）的句子明显地出现呼吸充置，每个语调群有对应的呼吸单位。这些表现说明韩语语调群与以前的看法不同不一定对应于呼吸单位，指出两者之间会有相当复杂的关系。通过这次实验得到的各个语调群音高模式与呼吸节奏的关系可以整理如下。

语调群音高模式	主要语调功能	呼吸充置
低平调（low level）	正式话语	无
低降调（low fall）	亲密	无
中平调（mid level）		无
低升调（low rise）	温和、亲切	有
降升调（fallrise）		有

虽然上面提到的 Lee(1990)在语调群 5 个音高模式的功能方面，除了低平调以外其他调型的分类有点模糊，但是我们要留意的是音高模式的确影响到呼吸的表现。不同的音高模式在不同的话语，那么这意味着不同的话语有不同的呼吸节奏表现。这也是以后的研究要进一步探讨的问题。

3 实验结果

通过本次实验得到的几个结论如下。

1）大呼吸单元与语义群(semantic group)之间有一对一的对应。人为了把信息自然地传达，先把握整个话语内容的认知结构然后运用适当的呼吸节奏。

2）与汉语的研究结果不同，韩语新闻朗读的呼吸单位相当正确地对应于每个停顿的地方。

3）实验语料中有两个常见的呼吸结构模型，命名为呼吸下倾（respiration declination）和呼吸上升（respiration uphilling）

4）与一般的看法不同，韩语语调群与呼吸群不一致。在语调群边界上音高调节的形式决定呼吸的重值。韩语的语调群边界常见的 5 个核调形式中， 低升（lowrise）和降升（fallrise）的两个调型引起呼吸的重值。这意味着特定音高模式会影响呼吸节奏。

韩语朗读语料的这些呼吸节奏表现表明，于在前的汉语语料的研究所提出的呼吸现象的定义相比，韩语呼吸节奏的大概的模式差异不大，这表示呼吸的调解基本上是人的生理学的结构来决定的，而不是语言类型学上的不同会区分出来的。

而且，不同语调群所产生的不同呼吸节奏表现引起我们的注意。我们通过本次实验，确认语调模式与呼吸之间相当复杂的关系，这里有我们以后的研究要进一步解释的问题所在。

参考文献

[1]谭晶晶, 孔江平. 新闻朗读的呼吸节奏初探，第七届中国语音学学术会议暨语音学前沿问题国际论坛, 2006

[2]O'Connor, J.D. and Arnold, G.F. 1973. *Intonation of Colloquial English*, 2nd edn. London: Longmans.

[3]Liberman,P. 1967. *Intonation, Perception, and Language.* MIT Press, Cambridge, Massachusetts.

[4]Seib Nooteboom. *The prosody of speech: Melody and Rhythm.* Handbook of Phonetic Sciences. Edited by Hardcastle and Laver, Blackwell Publishers Ltd, 1997,1999.

[5]Jun, Sun-Ah. 1993. *The Phonetics and Phonology of Korean Prosody.* Ph.D. dissertation, The Ohio State University.

[6]Myung-Sook Jung. 2002. *The Teaching Method of Korean Intonation by Basic Pattern.* Journal of Korean language Education 13-1:225~241.

[6]Lee, H.Y. 1990. *The Structure of Korean Prosody*, Seoul: Hanshin Publishing Co.(in Korean)

[7]Lee, H.Y. 1997. *Korean Prosody.* Laboratory of Korea.(in Korean)

日本‘能’中语音和嗓音的声学初探

A Pilot Study of Acoustic and EGG signals of Singing Voice of Japanese Noh

吉永郁代

Ikuyo Yoshinaga

提要 本文以日本歌舞‘能乐’中的‘能’为研究对象，采用提取语音信号中的时长和共振峰，分析文读的语音与‘能’的语音之间的关系。还采用提取 EGG 信号中的嗓音基频，开商参数和速度商参数，分析这 3 个参数之间的关系。得出的结论主要有：1）‘能’中语音的每个时长的长短是描叙各种风格的重要要素之一；2）共振峰频率值带抖动的趋势，主要是 F2、文读与‘能’之间的元音/u/的共振峰频率值呈比较大的变化；3）‘能’的歌声听觉上很低，但各基频平均值表明‘能’的基频高于文读的基频；4）唱‘能’时的开商比文读时的低；5）文读时基频与开商是正向相关，唱‘能’时其关系不一定是正向关系；6）速度商与基频、开商之间有负相关的趋势。

Abstract： The purpose of this study is to analyze the singing voice of Noh by comparing with its speaking voice, and also to reveal the style of expression of traditional culture in Noh singing. The acoustic and the electroglottography(EGG) signals are used for this study. The results show that 3 typical parts of Noh singing are different each other. When F0 of Noh singing is higher than that of speech, OQ of Noh singing is lower than that of speech, and SQ of Noh singing is basically lower than that of speech. The statistical analysis displays that the tendency of positive correlations between F0 and OQ both in speech and singing, and the tendency of negative correlations between SQ and F0, and SQ and OQ both in speech and singing.

关键词： 能乐；声学分析；基频；开商；速度商

Key words: Noh, acoustical analysis, F0, OQ, SQ

0 引言

‘能乐[1]’是在日本发展的历史悠久的传统艺能。2001 年登录到世界无形文化遗产的人类口传文艺的杰作。‘能乐’起源于中国的散乐。通过一些在日本国内的变化和外来的艺能的影响，变成了现在的‘能乐’。‘能’的歌声听起来很低，具有低沉的美[1]。关于‘能乐’有大量的研究成果，但以往对于其声学方面的研究很少。声学分析用的录音材料的不足是其理由之一。

对于歌声的研究，在国内外以往有不少研究，近年来歌声合成成为语音合成的热门话题。但大部分的研究集中在西乐唱法，对于其它唱法，其研究成果还不多。

本文主要讨论文读的语音和‘能’的语音之间的语音和嗓音问题。以往对于这方面的研究很少。在‘能’和‘狂言[2]’的元音对比研究中，采用共振峰，讨论各唱法和文读之间的元音音色[2]。在西乐唱法和日本传统唱法之间的对比研究中，关于‘能’的结论主要有：1）共振峰的变化很复杂，2)在一个元音内部，声音音质及音强会变化，3）辅音发得较长，往往有复杂而多样的变化[3]。对于‘能’、‘狂言’和Bel Canto唱法之间的距离，即各个唱法与文读语音音色之间的距离，结论提出：1）歌声与文读的语音之间的距离是Bel Canto>狂言>能，就是‘能’与文读的语音之间的关系比Bel Canto[3]、狂言与文读之间的关系还接近[4]。在嗓音研究方面，采用EGG信号，讨论基频、开商和速度商的关系，相关的结论主要有：1）速度商和音调高低成反比；2）典型的声带开相和闭相的比值为 1.21，开相稍大于闭相；3）开商与音调的高低成反比[5]。

1 实验说明

1.1 录音及分析软件

本次实验的录音设备是 Powerlab 肌电脑电仪及配套软件 ADInstruments 公司的 Powerlab Chart5 for Windows。本次实验采集了三路信号：第一通道是声音信号，第二通道是电子声们仪（EGG）信号 ，第三通道是通过 MLT1132 呼吸带传感器采集的呼吸信号，如图 1 所示。用 Chart5 来采集的 3 个通道信号如图 1。

[1] ‘能乐’演出一般包括‘能’和‘狂言’，是两个不同的歌舞形式。

[2] ‘能乐’中的另外一种歌舞形式。

[3] 起源于 18 世纪意大利的一种唱法。

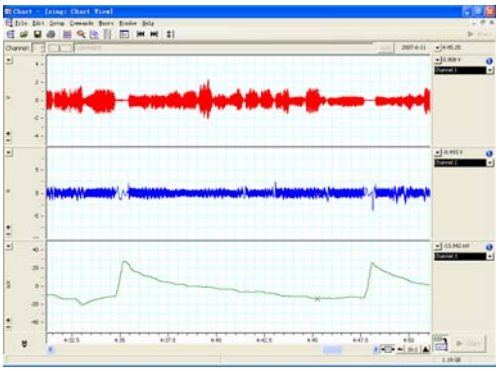


图 1　Chart5 采集的 3 个通道信号

信号主要使用美国 Kay 公司的 Real-Time EGG Analysis Model 5138、praat 和 SPSS 11.0 来进行分析。

1.2　发音人

发音人是 30 年代后期出身于日本北陆的一位男性。北陆是一个能乐很活跃的地方。他已经跟着宝生流[1]的老师学习‘能乐’19 年，唱‘能’的经验很丰富。

1.3　实验语料

发音材料是一部歌舞‘鹤亀[2]’和日语 50 音。‘鹤亀’由‘序破急[3]’组成的典型的一部歌舞。歌舞的录音材料有唱的与文读的两种形式。

实验后将记录的信号切成每一句，在每个音节的边界，声母和韵母的边界打了标记。录音信号中取出音节来提取出时长、基频、开商和速度商等的参数。

2.　‘能’的语音信号分析

2.1 摩拉时长

‘能’是由三个部分组成的。‘序’是开头部分，低速度地开始，到了‘破’开始乐器的演奏，往往这个时候能乐戏里的主角登场，和配角进行问答。所以在‘破’部分往往包括唱得好像说话一样的小段。最后‘急’部分又加速又加节奏感的高潮。这‘序破急’的概念是相等于绝句的起承转合。

下面采用时长参数进行分析这三部幕的各唱法的关系。发音材料是‘能’的三个部分‘序破急’部分的各 3 句，一共 9 句包含 185 个摩拉音。采用统计软件spss11.0。由于日语是摩拉语言，时长分析用摩拉单位[4]来进行分析。

2.1.1 文读和‘能’的单摩拉时长对比

表 1　文读和‘能’的单摩拉时长

	最小时长	最大时长	平均时长	标准差
文读	0.068	0.302	0.162	0.04
能	0.181	2.771	0.652	0.41

时长单位 秒

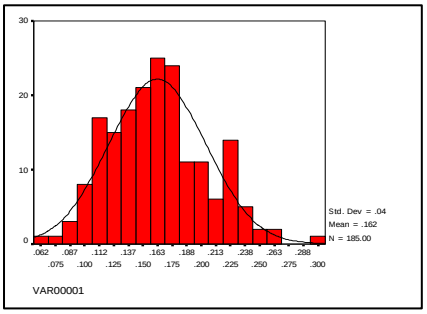


图 2　文读的时长概率分布

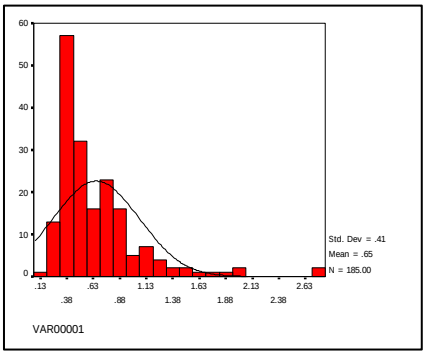


图 3‘能’的时长概率分布

从表 1 看，文读的平均时长为 0.162 秒，‘能’的平均时长为 0.652 秒。‘能’和文读之间的时长差异实在很大。‘能’的最小时长和最大时长之间的范围很宽，因此，‘能’的标准差也比文读高得多。看图 2 和图 3，文读的时长分布集中在大概平均值的位置。虽然‘能’的分布范围很宽，大多数的数据集中在 0.38~0.88 秒的位置。拉长的音是很少的几个音。文读中的元音，特别在擦音和塞音后往往呈清音化的现象，变得很短或者元音本身会脱落，摩拉时长变得更短。

2.1.2 ‘能’序，破，急之间的时长对比

2.1.2.1‘序’

[1]能有好几个流派，其中 5 个是大派，宝生流是其中一个。
[2]鹤亀’的内容是中国唐时代的一篇故事。
[3]‘序破急’是原来的能乐形式，现在也有不少不遵守这规则的。

[4]日语里的单音节具有由一个摩拉和两个摩拉组成的。把由两个摩拉组成的词的时长切成两个摩拉来进行分析。

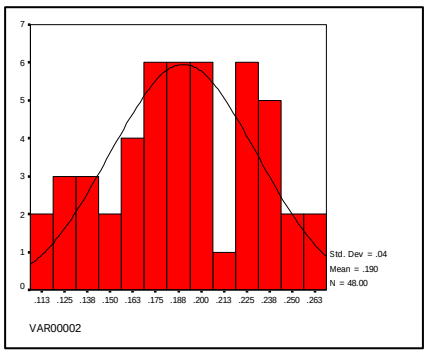


图 4　文读的时长概率分布

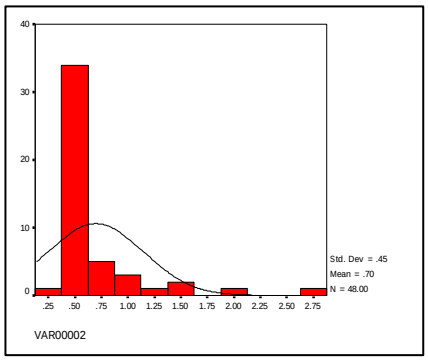


图 5　‘能’的时长概率分布

看图 4 和图 5，‘能’的数据集中在偏左的 0.5 秒左右的位置，平均为 0.69 秒。66 个样本当中的 30 多个，就是全数据的一半的音在这个位置。标准差为 0.45，时长的分布比较宽，拉长的音也不少，不过一半的时长集中在某一个位置，说明这‘序’是基本节奏上比较稳定的。文读的数据集中在图的中间的位置，标准差为 0.04，分布范围很窄。平均值为 0.19 秒。

2.1.2.2‘破’

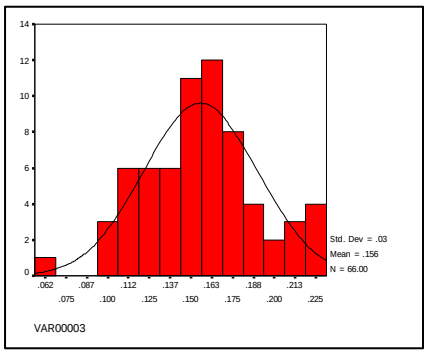


图 6　文读的时长概率分布

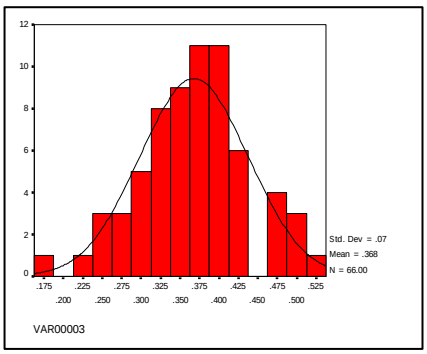


图 7　‘能’的时长概率分布

图 6 和图 7 可见，两个数据的平均在大概分布中间的位置，标准差低，就是数据的分布范围比较窄。‘能’的最大值为 0.526 秒，远远小于‘序’，‘急’的 2.7 秒。虽然文读的时长平均值为 0.156 秒，能的平均值为 0.368 秒，‘能’的比文读的长大约一倍，但两个的分布方式很有相似之处。因为‘破’是一种好像说话一样唱的一小段，结果‘破’的时长分布很像文读的分布。

2.1.2.3‘急’

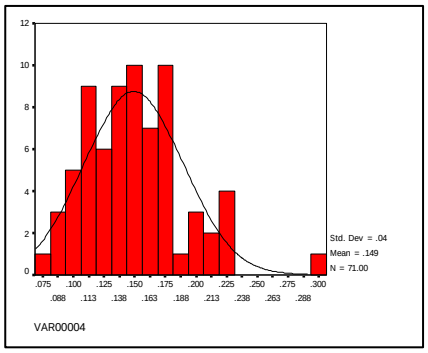


图 8　文读的时长概率分布

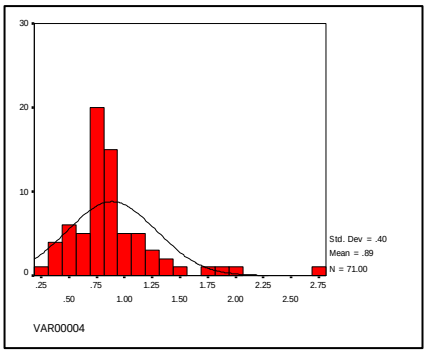


图 9　‘能’的时长概率分布

文读中的音的时长比较稳定，标准差值为 0.04 秒，平均值为大约 0.15 秒，‘能’的标准差值为 0.40，分布范围比较宽。但很多音的时长都集中在 0.8 秒左右的位置。这数值‘序，破，急’当中最长。

2.1.2.4 小结

表 2　序，破，急之间的时长对比表

	最小时长	最大时长	平均时长	标准差
文　读（序）	0.110	0.261	0.190	0.04
能（序）	0.365	2.771	0.697	0.45
文　读（破）	0.068	0.225	0.156	0.03
能（破）	0.181	0.526	0.368	0.07
文　读（急）	0.076	0.302	0.149	0.04
能（急）	0.283	2.769	0.886	0.40

见图 2 和图 3，文读的‘序破急’所有的标准差很低，每个摩拉的时长比较稳定。平均值大概在于分布中间的位置。‘能’的标准差很大，但其中‘破’的数值比较低。‘序’，‘破’，‘急’之间的时长还是有各个的特点。‘序’的标准差最大，其分布范围跟‘急’的分布范围没有差别，但每个摩拉的时长平均值比‘急’短大约 0.2 秒。‘序’是低速开始的戏剧的开头部分，3 部幕中最后一部幕‘急’的平均时长还是最长。第 2 部幕的‘破’往往包括跟说话一样的风格来唱的部分，标准差值为 0.07，在这三部幕当中时长最稳定的。平均值为 0.368 秒比‘序’，‘急’的平均值远远短。

这说明时间概念是描叙‘序，破，急’的风格的一个重要的要素。

2.2 共振峰

共振峰频率是元音声学特性的重要要素，跟元音音色很有密切的关系。文读和‘能’的元音，听觉上有音色的不同。分别取文读与‘能’中的 5 个元音，采取共振峰数据，进行平均。文读和能的发音材料都是持续元音/a/，/i/，/u/，/e/，/o/。文读的材料是稳定的平音元音段，每一个音 0.5 秒左右，‘能’的材料是稳定的元音段，每一个音 1.5 秒左右。

表 3　文读与‘能’的单元音共振峰对比（平均）

元音	文读			能		
	F1	F2	F3	F1	F2	F3
a	459.5	871.5	2796.7	556.9	749.4	2923.8
i	303.8	1785.6	2772.7	454.3	1891.2	2905.3
u	611.7	1933.9	3015.1	473.5	1520.9	2628.8
e	404.3	1836.3	2370.6	410.5	1875.4	2405.4
o	497.3	1824.6	2886.7	458.3	1826.3	2924.8

单位 Hz

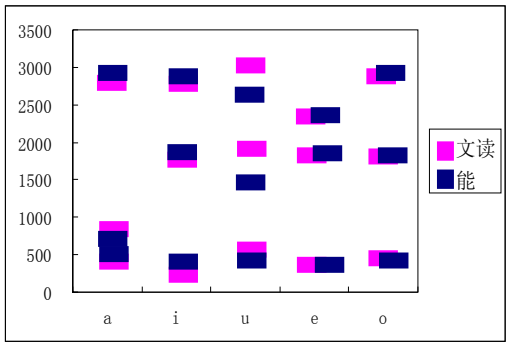


图 10　文读和‘能’的 5 个单元音共振峰对比　单位 Hz

声腔、共振峰频率和元音音色之间有很密切的关系。F1 和舌位高低很有密切的关系，舌位高，F1 就低。舌位低，F1 就高。F2 和舌位前后有关，舌位靠前，F2 就高。舌位靠后，F2 就低。F2 的上升和前共振腔的大小也有关，舌位往后移，前共振腔的空间变大，F2 就下降。舌位往前移，前共振腔的空间变小，F2 就升高。F3 和嘴唇的圆唇有关，圆唇而往前突出时，F3 就降低。

见图 10，‘能’和文读的五个元音/a/，/i/，/u/，/e/，/o/的 F1，F2，F3 平均值没有很大的差异，其中/u/呈比较大的差异。‘能’/a/的 F1 上升，这意味着开口度变大，F2 下降意味着‘能’的/a/是比文读的/a/靠后的元音。

‘能’的元音/i/的整个 F 值都稍微升上。开口度变大了以后 F1 上升。

元音/u/的整个 F 值都下降。日语的元音/u/是国际音标[ɯ]来描写的一个比较靠中间的，不太圆唇的元音。‘能’的元音/u/是靠后的音，舌位的后移后前共振腔的空间变大，这引起 F2 的下降。又圆唇化引起 F3 的下降，就是整个 F 值下降。

元音/e/和/u/在文读与‘能’之间没有大的差别。

‘能’的唱法是一种压下声带的发声类型。压下了声带以后声道就变长，整个 F 值会下降。不过实验结果来看共振峰在文读与‘能’之间没有很大的差别。元音/u/的共振峰变化比较大。

‘能’的元音的共振峰，特别是 F2 有抖动的趋势。照看图 11。

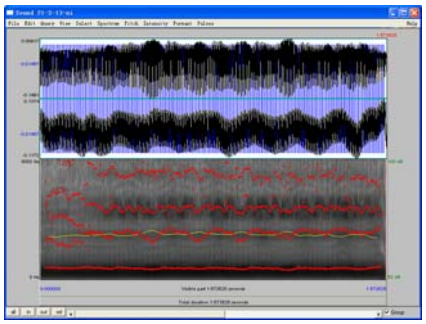


图 11　元音/u/的共振峰抖动

3. 嗓音分析

能的唱法是一种压下声带的状态下发声的很独特的唱法。在这主要讨论‘能’和文读的嗓音问题。在音调高低研究中，基频往往是音调特性的重点，利用基频参数来研究的成果有大量。以往的研究表明只利用基频参数不能解释所有的音调现象。本项研究中，采用声门阻抗信号提取嗓音信号，利用 EGG 软件来提取出嗓音信号，主要参数包括基频（FO），开商（OQ），速度商（SQ）。利用这三个嗓音参数来讨论它们之间的内在联系。

发音材料都是持续元音/a/，/i/，/u/，/e/，/o/。文读的材料是稳定的平音元音段，每一个音 0.5 秒左右，能的材料是稳定的元音段，每一个音 1.5 秒左右。

开商定义为：

开商等于开相和周期之比

（开商=开相/周期）

速度商定为：

速度商等于开启相和关闭相之比

（速度商=开启相/关闭相）

3.1 文读与‘能’的嗓音特征

表 4　文读与‘能’的嗓音特征（平均）

（）内显示标准差

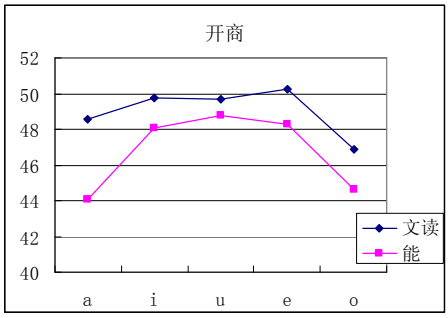


图 12　开商的文读与‘能’间对比

从数据表 4 看，文读和‘能’的开商都比较稳定，每个元音的标准差比较低，这说明文读和‘能’都在开商参数上比其它参数还稳定。文读的开商值在 46.852~50.243 之间，‘能’的开商数值在 44.070~48.784 之间。图 12 显示出了文读和‘能’的开商的关系，‘能’的所有的元音的开商比文读的低。而文读的每个元音之间的开商高低的关系与能的有相似之处。

文读的元音材料是平调的音，所以基频很稳定。虽然‘能’的元音材料也是比较平调的稳定的一段，但‘能’的唱法就是，无论拉长音的很稳定的一段，都包含一些抖动的调。结果基频的标准差比文读的高。关于基频，虽然能的歌声听起来很低，但表 4 显示‘能’的基频比文读的远远高。

速度商的标准差，文读和‘能’两个都比较高，而且关于速度商，看不出来文读和‘能’之间的明显的关系。

3.2 基频、开商、速度商的相关分析

元音	文读			能		
	F0	OQ	SQ	F0	OQ	SQ
a	113.6 (2.8)	48.5 (0.9)	411.2 (85.2)	154.9 (26.0)	44.0 (4.1)	355.4 (141.4)
i	131.0 (3.6)	49.7 (1.9)	434.6 (68.4)	191.6 (6.5)	48.0 (0.5)	295.0 (44.6)
u	126.4 (4.1)	49.7 (2.7)	457.7 (100.6)	211.5 (19.5)	48.7 (1.1)	396.8 (99.9)
e	109.6 (3.6)	50.2 (1.4)	438.6 (87.0)	193.4 (11.1)	48.3 (1.2)	373.9 (57.0)
o	126.2 (8.5)	46.8 (1.7)	320.7 (60.9)	168.6 (21.6)	44.5 (1.7)	348.8 (107.5)

主要讨论的是基频，开商，速度商三个参数中的各两个参数之间的相关关系。

3.2.1 元音/a/

表 5　文读/a/的基频及开商与速度商间相关分析结果

Correlations 文读/a/

		F0	OQ	SQ
F0	Pearson Correlatio	1	.686*	.249
	Sig. (2-tailed)	.	.000	.104
	N	44	44	44
OQ	Pearson Correlatio	.686*	1	-.288
	Sig. (2-tailed)	.000	.	.058
	N	44	44	44
SQ	Pearson Correlatio	.249	-.288	1
	Sig. (2-tailed)	.104	.058	.
	N	44	44	44

**. Correlation is significant at the 0.01 level (2-tailed).

表6 能/a/的基频及开商与速度商间相关分析结果

Correlations 能/a/

		F0	OQ	SQ
F0	Pearson Correlation	1	.772**	.141*
	Sig. (2-tailed)	.	.000	.015
	N	293	293	293
OQ	Pearson Correlation	.772**	1	.186**
	Sig. (2-tailed)	.000	.	.001
	N	293	293	293
SQ	Pearson Correlation	.141*	.186**	1
	Sig. (2-tailed)	.015	.001	.
	N	293	293	293

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

从表5和表6看，文读的基频与开商之间的相关系数为0.686，这说明基频与开商之间是比较高度的相关。‘能’的基频与开商之间相关系数为高度相关。文读的基频与速度商之间，还有开商与速度商之间相关性很低。‘能’的基频与速度商相关系数很低，只有0.141，但不相关的概率为0.015，小于0.05。速度商与开商也是相关系数很低，只有0.186，但不相关的概率为0.001，小于0.05。

3.2.2 元音/i/

表7 文读/i/的基频及开商与速度商间相关分析结果

Correlations 文读/i/

		F0	OQ	SQ
F0	Pearson Correlation	1	.590**	-.332*
	Sig. (2-tailed)	.	.000	.028
	N	44	44	44
OQ	Pearson Correlation	.590**	1	-.902**
	Sig. (2-tailed)	.000	.	.000
	N	44	44	44
SQ	Pearson Correlation	-.332*	-.902**	1
	Sig. (2-tailed)	.028	.000	.
	N	44	44	44

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

表8 能/i/的基频及开商与速度商间相关分析结果

Correlations 能/i/

		F0	OQ	SQ
F0	Pearson Correlation	1	-.121*	-.129*
	Sig. (2-tailed)	.	.044	.032
	N	278	278	278
OQ	Pearson Correlation	-.121*	1	.231**
	Sig. (2-tailed)	.044	.	.000
	N	278	278	278
SQ	Pearson Correlation	-.129*	.231**	1
	Sig. (2-tailed)	.032	.000	.
	N	278	278	278

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

看表7和表8，文读的基频与开商之间的相关系数为0.590，是比较高度的相关。基频与速度商之间的相关系数为-0.332，不相关的概率为0.028，小于0.05，是低度负相关。速度商与开商之间是相当高度的负相关关系。‘能’的基频与开商之间的相关系数为-0.121，但不相关的概率为0.044，小于0.05。开商与速度商之间也是相关系数比较低，但不相关的概率小于0.05，是有意的负相关关系。

3.2.3 元音/u/

表9 文读/u/的基频及开商与速度商间相关分析结果

Correlations 文读/u/

		F0	OQ	SQ
F0	Pearson Correlation	1	.773**	-.729**
	Sig. (2-tailed)	.	.000	.000
	N	44	44	44
OQ	Pearson Correlation	.773**	1	-.924**
	Sig. (2-tailed)	.000	.	.000
	N	44	44	44
SQ	Pearson Correlation	-.729**	-.924**	1
	Sig. (2-tailed)	.000	.000	.
	N	44	44	44

**. Correlation is significant at the 0.01 level (2-tailed).

表10 能/u/的基频及开商与速度商间相关分析结果

Correlations 能/u/

		F0	OQ	SQ
F0	Pearson Correlation	1	.620**	-.848**
	Sig. (2-tailed)	.	.000	.000
	N	230	230	230
OQ	Pearson Correlation	.620**	1	-.602**
	Sig. (2-tailed)	.000	.	.000
	N	230	230	230
SQ	Pearson Correlation	-.848**	-.602**	1
	Sig. (2-tailed)	.000	.000	.
	N	230	230	230

**. Correlation is significant at the 0.01 level (2-tailed).

从表9和表10看，文读的基频与开商之间的相关系数很高为0.773，是高度相关。‘能’的基频与开商之间也是比较高度的相关关系。文读的基频与速度商之间的相关系数为-0.729，是高度负相关。‘能’也基频与速度商之间是相当高的负相关。开商与速度商之间，文读和‘能’两个都是有意的负相关关系。

3.2.4 元音/e/

表11　文读/e/的基频及开商与速度商间相关分析结果

Correlations　文读/e/

		F0	OQ	SQ
F0	Pearson Correlation	1	.797**	-.007
	Sig. (2-tailed)	.	.000	.967
	N	44	44	44
OQ	Pearson Correlation	.797**	1	-.373*
	Sig. (2-tailed)	.000	.	.013
	N	44	44	44
SQ	Pearson Correlation	-.007	-.373*	1
	Sig. (2-tailed)	.967	.013	.
	N	44	44	44

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

表12　能/e/的基频及开商与速度商间相关分析结果

Correlations　能/e/

		F0	OQ	SQ
F0	Pearson Correlation	1	.740**	-.721**
	Sig. (2-tailed)	.	.000	.000
	N	298	298	298
OQ	Pearson Correlation	.740**	1	-.511**
	Sig. (2-tailed)	.000	.	.000
	N	298	298	298
SQ	Pearson Correlation	-.721**	-.511**	1
	Sig. (2-tailed)	.000	.000	.
	N	298	298	298

**. Correlation is significant at the 0.01 level (2-tailed).

见表11和表12，文读和‘能’都基频与开商之间是有意的相关关系，文读的相关系数为0.797，能的为0.740，是高度相关。开商与速度商之间也是有意的负相关关系，文读的相关系数为-0.373，‘能’的为-0.511，两个都是负相关。文读在基频与速度商之间毫无有相关性，但‘能’在基频与速度商之间是相当高的负相关性。

3.2.5 元音/o/

表13　文读/o/的基频及开商与速度商间相关分析结果

Correlations　文读/o/

		F0	OQ	SQ
F0	Pearson Correlation	1	.785**	.209
	Sig. (2-tailed)	.	.000	.174
	N	44	44	44
OQ	Pearson Correlation	.785**	1	.086
	Sig. (2-tailed)	.000	.	.581
	N	44	44	44
SQ	Pearson Correlation	.209	.086	1
	Sig. (2-tailed)	.174	.581	.
	N	44	44	44

**. Correlation is significant at the 0.01 level (2-tailed).

表14　能/o/的基频及开商与速度商间相关分析结果

Correlations　能/o/

		F0	OQ	SQ
F0	Pearson Correlation	1	-.009	.252**
	Sig. (2-tailed)	.	.869	.000
	N	366	366	366
OQ	Pearson Correlation	-.009	1	-.335**
	Sig. (2-tailed)	.869	.	.000
	N	366	366	366
SQ	Pearson Correlation	.252**	-.335**	1
	Sig. (2-tailed)	.000	.000	.
	N	366	366	366

**. Correlation is significant at the 0.01 level (2-tailed).

看表13和表14，文读基频与开商之间的相关系数为0.785，是高度相关，但*能*相关性很低。基频与速度商之间，文读的相关性很低。虽然能的基频与速度商之间的相关系数比较低。开商与速度商之间，文读的相关性比较低，‘能’的性关系数也比较，但不相关的概率为0，小于0.01，是负相关。

3.2.6 小结

5个元音的基频、开商、速度商的相关分析总结如下。

表15　3个参数之间的有意相关系数

	文读 a	文读 i	文读 u	文读 e	文读 o
F0 与 OQ 间	.686	.59	.773	.797	.785
F0 与 SQ 间		-.332	-.729		
OQ 与 SQ 间		-.902	-.924	-.007	
	能 a	能 i	能 u	能 e	能 o
F0 与 OQ 间	.772	-.121	.62	.74	
F0 与 SQ 间	.141	-.129	-.848	-.721	.252
OQ 与 SQ 间	.186	.231	-.602	-.511	-.335

=正向相关　=负相关

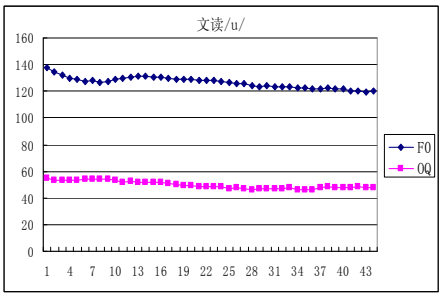


图13　文读/u/的基频与开商间对比

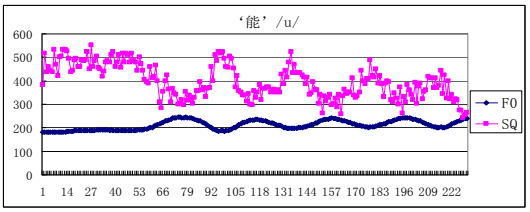


图14　能/u/的基频与速度商间对比

从表15看，统计结果说明文读的基频与开商

之间，5个元音当中，都有比较高度的正向相关（照看图13）。基频与速度商之间，只有元音/i/与/u/是负相关。开商与速度商之间，/i/，/u/，/e/三个元音是负相关。其中/i/和/u/的相关系数为-0.9以上的高度负相关。基频和开商是正向高度相关关系，元音/i/，/u/是跟开商有高度负相关关系，但速度商和基频之间的负相关没有和开商之间的密切联系性。‘能’的基频与开商之间的关系，一半以上是相关关系，但没有文读的稳定。基频与速度商之间，开商与速度商之间，负相关的趋势比较强（照看图14），但也有表示相关关系的，不过其关系系数比较低。

总之，基本上基频与开商之间成立正向相关关系。速度商与基频或者开商之间呈负相关的趋势。‘能’的唱法是一种压下声带的状态下唱的、包含强度抖动信号的唱法。这唱法引起基频，开商和速度商之间又低规则性又和文读不同的结果。

4 结论与问题

通过这次实验结果，得出以下几条结论：1）‘能’中语音的每个时长的长短是描叙各种风格的重要要素之一；2）共振峰频率值带抖动的趋势，主要是 F2、文读与‘能’之间的元音/u/的共振峰频率值呈比较大的变化；3）‘能’的歌声听觉上很低，但各基频平均值表明‘能’的基频高于文读的基频；4）唱‘能’时的开商比文读时的低；5）文读的基频与开商是正向相关，唱‘能’时其关系不一定是正向关系；6）速度商与基频、开商之间有负相关的趋势。

‘能’的唱法是一种独特的唱法与文读时的语音有所不同的。这种唱法产生的基频、开商和速度商的参数之间的内在规律还需要做进一步的研究。

参考文献

[1]小林保治. 能楽ハンドブック. 東京：三省堂，1993.
[2]長幡大介，柳田益造、中山一郎. 能と狂言における母音の違い. 音声文法研究会， 2000.
[3]中山一郎. 邦楽と洋楽の歌唱はどう違うか？. 日本音響学会誌，2000，56(5)：343-348.
[4]YOSHIOKA Noriko, NAGAHATA Daisuke, YANAGIDA Masuzo, et al. Differences among Vowels used in Noh, Kyogen and European Classical Singing. Technical report of IEICE, 2001, 1: 1-8.
[5]孔江平. 论语言发声. 北京：中央民族大学出版社，2001.

MRI-based Study of Morphological and Acoustical Properties of Mandarin Sustained Steady Vowel

Gaowu WANG12, Tatsuya KITAMURA3, Xugang LU2, Jianwu DANG2, Jiangping KONG1

(1) Phonetics Lab, Peking University, Beijing, China, 100871

（2）School of Information Science, Japan Advanced Institute of Science and Technology

1-1, Asahidai, Nomi, Ishikawa 923-1292, Japan

(3) Dept. of Information Science and Systems Engineering, Faculty of Science and Engineering, Konan University

8-9-1, Okamoto, Higashinada, Kobe, 658-8501, Japan

Abstract: To establish an elaborate vocal tract model, it is necessary to obtain more detailed 3D vocal tract shapes. Such work has been done on some languages, e.g., English, French, Japanese and Swedish, using Magnetic Resonance Imaging (MRI), while Mandarin has not been investigated in detail yet. In this study, we investigated the morphological and acoustic properties of Mandarin based on MRI measurements. MRI experiments were conducted to obtain 3D static morphologies of Mandarin vowels for one male and one female Chinese speaker. The teeth shape has been superimposed onto the volumetric vocal tract data to complement the loss of the bony tissue in MRI imaging, and then vocal tract shapes are extracted based on the 3D MRI data with the implanted teeth. A set of grid planes is designed to be perpendicular to the central line of the vocal tract, and the cross-sectional area functions are obtained by calculating the areas of the airway on the sliced grid planes. To evaluate the morphological measurements, a comparison is carried out between the formants estimated from the vocal tract area functions and the ones obtained from the real speech sound. The results showed that the MRI based formants were consistent with those from real speech sounds within 10% mismatch mostly.

1. Introduction

Magnetic resonance imaging (MRI) allows a tomographic view of body tissues in any plane of the human body, and gets the 3D shape of vocal tract, without known risks for the subject. MRI has been increasingly applied in speech research over the past 20 years (Baer et al. 1991; Lakshminarayan et al. 1991; Moore 1992; Dang et al. 1994; Yang and Kasuya 1994; Dang and Honda 1996; Demolin et al. 1996; Story et al. 1996; Tiede 1996; Alwan et al. 1997; Dang and Honda 1997; Badin et al. 1998; Honda and Tiede 1998; Engwall 1999; Fitch and Giedd 1999; Jackson and Shadle 2000; Stone et al. 2000). The articulatory data collected using MRI is valuable in understanding and modeling the vocal tract in three dimensions, particularly the pharynx area, the behavior of which during speech is traditionally hard to capture. And the volumetric production data is very important in articulatory synthesis, which the scientists have been involved for more than several decades.

MRI studies have been performed for several languages, e.g. English, French, Japanese and Swedish, and so on. However, few MRI studies work on Mandarin Chinese. In this study, we investigated the morphological properties of Mandarin vowels. The MRI experiments were conducted to obtain 3D static morphologies of Mandarin vowels for one male and female Chinese speaker. Accordingly the 3D acquisition is the main focus of this study.

2 Data Acquisition

2.1 Speech materials

Mandarin, also called Putonghua, is the official national standard spoken language of China, which is based on the principal dialect spoken in and around Beijing. This study covers the nine single vowels in Mandarin, /a o e i u ü (i)e (s)i (sh)i/, in Chinese Phoneticisation Scheme, whose IPA are / a o ɤ i u y e ɿ ʅ /, respectively. While another vowel er (/ər/), is still in dispute whether it is a single vowel, which is not covered in our present study.

The vowels are uttered in Chinese words, “啊喔屙衣乌淤噎思诗”, while in “噎思诗” we selected the stable sustained end part of the sound to do speech analysis. And all the words are in their first tone (high flat tone) to ensure the stability of the

sustained vowels.

2.2 Choice and training of subjects

There are several considerations about choosing qualified subjects:

1). The subjects should have no dyslogia or dysphonia, and are good at Mandarin, which means, they should be a native speaker of North Chinese dialects, especially those in and around Beijing, and has no dialect accent. Or he/she has passed the Putonghua Proficiency Test (PPT) and achieved Grade One, Level B (G1L2, the second highest grade, which is required for Mandarin teacher and television announcer) .

2). He/she should have no mental objects in body, and have no other diseases unsuitable for MRI experiment.

After the subjects are chosen, they should be trained to ensure the articulatory stability so as to obtain clear MRI images. The training consists of producing the speech materials in a supine position with MRI noise in the earphones. Enough practices make better imaging results.

At present, we have two subjects, both of them are North Chinese around Beijing, and the female one has passed the PPT G1L2.

2.3 MRI equipment and scan specifications

MRI data were acquired by the Shimadzu-Marconi ECLIPSE 1.5T PowerDrive 250 installed at the ATR Brain Activity Imaging Center (ATR-BAIC).

Traditionally, a long standing drawback of MRI is the image acquisition time, which was several minutes per speech taken in the earliest studies.

Recently, acquisition time was shortened to around 30s, which depends mainly on the number of image planes used, the development of the MR machines and the exploration of technical possibilities done by physicists controlling the MR machines. However, this still requires that subjects hold articulatory configurations steady for durations which are not only extra-linguistic but are also motorically difficult to produce. This might result in articulatory instability and subject motion during image acquisition, which in turn might create image artifacts.

At present, a synchronized sampling method (SSM) with external trigger pulses has been developed by (Masaki et al. 1996) to record movements of the speech organs as a set of sequential images. And this method can also be used in acquiring static 3D shape of vowels. The subject repeats the vowel about 30~36times, each time sustained 3s, which can allow subject to articulate stably and naturally. Figure 1 shows the experimental setup. The trigger device presents noise burst trains to the subject through a headset and outputs the scan pulses to the MRI scanner to synchronize the data acquisition. The subject listens to the noise burst trains to pace the utterance, while the MRI scanner initiates data acquisition synchronized with the trigger pulses.

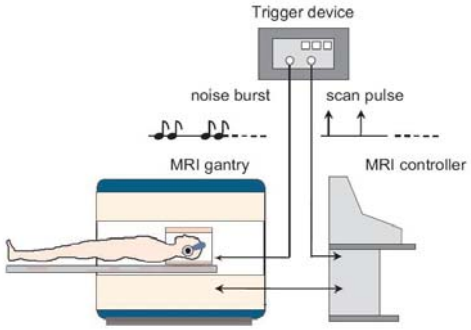


FIG. 1. Experimental setup for the synchronized sampling method. The trigger device, PC-based circuitry, controls the timing of a subject's utterance and that of MRI scanning. The device represents noise burst trains to a subject through the headset, and it sends scan pulses to the MRI controller. The noise burst trains and scan pulses synchronize with each other. after (Takemoto et al. 2006).

The parameters used in the SSM MRI scans were as follows: 3.4 ms echo time (TE), 2200 ms relaxation time (TR), 44~51 sagittal slice planes, 1.5mm slice thickness, 1.5 mm slice interval (which means the gap and overlap between slices are 0 mm), 256*256 mm field of view (FOV), and 512*512 pixel image size. The data thus obtained consist of 44~51 sagittal slices, stored as the DICOM file format.

3. MRI data processing

3.1 Image preprocessing

The images were converted from DICOM to TIFF and denoised using ImageJ software, which is released by NIH(National Institutes of Health, USA). Figure 2 shows an example of image denoising effect.

FIG. 2. An example of image denoising effect. The noise spots have been eliminated in the right panel.

3.2 Tooth superimposition

MRI has a disadvantage in imaging the bony structure because the calcified structures lacking mobile hydrogen give no resonance signals. Accordingly, the region of the teeth is found as dark as the air space. It is therefore necessary to obtain the teeth-air boundary to measure the vocal tract area function accurately from MRI data.

We referred to a method for teeth imaging to solve this problem (Takemoto, Honda et al. 2006). The same subject lay prone in the MRI gantry holding multi mineral juice in the mouth as a contrast medium, while static MRI scan was performed with the following parameters: 11 ms echo time (TE), 3000 ms relaxation time (TR), sagittal slice plane, 51 images, 1.5 mm slice thickness, a 1.5 mm slice interval, no averaging, 256*256 mm field of view (FOV), and 512*512 pixel image size. The images thus obtained demonstrated the oral cavity with high pixel values (bright), while the teeth and jaws appeared with low pixel values (dark). This contrast makes it easy to segment the teeth with the supporting rigid structures from the oral cavity. The maxilla and the mandible with the teeth were reconstructed to obtain the "digital jaw casts," which were then manually superimposed onto the MRI volumes. Figure 3 shows the result of tooth imposition.

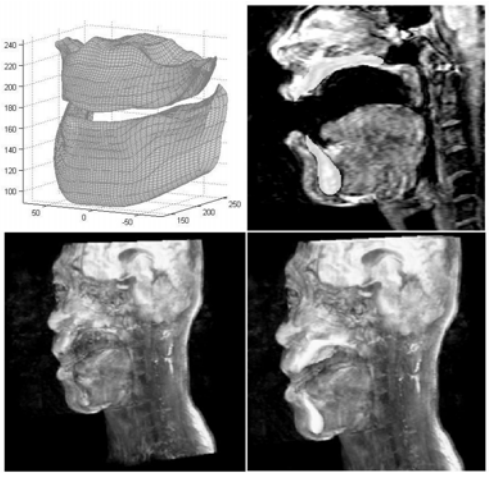


FIG. 3.The results of tooth superimposition. Upper-left is the 3D digital jaw casts; upper-right is the mid-sagittal plane after imposition; lower-left and right are the 3D show of tooth superimposition.

3.3 Extracting vocal tract area function

Vocal tract area functions were extracted from the reconstructed volumes with the teeth. The extraction was performed in three steps. First, the vocal tract midline was semi-automatically calculated in the mid-sagittal image. Along the midline, then, images perpendicular to the midline were resliced at 1~5 mm intervals. Finally, the area of the vocal tract region in each section was measured to obtain the area function. We use the method proposed by Takemoto et al. 2006. Figure 4 shows the results of calculating the vocal tract midline on an actual image: the locations of cross sections from which the vocal tract area function is measured.

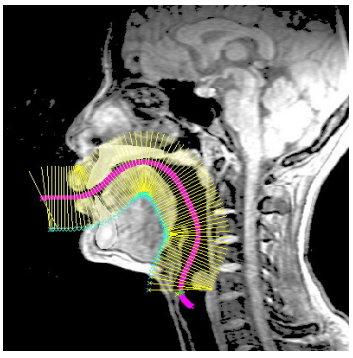


FIG. 4. The locations of grid planes

One remaining problem is how to measure the cross sectional area of the vocal tract near the lip end, where the upper and lower lips are separated and a complete circumferential outline of the vocal tract section cannot be determined. The solution used in this study followed Takemoto et al. 2006, to apply the measurable area nearest to the problematic sections.

In the furthest section from the glottis where the circumferential area could be measured, the length of the section was extended to half the distance from the end of that section to the last section where the upper and lower lips could still be observed.

3.4 Calculating transfer function

The vocal tract transfer functions were calculated for all the volumes obtained by MRI using a transmission line model, which is detailed in (Dang and Honda 1996).

3.5 Acoustic recording and analyses

Speech sounds were recorded from the subject in a soundproof room. The subject lay supine on the floor with a headset to listen to the noise burst trains. This is to reproduce the environment of MRI acquisition. In this situation, the subject is asked to repeating the vowels in the same way did in MRI experiment as possible. The speech signals and noise bursts were recorded with a record system, which consists of:

Microphone: SONY ECM-G5M

Soundcard: CREATIVE SOUNDBLASTER AUDIGY2NX

Computer: DELL XPS M1210

We selected the stable segment of the recorded vowels, and use Praat software to extract the lower four formants.

4 Results

4.1 3D inside view of Mandarin articulation

To explore the inside view, we reconstructed cutaway views for 9 Mandarin vowels based on volumetric MRI data. Thus, we obtained the 3D shape of Mandarin vowels and show the inside vocal tract and articulators, which have never been achieved by other studies of Mandarin, as we know present.

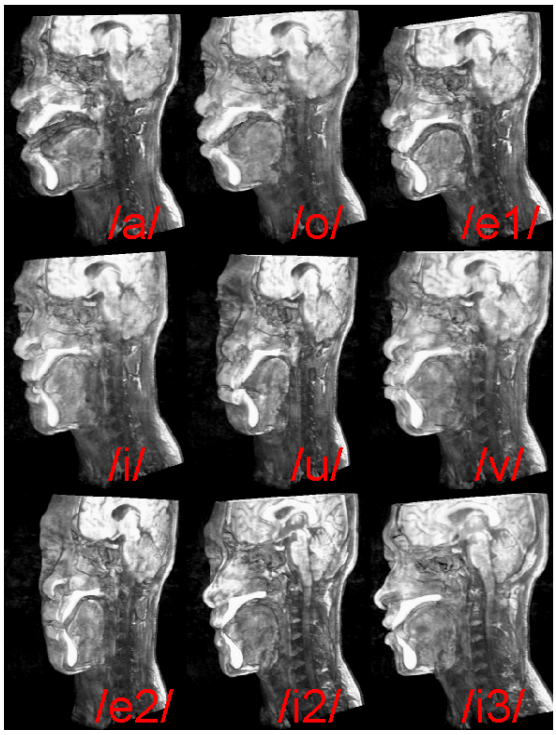


FIG. 5. The 3D shapes of 9 single vowels in Mandarin

4.2 The vocal tract shape and cross sectional areas

In order to evaluate the reliability of the area functions extracted from the 3D MRI data, the areas of the cross section had been extracted, and the corresponding formant frequencies had been derived and compared with those of real speech sound.

Figure 6 shows the results of the straightened VT volume of vowels /a i u/, and their binarized grid planes

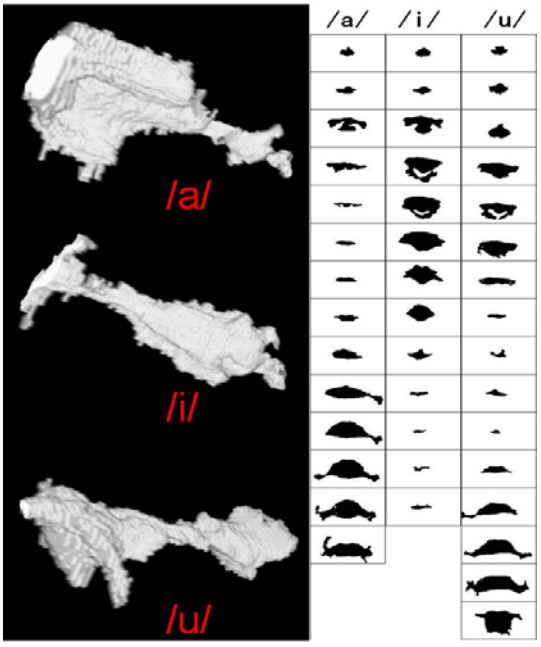


FIG. 6. The straightened 3D VT shapes of /a i u/ and their cross

sectional areas

Table 1 shows the first four formants, which are compared with those of natural speech sounds. The absolute percent error between all the formant data from the transfer functions and natural speech sounds is less than 10%. Relatively larger errors are found regionally.

Table 1: the natural and calculated speech formants of the female subject, and the percent errors for the latter relative to the former. “n” denotes the natural speech, “c” the calculation, and “d” percent errors.

	/a/	/o/	/e1/	/i/	/u/	/v/	/e2/	/i2/	/i3/
nF1	814	590	600	325	390	320	560	420	410
nF2	1312	1000	1225	2660	770	1960	2120	1450	1810
nF3	3214	3160	3140	3460	2950	2470	2850	3170	2550
nF4	4354	4380	4380	4550	4150	3800	4430	4170	3370
cF1	737	550	554	321	382	300	504	440	442
cF2	1512	880	1382	2776	832	2015	1907	1385	1790
cF3	3307	2855	3474	3348	2984	2566	2649	3315	3162
cF4	3846	3845	4441	4368	3745	4030	3876	4308	3726
dF1	-9.5	-6.8	-7.7	-1.2	-2.1	-6.3	-10.0	4.8	7.8
dF2	15.2	-12.0	12.8	4.4	8.1	2.8	-10.0	-4.5	-1.1
dF3	2.9	-9.7	10.6	-3.2	1.2	3.9	-7.1	4.6	24.0
dF4	-11.7	-12.2	1.4	-4.0	-9.8	6.1	-12.5	3.3	10.6

5 Conclusions and Discussions

In this study, the 3D static morphologies of 9 sustained Mandarin vowels had been investigated. The production data, in which the different articulators are detailed depicted, will help us to establish an articulatory vocal tract model.

The 3D inside view model will also be helpful in Visual Speech Training Aid for those disable persons who cannot hear but learn speech only by mimic the positions and movements from what they can look.

As for the formants calculated from the production data have an error no more than 10% mostly, which is advanced than early research. (Bao 1983) had calculated the formants of Mandarin of a female subject, as listed in table 2.

Table 2. The results of Bao (1983). ‘r’ denotes the revised values

	/a/	/o/	/e1/	/i/	/u/	/v/	/e2/	/i2/	/i3/
nF1	928	795	686	328	428	338	673	490	465
nF2	1300	805	1130	3000	606	2900	2575	1768	2010
cF1	621	534	487	219	304	203	467	275	297
cF1r	766	674	625	355	438	340	604	409	431
cF2	1263	1212	1542	2222	827	2070	1806	1616	1869
dF1	-33.1	-32.8	-29.0	-33.2	-29.0	-39.9	-30.6	-43.9	-36.1
dF1r	-17.5	-15.2	-8.9	8.1	2.2	0.5	-10.2	-16.6	-7.4
dF2	-2.8	50.6	36.5	-25.9	36.5	-28.6	-29.9	-8.6	-7.0

Bao (1983) obtained the area function from the X-ray photo of the 9 vowels of a female subject, and used a simplified transmission line model to calculate the formants. The percent errors of F1 are more than 30% systematically. So he adopted the experiential revised formula propose by (Lonchamp et al. 1983) to reduce the errors.

Compared with Bao (1983), our improvement may be attributed to these reasons:

1). The X-ray photo can only show the mid-sagittal plane, so Bao has to use an experiential formula to calculate the cross sectional areas from mid-sagittal dimensions. While now we can measure the real cross sectional areas using MRI.

2). Bao (1983) used the simplified transmission line model, while we adopt a model including wall impedance.

In the future, we will get more detailed vocal tract shape of Mandarin vowels and consonants, and the branched parts such as nasal cavity, piriform fossa, which will help us to establish an elaborated articulatory vocal tract model.

Acknowledgement

This study is supported by Grant-in-Aid for Scientific Research of Japan (No. 17300182) and in part by SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan.

References

[1]Alwan, A., S. S. Narayanan, et al. (1997). "Toward articulatory-acoustic models for liquid consonants based on MRI and EPG data. Part II: The rhotics." Journal of the Acoustical Society of America 101: 1078-1089.

[2]Badin, P., G. Bailly, et al. (1998). A three-dimensional linear articulatory model based on MRI data. Proceedings of the Third ESCA/COCOSDA International Workshop on Speech Synthesis: 249-254.

[3]Baer, T., J. C. Gore, et al. (1991). "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels." The Journal of the Acoustical Society of America 90(2): 799-828.

[4]Bao, H. Q. (1983). "On the relationship between cross-sectional area function of vocal tract and formant frequencies of vowel: A preliminary report." Report of Phonetic Research, Phonetics lab,

CASS.
[5]Dang, J. and K. Honda (1996). "Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation." The Journal of the Acoustical Society of America 100(5): 3374-3383.
[6]Dang, J. and K. Honda (1997). "Acoustic characteristics of the piriform fossa in models and humans." The Journal of the Acoustical Society of America 101(1): 456-465.
[7]Dang, J., K. Honda, et al. (1994). "Morphological and acoustical analysis of the nasal and the paranasal cavities." The Journal of the Acoustical Society of America 96(4): 2088-2100.
[8]Demolin, D., T. Metens, et al. (1996). Three-dimensional measurement of the vocal tract by MRI. Proceedings of 4th ICSLP. Philadelphia: 272-275.
Engwall, O. (1999). "Vocal tract modeling in 3D." KTH STL-QPSR: 31-38.
[9]Fitch, W. T. and J. Giedd (1999). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging." The Journal of the Acoustical Society of America 106(3): 1511-1522.
[10]Honda, K. and M. Tiede (1998). An MRI study on the relationship between oral cavity shape and larynx position. ICSLP 5th.
[11]Jackson, P. J. B. and C. H. Shadle (2000). Aero-Acoustic Modelling of Voiced and Unvoiced Fricatives based on MRI Data. 5th Seminar on Speech Production: Models and Data. Kloster Seeon, Bavaria, Germany: 185-188.
[12]Lakshminarayan, A. V., S. Lee, et al. (1991). " MR imaging of the vocal tract during vowel production." Journal of Magnetic Resonance Imaging 1: 71-76.
[13]Lonchamp, F., J. P. Zerling, et al. (1983). Estimation vocal tract area functions: A progress report. Proceding of 10th ICPS: 3271.
[14]Masaki, S., M. K. Tiede, et al. (1996). "MRI-based speech production study using a synchronized sampling method." J. Acoust. Soc. Jpn.(E) 20: 375-379.
[15]Moore, C. A. (1992). "The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images." Journal of Speech and Hearing Research 35: 1009-1023.
[16]Stone, M., D. Dick, et al. (2000). Modelling the Internal Tongue using Principal Strains. 5th Seminar on Speech Production: Models and Data. Kloster Seeon, Bavaria, Germany: 133-136.
[17]Story, B. H., I. R. Titze, et al. (1996). "Vocal tract area functions from magnetic resonance imaging." The Journal of the Acoustical Society of America 100(1): 537-554.
[18]Takemoto, H., K. Honda, et al. (2006). "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data." Journal of the Acoustical Society of America 119(2): 1037-1049.
[19]Tiede, M. (1996). "An MRI-based study of pharyngeal volume contrasts in Akan and English." Journal of Phonetics 24: 399-421.
[20]Yang, C. S. and H. Kasuya (1994). Accurate measurement of vocal tract shapes from magnetic resonance images of child, female and male subjects. ICSLP. Yokohama, Japan: 623-626.

基频因素对北京话阴平、阳平声调在语境中的感知的作用*

The influence of F0 on the perception of Tone 1 and Tone 2 of Mandarin in the first syllable of disyllables

杨若晓 刘朋

Yang Ruoxiao, Liu Peng

摘要：实验考察了基频因素对北京话阴平、阳平声调在双音节中感知的作用。采用PSOLA合成语音的方法制作语音样本，并进行听辨实验。实验结果显示被试对位于前一音节的阴平、阳平声调的判断会受到后一音节起点、终点、拐点和整体基频变化的影响，二者之间的关系可以分为四种类型；其次，发生听辨错误的声调多数落在上声，这和上声“低”的特征有关。

Abstract：This experiment investigated how F0 would influence the perception of Tone 1 and Tone 2 of Mandarin in the first syllable of disyllables. The perception materials were made using PSOLA by Praat, and the perception test was carried by 10 subjects. The results indicate that with the variation of F0 in the beginning, end, the lowest of the F0 contour, as well as the variation of the whole contour, the subjects' perception of Tone 1 and Tone 2 in the first syllable changes. The relations between the variation of F0 contour and the perception can be divided into four types. Moreover, when the subjects judge Tone 1 and Tone 2 in the first syllable falsely, most of the wrong choices are Tone 3, which may be related to the “low” distinctive features of Tone 3.

关键词：语音感知 基频 阴平 阳平

Key words：phonetic perception, F0, Yinping, Yangping

0 引言

林焘、王士元（1984）在《声调感知问题》中考察了声调音域的变化对北京话声调信息传递的影响。具体来说，他们主要考察了“阴平+轻声”、“阴平+阴平”和“阴平+阳平”几种音节组合中位于前一音节的阴平声调的感知是否随后一音节基频和时长的改变而变化。实验的结果显示随着 D 值（第二音节的音高起点和第一音节的音高终点之间的音高差，文中实为基频差）的增加，位于前一音节的阴平音节往往被听为升调阳平声，如果同时前一音节的阴平音节时长较长，那么就有可能被听为上声。对于实验结果，他们进一步指出这和 Ladefoged P. & Broadbent D. E.（1957）对英语元音在共振峰变化的语句中的感知研究一样，说明同一个言语刺激由于受到了话语中其他结构的影响，可以被感知为不同的语言范畴；对于北京话阴平声调而言，它随后字的变化而被感知为阳平或者上声可能是由于听错觉造成的。

本文所报告的实验在以上实验的基础上进行，主要考察在后一音节为变化的四声时，位于前一音节的阴平、阳平音节在感知上是否会发生变化，不过本次实验中只考虑基频一个变量对听感的影响。

1 方法

以下说明本次实验的参与者、测量方法和实验程序。

1.1 参与者

首先，听辨样本的制作基于一位北京男生的录音样本，该位发音人一直在北京生活。发音人资料如下：张昊，男，北京人，24 岁。

其次，本次实验一共有 10 名听辨被试参加，5 男 5 女。所有被试都通过网络招募。10 名被试都一直在北京生活，在日常生活中使用北京话。10 名被试在完成实验后都能得到一份 15 元的礼物作为报酬。

1.2 样本制作

实验样本用软件 Adobe Audition1.5 和 Praat 制作。

首先，用 Adobe Audition1.5 录制发音人的发音，发音词表中每个项目读两遍，采样速率为 22050Hz，双通道录制，左通道录制声音信号，右通道录制喉头仪信号（EGG）。发音词表如下

*本文得到国家自然科学基金 A040511 的资助。

（其中/baba/音节为无意义的音节，/xian1ji1/音节为有意义的音节）：

表 1：发音词表

/ba1X/	/ba2X/	/xian1ji1/
巴的	跋的	鲜鸡
扒巴	拔巴	
扒拔	拔拔	
扒靶	拔靶	
扒爸	拔爸	

其次，利用 Adobe Audition1.5 的“伸展”功能对切好的声音样本进行时长归一化的处理。对于非轻声音节，例如/ba1ba1/，我们使第一音节的有声段时长为 300ms，第二音节的有声段时长为 400ms；对于轻声音节，例如/ba1de/，我们使第一音节的有声段时长为 300ms，轻声音节的有声段时长为 150ms。以上时长的设定主要基于自然语言中孤立的双音节词两音节之间的时长关系（林茂灿等，1984；王韫佳，2004）。

时长归一化后的样本再输入到 Praat 中进行基频（下面称 F0）的调整和样本合成。首先，我们根据语音样本的实际 F0 设定了阴平、阳平、上声、去声和轻声的起点、终点 F0，上声还有拐点（/ba3/音节中指最低的一个基频点）F0，见表 2，我们把这些样本称为基准音节。除 3.1 节第一组/ba1de/音节外，以后的调整都在这些点上进行。本次实验中只调整了 F0，调整的步长为 10Hz，调整时分三种情况：只调整起点处 F0（即起点每次升高或者降低 10Hz，终点、整体基频的调整方法与此相同），保持终点 F0 不变；只调整终点 F0，保持起点处 F0 不变；同时调整起点和终点处 F0（上声还包括只调整拐点 F0 的情况），调整后的语音样本由 Praat 采用 PSOLA 方法合成，不同声调组合的音节合成得到的样本数目不等，一共得到 455 个样本。

表 2：基准音节的起点、终点、拐点 F0

	/ba1/	/ba2/	/ba3/	/ba4/
起点 F0(Hz)	167	110	100	185
终点 F0(Hz)	167	154	105	90
拐点 F0(Hz)			80	

	/de/位于/ba1/后	/de/位于/ba2/后	/xian1/	/ji1/
起点 F0(Hz)	140	154	167	167
终点 F0(Hz)	90	90	167	167
拐点 F0(Hz)				

表 3：基于基准音节合成的样本数

	基于基准音节合成的样本数
第一组/ba1de/	5
第二组/ba1de/	55
/ba1ba1/	5
/ba1ba2/	35
/ba1ba3/	49
/ba1ba4/	62
/ba2de/	54
/ba2ba1/	13
/ba2ba2/	43
/ba2ba3/	64
/ba2ba4/	52
/xian1ji1/	18
总计	455

1.3 实验程序

455 个语音样本通过 Matlab 自编程序随机播放，播放声音后呈现出单个汉字或者两字组构成的选项让被试进行强迫性选择，每种情况下都有四个选项，分别对应北京话四声。具体来说，这四个选项分成三类情况：第一类针对轻声音节，所给出的四个选项为“1）巴的　2）跋的　3）把的　4）霸的”；第二类针对非轻声音节/baba/，所给出的四个选项“1）巴　2）跋　3）把　4）霸”，要求被试对所听到的样本的第一个音节做出判断；第三类针对/xianji/音节，所给出的四个选项为“1）鲜鸡　2）咸鸡　3）险机　4）献鸡”。被试在选择时可以有三次机会重听，这由被试自己控制。整个实验持续 30 到 40 分钟，中间没有休息。听辨实验在北京大学语言学实验室录音室进行。

被试的选择通过 Matlab 自编程序保存于 Excel 中。下面的数据分析在 Excel 中进行，主要统计了听辨人在后一音节不同基频下判断/ba1/、/ba2/错误的概率和被错判的声调在每组选择中分布的概率。统计时区分男女被试进行。

2 结果和讨论

2.1 第一组/ba1de/音节的听辨结果

首先，我们采用 2.2 节中所描述的方法根据林焘、王士元（1984）文中“合成语音实验”中所给出的轻声音节基频值在/ba1de/音节的基础上合成了五个测试音节。合成这五个音节的目的在于验证林焘、王士元（1984）实验中所得出的“随着第二个音节音高的升高，原来为平调的第一个音节可能被听为别的声调”的结论。下面首先给出这五个音节的基频值（由于合成中采用设定起点和终点基频，其余点基频由 praat 软件采用 PSOLA 方法自动生成，所以以下表格中只给出起点和终点的基频值，采用“X1（起点基频值）-X2（终点基频值）”的形式）。见表 1：

表 4：第一组/ba1de/音节的基频

/ba1/(Hz)	/de/(Hz)
167-167	110-70
	120-80
	130-90
	140-100
	150-110

听辨结果显示，随着第二个音节/de/整体基频值的提高，听辨人对第一个音节声调的判断发生了变化，下面的图 1 和图 2 分别展示了男女听辨人在不同基频值判断/ba1/错误的概率。从图中可以看到随着/de/基频值的提高，听辨人对/ba1/音节判断错误的概率也在升高，男女表现出了一些差异。这里的结果验证了林焘、王士元（1984）实验的结论。但是，本次实验的听辨结果显示/ba1/不仅被错判为 2 声，还被错判为 3 声，这和林焘、王士元（1984）的实验结果不太一致，他们的实验中在以上基频条件下没有被判断为 3 声的情况。

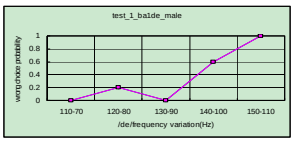


图 1：第一组/ba1de/中听辨/ba1/错误的概率（男）

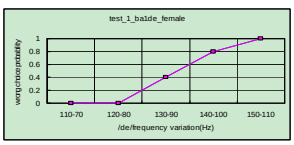


图 2：第一组/ba1de/中听辨/ba1/错误的概率（女）

2.2 第二组/ba1de/音节的听辨结果

与 3.1 节所说的样本不同，本次实验中，我们不仅对/ba1de/中第二个音节的整体音高进行了调整，还分别对其起点、终点进行了基频的逐步调整，采用这种方法调整的样本称为第二组/ba1de/音节。听辨结果显示，在只改变终点基频时，听辨人对/ba1/音节声调的判断基本不发生错误，但是在改变起点或整体基频的情况时，/ba1/音节声调被判断错误的概率增加了。

2.3 /ba1ba1/、/ba1ba2/、/ba2ba1/和/xian1ji1/音节的听辨结果

/ba1ba1/、/ba1ba2/、/ba2ba1/和/xian1ji1/音节的听辨结果则显示：只改变起点或终点基频时，听辨人对第一个音节声调的判断基本不发生错误，但是在改变整体基频的情况下，对声调判断错误的概率逐渐增加。

不过，需要说明的是在改变/ba1ba2/中 /ba2/的起点基频时，当起点基频值升高至 200Hz 时，男女听辨人对第一个音节/ba1/声调辨认错误的概率也提高了，这种表现和 3.2 节/ba1de/音节中改变/de/起点基频时的听辨结果相似。这在一定程度上说明起点处基频的高低对正确辨别声调具有一定的作用，这种作用大于终点处的基频高低，但是不如整体基频的改变那么重要。从前后音节声调的对比来讲，这一结果可能是因为起点处的基频值和前一个音节的终点基频值因为距离近而容易产生对比，这样起点处基频的高低比起终点处要容易影响到听感。

2.4 /ba1ba4/、/ba2ba2/和/ba2ba4/音节的听辨结果

与以上听辨结果不同，当采用同样的方法改变/ba1ba4/、/ba2ba2/和/ba2ba4/音节中第二音节的起点、终点和整体基频时，听辨者对第一个音节声调判断错误的概率都在某个特定值处发生了改变，也就是说在这三个音节中，随着第二音节起点、终点和整体基频值的改变，听辨人对第一个音节声调判断错误的概率都在增加。

2.5 /ba1ba3/、/ba2de/和/ba2ba3/音节的听辨结果

当/ba1/和/ba2/后接/ba3/时，我们采用同样的方法修改了/ba3/的起点、终点、整体基频以及拐点（在/ba3/音节中指最低的一个基频点）处的基频，这几套样本的听辨结果与以上几组的听辨结果都不相同。具体来说，无论改变/ba3/起点、终

点、拐点还是整体基频合成出的样本，听辨人对/ba1/音节声调的判断都没有发生错误。

而对基于/ba2ba3/和/ba2de/音节合成的样本，随基频变化被试对/ba2/音节声调判断错误的概率变化没有明显规则可循。

2.6 /ba1/和/ba2/被错判为其他声调时的概率分布

以上五节主要说明作为听辨目标音节的位于前一音节的阴平、阳平音节，随着后一音节基频的改变，被试对听辨音节的声调判断也会发生变化，这种变化可以分为 3.2 到 3.5 节所描述的四种类型。在此基础上，我们统计了被错判的第一音节声调分布于阴平（当第一音节为阳平时）、阳平（当第一音节为阴平时）、上声和去声的概率，发现发生错判的声调有很大部分落在上声上。下面以男生被试为例，给出具有代表性的被错判声调在四声上的分布概率图示（图中只画出被错判的声调的概率分布，不画出判断正确的目标声调的概率分布，因而每幅图中只有三条曲线，被错判为上声的概率分布曲线为方形格示曲线）。

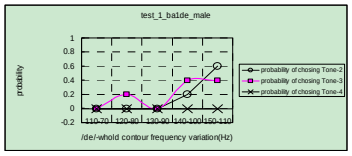


图 3：第一组/ba1de/中改变/de/整体基频时/ba1/音节被错判声调的概率分布（男）

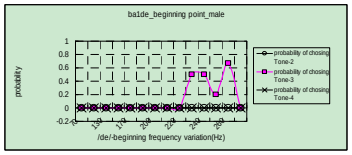


图 4：第二组/ba1de/中改变/de/起点基频时/ba1/音节被错判声调的概率分布（男）

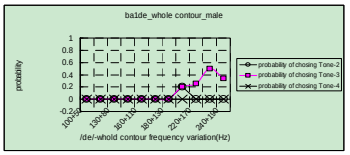


图 5：第二组/ba1de/中改变/de/整体基频时/ba1/音节被错判声调的概率分布（男）

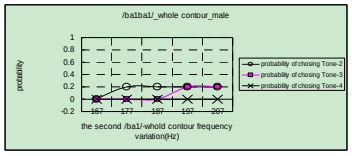


图 6：/ba1ba1/中改变后一/ba1/音节整体基频时前一/ba1/音节被错判声调的概率分布（男）

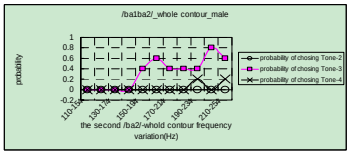


图 7：/ba1ba2/中改变后一/ba2/音节整体基频时前一/ba1/音节被错判声调的概率分布（男）

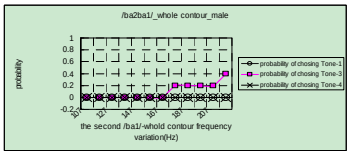


图 8：/ba2ba1/中改变后一/ba1/音节整体基频时前一/ba1/音节被错判声调的概率分布（男）

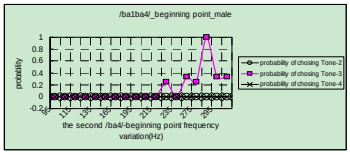


图 9：/ba1ba4/中改变后一/ba4/音节起点基频时前一/ba1/音节被错判声调的概率分布（男）

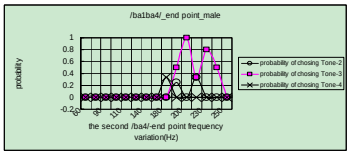


图 10：/ba1ba4/中改变后一/ba4/音节终点基频时前一/ba1/音节被错判声调的概率分布（男）

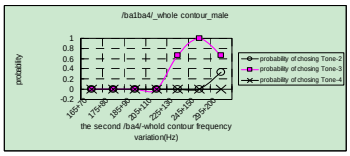


图 11：/ba1ba4/中改变后一/ba4/音节整体基频时前一/ba1/音节被错判声调的概率分布（男）

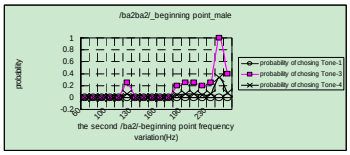


图 12：/ba2ba2/中改变后一/ba2/音节起点基频时前一/ba2/音节被错判声调的概率分布（男）

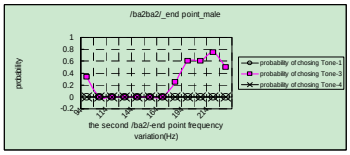


图 13：/ba2ba2/中改变后一/ba2/音节终点基频时前一/ba2/音节被错判声调的概率分布（男）

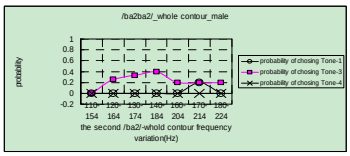


图 14：/ba2ba2/中改变后一/ba2/音节整体基频时前一/ba2/音节被错判声调的概率分布（男）

3 结论

从以上实验结果中可以得到两个结论。

首先，在双音节环境中，位于前一音节的阴平、阳平的感知会受到后一音节基频变化的影响。具体来说，本次实验再次证明林焘、王士元（1984）实验中关于阴平声调在双音节前一音节中的感知会随着后一音节基频的变化而发生改变的论述。此外，由于我们将要测试的阴平、阳平音节放置在了五种声调环境中，并设置了改变起点、终点、拐点和整体基频四种变化模式，所以本次实验的结果显示了 10 种声调组合中需要被试判断的阴平、阳平声调被错判的概率和四种变化模式的关系可以分为四类：改变后一音节起点和整体基频时会影响被试对前一音节声调的判断（3.2 节）；改变后一音节的整体基频时会影响被试对前一音节声调的判断（3.3 节）；改变后一音节起点、终点和整体基频时都会影响到被试对前一音节声调的判断（3.4 节）；改变后一音节起点、终点、拐点和整体基频时基本不影响（除一些个别的点外）被试对前一音节声调的判断和不规则的情况（3.5 节）。

错判发生的原因如林焘、王士元（1984）所说的那样，可能与听错觉有关，也就是说，在声调语言中，听觉在刚刚接受到一个音节的基频刺激时，不可能立刻判定它的音域并且把它归入到已习得的调类中去，而要和前后音节的音高进行对比才能做出决定，因而位于双音节中前一音节的声调需要和后一音节的声调对比才能确定其调类。这时，如果后一个音节基频很高（根据本次实验，可能是起点、终点和整体的基频），被试可能由于这个很高的基频的影响而产生前一音节较低的错觉，从而把前一音节的调类归入具有“非高”特征的调类中，于是很大一部分样本的前一音节被判断为上声和阳平，其中判断为上声居多。

其次，本次实验的另一个发现是即使在没有时长因素的影响下，当后一音节基频发生改变时，被错判的阴平、阳平声调大部分落在上声上（3.6 节）。这一结果应和北京话上声的特点有关。以往研究已经表明上声是北京话中唯一一个具有稳定的低的特征的声调，其深层形式中也包括低的特征（凌峰、王理嘉，2003），因而当后一音节相比前一音节基频高出许多时，被试很可能由于听错觉而主要感知到了前一音节声调“低”的特征，于是将其判断为上声。同时，反过来说，这次实验的结果也在一定程度上说明：上声被感知时的最重要的一个征兆很可能就是“低”。

不过，本次实验也还存在很多不足。例如，没有系统考虑时长的因素在声调感知中的作用；各个组别中合成的音节数未作统一，这使得不同组间的可比性降低；也未能探讨后一音节声调基频变化影响到前一音节声调感知的阈限问题；只考察了阴平和阳平两个声调，上声和去声的情况还有待研究；实验被试数目有限，更大范围的考察将更有利于反映更加普遍的事实。这些都是今后将继续考察的方向。

参考文献

[1]林茂灿，颜景助，孙国华. 1984. <北京话两字组正常重音的初步实验>. 《方言》. 1: 57-73

[2]林焘、王士元. 1984. <声调感知问题>. 《中国语言学报》2: 59-69

[3]凌峰、王理嘉. 2003. <普通话上声深层形式和表层形式>. 《第六届全国现代语音学学术会议论文集（下）》. 413-416

[4]王韫佳. 2004. <音高和时长在普通话轻声知觉中的作用>. 《声学学报》. 29.5: 453-461

[5]P. Ladefoged and D. E. Broadbent. 1957. Information Conveyed by Vowels. Journal of the Acoustical Society of America. 29: 98-104

（杨若晓 北京大学中文系语言学实验室 100871 yangruoxiao@pku.edu.cn
刘朋 IBM 中国有限公司中国系统科技实验室 100085 liupcdl@cn.ibm.com）

Lip Contour Extraction Based on Support Vector Machine

Xiaosheng Pan, Jiangping Kong, Alan Wee-Chung Liew
Peking University, Chinese University of Hong Kong, Griffith University
Tianchi03@163.com

Abstract: Lip contour extraction was a useful technique for obtaining a mouth shape in an image, and was one of the most important techniques for human-machine interface applications, such as lip reading and speech recognition. In this paper, a new method to extract lip contour from video was propose based on the fact that the lip color and skin-color was varied in the different color spaces. We first extracted frames from digital video first; then we classified face into lip area and non-lip area by the Support Vector Machine. At last we obtained some parameters from the lip area to reconstruct the lip contour. The experiment results proved that the proposed method was accurate and robust.
Keywords: Lip contour, Segment, Color space, Support Vector Machine

1. Introduction

Continuous lip information from a speaker is useful in speech recognition system especially in a noisy environment. In color image, we usually use lip and skin color information to separate lip and non-lip area. Many image segmentation techniques have been proposed [1][2][3][7]. For color image segmentation, histogram-based and clustering-based methods have been widely used. In [2], [3]Cheng et al. proposed a histogram segmentation technique which involves performing a fuzzy partition on a two-dimensional(2-D) histogram based on the maximum fuzzy entropy principle. In [7], alan et al. proposed a segmentation method which use fuzzy clustering and also take account for spatial information.

There are two main problems in lip segment. First, the lip color is similar with tongue color, so inner contour of the lip is difficult to separate from the tongue. Second, the outer contour of lip is often blurred due to the movement of the lip. The algorithm we proposed can successfully resolve the second problem but not the first one.

This paper is organized in four sections. Section two describes the approach proposed here. Section three explains the experimentation setup for testing the approach and gives the results on classification. Section four outlines major conclusions as well as gives directions for future work.

2. The method

2.1. Color Space

As we know, there are many kinds of color spaces that have been provided in existing research work, such as RGB, HSV, CMY, and Chromatic color space[4]. The original lip images are in the RGB color format. It is well known that the value [R G B] of the color image in the color space RGB not only represent chrominance but also luminance; they have high correlation. In most cases, mapping the original image into an adaptive color space can let the feature extraction task easier to be taken. After the comparison of the color spaces listed above, the chromatic color space is chosen as the color space applied in this paper. Three approximately color spaces are the 1976 CIELAB color space (L,a,b), the 1976 CIELUV color space (L,u,v) and YIQ color space (Y,I,Q). Every color component of lip image can be calculate by equation from[4]. Every color component of lip image is list in Figure.1. Because the component L in CIELAB and CIELUV has same value, so we only chose one of them. It is also found that the lip contour of the component v in color space CIELUV and the component b in color space CIELAB can be directly observed dissimilar with the original one, in particular in the area of the lip corners.

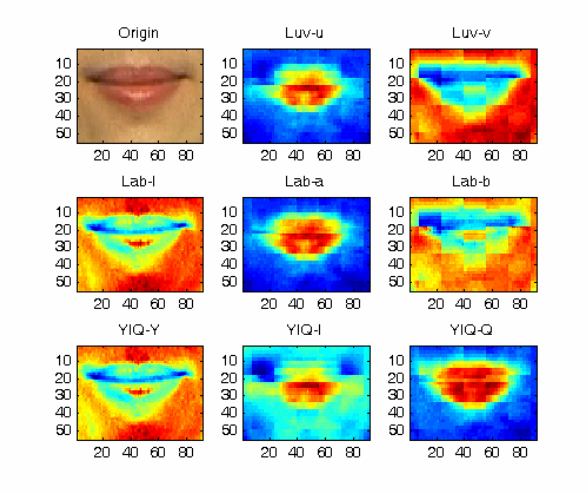


Figure.1 Different color component of lip image

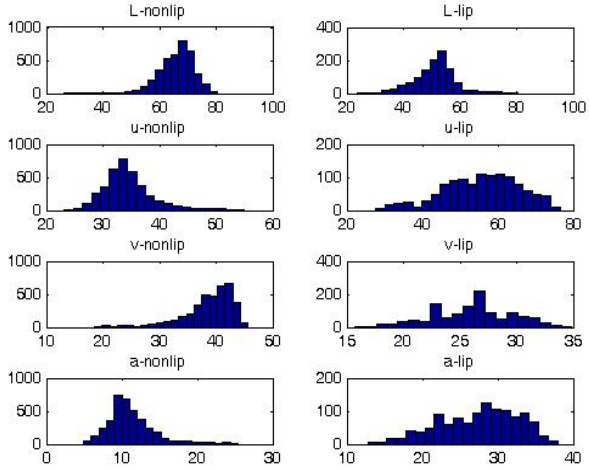


Figure.2 the histogram of color component

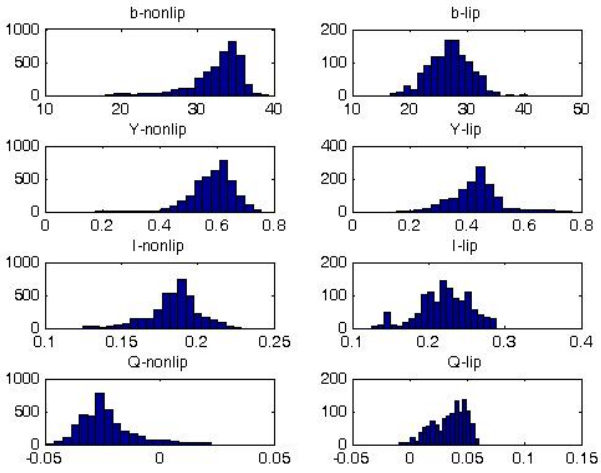


Figure.3 the histogram of color component

2.2. Histogram of lip and non-lip area

We know histogram represents the color distribution of lip and non-lip area. Figure.2 and Figure.3 show the lip area and non-lip area of different color components. The histograms in the left column represent the non-lip area pixel distribution, whereas that in the right column indicated the lip area pixel distribution. If the overlap between the pixel distraction histogram of the lip area and that of the non-lip area exceeds a certain magnitude, then it is very difficult to distinguish lip and non-lip areas. The histograms of feature v have some overlaps between lip and non-lip area. Thus it is not useful for lip contour extraction, so is that of feature b. Then feature vector defined as below:

$$F = \{L, u, a, Y, I, Q\}$$

(1)

2.3. Support Vector Machine

Support Vector Machine (SVM) has recently been successfully applied to a number of applications ranging from face identification to text categorization. The approach is systematic and motivated by statistical learning theory[5]. SVM are based on the structural risk minimization principle, closely related to regularization theory. Here we focus on SVM for two-class classification. The mathematic expression of SVM is:

$$\min_{w} \phi(w,b) = \frac{1}{2}(w \cdot w)$$

（2）

subject to the constraints:

$$y_i \,|\, (w \cdot x_i + b) \,| \geq 1, i = 1, \ldots, n$$

（3）

And the learning task can be reduced to minimization of primal Lagrangian:

$$L = \frac{1}{2}(w \cdot w) - \sum_{i=1}^{n} a_i (y_i (w \cdot x + b) - 1)$$

(4)

Where a_i are Lagrangian multipliers, hence $a_i \geq 0$. Taking the derivatives with respect to b and w and resubstituting back the primal gives the Wolfe dual Lagrangian:

$$W(a) = \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i=1}^{n} a_i a_j y_i y_j (x_i \cdot x_j)$$

(5)

which must be maximized with respect to a_i subject to the constraint:

$$a_i \geq 0 \qquad \sum_{i=1}^{n} a_i y_i = 0$$

(6)

The decision function is then given by:

$$f(x) = \mathrm{sgn}((w \cdot x) + b)$$
$$= \mathrm{sgn}(\sum_{i=1}^{n} a_i^* y_i (x_i \cdot x) + b^*)$$

(7)

An SVM is mainly characterized by its kernel function. In experiment, we use linear kernel function with parameter C=1.

3. Implementation and result

The lip video is provided by the Linguistics Lab of Department of Chinese Language and Literature of Peking University.

The lip contour extraction procedure is as following:

Step1. Get frames from the video.

Step2. Mark lip contour by handwork[6], separate the face into two component, lip and non-lip. And obtain the axes of every pixel.

Step3. Map the origin lip images into other color space and extract the feature vector F.

Step4. Select appropriate pixels as training data[8], then use SVM classify the lip images.

Step5. Extract parameters from lip area and use lip model provided from [6] to obtain a lip contour with smooth edge.

The result of the lip segment from the continuous frame is shown in Figure.4. The left column is four continuous frames extract from video. The middle column is the result of we classified the face into lip area and non-lip area by the Support Vector Machine. We are not very satisfied with the result. Because the lip edges we get are not smooth enough, and it is found that the non-lip area in the lip corners with dark color is misidentified as lip area. It is also the case for the concave non-lip area immediately below the under lip. Because the color in those position is more similar to lip area.

Alan et al. provide a lip model[6], the equations describing the lip shape Figure.5 are given by:

$$y_1 = \frac{-h_1}{(w - x_{off})^2}(|x - sy_1| - x_{off})^2 + h_1$$

(8)

$$y_2 = h_2\left(\left(\frac{x - sy_2}{w}\right)^2\right)^{1+\delta}$$

(9)

As its origin is at $(0,0)$, $x \in [-w, w]$. The parameter shows the skewness of lip shape. The skewing is attained with the ways of using the transformation matrix $\begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$. The exponent δ describes the deviation of y_2 from a quadratic curve and it allows lip with different degree of curvature to be captured accurately.

Then we extract parameters from middle column and use equations (8) and (9) to get the lip contour. The result of lip contour is list in the right column of Figure.4.

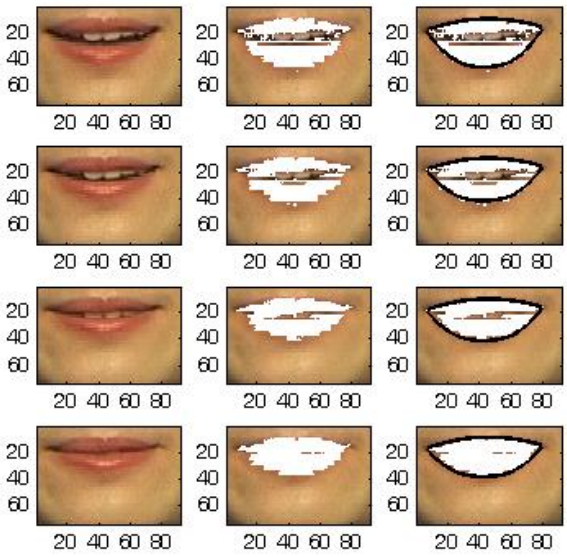


Figure.4 The left column is the lip image sequence of speaking. The middle column is corresponding the results of classify face into lip and non-lip area. The

right column is the lip contour optimized by lip model.

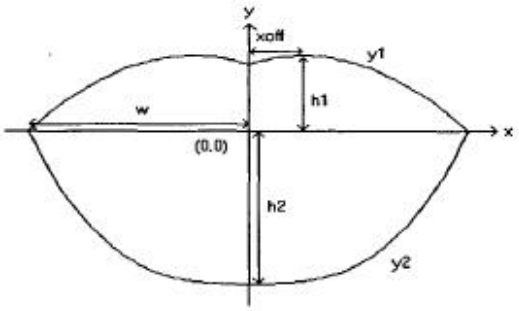


Figure.5 Lip model

4. Discussion

There still something can be done to improve the accuracy of my lip segment program. For example inner lip contour is still can not be extracted for the tongue color is similar to lip color, and get inner lip contour is our next stage target. And make some preprocessing to origin image maybe can eliminate the shadow at the underlip. Then the result of lip segment will be improved probably.

When using Support Vector Machine , we need to choose a kernel function, and decide the parameter values to kernel function. There is still no theory to prove which kernel function is best, so we need to do more experiment to find out which kernel function is the most suitable for our algorithm.

5. Reference

[1]Lewis TW, Powers D M, "Lip Feature Extraction Using Red Exclusion", In Proc. Selected papers from Pan-Sydney Workshop on Visual Information Processing, 2000, pp.61-67.

[2]. H. D. Cheng, Y. H. Chen, and X. H. Jiang, "Thresholding using two dimensional histogram and fuzzy entropy principle," IEEE Trans. Image Processing, vol. 9, pp. 732–735, Apr. 2000.

[3] H. D. Cheng, J. R. Chen, and J. Li, "Threshold selection based on fuzzy C-partition entropy approach," Pattern Recogn., vol. 31, no. 7, pp. 857–870, 1998

[4]Adrian Ford and Alan Roberts, "Colour Space Conversions", August 11, 1998 (coloureq.pdf)

[5] V. N. Vapnik. Statistical Learning Theory. Wiley, 1998.

[6] A.W.C. Liew, S.H. Leung, W.H. Lau, "Lip contour. extraction using a deformable model," In Proc. IEEE ICIP, Vol. 2, pp. 255-258, 2000.

[7]Liew A, Leung S H,Lau W H, et al. Segmentation of Color Lip Images by Spatial Fuzzy Clustering[J].IEEE Trans. on Fuzzy Systems,2003,11(4):542-549.

[8]边肇琪 张学工等编著 《模式识别》第二版 清华大学出版社 2002 年 10 月

汉语普通话声母辅音唇读认知策略的实验检验

A Cognitive Experiment on the Lip-reading strategies of Consonant Phoneme in Mandarin

吴君如

Wu Junru

摘要：在单音节条件下，唇读者不能分辨汉语普通话音系声母辅音发音方法特征中的大部分，在分辨正确率上和非唇读者没有明显差异。实验结果从一个方面验证了以下观点——汉语普通话唇读者通过唇读认知有声语言时采用了一种有取舍的认知策略，而非还原全部音系信息。此外，实验显示，唇读者面对不能分辨的对立对时做出的选择往往不是随机的，而体现着某种倾向性。

Abstract：While lip-readers of Mandarin could understand a whole sentence through lip-reading, they are unable to distinguish most pairs syllables with minimal distinct features, when given the video of a separate syllable, which shows little significant difference from the result of the non-lipreaders. A conclusion could be draw that the cognitive strategies lip-readers use to understanding speech is not to get all the phonemic information from the lip but to refer to other levels of the language.

关键词：唇读、认知

Key words：lipreading, congnitive

0 引言

所谓唇读(lipreading) ,是指通过观察说话者的口型变化, “读出”或“部分读出”其所说的内容。唇读技术源于听力弱者或者听力障碍者学习、了解正常人的表达的一种技巧,它亦可用于特定场合的信息获取(如情报等)。[1]所有人都具有一定的唇读能力，但只有部分人具有唇读技术。

近期研究表明，听力损失本身并不是具有较强唇读能力的充分条件，唇读能力也并不和听力损失程度有确定联系，而与早发性听觉受损(Early-Onset Hearing Impairment)，在听力障碍条件下使用有声语言进行交流的经历以及阅读能力等因素有显著联系。[2]本实验对被试的选择尽量满足以上条件。

前人实验证明在使用句子作为测试材料的实验中，早发性听力受损的唇读者相对于听人显示出显著的优势[2]。

若以单音节为测试材料，这种优势是否依然存在呢？本实验表明，具有唇读技术的聋人（本文称为“唇读者”）和不具有唇读技术的听人（本文称为“非唇读者”）在分辨仅声母发音方法不同的普通话单音节时正确率没有明显差异。并且，唇读者也不能分辨汉语普通话音系声母辅音发音方法特征中的大部分。

因此，唇读者对不同层面的语言信息的认知能力是有所不同的。

结合前人实验推论：汉语普通话唇读者通过唇读认知有声语言时可能采用了一种有取舍的认知策略，而非简单还原全部音系信息。

1 实验目的

假设汉语普通话唇读者对有声语言声母辅音发音方法的认知策略有两种可能：A 对声母辅音发音方法作全面认知，按照汉语普通话音系系统进行认知，例如准确分辨出/pa55/,/pha55/,/ma55/；B 对声母辅音发音方法作选择认知，利用语言系统的冗余性补充本层面信息的不足。例如不辨别/pa55/,/pha55/,/ma55/的唇形，而在语境中处理为不同的语素。

本实验对上述问题做出检验。测试唇读者能否分辨某个特征的最小对立对，若能够分辨，则认为唇读者对本特征采取了全面认知策略，若不能分辨则认为唇读者对本特征采取了选择认知策略。例如，如果唇读者能够分辨/pa55/,/pha55/，则认为在/-a55/条件下唇读者对双唇辅音的送气与否采取全面认知策略，反之则认为唇读者采取选择认知策略。

实验设计没有考虑协同发音因素，排除掉唇读者

借助双音节或多音节间的协同发音区别出无法从单音节中分辨出来的特征的情况，本实验的结果对检验上述两种认知策略才是有效的。

2 实验设计

播放录像，被试从两个选项里选出所看到的“字”，必须选择一个。每个对立对设置两道题。例如：

对立对/pa55/-/pha55/

题号	录像	0	选项 A	1	选项 B
0	ba	ba	八	pa	趴
1	pa	ba	八	pa	趴

在实验中每个被试都会看到/pa55/和/pha55/（表格中用汉语拼音，下同）的两段无声录像（两段录像的播放顺序在所有题目中是随机的）。记录下每个被试完成本题的正误。

考察的有以下几组对立对：1.送气不送气（展声门），如/pa55/-/pha55/，2.浊音非浊音（浊音），如/ʂu55/-/ʐu35/（实际上是同部位的擦音和通音），3.鼻音-塞音/边音（鼻音），如/pa55/-/ma55/,/la55/-/na35/，5. 塞音/塞擦音-擦音（连续），如/ka55/-/xa55/,/tɕi55/-/ɕi55/。（括弧里是对立对体现的特征）。共 51 对 102 道题。

被试两组 A 唇读者组，聋人，具有用唇读理解语言的能力，裸眼或矫正视力正常，33 人，7 人为大学生，其余为初中生，大部分是北方人。B 非唇读者组，对比组，听力正常，裸眼或矫正视力正常，33 人，大学本科及以上学历。聋人组初中生来自以口语教学为主的北京市一聋，日常以唇读和手势汉语结合进行交流，7 名大学生亦受过多年语言康复训练。两组被试教育程度有一定差异，可能对实验结果有一定影响。

录像由北京大学语言学实验室录制，发音人金子，图像格式：avi，504*382，25 帧/秒，386kbps，24 位，压缩 FFDS。

3 检验及其意义

3.1 二组检验

就每一道题对唇读者组数据和非唇读者组数据进行二组检验，观察两组之间是否有显著差异。

3.2 假设检验

本检验统计出唇读者完成某道题的正确率，并由样本检验总体正确率是不是 50%。

如果是，则意味着这道题的选择是随机的，反之这道题的选择不是随机的。在选择随机的情况下，唯一的解释是被试不具有分辨两个选项的能力。在选择不是随机的情况下，正确率可能大于 50%也可能小于 50%，正确率小于 50%时，也只能解释为被试不具有分辨两个选项的能力，至于选择为什么不是随机的，有待进一步探讨，正确率大于 50%时并不意味着被试必然能够分辨选项，因为如果一对题里只要有一道正确率等于或小于 50%，就不能说被试可以分辨这对区别特征。

1.如果一对题里的两道的选择都是随机的，那么被试不能分辨这个最小对立对，他们对这个对立对的认知是选择认知。

2.如果一对题里两道题都不是随机作出选择的，并且总体正确率都大于 50%，那么可以认为被试可以分辨这个对立对。

3.如果一对题里两道题都不是随机选择，只要其中一道存在总体正确率等于或低于 50%的情况，就不能认为被试可以分辨这个对立对。其中有两对题出现两道题正确率都低于 50%的情况。这两对题都属于由于音系的原因，找不到最小对立对而使用声调不同的音节的情况。例如/ta55/-/na35/，实验结果意味着看到/ta55/时，被试大多以为是/na35/，看到/na35/时又大多以为是/ta55/。这是很奇怪的，是否与声调因素影响有关，尚待考察。需要提到的是，听人组并没有出现这种双低的情况。

3.3 配对检验

每一个被试都完成每一对题中的两道题，这两道题测试的是被试对同一个对立对的分辨能力，配对检验研究的因素是被试受到对立对的哪一项的刺激。对唇读组的这个影响因素作配对检验。如果结果显示被试对这对的两道题反应不同的话，就可以认为使用对立对的哪一项进行刺激对被试分辨对立对有影响，也就是说面对对立对，被试选择是有一定倾向的。结合前面的假设检验结果讨论.

1. 刺激项对被试分辨对立对无影响。绝大

多数情况属于此类，被试没有倾向。

2. 刺激项对被试分辨对立对有影响，两道题都不是随机选择，一道高于 50%，另一道低于 50%，这时候我们可以说，被试倾向于总体正确率高的那个选项。

3. 刺激项对被试分辨对立对有影响，其中一道题随机选择，另一道高于 50%。这也暂时解释为被试倾向于总体正确率高的那个选项。

4. 刺激项对被试分辨对立对有影响，但两道总体正确率都为 50%。这也暂时解释为被试倾向于总体正确率高的那个选项。

5.刺激项对被试分辨对立对有影响，一道总体正确率为 50%，另一道总体正确率小于 50%。这也暂时解释为被试倾向于总体正确率高的那个选项。

6. 刺激项对被试分辨对立对有影响，两道题的总体正确率均高于 50%。认为被试可以分辨对立对，但有偏好。

配对检验的结果有一些地方难以解释。

例如，

题号	录像	0	选项 A	1	选项 B	lipreader correct rate	lipreader 组 t 检验 sig 值	配对检验
20	zha	zha	渣	cha	插	0.758	0.002	0.006
21	cha	zha	渣	cha	插	0.455	0.609	

如果被试倾向于/tʂa55/，为什么被试在看到/tʂha55/时没有特别倾向于选/tʂa55/呢？实验中题目顺序打乱，但每个被试看到的题目顺序是一致的，被试先看到/tʂha55/，过一会儿又看到/tʂa55/。一种猜想是，被试在没有看过整个对立对之前不能分辨/tʂha55/，作随机选择，但过一会儿当他看到/tʂa55/的录像时认识到相对于前面看到的/tʂha55/这个录像更接近于/tʂa55/，他就选对了。这有可能是被试选择的倾向受到题目出现顺序的影响。这点猜想还需要进一步验证。验证结果对这个实验的讨论有影响，因为如果猜想成立，就不能说每个人完成每道题目都是独立事件了。另外，这也提醒我们唇读者对说话者的唇形有个熟悉过程，对某个发音人的唇形分辨能力有可能是可以提高的，这也需要验证。

简单验证如下：

对所有 51 对题作列联表。x 为测试顺序先后；y 为题对正确率高低；n（对）

y \ x	先-后	后-先	边缘和
高-低	7	14	21
低-高	18	7	25
等	3	2	5
边缘和	28	23	51

（1）先-后—高-低；后-先—低-高

先出现的题目正确率高，后出现的录像正确率低。7+7=14（对）

（2）先-后—低-高；后-先—高-低。

先出现的题目正确率低，后出现的题目正确率高。14+18=32（对）

（3）原词表按 list1 顺序排列，排列规则为[+特征]在[-特征]后，list2 记录题目出现顺序。因此，先-后意味着具有某特征的录像题后出现，后-先意味着具有某特征的录像题先出现

由此可见，先出现的题目正确率低，后出现的题目正确率高的情况具有明显优势（32> 14），这种优势在具有某特征的录像题后出现的情况下更为明显（18>14）。

这也就是说，一道题目在一对题目中出现的先后会影响其正确率，出现选择的倾向性的原因很可能是题目对被试的训练效应。这一方面提醒我们要在进一步的实验中应当尽量排除题目出现先后的影响，比如对每个被试使用不同的随机题序进行测试；另一方面也提示我们需要另行设计实验验证对发音人的熟悉程度对唇读分辨的影响。

4 统计结果

二组检验。在 102 道题中，总共有 24 道题唇读者和非唇读者的有显著差异，其中 17 道都伴随唇读者组总体正确率偏离 50%或对一对题反应不同的情况。（唇读者和非唇读者 102 道题总平均正确率分别为 0.534 和 0.537）

假设检验。唇读组大部分由样本检验总体正确率为 50%，即选择为随机选择，/la55/-/na35/,/li51/-/ni35/,/ka55/-/xa55/ 三对两道题正确率高于 50%，其余正确率不为 50%的题对均存在其中一道题正确率等于或低于 50%的情况。

配对检验，大部分无倾向性，不少题目出现受刺激因素影响的倾向性。

具体如下 ：

唇读者组和非唇读者组正确率不同的单题（录像对应的音节在前）/pu51/-/phu55/，/ta55/-/tha55/，/tsa35/-/tsha55/，/zɻ51/-/ʂɻ55/，/mi55/-/pi55/，/pu51/-/mu51/，/ma55/-/pha55/，/ta55/-/na35/，/na35/-/ta55/，/ti55/-/ni35/，/tha55/-/na35/，/thi55/-/ni35/，/la55/-/na35/，，/na35/-/la55/，/ni35/-/li51/，/lu35/-/nu35/，/ku55/-/xu55/，/xu55/-/ku55/，/tɕi55/-/ɕi55/，/tsɻ55/-/sɻ55/，/tsha55/-/sa55/，/sɻ55/-/tshɥ/

唇读者一道正确率大于 50%另一道小于50%的对立对 /ta55/-/tha55/，/tsɻ55/-/tshɥ55/，/khu55/-/xu55/。

唇读者两道正确率都大于 50%的对立对/la55/-/na35/，/li51/-/ni35/，/ka55/-/xa55/

唇读者一道正确率大于 50%另一道等于50%的对立对/pi55/-/phi55/，/pu51/-/phu55/，/ti55/-/thi55/，/tʂa55/-/tʂa55/，/pha55/-/ma55/，/lu35/-/nu35/，/ku55/-/xu55/，/kha55/-/xa55/，/zu55/-/su55/，/tshɥ55/-/sɻ55/。

唇读者一道正确率等于 50%，一道低于 50%的对立对 /tɕi55/-/tɕhi55/，/tʂu55/-/tʂhu55/，/tsa35/-/tsha55/，/pa55/-/ma55/，/pi55/-/mi55/，/thi55/-/ni35/，/tʂɻ55/-/ʂɻ55/。

唇读者两道正确率都小于 50%的对立对/ta55/-/na35/，/ti55/-/ni35/。

配对检验显示刺激项对被试分辨对立对有影响/ta55/-/tha55/，/tʂa55/-/tʂha55/，/tʂɻ55/-/tʂh55/，/tsa35/-/tsha55/，/tsɻ55/-/tshɥ55/，/pa55/-/ma55/，/phu55/-/mu51/，/la35/-/na35/，/lu35/-/nu35/，/ku55/-/xu55/，/kha55/-/xa55/，/khu55/-/xu55/，/tɕi55/-/ɕi55/。

5 结语

通过对统计数据进行检验，可以得到以下几点：

1.根据假设检验结果，汉语普通话唇读者对绝大多数对立对没有分辨能力，可以认为，唇读者对大多数声母辅音发音方法作选择认知，对单音节内辅音的大多数发音方法特征不作区分，可能通过其他层面的信息来弥补这部分信息的损失。

2.根据假设检验结果，汉语普通话唇读者对以下对立对有分辨能力，作全面认知：/la55/-/na35/,/li51/-/ni35/,/ka55/-/xa55/

所谓“全面认知”也是受到很大限制的。对声母辅音发音方法的分辨受到声母发音部位和韵母的条件的限制。比如并非所有连续辅音都能和同部位的非连续辅音分辨开来，而只有/k/能和/x/部分地分辨开来，所谓/k/-/x/能分辨也不是和任何韵母搭配都可以分辨，而只是和/a /搭配可以分辨。/l/-/n/只有在开口呼和齐齿呼条件下可以分辨，合口呼条件下/lu35/-/nu35/一个总体分辨率高于 50%另一个等于 50%，呈现出一种倾向性（或许可以认为这个对立的分辨是要经过对比才能作出的）。此外，由于音系的原因，这里选择的是声调不同的对立对，并不严格满足最小对立原则，所以分辨也有可能是声调不同造成的。

3.假设检验结果显示，尽管汉语普通话唇读者对绝大多数对立对没有分辨能力，但唇读者对一些对立对并没有作作随机选择，分辨率有高-中、低-中、低-低三中模式，暗示他们选择时具有一定的倾向性。不过造成倾向的原因有可能不全是音系上的，而是一对题目中先出现的对被试产生了训练效应。

4.配对检验显示 51 对 102 道题中有 13 对 26 道出现刺激项对被试分辨对立对有影响的情况，虽然受到题目测试先后的干扰，但唇读者对对立对的倾向还是有音系上的规律的，因此也不能排除唇读者面对对立对进行选择时是有偏好的可能性。这需要在进一步实验中排除题目测试先后因素干扰，再进行确认。

音系上的规律如下：

从一道题正确率高于 50%，另一道低于 50%的对立对可以看出对于塞音唇读者倾向于不送气（非展声门）的对立项，不连续的对立项：/ta55/>/tha55/, /tsɻ55/>/tshɻ55/, /ku55/>/xu55/

其他情况包括一道题随机选择，另一道高于

50%；两道总体正确率都为 50%；一道总体正确率为 50%，另一道总体正确率小于 50%，从中可以看出对于塞擦音唇读者倾向于送气（展声门）的对立项和连续的对立项，倾向非鼻音对立项：/tʂa55/>/tʂha55/，/tʂhɿ55/>/tʂɿ55/，/tsha55/>/tsa35/，/tsɿ55/>/tshɿ55/，/pa55/>/ma/，/phu55/>/mu51/，/lu35/>/nu35/，/ku55/>/xu55/，/ɕi55/>/tɕi55/。(/tsɿ55/>/tshɿ55/跟测试字使用较冷僻的“疵”有关）

3.二组检验的主要目的是观察在单音节层面进行分辨时唇读者和非唇读者有无明显差异，检验结果显示，虽然在具体题目的正确率上时有差异，但双方都不能分辨绝大部分对立对是一致的，唇读者并没有比非唇读者显示出特殊的优势，那么他们的唇读能力应该在单音节层面之外。此外，分别求两组被试所有题目的平均正确率标准差，唇读者标准差较低（3.520 >2.691），说明唇读者的正确率比较稳定。

参考文献

[1] 姚鸿勋、高文、王瑞、郎咸波 2001《视觉语言——唇读综述》《电子学报》2001 Vol 29. No2.

[2] Edward T Auer Jr, Lynne E Bernstein. (2007). Enhanced Visual Speech Perception in Individuals With Early-Onset Hearing Impairment. Journal of Speech, Language, and Hearing Research, 50(5), 1157-65.

通什黎语声调的实验研究

An Experimental Study on the tones of Tongzha Li Language

钱一凡

Qian Yifan

摘要：黎族是我国岭南的少数民族之一，黎语的声调一直是学界争论的问题之一。通什黎语属于黎语杞方言，共有 9 个单字调，其中舒声调 6 个，促声调 3 个。本文用实验的方法对其单音节声调和双音节组合调进行了考察。

Abstract: Li minority is one of the Lingnan minorities of China. The tones of the Li language is a sustained argument in the academic circle of ethnic languages. Tongzha dialect is a part of Qi dialect of Li language. There are 9 monosyllabic tones, in which there are 6 long tones and 3 short tones. We examined the monosyllabic tones and disyllabic tones of Tongzha dialect through experimental means.

关键词：黎语 声调

Key words: Li language, tones

1 黎族和黎语

黎族是我国岭南的少数民族之一，主要生活在海南省中部以五指山为中心的乐东、保亭、白沙、东方、昌江、崖县、陵水、琼中八个县内，共约 120 多万人。黎语属汉藏语系壮侗语族黎语支，与黎语同属壮侗语族的其他语言还有壮傣（台）语支的状语、布依语、泰语、临高话和侗水语支的侗语、水语、仫佬语、拉珈语和莫话等。黎语共分五个方言：侾、杞、美孚、本地、加茂。其中加茂方言与其他四个方言差别较大。

2 研究简介

通什位于海南黎族苗族自治州的中部，通什黎语属于黎语的杞方言。黎族与汉族分化时间较早，黎语的早期汉语借词比较少，与汉语调类的对应关系不明显，因此，黎语的调类一直是学界争论的一个问题。欧阳觉亚（1983）对通什黎语调类的划分如下：

序号	调值	例字	词义
1	33	ɬa^{33}	鱼
		-ʔa:u^{33}	人
2	121	ta^{121}	田
		-ɬiaŋ^{121}	手指
3	55	ka^{55}	柴刀
4	11	pa^{11}	五
		pei^{11}	婶母
5	51	tha^{51}	饭
		n̥u:n^{51}	推
6	14	ka^{14}	马
7	55	khat^{5}	鼻子
		tat^{5}	鸟
8	13	fa:t^{13}	穷
9	43	li:p^{53}	眨眼
		ʔo:ʔ^{53}	喝（水）

欧阳觉亚将通什黎语分成 6 个舒声调和 3 个促声调，其中，1-6 调是舒声调，7-9 调是促声调。

欧德里古尔（1984）对通什黎语的各舒声调进行了研究，并用村话与黎语各方言的声调进行对比的方法分析了黎语声调的古今分化。在对各舒声调的描述中，除了将欧阳觉亚（1980）的第二调描写为 131 之外，其余各调的调值描写均与前者无异。

罗美珍（1986）使用与黎语同语族的傣语跟黎语的通什话和保定话进行对比，认为黎语的声调是在不同的层次上产生的。1、3、7、4（罗称为 1’调）调有同源词证明是与侗-台语相关的，反映的是发生学上的“亲缘”关系；5、2、6、8、9 各调没有更多的同源词证明与侗-台语相关，反映的是后来发展或影响的结果。罗认为不能仅用数字来表示调类，容易混淆两种不同层次的关系。

3 实验方法

3.1 词表

本文沿用欧阳觉亚（1980）对通什黎语调类的划分，以考察每个调的实际调值。选词时，单音节每个声调选词 4-8 个，双音节组合根据选词难度的不同，选择 1-10 个词不等。双音节的组合中，1+9、2+4、2+7、2+8、2+9、4+2、4+7、4+8、5+2、5+5、5+7、5+9、6+2、6+9、7+8、7+9、8+2、8+4、8+5、8+6、8+8、8+9、9+1、9+2、9+3、9+4、9+6、9+7、9+8、9+9 的组合有类无词，有词的组合共 51 个。

3.2 实验设备与实验步骤

本次实验采用 Adobe® Audition™ 1.5 作为录音软件，双通道录制，左通道录制语音信号，有通道录制 EGG 信号，采样率为 16bit，22kHz。

后期分析采用北大语音乐律实验室自编的 Wavefinal 软件和基于 MATLAB 的北大语音乐律平台中的声调分析程序，提取经过时长归一化后的样本基频数据。然后采用 Microsoft Excel 对数据进行整理和分析，得出各调平均的基频数值。

在数值分析之前，我们使用 T 值法将基频值转换为五度值，以使结论更具备语言学意义，关于 T 值法的具体算法，可以参考石锋（1986）的论述。

4 实验结果

4.1 单音节声调

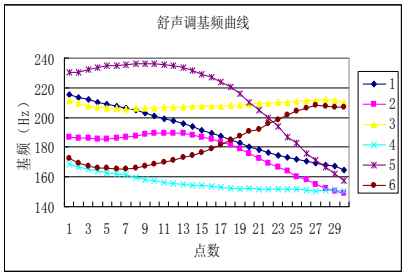


图 1 单音节舒声调基频曲线

图 1 是舒声调 1-6 调的基频曲线，其中横坐标是点数，纵坐标是基频值。由图 1 可得，各调的基频曲线相差很大，其中，1 调是一个降调，2 调是一个中平降调，3 调是一个平调，4 调是一个低降调，5 调是一个高平降调，6 调是一个升调。

我们把舒声调各调的基频值用 T 值法转换为五度值，得到图 2：

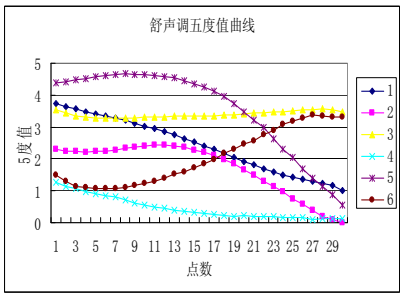


图 2 单音节舒声调五度值曲线

图 2 是舒声调 1-6 调的五度值曲线，其中横坐标是点数，纵坐标是五度值。根据图 2，我们可以构拟单音节舒声各调的五度值，并与欧阳觉亚（1980）的结论相比较如下：

调类	1	2	3	4	5	6
调值	42	331	44	21	551	24
欧阳	33	121	55	11	51	14

我们发现，在舒声调上，我们的结果与欧阳觉亚（1980）的结果有分歧，其中 1、2、4 调分歧较大：1 调欧阳描写为中平，我们的结果是降调，2 调欧阳描写为低升降调，我们的结果是中平降调，4 调欧阳的描写为低平调，我们的结果为低降调。

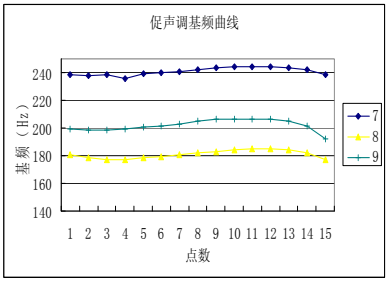


图 3 单音节促声调基频曲线

图 3 是促声调 7-9 调的基频曲线，其中横坐标是点数，纵坐标是基频值。由图 3 可得，3 个促声调基频形状相似，都是平调，不同的是高低位置，其中 7 调最高，8 调最低，9 调居中。

我们把促声调各调的基频值用 T 值法转换为五度值，得到图 4：

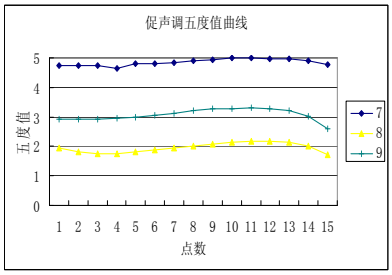


图 4 单音节促声调五度值曲线

图 4 是促声调 7-9 调的五度值曲线，其中横坐标是点数，纵坐标是五度值。根据图 4，我们可以构拟单音节促声调的五度值，并与欧阳觉亚（1980）的结论比较如下：

调类	7	8	9
调值	5	3	4
欧阳	55	13	43

在促声调上，我们的结果与欧阳觉亚（1980）的结果也有分歧，7 调与欧阳的构拟完全一致，8 调欧阳构拟为低升调，我们的结果是中平调，9 调欧阳构拟为降调，我们的结果是平调。

4.2 双音节声调

（1）1 调

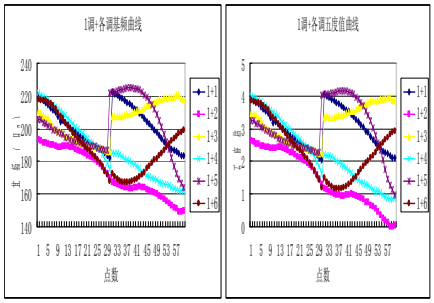


图 5 1 调+舒声各调的基频和五度值曲线

图 5 的左图是首字为 1 调，末字为其余各舒声调的基频值曲线，右图为相应的五度值曲线。由图 5 可得，当双音节末字为舒声各调时，作为首字的 1 调调型统一，可拟为 42 调，与单音节时的调值相同，未发生变调。

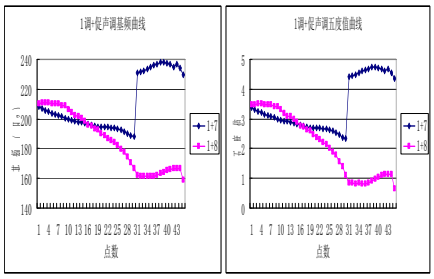


图 6 1 调+促声各调的基频和五度值曲线

图 6 的左图是首字为 1 调，末字为各促声调的基频值曲线，右图为相应的五度值曲线。由图 6 可得，当双音节末字为促声各调时，作为首字的 1 调调型统一，可拟为 42 调，与单音节调值相同，未发生变调。

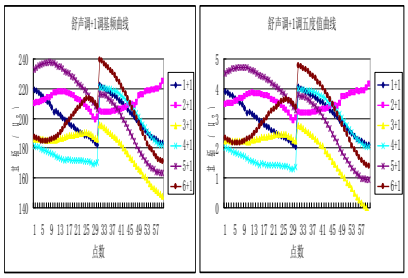


图 7 舒声各调+1 调的基频和五度值曲线

图 7 的左图是首字为舒声各调，末字为 1 调的基频值曲线，右图为相应的五度值曲线。由图 7 可得，当双音节首字为舒声各调时，作为末字的 1 调调型分歧较大，其中，1+1，4+1，5+1 的组合中，作为末字的 1 调调值为 42，2+1 的组合中，1 调的调值为 34，3+1 的组合中，1 调的调值为 31，6+1 的组合中，1 调的调值为 52。

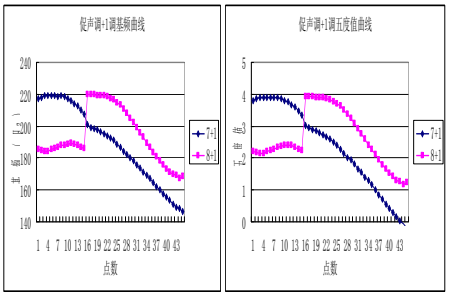


图 8 促声各调+1 调的基频和五度值曲线

图 8 的左图是首字为促声 7 调、8 调，末字为 1 调的基频值曲线，右图为相应的五度值曲线。由图 8 可得，当双音节首字为促声调时，作为末字的 1 调调型有分歧，其中，7+1 组合中，1 调的调值为 31；8+1 组合中，1 调的调值为 42。

（2）2 调

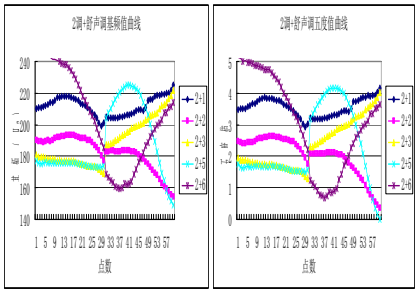


图 9 2 调+舒声各调的基频和五度值曲线

图 9 的左图是首字为 2 调，末字为舒声各调的基频值曲线，右图是相应的五度值曲线。由图 9 可得，当双音节末字为舒声各调时，作为首字的 2 调调型有分歧，其中，2+1 组合中，2 调的调值为 343；2+2 的组合中，2 调的调值为 33；2+3、2+5 的组合中，2 调的调值为 22；2+6 的

组合中，2 调的调值为 53。

由于我们的词表中没有 2 调+促声各调的组合，因此暂时缺少可供分析的数据。

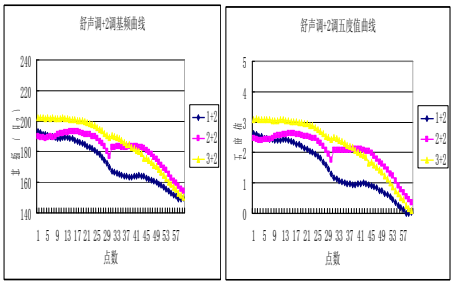


图 10 舒声各调+2 调基频和五度值曲线

图 10 左图是首字为舒声 1、2、3 调，末字为 2 调的基频值曲线，右图是相应的五度值曲线。由图 10 可得，首字为舒声各调时，作为末字的 2 调调型有分歧，其中，1+2 组合中，2 调的调值为 21；2+2、3+2 组合中，2 调的调值为 31。

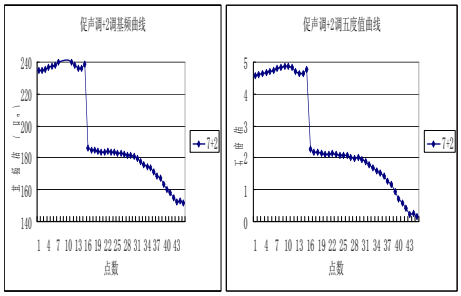


图 11 促声 7 调+2 调基频和五度值曲线

图 11 左图是首字为促声 7 调，末字为 2 调的基频值曲线，右图是相应的五度值曲线。由图 11 可得，首字为 7 调时，作为末字的 2 调调值为 31。

（3）3 调

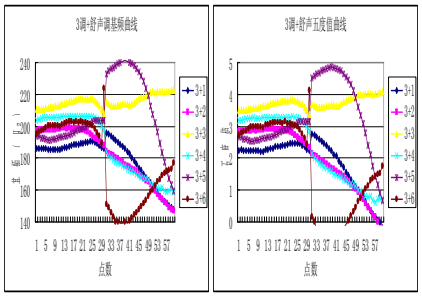


图 12 3 调+舒声各调基频和五度值曲线

图 12 左图是首字为 3 调，末字为舒声各调的基频值曲线，右图是相应的五度值曲线。由图 12 可得，当双音节末字为舒声各调时，作为首字的 3 调调型统一，可构拟为 44 调，与单音节声调相同，未发生变调。

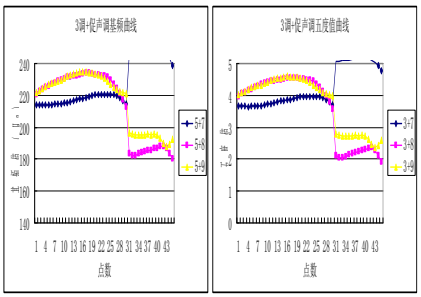


图 13 3 调+促声各调基频和五度值曲线

图 13 左图是首字为 3 调，末字为促声各调的基频曲线，右图是相应的五度值曲线。由图 13 可得，双音节末字为促声各调时，作为首字的 3 调调型统一，可构拟为 44 调，与单音节调值相同，未发生变调。

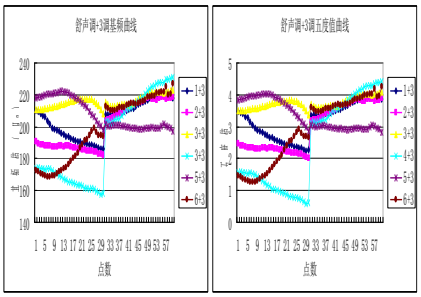


图 14 舒声各调+3 调基频和五度值曲线

图 14 左图是首字为舒声各调，末字为 3 调的基频值曲线，右图是相应的五度值曲线。由图 14 可得，双音节首字为舒声各调时，作为末字的 3 调调型有分歧，其中 1+3、2+3、3+3、4+3、6+3 组合中，3 调的调值为 35；5+3 组合中，3 调的调值为 33。

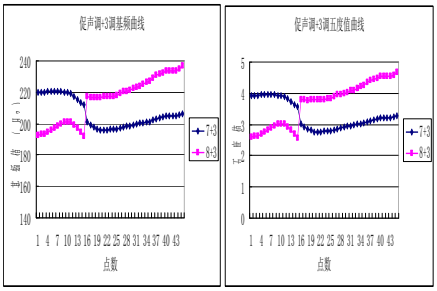


图 15 促声各调+3 调基频和五度值曲线

图 15 左图是首字为促声 7、8 调，末字为 3 调的基频值曲线，右图是相应的五度值曲线。由图 15 可得，双音节首字为促声调时，作为末字的 3 调调型有分歧，其中，7+3 的组合中，3 调的调值为 34，8+3 的组合中，3 调的调值为 45。

（4）4 调

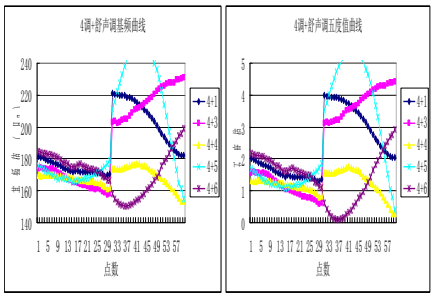


图 16 4 调+舒声各调基频和五度值曲线

图 16 左图是首字为 4 调，末字为舒声各调的基频值曲线，右图是相应的五度值曲线。由图 16 可得，双音节末字为舒声各调时，作为首字的 4 调调型统一，可构拟为 21 调，与单音节时的调值相同，未发生变调。

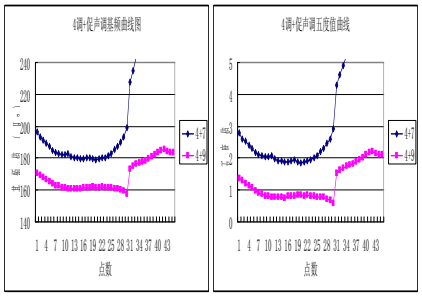


图 17 4 调+促声各调基频和五度值曲线

图 17 左图是首字为 4 调，末字为促声 7、9 调的基频值曲线，右图是相应的五度值曲线。由图 17 可得，双音节末字为促声各调时，作为首字的 4 调调型有分歧，其中，4+7 的组合中，4 调的调值为 32；4+9 的组合中，4 调的调值为 21。

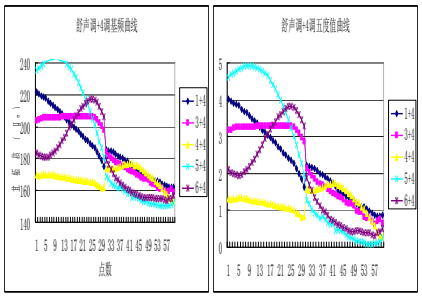


图 18 舒声各调+4 调基频和五度值曲线

图 18 左图是首字为舒声各调，末字为 4 调的基频值曲线，右图是相应的五度值曲线。由图 18 可得，双音节首字为舒声各调时，作为末字的 4 调调型统一，可构拟为 21 调，与单音节时的调值相同，未发生变调。

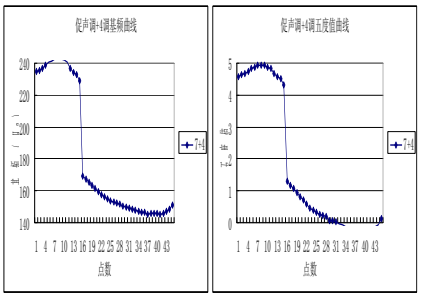


图 19 促声 7 调+4 调基频和五度值曲线

图 19 左图是首字为促声 7 调，末字为 4 调的基频值曲线，右图是相应的五度值曲线。由图 19 可得，双音节首字为 7 调，末字为 4 调时，4 调的调值可构拟为 21，与单音节时的调值相同，未发生变调。

（5）5 调

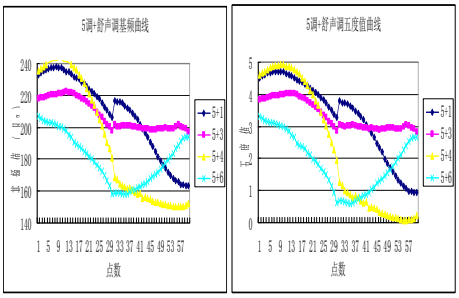


图 20 5 调+舒声各调基频和五度值曲线

图 20 左图是首字为 5 调，末字为舒声各调的基频值曲线，右图是相应的五度值曲线。由图 20 可得，双音节末字为舒声各调时，作为首字的 5 调调型有分歧，其中，5+1、5+3、5+4 的组合中，5 调的调型为 53；5+6 的组合中，5 调的调型为 42。

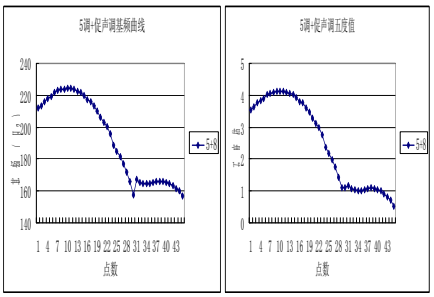


图 21 5 调+促声 8 调基频和五度值曲线

图 21 左图是首字为 5 调，末字为促声 8 调的基频值曲线，右图是相应的五度值曲线。由图 21 可得，当末字为促声 8 调时，作为首字的 5 调调值可构拟为 551，与单音节时的调值相同，未发生变调。

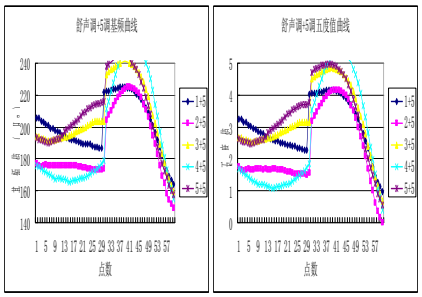


图 22 舒声各调+5 调基频和五度值曲线

图 22 左图是首字为舒声各调，末字为 5 调的基频值曲线，右图是相应的五度值曲线。由图 22 可得，当首字为舒声各调时，作为末字的 5 调调型统一，可构拟为 551 调，与单音节时的调值相同，未发生变调。

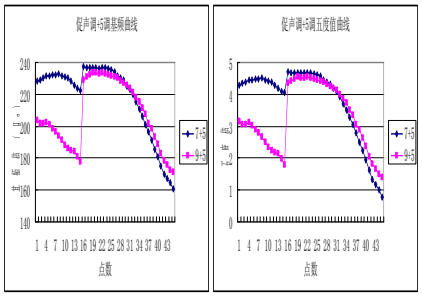


图 23 促声各调+5 调基频和五度值曲线

图 23 左图是首字为促声 7、9 调，末字为 5 调的基频值曲线，右图是相应的五度值曲线。由图 23 可得，当首字为促声 7 调和 9 调时，作为末字的 5 调调型统一，可构拟为 551 调，与单音节时的调值相同，未发生变调。

（6）6 调

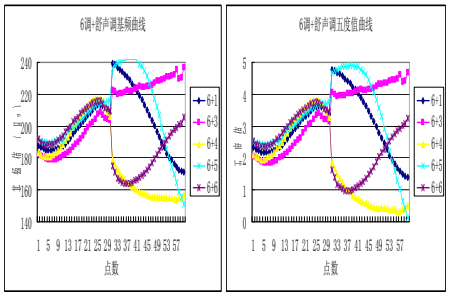


图 24 6 调+舒声各调基频和五度值曲线

图 24 左图是首字为 6 调，末字为舒声各调的基频值曲线，右图是相应的五度值曲线。由图 24 可得，当末字为舒声各调时，作为首字的 6 调调型统一，可构拟为 24 调，与单音节是的调值相同，未发生变调。

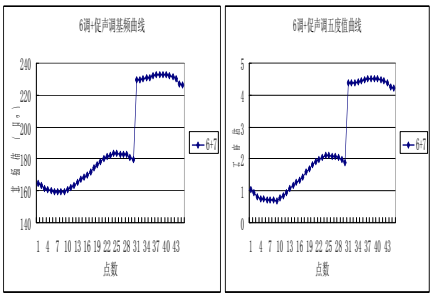


图 25 6 调+促声 7 调基频和五度值曲线

图 25 左图是首字为 6 调，末字为促声 7 调的基频值曲线，右图是相应的五度值曲线。由图 25 可得，当末字为促声 7 调时，作为首字的 6 调可构拟为 13 调。

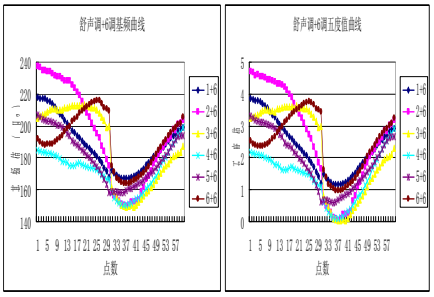


图 26 舒声各调+6 调基频和五度值曲线

图 26 左图是首字为舒声各调，末字为 6 调的基频值曲线，右图是相应的五度值曲线。由图 26 可得，当首字为舒声各调时，作为末字的 6 调调型统一，可构拟为 13 调。

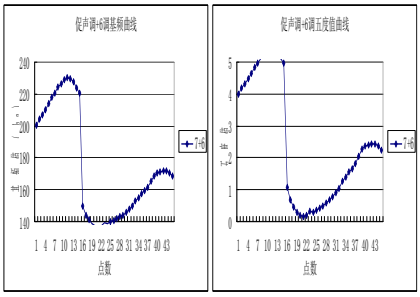


图 27 促声 7 调+6 调基频和五度值曲线

图 27 左图是首字为促声 7 调，末字为 6 调的基频值曲线，右图是相应的五度值曲线。由图 27 可得，当首字为促声 7 调时，作为末字的 6 调可构拟为 13 调。

（7）7 调

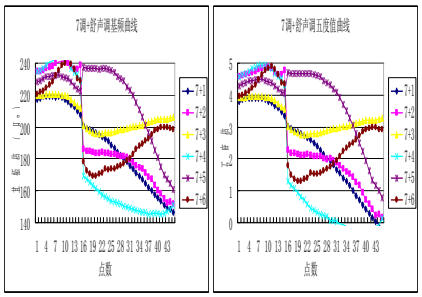


图 28 7 调+舒声各调基频和五度值曲线

图 28 左图是首字为 7 调，末字为舒声各调的基频值曲线，右图是相应的五度值曲线。由图 28 可得，当末字是舒声各调时，作为首字的 7 调调型统一，调值可构拟为 5，与单音节时的调值相同，未发生变调。

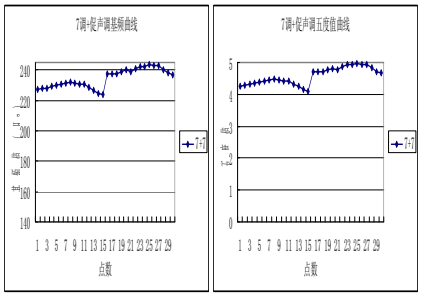


图 29 7 调+促声 7 调基频和五度值曲线

图 29 左图是首字为 7 调，末字为促声 7 调的基频值曲线，右图是相应的五度值曲线。由图 28 可得，当末字为促声 7 调时，作为首字的 7 调调值可构拟为 5，与当音节是的调值相同，未发生变调。

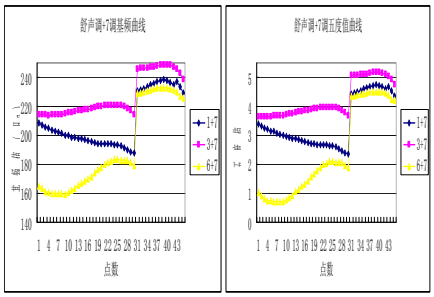


图 30 舒声各调+7 调基频和五度值曲线

图 30 左图是首字为舒声各调，末字为 7 调的基频值曲线，右图是相应的五度值曲线。由图 30 可得，当首字为舒声各调时，作为末字的 7 调调值可构拟为 5，与单音节时的调值相同，未发生变调。

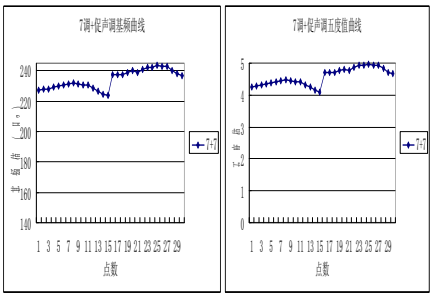


图 31 促声 7 调+7 调基频和五度值曲线

图 31 左图是首字为促声 7 调，末字为 7 调的基频值曲线，右图是相应的五度值曲线。由图 31 可得，当首字为促声 7 调时，作为末字的 7 调调值可构拟为 5，与单音节时的调值相同，未发生变调。

（8）8 调

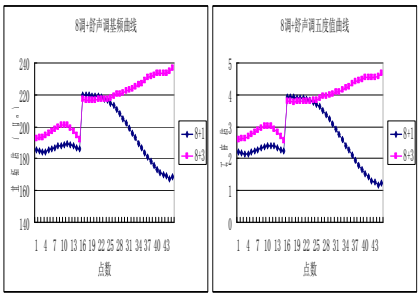


图 32 8 调+舒声各调基频和五度值曲线

图 32 左图是首字为 8 调，末字为舒声 1、3 调的基频值曲线，右图是相应的五度值曲线。由图 32 可得，当末字为舒声 1、3 两调时，作为首字的 8 调调型统一，调值可构拟为 3，与单音节时的调值相同，未发生变调。

由于我们的词表中没有 8 调+促声各调的组合，因此暂时缺乏可供分析的数据。

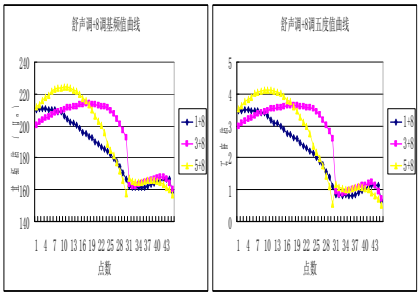


图 33 舒声各调+8 调基频和五度值曲线

图 33 左图是首字为舒声各调，末字为 8 调的基频值曲线，右图是相应的五度值曲线。由图 33 可得，当首字为舒声各调时，作为末字的 8 调调型统一，调值可构拟为 1。

由于我们的词表中没有促声各调+8 调的组

合，因此暂时缺乏可供分析的数据。

（9）9 调

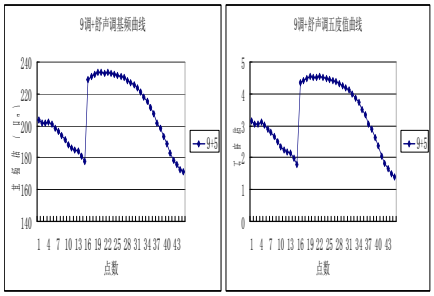


图 34 9 调+舒声 5 调基频和五度值曲线

图 34 左图是首字为 9 调，末字为舒声 5 调的基频值曲线，右图是相应的五度值曲线。由图 34 可得，当末字为 5 调时，作为首字的 9 调调值可构拟为43。

由于我们的此表中没有 9 调+促声各调的组合，因此暂时缺乏可供分析的数据。

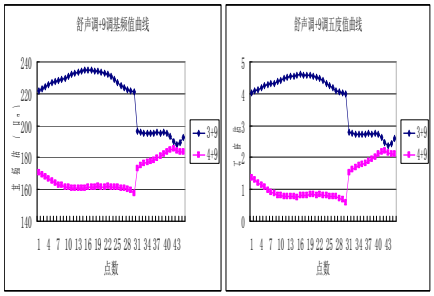


图 35 舒声各调+9 调基频和五度值曲线

图 35 左图是首字为舒声 3、4 调，末字为 9 调的基频值曲线，右图是相应的五度值曲线。由图 35 可得，当首字为舒声 3、4 两调时。作为末字的 9 调调型统一，调值可构拟为 3。

由于我们的词表中没有促声各调+9 调的组合，因此暂时缺乏可供分析的数据。

5 结论

末字 / 首字		1	2	3	4
		42	331	44	21
1	42	42+42	42+21	42+35	42+21
2	331	343+34	33+31	22+35	+21
3	44	44+31	44+31	44+35	44+21
4	21	21+42	21+	21+	21+21
5	551	53+42		53+33	53+21
6	24	24+52		24+35	24+21
7	5	5+31	5+31	5+34	5+21
8	3	3+42		3+45	
9	4				

（续表）

末字 / 首字		5	6	7	8	9
		551	24	5	3	4
1	42	42+551	42+13	42+5	42+1	42+3
2	331	22+551	53+13			
3	44	44+551	44+13	44+5	44+1	44+3
4	21	21+551	21+13	32+5		21+3
5	551		42+13		551+1	
6	24		24+13	13+5		
7	5	5+551	5+13	5+5		
8	3					
9	4	43+551				

上表对通什黎语单音节和双音节声调的调值进行了归纳，其中空格的部分是我们的词表中没有的组合。黎语杞方言通什话的双音节组合调中存在大量的连读变调现象。

参考文献

[1] 中国语言地图集，香港，朗文出版社，1987

[2] 欧阳觉亚，黎语调查研究，北京，中国社会科学出版社，1983

[3] [法]欧德里古尔，海南岛几种语言的声调，《民族语文》1984 年第 4 期

[4] 罗美珍，黎语声调刍议，《民族语文》1986 年第 3 期

仙居方言中的浊内爆音

The Voiced Implosives of Xianju Dialect

宋益丹

Song Yidan

摘要： 本文运用麦克风和电子声门仪对吴语仙居方言的浊内爆音进行声学分析和嗓音特性分析，得出如下结论：（1）典型的浊内爆音开商值小，具有紧喉特征，弱化的浊内爆音紧喉特征消失。（2）浊内爆音的平均开商值与时长之间存在强负相关关系。（3）浊内爆音的平均开商值与平均基频值之间存在弱负相关关系。因此，浊内爆音在搭配高调时仍可能形成较低的起始基频。（4）仙居方言的浊内爆音正处在消失的过程中，演变方向为同部位的清塞音 p、t。这可以从自然音变的角度进行解释。

Abstract： This paper analyzes the acoustic and phonatory features of the voiced implosives of Xianju dialect affiliated to the Wu Dialect in China. Some conclusions are drawn. Firstly, the typical voiced implosives with small open quotient (OQ) are considered to be laryngealized whereas the reduced voiced implosives have no laryngeal features. Secondly, the averaged OQ of the voiced implosives has a strong negative correlation with the duration of the implosive. Thirdly, the averaged OQ of the voiced implosives is weakly negatively correlated with the averaged fundamental frequency. Thus low initial fundamental frequency comes into being when the implosives is specified with a high tone. Fourthly, the voiced implosives of Xianju dialect are in the process of on-going sound change, neutralizing with the voiceless plosive /p/ and /t/ conditioned by the places of articulation, which can be explained away by the theory of natural sound change.

关键词： 仙居方言 浊内爆音 开商 速度商

Key words： Xianju dialect, voiced implosive, open quotient, speed quotient

0 引言

据 Maddieson（1984）[1]统计，世界语言中有 10.1%存在内爆音。中国的侗台语和汉语东南方言（吴语、闽语、粤语等）有浊内爆音的分布。[2]吴语中浊内爆音的分布主要集中在南部吴语，且绝大部分都属山区。此外，上海周边各县也有分布。[3]

浊内爆音的判定可以从生理信号和声学信号两个方面进行。生理特征表现在浊音爆破时口腔形成负气压，这可以通过气流气压计测得。声学特征表现为成阻段的声波振幅渐强，而普通浊爆音成阻段的振幅是渐弱的。[4]

对于浊内爆音的嗓音情况，目前尚未见实验报导。传统看法是喉塞或紧喉。国内学术界曾用“先/前喉塞浊音”（李方桂，1943；陈忠敏，1995）、紧喉浊塞音（郑张尚芳，1988）等术语来描写浊内爆音。

本文通过分析声学信号和声门阻抗信号（EGG），对仙居方言浊内爆音的声学特征和嗓音特性进行研究，以期用科学的方法来描写浊内爆音的喉部机制。

仙居县位于浙江省东南部，属于吴语台州片。据游汝杰（2002）调查，有内爆音 ɓ 和 ɗ（原文用先喉塞 ʔb、ʔd 描写），分别来源于中古的帮母和端母。[5]

1 研究方法

1.1 实验仪器

本文语音材料于 2007 年 11 月在浙江省仙居县录制，录音软件为 Adobe Audition。实验仪器包括指向性麦克风、电子声门仪（Electroglottograph）、外置声卡、调音台、手提电脑。录音时，同时采集两路信号：左通道为声音信号，右通道为声门阻抗（EGG）信号。采样频率 22050Hz。

1.2 发音人

发音人有 4 位，两男（M1、M2）两女（F1、F2），年龄在 59-64 岁之间，均为仙居城关人。

1.3 实验词表

仙居方言共 8 个调类，浊内爆音出现在 4 个阴调类。本文只选用阴去和阴入调的字，每个调帮母和端母各选 3 个，得到 12 个单音节词。见表 1。每个词读两遍，共得到 12×2×4=96 个样本。

表 1：实验材料

	阴去	阴入
帮母	拜/a/	百/aʔ/
	闭/i/	笔/iʔ/
	布/u/	北/oʔ/
端母	带/a/	搭/aʔ/
	蒂/i/	跌/iʔ/
	妒/u/	督/oʔ/

1.4 分析方法

（1）在 Praat 中观察声波图和宽带语图，并对声母时长进行标注和计算。

（2）用 Kay 公司提供的配套软件 Real-Time EGG Analysis 提取每个音节的基频、开商、速度商。

（3）用 Matlab 编写程序进行数据归一化处理。用线性插值的办法，对声母前浊段提取 10 个点，元音段提取 20 个点。

（4）用 spss 和 Excel 进行统计分析与作图。

2 实验分析及结果

2.1 语图表现

仙居方言浊内爆音的声学表现与以往研究一致。从声波图上看，浊内爆音的声波振幅随时间增大，到爆破时达到最大值。从宽带语图上看，一般有较长的浊横杠，浊横杠和元音共振峰之间有的可见冲直条。见图 1。

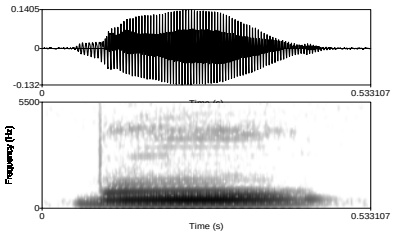


图 1：布/ɓu/的声波图和宽带图

并不是所有的实验词都被发成浊内爆音，有部分词读成了清塞音 p 和 t。四位发音人中，女声发浊内爆音的较多，F1 有 17 例浊内爆音，F2 所有样本均为浊内爆音；男声 M1 只有 2 例浊内爆音，M2 有 12 例。这种人际差别反映了浊内爆音正处在消失的过程中。据了解，仙居一些年轻人已全部读成清塞音了。

2.2 时长

我们统计了四位发音人的浊内爆音 ɓ、ɗ 在阴去（舒声调）、阴入（促声调）中的前浊段时长。表 2 中，5 代表所在音节为阴去调，7 代表所在音节为阴入调。n 为样本数，m 为时长均值，单位毫秒。

表 2：浊音段时长

		ɓ5	ɓ7	ɗ5	ɗ7
F1	n	5	6	0	6
	m	18	76		60
F2	n	6	6	6	6
	m	78	103	125	100
M1	n	0	2	0	0
	m		28		
M2	n	2	5	2	3
	m	37	62	44	53

不同发音人之间、同一个发音人的不同调类之间，前浊段时长差异很大。时长的这种离散性同样反映了浊内爆音的不稳定状态。

整体而言，促声调中的前浊段时长比舒声调中的更长（F2 的 ɗ 声母例外）。阴去调中，ɗ 长于 ɓ；阴入调中，ɓ 长于 ɗ。

2.3 基频

由于开商、速度商和基频之间存在相对关系，所以研究嗓音特性需要分析浊内爆音所在音节的基频变化模式。

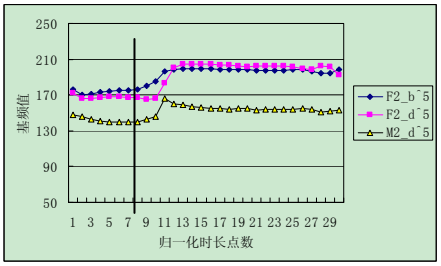


图 2：阴去调基频曲线

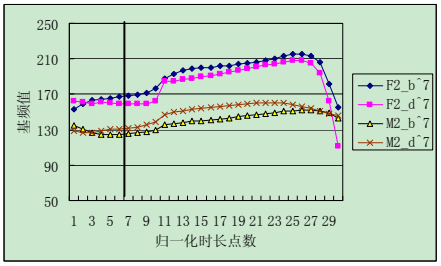


图 3：阴入调基频曲线

图 2 和图 3 分别给出了 F2、M2 两位发音人在阴去调、阴入调中的基频走势。M2 阴去调的 ɗ 起始周期能量较弱，基频提取有问题，故不在统计之列。图中，bˆ代表 ɓ，dˆ代表 ɗ，图中有一垂直标线，标线前为前浊段，标线后为元音段。下文同。

对照两图，最显著的特征是浊内爆音的基频要显著低于后接元音，基频走势平坦或微升。此外，还有如下规律：（1）两发音人的基频走势一致，女声基频高于男声。（2）同一调类中，6、d 所在音节的基频曲线走势一致。（3）两调类调型段的基频值接近，基频曲线走势略有不同，阴去平坦，阴入略升，后有喉塞引起的降尾。

阴去和阴入在仙居方言声调系统中都是高调，可以避免低调造成的发声类型的改变从而影响实验结果。前浊段和元音段基频走势都比较平坦，可以忽略基频变化对开商和速度商的影响。这也是我们选择这两个调类进行实验的原因。

2.4 开商

开商的定义为声门的开相（open phase）和周期之比。在语言学中，开商用于描述声门的松紧度。开商值越大表明声门的打开度越大。正常嗓音的平均开商值为 55。紧喉音的平均开商值为 48。[6]p181-182

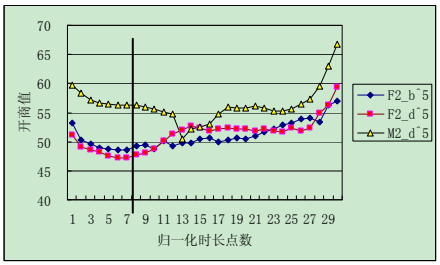


图 4：阴去调开商曲线

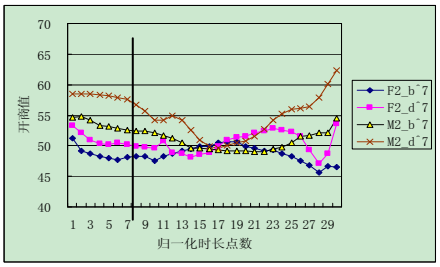


图 5：阴入调开商曲线

从图 4 和图 5 可知， F2、M2 的起始开商值都在 50 以上，属于正常嗓音的范围。因为浊音的起始需要一个相对大的开商值。在前浊段，开商值随时间呈递减趋势，表明声门打开度越来越小。这是由于喉头下压，拉紧声带造成的。

表 3 给出了两位发音人前浊段的平均开商。F2 的平均开商值均小于正常嗓音，而和紧喉嗓音接近，部分点的开商达到紧喉嗓音的范围。M2 的开商值普遍大于 F2，全部属于正常嗓音的范围。可见浊内爆音的喉部状态是比较灵活的，从紧喉到正常嗓音都存在。紧喉特征并不是浊内爆音的必要条件。

表 3 前浊段平均开商值

	6̃5	6̃7	d̃5	d̃7
F2 平均开商	49.51	48.66	48.32	50.18
M2 平均开商		53.27	56.88	57.42

为了解前浊段时长和平均开商值之间的关系，我们在 SPSS 中做了相关性分析。表 4 结果显示，两者的皮尔逊系数为-0.553，概率为 0.002，表示在 0.01 水平上，两者呈强负相关关系。即时长和开商值之间密切关联，时长越长，平均开商值就越小。

表 4：前浊段时长与开商相关系数表

Correlations

		开商	时长
开商	Pearson Correlation	1	-.553**
	Sig. (2-tailed)	.	.002
	N	29	29
时长	Pearson Correlation	-.553**	1
	Sig. (2-tailed)	.002	.
	N	29	29

**. Correlation is significant at the 0.01 level

由此，我们可以归纳浊内爆音的典型状态是成阻段时长较长，开商随时间逐渐减小，达到紧喉状态。浊内爆音的弱化形式为成阻段时长变短，开商随时间渐小，但无紧喉特征。两位发音人中，F2 的浊内爆音比较典型，而 M2 则体现出弱化特征。

表5给出了前浊段平均开商值与平均基频值之间的相关关系。两者的皮尔逊系数为-0.405，概率为0.029，表示在0.05水平上，两者呈弱负相关关系。即开商值越小，基频越高。这就可以说明为什么浊内爆音出现在阴调类而不是阳调类中。发浊内爆音时，喉头的下压动作造成声带拉伸，喉部肌肉紧张，开商值变小，基频提升。

表 5：前浊段基频与开商相关系数表

Correlations

		基频	开商
基频	Pearson Correlation	1	-.405*
	Sig. (2-tailed)	.	.029
	N	29	29
开商	Pearson Correlation	-.405*	1
	Sig. (2-tailed)	.029	.
	N	29	29

*. Correlation is significant at the 0.05 level (2-tailed).

2.5 速度商

速度商的定义为声门的开启相（opening phase）和关闭相之比。通过速度商的测量，我们可以了解声门开启和关闭的速度。

通过研究我们发现，浊内爆音速度商的变化比较复杂，暂时还没有找到固定的速度商变化模式。

3 总结

（1）典型的浊内爆音开商值小，具有紧喉特征，弱化的浊内爆音紧喉特征消失。

（2）浊内爆音的平均开商值与时长之间存在强负相关关系。时长越长，开商值越小，紧喉特征越明显。

（3）浊内爆音的平均开商值与平均基频值之间存在弱负相关关系。因此，虽然发音时需要下压喉头，拉伸声带，但浊内爆音在搭配高调时仍可能形成较低的起始基频。

4 余论

仙居方言的浊内爆音出现了弱化，正处在消失的过程中，演变方向为同部位的清塞音 p、t。这和有些方言（如附近的永康）演变为鼻音或边音的情况不同。

从自然音变的角度解释，浊内爆音弱化时首先成阻段时长变短，由于和高调相配，声母的浊感减弱，同时又受到系统简化的驱动，[+浊]的区别特征消失，演变为同部位的清塞音，和舌根音 k 成为一组，共同构成清塞音序列。

参考文献

[1]Ian Maddieson, Patterns of sounds, Cambridge University Press, 1984.

[2]朱晓农，内爆音，《方言》，2006，1。

[3]郑张尚芳，浙南和上海方言中的紧喉浊塞音声母 ʔb、ʔd 初探，《吴语论丛》，1988。

[4]Peter Ladefoged and Ian Maddieson, The Sounds of the World’s Languages, Blackwell Publishers,1996.

[5]游汝杰，吴语声调的实验研究，复旦大学出版社，2002。

[6]孔江平，论语言发声，中央民族大学出版社，2001