语音乐律研究报告

Status Report of Phonetic and Music Research

2008

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国语言文学系语言学实验室年度报告的第三期，收录了实验室 2008 年发表和未发表的论文 6 篇，以及对实验室现有实验仪器的说明报告 9 篇。2008 年，实验室主要进行了如下活动：

一、　科研进展

实验室的研究主要包括以下几个方面：基础理论和方法、普通语音学、生理语音学、民族语音学、病理语音学等，详见本论文集内容。

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 Berkeley、JAIST、CNRS、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。

2008 年，实验室邀请了法国国家科学院语音学和音系学实验室（CNRS Laboratoire de Phonétique et Phonologie）、巴黎第三大学（L'Université Sorbonne Nouvelle - Paris 3）教授杰奎琳 • 维西尔（Jacqueline Vaissière）来访，并做了相关的研究报告。

实验室的大部分人员参加了 2008 年 4 月在北京举行的“第八届中国语音学学术会议暨庆祝吴宗济先生百岁华诞语音科学前沿问题国际研讨会”。

北京大学中国语言文学系语言学实验室
Linguistics Lab, Department of Chinese Language and Literature, Peking University
Tel: 86-01-62753016
Email: ell@pku.edu.cn
Website: http://www.phonetics.ac.cn/

实验室人员		联系方式
孔江平	教授	jpkong@pku.edu.cn / kongjp@gmail.com
李锋	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
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
杨锋	08 级博士生	yangzihuai@163.com
钱一凡	06 级硕士生	yifanqian@gmail.com
杨若晓	06 级硕士生	yangruoxiao@pku.edu.cn
叶泽华	07 级硕士生	zehuay@gmail.com
吴君如	08 级硕士生	yuerr@163.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 10 名。）

本期执行编辑：杨若晓　杨锋

目　录

语言发声研究的基本方法

孔江平

0 引言

言语产生的理论将语音发音（speech production）分为调音（articulation）和发声（phonation）两部分。其中，调音主要是指各部位发音器官的协调运动形成的声道形状，然后共鸣而产生的不同的语音，比如，通常语言里最常见的元音/a/，/i/，/u/等，都是由于不同的声道形状共鸣而产生的不同语音。发声是指声带在气流的作用下，以不同的振动方式而产生的声源，声源主要包括了声带振动的频率，即振动的快慢，以及声带振动的方式，比如，正常嗓音发声，气嗓音发声，挤喉嗓音发声，气泡嗓音发声，假嗓音发声等。

在早期的语音学研究中，人们对语言的调音有较多的认识，如发音部位的定义和发音方法的定义等，根据这些语音学中的基本定义，产生了元音、辅音等基本概念，从而建立起了语音学或者说基于调音的语音学基础理论和方法，这些理论和方法在语言学的研究中起了极为重要的作用。然而，随着语言学、言语声学、言语生理学的发展，人们对语音产生的认识有了很大飞跃。另外，在语言学研究中进行的大量语言田野调查使人们发现了大量具有语言学意义的发声类型，如中国的彝语、哈尼语、景颇语、载瓦语等（孔江平，2001），中国民族语言学中常用的“松元音”和“紧元音”这些语言学概念，就是对不同发声类型的语言学定义，这些都使得语言学家和语音学家逐步认识到了语言发声的重要性。

在语言发声类型的研究中，首先是信号的问题，也就是说研究发声类型要采集什么信号，通常我们研究语音最为重要的信号是声音信号，另外，在早期调音的研究中，X-光声道和腭舌的信号对于语音调音的研究起了非常重要的作用，当然现在有更多的信号，如核磁共振（MRI）、螺旋CT、动态电子腭位信号等。在发声的研究中，可以使用的信号主要有：1）语音信号；2）声门阻抗信号；3）气流气压信号；4）高速数字成像信号等，这些信号可以用不同的信号处理算法来处理，从中提取出有用的可以反映各种语言发声类型的参数，从而解释语言发声类型的性质和揭示语言发声的生理学、物理学和语言学本质。

发声研究的方法很多，有的主要用于嗓音生理学和病理学的研究，有的主要用于语言声学的研究，还有些方法可以用于心理学语音学的研究。本文主要根据语音学的需求，介绍一些能用于嗓音发声类型研究的基本方法。它们包括：1）谐波分析法；2）逆滤波分析法；3）频谱倾斜率分析法；4）多维嗓音分析法；5）声门阻抗分析法；6）动态声门分析法；7）发声起始状态分析法；8）嗓音音域分析法；9）嗓音的合成方法。

1 发声的生理和物理基础

本节简单介绍一下喉头的解剖，喉头主要是由软骨和肌肉组成，主要的软骨有会厌软骨（epiglottis cartilage）、甲状软骨（thyroid cartilage）、环状软骨（cricoid cartilage）和一对杓状软骨（arytenoid cartilages）。

环甲节（cricothyroid joint）连接甲状软骨和环状软骨，环杓节（cricoarytenoid joint）连接杓状软骨和环状软骨，杓状软骨和环状软骨之间的运动改变声带的长度，杓状软骨沿着环杓节纵向运动导致声带的打开和并拢，(Sawashima, 1983; Titze, 1994; Hirose, 1997; Kong, 2007)，见图 1。[1]

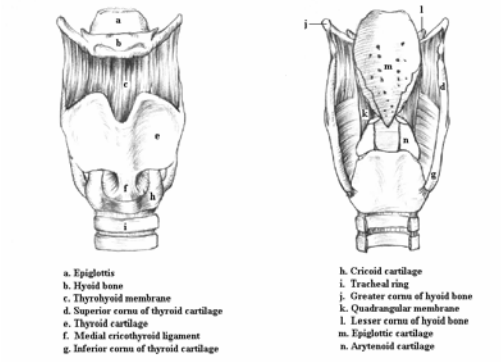


图 1 喉头的整体框架：前视图和后视图

[1] 图片参考了Titze (1994)绘制而成。

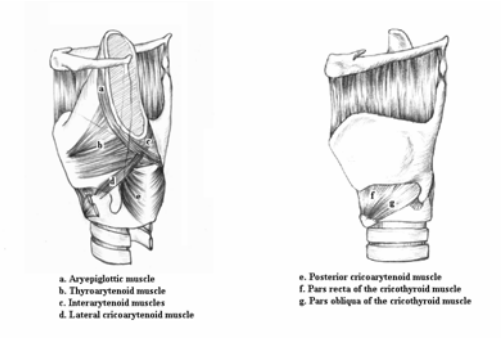


图 2 喉内肌：后侧视图和前侧视图

喉头大约有十几条肌肉，它们主要是喉内肌、环甲肌(CT)、外展肌、内收肌、后环杓肌(PCA)、内杓肌(INT or IA)、侧环杓肌(LCA)、甲杓肌(TA)、发声肌（VOC)和外喉肌。喉内肌由一组肌肉组成，喉外肌则包括舌骨上肌（suprahyoid muscle）和舌骨下肌（infrahyoid muscle），见图 2。

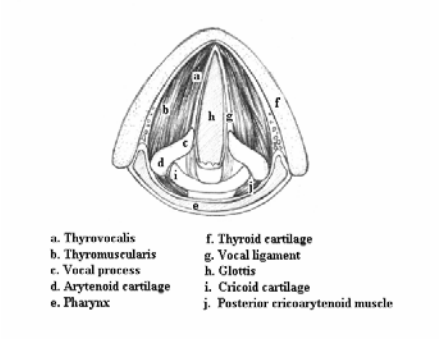


图 3 上视图（声带切面）

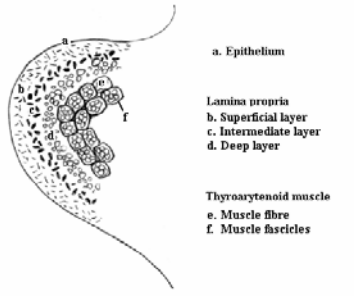


图 4 声带的结构

环甲节和环杓节的运动由喉内肌控制，喉升高和降低的运动由喉外肌控制，如，舌骨上肌和舌骨下肌，这些肌肉支撑着整个喉头，也可能导致音调的变化。杓状软骨的运动由外展肌（PCA）控制和内收肌（INT、LCA 和 TA）控制，这些肌肉导致声带的外展和内收。CT 的收缩导致声带的拉伸，VOC 在某种程度上控制声带的有效质量和紧张度。

声带由软组织构成，包括：1) 上皮细胞（epithelium）；2) 表面层（superficial layer）； 3) 中间过渡层（intermediate layer）；4) 深层（deep layer）；5) 肌肉（muscle）。声带结构也可以分为：1) 粘膜（mucosa），它包括上皮细胞和表面层；2) 韧带（ligament），它包括中间过渡层和深层；3) 肌肉。 声带的结构还可以简单地分成：1) 声带的覆盖层, 包括上皮细胞、表面层和中间过渡层；2) 声带体，它包括深层和肌肉。见图 3、4 中的细节。

发声的生理结构和发声的物理原理是直接相关的，同时，声源的产生涉及到声道的共鸣。在言语产生过程中，由于声源是通过声道发出来的，因此又叠加上了共鸣的特性，这使得研究声源变得很复杂，因为要从声波中提取声源首先就要把声道共鸣的特征去掉，在现代信号处理技术中这仍然是一个没有很好解决的问题和技术上的难点。

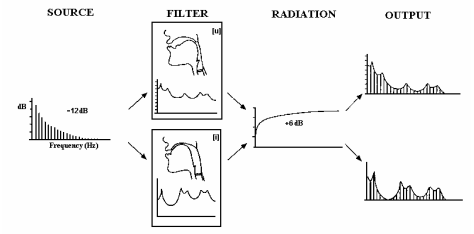


图 5 言语产生的三个部分和过程(Hardcastle et al, 1997)

图 5 是言语产生的一个基本原理的示意图，严格地讲言语产生可以分为三个部分，第一个部分是声源，用线谱图表示，它的基本特性是每个倍频程下降 12dB。第二个部分是共鸣，声源经过声道共鸣到达口部，这个过程改变了声源的共鸣特性形成了不同的元音，如图中口腔侧面图和共振峰包络图所示，一个表示/i/的形成过程，而另一个表示/a/的形成过程，。第三个部分是唇辐射，用曲线图表示，其物理特性是每个倍频程提高 6dB，最后形成语音，用共振峰包络图和线谱图表示。在言语产生的过程中，唇辐射的特性极其重要，需要真正理解。

2 谐波分析法

在语言发声研究中，谐波分析是一种最简单易行的方法，也是最早被语音学研究人员使用的方法，说它简单是因为只要有语言的录音，就可以进行声学分析，而声学分析只要做最简单的功率普分析即可。因此在上个世纪 70 年代，美国 UCLA 的语音学家用此方法分析和研究了许多语言的发声类型，其中就包括中国的一些民族语言，(Ladefoged, 1973, 1988; Laver,1980; Ladefoged et al, 1987a, 1987b; Kirk et al，1984; Therapan, 1987;

Anthony, 1987; Maddieson et al，1985；孔江平，2001）。

谐波分析法在声学原理上主要是根据声源能量的大小，即声源谱的特性，声源谱高频能量强会导致第二以上谐波的能量大于第一谐波的能量，因此可以通过测量第一、二谐波的能量来判断嗓音发声类型的不同，一般是使用第一、二谐波之比的方法。

其优点是简便易行，但这种方法也存在很多缺点，其中最主要的缺点在测量数据时共振峰对谐波能量会有影响，因此，有经验的语音学家在使用此方法时，往往选择元音/a/作为分析的样本，这是因为/a/的第一共振峰比较高，因而对第一、二谐波的能量影响比较小或者没有影响，这样就可以得到比较稳定和有规律的数据。如果使用了/i/、/y/、/ u /作为测试样本，第一共振峰比较低高，基频的能量和第一种振峰的能量会重叠，因而就得不到真实的嗓音数据。

为了解决其不足，研究人员往往会使用第二共振峰的能量和第一或第二谐波能量的比值来判断嗓音的发声类型，这种补救的方法在通常情况下对分析语言的不同发声类型也都会很有效。但是在语言发声类型的研究中，以不同发声类型作为最小对立时，声道的形状不一定完全相同，往往会有一定的差别，这就导致了要研究的嗓音发声的最小对立元音的共振峰不同，从而影响到数据的测量导致数据的误差，在这种情况下往往要考虑其它的研究方法。

3 逆滤波分析法

图 6 是言语产生和逆滤波的原理示意图，分为上中下三张图，上图是言语产生的原理，与图五不同的是，共鸣部分用滤波器来说明，唇辐射和语音输出用线谱图和共振峰包络图表示。中图为逆滤波的原理，一段语音经过逆滤波得到声源，逆滤波是将原共鸣特性反过来设计逆滤波的滤波器，这样就可以将语音中的共鸣去掉，最终得到声源(Alku P. (1991).; Lindestad P-A et al. (1999).; 方特.G、高奋.J，1994，)。

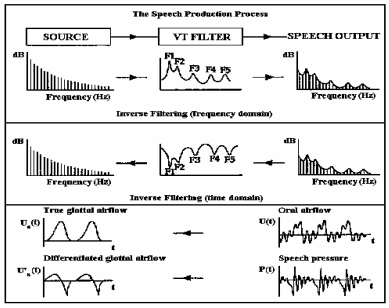


图 6 言语产生和逆滤波原理示意图

下图为声源和语音的关系，这一部分对于理解逆滤波最为重要，我们通过发声的生理研究知道，肺部气流冲破关闭的声带使声带振动产生声源，气流冲开声带形成的空隙称为声门，单位时间内通过这个空隙的气流为声门气流，一般用体积流速度表示。从下图的上半部分可以看到，如果对口腔气流进行逆滤波，得到的是声门气流，口腔气流一般是通过口腔面罩或气流计采集，逆滤波后得到声门气流。通常情况下很难直接采集到声门气流，因为很难将采集器放到声带（声门）的上方。声门气流除了从口腔气流中逆滤波得到以外，实际上还可以从声门面积推算出来。根据目前的技术，采集声门面积要用高速数字成像和图像处理的技术（孔江平，2007）。下图下半部分是从声压（语音）经过逆滤波得到声源，这个声源的形式是声门气流的微分，从中可以看出，口腔气流的微分形式基本就是声压。在发声的研究中，通常很少用口腔气流经逆滤波得到声源信号，一般都是对声压进行逆滤波获取声源。

下面我们来讨论一下从声压逆滤波获取声源的基本方法、过程和存在的问题。要进行逆滤波首先要提取声道共鸣的特性，在现代信号处理技术中一般使用自回归（AR）模型，具体来讲就是线性预测（LPC），

$$s(n)=\sum_{k=1}^{p}a_k s(n-k)+Gu(n)$$

以上公式定义了言语产生的基本时间离散模型，其中 s(n)为言语信号， u(n)为激励源， G 为增益， k{ak}为滤波器系数，k 为信号的延时。

从上面的公式可以看出，逆滤波首先是对一段语音进行线性预测，计算出线性预测系数，我们知道线性预测系数代表的就是共鸣特性，那么将得到的系数直接用于逆滤波就能得到声源。

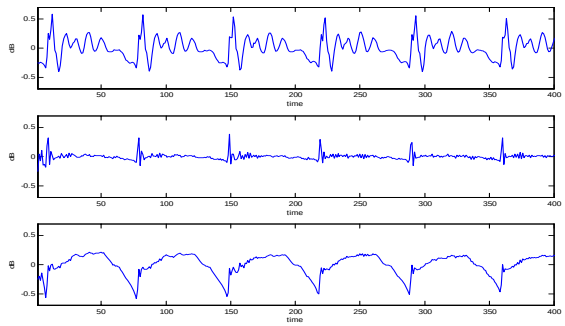


图 7 两种不同的逆滤波方法

图 7 显示了两种不同的线性预测逆滤波方法，图上为语音波形/a/，图中为直接使用线性预测系数逆滤波后的结果，在言语工程上通常使用这种方法，称为语音的残差信号，对其进行积分可以得到声门气流的波形。图下为另外一种逆滤波的方法，这种方法是首先将语音信号进行预加重（preemphasis）处理，然后提取出线性预测系数，将提取出的线性预测系数对同一段未进行预加重的语音信号进行逆滤波，得到的就是图下中的信号，这种声源信号通常用于语音学和言语产生理论的研究，这是因为利用这种微分的声源信号能够更好地解释语音声源的物理意义和语言学意义，例如，著名的 LF 模型（Fant et la, 1985）和 Klatt88 串并联共振峰参数合成器都是利用声门气流的微分形式来建立模型的。

虽然线性预测逆滤波有算法简单快捷等优点，在工程也广泛使用，但在对语音声源的研究方面，它也存在许多缺点至今不能很好地解决。这是因为线性预测是全极点（共鸣）模型，因此无法很好地提取出语音中的零点（反共鸣），而人的言语产生系统中始终都存在零点。首先，由于人的鼻腔和口腔同时参加共鸣时会出现耦合现象，因此产生了零点(Dang, J., K. Honda, et al, 1994)；第二，由于咽腔底部梨状窝的存在，发音时也会导致出现零点（Dang, J. and K. Honda,1997）；第三，声门下气管在声门打开时同样会产生耦合，以至于产生零点；这三个方面都影响逆滤波的精度和效果，因此，目前在通常的研究中，可以通过选择元音，特别是/a/元音的办法避开由于鼻音导致逆滤波不精确的问题，但总不是解决问题的根本办法。要想根本解决逆滤波的中零点的问题，就必须提取零点将其加入到逆滤波参数中。

逆滤波还可以用滤波器组的方法，这种方法通过改进界面的友好性，可以加入人为干涉的功能，通过人工干预极点和零点的参数，达到理想的逆滤波效果。一个简单的方法就是将 Klatt88 串并联共振峰参数合成器反过来写就是一个很好的零极点的逆滤波器。另外，我们目前正在研究开发的“基于频谱反转的逆滤波系统”也得到了很好的结果，但仍需要进一步改进才能完善。相信随着逆滤波技术的发展和改进，通过逆滤波提取语音声源将会越来越准确和简便易行。

4 频谱倾斜率分析法

人类言语嗓音的特性是每个倍频程下降 12 个分贝，由于发音人、性别和语言不同，这个数字会发声变化，如果一种语言中有不同的发声类型，嗓音的频谱倾斜率就会有较大的差别，这为我们研究嗓音发声类型提供了实证测量和研究的可能，但言语声波是通过共鸣以后发出来的，因此加入了声道的共鸣特性。如果要测量嗓音的频谱倾斜率，首先就要对语音进行逆滤波，在去掉了语音的共鸣特性后，才可以对信号进行频谱倾斜率的测量，这就是语音分析前的预处理。在前面逆滤波一节中我们讲过，语音常常有零点存在，如果语音处理不能将所有的零点都去掉，可能会影响到最终的测量结果。但从某种意义上讲，如果使用一种相同的逆滤波的方法，即使有部分零点没有去掉，只要条件相同，不会对结果有很大影响。

频谱倾斜率分析方法的具体做法是 1）对原始的语音信号进行逆滤波，提取出声源信号，逆滤波的方法可以使用提取语音参差的方法，也可以使用提取声门气流微分形式的方法，但这两种方法对后面提取的参数会有很大的影响，大致会有每个倍频程 6 个分贝的差别，前者的谱倾斜率要小，后者的谱倾斜率要大；2）对逆滤波后的信号分帧计算功率谱，然后对每一帧的功率谱作局部最大值检测；3）使用多项式拟合的方法对检测出的局部最大值进行曲线拟合，通常情况下使用二次多项式拟合；4）根据拟合出的曲线计算出嗓音每个倍频程下降的分贝数作为最终的结果，单位是每个倍频程下降多少分贝（- dB/oct）。

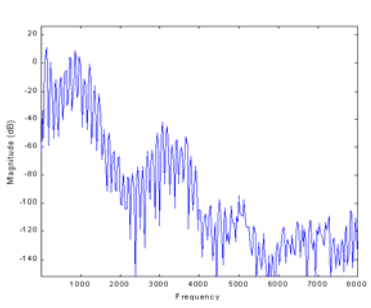


图 8 普通语音功率谱

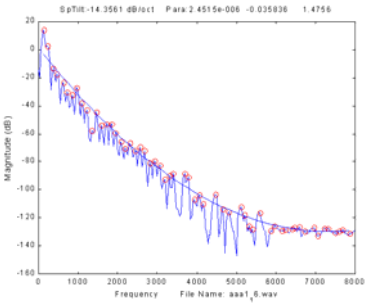


图 9 频谱倾斜率示意图

图 8 是语音的普通功率普，图 9 是频谱倾斜率示意图。可以看出，图 9 中的频谱下倾很平滑，这是因为经过了逆滤波，去掉了共振峰，小的圆圈是自动检测出的局部最大值，平滑线是根据局部最大值得到的二次多项式拟合后的曲线，图上方的参数分别是二次多项式曲线的截距、斜率、曲率和频率倾斜率（-dB/oct）。

频谱倾斜率分析的方法可以用在许多方面，这些研究主要包括：1）对一种语言的频谱特性进行定量分析，以便在通信等领域中作为这种语言的基本参考；2）对一种语言不同发声类型的持续元音进行嗓音特性的定量分析；3）对病变嗓音的不同类型进行性质的描写；4）对有语言学意义的发声类型进行分析和感知方面的研究；5）为语言韵律研究和韵律的建模提供基础数据；6）作为嗓音合成的基本参数。总之，嗓音频谱倾斜率（频谱下倾）在嗓音发声类型研究中是一种重要的方法。

5 多维嗓音分析法

多维嗓音分析最早是从嗓音病理领域发展出来的一种通过声音检测嗓音质量、发声类型和诊断嗓音病变的声学方法，在国际上，特别是在欧美等发达国家，医院里常常患者用此方法对嗓音患者进行初步诊断并以检测的数据和图标作为嗓音的病例。

多维嗓音分析主要分信号录音、算法和参数 3 个方面，第一是录音，多维嗓音分析要求录音必须是二至三秒钟的持续元音，而且需要 44-48k 的采样频率。另外，根据我们的实践，声门阻抗信号也可以用于多维嗓音的分析，但需要进行一些预处理，最好是用声门阻抗信号的微分形式。第二是算法，多维嗓音分析的算法很多，这里我们介绍两个最重要也是最有特色的算法[2]，绝对频率抖动和频率抖动百分比。

绝对频率抖动将一段浊音音调周期之间的变化定义为：

$$Jita = \frac{1}{N-1}\sum_{i=1}^{N-1}|T_0^{(i)} - T_0^{(i+1)}|$$

其中，T0(i)的 i=1, 2, 3, ... N 是提取的音调周期参数，N 等于提取的音调周期的个数。

频率抖动百分比将一段浊音音调周期之间的相对变化定义为：

$$Jitt = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_0^{(i)} - T_0^{(i+1)}|}{\frac{1}{N}\sum_{i=1}^{N}T_0^{(i)}}$$

其中， T0(i)的 i=1, 2, 3, ... N 是提取的音调周期参数，N 等于提取的音调周期的个数。

常用的多维嗓音分析参数有六类 33 项 6 大类，第一类是“基音基础参数”，包括：1）平均基频(Fo. Hz)，2）平均音调周期 (To. Ms)，3）最高基频 (Fhi. Hz)，4）最低基频 (Flo. Hz)，5）F0 标准偏差 (STD. Hz)，6）基频半音范围 (PFR)。第二类是“频率抖动参数”，包括：7）F0 抖动频率(Fftr．Hz)，8）振幅抖动频率(Fatr．Hz)，9）分析样本时长(Tsam s)，l0）绝对频率抖动(Jita. Us)，11）频率抖动百分比（Jiit.%），12）相对平均扰动(RAP.%)，13）音调扰动商(PPQ.%)，14）平滑音调扰动商(sPPQ.%)，15）基频变化率 (vFo. %)。第三类是“振幅抖动参数”，包括：16）振幅抖动(ShdB dB)，17）振幅抖动百分比(Shim dB)，18）振幅扰动商(APQ%)，19）平滑振幅扰动商(sAPQ%)，20）振幅变化率(vAm%)。第四类是“嗓音指数”，包括：21）清浊率(NHR)，22）嗓音骚动(VTI)，23）软发声指数(SPI)，24）F0 抖动强度指数(FTRI%)，25）振幅抖动强度指数(TRI %)。第五类是“嗓音清化参数”，包括：26）嗓音破裂级(DVB)，27）次和谐级(DSH)，28）清声级(DUV)，29）嗓音破裂数(NVB)，30）次和谐音段数(NSH)，31）非浊音段数(NUV)。第六类是“基本参数”，包括：32）计算音段数(SEG)，33）总测定音调周期(PER)。图 10、11 是一个正常嗓音和一个气嗓音的图形表示。

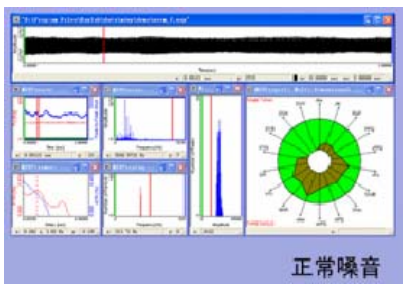


图 10 一个正常嗓音的图形表示

[2] 见美国KAY公司多维嗓音分析选件的使用手册。

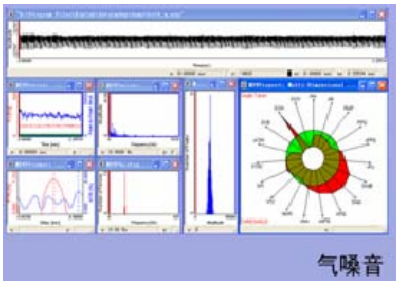


图 11 一个气嗓音的图形表示

从以上的数据可以看出，多维嗓音分析是一种从声学的角度描写个人嗓音特性、区分嗓音性别、鉴定嗓音声纹、量化不同语言嗓音和诊断嗓音病变的有效方法。虽然多维嗓音的有些算法还需要改进，但大部分的算法都是公认的已经很稳定的算法。我们在区分汉语、藏语、蒙语和彝语四种语言的嗓音特性中得到了很好的结果（孔江平，2001；Shen Mixia and Kong Jiangping. (1998),；Kong Jiangping, Caodao Barter, Chen Jiayou and Shen Mixia. (1997).； Hall K. D., Yairi E., 1992,；Horii Y. , 1985,）。

6 声门阻抗分析法

声门阻抗信号（signal of electroglottalgraph）是通过声门仪（laryngography,通常称“喉头仪”）采集的涉及声门变化的生理电信号，这个仪器最早是由英国伦敦大学的佛森教授发明和研制的。众所周知，直到目前人们从语音信号中提取声源信号还有很多困难，因此在研究语音的声源方面还存在许多障碍，声门仪的出现确实在很大程度上推动了嗓音声源的研究，特别是在言语嗓音生理和嗓音病理的研究和诊断方面得到了很大地发展。

声门阻抗信号的出现，给研究者开辟了一个新的领域，使人们对声门的变化、声带的振动方式和嗓音声源的关系研究有了很大的发展，特别是对语言的发声类型有了更好的认识。在语音学研究方面，从声门阻抗信号中提取出来的参数可以很好地用来描写不同语言的发声类型，因而被语音学家广泛使用。从声门阻抗信号中可以提取出许多参数用于嗓音发声的描写、研究和建模。其中有三个参数最为重要，它们是：1）基频，2）开商，3) 速度商。实际上，基频、开商和速度商不仅仅是从声门阻抗信号中可以提出，从语音声源信号的积分形式中也能提取出来，方特教授著名的LF嗓音模型中使用的开商和速度商就是指从语音信号中提取出的开商和速度商 。[3]

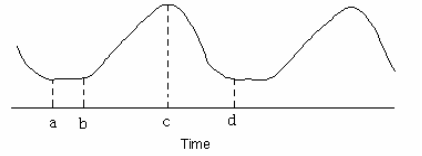


图 12 声源信号的基频、开商和速度商的基本定义

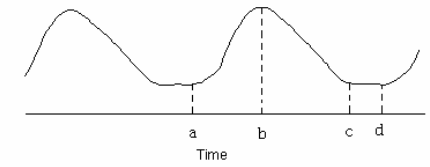


图 13 声门阻抗信号的基频、开商和速度商的基本定义

从言语信号的物理意义上讲，基频是周期的倒数，这个比较清楚。开商是指声门打开相比整个周期，同样的物理意义，我们也可以用接触商来表示，只是数据不同[4]。速度商是指声门的正在打开相比声门的正在关闭相。

图 12 为从语音信号中提取出来的嗓音信号的积分形式，通常情况下其波形的峰值是右倾的，图中 ad 为周期，ab 为闭相，bd 为开相，bc 为声门正在打开相，cd 为声门正在关闭相。基频、开商和速度商可以用以下公式来定义：

基频 = 1/周期(ad)

开商 = 开相(bd)/周期(ad) ×100%

速度商 = 声门正在打开相(bc)/声门正在关闭相(cd)×100%

图 13 为语音声门阻抗信号的积分形式，也是原始形式，通常情况下其波形的峰值是左倾的，图中 ad 为周期，ac 为闭相，cd 为开相，bc 为声门正在打开相，ab 为声门正在关闭相。基频、开商和速度商可以用以下公式来定义：

基频 = 1/周期(ad)

开商 = 开相(cd)/周期(ad)×100%

速度商 = 声门正在打开相(bc)/声门正在关闭相(ab)×100%

使用基频、开商和速度商可以用来描写和定义不同的发声类型。在语言发声类型的研究方面，这些定义可以用来描写汉语声调的嗓音发声模型、民族语言中元音的发声类型、汉语韵律研究的嗓音模型、病变嗓音的性质、声纹鉴定、声乐研究中的不同唱法和唱腔等。限于篇幅本文只是

[3] 这几年有很多同行问我关于开商和速度商的定义，大家觉得有些文献上讲的有出入，不是很清楚，主要的问题就是因为这两个定义不仅用在声源信号上，而且用在了声门阻抗信号上，因而产生了一点混淆。

[4] 见Kay公司的EGG相关分析软件。

简单介绍一下，给出这些定义的嗓音区别性特征和声学发声图。

	气泡音	气嗓音	紧喉音	正常音	高音调嗓音
音调	1	2	3	4	5
速度商	4	1	5	3	2
开商	5	4	1	3	2
音调抖动	4	5	2	1	3

表 1 发声类型特征表

	气泡音	气嗓音	紧喉音	正常音	高音调嗓音
音调	－	－	－	＋－	＋
速度商	＋	－	＋	＋－	－
开商	＋	＋	－	＋－	－
音调抖动	＋	＋	＋	＋－	＋

表 2 发声类型区别特征表

表 1 是发声类型特征表，列出五种发声类型及嗓音参数大小顺序的数值，参数根据基频参数的大小排序，从数据的矩阵中可以看出这五种发声类型可以完全区分开来，因此，根据这一性质，我们可以将其转换为区别性特征来描写语言的不同发声类型，因为通常情况下，这五种发声类型能涵盖大部分语言的发声类型现象。

表 2 是五种语言发声类型及其参数的区别性特征表，特征符号“＋”和“－”是根据正常嗓音的参数来区分的，即，正常嗓音定义为“＋－”，大于正常嗓音的参数定义为“＋”，而小于正常嗓音的参数定义为“－”。从表中可以看出，紧喉音可以描写为：音调“－”，速度商“＋”，开商“－”，音调抖动“＋”，见表二。当然根据嗓音的这些数据，还可以采用其他的方法来描写嗓音发声类型，建立更符合某种语言音位系统的区别性特征系统。

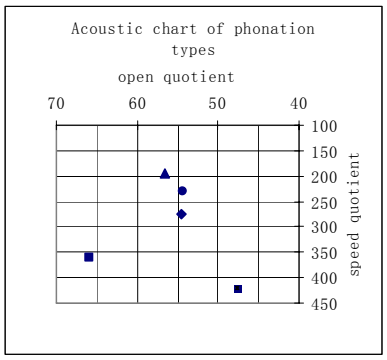


图 14 声学发声图

在语音学研究中，通常用声学元音图来描写元音的位置及其特性，根据嗓音的参数，我们提出用“声学发声图”或者也可以成为“声学嗓音图”来描写语言嗓音发声类型的特性（孔江平，2001），图中横轴为开商，纵轴为速度商，即根据这种体系和方法，图中菱形为正常嗓音，右下角的小方形表示紧候音，图左下角的大方形是气泡音，正常嗓音上边的圆形是高音调嗓音，三角形是气嗓音，同样我们也可加上基频画出三维的声学发声图来。另外，用“声学发声图”不仅可以从声门阻抗信号中获取参数，也可以从声学信号中获取参数，如果参数是从生理信号中获得的也可以称为“生理发声图”或者“生理嗓音图”，其内容是有区别的，但描写语言嗓音发声类型的主导思想完全一致。

7 动态声门分析法

动态声门分析方法的设备和技术基础是高速数字成像和数字图像处理，在设备和技术上对文科背景的语音学研究人员来说有一定的困难，但随着技术的发展，技术问题会越来越简单。众所周知，从传统的语音学到科学的语音学的一个最为重要的标志是在 X 光出现后被及时地应用到了语音学中，从而发现了舌位高低前后和语音发音的重要关系，大大推动了语音学向着科学的方向发展，但直到今天我们对声带的生理机制还知之甚少。上世纪七八十年代高速数字成像技术的进展，使我们有条件对声带的振动进行观察和研究，可以说高速数字成像技术的应用一定会大大推进语言嗓音发声的研究，其意义就在于此。

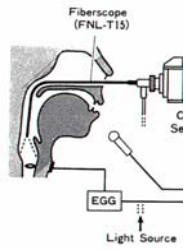


图 15 高速数字成像设备 (Kiritani et al, 1993)

图十五是高速数字成像设备的示意图，内窥镜可以是软的导线也可以是硬的，前端装有镜头和冷光源，另外，还有话筒和电声门仪，通常情况下三路信号同时采集。其中图像的采样频率可以高达 4501 帧/秒(最大采样频率 9000 帧/秒)，文件格式一般为 256´256 像素 8 比特灰度级，随着计算机速度的提高和内存的增加，现在已经有了彩色的高速数字成像系统。

在信号处理技术层面，主要是图像处理和语音信号处理。数字图像信号处理的目的主要是提取出视频信号中声门的面积，然后根据面积提取各种用于研究的参数。语音信号的主要目的是提取出有用的语音参数，如，基频、共振峰、开商、速度商、振幅等。有了这些参数就可以进行动态声门和语音关系的研究。

图 16 图像处理的简单过程

从图 16 可以看出，为了处理的方便，一帧图像可以先加一个小窗用来确定声门的面积，然后经过调节对比度和抽取声门面积等方法最终得到动态的声门面积，当然也可以用自动的方法。这只是最简单的过程，因为实际的高速数字视频会出现光线灰暗、抖动和漂移等现象，都需要加以处理才能得到较好的动态声门。

动态声门的参数提取比较复杂，不是技术的问题而是怎样定义的问题，因为，提取声门面积有很多方法，也可以定义很多参数，这需要根据研究的目的来确定。因为高速数字成像不仅可以用于声带振动的基础研究，而且还可以用于嗓音病理的诊断和研究以及言语工程的生理合成研究。因此，只有确定了研究的目的，才能断定提取的参数。在基础研究方面有以下基本参数可以比较全面地对动态声门进行研究和建模，见图 17。

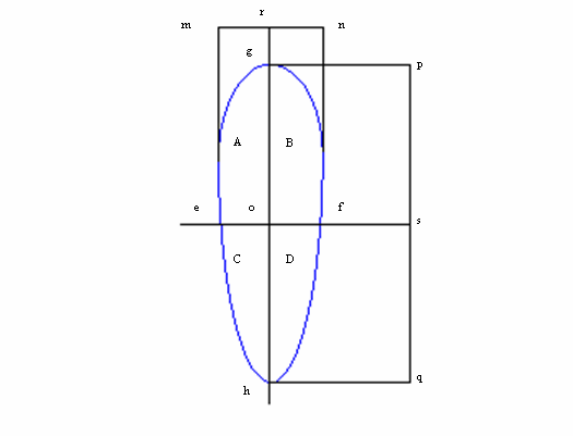


图 17 参数定义示意图

第一类是基本参数，包括：1）声门面积最大值位置；2）声门开启点；3）声门关闭点；4）绝对声门长度；5）绝对声门宽度；6）声门面积长宽比。

第二类是声门面积参数，包括：7）声门总面积；8）左声门面积；9）右声门面积；10）上声门面积；11）下声门面积。

第三类是声门长宽参数，包括：12）声门长；13）声门；14）前声门长；15）后声门长；16）左声门宽；17）右声门宽。

第四类是声门面积函数参数，包括：18）声门面积函数周期；19）声门面积函数；20）声门面积函数开相；21）声门面积函数闭相；22）声门面积函数开商；22）声门面积函数速度商；24）直流分量基础参数。

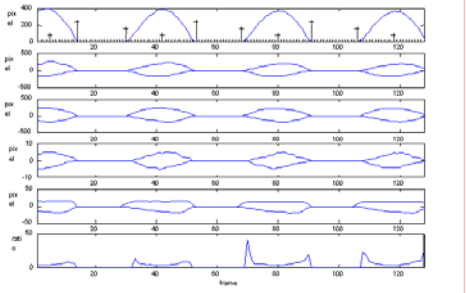


图 18 正常嗓音声门参数示意图

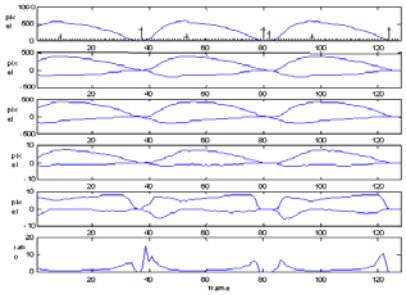


图 19 气常嗓音声门参数示意图

图 18 和 19 是正常嗓音和气嗓音声门参数示意图，图中给出了十三个参数，从上至下分别是声门面积、左右声门面积，前后声门面积、左右声门宽度、前后声门宽度、长宽比，另外，根据声门面积可以计算出基频、开商和速度商，这样一共是十三个参数。这十三个参数是经过验证的，比较稳定，而且和具有物理意义的语音参数有比较密切的关系。

动态声门技术是一项比较新的技术，通过应用这一技术和研究方法，可以探索语言嗓音发声类型生理和物理之间的关系，追寻语言发声类型的本质，因而它在语言发声研究领域不仅具有理论意义，而且具有实际的应用价值，另外，这一技术在嗓音医学领域和言语工程领域具有很好的开发应用前景（Kong Jiangping，2007[5]）。

8 发声起始状态分析法

发声起始状态分析是指对声带振动起始过程的分析，大家知道，由于各人声带条件的不同、语言发声类型的不同和前边声母条件的不同，声带从静止到振动的过程会发生变化，从而导致各种不同的声带振动起始方式。生理上将声带振动到声带完全闭合（接触）称为“声带接触时间（Vocal attack time，VAT）”(Baken RJ, Orlikoff RF.1998；R.J.Baken et la, 2007[6])，通过测量VAT可以帮助我们确定嗓音的发声类型，下面对这种方法进行一些简单的介绍。

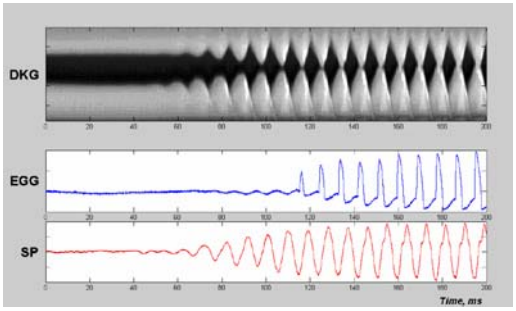


图 20 VAT 原理示意图

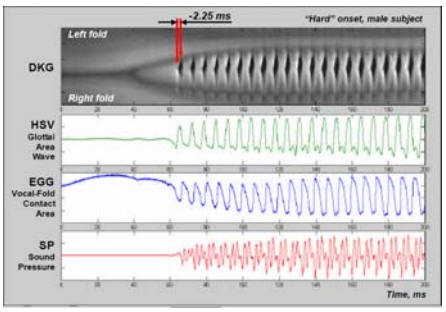


图 21 硬嗓音 VAT 原理示意图

图 20 是 VAT 原理示意图，图分为上中下三张，上图是声带高速数字成像单线图（kymography），其成像原理是在声带振动的高速视频图像中选一条线，然后将它们排列成一张图片，横轴是帧数或者时间，纵轴是每一帧所取的画面，通常是取声门的中间线，从上图可以看出声带振动的变化过程。中图是声门阻抗信号，从声门阻抗信号的原理我们知道，一旦声带接触，声门阻抗信号会突然增大，因为声带接触时阻抗会变得很小。下图是经过带通滤波的语音信号。

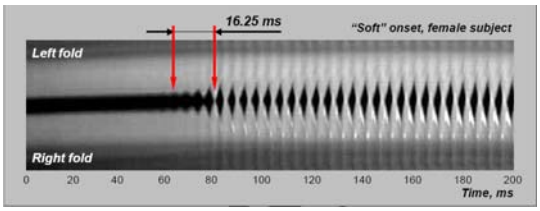


图 22 软嗓音 VAT 原理示意图

从图 20[7]可以看出，声带从开始振动到声带完全接触有一个较长的过程，大概有六七个振动周期，结合其他信号可以明显看出，声带一开始振动声压就出现了，但声门阻抗信号特别小，但当声带接触时声门阻抗信号就突然变大。根据这一原理，我们就可以测量出VAT来。从图 21 可以看出，声带的振动是突然启动的，它在第一个振动周期声带就完全闭合了，体现出了另一种不同的声带启动方式。从这两张图可以看出，VAT的定义是“声带开始振动到声带接触的时间”。根据这个定义，图 21 的VAT为负值（-2.25ms）。图 22 是另一张嗓音VAT示意图，从中可以看出VAT的数值为正（16.25ms）。从视频可以看出前者为硬启动嗓音（hard voice），后者为软启动嗓音（soft voice）。

VAT 是一种最近才发展出来的语言嗓音研究的方法，它有很好的研究和应用前景，虽然目前还主要是应用在病理嗓音的诊断、分析和研究方面。很显然，这一方法经过改进可以用于语音发声类型的共时描写、发声类型和声母生理关系的

[5] 有关利用高速数字成像研究声带振动和动态声门的文章很多，有大量参考文献，但专著很少，在这本书中列出了上个世纪能找到的所有文献，可供参考。

[6] 正在发表中。

[7] 这几张图取自R.J.Baken et la, 2007。

分析、语言发声类型在历史音变中的作用等方面的研究。

9 嗓音音域分析法

研究和测定嗓音有很多方法，其中有一种算法上很简单，但很有用的方法，即嗓音音域分析法，这种方法主要是通过测定发音人的音域范围达到确定一个人嗓音特性的目的。具体的方法是合成一个特定音阶(通常是钢琴键盘上的某个音)音高的声音，让发音人模仿其音高，同时发音人的发音从最弱变到最强。根据语音分段计算出基频和振幅（分贝），然后将其划在二维图上，其中 x 轴为基频，y 轴为振幅，颜色变化为频度。

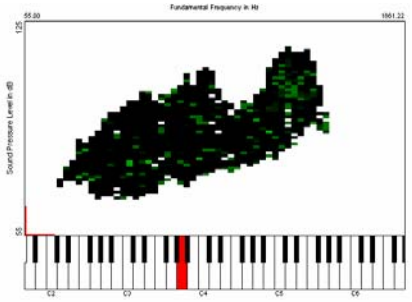


图 23 男声宽音域数据示意图

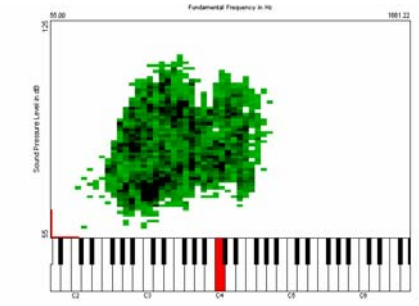


图 24 男声窄音域数据示意图

图 23 和 24 四是两个不同男声的音域示意图 ，从第一张图可以看出，这个男声的音域很宽，有将近四个八度，几十个分贝的分布。从第二张图的数据可以看出，其音域较窄，不到三个八度，但分贝数值和第一个男声大致相同。很显然，利用这一方法可以很好地测定一个人的音域范围，了解其声带和嗓音的自然条件，在声乐考试和教学中都会很有用处。

对嗓音音域分析方法进行一些改进就可以对一个人或一种语言的音域范围进行定性的描写、研究和建模。比如，对两种不同语言的大量语音样本进行计算就能得到该语言的嗓音分布范围，因为我们知道不同的语言在发声上有很大不同，这种差别体现了语言发声的特点。如果加上开商和速度商等参数就可以对嗓音进行建模研究，这种模型对语音参数合成十分有用。

10 嗓音的合成方法

从嗓音的研究方法上讲，嗓音的合成方法可以分为物理和生理两种，但从嗓音最终的物理性质上讲，嗓音的合成最终都要到声学层面。这里从声学和生理两个方面介绍一下嗓音合成的基本方法。

从声学的角度来合成嗓音，首先要从语音中提取出声源信号，然后对这些信号进行分析和建模，最后次采用一种声学模型来合成出想要的嗓音。根据这一原则，人们进行了大量的研究，其中有些是用嗓音的积分形势来建立嗓音模型（Rosenberg A.E. 1971; Hedelin P.,1984; Fant G.,1979），而有些是用嗓音的微分形式来建模(Fant G. ,1982; Ananthapadmanabha T.V. ,1984; Fant et al, 1985; Ljungqvist et al, 1985)。

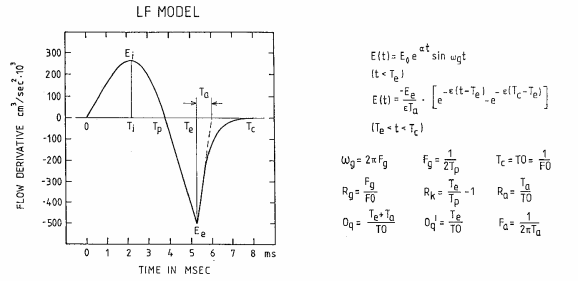


图 25 方特等的 LF-模型

图 25 给出了方特教授的著名的嗓音声学模型，这个模型是用两部分组合来模拟一个周期内部的嗓音，第一部分是由一条正弦曲线和一条指数曲线的乘积所产生的新的曲线来模拟，第二部分是由一条指数曲线来模拟，两条曲线构成了一个周期内部完整的嗓音。图中左半部分是嗓音曲线，右半部分是定义。可以看出，方特教授等的模型只有四个参数，并且和嗓音的物理意义有明确的关系，是一种简单明了又能很好反映嗓音特点的模型，可以合成不同的嗓音发声类型，因此，这种模型可以很好地应用在语音参数合成系统中。

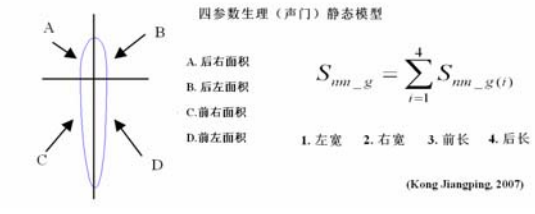


图 26 四参数生理（声门）模型

图 26 是四参数嗓音生理静态声门模型，建模的参数时通过高速数字成像系统采集到的声带高速视频信号，通过处理这些信号得到动态声门，经过统计分析和语言发声类型的分析，最终提出的一个建模方案。从图中可以看出，整个声门面积是由四个四分之一椭圆组成，即，声门后右面积，声门后左面积，声门前右面积和声门前左面积。这四个面积是由四条边长计算出来的，因此，最终的参数只有四个声门边长参数。在合成嗓音时还要结合动态声门的参数，详见 Laryngeal Dynamic and Physiological Models（Kong Jiangping, 2007）。

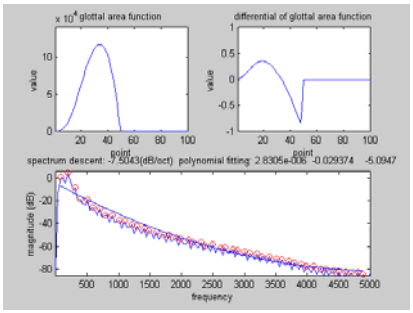


图 27 正常嗓音生理合成参数图

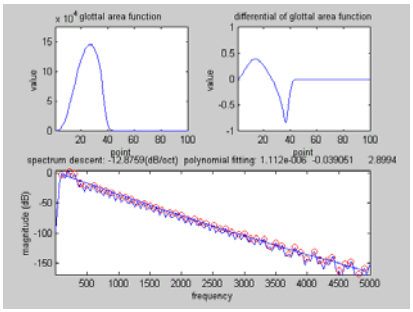


图 28 漏气嗓音生理合成参数图

图 27 和 28 是利用该生理模型合成的嗓音例子，前者是正常嗓音，上左图是声门面积函数，上右图是微分形式，下图是嗓音频谱倾斜率及参数，从中可以看到频谱倾斜率为-7.5/倍频。后者是一个有点漏气的嗓音，这个嗓音的频谱倾斜率为-12.87/倍频。

根据动态声门建模和合成嗓音，除了可以用于语音参数合成以外，就目前来说，其最主要的目的是用来模拟声带振动后形成的动态声门，这样就可以对声带的振动进行仿生的研究，例如，调节左右声门的合成频率，可以使左右声门的频率产生一个微小的差别，这样合成出来的嗓音就会出像噪音或者同时出现大小周期，这种嗓音发声类型在病变嗓音中经常出现，这就为我们研究嗓音生理机制、病变嗓音特性、病变嗓音手术方案和特殊嗓音合成奠定了理论基础。

11 结束语

语言发声研究的进展始终随着嗓音生理和声学研究技术的发展，两者相辅相成，缺一不可，从而推动了语言发声研究的进步。然而，在我们选择研究方法时，并不一定非要选择最复杂的研究方法，而是要根据研究对象和研究目的选择最适当的方法，这样才能得到最为有用和可靠数据，从而揭示嗓音发声的内在规律，达到研究的目的。在我国，发声研究方法的进步促进了嗓音发声研究的发展。总的来说，这些研究方法可以应用在发声语音学、生理语音学、嗓音病理学、言语声学、声纹鉴定等相关领域。

在面向语言学的语音学领域，新的研究方法使我们可以开创新的研究热点，例如，VAT 的应用不仅可以用来嗓音发声的启示状态，而且可以利用此方法研究辅音和嗓音发声生理的内在机制，从而达到解释语言历史音变生理制约的基本规律。又如，基于高速数字成像的动态声门研究可以使我们有可能建立声带振动的生理模型，从而模拟语言嗓音发声类型的产生和基本的特性。也可以用来模拟病变嗓音的振动方式和制定嗓音病变手术的方案以及模拟术后的嗓音。另外，嗓音的生理模型的不断完善，将会大大推动基于仿生学的语音生理合成的完善和建立高质量的合成系统，推动语言嗓音发声基础和应用研究的发展和进步。

参考文献

[1] 方特.G、高奋.J，1994，言语科学与言语技术，商务印书馆，北京。

[2] 孔江平（2001），《论语言发声》，中央民族大学出版社，北京。

[3] Alku P. (1991). Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Pro. EUROSPEECH `91, 1081-1084.

[4] Ananthapadmanabha T.V. (1984). Acoustic analysis of voice source dynamics. Speech Transmission Laboratory – Quarterly Progress and Status Report, 2-3, 1-24. Roya Institute of Technology, Stockholm.

[5] Anthony Traill and Michel Jachson, 1987. Speaker variation and phonation types in Tsonga nasals, UCLA Working Papers in Phonetics 67, June.

[6] Baken RJ, Orlikoff RF. Estimating vocal fold adduction time from EGG and acoustic records. In: Schutte HK, Dejonckere P, Leezenberg H,

Mondelaers B, Peters HF, eds. Programme and abstract book: 24th IALP congress, Amsterdam; 1998:15.

[7] Dang, J., K. Honda, et al. (1994). "Morphological and acoustical analysis of the nasal and the paranasal cavities." Journal of the Acoustical Society of America 96(4): 2088-2100.

[8] Dang, J. and K. Honda (1997). "Acoustic characteristics of the piriform fossa in models and humans." Journal of the Acoustical Society of America 101(1): 456-465.

[9] Fant G. (1979). Glottal source and excitation analysis. STL-QPSR, No. 1, pp. 85-107.

[10] Fant G. (1982). The voice source, acoustic modeling. STL-QPSR, No. 4, pp. 28-48.

[11] Fant G., Liljencrants J. and Lin Q., 1985, A four parameter model of glottal flow. STL-QPSR, No. 4, 1985, pp 1-13.

[12] Hall K. D., Yairi E., 1992, Fundamental frequency, jitter, and shimmer in preschoolers who stutter, Journal of Speech and Hearing Research, Volume 35, 1002-1008, October.

[13] Hedelin P. (1984). A glottal LPC-vocoder. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1.6.1-1.6.4. San Diego.

[14] Hirose H. (1997). Investigation the Physiology of Laryngeal Structures. Chapter 4. The Handbook of Phonetic Sciences, edited by William J. Hardcastle and John Laver. Blackwell Publishers.

[15] Horii Y. , 1985, Jitter and shimmer in sustained vocal fry phonation, Folia Phoniatrics, Vol. 37, 81-86.

[16] Kirk P.L., Ladefoged P. and Ladefoged J., 1984, Using a spectrograph for measures of phonation types in a natural language, UCLA Working Paper in Phonetics.59, pp 102-113.

[17] Kong Jiangping，2007，Laryngeal Dynamic and Physiological Models: High-Speed Imaging and Acoustical Techniques, Peking University Press.

[18] Kong Jiangping, Caodao Barter, Chen Jiayou and Shen Mixia. (1997). Acoustic study of jitter, shammer and tremmors in Mandarin. Theory and Application of Signal Processing. Conference on Speech, Image and communication Signal Processing of China (SICS'97) , Oct., 1997, Zhengzhou, China. (in Chinses).

[19] Ladefoged P., 1973, The features of larynx. Journal of Phonetics. 1, pp.73-84.

[20] Ladefoged P., 1988 , Discussion of Phonetics: A note on some terms for phonation types, "Vocal Fold Physiology, Vol. 2, Vocal Physiology: Voice Production, Mechanisms and Functions", Raven Press, New York.

[21] Ladefoged P., Maddieson I., and Jackson M., 1987a, Investigating phonation types in different languages, UCLA Working Papers in Phonetics 67, June.

[22] Ladefoged P., Maddieson I., Jackson M. and Huffman M., 1987b, Characteristics of the voice Source, UCLA Working Papers in Phonetics 67, June 1987. [To be appeared in the Proceedings of the European Conference on Speech Technology, Edinburgh, 2-4 September.

[23] Laver J., 1980, The Phonetic Description of Voice Quality, Cambridge University Press.

[24] Lindestad P-A, Sodersten M., Maerker B. and Granqvist S. (1999). Voice source characteristics in Mongolian "Throst singing" studied with hgih-speed imageing technique, acoustic spectra and inverse filtering. Phoniatric and Logopedic Progress Report, No 11. Department of logopedics and phoniatrics, Karolinska institute, Huddinge university hospital, Sweden.

[25] Ljungqvist M. and Fujisaki H. (1985). A comparative study of glottal waveform models. Technical Report of the Institue of Electronics and Commmunications Engineers. Japan, EA85-58, 23-9.

[26] Maddieson I. And Ladefoged P., (1985), 'Tense' and 'Lax' in four minority languages of China, UCLA Working Paper in Phonetics. 60-99. 59-83.

[27] Rosenberg A.E. (1971). Effect of glottal pulse shape on the quality of natural vowels. Journal of the Acoustical Society of America, 49,583-98.

[28] Sawashima M. and Hirose H.. (1983). Laryngeal gestures in speech production. Chapter 2. The Production of Speech, edited by Peter F. MacNeilage, Springer-Verlag, New York, Heideberg, Berlin.

[29] Shen Mixia and Kong Jiangping. (1998),

MDVP(multi-dimensional voice processing) study on sustained vowels of Mandarin through EGG signal. Proceedings of Conference on Phonetics of the Languages in China, May 28, Hong Kong.

[30] Titze I.R. (1994). Principle of Voice Production. Prentice Hall, Englewoods cliffs, New Jersey 07632.

[31] Therapan L. Thongkum, 1987 “Phonation types in Mon-Khmer languages”, UCLA Working Papers in Phonetics 67.

[32] William J. Hardcastle and John Laver (edit), (1997), Handbook of Phonetic Sciences, Blackwell Publishers.

（孔江平 北京大学中国语言文学系语言学实验室 100871 kongjp@gmail.com）

古藏语音位系统的结构和分布[1]

孔江平

0 引言

传统的语言学及其相关学科通常被列入人文科学，在方法上也大都采用思辨的方法，然而，进入了 21 世纪后，语言学及其相关学科明显出现了分化，即从语言学的各个领域分化出了用纯科学方法研究语言的新领域甚至新的学科。例如，传统的语音学先是采用了 X 光技术研究语音的发音动作，声谱仪发明以后又广泛采用了声学分析的实证方法，进入本世纪后，语音学又进入了脑科学领域，在近十年的《科学》和《自然》这两本最重要的科学期刊中，有关语音学和语言学的论文几乎都涉及脑科学，它们包括语音习得，语义认知，甚至句法结构。很显然，人们正在通过不同的语言表层追寻语言活动的本质。然而，传统语言学的丰富知识、理论框架和思路始终都是语言科学研究的基础，其中，音位“功能负担（functional load）”就是传统语言学中一种极具科学的方法。

藏语是我国一种历史悠久的民族语言，属汉藏语系藏缅语族藏语支。藏语使用人口众多，分布地域广大，语言环境相对纯净。同时，藏文是一种拼音文字，又有丰富的古代文献，这为音位功能负担的研究创造了很好的条件。另外，藏语方言有从无声调到有声调的完整发展过程，这对研究藏语乃至汉藏语声调的起源有很重要的意义。本文将从语音学、音系学和语言学的角度，以古藏语的音位系统[2]为研究对象，对古藏语音韵系统中音位的结构和分布进行分析和研究，其目的是想通过对古藏语音位系统、结构和分布的研究，为古藏语音位功能负担的研究奠定语言科学的基础。

[1] 在跟王士元教授读博士的期间曾经读了王老师关于“功能负担（functional load）”方面的文章。王老师的这些论述无论是在语言学领域还是在现代信息技术领域都有很重要的理论意义和现实意义，但中国这方面的研究还很少。近几年和西北民族大学的于洪志教授及李永宏博士一同在建立藏文和藏语的数据库，有了做藏语功能负担研究的条件，因此，写了这篇很不成熟的文章一来为纪念王先生在这方面的贡献，二是希望能有更多的国人了解“功能负担”的研究在语言学和现代信息技术领域的重要性。

[2] 藏语是拼音文字，它从创立到现在变化不是很大，因此，在藏语学界一般都认为藏文所反映出来的音位系统基本上就是古藏语（7-9 世纪）的音位系统。

1 音位功能负担

音位功能负担的概念和研究可以追溯到早期的布拉格学派时期（Mathesius, 1929; Jakobson, 1931; Trubetzkoy, 1939），当时主要注重于音位的二元对立。在功能负担的语言学研究方面，50 年代，主要有霍凯特的研究（Hockett，1955，1961）和格林博格(Greenberg, 1959)的研究。霍凯特认为：功能负担的重要性在于它对描写音韵系统有重要的价值，从而使我们可以有一个尺度来认识语言信息、语言冗余度和言语识别。格林博格认为：功能负担以通用的方式反映了一组音位或一组对立特征各成员之间的对有区别意义信号的贡献。在 60 年代，主要有赫厄希斯瓦尔德关于功能负担和音变的研究（Hoenigswald, 1960），他认为：功能负担和语言的音变有关，并提出了一个假说，即“在一种语言里，如果一种对立用的很少，它的消失对系统造成的危害要小于功能负担大的对立”。60 年代还有王士元教授（Wang, 1967)有关功能负担的著名研究（详见下）和京•罗伯特（King，1965，1967，1967）的研究，京•罗伯特将音变和功能负担一同进行研究，并着重研究了音位功能和语音音变的关系，发现在日耳曼语中，功能负担和历史音变的关系不大。在本世纪有苏仁德兰和尼育基（Surendran and Niyogi，2003)的研究和苏仁德兰和利佛（Surendran and Levow，2004) 的研究，苏仁德兰和利佛在其研究中不仅讨论了霍凯特的定义，还讨论了音位、区别性特征和超音段特征的功能负担，同时，他们还研究了汉语声调的功能负担，发现汉语声调的功能负担与元音同样高。

在 60 年代最具有代表性的是王士元教授的研究，他首次实施了大文本语料功能负担的计算和指出了计量功能负担的困难，并给出了解决这些困难的方法。首先他讨论了音位系统中常见的三种分布、霍凯特与格林博格的测量方法以及这些方法和香农等（Shannon and Weaver, 1949；Shannon,1951；Kucera,1963）的通信理论及各种语言学概念的关系；其次在这些背景知识的基础上，王士元教授讨论了功能负担计量必须满足的五个条件；最后他系统地发展了四种计量功能负担的方法。另外，王士元教授还指出“音变严重受到其他许多因素的影响，如音位之间语音的相似度和语言的接触等，但

正如许多历史语言学家相信的那样，如果功能负担在音变中确实起作用的话，那么用量化的解释至少可以从一个方面阐明音变这一难题”。

关于音位系统的分布，王士元教授指出：“对任意一个语音序列，有三种相互关联的分布，即相似性分布、交叉性分布和互补性分布。当语音序列中每一个音位的排列都确切共有一组相同的环境时，它就处于相似性分布中；当音位的排列共有一些相同的环境，而不是所有的环境时，该序列处于交叉性分布中；当音位排列没有任何共有环境时，该序列处于互补性分布中”。王士元教授的研究为后来功能负担的研究建立了一个理论上的基本框架。

以往的功能负担研究都是在大文本的基础上进行音位的功能负担统计和量化以及计算该语言的熵值和冗余度，对于藏语来说这应该是下一步的工作，因为我们首先应该先对古藏语音位系统内部的结构和分布进行研究，这是本文研究的重心。在此我们将一种语言音位系统内部的“音位结构、音位分布和音位功能负担”称为“音位结构功能负担（structural functional load）”，这种音位内部的“结构功能负担量”是封闭性的。从性质上看，结构功能负担更能体现一个语言音位系统的性质，为语言历史音变的研究开奠定基础。

2 古藏语的音位系统

根据藏语的史书记载[3]，藏文大约创制于七世纪，藏文记载了大量藏族的历史、藏族各领域的知识和文化以及民间传说和史诗，传承了丰富多彩的藏民族文化。

2.1 藏文的起源和文字系统

本实验采用的处理程序是北京大学中文系语音乐律平台，在 Windows 平台下用 Matlab 编程实现。

关于藏文的起源，目前学者们还有一些不同的看法，但一般都认为在七世纪中叶，藏王松赞干布派大学者图弥桑布扎赴印度学习梵文回国后，仿效梵文创制了古藏文。另外，有学者认为统一的藏文创制前西藏地区已经有文字在使用，主要和古代的象雄文有关[4]。统一的藏文一共有 30 个辅音字母，4 个元音符号，加一个零位ཨ(a)，共 5 个元音，可以说藏文基本上反映了古藏语的音韵系统。见表 1-2。

表 1：藏文字母表

藏文	转写	序号	藏文	转写	序号	藏文	转写	序号	藏文	转写	序号	藏文	转写
ཀ	k	7	ཉ	j	13	ད	p	19	ཛ	dz	25	ར	r
ཁ	kh	8	ཊ	ny	14	ཕ	ph	20	ཝ	w	26	ལ	l
ག	g	9	ཋ	t	15	བ	b	21	ཞ	zh	27	ཤ	sh
ང	ng	10	ཌ	th	16	མ	m	22	ཟ	z	28	ས	s
ཅ	c	11	ཎ	d	17	ཙ	ts	23	འ	v	29	ཧ	h
ཆ	ch	12	ཏ	n	18	ཚ	tsh	24	ཡ	y	30	ཨ	–

表 2：藏文元音符号表

序号	元音	转写	序号	元音	转写
1	ཨ	a	4	ེ	e
2	ི	i	5	ོ	o
3	ུ	u			

在音节结构上，藏语分为前加字、上加字、基字、下加字、元音、后加字和再后加字，见图 1。

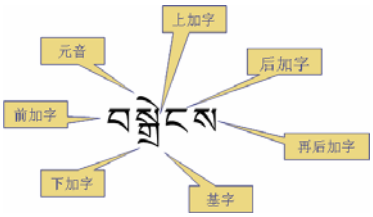


图 1 藏语最长的音节结构图

2.2 藏文历史上的三次厘定

藏文在历史上有三次厘定，第一次厘定是在七世纪中叶图弥桑布扎创制藏文起到九世纪中叶的二百多年里，见根敦群培的《白史》(《དེབ་ཐེར་དཀར་པོ》)；第二次厘定是在九世纪中叶藏王赤热巴巾时期，主要是当时的大译师，如噶哇白泽等，根据当时藏语的发展情况进行的厘定；第三次藏文厘定是从十一世纪末阿里王意希畏时期到十五世纪初的三百多年中，由众多大译师在从事佛经的翻

[3] 这些历史文献主要有《敦煌本吐蕃历史文书》，作者：王尧、陈践译注，责任编辑：工布吉村　民族出版社，1980 年 10 月出版；《玛尼全集》、《布敦教史》、《红史》、《西藏王臣记》、《贤者喜宴》、《白史》等。

[4] 如根敦群培的“藏文的由来与演变”；才让太的“藏文起源新探”；群培多杰的“藏文渊源初探”等。

译中进行的厘定，如仁钦桑布等。这三次藏文的厘定在仁钦扎西的《丁香帐》(《ལི་ཤིའི་གུར་ཁར་》)都有记载，其中以第二次厘定最为重要，史料记载也较为详细。

在七世纪图弥桑布扎创制藏文时撰写了《文法根本三十颂》（ལུང་སྟོན་པ་རྩ་སུམ་ཅུ་པ）（又译《三十颂》）和《字性法纲要》（རྟགས་ཀྱི་འཇུག་པའི་རྩ་བ）（又译《字性组织法》）两本有关藏文文法的经典著作，九世纪大规模进行藏文厘定时，对这两本书也进行了修订，流传至今的这个蓝本是藏语古音韵及文法的基本理论框架，古藏语时期没有声调，藏语古音韵的音位系统基本反映和代表了七世纪的藏语语音的面貌。

2.3 藏文语法文献

《三十颂》和《字性组织法》是藏语古代文献中最有名的文法著作，由图弥桑布扎所著。前者主要是研究藏文辅音和元音之分类、格、虚词、动词形态变化等语法范畴和归类；后者主要是研究语音结构、字母组合搭配、动词屈折变化规律等，可以看出藏语的音位系统和语法系统是有机地组合在一起的。

在以后的藏文文法著作中比较著名的还有《司徒文法详解》（司徒•曲吉迥乃，成书时间不详）。司徒•曲吉迥乃(1699—1744 年,又名曲吉朗瓦)是 17 世纪著名藏文文法学者，他调查和研究了藏族的语言和方言，根据藏语的实际语音详尽解疏了《三十颂》和《字性组织法》，最终写成了这本历史上具有权威性的文法著作。

现代藏文文法著作有《藏语语法》（西北民族学院，语文室，藏语组，1958）、《简明藏文文法》（胡书津，1995）和《实用藏文文法教程》（格桑居冕，格桑央京，2004 年）。它们都对藏文的音位系统和文法有详尽的解释，其中格桑居冕教授的《实用藏文文法教程》不仅详细解疏了《三十颂》和《字性组织法》，还根据现代语音学和语言学的理论提出了自己的看法。另外，还有胡坦教授的《藏语研究论文》（胡坦，2002），这本文集比较全面地涵盖了藏语研究的各个领域。

2.4 古藏文的音位系统

根据《藏语方言概论》（格桑居冕，格桑央京，2002 年），古藏文有 213 个声母和 77 个韵母，这个系统是一个从简原则的系统。格桑居冕教授在其《实用藏文文法教程》中，还有一个古藏语的音位系统。根据这个系统，古藏语的声母一共有 220 个，其中单辅音声母 30 个，二合辅音 115 个，三合辅音 69 个和四合辅音 6 个。古藏语韵母一共有 98 个，其中有 5 个单元音韵母，8 个复元音韵，50 个单尾韵和 35 个辅尾韵[5]，这个系统比较全面地反映了古藏语的音位情况。

我们根据藏文现有的词典[6]，建立了一个藏文词汇的数据库，整个数据库的藏语常用词汇有约 9 万多条，其中单音节词（不包括梵文和借词）有 6273 条，双音节词汇有约 43000 条左右，三音节词汇约有 20000 余条，四音节词汇约有 16000 余条，四音节以上词汇约有 5000 余条，其中单音节词约占总词汇量的 6.3%。根据我们的数据库，本文将古藏语的声母定为 220 个，和格桑居冕教授的声母数目相同，但韵母数目不同，我们的韵母有 100 个，比格桑居冕教授的多 2 个，但实际上有 7 个韵母不同[7]。下文关于古藏语音位的统计是建立在 6273 个单音节词的基础上。

3 古藏语音位系统的结构、分布和负担

由于语言的差异，在建立语言的音位系统时也会采取不同的原则和处理的方法，通常的音位学理论基本上多采用音素音位的方法，然而，中国古代对汉语音韵系统的描写就采用了声韵调的方法。对于藏语来说，采用的方法又有所不同，由于藏语构词方法的特殊性，藏语的音位系统对一个音节的描写采用前加字、上加字、基字、下加字、元音、后加字和再后加字的方法。从言语产生的线性原则上看，任何一种语言都可以采用音素音位的方法来建立其音位系统，但不是所有的语言都可以用声韵母系统来进行音位系统构建。下面我们将先从音素音位系统的角度来分析古藏语音位系统的结构和音位之间的分布关系，然后，再对不同的方法进行讨论。

3.1 前加字的结构和分布

根据藏文文法[8]，藏文的前加字共有 5 个，分别是ག(g), ད(d), བ(b), མ(m), འ(v)，古藏语文法认为前加字和时态变化、能所关系等有密切的关系。其中，前加字བ(b) 能和基字ཀ(k)，ཅ(c)，ཏ(t)，ཙ(ts)，ག(g)，ང(ng)，ཇ(j)，ཉ(ny)，ད(d)，ན(n)，ཛ(dz)，ཞ(zh)，ཟ(z)，ར(r)，ཤ(sh)，ས(s)相拼。前加字འ

---

[5] 另见格桑居冕先生的“藏语字性法与古藏语音系”一文（民族语文 1991 年，6 期）。

[6] 我们的数据库收录了《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》和《藏语文课本（小学 12 册、初中 6 册、高中 6 册）》等 6 本词典、字典和课本。

[7] 详细见我们即将编辑出版的《藏语方言调查字表》。

[8] 本文参考的藏文文法，主要是指上文列出的这些文献。

(v)能和基字ག(g), ཇ(j), ད(d), བ(b), ཛ(dz), ཀ(k), ཆ(ch), ཐ(th), ཕ(ph), ཚ(tsh)相拼。前加字ག(g)能和基字ཅ(c), ཏ(t), ཙ(ts), ཉ(ny), ད(d), ན(n), ཛ(dz), ཡ(y), ཟ(z), ཤ(sh), ས(s)相拼；前加字ད(d)能和基字ཀ(k), པ(p), ག(g), ང(ng), བ(b), མ(m)相拼。前加字མ(m)能和基字ཁ(kh), ཆ(ch), ཐ(th) , ཚ(tsh), ག(g), ཇ(j), ད(d), ཛ(dz), ང(ng), ཉ(ny), ན(n)相拼。前加字的前向环境是空位，基字和上加字是前加字的后向环境。前加字的环境反映了每个前加字在整个音位系统中的组合关系，而所有的环境又构成了前加字音位系统的结构。

五个前加字中བ(b)在单音节词中的出现频率最高有 947 次，前加字མ(m)的出现频率最低，只有 224 次，前加字འ(v)出现 776 次，前加字ག(g)出现 435 次，前加字ད(d)出现 249 次，前加字总共出现 2631 次，见表 3。前加字的出现频率反映了每个前加字在整个音位系统中的地位。

表 3 古藏语前加字音位出现频率表

序号	藏文	转写	频率
1	བ	b	947
2	འ	v	776
3	ག	g	435
4	མ	m	224
5	ད	d	249
合计			2631

从前加字的环境上看，由于本文讨论的结构和分布是在藏语 6273 个单音节内部，所以，前加字只有后向环境的分布，在藏文文法著作中一般都会讨论前加字和基字的组合规律，从来没有讨论和上加字的组合关系，但从实际的语音组合我们知道，上加字有 3 个，它们是ར(r), ལ(l), ས(s)，也知道上加字和基字有一定的组合规律，因此，前加字和上加字的组合情况和分布是可以推导出来的，详见下一节。从藏语音位结构可以看出，由于每个前加字不和同部位的基字相拼，因此，每个前加字所具有的环境不完全相同，前加字བ(b)有 15 种环境，前加字འ(v)有 10 种环境，前加字ག(g)有 11 种环境，前加字ད(d)有 6 种环境，前加字མ(m)有 11 种环境。从分析可以看出，前加字བ(b)的组合能力最强，可以和 15 个基字组合，而且出现频率也最高。这些环境构成了前加字和基字的分布关系，很显然，其中存在大量互补的分布。

3.2 上加字的音位结构、分布和负担

藏语有 3 个上加字，分别是ར(r)，ལ(l)，ས(s)。上加字和基字的组合关系在现有的文法著作中鲜有提及[9]，只能从现有的藏文词汇中归纳出来。其中上加字ར(r)可以和基字ག(k)，ཏ(t)，ཙ(ts)，ག(g)，ཇ(j)，ད(d)，བ(b)，ཛ(dz)，ང(ng)，ཉ(ny)，ན(n)，མ(m)相拼；上加字ལ(l)能和基字ག(k)，ཅ(c)，ཏ(t)，པ(p)，ག(g)，ཇ(j)，ད(d)，བ(b)，ཧ(h)相拼；上加字ས(s)能和基字ག(k)，ཏ(t)，པ(p)，ཚ(tsh)，ག(g)，ད(d)，བ(b)，ང(ng)，ཉ(ny)，ན(n)，མ(m)相拼。另外，上加字和前加字的组合关系在藏文文法著作中也很难找到，从我们的数据库中检索的结果来看，只有前加字བ(b)和上加字ར(r)，ས(s)，ལ(l)可以相拼。

三个上加字在单音节词汇中的出现频率都比较大，见表 4，其中上加字ས(s)出现的最多为 971 次，上加字ར(r)出现了 554 次，上加字ལ(l)出现了 223 次，总共 1748 次。从中可以看出上加字ས(s)的组合能力最强。另外，上加字ར(r)，ས(s)和ལ(l)和前加字བ(b)组合只形成 3 了个词，功能负担很小。

表 4 古藏语上加字及音位出现频率表

序号	藏文	转写	频率
1	ར	r	554
2	ལ	l	223
3	ས	s	971
合计			1748

从以上的组合关系可以看出，基字为上加字的后向环境，其中，上加字ར(r)有 12 种环境，上加字

[9] 上加字在文献中出现的很少，如，在《柱间史》(ཆོས་རྒྱལ་སྲིད་བཙན་སྒམ་པའི་བཀའ་ཆེམས་ཀ་ཤེལ་མ་)中有“而系足字及其他分支结构则完全是智慧的独创”的描写。另外，在《西藏王统记》(རྒྱལ་རབས་གསལ་བའི་མེ་ལོང་)中有“共母字有三个，即ར་ལ་ས་(原注：这三个字可以作所有系字的母字)”的记载。

ལ(l)有 10 种环境，上加字ས(s)有 11 种环境。除了基字བ(b)外，上加字和许多基字永远都不会相拼，形成为互补分布；前加字བ(b)可以和上加字ར(r)，ལ(l)，ས(s)相拼，其他的也都永远不能组合；从上加字的角度看，前加字བ(b)是上加字ར(r)，ལ(l)，ས(s)的前向环境，其组合能力比较小。从分析可以看出，上加字和基字的组合能力较强，但和前加字的组合能力就十分有限，它们可以是基字的一个小的分类。

3.3 基字的音位结构、分布和负担

在藏语里，30 个藏文辅音字母都可以做基字，基字的后向环境是下加字和元音，30 个基字和所有的元音都能相拼，因此，组合规律比较简单，是相似性分布。基字的前向环境是 3 个上加字和 5 个前加字，从数据库统计的结果看，有 26 个基字可以和前加字、上加字或前加字+上加字的组合相拼，其中，基字ཡ(y)可以和前加字ག(g)和上加字ལ(l)相拼；基字ཏ(t)可以和前加字ག(g)，བ(b)，上加字ར(r)，ས(s)，ལ(l)，前加字+上加字组合བར(br)，བས(bs)，བལ(bl)相拼；基字ཆ(ch)可以和前加字མ(m)，འ(v)相拼；基字ཙ(ts)可以和前加字ག(g)，བ(b)，上加字ར(r)，ས(s)，前加字+上加字组合བར(br)相拼；基字ཤ(sh)可以和前加字ག(g)，བ(b)相拼；基字ཧ(h)可以和上加字ལ(l)相拼；基字ཉ(ny)可以和前加字ག(g)，མ(m)，上加字ར(r)，ས(s)，前加字+上加字组合བར(br)，བས(bs)相拼；基字ཐ(th)可以和前加字མ(m)，འ(v)相拼；基字ད(d)可以和前加字ག(g)，བ(b)，མ(m)，འ(v)，上加字ར(r)，ས(s)，ལ(l)，前加字+上加字组合བར(br)，བས(bs)，བལ(bl)相拼；基字ཕ(ph)可以和前加字འ(v)相拼；基字ཚ(tsh)可以和前加字མ(m)，འ(v)相拼；基字མ(m)可以和前加字ད(d)，上加字 ར(r)，ས(s)相拼；基字ཞ(zh)可以和前加字མ(m)，འ(v)，上加字ར(r)，前加字+上加字组合བར(br)相拼；基字ཇ(j)可以和上加字ར(r)，ལ(l)，前加字མ(m)，འ(v)，前加字+上加字组合བར(br)相拼；基字ཞ(zh)可以和前加字ག(g)，བ(b)相拼；基字ཟ(z)可以和前加字ག(g)，བ(b)相拼；基字ས(s)可以和前加字ག(g)，བ(b)相拼；基字བ(b)可以和前加字ད(d)，འ(v)，上加字ར(r)，ས(s)相拼；基字ར(r)可以和前加字བ(b)相拼；基字ཁ(kh)可以和前加字མ(m)，འ(v)相拼；基字ག(g)可以和前加字མ(m)，ད(d)，འ(v)，བ(b)，上加字ར(r)，ས(s)，ལ(l)，前加字+上加字组合བར(br)，བས(bs)相拼；基字ན(n)可以和前加字ག(g)，མ(m)，上加字ར(r)，ས(s)，前加字+上加字组合བར(br)，བས(bs)相拼；基字ཀ(k)可以和前加字བ(b)，ད(d)，上加字ར(r)，ས(s)，ལ(l)，前加字+上加字组合བར(br)，བས(bs)相拼；基字ཅ(c)可以和前加字ག(g)，བ(b)，上加字ལ(l)相拼；基字པ(p)可以和前加字ད(d)，上加字ས(s)，ལ(l)相拼；基字ང(ng)可以和前加字མ(m)，ད(d)，上加字ར(r)，ས(s)，ལ(l)，前加字+上加字组合བར(br)，བས(bs)相拼。

表 5 古藏语基字音位出现频率表

序号	藏文	转写	频率	序号	藏文	转写	频率	序号	藏文	转写	频率
1	ཀ	k	496	11	ད	d	475	21	ཞ	zh	153
2	ཁ	kh	297	12	ན	n	185	22	ཟ	z	159
3	ག	g	654	13	པ	p	200	23	འ	v	35
4	ང	ng	147	14	ཕ	ph	238	24	ཡ	y	104
5	ཅ	c	156	15	བ	b	499	25	ར	r	120
6	ཆ	ch	135	16	མ	m	215	26	ལ	l	68
7	ཇ	j	125	17	ཙ	ts	203	27	ཤ	sh	157
8	ཉ	ny	226	18	ཚ	tsh	147	28	ས	s	308
9	ཏ	t	304	19	ཛ	dz	127	29	ཧ	h	130
10	ཐ	th	151	20	ཝ	w	22	30	ཨ	–	37

表 5 为 30 个基字及在单音节词中的出现频率，从表中可以看出，基字在单音节词汇中出现频率最高的是ག(g)为 654 次，最低的是ཝ(w)只有 22 次。很显然，基字和前加字、上加字以及前加字+上加字的组合在组合能力上差别较大，是很不均衡的，而基字却很重要，其中ག(g)在古藏语的音位系统中是最重要的一个音位，它的地位在后加字中也十分重要，详见下。

从前向环境上看，基字ཡ(y)有 2 种环境，基字ཏ(t)有 8 种环境，基字ཆ(ch)有 2 种环境，基字ཙ(ts)有 5 种环境，基字ཤ(sh)有 2 种环境，基字ཧ(h)有 1 种环境，基字ཉ(ny)有 6 种环境，基字ཐ(th)有 2 种环境，基字ད(d)有 10 种环境，基字ཕ(ph)有 1 种环境，基字ཚ(tsh)有 2 种环境，基字མ(m)有 3 种环境，

基字ཛ(dz)有 4 种环境，基字ཇ(j)有 5 种环境，基字ཞ(zh)有 2 种环境，基字ཟ(z)有 2 种环境，基字ས(s)有 2 种环境，基字བ(b)有 4 种环境，基字ར(r)有 1 种环境，基字ཁ(kh)有 2 种环境，基字ག(g)有 9 种环境，基字ན(n)有 6 种环境，基字ཀ(k)有 7 种环境，基字ཅ(c)有 3 种环境，基字པ(p)有 3 种环境，基字ང(ng)有 7 种环境。可以看出，基字不同的环境体现了基字在整个音位系统中的结构和功能。基字的后向环境主要是元音，因此，组合能力是完全相同的。总之，基字的前后环境构成了古藏语音节结构的核心。

3.4 下加字的音位结构、分布和负担

通常认为藏语有 4 个下加字，分别是ར(r)，ལ(l)，ཡ(y)和ྭ(w)。但从藏语的词汇上看，下加字ར(r)和ཡ(y)可以和ྭ(w)组合构成复合下加字，这样一共有 6 个。在结构上，下加字ཡ(y)可以和基字ཀ(k)，ཁ(kh)，ག(g)，པ(p)，ཕ(ph)，བ(b)，མ(m)相拼；下加字ར(r)可以和基字ཀ(k)，ཏ(t)，པ(p)，ཁ(kh)，ཕ(ph)，ག(g)，ད(d)，བ(b)，མ(m)，ས(s)，ཧ(h)相拼；下加字ལ(l)可以和基字ཀ(k)，ག(g)，བ(b)，ར(r)，ས(s)，ཟ(z)相拼；下加字ྭ(w)可以和基字ཀ(k)，ཁ(kh)，ག(g)，ཉ(ny)，ད(d)，ཚ(tsh)，ཞ(zh)，ཟ(z)，ར(r)，ལ(l)，ཤ(sh)，ས(s)，ཧ(h)相拼。

从表中可以看到，ར(r)，ལ(l)，ཡ(y)在单音节词汇中的出现频率分别为 854，226 和 688，而ྭ(w)只有 29 个。下加字ར(r)，ཡ(y)和ྭ(w)构成的下加字组合的词汇很少，其中ྲ(rw)只有 4 个词，而ཡྭ只有 2 个词，因此，我们认为这两个下加字的组合的地位需要进一步的研究和考证，最主要的它们在古代是否真正发音，因为，下加字ྭ(w)可能是和梵文有关，不一定有实际的音位功能，见表 6。

表 6 古藏语下加字音位出现频率表

序号	藏文	转写	频率	序号	藏文	转写	频率
1	ར	r	854	4	ལ	l	226
2	རྭ	rw	4	5	ཡ	y	688
3	ྭ	w	29	6	ཡྭ	yw	2

由下加字和基字的组合关系可以看出，下加字的后向环境是元音，下加字ར(r)，ལ(l)，ཡ(y)三个能和所有 5 个元音相拼；ྭ(w)，ྲ(rw)，ླྭ(lw)三个下加字组合只和元音ཨ(a)相拼，和另外 4 个元音不能组合。下加字的前向环境是基字，其中，下加字ཡ(y)有 7 种基字环境，下加字ར(r)有 11 种基字环境，下加字ལ(l)有 6 种基字环境，下加字ྭ(w)有 13 种基字环境，这些环境形成了下加字的音位结构和分布。

3.5 元音的音位结构、分布和负担

古藏语元音一共只有 5 个，因此，在单音节词中的出现频率很高，见表 7，其中元音ཨ(a)的出现频率最高为 1710 次，元音ཨི(i)的出现频率最低位 907 次，元音ཨོ(o)的出现频率为 1475 次，元音ཨུ(u)的出现频率为 1189 次，元音ཨེ(e)的出现频率为 992 次。从组合规律上看，元音能和所有的基字相拼，因此，元音的前向环境组合能力很强。基字ྭ(w)它只能和元音ཨ(a)结合，见上一节。

表 7 古藏语韵腹元音音位出现频率表

序号	声母	转写	频率	序号	声母	转写	频率
1	ཨ	a	1710	4	ཨེ	e	992
2	ཨི	i	907	5	ཨོ	o	1475
3	ཨུ	u	1189				

古藏语元音的后向环境是后加字和后加字+再后加字的组合，后加字有 10 个，其中，后加字ཝ(v)不发音，这样实际只有 9 个。另外，还有 3 个元音韵尾和 7 个后加字+后加字组合，总共 19 个，它们也都能和 5 个元音相拼，这样古藏语 5 个元音各有 19 种环境，形成相似性分布，详见下。

3.6 后加字和再后加字的音位结构、分布和负担[10]

藏文的后加字有 10 个，分别是ག(g)，ང(ng)，ད(d)，ན(n)，བ(b)，མ(m)，ཝ(v)，ར(r)，ལ(l)，ས(s)，其中后加字ཝ(v)不发音。元音韵尾有 3 个，分别是ོ(–o)，ུ(–u)，ི(–i)。再后加字有两个，分别是ས(s)和ད(d)，再后加字ས(s)能和ག(g)，ང(ng)，བ(b)，མ(m)四个后加字相拼，形成 4 个后加字+再后加字组合，它们分别是ངས(ngs)，གས(gs)，བས(bs)，མས(ms)，再后加字ད(d)能和后加字ན(n)，ར(r)，ལ(l)相拼，形成 3 个后加字+再后加字组合，它们分别是ནད(nd)，རད(rd)，ལད(ld)。藏文文法认为后加字和后加字+再后加字组合在语音、词汇和

[10] 由于再后加字只有两个，我们在此将它们放在后加字中一起讨论。

语法方面有重要作用。

表 8 古藏语后加字和元音韵尾音位出现频率表

序号	藏文	转写	频率	序号	藏文	转写	频率	序号	藏文	转写	频率
1	ག	g	957	6	མ	m	548	11	ི	i	230
2	ང	ng	833	7	ར	r	447	12	ུ	u	156
3	ད	d	453	8	ལ	l	412	13	ོ	o	20
4	ན	ng	475	9	ས	s	506				
5	བ	b	546	10	ཝ	v	37				

从表 8 可以看出，在单音节词中出现频率最高的后加字是ག(g)，一共出现了 957 次，出现频率最少的是后加字ང(ng) 为 412 次。元音韵尾ི(i)出现的频率最高为 230 次，ོ(o)出现的频率最低为 20 次，另外ཝ(v)在韵尾不发音。从表 9 中可以看出，再后加字ས(s)的出现频率很高，一共 1146 次，但再后加字ད(d)的出现频率却很低，只有 15 次。

表 9 古藏语再后加字音位出现频率表

序号	声母	转写	频率
1	ས	s	1146
2	ད	d	15

从以上的结构可以看出，后加字和元音韵尾一共是 13 个，这 13 个韵尾可以和所有元音相拼，各有 5 个环境，形成相似性分布。再后加字ས(s)能和ག(g)，ང(ng)，བ(b)，མ(m)相拼形成 4 种组合，其余不能组合。再后加字ད(d)能和ང(ng)，ལ(l)，ར(r)相拼形成 3 种组合，其余不能组合。元音韵尾能和所有元音组合。由于本文只讨论单音节词内部的音位结构和分布，后加字和后加字+再后加字组合的后向环境是空位。

4 基于结构和分布的音位负担及冗余度分析

在第四节分别分析了古藏语前加字、上加字、基字、下加字、元音、后加字和再后加字音位的结构、分布和负担，本节将从古藏语音位的结构和分布来讨论音位体系的基本信息框架。表 10 给出了能出现在不同位置的古藏语辅音音位的频率表，其中不包括那些只出现在基字位置的出现频率，其数据可以从以上的表中找到。从表中的数据可以看到音位通过结构和分布体现出功能，例如ཝ(v)在基字和后加字位置上只分别出现了 35 次和 37 次，这个数据量在整个系统中是很小的，显得微不足道，但在前加字位置上ཝ(v)却出现了 776。我们知道ཝ(v)在作韵尾时是不发音的，只是用于确定前边的字母是否是基字。另外，整体上看前加字和后加字的结构和分布同样体现了古藏语形态音位功能的在整个系统的重要性，也体现了藏语音位体系的一个更深层次的功能体系。

表 10 古藏语辅音音位出现频率表

序号	藏文	转写	基字	前加字	上加字	下加字	后加字	再后加字	汇总
20	ཨ	w	22			35			57
12	ན	n	185				475		660
24	ཡ	y	104			688			792
23	ཝ	v	35	776			37		848
26	ལ	l	68		223	226	412		929
4	ང	ng	147				833		980
16	མ	m	215	224			548		987
11	ད	d	475	249			453	15	1192
25	ར	r	120		554	854	447		1975
15	བ	b	499	947			546		1992
3	ག	g	654	435			957		2046
28	ས	s	308		971		506	1146	2931
汇总				2631	1748	1803	5214	1161	

表 11 是古藏语元音音位的分布，从数据上看元音系统比较简单，除了五个元音外，有三个元音可以作韵尾，在功能上语音韵尾来源于语素附加成分元音缩减，而且在整个体系中出现的次数很少。

表 11 古藏语元音音位出现频率表

序号	藏文	转写	元音	韵尾	汇总
1	ཨ	a	1710		1710
2	ཨི	i	907	230	1137
3	ཨུ	u	1189	156	1345
4	ཨེ	e	992		992
5	ཨོ	o	1475	20	1495
合计			6273	406	6679

根据以上的数据和分析，我们来讨论一下根据音位结构、分布和负担所体现出来的基本的音位信息体系。在结构主义的音位学中，最为主要的原则是对立原则和互补原则，但这种原则不能完全反映出语音音位系统的信息体系和功能体系，从信息量的角度来看，古藏语有 35 个音位，单音节长度最长的有 7 个音位，理论上最大可能的组合为 35 的 7 次方（64339296875），根据我们的统计，6273 个单音节词中音位共出现了 25509 次，因此，每个实际的音位所占有的理论上的信息空间为 2522219（64339296875 / 25509），即一个音位有 2522219 可选择的位置，在此我们称其为“无条件制约”或“零级制约”，很显然在零级制约条件下，音位的冗余度很大。

从藏语音位的结构来看，前加字能出现 5 个音位，上加字能出现 3 个音位，基字能出现 30 个音位，下加字能出现 6 个音位，元音 5 个音位，后加字 10 个音位加上 3 个元音韵尾共 13 音位，再后加字能出现 2 个音位，这些数字相乘为 351000，将其除以实际音位的出现数为 13.7598（351000 / 25509），我们将这种制约称为“结构制约”或“一级制约”，很显然，在结构制约条件下，一个音位能选择的信息空间约有 14 个，和理论上的信息空间相比已经大大降低了，音位的冗余度也大大降低。从上一节的数据分析可知结构制约的本质是音位的结构，结构在规则上一般都很简单明了，易于建立模型，而且功能显著。从 2522219 降到 13.7598 只用 7 条规则即可说明。

上文的数据分析描写了相邻两个音位的组合关系，见第四节，根据这些关系我们可以得到这样的分布空间：1）前加字-上加字之间有 3 种组合形式；2）前加字-基字之间有 54 中组合形式；3）上加字-基字之间有 33 种组合形式；4）基字-下加字之间有 37 种组合形式；5）基字-元音之间有 150 种组合形式；6）下加字-元音之间有 16 种组合形式；7）元音-后加字之间有 65 种组合形式；8）后加字-再后加字之间有 7 种组合形式，见图 2。

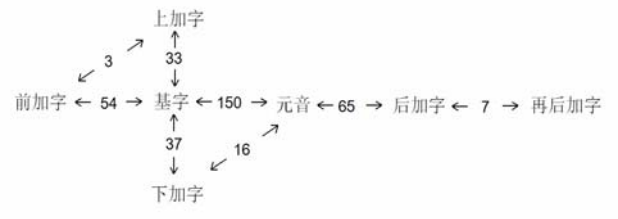


图 2 古藏语相邻音位组合

这些组合关系进一步对具体的音位进行了制约，在此我们将其称为“相邻组合关系制约”或“二级制约”。 与一级制约相比每一个音位在音位系统种的可选信息空间会进一步降低，由于存在交叉组合关系，怎样计算这种制约所形成的音位冗余度或者信息量，还需要进一步研究，在此我们先用两个相邻音位组合关系系数来描写一下二级制约对音位信息空间的制约情况。前加字和上加字之间的理论组合为 15，而实际组和为 3，其系数为 0.2。上加字和基字之间的理论组合为 9 0，实际组和为 3 3，其系数为 0.3667。前加字和基字之间的理论组合为 150，实际组和为 54，其系数为 0.36。下加字和基字之间的理论组合为 180，实际组和为 37，其系数为 0.0206。基字和元音之间的理论组合为 150，实际组和为 150，其系数为 1。下加字和元音之间的理论组合为 30，实际组和为 16，其系数为 0.5333。元音和后加字之间的理论组合为 65，实际组和为 65，其系数为 1。后加字和再后加字之间的理论组合为 26，实际组和为 7，其系数为 0.2692。这样平均的组合系数为 0.4687，用结构信息总空间 351000 乘以这个系数得到相邻组合关系制约的信息总空间 164513.7，然后再用这个数除以音位实际出现的次数 25509 得到每个音位的信息空间 6.4492，即每个音位有约 6.5 个位置可选择。这个数显然小于一级制约的 13.7598。

从上图看，“相邻组合关系制约”需要用 8 条或者 8 组规则描写，还不算太复杂，也可以进一步确定每个音位在系统中的性质和功能，然而，这种相邻组合关系并不能体现出非相邻音位的组合关系，例如，前加字和基字有 54 种组合关系，但并不能推测出这 54 种组合和 37 种下加字的组合关系。如果要建立整个音位系统中每个音位组合关系

的制约规则，在音素音位体系下，整个系统会十分复杂。因为每个音位的结构和分布要被精确描写，古藏语大概要用上千条规则才会使其冗余度会等于 1（25509/25509）。在此我们将这种制约称为“系统组合关系制约”或“三级制约”。

从古藏语的音位结构和分布可以看出，音位的位置不同，功能和作用也不同，有的只具有普通音位的作用，而有的具有语素音位的作用，其音位的负担量也差别很大，如，前加字出现频率最高的是ད(b)为 947，出现频率最小的是མ(m)为 224；上加字出现频率最高的是ས(s)为 971 ，出现频率最小的是ལ(l)为 223；基字出现频率最高的是 ག (g) 为 510，出现频率最小的是ཝ(w)为 22；下加字出现频率最高的是ར(r)为 854，出现频率最小的是ཡཝ(yw)为 2；元音出现频率最高的是ཨ(a)为 1710，出现频率最小的是ཨི(i)为 907；后加字出现频率最高的是ག(g)为 957，出现频率最小的是ཨོ(o)为 20；再后加字出现频率最高的是ས(s) 为 1146，出现频率最小的是ད(d)为 15。虽然古藏语音位负担的功能和分布有很大不同，但音位相互之间却有很好的组合十分规律，这是因为古藏语在组合形式上不是任意的，而有字性（语音发音生理的性质）的制约[11]。

从以上古藏语的音位体系可以看出，虽然古藏语音位系统的结构和分布有规律可循，但相当复杂。按照音素音位系统来计算古藏语的功能负担，虽然每个音位在系统中的地位已经十分清楚，但整个系统十分庞大。我们在此讨论的只是音位的结构、分布和负担的问题，还没有涉及古藏语字性（藏语语音组合相拼的语音性质和制约关系）的问题，如果涉及到字性，整个音位系统更复杂，而且要结合方言和藏语历史音变才能进行讨论。从以上的分析可以得出的结论是：音位系统内部结构、分布和负担的研究能使我们更深入地认识古藏语音位系统的本质。

5 音位结构和分布与功能负担的讨论[12]

在了解了古藏语音位的结构和分布后，我们想进一步知道这样复杂的系统用什么方法能更好地来描写、量化和反映一个语言共时和历时的本质和变化规律。下面我们将从音素音位系统、声韵调音位系统和藏语传统音位系统三个方面来进行一些讨论。

5.1 音素音位系统与功能负担

按照音素音位的方法，古藏语音位系统一个音节最长可以有 7 个语音单元，即，前加字、上加字、基字、下加字、元音、后加字和再后加字。按照功能负担通常的计算方法，两个语音单元之间的关系或环境都会体现的计算的结果中。按照双音子来计算，在一个藏语音节中最多就会出现 8 种组合：1）“空位+前加字”；2）“前加字+上加字”；3）“上加字+基字”；4）“基字+下加字”；5）“下加字+元音”；6）“元音+后加字”；7）“后加字+再后加字”，8）“再后加字+空位”。按照三音子来计算，在一个藏语音节中最多就会出现 7 种组合：1）“空位+前加字+上加字”，2）“前加字+上加字+基字”；3）“上加字+基字+下加字”；4）“基字+下加字+元音”；5）“下加字+元音+后加字”；6）“元音+后加字+再后加字”，7）“后加字+再后加字+空位”。

现代语音识别和语音合成技术都运用了功能负担的理论，其中最常用的是双音子和三音子的建模方法，这种方法简单，但运算量大，比较适合计算机大存储量和高速运算的特点，然而，这种方法和藏语的事实有一定的差距，例如，前加字主要具有某些语法功能，它主要和基字有一定的分布关系，但按照双音字和三音子的方法，计算的结果中暗含了前加字和上加字的信息和分布关系，这些信息是否是有用的信息或者说对什么有用还需要进一步研究，但从语言学的角度看，这些信息和藏语的语言实际有一定的差距。从表面上看对语音识别和合成技术来说是可行的一种方案，但不是最佳的方案。

从现代藏语方言的实际读音来看（瞿霭堂，谭克让，1983；华侃，2002；格桑居冕，2002），有的前加字和上加字已经不是独立的音素了，如，安多方言中，前置辅音ད(b)在很多情况下只保留了双唇的特征（孔江平，1991），另外，下加字在现代藏语方言中也大多和基字合成了一个音素。从以上的分析可以看出，基于音素音位功能负担的计算，虽然音位的实际数目小了很多，但计算量和音位之间的相互关系极为复杂和冗余，这在语言学和信息技术层面都不是最合理的。

5.2 声韵调音位系统与功能负担

单音节是汉藏语系语言的一个基本语音特征，

[11] 传统藏语十分重视字性的研究，有许多古代文献讨论这个问题，字性反映的是语音生理在音位组合中的制约性和组合规律。

[12] 我们对声韵调音位系统和传统藏语音为系统也做了音位结构和分布的计算，但由于内容过多，三种计算结果的对比研究将另文讨论，在此只进行一些方法上的讨论。

和其相对应的语音单位是声韵调，如果用这种语音单位来研究藏语的功能负担，就会产生和音素音位不同的结果，例如，声母出现频率最高的是ཐ(th)为 72，出现频率最小的为 1 个，如ཁྷ(khw)等；韵母出现频率最高的是ཨ(a)为 210，出现频率最小的为 1 个，如ཨནྡ(and)等，这和音素音位计算出来的负担是不同。另外，在功能上藏文的前加字、后加字和再后加字都有一定的语法功能，完全按照声母和韵母这样的语音单位来考虑古藏语功能负担的计算和理论框架会有问题，因为前加字会被包括在声母中，后加字和后加字+再后加字的组合会被包括在韵母中。藏语的 5 个前加字和基字有一定的组合关系，但不是所有的 5 个前加字和所有的基字都能组合，如果用声母和韵母作为计算功能负担的单位，这些组合关系就都被包含在声母中。后加字和后加字+再后加字的组合都具有语法功能，如果运用韵母作为语音单位来计算功能负担，作为语素音位的形式被包含在了韵母中，虽然语素音位的功能信息量没有丢失，但解释性就会比较差。从而在方法论上又从科学（内在规律）层面退到了技术层面（统计模型）。

从另一个方面来看，采用声韵调作为语音单位对研究声调功能负担和从超音段功能负担研究声调起源也许会比较方便，因为采用的语音单位比较简单，因此，方便计算超音段特征，但我们知道声调的来源有时可能只是来源于声母的一个音素或者特征，而不是整个声母，这就需要细化到音素和区别性特征，例如，藏语声调的高低和声母的清浊有直接关系，但又和上加字ས(s)有关，采用声母、韵母和声调作为音位的单位又不能详细反映其语音历史音变的特点。

5.3 古藏语音位系统与功能负担

根据古藏语音位系统和音位功能的实际，音位的功能负担单位可以采用“前加字”、“基字丁（上加字+基字+下加字）”、“元音”和“韵尾（后加字、元音韵尾和后加字+再后加字）”4 个音位单元。如果采用双音子的方法计算，在词汇内部一共有 5 种双音子组合：1）“空位+前加字”，2）“前加字+基字丁”，3）“基字丁+元音”，4）“元音+韵尾”，5）“韵尾+空位”。如果采用三音子的方法，一共有 4 种三音子组合：1）“空位+前加字+基字字丁”，2）“前加字+基字字丁+元音”，3）“基字字丁+元音+韵尾”，4）“元音+韵尾+空位”。

根据这种语音单位来计算功能负担，其结果和上面两种会有很大不同，我们在前面已经列出了前加字、上加字、下加字、元音、后加字和再后加字的负担，这里我们给出基字丁和韵尾的负担，基字丁出现频率最高的是ད(d)为 200，出现频率最小的为 1 个，例如ཟྭ(rw)；韵尾出现频率最高的是ག(g)为 957，出现频率最小的是ནྡ(nd)为 5。根据传统藏文音位的定义和计算结果比较符合藏语的语言实际，既保留了语素音位的独立性，又简化了整个音位系统，它在音位系统归纳和研究上由一定的理论意义。同时，对藏语文本的处理、语音合成和语音识别系统的建立都有很重要的参考价值。

从上面的讨论可以看出，音位单位和音位功能负担的计算有密切的关系，其结果也会有很大差异。这种结果和差异在共时层面比较容易看出，也可以将其应用在藏语信息技术的开发和应用上。但在历时语言学研究层面，直到目前我们还无话可说，这方面还需要进一步的研究，例如，用什么样的音位单位和什么样的计算方法，能够较好地反映出藏语历史音变的轨迹和本质，这是我们语言学家最想知道和为之而努力的。

6 结束语

本文从音位结构、音位分布和音位功能负担的角度，以古藏语音位系统为研究对象，分析了古藏语音位系统的内部结构、分布和负担，提出了“音位结构功能负担”的研究思路，从而勾画出了古藏语音位系统结构、分布和负担的基本框架，从这个框架可以看出，古藏语音位的结构、分布和负担十分复杂，构成了一个庞大而复杂的系统，音位之间有分布的差别、功能的差别和负担量的差别，它们反映出了每个音位在一个系统中有着各自的地位，这种系统用传统音位学是很难解释的，但这种复杂的系统在计算机上进行统计学的或语言学的建模和实现却不困难。文章还讨论了古藏语音位的结构制约、相邻组合关系制约和系统组合关系制约，很显然，这些音位的不同制约反映了藏语音位在结构、分布和负担上的基本特性，对于信息的承载和传递有重要意义。文章最后从音素音位系统、声韵调音位系统和藏语传统音位系统三个方面，讨论了音位功能负担的差异和语言学的意义。研究表明，计算古藏语的音位功能负担首先要考虑语言的具体性质，从而确定计算的单位。从语言功能负担的本质来看，采用功能负担对一种语言进行研究，首先要明确研究的目的和性质，这样才能确定研究功

能负担的单位，进而得到理想和正确的结论。古藏语到现代方言的基本音变情况主要是复辅音声母脱落后逐渐产生声调，显然结构和分布构成的音位负担在其中起了很重要的作用，以古藏语的音位系统来做音位的结构和分布的研究是想探讨和构建语言音位功能负担的一个基本理论框架，从而为藏语的各个方言音位结构负担的研究奠定可行的基础，其目的是为了从音位功能负担研究的角度探讨古藏语的每一个音位到现代方言的变化脉络，这才是我们的最终目标，但怎样从理论上解释还需要进一步建立大规模的方言数据库，需要精确的音位负担量的计算和历史语言学的研究，最终来验证音位功能负担是否是语言历史音变的一种主要原动力。科学技术的进步不仅对语言学提出了更高的要求，也开辟了更多新的领域，相信语言功能负担的研究会在共时语言学、历时语言学和语言信息科学方面有新的贡献。

参考文献

[1] 《司徒文法详解》，噶玛司徒，青海人民出版社出版了，1984 年

[2] 《藏语语法》，西北民族学院，语文室，藏语组，1958 年 5 月，油印本

[3] 《阿里藏语》，瞿霭堂 谭克让，中国社会科学出版社，1983 年 2 月

[4] “道孚藏语双擦音声母的声学分析” 孔江平，1991，《民族语文》第 3 辑

[5] 《简明藏文文法》，胡书津，云南民族出版社，1995 年 12 月

[6] 《藏语安多方言词汇》，华侃，甘肃人民出版社，2002 年 6 月

[7] 《藏语研究论文》，胡坦，中国藏学出版社，2002 年 12 月

[8] 《藏语方言概论》，格桑居冕，格桑央京，民族出版社，2002 年 7 月

[9] 《实用藏文文法教程》（修订本），格桑居冕，格桑央京，四川出版集团，四川民族出版社，2004 年 11 月

[10] Vilem Mathesius. 1929. La Structure Phonologique du Lexique du Tch`eque Moderne. Travaux du Cercle Linguistique de Prague, 1:67–84.

[11] Jakobson R. 1931. Prinzipien der Historischen Phonologie. Travaux du Cercle Linguistique de Prague, 4:246–267.

[12] Trubetzkoy N. 1939. Grundz¨uge der Phonologie. Travaux du Cercle Linguistique de Prague, 7.

[13] Shannon, C.E. and Weaver, W.: The mathematical theory of communication (University of Illinois Press, Urbana 1949).

[14] Shannon, C.E.: Prediction and entropy of printed English. Bell System Technical Journal, 30: 50--64 (1951).

[15] Hockett, C.F.: A Manual of Phonology; p.218 (Baltimore Waverly Press, 1955).

[16] Greenberg, H.H.: A method of measuring functional yield as applied to tone in Mrican languages. Georgetown University Monograph Series on Languages and Linguistics 12: 7-16 (1959).

[17] Hoenigswald, H.M.: Language change and linguistic reconstruction, pp.79-80 (University of Chicago Press, Chicago 1960).

[18] Hockett, C.F.: The quantification of functional load. Rand Report, p.2338, Santa Monica (1961).

[19] Kucera, H.: Entropy, redundancy and functional load. American Contributions to the Fifth International Congress of Slavists, pp. 191-219, Sofia (1963).

[20] King, R.D.: Functional load: its measure and its role in sound change. University of Wisconsin PhD dissertation (1965). A version of this will appear in 'Language'.

[21] King, R.D.: 1967. Functional Load and Sound Change. Language, 43.

[22] King, R.D.: 1967. A Measure for Functional Load. Studia Linguistica, 21:1–14. Ioannis

[23] William S-Y. Wang. 1967. The Measurement of Functional Load. Phonetica, 16:36–54.

[24] Surendran D. and Niyogi P. 2003. Measuring the Usefulness (Functional Load) of Phonological Contrasts. Technical Report TR-2003-12., Department of Computer Science, University of Chicago.

[25] Surendran D. and Levow G-A. 2004. The Functional Load of Tone in Mandarin is as High as that of Vowels. In Proceedings of the International Conference on Speech Prosody 2004, pages 99–102, Nara, Japan.

（孔江平 北京大学中国语言文学系语言学实验室

100871 kongjp@gmail.com）

武定彝语松紧音研究*

A Study of Lax/tense Voice in the Wuding Yi

汪锋 孔江平

Wang Feng Kong Jiangping

摘要：本文从实验语音学的角度对武定彝语中的松紧音进行了分析，探讨了松紧对立在音高、时长、谐波振幅、共振峰、开商和速度商等参数上的表现，发现在不同声调中松紧对立的实质并不相同，在 1 调中主要区别在于发声状态，即紧喉音与普通嗓音的不同；而在 3 调中，对立的基础主要是时长的差异，即短促调与舒声调的区别。据此，本文认为：在音系上可以统一处理为“松紧”的对立并不蕴含语音实质上的同一。

Abstract: This paper conducts an experimental phonetic study of lax/tense voice in the Wuding Yi. The representations of the lax/tense distinction in different parameters such as pitch, duration, amplitude, formant, open quotient and speed quotient, have been analyzed. It is found that the qualities of lax/tense distinction under different tonal conditions are not the same. Under the tone 1, the distinction lies on difference of phonation types, i.e. creaky vs. modal voice; While under the tone 3, the distinction bases on duration differences of tones. Therefore, it should be noted that the ‘lax/tense’ contrast in phonology does not imply the same phonetic basis.

关键词：松紧音 发声类型 武定彝语

Keywords: Lax/tense voice, phonation type, the Wuding Yi

0 绪言

对中国境内少数民族语言中的特殊发声类型的注意早在 20 世纪 50 年代就有了，例如：马学良(1948)、胡坦、戴庆厦(1964)、戴庆厦(1979)等。从马学良(1948)开始，普通嗓音与特殊发声类型的区别就被概括为“松紧音”的对立，早期的研究主要集中在音系描写和历史比较方面，对特殊发声类型的物理特性和发声的生理机制尚未涉及，因此，“松紧音”在语音实质上对应物是否相同，是否存在跨语言的变异等问题都尚不清楚。国外率先开始了一些实验语音学和声学的研究(Maddieson and Ladefoged 1985)，但仍是零星研究，直到 20 世纪 90 年代，中国社会科学院民族所语音实验室等才开始了系统的研究(孔江平 2001)。

本文主要关注武定彝语中松紧音的语音实质。武定彝语在《彝语简志》中划归为彝语东部方言 (陈士林 等 1985)。本文以云南省楚雄彝族自治县武定县发窝乡大西邑村为代表点，两位发音人均为中年男性。

1 彝语武定话音位系统

根据《藏缅语族语音词汇》，武定彝语的音位系统如下：

声母 (46)

p	p^h	b	m	f	v		
t	t^h	d	n		l	ɬ	
ʈ	$ʈ^h$	ɖ					
k	k^h	g	ŋ	x	ɣ		h
ts	ts^h	dz		s	z		
tʂ	$tʂ^h$	dʐ	ɳ	ʂ	(ʐ)		
tɕ	$tɕ^h$	dʑ	ɲ	ɕ	j		
mp^h	nts^h	nt^h	$ɳtʂ^h$	$ɲtɕ^h$	$ŋk^h$		

韵母 (20)

i	e	ɑ	ɒ	ɔ	u	ʏ	ɤ	ɯ	ɿ
i̱	e̱	ɑ̱	ɒ̱	ɔ̱	u̱	ʏ̱	ɤ̱	ɯ̱	ɿ̱

“说明： 1. e̱的实际音值为ɛ̱；2. ɿ ɻ与舌尖后音相

* 本研究得到国家社科基金项目[#07CYY025]、中国博士后科学基金[#20070420555]和国家自然科学基金[#60773159]的资助。本研究得到云南的各位发音合作人和朋友们的大力支持，他们是陆燕、朱学溢、杨国溢、李永宏、邝剑菁。李永宏为本研究编写了诸多 Matlab 程序，谨此一并致谢。本文所余错漏概由作者负责。

拼为ɿʅ; 3. u u̱的实际音值为口腔略开的ʊ ʊ̱。”(《藏缅语族语音词汇》)在本研究核对材料后，没有找到在ɑ和ɒ上的松紧最小对立，前者通常在紧音中出现，后者在松音中出现，也没有发现v与v̱的最小对。但发现了在ɚ上有松紧的对立，在以上归纳中没有提到。

声调 (4)

55　33　2　11

“说明：2 为短调。主要出现紧元音韵母音节中；11 调只出现在松元音韵母音节中。2 调和 11 调互补，也可以合并为一个调。” (《藏缅语族语音词汇》) 因此，可以将声调系统处理为：1 调 [55]；2 调[33]；3 调[2]和[11]。

2 实验材料

在 2 调上没有发现松紧音的对立，本研究包括 7 对在 1 调上的松紧对立，7 对在 3 调上的松紧对立，具体情况如下：

	汉义	语音	汉义	语音
i	七	ɕi1		
i̱	新	ɕi̱1		
ɿ	撒(种)	ʂɿ1		
ʅ	拧	ʂʅ1		
e			漏(水)	dze3
e̱			忌嘴	dze̱3
ɒ	菜	ɣɒ1	瓦	ŋɒ3
ɑ	编	ɣɑ1	鸟	ŋɑ3
ɔ			北	kʰɔ3
ɔ̱			勒	kʰɔ̱3
ɤ	喉结	lɤ1 (tʰɤ 33)	收割	kɤ3
ɤ̱	锥子	lɤ̱1	样子	kɤ̱3
u	裤子	ɬu1	老虎	lu1
u̱	放牧	ɬu̱1	够	lu̱1
			布	pʰu3
			翻身	pʰu̱3
ɯ	下	kɯ1	告状	tɯ3
ɯ̱	会(写)	kɯ̱1	把(米)	tɯ̱3
ɚ			辣	pʰɚ3
ɚ̱			剥	pʰɚ̱3

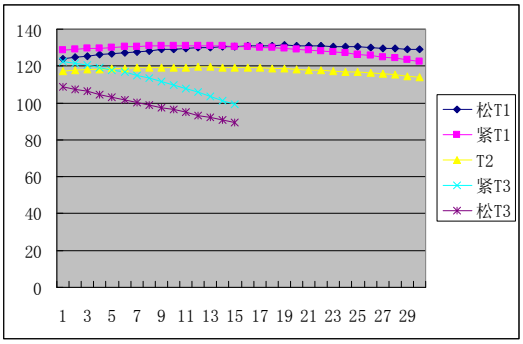
本研究同时采集了上述实验材料的语音信号和嗓音信号，嗓音信号是通过喉头仪采集的，利用 Kay 公司的 Real Time EGG 软件提取开商、速度商等参数。

3 参数分析

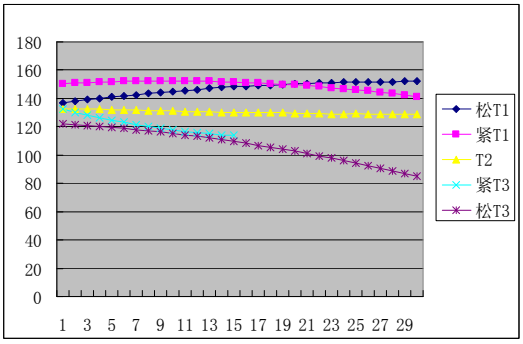
3.1 基频

根据上述材料，声调与松紧的搭配，可以得到 5 组：松 T1、紧 T1、T2、紧 T3、松 T3；每组选取 5 个样本，每个样本发音两遍。长调基频选取 30 个点，而短调为 15 个点，以自相关方法提取。数据提取以自行编写的 Matlab 程序实现。二位发音人的数据分别平均，基频情况如下：

男 1



男 2

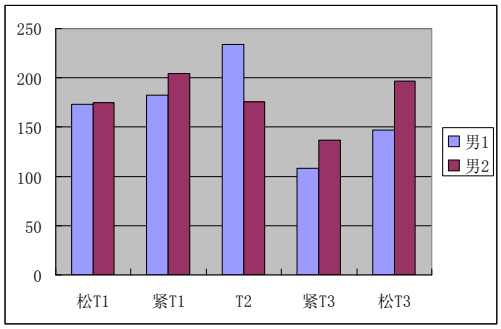


在 1 调上，紧音的基频在前半段略高，在后半段低下去；而如果以 5 度标调，则大致上紧T3 可为 42，而松T3 可为 31，也就是，在此对立中，紧元音在基频上比松元音略高。

3.2 时长

	男 1	男 2
松 T1	173.3	175.2
紧 T1	182.1	204.4

T2	233.75	175.375
紧 T3	107.8	136.9
松 T3	146.7	197.1



3.3 谐波

为了全面考察谐波的特征，每个音节提出 30 个点，这样可以观察在整个韵母部分上声带状态变化的情况。孔江平 (2001：103)在研究阿细彝语时指出：h2/h1 的计算方法能反映松紧元音的相对值，而且便于进行不同元音之间松紧的比较研究。

在 1 调情况，男 1 数据如下（限于篇幅，只列出第 13-23 点，下同）：

相比较而言，紧 T3 明显是最短促的。

语音	汉义	13	14	15	16	17	18	19	20	21	22	23
ɕi1	七	1.04598	1.04598	1.03448	1.02273	1.04598	1.04598	1.04598	1.04598	1.05814	1.05814	1.07059
ɕi1		1.05814	1.05814	1.07059	1.07059	1.07059	1.08333	1.08333	1.08333	1.08333	1.07143	1.05952
ɕi̱1	新	1.17073	1.17073	1.18519	1.17284	1.17284	1.18750	1.17284	1.17284	1.16049	1.17500	1.14815
ɕi̱1		1.15190	1.16667	1.15385	1.15385	1.15385	1.18421	1.11392	0.97701	0.93258	0.95402	0.95402
ɬu1	裤子	1.04494	1.04494	1.04494	1.03371	1.02247	1.02247	1.02247	1.02247	1.01124	1.01124	1.01136
ɬu1		1.06667	1.08989	1.10227	1.09091	1.09091	1.10345	1.10345	1.09195	1.09195	1.10465	1.10465
ɬu̱1	放牧	1.13253	1.12048	1.10843	1.09524	1.08333	1.07143	1.08434	1.08434	1.10976	1.11111	1.23457
ɬu̱1		1.10976	1.11111	1.09877	1.09877	1.11392	1.04348	1.01020	1.02062	1.04301	1.10000	1.10000
kɯ1	下	1.10989	1.10989	1.12222	1.13333	1.12222	1.12222	1.09890	1.11111	1.10000	1.10000	1.10112
kɯ1		1.07692	1.07692	1.06593	1.05495	1.05556	1.05556	1.05556	1.05556	1.06742	1.06742	1.06742
kɯ̱1	会(写)	1.22093	1.22093	1.23529	0.98980	0.88571	1.25000	1.21176	1.22619	1.21429	1.20238	1.19048
kɯ̱1		1.12222	1.13333	1.13333	1.13333	1.13333	1.13333	1.12222	1.11111	1.12360	1.11236	1.10112
lʏ1	喉结	1.05556	1.05556	1.05556	1.04444	1.03333	1.03333	1.03333	1.04494	1.04494	1.03371	1.03371
lʏ1		1.05747	1.05747	1.05747	1.06977	1.06977	1.05814	1.07059	1.05882	1.07143	1.05952	1.06024
lx̱1	锥子	1.07229	1.07229	1.08434	1.10976	1.10843	1.10843	1.10843	1.10843	1.10843	1.10843	1.09639
lx̱1		1.05882	1.05952	1.05952	1.05952	1.05952	1.07229	1.07229	1.07229	1.07229	1.07229	1.07229
lu1	老虎	1.09195	1.09195	1.09195	1.07955	1.06818	1.06818	1.06897	1.06897	1.05747	1.05747	1.05747
lu1		1.13953	1.16279	1.12791	1.07059	1.05882	1.07059	1.07143	1.05952	1.05952	1.05952	1.07229
lu̱1	够	1.09412	1.09412	1.09412	1.09412	1.08235	1.09524	1.09524	1.09524	1.09524	1.09524	1.08333
lu̱1		1.10843	1.10976	1.09756	1.13415	1.19512	1.20988	1.22222	1.17500	1.08750	1.08750	1.08750
ʂʅ1	撒(种)	1.04706	1.04706	1.03529	1.03529	1.02353	1.02353	1.01176	1.01176	1.01190	1.01190	1.01190
ʂʅ1		1.08434	1.08434	1.07143	1.07143	1.08434	1.07229	1.07229	1.07229	1.07229	1.07229	1.06024
ʂḻ1	拧	1.15854	1.17073	1.17073	1.18519	1.17284	1.16049	1.16049	1.14815	1.14815	1.13580	1.12346
ʂḻ1		1.14815	1.14815	1.13415	1.13415	1.12195	1.12195	1.12195	1.12195	1.12195	1.12346	1.12346
ɣɒ1	菜	1.04598	1.04651	1.03488	1.02326	1.01176	1.01190	1.01205	1.01205	1.01205	1.02410	1.03614
ɣɒ1		1.03529	1.03488	1.04706	1.04706	1.04706	1.04706	1.04651	1.04494	0.97802	0.93333	1.01163
ɣo̱1	编	1.06897	1.05747	1.06977	1.05814	1.03488	1.02326	1.02326	1.02326	1.02326	1.01163	1.01163
ɣo̱1		1.07143	1.07143	1.07143	1.05952	1.05952	1.07229	1.06024	1.07317	1.07317	1.06098	1.07407

这些数据非常明显的显示在武定彝语的 8 对松紧音中，紧音的h2/h1 要大于对应的松音。但要注意的是，其中有一些小的变化，(1) ɕi1‘新’第 2 个样本的第 20-24 点，比值突然降低，比对应的松音都要低；kɯ1‘会(写)’ 在第 16-7 点和第 26-30 节中，le1‘锥子’第 2 个样本，在 16-7 点；lu1 在 13-15 点；也有类似情况；(2) ɬu1‘放牧’的两个音节中，在第 17-26 点出现了比值比松音低的情况；ɣɑ1 在第 17-30 点；都有类似情况。

变异情况(1)应该属于标记或者计算中的问题，可以忽略不计，这一点可以通过其他数据来校正，详见后文。变异情况(2)则很可能说明该紧音开始不稳定了，从贯穿整个韵，变到部分韵段上已经不具有这一区别特征了，基本上都从中段以后开始不稳定，或许这是“紧”特征消失的一个基本规律。

男 2 的数据如下：

语音	汉义	13	14	15	16	17	18	19	20	21	22	23
ɕi1	七	1.06593	1.06593	1.06593	1.06593	1.07778	1.07778	1.07778	1.07778	1.06667	1.07865	1.07865
ɕi1		1.02128	1.03226	1.03226	1.03226	1.03226	1.03226	1.03226	1.03226	1.03226	1.04348	1.03261
ɕi1	新	1.16279	1.16471	1.16471	1.16667	1.16667	1.15476	1.15476	1.14286	1.13095	1.11905	1.10714
ɕi1		1.16667	1.15476	1.15476	1.14286	1.13095	1.12048	1.12195	1.09756	1.08537	1.08537	1.09877
ɬu1	裤子	1.1	1.1	1.08889	1.08889	1.08889	1.10112	1.08989	1.07865	1.07865	1.07865	1.06742
ɬu1		1.09091	1.10227	1.10227	1.10227	1.09091	1.09091	1.09091	1.10345	1.10345	1.10345	1.09195
ɬu1	放牧	1.14458	1.13095	1.13095	1.13095	1.13095	1.13095	1.11905	1.11905	1.10714	1.09524	1.08235
ɬu1		1.07059	1.07059	1.07059	1.07059	1.07059	1.07059	1.09524	1.09524	1.09524	1.09524	1.10714
kɯ1	下	0.87805	0.84146	1.32258	0.81481	1.02703	0.74699	0.80247	0.85897	0.71951	0.75309	0.80769
kɯ1		1.06522	1.05435	1.04348	1.04348	1.06593	1.07692	1.08889	1.08889	1.08989	1.07865	1.06742
kɯ1	会(写)	0.90991	0.82883	0.79279	0.81818	0.81982	0.81818	0.82569	0.78899	0.77778	0.96154	0.74257
kɯ1		1.1573	1.1573	1.1573	1.17045	1.15909	1.14773	1.13636	1.13793	1.12644	1.13953	1.12791
lʏ1	喉结	1.08989	1.07865	1.07865	1.07865	1.09091	1.09091	1.09195	1.10465	1.13095	1.13095	1.13253
lʏ1		1.12791	1.12791	1.12791	1.13953	1.13953	1.15294	1.15294	1.15294	1.16667	1.16667	1.16667
lɤ1	锥子	1.08046	1.08046	1.0814	1.06977	1.05814	1.04651	1.03448	1.04651	1.04651	1.05882	1.07143
lɤ1		1.14458	1.13095	1.14458	1.13253	1.13253	1.13253	1.12048	1.13415	1.12195	1.12195	1.1358
lu1	老虎	1.08791	1.08889	1.10112	1.10227	1.10227	1.11494	1.10345	1.10465	1.0814	1.04598	1.03409
lu1		1.07865	1.07865	1.07865	1.08989	1.07778	1.08889	1.10112	1.10112	1.10227	1.10227	1.10227
lu1	够	1.1358	1.15	1.1519	1.15385	1.13924	1.1519	1.1519	1.1519	1.1519	1.1519	1.1375
lu1		1.10976	1.10843	1.10843	1.12048	1.12048	1.13253	1.14634	1.14634	1.16049	1.1875	1.1875
ʂʅ1	撒(种)	1.04494	1.04545	1.04598	1.03448	1.02299	1.02326	1.01176	0.98824	0.97647	0.97647	0.97647
ʂʅ1		1.06897	1.06897	1.06897	1.05747	1.05747	1.05747	1.04598	1.04598	1.05814	1.04651	1.04651
ʂʅ1	拧	1.17073	1.17073	1.17073	1.15854	1.13415	1.12195	1.12048	1.10843	1.08434	1.05952	1.07143
ʂʅ1		1.15116	1.16471	1.16667	1.15294	1.15476	1.13095	1.12048	1.10843	1.10843	1.13415	1.14634
ɣɒ1	菜	1.09091	1.10227	1.09091	1.10345	1.09195	1.08046	1.05747	1.04598	1.02299	1.01149	1.02299
ɣɒ1		1.10714	1.10714	1.12048	1.12048	1.10714	1.10714	1.09412	1.10714	1.11905	1.11905	1.11905
ɣɑ1	编	1.13253	1.13415	1.14815	1.16049	1.16049	1.14815	1.13415	1.10843	1.10843	1.08333	1.09639
ɣɑ1		1.12346	1.10976	1.10976	1.08434	1.08434	1.08434	1.08434	1.09756	1.09756	1.08537	1.09877

与男 1 的情况非常相似，在 7 对松紧音中，紧音的h2/h1 基本上都要大于对应的松音。但其中有一些不稳定的情况值得注意，如：(1) kɯ̲1‘会(写)’中比值突然降低，比对应的松音都要低；lu̲1‘够’在 13-15 点；也有类似情况；(2) ɬu̲1‘放牧’的两个音节中，在第 16-27 点出现了比值比松音低的情况；ɣɑ̲1‘编’在第 9 点、第 11 点、第 15-30 点。(3) le̲1‘锥子’整体都没有规则性。

二者比较，紧音不稳定的情况出现的音节并不完全一致，“紧”特征丧失的音节各不相同，显示出词汇扩散的特点(Wang 1969)。

在 3 调中，男 1 的表现如下：

语音	汉义	13	14	15	16	17	18	19	20	21	22	23
dɘ3	堤	1.01190	1.03571	1.05952	1.05952	1.02381	1.04762	1.16667	1.14286	1.01205	0.95122	1.07317
dɘ3		1.15663	1.21429	1.14286	1.04762	0.98795	0.96341	0.98795	1.10714	1.19048	1.09756	1.02500
dɘ̲3	割	1.05682	1.13580	1.11111	1.12658	1.19737	1.00000	0.93407	1.14607	1.13793	1.04762	1.06579
dɘ̲3		0.95699	1.02247	1.09524	1.09524	1.08537	1.10843	1.17978	1.24444	1.18182	1.04651	1.01220
dze3	漏	1.03571	0.95238	0.95181	0.97590	1.01205	1.03659	1.03659	1.03659	1.04938	1.04938	1.04938
dze3		1.20930	1.14118	1.03571	0.97590	0.97590	1.06098	1.14815	1.12346	1.03704	1.00000	1.05000
dze̲3	忌嘴	0.94565	0.96703	0.94565	0.94505	1.02326	1.08750	1.02469	1.16667	1.24390	1.02062	0.97802
dze̲3		1.08235	1.08235	1.07143	1.05952	1.09302	1.16854	1.19780	1.13043	1.05556	1.03448	1.04762
kʏ3	收割	1.03333	1.03333	1.03333	0.98901	0.97802	1.08791	1.12222	0.98876	0.94318	1.02273	1.12644
kʏ3		1.06522	1.06667	1.06818	1.05747	1.05747	1.05747	1.03409	1.01136	1.00000	0.98864	1.01136
kɤ̲3	样子	1.09412	1.09524	1.09639	1.12791	1.18681	1.20430	1.19565	1.18681	1.16129	1.10526	1.08602
kɤ̲3		1.11905	1.11364	1.09677	1.09474	1.09574	1.10753	1.11957	1.13333	1.10227	1.04706	1.00000
kʰɔ3	北	1.05952	1.05952	1.04762	1.04762	1.06024	1.04762	1.06024	1.05952	1.05952	1.07143	0.98889
kʰɔ3		1.05814	1.05814	1.05814	1.08235	1.04598	0.96703	0.95604	1.03488	1.05952	1.03571	1.03571
kʰɔ̲3	勒	1.05747	1.05882	1.07143	1.12644	1.19780	1.20225	1.15000	1.15789	1.21951	1.26966	1.25275
kʰɔ̲3		1.11111	1.12500	1.12658	1.14103	1.15000	1.16667	1.17978	1.19565	1.23077	1.22472	1.17442
ŋɒ3	瓦	1.02353	1.04762	1.03529	1.00000	0.98864	1.01163	1.02410	1.02439	1.08235	1.12791	1.02353
ŋɒ3		1.09412	1.03750	1.02532	1.08537	1.12791	1.06818	1.02353	1.10127	1.04938	0.93182	0.89888
ŋɑ̲3	鸟	1.07317	1.02299	1.01163	1.04938	1.06494	1.09333	1.10390	1.08434	1.06977	1.04762	1.03896
ŋɑ̲3		1.02326	1.03614	1.07595	1.09091	1.07792	1.07792	1.07792	1.09091	1.07595	1.04938	1.02439
pʰɘ3	辣	1.02410	1.02439	1.03659	1.03659	1.03659	1.02439	1.01235	0.98765	1.00000	1.02532	1.02500
pʰɘ3		0.98750	1.00000	1.01235	1.03750	1.03797	1.02532	1.01282	1.02597	1.06410	1.12821	1.12658
pʰɘ̲3	剥	1.09524	1.08333	1.08434	1.08537	1.09524	1.14943	1.18889	1.21978	1.20879	1.21348	1.23864
pʰɘ̲3		1.05882	1.05882	1.07143	1.07143	1.07143	1.05952	1.04762	1.04762	1.07059	1.10465	1.13953
pʰu3	布	1.17857	1.15663	1.12346	1.06250	1.00000	0.96471	0.96667	0.97849	0.98936	1.01099	1.03488
pʰu3		1.07500	0.98765	1.00000	1.10000	1.17500	1.14634	1.09639	1.06098	1.03750	1.03846	1.06329
pʰu̲3	翻身	1.00000	1.01020	1.03158	1.06593	1.09195	1.13253	1.13415	1.13253	1.11628	1.08889	1.05319
pʰu̲3		1.08046	1.10714	1.12195	1.12346	1.10843	1.10588	1.07865	1.06452	1.04167	1.04124	1.04167
tɯ3	告状	1.05618	1.03333	1.03333	1.04494	1.06897	1.05814	0.98851	0.86813	0.82609	0.91011	1.05952
tɯ3		1.07955	1.05618	1.06742	1.09091	1.09195	1.05814	1.04706	1.09756	1.10976	0.98864	0.87097
tɯ̲3	把(米)	1.10000	1.10000	1.11111	1.11236	1.11236	1.13636	1.17442	1.17442	1.07865	0.90625	0.82000
tɯ̲3		1.11236	1.03226	0.95876	0.90909	0.91753	0.93684	0.94681	0.94681	0.95745	0.93684	0.93684

的 6 对在h2/h1 值上没有发现规律。

男 2 的数据如下：

在 7 对中，紧音的h2/h1 大于对应的松音的例子只有 1 对：tɯ3‘告状’ / tɯ̲3‘把（米）’。其他

语音	汉义	13	14	15	16	17	18	19	20	21	22	23
dze3	漏	1.03371	1.02222	0.95604	0.95604	1.03409	1.04598	1.03488	1.04651	1.03333	1.02198	1.03409
dze3		1.05952	1.08537	1.07143	1.04545	1.04494	1.03371	1.04494	1.05747	1.08537	1.0875	1.08642
dze̲3	忌嘴	1.08235	1.09524	1.10843	1.10976	1.10976	1.05682	1.04598	1.06173	1.04878	1.04878	1.04878
dze̲3		0.94737	0.97802	1.08235	1.13415	1.13415	1.10976	1.11111	1.09877	1.07317	1.06098	1.09639
kɤ3	收割	1.02198	1.01099	1.01099	0.98901	0.95604	0.94444	0.94382	1	1.07865	1.08046	1.05952
kɤ3		0.98864	0.98864	0.97727	0.96591	1	1.05618	1.06818	1.03448	1.02353	1.0241	1.06098
kɤ̲3	样子	1.2125	1.12644	1.05263	1.07609	1.11111	1.10345	1.1375	1.16883	1.08333	1.03488	1.13158
kɤ̲3		1.09524	1.08333	1.09639	1.08434	1.08434	1.08434	1.07143	1.05952	1.03488	1	1
kʰɔ3	北	1	1.03529	1.05882	1.08333	1.08333	1.08333	1.08333	1.09524	1.10714	1.10714	1.11905
kʰɔ3		1.09524	1.09639	1.08434	1.08434	1.07143	1.07143	1.05952	1.05952	1.04762	1.03571	1.02381
kʰɔ̲3	勒	1.075	1.06173	1.07407	1.1125	1.1125	1.03614	0.96512	0.95349	0.9881	1.02439	1.05
kʰɔ̲3		1.03659	1.03659	1.02439	1.02439	1.0241	1.02439	1.02469	1	1	1	1
ŋɒ3	瓦	1.04938	1.04938	0.95294	1.01205	1.03797	1.02532	1.025	1.0375	1.05063	1.09091	1.0641
ŋɒ3		1.08642	1.10465	1.02247	1.03571	1.12987	1.11688	1.16867	1.09302	0.98795	1.05128	1.05195
ŋɑ̲3	鸟	1.02353	1.02353	1.01176	1.01176	1.0119	1.0119	1.0241	1	0.95238	0.96429	1.02469
ŋɑ̲3		1.14474	1.14474	1.15476	1.07955	0.91954	0.88506	0.98795	1.11688	1.1039	1.07407	1.0119
pʰɚ3	辣	1.03371	1.02273	1.02273	1	0.97727	1	1.01176	1.03614	1.06173	1.10256	1.08861
pʰɚ3		0.97727	0.96591	1	1.03448	1.02326	1.0119	1.03659	1.075	1.10256	1.08974	1.0625
pʰɚ̲3	剥	1.06024	1.08537	1.09877	1.09877	1.125	1.13924	1.16667	1.18182	1.17949	1.15385	1.11392
pʰɚ̲3		1.08642	1.07407	1.06173	1.04938	1.04938	1.075	1.075	1.0875	1.075	1.0625	1.05
pʰu3	布	1.06667	1.07692	1.03409	1.03409	1.04545	1.03448	1.04545	1.07778	1.07865	1.05814	1.04819
pʰu3		1.09091	1.08434	1.1	1.09524	1.06667	1.06742	1.06742	1.08235	1.10127	1.075	1.06024
pʰu̲3	翻身	1.15663	1.17073	1.1875	1.18987	1.21519	1.26582	1.31169	1.28205	1.17647	1.05208	1.0303
pʰu̲3		1.08791	1.1	1.10345	1.10976	1.12821	1.12821	1.13924	1.15385	1.16883	1.17105	1.17333
tɯ3	告状	1.04444	1.04444	1.04494	1.04494	1.04494	1.07865	1.09091	1.05618	1.02151	1.01064	0.98925
tɯ3		1.03333	1.03333	1.03333	1.03371	1.04494	1.1236	1.08889	1.03297	1.05376	1.03261	1.04545
tɯ̲3	把(米)	1.17241	1.17241	1.17241	1.17241	1.16092	1.17442	1.16279	1.13793	1.13953	1.13953	1.13953
tɯ̲3		1.1236	1.13636	1.14943	1.16092	1.17442	1.17442	1.16279	1.16279	1.16279	1.16279	1.16471

与男 1 类似，在 7 对中，只有 1 对tɯ3‘告状’/ tɯ̲3‘把（米）’保持了区分，其他情况都失去了h2/h1 值的区分度。

小结：

在 1 调中声带松紧的音质仍旧保留，但在 3 调中，紧调的短促是音节的主要区别特征。在tɯ3‘告状’ / tɯ̲3‘把（米）’的对立中，不仅在声调上有区别，在松紧上也有区别，属于双向对立的情况。但从上面的数据看出，大部分此类对立中，在声带松紧上没有区别，主要靠声调的对立来区别音节。据此，大致可以推测，早期在此类音节中，松紧音质是主要区别特征，而声调上的区分是次要的，但随后声调上的区分占据主导地位，松紧音质的区分特征逐渐消失。为什么不能解释为：在tɯ3‘告状’ / tɯ̲3‘把（米）’对立中，松紧上的区分是由于声调的短促化造成的呢？如果是这样，那么松紧区分是一种后起现象，

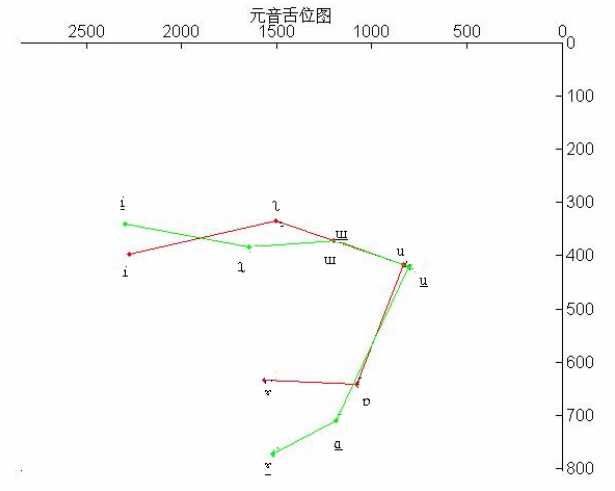
作为一种趋势，应该不断增多，因为紧 3 调这种产生条件一直稳定存在，而事实上，现在观察到的是松紧音质正在消失中，不仅在武定彝语中是这样，在其他彝语中也如此， 松紧音质作为特殊发声类型，从类型学的角度上看，也处于不稳定状态，容易变化，如白语中的情况(参见 汪锋 2007)。另外从武定彝语的音韵系统的情况也可以得出同样的结论：(1)因为在 1 调中，ɯ的松紧对立是唯一的区别特征，不是羡余的，可以追溯到早期彝语的情况，也就是说，松紧对立是较早阶段的。(2)武定 3 调有两个来源*6 调和*8，8 调在其他彝语中也表现为紧元音，对应 6 调的在其他彝语中都是松音，而且是受限分布，是可以根据规则推出来的(参见 汪锋 2006)。

3.4 共振峰

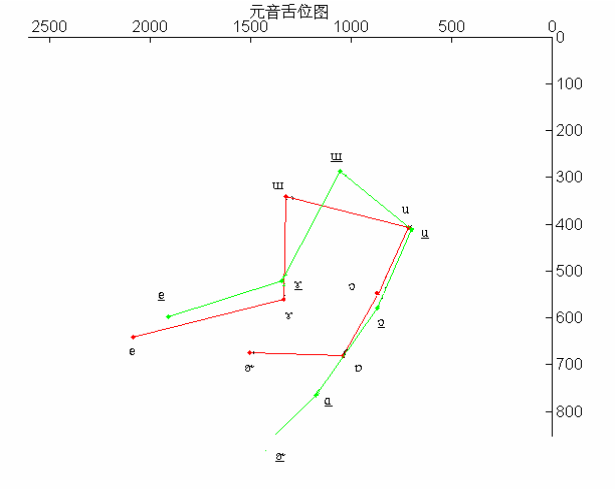
每条共振峰取 30 帧，平均后数值为第 1 组，而如果在共振峰的稳定段取点，本文取 30 帧中的第 16 帧，则可得到第 2 组数据，数据详见附录(二)。

从共振峰数据看，无论是在 1 调中，还是在 3 调中，表现出规律性的是：ɒ与ɑ̲相比，后者F1和F2 都要大，即后者开口度大，舌位靠前，这也符合我们田野调查时的听感。而其他的松紧对，都看不出与松紧音相应的有规则的明显对立，因此，可以认为武定彝语中的这些松紧对除了ɒ与ɑ̲外，都不是以调音的差异为区别特征的。

在 1 调情况下，以男 1 数据为例画出元音舌位图：



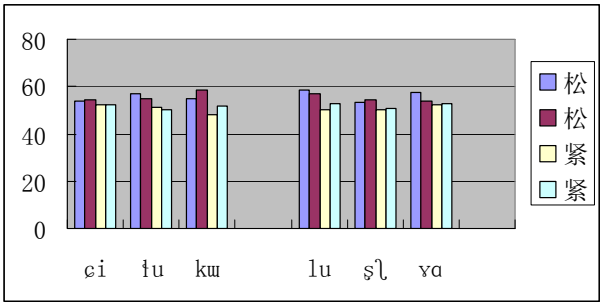
在 3 调情况下，以男 2 的数据画出元音舌位图如下：



3.5 开商

开商的算法是：开商=开相/周期。采用平均的办法处理数据，两个发音人的每个发音样本都分开计算，以观察更为细致的个体变异。

(1) 在 1 调的情况下，松音的开商都要大于紧音的开商，也就是说，松音声带开启所要的时间比相对的紧音要长；或者说，紧音声带开启比松音快。以男 2 为例，相差明显：

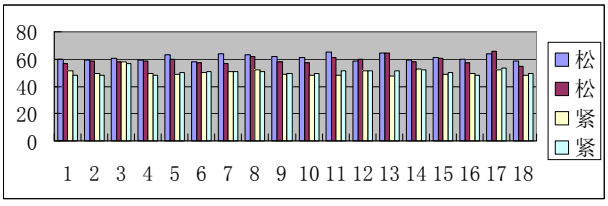


具体的开商统计数值如下：

语音	汉义	开商(男 1)	开商(男 2)
ɕi1	七	57.26	53.79
ɕi1		60.32	54.23
ɕi̱1	新	58.27	52.46
ɕi̱1		56.99	52.34
ɬu1	裤子	60.53	57.2
ɬu1		57.79	55.11
ɬu̱1	放牧	50.18	51.03
ɬu̱1		51.73	50.24
kɯ1	下	57.61	54.87
kɯ1		57.65	58.53
kɯ̱1	会(写)	50.31	48.12
kɯ̱1		51.62	51.81
lɤ1	喉结	60.21	58.09
lɤ1		59.63	54.63
lɤ̱1	锥子	58.46	51.01
lɤ̱1		55.17	51.1
lu1	老虎	60.75	58.35
lu1		60.65	56.88
lu̱1	够	53.39	50.42
lu̱1		55.55	52.78
ʂɭ1	撒(种)	54.78	53.11
ʂɭ1		55.1	54.23
ʂɭ̱1	拧	47.87	50.22
ʂɭ̱1		49.24	50.64
ɣɒ1	菜	57.77	57.36
ɣɒ1		56.37	53.88
ɣɑ̱1	编	52.43	52.44
ɣɑ̱1		58.93	52.79

在男 1 数据统计中，有两个音节的紧音突然跳高，具体原因尚待研究。

(2) 3 调中，数据显示，无一例外的，松音的开商都要大于紧音的开商。以男 2 为例，图示如下：



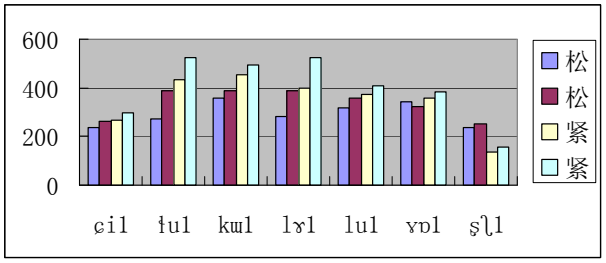
详细数据如下：

语音	汉义	开商(男 1)	开商(男 2)
dze3	漏	57.59	59.28
dze3		60.44	58.82
dze̱3	忌嘴	55.73	49.35
dze̱3		56.65	48.37
kɤ3	收割	58.78	62.95
kɤ3		61.24	59.82
kɤ̱3	样子	48.01	48.55
kɤ̱3		48.01	50.25
kʰɔ3	北	58.11	58.02
kʰɔ3		58.07	57.35
kʰɔ̱3	勒	52.12	50.38
kʰɔ̱3		46.32	51.04
ŋɒ3	瓦	57.33	61.28
ŋɒ3		57.7	57.17
ŋɑ̱3	鸟	53.19	48.41
ŋɑ̱3		52.4	49.62
pʰɚ3	辣	63.47	64.64
pʰɚ3		63.79	64.25
pʰɚ̱3	剥	56.65	47.16
pʰɚ̱3		60.25	51.25
pʰu3	布	59.36	59.48
pʰu3		59.9	57.89
pʰu̱3	翻身	51.47	52.42
pʰu̱3		53.25	51.75
tɯ3	告状	62.06	58.69
tɯ3		61.08	54.87
tɯ̱3	把(米)	51.58	48.41
tɯ̱3		52.2	49.35

3.6 速度商

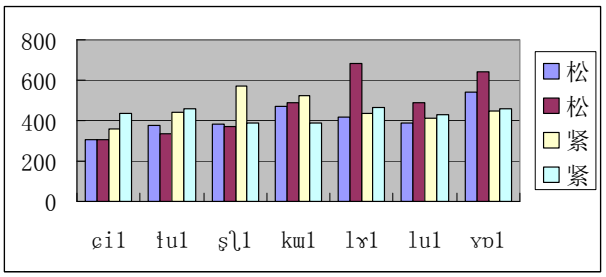
速度商的算法是：速度商=开启相/关闭相。同样采用平均法提出数据，分别得到两个发音人各次发音的速度商值。

(1) 在 1 调中，紧音的速度商基本都比松音的大，以男 1 为例，如下所示：



只是在ʂʅ1 音节上，正好相反，松音的速度商反而要大，原因未知，尚待进一步研究。

在男 2 中，不规则的情况开始增多，只有三对是紧音速度商大于松音，在 kɯ1, lɤ1, lu1 上没有规律，如果观察其波形，会发现波形的异常，对于此类波形产生的原因尚待进一步研究，在此处暂时可看作个体变异现象；而在 ɣɒ1 松音的速度商则大于紧音了。



详细数据如下：

语音	汉义	速度商(男 1)	速度商(男 2)
ɕi1	七	237.73	303.75
ɕi1		261.95	303.73
ɕi̱1	新	266.98	361.37
ɕi̱1		295.59	434.11
ɬu1	裤子	274.71	373.76
ɬu1		386.64	333.33
ɬu̱1	放牧	433.6	438.27
ɬu̱1		525.63	456.58
kɯ1	下	359.32	
kɯ1		388.93	490.74
kɯ̱1	会(写)	452.95	526.23
kɯ̱1		494.28	387.98
lɤ1	喉结	283.62	418.87
lɤ1		388	679.87
lɤ̱1	锥子	399.17	436.67
lɤ̱1		526.73	464.45
lu1	老虎	315.38	385.78
lu1		360.17	490.86
lu̱1	够	371.16	412.67
lu̱1		406.89	429.93
ʂʅ1	撒(种)	236.37	379.97
ʂʅ1		250.72	372.64
ʂʅ̱1	拧	137.71	567.8
ʂʅ̱1		157.88	388.73
ɣɒ1	菜	340.47	539.98
ɣɒ1		321.21	643.26
ɣɑ̱1	编	359.1	448.43
ɣɑ̱1		384.36	458.4

(2) 3 调中，紧音速度商大于松音的规律基本上丧失，在男 1 和男 2 两个发音人在不同的少数音节上保存这一规律，男 1 一对音节(tɯ3)，男 2 两对音节(kɤ3，ŋɒ3)。详细数据如下：

语音	汉义	速度商(男 1)	速度商(男 2)
dze3	漏	301.55	495.3
dze3		359.06	682.52
dze̱3	忌嘴	268.7	507.56
dze̱3		334.76	630.29
kɤ3	收割	339.87	345.84
kɤ3		263.89	435.92
kɤ̱3	样子	302.86	479.74
kɤ̱3		242.23	571.82
$k^hɔ3$	北	419.45	479.44
$k^hɔ3$		411.26	669.6
$k^hɔ̱3$	勒	332.2	485.33
$k^hɔ̱3$		246.91	569.33
ŋɒ3	瓦	383.58	421.33
ŋɒ3		374.38	469.95
ŋɑ̱3	鸟	342.37	484.74
ŋɑ̱3		253.71	515.71

pʰɚ3	辣	462.82	575.47
pʰɚ3		466.79	599.31
pʰə̲3	剥	481.76	533.31
pʰə̲3		456.82	450.88
pʰu3	布	456.82	433.07
pʰu3		386.67	456.41
pʰu̲3	翻身	366.49	425.14
pʰu̲3		396.79	467.95
tɯ3	告状	364.83	374.52
tɯ3		353.12	1058.55
tɯ̲3	把(米)	546.97	562.09
tɯ̲3		419.89	548.25

4 结论

根据(孔江平 2001:188)的研究，紧喉音与其他发声类型在开商和速度商上的区别如下所示：

	气泡音	气嗓音	紧喉音	正常音	高音调嗓音
速度商	+	-	+	+ -	-
开商	+	+	-	+ -	-

根据上节的讨论，从速度商和开商的表现来看，在 1 调中，武定彝语松紧音对立的语音实质是紧喉音和正常音的对立，在两个样本中这一特征很明显；而在 3 调中，松紧音的对立就不是典型的紧喉音和正常音的对立，因为在速度商上，体现不出明显的对立规则。简而言之，武定彝语紧喉音和正常音的对立保持在 1 调中，而在 3 调中起区别作用的并不是发声类型。

结合前文的声学分析结果，在 3 调音节中起区别作用的应该是时长，也就是传统上认为“紧”的音现在并不体现为发声类型上的紧喉，而是时长上的短促特征。而与短促密切相关的是开商要小，短调的开商比相应的长调要小，也就是短促调与声带开启时间短也相关；而紧喉音的一个特征就是开商小，这一特征可能就是音变的桥梁。设想在紧喉音时，紧音开商小，也就是开启时间短，在非持续调的情况下，会造成短促的效果，主要表现为时长相应缩短，一旦该羡余特征得到发展，就有可能取代其他方面的特征而成为主要区别特征，这时候，表现为紧音的其他方面开始不再对立，比如速度商上，紧音不一定比松音大，这在 3 调的松紧对立上表现很明显，如上 3.6 所示。而在 1 调，高平调的情况下，短促的情况没有得到发展，因此保持音节区别的主要特征仍旧是声带的“松紧”状态，紧音的开商小，速度商大与声学上的 h2/h1 值大相匹配。

这样，可以得到比较重要的两点启示：(1) 在不同的音高条件下，松紧音性质可以有不同的变化。也就是说，在音系上可以统一处理为“松紧”的对立并不蕴含语音实质上的同一，“松”通常对应所谓的“正常音”(modal voice)，“紧”通常对应“特殊音”(non-modal voice)。从武定彝语的情况可以看到，“紧”作为音系特征，至少涵盖了 1 调中的“紧喉”嗓音和 3 调中的“短促”时长，因此，“紧”是一个语言学或者音位学平面上的概念，而不是语音学平面的发声类型的概念。(2) 发音人的个体在利用不同的语音特征上会有不同，表现在不同发音人中体现特征变化的音节不同。而这两点分别说明了语音变化的条件性和在音变实现上的词汇扩散性。

参考文献

[1] 陈士林、边仕明、李秀清. 1985. 彝语简志. 北京：民族出版社.

[2] 戴庆厦 1979 我国藏缅语族松紧元音来源初探，民族语文，第 1 期.

[3] 胡坦、戴庆厦 1964 哈尼语元音的松紧，中国语文，第 1 期.

[4] 黄布凡 等. 1992. 藏缅语族语言词汇.北京:中央民族学院出版社.

[5] 孔江平. 2001. 论语言发声. 北京:中央民族学院出版社.

[6] 马学良 1948 倮文作祭献药供牲经译注，中央研究院历史语言所集刊.20.

[7] 汪锋. 2006. 《白、汉、彝比较研究——兼论白语的源流》，北京大学博士后报告.

[8] 汪锋. 2007. 白语方言中特殊发声类型的来源与演变. 汉藏语学报. 第 1 期.

[9] Maddieson Ian and Ladefoged Peter. 1985. ‘Tense’ and ‘Lax’ in four minority languages of China, UCLA Working Papers in Phonetics. 60-99.

（汪锋 北京大学汉语语言学研究中心/中文系

100871 tbwfwf@gmail.com）

（孔江平 京大学汉语语言学研究中心/中文系

100871 kongjp@gmail.com）

（原发表于《中国语言学》第 2 期，济南：山东教育出版社）

A Study of Mandarin Chinese Using X-Ray and MRI

Gaowu WANG, Xugang LU, Jianwu DANG,
Huaiqiao BAO, Jiangping KONG

Abstract: This paper describes a primary study to establish a dynamic articulatory model by combining MRI technique and X-ray data, where the former is used to refine the detailed real shape of the vocal tract and the latter provides the dynamic information of articulation. In this study, MRI experiments were conducted to obtain 3D static morphologies of 9 single vowels of Mandarin, and the vocal tract shapes were investigated. A set of coefficients of the alpha-beta model has been derived from MRI data. The articulatory movement was obtained from a Mandarin X-ray video (cineradiography) database, which is the only available corpus for Mandarin, and the cross-sectional areas were calculated using the MRI based alpha-beta coefficients. For an evaluation, the formants estimated from the vocal tract area functions of both MRI and X-ray were compared with those obtained from real speech sound. The estimation was consistent with the real speech sound with a mismatch of about 10% and 15%, respectively.

Keywords: Mandarin X-ray MRI vocal tract

1. INTRODUCTION

To better understand speech production, from the phonological inputs to acoustic signals, going through motor control and physiological processes, it is important to establish an elaborate articulatory and vocal tract model. To do so, it is necessary to get a fine morphology, especially the detailed real shape of the vocal tract during speech.

So far, the technologies adopted in speech production to get vocal tract shape include X-ray photos and dynamic video (cineradiography), Electropalatography (EPG), computed tomography (CT), Ultrasonics, Magnetic Resonance Imaging (MRI), and Electromagnetic articulography (EMA). Each of them has its own advantages and disadvantages. For example, MRI can provide 3D static vocal tract shape data with high spatial resolution but poor temporal resolution, while cineradiography has higher temporal resolution but can only show 2D sagittal dynamic movement. It is required a mapping from widths in 2D plane to the area function if the transmission line model is employed in acoustic modeling.

In this research, we utilize the advantages of X-ray cineradiography and MRI observation. The latter can provide a mapping for the former by using the 2D and 3D information obtained from MRI observation. To do so, we draw a set of alpha-beta coefficients from 2D and 3D static morphologies in MRI data. These coefficients are used to estimate the cross-sectional area function from the midsagittal width in X-ray movie. So, we can establish a mapping from 2D widths to 3D areas for articulatory movements in X-ray movie. The estimation result is evaluated by using the transmission line model in acoustic modeling.

1.1 Introduction of X-ray studies in speech research

Research on speech using X-ray has a long history. According to Dart [1], as early as in 1907, Barth and Grunmach used still X-ray to study speech.

In 1942, Chiba [2] combined X-ray photography, palatography, and laryngoscopy to measure the vocal tract shape in their pioneering research.

Thereafter, Fant completed the theory of speech production based on the extensive analysis of X-ray data [3]. Also, some important studies [3-8] have been carried out using X-ray data, successively.

Although other new technologies such as MRI, CT, and EMA were developed, nowadays, the full sagittal view of vocal tract articulators during running speech provided by cineradiography remains unsurpassed by more modern techniques. Currently, the collection techniques are of two types: one permits real time tracking, but is limited to a few coplanar points (microbeam, magnetometer); the other gives full volume vocal tract images, but is limited to a static, sustainable configuration (MRI, CT). The data obtained from the cineradiography are still used in many recent research [9, 10].

In China, Zhou and Wu [11] recorded X-ray photos for Mandarin vowels and consonants uttered by a female native speaker. Bao and Yang [12] had measured the articulatory movements of five consecutive Mandarin vowels using the cineradiography. Bao [13] used X-ray still images to give a physiological explanation for the classification of Mandarin vowels. Also, some preliminary studies were performed on the relationship between the cross-sectional area function of the vocal tract and formant frequencies of vowels [14], and the synthesis of Mandarin single vowels by articulatory parameters [15].

1.2 Introduction of MRI studies

MRI allows a tomographic view of body tissues in any plane within the human body, and gets the 3D shape of the vocal tract without known risks for the subject. So, it

has been increasingly applied in speech research over the past 20 years [16-31].

The articulatory data collected using MRI are valuable in understanding and modeling the vocal tract accurately, particularly the pharynx, the behavior of which during speech is traditionally hard to capture. Also, the volumetric production data is very important in articulatory synthesis, in which scientists have been involved for several decades.

At present, MRI studies have been carried out on several languages, e.g. English, French, Japanese and Swedish, and so on. However, few MRI studies have been conducted on Mandarin Chinese. In this study, we investigated the morphological properties of 9 Mandarin single vowels.

2. DATA PROCESSING

In this section, we introduce the procedures for marking X-ray movie and obtaining MRI data.

2.1 X-ray video processing

An X-ray video database provided by Bao [13, 14], is the only available cineradiographs corpus of Mandarin. Both sagittal cineradiography and frontal cine-photographs of the lips are given simultaneously. In total 786 utterances of 2 male and 2 female subjects are shown to illustrate most of the sounds of Mandarin, covering 217 syllables, besides "retroflexed" syllables and tone contrasts. This video has been segmented to syllables in avi files at 30 fps.

A platform 'VocalMarker' in Matlab is programmed to trace the midsagittal articulatory movements. Figure 1 shows an example image of the marking of articulatory movement, from which the midsagittal width of the vocal tract has been measured.

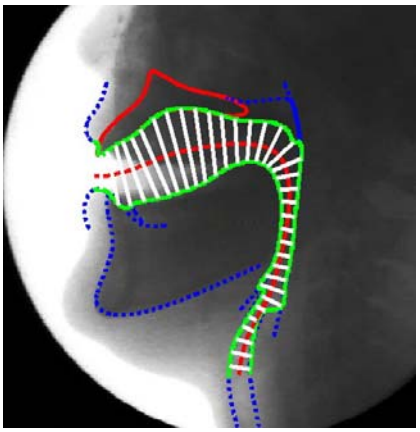


Figure 1: The midsagittal view of the marked X-ray video

2.2 MRI data acquisition and processing

2.2.1 Speech materials

This study covers the 9 single vowels in Mandarin, /a o e i u ü (i)e (s)i (sh)i/, in the Chinese Phoneticisation Scheme, whose IPA are / a o ɤ i u y e ɿ ʅ /, respectively, while another vowel /er/ (/ər/), whose status as a single vowel is still in dispute, is not included in our present study.

The 9 vowels are uttered in Chinese single syllable words, "啊喔嗣衣乌淤椰思诗", respectively, while the stable posterior parts in "椰思诗" were segmented and treated as independent vowels in the following speech analysis. In addition, all the characters are produced in the first tone (high flat tone) to ensure the stability of the sustained vowels.

2.2.2 Selection and training of subjects

The subjects have no speech or voice problems. They are native speakers of North Chinese dialects, live in or around the Beijing area, and have no other dialect accents. Alternatively, the subjects had passed the Putonghua Proficiency Test (PPT) and achieved Grade One, Level B (G1L2, the second highest grade, which is required for a Mandarin teacher and a television announcer).

After the subjects are selected, they are trained to ensure articulatory stability for obtaining clear MRI images. In the training, we require the subjects to produce the speech materials in a supine position with MRI noise in the earphones. Sufficient practice yields better imaging results.

At present, we have two subjects. Both of them are North Chinese from around Beijing area, and the female subject has passed the PPT G1L2.

2.2.3 MRI equipment and scan specifications

The MRI data was acquired with the Shimadzu-Marconi ECLIPSE 1.5T PowerDrive 250 installed at the Brain Activity Imaging Center, Advanced Telecommunications Research Institute (ATR-BAIC), in Kyoto, Japan.

Originally, a long standing drawback of MRI is the image acquisition time, which was several minutes per speech required in the earliest studies, while the acquisition time of this MRI machine was around 30 seconds for a 3D scan of the vocal tract, and the time varies with the number of image slices. To obtain a high quality image, it is still required that subjects maintain stable articulatory configurations during the image acquisition period, which might result in articulatory instability and subject motion during image acquisition, which in turn might create artifacts in the images.

Recently, a synchronized sampling method (SSM) with external trigger pulses developed by Masaki [32] was adopted in recording the movements of the speech organs as a set of sequential images. This method can also be used in acquiring the static 3D shape of vowels. The subject repeats the vowel about 30~36 times, each time sustains 3 seconds, which enables the subject to articulate in a stable manner. Figure 2 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.

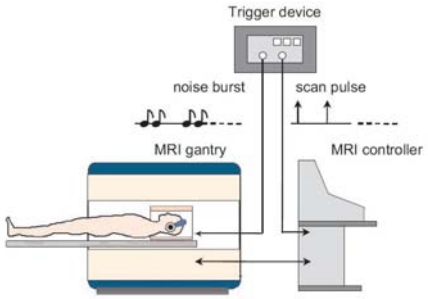


Figure 2: Experimental setup for the synchronized sampling method. (after Takemoto [33]).

The parameters were as following: a 3.4 [ms] echo time (TE), a 2200 [ms] relaxation time (TR), 44~51 sagittal slice planes, a 1.5 [mm] slice thickness, a 1.5 [mm] slice interval, a 256*256 [mm] field of view (FOV), and a 512*512 pixel image size. The data are stored in the DICOM file format.

2.2.4 MRI image preprocessing

The images were converted from DICOM into TIFF and denoised using ImageJ software, which is released by the NIH (National Institutes of Health, USA). Figure 3 shows an example of the image denoising effect, in which the articulatory structures are maintained with carefully checking.

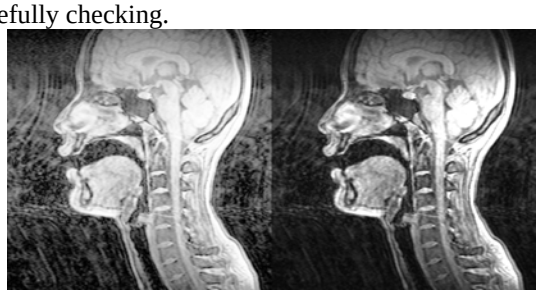


Figure 3: An example of the image denoising effect. Some noise spots (left panel) have been eliminated (right panel).

2.2.5 Teeth superimposition

MRI has a disadvantage in imaging the bony structure because the calcified structures lacking mobile hydrogen produce no resonance signals. Accordingly, the region of the teeth shows the same brightness as that of the air space. To build up an elaborate vocal tract model, however, it is necessary to obtain the teeth-air boundary to accurately reconstruct the vocal tract shape from the MRI data.

To solve this problem, we measured the structure of the teeth before or after obtaining the articulation data [33]. The subjects were asked to fill their mouth with a multi-mineral juice as a contrast medium and to lay prone in the MRI machine. An MRI scan is performed with the following parameters: an 11 [ms] echo time (TE), a 3000 [ms] relaxation time (TR), 51 sagittal slice planes, a 1.5 [mm] slice thickness, a 1.5 [mm] slice interval, a 256*256 [mm] field of view (FOV), and a 512*512 pixel image size. The data are stored in the DICOM file format.

In the teeth scan, the images showed the oral cavity with high brightness due to the contrast medium, while the teeth and jaws appeared with low brightness. This contrast makes it easy to extract the teeth and their supporting rigid structures (maxilla, mandible) 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 original MRI volumes. Figure 4 shows the result of teeth extraction and imposition. We resliced the “digital jaw casts” to the same slices as the slices of the vowels data, and located the upper and lower teeth manually in the midsagittal plane, respectively, to minimize the superimposition error, as shown in the right panel of Figure 4.

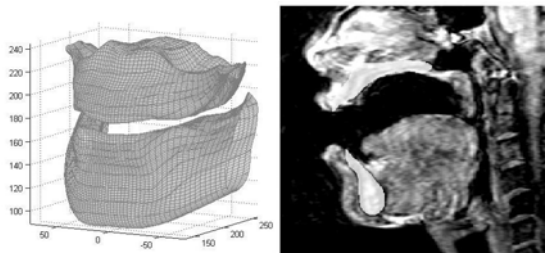


Figure 4: Left is the 3D digital jaw casts; Right is the midsagittal view after superimposition

2.2.6 Extracting the vocal tract area function

Vocal tract area functions were extracted from the reconstructed volumes with the teeth. Similar to [33], the extraction was performed in three steps. First, the vocal tract midline was semi-automatically calculated in the midsagittal 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. Figure 5 shows the midline on an actual image of the vowel /i/, and the cross-sections from which the vocal tract area function is measured.

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. We followed the method in [33]: as seen in Figure 5, we determined the furthest section from the glottis where the circumferential area could be measured as the last section (slice “h”), and the length of this section was extended to halfway from the end of this section to the last section where the upper and lower lips could still be observed (slice “i”).

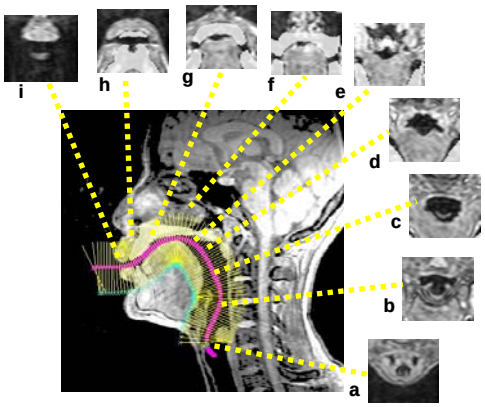


Figure 5: The results of extraction

2.2.7 Calculating transfer functions

The vocal tract transfer functions were calculated and the formants were estimated for all the volumes obtained by MRI using a transmission line model, which is detailed in [21].

2.2.8 Acoustic recording and analysis

Speech sounds were recorded from the subject in a soundproof room as the natural speech sound. 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 repeat the vowels as much in the same way as in the MRI experiment. The speech signals and noise bursts were recorded with a recording system, consisting of a SONY ECM-G5M Microphone, a BEHRINGER EURORACK UB502 Mixer, a CREATIVE AUDIGY2NX Soundcard, and a DELL XPS M1210 Computer.

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

3. RESULTS

3.1 MRI 3D inside view of Mandarin articulation

To explore the inside view, we reconstructed cutaway views based on volumetric MRI data for 9 Mandarin vowels. As a result, we obtained 3D shapes of Mandarin vowels, and show the inside vocal tract and articulators in Figure 6. The 3D images are produced by the 3DMed toolkit released by the Medical Image Processing Group, IA, CAS. To our knowledge, this is the first time to show a 3D shape for Mandarin vowels with the inner view. For this reason, we enlarge the special apical vowel of Chinese at the bottom of the figure. Such a result can be applied to the deaf or to people who have difficulties in learning Mandarin.

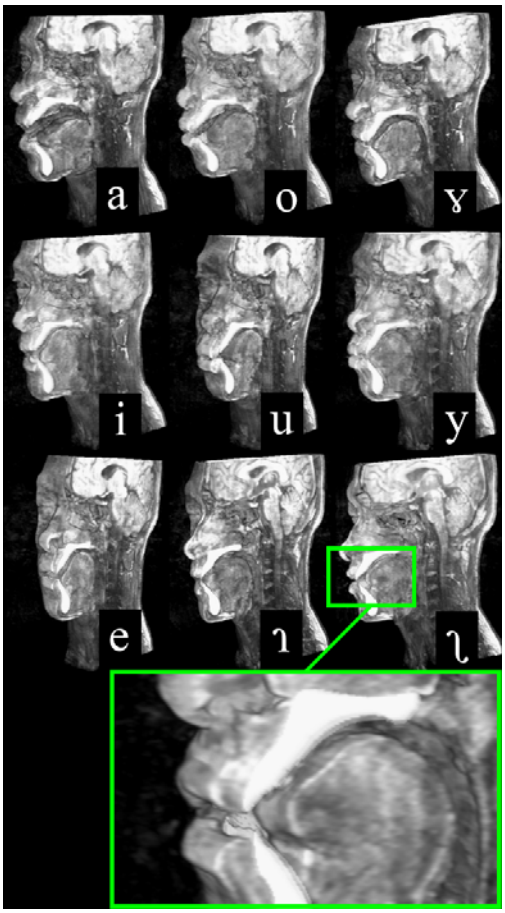


Figure 6: The 3D shapes of 9 single vowels in Mandarin.

3.2 MRI 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 have been extracted, and the corresponding formant frequencies have been estimated and compared with those of natural speech sound. Figure 7 shows the 3D vocal tract shape for vowel /a/.

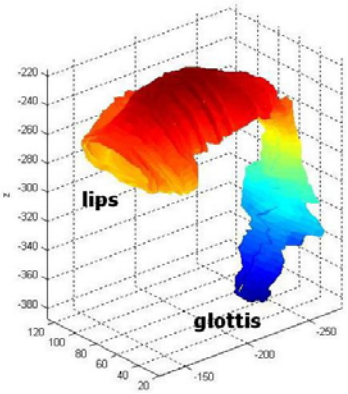


Figure 7: The 3D Vocal Tract for the vowel /a/

Table 1 shows the frequencies of first four formants, which are compared with those of natural speech sound. The percentage errors between them are less than 10%. Relatively large errors are found locally. The mean

absolute percentage error is 7.4%. This result is better than [34], whose mean absolute percentage error is 12.2% for English, but no so good as [33], whose error is 4.5%, for Japanese. While the Difference Limen (DL) for formant frequency discrimination range between 3% and 14%, which were reported in [35-37]

Therefore, in the future work, we will decrease the errors, and synthesize Mandarin single vowels using the estimated formants to perform a perceptual evaluation.

Table 1: The natural and calculated speech formants of the female subject in MRI, and the percentage errors for the latter relative to the former. "n" denotes the natural speech, "c" the calculation, and "d" percentage 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

3.3 From the X-ray midsagittal widths to cross-sectional areas using the alpha-beta model

Because X-ray video can only show the sagittal view of the vocal tract, it is necessary to find a method to estimate cross-sectional areas. At present, most transformations going from the midsagittal distance to the cross-sectional area are based on the original transformation defined by [5], which is the α β (alpha-beta) Model.

$$A(x) = \alpha(x) d(x)^{\beta(x)} \qquad (1)$$

where 'd' is the midsagittal distance, 'A' the cross-sectional area, 'x' the position along the vocal tract mid-line, and 'α' and 'β' are the two coefficients of the transformation, which are also functions of the variable 'x'.

In this study, a set of 'α' and 'β' coefficients has been calculated using the 'd' and 'A' from real MRI 3D data for 9 vowels, minimizing the estimation errors:

$$\arg\min_{\alpha(x),\beta(x)} \left\{ \sum_{V=/\,aoeiuv\ldots/} \left[A(x,V) - \alpha(x) d(x,V)^{\beta(x)} \right]^2 \right\} \qquad (2)$$

where 'V' represents the 9 single vowels in Mandarin. It means that 'A' and 'd' are phoneme dependent.

This set of alpha-beta coefficients reflects the morphological characteristics of this subject in MRI, so that we use Eq. (1) to estimate cross-sectional areas from midsagittal width of different vowels of this subject within limited errors.

And this set of alpha-beta coefficients is applied to the X-ray video data, using Eq. (1), to estimate 'A' from 'd' of the female subject in X-ray movie. From the estimated cross-sectional areas for the female subject in the X-ray database, we calculated the lower four formants of vowels (at present only /a i u/, due to the arduous X-ray video tracing), as shown in table 2, generally with a 15% mismatch as compared with the formants obtained from the real speech sound. Although the subjects are different, this is better than the empirical formula used in [14].

Table 2: The natural and calculated speech formants of the female subject 1 in the X-ray database

	/a/	/i/	/u/
nF1	911	390	450
nF2	1364	2862	921
nF3	3650	3699	3365
nF4	4232	4260	4165
cF1	786	398	489
cF2	1289	2476	977
cF3	3393	3423	2701
cF4	4027	4078	3386
dF1	-13.7	2.1	8.7
dF2	-5.5	-13.5	6.1
dF3	-7.0	-7.5	-19.7
dF4	-4.8	-4.3	-18.7

4. CONCULUSIONS AND DISCUSSION

This paper established a mapping from 2D widths to area functions based on MRI data, and applied to 2D articulatory movements in X-ray movie. By utilizing the advantages of X-ray movie and MRI data, we make use of morphology information in both of them.

At first, we extracted the articulatory movements from a Mandarin X-ray database, and measured the midsagittal widths of the vocal tract. Next, a set of alpha-beta coefficients were estimated from the MRI data. Finally, the area function of the vocal tract of X-ray was calculated using this set of alpha-beta coefficients. The

acoustic comparison showed that the combination of MRI and X-ray is available for further development.

However, the alpha-beta coefficients are speaker dependent, which will bring error when we apply them on the speaker in X-ray movie. In the future we will include more phonemes and syllables in Mandarin, not only some cardinal vowels /a i u/ in this paper, to evaluate the alpha-beta coefficients in both static and dynamic ways.

5. ACKNOWLEDGES

We sincerely thank the anonymous reviewers for their helpful comments and constructive suggestions.

This study is in part supported by a Grant-in-Aid for Scientific Research of Japan (No. 20300064) and SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC) of Japan.

6. REFERENCES

[1]Dart, S.N., A bibliography of X-ray studies of speech. UCLA Working Papers in Phonetics, 1987. **66**: p. 1-97.

[2]Chiba, T. and Kajiyama, M., *The Vowel: Its Nature and Structure*. 1942, Tokyo.

[3]Perkell, J.S., *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. Research Mono. No. 53. 1969, Cambridge, MA.: MIT press.

[4]Ohman, S.E.G. and Stevens, K.N., Cineradiographic Studies of Speech: Procedures and Objectives. Journal of the Acoustical Society of America, 1963. **35**: p. 1889.

[5]Heinz, J.M. and Stevens, K.N., On the Derivation of Area Functions and Acoustic Spectra from Cineradiographic Films of Speech. Journal of the Acoustical Society of America, 1964. **36**(1): p. 37.

[6]Moll, K.L. and Daniloff, R.G., Investigation of the Timing of Velar Movements during Speech. The Journal of the Acoustical Society of America, 1971. **50**(2B): p. 678-684.

[7]Mermelstein, P., Articulatory model for the study of speech production. The Journal of the Acoustical Society of America, 1973. **53**(4): p. 1070-1082.

[8]Harshman, R., Ladefoged, P., and Goldstein, L., Factor analysis of tongue shapes. Journal of the Acoustical Society of America, 1977. **62**(3).

[9]Beautemps, D., Badin, P., and Bailly, G., Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling. The Journal of the Acoustical Society of America, 2001. **109**(5): p. 2165-2180.

[10]Iskarous, K., Patterns of tongue movement. Journal of Phonetics, 2005. **33**(4): p. 363-381.

[11]Zhou, D.F. and Wu, Z.J., *The Articulation Album of Mandarin*. 1963.

[12]Bao, H.Q. and Yang, L.L., Study on cineradiography of consecutive vowels. Report of Phonetic Research, Phonetics Lab, CASS, 1982.

[13]Bao, H.Q., A physiological account of single vowels in Putonghua. ZHONGGUO YUWEN, 1984.

[14]Bao, H.Q., 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, 1983.

[15]Zu, Y.Q., The primary study of synthesizing vowels with articulatory parameters Report of Phonetic Research, Phonetics lab, CASS, 1983.

[16]Tiede, M., An MRI-based study of pharyngeal volume contrasts in Akan and English. Journal of Phonetics, 1996. **24**: p. 399-421.

[17]Demolin, D., Metens, T., and Soquet, A. Three-dimensional measurement of the vocal tract by MRI. in *Proceedings of 4th ICSLP*. 1996. Philadelphia.

[18]Baer, T., et al., Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. Journal of the Acoustical Society of America, 1991. **90**(2): p. 799-828.

[19]Lakshminarayan, A.V., Lee, S., and McCutcheon, M.J., MR imaging of the vocal tract during vowel production. Journal of Magnetic Resonance Imaging, 1991. **1**: p. 71-76.

[20]Moore, C.A., The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images. Journal of Speech and Hearing Research, 1992. **35**: p. 1009-1023.

[21]Dang, J. and Honda, K., Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation. Journal of the Acoustical Society of America, 1996. **100**(5): p. 3374-3383.

[22]Dang, J. and Honda, K., Acoustic characteristics of the piriform fossa in models and humans. Journal of the Acoustical Society of America, 1997. **101**(1): p. 456-465.

[23]Dang, J., Honda, K., and Suzuki, H., Morphological and acoustical analysis of the nasal and the paranasal cavities. Journal of the Acoustical Society of America, 1994. **96**(4): p. 2088-2100.

[24]Story, B.H., Titze, I.R., and Hoffman, E.A., Vocal tract area functions from magnetic resonance imaging. Journal of the Acoustical Society of America, 1996. **100**(1): p. 537-554.

[25]Alwan, A., Narayanan, S.S., and Haker, K., Toward articulatory-acoustic models for liquid consonants based on MRI and EPG data. Part II: The rhotics. Journal of the Acoustical Society of America, 1997. **101**: p. 1078-1089.

[26]Badin, P., et al. A three-dimensional linear articulatory model based on MRI data. in *Proceedings of the 3rd ESCA/COCOSDA International Workshop on Speech Synthesis*. 1998.

[27]Honda, K. and Tiede, M. An MRI study on the relationship between oral cavity shape and larynx position. in *Proceedings of 5th ICSLP*. 1998.

[28]Engwall, O., Vocal tract modeling in 3D. KTH STL-QPSR, 1999: p. 31-38.

[29]Fitch, W.T. and Giedd, J., Morphology and development of the human vocal tract: A study using magnetic resonance imaging. Journal of the Acoustical Society of America, 1999. **106**(3): p. 1511-1522.

[30]Jackson, P.J.B. and Shadle, C.H., Frication noise modulated by voicing, as revealed by pitch-scaled decomposition. Journal of the Acoustical Society of America, 2000. **108**(4): p. 1421-1434.

[31]Stone, M., et al. Modelling the Internal Tongue using Principal Strains. in *5th Seminar on Speech Production: Models and Data*. 2000. Kloster Seeon, Bavaria, Germany.

[32]Masaki, S., et al., MRI-based speech production study using a synchronized sampling method. J. Acoust. Soc. Jpn.(E), 1996. **20**: p. 375-379.

[33]Takemoto, H., et al., Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. Journal of the Acoustical Society of America, 2006. **119**(2): p. 1037-1049.

[34]Story, B.H., Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002. The Journal of the Acoustical Society of America, 2008. **123**(1): p. 327-335.
[35]Flanagan, J., *A Difference Limen for Vowel Formant Frequency*. 1955, ASA. p. 613-617.
[36]Mermelstein, P., Difference limens for formant frequencies of steady-state and consonant-bound vowels. The Journal of the Acoustical Society of America, 1978. **63**(2): p. 572-580.
[37]Kewley-Port, D. and Zheng, Y., *Auditory models of formant frequency discrimination for isolated vowels.* 1998, ASA. p. 1654-1666.

Gaowu WANG, Jiangping KONG, Phonetics Lab, Peking University, Beijing, China, 100871
Xugang LU, ATR Spoken Language communication Research Labs, National Institute of Information and Communications Technology, Japan
Jianwu DANG, School of Information Science, Japan Advanced Institute of Science and Technology, Nomi, Ishikawa, Japan, 923-1292
Huaiqiao BAO, Institute of Ethnology and Anthropology, CASS

Estimation of Vocal Tract Area Function for Mandarin Vowel Sequences Using MRI

Gaowu Wang [12], *Jianwu Dang* [1], *Jiangping Kong* [2]

[1] School of Information Science, Japan Advanced Institute of Science and Technology
[2] Phonetics Lab, Peking University, Beijing, China
wanggaowu@pku.edu.cn, jdang@jaist.ac.jp, jpkong@pku.edu.cn

Abstract

To fully explore the dynamic properties of speech production and investigate the relation between vocal tract geometry and speech acoustics, estimation of vocal tract area functions from measurements of the sagittal plane is an important step. In this study, we investigated the relation between the measurements on two dimensional (2D) and three dimensional (3D) MRI data and used an alpha-beta model to describe this relation. As a result, a set of parameters were derived from 3D static MRI data, and applied to time-varying vocal tract widths derived from 2D MRI movies, to synthesize Mandarin vowel sequences. An acoustic evaluation comparing the natural and calculated formants shows that the alpha-beta model can represent dynamic states of articulatory movements of vowel sequences, as well as those of the sustained vowels.

Index Terms: MRI, speech production, Mandarin, vowel, alpha-beta model

1. Introduction

Magnetic resonance imaging (MRI) allows a tomographic view of body tissues in any plane of the human body, and yields the 3D shape of the vocal tract, without any known risk for the subjects. MRI has been increasingly applied in speech research over the past 20 years [1-6]. The articulatory data collected using MRI is valuable in understanding and modeling the vocal tract accurately, particularly the pharynx area, since the behavior of this area is hard to capture during speech by traditional approaches. The volumetric morphological data is very important for articulatory synthesis, in which scientists have been involved for several decades.

At present, MRI studies have been carried out on several languages. However, few MRI studies have been conducted on Mandarin. MRI will be very valuable for research on Mandarin, since our final goal is to develop a dynamic 3D articulatory model and a visual aid system for Mandarin learning, which requires adequate morphological and articulatory data. However, due to the high cost of MRI experiments for obtaining 3D data of vocal tract, the 2D articulatory data in mid-sagittal plane are more attractive for speech research, where the generation of area functions from measurements of the sagittal section is an important step in the study of the relation between 2D and 3D geometry of the vocal tract in acoustic aspect. Several such generation methods have been proposed in the past [7-10] and the alpha-beta model is most famous.

In this study, we performed MRI experiments to get static and dynamic data of vocal tract, and uses alpha-beta model to estimate area functions of Mandarin vowel sequences.

2. Data Acquisition

In this section, we introduce the procedures for obtaining MRI data, including the selection of speech materials and subjects, the paradigm of MRI experiments, and the pre-processing of the MRI data. Due to the laborious processing of MRI data, at present, we only show the results of one female subject.

2.1. Speech materials

Mandarin, also called Putonghua, is the official national standard spoken language of China and is derived from the principal dialect spoken in and around the Beijing area.

As for the sustained vowels, our study covered the nine single vowels in Mandarin, /a o e i u ü (i)e (s)i (sh)i/, in the Chinese phoneticization scheme, whose IPA symbols are / a o ɤ i u y e ɿ ʅ /, respectively. There is one more vowel, “er” (/ər/), whose status as a single vowel is still in dispute. Therefore, we did not include this vowel in the present study. The vowels were uttered by saying the Chinese characters “啊喔嗣衣乌淤噎思诗”, where the vowels / e ɿ ʅ / in “噎思诗” are consonant-dependent and appear stably with preceding specific consonants or semivowels. For this reason, we asked subjects to pronounce the syllables with these three vowels instead of the isolated vowels, and we used the stable segments of the vowels in the following speech analysis. In addition, all the characters were produced in the first tone (high flat tone) to ensure the stability of the sustained vowels.

As for the dynamic movements, we selected 39 syllables, including diphthongss (e.g. /ai ei ao/), triphthongs (e.g. /iao iou uai/), and CV syllables (e.g. /ba bi bu/). 3 syllables were uttered in each MRI section, e.g. /ai ei ao/ which were uttered by saying the Chinese characters “哀诶凹”.

2.2. Selection and training of subjects

The subjects have no speech or voice problems. They are native speakers of North Chinese dialects, live in and around the Beijing area, and have no other dialect accents. Alternatively, the subjects have 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 announcers).

The subjects were trained to ensure articulatory stability so that we could obtain clear MRI images. In the training, we required the subjects to produce the speech materials in a supine position with MRI noise in the earphones. Sufficient practice yielded better imaging results.

2.3. MRI equipment and scan specifications

The MRI data were acquired using the Shimadzu-Marconi ECLIPSE 1.5T PowerDrive 250 installed at the Brain Activity Imaging Center, Advanced Telecommunications Research Institute (ATR-BAIC), Kyoto, Japan.

As is well known, the major drawback of MRI is its poor time resolution for speech research. At present, a synchronized sampling method (SSM) developed by [11] is adopted in recording the movements of the speech organs as a set of sequential images. This method can also be used for acquiring the static 3D shape of vowels.

In this study, the parameters used in the SSM MRI scans for sustained vowels are shown in Table 1.

Table 1. *Parameters for sustained vowels*

Echo time (TE)	3.4 ms
Relaxation time (TR)	2200 ms
Number of slices	44-51 sagittal slice planes
Slice thickness	1.5 mm
Slice interval	1.5 mm
Field of view (FOV)	256*256 mm
Image size	512*512 pixels
Image data format	DICOM file

The parameters of SSM MRI for dynamic movement of vowel sequences were as follows: a 2200 [ms] sequence length with 128 phases, the mid-sagittal slice, an FOV of 256*256 [mm], an image size of 256*256 pixels. As the result, for a vowel sequence, e.g. /ai ei ao/, we got 128 frames in mid-sagittal plane, which formed a real time movie at 60fps.

2.4. Image preprocessing

The images were converted from DICOM to TIFF and denoised using ImageJ software, which was produced by the National Institute of Health, USA.

2.5. Teeth superimposition

MRI has a disadvantage in imaging bony structures because calcified structures that lack mobile hydrogen produce no resonance signals. Accordingly, the region of the teeth has the same darkness as the air space. To solve this problem, we measured the structure of the teeth using the teeth imaging method proposed in [12]. The maxilla and the mandible with the teeth were reconstructed to obtain “digital jaw casts”, which were then manually superimposed onto the original MRI images. Figure 1 shows the result of teeth extraction and imposition. We resliced the digital jaw casts to match the slices of the vowel data, and we located the upper and lower teeth manually in the mid-sagittal plane, to minimize superimposition error, as shown in the right panel of Figure 1.

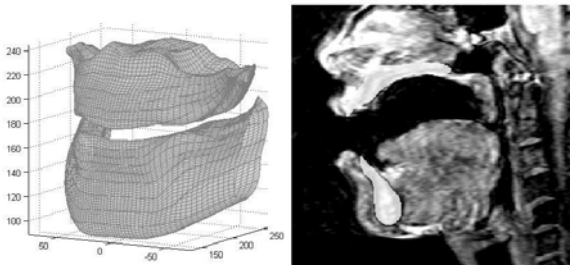


Figure 1: *left: 3D digital jaw casts; right: mid-sagittal plane after imposition.*

2.6. Data processing for sustained vowels

To study the acoustic properties of the vowels based on MRI data, we use the classical transmission line model [13-15]. This requires us to obtain the vocal tract area functions, which can be calculated from the reconstructed vocal tract shape with the teeth shown above.

The calculation was performed in three steps. First, the vocal tract midline was semi-automatically calculated in the mid-sagittal image. Then, along the midline, images perpendicular to the midline were resliced at 1~5 mm intervals. Finally, the cross-sectional area of the airway of the vocal tract in each slice was measured. The vocal tract area function is defined as a series of cross-sectional areas along the vocal tract. Figure 2 shows the midline of the vocal tract on the midsagittal plane, the grid lines for this vocal tract, and the cross-sectional images sliced according to the grid lines. The vocal tract area function is derived from these slices.

As seen in Figure 2, close to the lip end, the upper and lower lips are separated and a complete circumferential outline of the vocal tract section cannot be determined. It was necessary to answer the following two questions. Where is the end of the vocal tract in terms of acoustic meaning? How should the area of the incomplete shape be measured? In this study we adopted a method proposed by Takemoto [12]. As shown in Figure 2, we determined the furthest section from the glottis where the circumferential area could be measured as the last section (slice “h”), and the length of this section was extended to halfway from the end of this section to the last section where the upper and lower lips could still be observed (slice “i”).

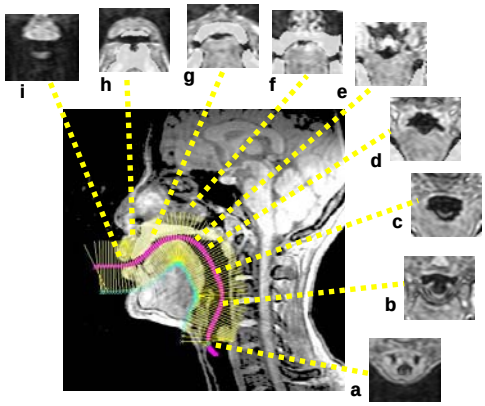


Figure 2: *Results of extraction of the cross-sectional slices.*

Subsequently, the vocal tract transfer functions were calculated based on the cross-sectional area function of the vocal tract obtained from MRI using a transmission line model, which is detailed in [15]. The calculated formant frequencies are listed in Table 2.

The speech sounds of the subjects were recorded in a soundproof room to remove the noise of MRI machine. In order to reproduce the situation of MRI acquisition, the subjects lay supine on the floor with a headset that fed the noise burst trains to the subjects. The subjects were then asked to repeat the vowels as similarly as possible to the way they spoke in the MRI machine.

We selected the stable segment of the recorded vowels, and used the Praat software to extract the four lower formants. Table 2 shows the first four formants calculated from MRI data and the formants of natural speech sounds for comparison. The absolute percentage error between all the

formant data from the transfer functions and from natural speech sounds is about 10%. The mean absolute percent error is 7.4%. This result is not as good as [12], in which the mean absolute percent error was 4.5% for Japanese vowels, but is better than a recent MRI study [16], in which the mean absolute percent error was 12.2%.

Table 2. *Comparison of the natural and calculated speech formants, "n" denotes natural speech, "c" the calculation, and "d" the percentage error.*

	a	o	e	i	u	ü	(i)e	(s)i	(sh)i
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

2.7. Data Processing for vowel sequences

A program 'VocalTractMarker' in Matlab was designed for semi-automatic tracing of the articulatory movements in the mid-sagittal plane. Figure 3 shows an example image of this marking. From the glottis to the lips along the midline of the air way, we measured the widths of the vocal tract at certain intervals.

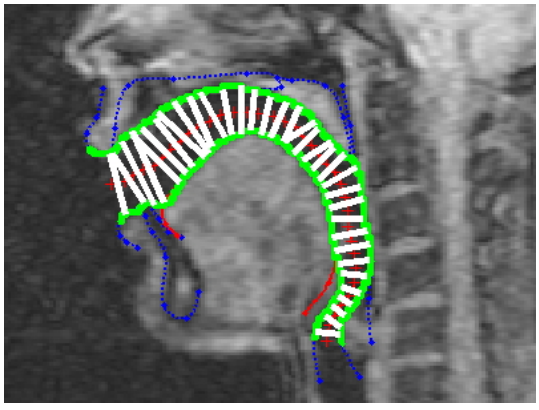


Figure 3: *Measuring the widths in the mid-sagittal plane in MRI movie.*

3. Applying alpha-beta model

3.1. For sustained vowels

Because the 2D dynamic MRI movie can only show 2D information of the vocal tract in the sagittal view, it is necessary to find a method to estimate cross sectional areas. At present, many transformations going from the mid-sagittal width to the cross-sectional area are based on the original transformation defined by [17], which is called the "α β" (alpha-beta) Model.

$$A(x) = \alpha d(x)^{\beta}, \quad (1)$$

where 'd' is the width of the vocal tract in the mid-sagittal plane, 'A' is the cross-sectional area, 'x' is the distance from the glottis along the vocal tract midline, and 'α' and 'β' are the two parameters of the transformation, which are also the functions of variable 'x'.

A set of 'α' and 'β' parameters was calculated using 'd' and 'A' from real MRI 3D data of 9 vowels, by minimizing the estimation errors. This set of alpha-beta parameters reflects the morphology characteristics of this subject, so that we can use Eq. (1) to estimate cross-sectional areas from the mid-sagittal widths of different vowels of this subject with a mean absolute error of 0.29 [cm^2] for all 9 vowels, while the errors are 0.32, 0.57, 0.24, 0.18, 0.44, 0.11, 0.24, 0.28, 0.21 [cm^2] for / a o ɤ i u y e ɪ ɻ /, respectively.

3.2. For vowel sequences

We used this set of alpha-beta parameters to calculate the cross-sectional areas of the vowel sequences, from which the formants were calculated and compared to the formants of real speech.

Figure 4 shows the movement of articulators during the diphthongs /ai/ in vowel sequence /ai ei ao/. From the articulatory posture of /a/ to that of /i/, the varying cross-sectional areas have been calculated by the alpha-beta model. One can see that while producing /ai/ the contours of the tongue have an invariant point, which is reflected in the area functions also.

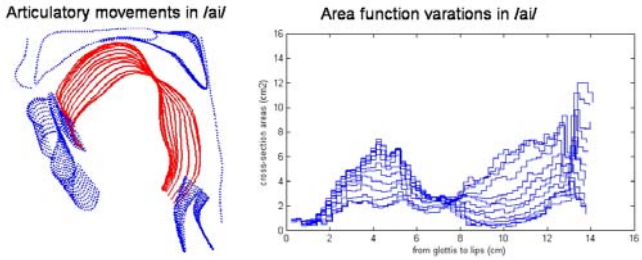


Figure 4: A*rticulatory movements and area function variations in diphthongs /ai/.*

4. Conclusions and discussion

For a clear contrast, we compared the formant trajectories in the acoustic plane and in the time domain, taking /ai ei ao/ as an example. In Figure 5, the large black dots indicate the natural formants of the 9 sustained vowels. The large blue asterisks indicate the formants calculated from the cross-section area functions estimated using the alpha-beta model on 2D data. And the small black dots demonstrate the trajectories of the formants of the recorded natural speech in

the acoustic plane. The small blue asterisks are the calculated formants from the 2D dynamic MRI movie. The upper panel shows the trajectories of F1 and F2 for /ai ei ao/ in the acoustic space, while the lower panel shows the trajectories in the time domain.

In the upper panel, two of the four open circles indicate the articulation places of /a/ and /i/ in diphthongs /ai/, respectively. One can see that the /a/ in /ai/ is more forward than single vowel /a/, and the articulation place of /i/ is not fully achieved in diphthongs /ai/.

In the lower panel, the formant trajectories calculated using alpha-beta model are consistent with those from natural speech. This shows that the alpha-beta model also performed well on dynamic vocal tracts, and the coarticulation trajectory information has been preserved.

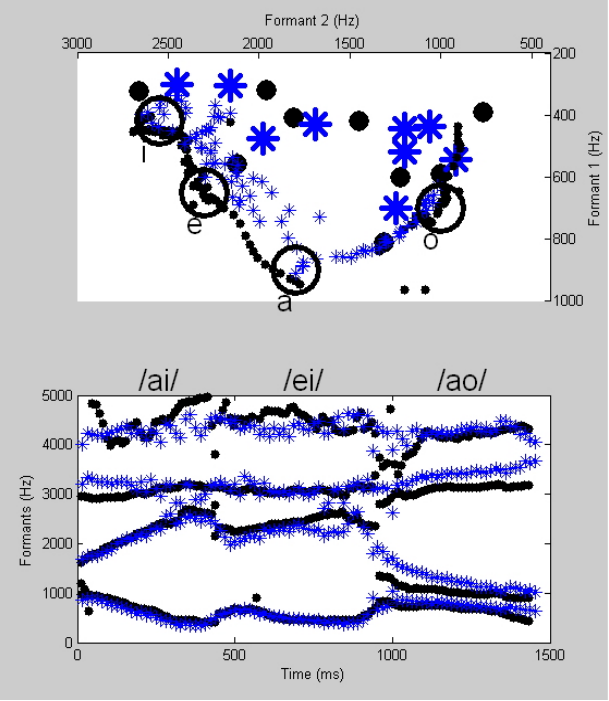


Figure 5: *Comparison of the real and calculated formant trajectories in the acoustic plane and in the time domain for vowel sequence /ai ei ao/.*

However, as shown in the lower panel of Figure 5, in the diphthongs /ao/, there is a large difference between the formant trajectories of the real speech sound and the calculation. This may be introduced from the difference between the states of the subject in the sound proof room and the MRI room. The noise in the MRI room and fatigue prevent the subjects from maintaining stable articulation. In the case of /ao/, it was confirmed that the trajectory for sounds during the MRI experiment was close to the calculation, but the trajectories were generally not clear due to the MRI noise. This will be investigated in future research.

Due to the laborious processing of MRI data, in this study, at present, we only show some examples of one female subject. In the future, for more generality, we will apply this study on more subjects.

5. Acknowledgement

This study is supported in part by SCOPE (071705001) of the Ministry of Internal Affairs and Communications (MIC) of Japan.

6. References

[1] T. Baer, J. C. Gore, L. C. Gracco, and P. W. Nye, "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," *Journal of the Acoustical Society of America*, vol. 90, pp. 799-828, 1991.

[2] J. Dang, K. Honda, and H. Suzuki, "Morphological and acoustical analysis of the nasal and the paranasal cavities," *Journal of the Acoustical Society of America*, vol. 96, pp. 2088-2100, 1994.

[3] B. H. Story, I. R. Titze, and E. A. Hoffman, "Vocal tract area functions from magnetic resonance imaging," *Journal of the Acoustical Society of America*, vol. 100, pp. 537-554, 1996.

[4] M. Tiede, "An MRI-based study of pharyngeal volume contrasts in Akan and English," *Journal of Phonetics*, vol. 24, pp. 399-421, 1996.

[5] K. Honda and M. Tiede, "An MRI study on the relationship between oral cavity shape and larynx position," in *Proceedings of 5th ICSLP*, 1998.

[6] O. Engwall, "Vocal tract modeling in 3D," *KTH STL-QPSR*, pp. 31-38, 1999.

[7] G. Fant, "Vocal tract area functions of Swedish vowels and a new three-parameter model," in *the International Conference on Spoken Language Processing*,, Banff, 1992, pp. 807-810.

[8] J. Sundberg, "On the problem of obtaining area functions from lateral X-ray pictures of the vocal tract " *STL QPSR.*, pp. 43-45, 1969.

[9] P. Perrier, L. J. Boe, and R. Sock, "Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: modeling the transition with two sets of coefficients. ," *J. Speech Hear. Res.*, vol. 35, pp. 53-67, 1992.

[10] A. Soquet, V. Lecuit, T. Metens, and D. Demolin, "Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with MRI," *Speech Communication*, vol. 36, pp. 169-180, Mar 2002.

[11] S. Masaki, M. K. Tiede, K. Honda, Y. Shimada, I. Fujimoto, Y. Nakamura, and N. Ninomiya, "MRI-based speech production study using a synchronized sampling method," *J. Acoust. Soc. Jpn.(E)*, vol. 20, pp. 375-379, 1996.

[12] H. Takemoto, K. Honda, S. Masaki, Y. Shimada, and I. Fujimoto, "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," *Journal of the Acoustical Society of America*, vol. 119, pp. 1037-1049, Feb 2006.

[13] G. Fant, *Acoustic Theory of Speech Production: With Calculation Based on X-Ray Studies of Russian Articulation*. Mouton: The Hague, 1960.

[14] J. Flanagan, L., *Speech Analysis Synthesis and Perception*. New York: Spinger, 1972.

[15] J. Dang and K. Honda, "Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation," *Journal of the Acoustical Society of America*, vol. 100, pp. 3374-3383, 1996.

[16] B. H. Story, "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002," *The Journal of the Acoustical Society of America*, vol. 123, pp. 327-335, 2008.

[17] J. M. Heinz and K. N. Stevens, "On the Derivation of Area Functions and Acoustic Spectra from Cineradiographic Films of Speech," *Journal of the Acoustical Society of America*, vol. 36, p. 37, 1964.

Statistical Correlation Analysis between Lip Contour Parameters and Formant Parameters for Mandarin Monophthongs

Junru Wu[1], Xiaosheng Pan[1], Jiangping Kong[1], Alan Wee-Chung Liew[2]

1Linguistic Lab, Dept. of Chinese Language & Literature,

Peking University, Beijing, China

2School of Information & Communication Technology, Griffith University,

Gold Coast Campus, QLD 4222, Australia

Email: yuerr@163.com, itol_xs@pku.edu.cn, kongjp@gmail.com, a.liew@griffith.edu.au

Abstract

In this study we examine quantitatively the correlation between the geometric lip contour parameters and the formant parameters for Mandarin monophthongs, and carry out a multiple linear regression study between the two parameters. We explicitly analyze the relationship between different geometric lip parameters and the formant parameters, which have some linguistic significance instead of the usual acoustic parameters such as MFCC. We analyzed the linguistic meaning of the regression formula, and found it in accord with the classical result on the relationship between vocal-track and speech acoustics. And those regions with relatively poor effect of estimation are related to specific phonetic conditions. **Index Terms**: facial motion, monophthong, canonical correlation analysis, multiple linear regression

1. Introduction

The correlation between acoustics and visual features of speech is a fundamental issue in the field of audiovisual speech processing and an important aspect in phonetics. To examine this correlation, effective methods for the extraction and quantification of visual lip parameters, as well as statistical models are necessary.

In this study we examine quantitatively the correlation between the geometric lip contour parameters[1] and the formant parameters for Mandarin monophthongs, and carry out a multiple linear regression study between the two sets of parameters.

Several questions associated with speech acoustics and the geometric lip contour parameters are addressed in this paper: (1) what is the explicit relationship between different geometric lip parameters and formant parameters? (2) Is the non-linearity that exists between the two sets of features (lip parameters and formant parameters) distributed randomly or in accordance with some phonetics rules?

The widely cited study of Yehia, et al in 1998 [2] examined the degrees of correlation among vocal-tract and facial movement data and the speech acoustics. Using two corpuses of sentences in two languages, the 3D position data of markers placed on the face and in the vocal-tract was extracted. LSP coefficients and RMS amplitude of the signal were extracted from the acoustic signals in a separate experimental session. After temporal aligning the frames of different sets, for each two set of parameters linear estimation was carried out to build an estimator to recover a matrix, and then the mean correlation coefficient between measured and recovered vectors was measured through all possible combinations of training and test data. It was noted that estimation of the speech acoustics (f) from facial measures (x) is considerably better than from vocal-tract measures (y). Then, dimensionality analysis was carried out. Principle Component Analysis was used to reduce the dimensionality. However, there is not a priori reason to believe that the dimensionality reduction achieved in one space is optimum for describing the data measured in another space. Thus the authors perfomed singular value decomposition to map the data for the vocal-tract, facial and acoustic spaces onto a common coordinate system. It shows that between 4 to 8 components are sufficient to represent the mappings examined. In the discussion, the authors noted that in their study no temporal analyses were done; all correlations are based only on spatial properties of the data. They believe that the resulting global correlations suggest that correlated tongue-jaw behavior is basic to producing all speech rather than the result of some higher level phonetic control.

In later studies other data sources and parameters were tried. 2D face motion data has been examined by Almajai[3], Barker[4], Barbosa[5] and Jiang[6]. Some researchers tried to combine the study with visual feature extraction technique, for example 2-D DCT and cross-DCT (Almajai et al.) and 'Chroma-Key' processing (Barker et al.). As for Yehia's later study with Barbosa, a search algorithm is used for tracking the 2D facial motion of markers painted on the speaker's face.

As for acoustic data, Barker[4] tried LP, LSP and RMS parameters, and showed that the strongest correlations are achieved using the LSP parameterization. Almajai et al. used mel-scale filterbank vectors and the first four formant frequencies extracted using a combined linear predictive analysis-Kalman filter. Formant frequencies are closely related to speech production and correspond to resonant frequencies in the vocal tract. It has been shown that filterbank features exhibit higher correlation to visual features than formant frequencies. However, mel-scale filterbank parameters are hard to explain linguistically. Moreover, there

is a lack of good method to extract linguistically meaningful formants in from these parameters.

As Yehia et al. [2] mentioned in their study, certain type of non-linearity exists between the acoustics and the visual features of speech. As Barker et al. [4] mentioned, Examination of the error distributions of the LP parameters reveals them to be multimodal i.e. clearly non-Gaussian. This shows that the linear estimates are essentially an inadequate model of the true mapping. No temporal analyses were done in [2]. Later studies tried to address these two points. Barbosa and Yehia[5] used time-invariant and time-varying linear models, as well as nonlinear (neural network) models (Levenberg-Marquardt algorithm) in their study. As a result, the correlation coefficients between measured and estimated trajectories are as high as 0.95. This estimation of facial motion from speech acoustics indicates a way to integrate audio and visual signals for efficient audiovisual speech coding. On the other hand, relatively little studies from the perspective of linguistic phonetic principles have been done on this subject. In fact, there is complicated structure inside the so called 'non-linearity' related to linguistic phonetic notions.

Barker et al. [4] used a corpus of isolated nonsense words having a VCCV vowel-consonant structure (the systematic structure of the corpus allows the audio-visual correlation to be separately analyzed for both consonants and vowels). When examining the size and distribution of the errors, it is found that the lips can be more reliably reconstructed during vowels than during consonants. Jiang et al. [6] noted that, in general, predictions for CV syllables are better than those for sentences. Almajai et al. [3] measured the audio-visual correlation within each phoneme and then averaging the correlation across all phonemes and compared the result to the measurement across the whole corpus of sentences. As a result, there is an increase in correlation to R=0.9 when the audio-visual correlation is measured within each phoneme. All these indicated that the linearity of correlation is higher inside each phoneme than across different phonemes. In other words, different phonemes have different audio-visual correlations. A universal model would do well in some phonemes but do quite poorly in other phonemes.

2. Experiment

Authors should observe the following rules for page layout.

2.1. Database

The experiments for data acquisition are carried out for one female speaker of Chinese Mandarin. The data are acquired using a corpus of 61 Mandarin initials.

Mandarin has 22 initials. We choose 3 monophthongs /a, i, u/ for each initial and get 61 syllables (see Table 1). For those syllables that don't exist in Mandarin phonemic system the vowels /i/, /ɿ / or /ʅ / /y/ are used instead.

G1	b /p/	p /ph/	m /m/	f /f/	**G2**	d /t/	t /th/	n /n/	l /l/
a	ba1	pa1	ma2	fa1	a	da1	ta1	na2	la1
i	bi1	pi1	mi2	—	i	di1	ti1	ni2	li1
u	bu1	pu1	mu2	fu1	u	du1	tu1	nu2	lu1

G3	g /k/	k /kh/	h /x/	**G4**	j /tɕ/	q /tɕh/	x /ɕ/	**G7**	∅
a	ga1	ka1	ha1	a	jia1	qia1	xia1	1	a
i	—	—	—	i	ji1	qi1	xi1	2	i
u	gu1	ku1	hu1	u	ju1	qu1	xu1	3	u

G5	zh /tʂ/	ch /tʂh/	sh /ʂ/	r /ʐ/	**G6**	z /ts/	c /tsh/	s /s/
a	zha1	cha1	sha1	—	a	za1	ca1	sa1
u	zhu1	chu1	shu1	ru4	u	zu1	cu1	su1
/ʅ /	zhi1	chi1	shi1	ri4	/ɿ /	zi1	ci1	si1

Table 1: The 61 syllables used in the study (Group1 (G1) Bilabial, Group2 (G2) Alveolar, Group3 (G3) Velar, Group4 (G4) Coronal, Group5 (G5) Retroflex, Group6 (G6) Alveolar2)

These AVI clips are part of an existing audio-visual corpus recorded by our laboratory in Peking University. The two domains of data are taken simultaneously using Adobe Premiere Prof 1.5 in AVI format (Video: 720×576 pixel, 24bits, 25 fps; Audio: 639kbps, PCM 16bits, 32 kHZ, 1024kbps) and then separated using Virtual Dub 1.7.0.1c1.x by Avery Lee. Then the Audio was re-sampled to 11025 HZ, and the area of Lip (96×80 pixel) was extracted from the original video.

2.2. Parameterization

At this point, the data available are the audio and video signal. This section describes suitable parametric representations that will help in the study of the relationship between the two domains.

2.2.1. Lip contour extraction

The lip contour parameters are characterized by using the deformable template by Liew et al. [1] from video images of the speaker's face. The parameters of the model are adjusted manually to correct any fitting error since the aim of this study is to investigate the relationship between visual lip features and monophthongs, instead of lip segmentation methodology.

In [1], a robust deformable model-based technique for lip contour extraction from a color RGB lip image is proposed. The method uses a region-based stochastic cost function to find an optimum partition of a given lip image into lip and nonlip regions. Spatial fuzzy clustering using both luminance and chrominance features from the CIELAB and CIELUV color spaces is used to produce a probability map of the lip image. The optimum model parameters are then found by performing a conjugate gradient search on the cost function. Extensive experimental results show the feasibility of their approach.

Given a lip model as shown in Figure 1 , the equations describing the lip shape are given by [1]:

$$y_1 = \frac{-h_1}{(w - x_{off})^2}\left(|x - sy_1| - x_{off}\right)^2 + h_1 \qquad (1)$$

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

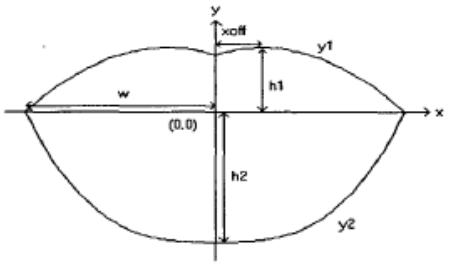


Figure 1: Geometric lip model

The parameters we used are a little different from the one presented in [1]. In this study we added the inner lip parameters into the lip model of [1]. Our lip model are as shown in Fig.2 and is parameterized as follows: the horizontal position of the lip in the rim (y1), the vertical position of the lip in the rim (y2), the width of the outer contour of the lip (y3), the distance from the lower outer contour to the level line (y4), the distance from the upper outer contour to the level line (y5), the degree of concavity of the philtrum (y6), the curvature of the lower lip counter (y7), the obliquity of the lip (y8), the width of the inner lip (y9), the distance from the lower inner contour to the level line (y10), the distance from the higher inner contour to the level line (y11), the Skewness of the lip (y12), enantiomorphism (y13). Finally the set of lip parameters are given by

Y= {y1, y2, y3, y4, y5, y6, y7, y8, y9, y10, y11, y12, y13} (3)

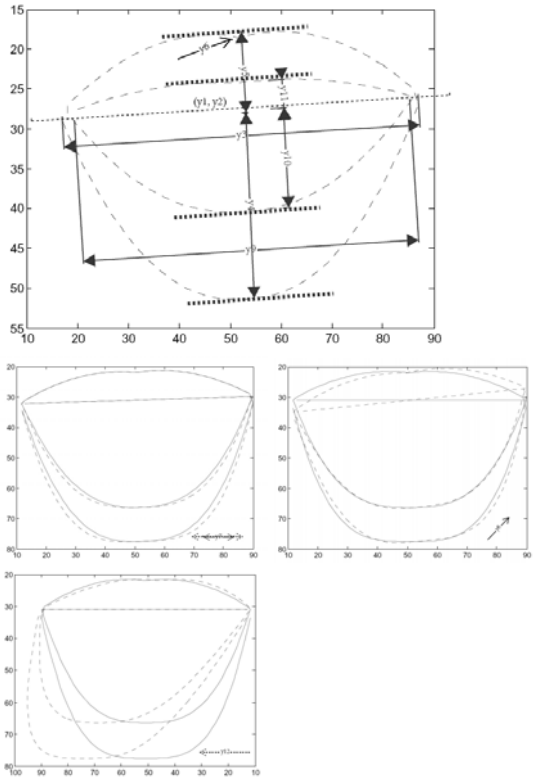


Figure 2: Geometric lip model (13 parameters). First plot: y1, y2, y3, y4, y5, y6, y9, y10, y11; Second plot: y7; Third plot: y8; Last plot: y12

2.2.2. Formants

The first three formant parameters (fmt1, fmt2, fmt3) are obtained from audio data taken simultaneously with the video data using the method of LPC (period in-phase), using *Wavefinal*, which is written by our laboratory. The boundary of the consonant and the vowel in a syllable is manually marked. Only the vowel frames are used:

F= {fmt1, fmt2, fmt3} (4)

2.2.3. Alignment

Since the video sampling rate of 25 fps is lower than the audio sampling rate (11025HZ), we performed appropriate time alignment of the two sets of parameters. The formants taken using LPC is period in-phase, and the interval between pulses is always changing, which is known as jitter. As a result, the frames of formant data don't have identical time length, and an indefinite number of formant frames are related to a frame of lip contour parameters.

We calculate the average of F through all the frames of formants and align it with the relative frame of lip contour parameters. The relative bandwidth (bw) is also extracted. For those formant frames at the beginning or end of the vowel with null, the next or the former data was filled into the null. Table 2 shows a frame of the aligned parameters.. 800 frames of data are obtained from the vowel parts of these syllables.

pinyin	**mark_inx**	**y1**	**y2**	**y3**	**y4**	**y5**	**y6**	
a1	1	47.9	38.6	34.33	39.43	11	6.84	…

y7	**y8**	**y9**	**y10**	**y11**	**y12**	**y13**	
1.1	0.03	33	25.95	0.62	0	0	…

fmt0	**fmt1**	**bw1**	**fmt2**	**bw2**	**fmt3**	**bw3**	**lipmark_idx**	**position**
239	1204	69	1702	94	3227	122	24	0.96

Table 2. Tthe first frame of data from /a1/

2.3. Canonical correlation analysis

Canonical correlation was first established by Hotelling in 1936 to analysis the correlation between two sets of random variables. The idea of reducing dimensionality is borrowed from PCA. The correlation between two sets of parameters is reduced to the correlation between two canonical variables.

We perform statistical matrix analysis on the two sets of parameters. The correlation analysis result for the set of lip parameters (Set1) shows that many of the lip parameter pairs have Pearson correlation coefficient larger than 0.5 (y2-

y3,y2-y5,y2-y9,y3-y4,y3-y5,y3-y6,y3-y9,y3-y11,y3-y12,y4-y6,y4-y9,y5-y6,y5-y9,y5-y11,y6-y9,y6-y11,y6-y12,y9-y11,y9-y12). For the set of formant parameters (Set2), no Pearson correlation coefficient is larger than 0.5. This shows that there is much redundancy between the lip parameters whereas the formant parameters are largely independent. In the correlation between the set of lip parameters and the set of formant parameters, several pairs have Pearson correlation coefficient larger than 0.5 (y3-fmt1 (R=0.5712), y3-fmt2 (R=0.6109), y4-fmt1 (R=0.7571), y5-fmt2 (R=-0.5826), y9-fmt1 (R=0.5818), y9-fmt2 (R=0.6461), y10-fmt1 (R=0.7414)), indicating that there are strong relationships between the two sets of parameters. Canonical Correlation analysis is performed on the two sets of parameters and the results are verified using hypothesis testing (Wilk's and Chi-Sq. tests). The first canonical correlation coefficient (L1-F1) is 0.911 which is larger than any correlation coefficient between any two individual variables from Set1 and Set2. Standardized canonical coefficients for Set1 indicates that L1 = 0.018y1- 0.039y2+ 0.328y3+ 0.119y4+ 0.049y5- 0.046y6- 0.013y7- 0.086y8- 0.972y9- 0.531y10- 0.291y11+ 0.047y12, whereas standardized canonical coefficients for Set2 indicates that F1=-0.778fmt1-0.648fmt2+0.052fmt3.

Canonical loadings for Set1 shows that y3, y4 y6, y7, y8, y9, y10 have negative correlation to L1. Canonical loadings for Set2 shows that fmt1, fmt2, fmt3 have negative correlation to F1. Cross loadings for Set1 shows that y3, y4, y6, y9, y10, y12 can be better estimated by F1，y5 can also be estimated by F1 to certain extent. Cross loadings for Set2 shows that fmt1, fmt2 can be better estimated by L1. The difference of sign between canonical coefficient and canonical loadings of y3, y4, y11 indicates that they may be compensation parameters for y9, y10, y5.

Redundancy analysis shows that the proportion of variance of Set1 explained by its own canonical variant is 41.7%. The proportion of variance of Set2 explained by its own canonical variant is 34.3%. The proportion of variance of Set1 explained by opposite canonical variant is 34.6%. The proportion of variance of Set2 explained by opposite canonical variant is 28.5%. Hence, the formant parameters are better in explaining the lip parameters than vice versa.

Canonical correlation analysis is later carried out separately between the inner lip parameters (y9-y10-y11) and formant parameters or outer lip parameters (y3-y4-y5) and formant parameters. The parameters in the set are chosen in accordance with the above-mentioned CCA. It was found that the set of inner lip parameters is also better in explaining F than the set of outer lip parameters.

2.4. Multiple Linear Regression

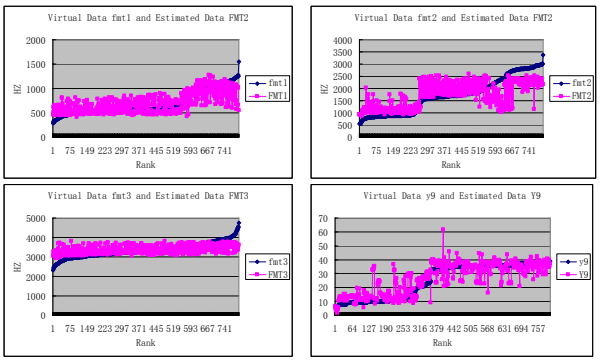
Multiple linear regression (MLR) is used to fit the linear combination of the components of the multiple-dimension vector L, i.e. (y3, y4, y5) or (y9, y10, y11) as independent variables, to the single-dimension vector fmt1, or fmt2, or fmt3, which is the dependent variable. It is also used to fit the linear combination of the components of the multiple-dimension vector F (fmt1, fmt2, fmt3) as the independent variables to the single-dimension vector, i.e. y3, or y4, or y5, or y9, or y10, or y11, as the dependent variable. In this process, the method of stepwise regression is used to exclude the independent variables which do not fit the criterion and include the independent variables which contribute most to the dependency. So the independent variable fmt3 is excluded from the regression formula of F and y3, y9 is excluded from the regression formula of L (y9, y10, y11) and fmt1, y11 is excluded from the regression formula of L (y9-y10-y11) and fmt2. (see Table 3)

	formula	R	AESq	coll
1	fmt1=25.024y4+19.685y5+8.758y3-494.021	0.779	0.606	+
2	fmt2=89.974y3-34.583y4-47.513y5+296.385	0.683	0.645	+
3	fmt3=-46.532y5-15.672y4+12.456y3+3919.007	0.483	0.230	+
4	y3=0.005fmt2+0.013fmt1+14.838	0.838	0.701	
5	y4=0.022fmt1+0.004fmt2-0.001fmt3+10.729	0.821	0.673	
6	y5=-0.002fmt2-0.002fmt3-0.003fmt1+25.022	0.647	0.417	
7	fmt1=24.461y10+26.840y11+392.405	0.792	0.626	+
8	fmt2=65.313y9-41.547y10+578.949	0.698	0.486	+
9	fmt3=-45.985y11-19.298y10+14.908y9+3225.469	0.513	0.260	+
10	y9=0.011fmt2+0.028fmt1+0.002fmt3-18.616	0.872	0.760	
11	y10=0.026fmt1+0.005fmt2-0.001fmt3-12.254	0.858	0.736	
12	y11=-0.001fmt2-0.003fmt3-0.001fmt1+11.745	0.546	0.295	

Table 3. Regression formula, where ARSq denotes adjusted R-Square, coll denotes collinearity.

Similar to the result of canonical correlation analysis, the formant parameters works better than the lip parameters, and the inner lip parameters works better than the outer lip parameters as the independent variables in that the Adjusted R-Square is larger.

Figure 3 shows all the frames of virtual data (i.e., predicted from regression formula) and estimated data (LPC estimated) when the virtual data is ranked in ascending order.



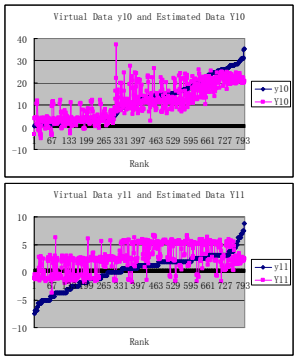


Figure 3. All the frames of virtual data and estimated data. The virtual data are ranked in ascending order.

3. Discussion

3.1. The phonetic meaning of the coefficients

The sign of the coefficients in the regression formula show that the larger the opening of the lower lip is, the larger the first formant is; the higher the opening of the upper lip is (from the level line), the larger the first formant is; the wider the lip is, the larger the second formant is; the smaller the opening of the lower lip is, the larger the second formant is; the wider the lip is, the larger the third formant is; the smaller the opening of the upper lip is, the larger the third formant is; the smaller the opening of the lower lip is, the larger the third formant is. The finding is in harmony with the study of speech articulation and formant frequencies in linguistic research, which state that the first formant is positively related to the opening aperture of the mouth, the second formant is negatively related to the posteriori of the tongue and the roundness of the mouth, and the second and the third formant come closer when the lip is rounded. Figure 4 shows the fmt1, fmt2, and fmt3 plot for different vowels, whereas Figure 5 shows the plot for different vowels as a function of lip parameters y9, y10, and y11. It can be seen that the vowels can be separated to some extend based on the three lip parameters, although the separation is not as good as that using fmt1, fmt2, and fmt3.

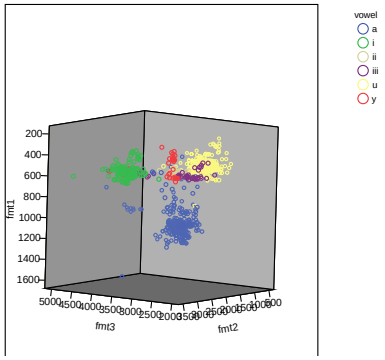


Figure 4. fmt2 as x-axis, fmt1 as y-axis, fmt3 as z-axis

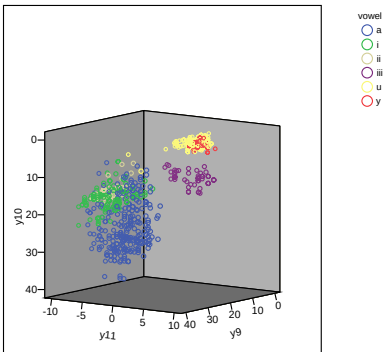
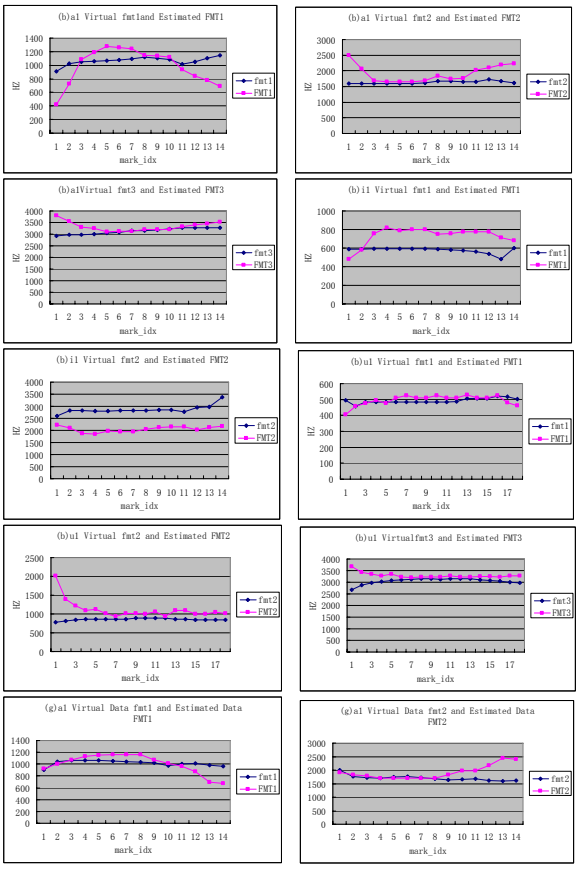


Figure 5. y9 as x-axis, y10 as y-axis, y11 as z-axis

3.2. The Distribution of Non-linearity

Although the MLR analysis reveals some useful insights, the linearity assumption has its limitation. First, we observed that the distribution of fmt1, fmt2 is far from being Gaussian. Second, there is noticeable co-linearity between y3 and y4, y3 and y5, y9 and y10, y9 and y11 and the set of outer lip parameters and inner lip parameters can be recovered from each other. Third, the distribution of the residuals is not Gaussian, indicating that there is residual correlation not extracted by MLR. This residual correlation may be due to the non-linear relationship between the two sets of parameters.

It is important to note that this non-linearity neither distribute randomly within a phoneme nor across different phonemes. Figure 6 shows some examples of the virtual and estimated formant data in specific vowels.



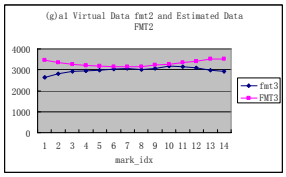


Figure 6 the virtual and estimated data in (b)a, (b)i, (b) u, and (g)a

First, the position of a specific frame in the vowel influences the relation between its lip and formant parameters. We could see from the figures that the linear estimators do quite well in the middle of the vowel but poorly at the beginning or at the end of the vowels. The estimated loci usually take the shape of an arc, which is related to the loci of the lip parameters which indicate the movement feature of the lips. While pronouncing, the muscles first move fast in acceleration to a specific point, then keep the state until the fast release comes. At the same time, the formants do not experience such dramatic change but keep a relatively stable state.

Second, the estimators perform differently depend on the interaction between the vowel and the initial circumstance. As for /a/, the changes of lips are very sensitive to the initial circumstance. As shown in the figure, the lip parameters change dramatically after (b)/p/, but not so much after (g)/k/, which lead to distinctly different estimated formant loci between these two /a/s, while the virtual formant loci are less different. This is obviously not a linear correlation. In /u/, the state of lips does not change much while pronouncing this vowel, regardless of the initial circumstance, so linear estimators fit well.

From the viewpoint of phonetics and the study of vocal tract, different articulation can be used to produce the same set of formants. For example, to produce /u/, which is marked with lower second and third formant, the speaker may have more posterior tongue position and less rounded lip or more rounded lip but less posterior tongue position. On the other hand, since the same lip shape may combine with different tongue position, different formant structures can be obtained. For example, /u/ and /y/ have almost the same lip contour parameters but different formant set. This may explain why it is not as effective to estimate formants from lip contour parameters as vice versa. Hence, combining it with the study of vocal track may be a useful attempt. [7-9]

4. Conclusions

Canonical Correlation Analysis and Multiple Linear Regression are carried out across the lip contour parameters and formant parameters of Mandarin monophthongs and the effect is evaluated. It was observed that formant parameters works better than the lip parameters, the inner lip parameters works better than the outer lip parameters in estimation, and the first and second formants are better estimated than the third one. These are in accordance with former studies. It was also found that the phonetic meanings of the coefficients of MLR are in harmony with the study of speech articulation and formant frequencies in linguistic research. The distribution of non-linearity of the correlation is not random, but influenced by the position of the frame in the vowel and the interaction between vowel and initial circumstance. Future work would be to incorporate this findings into the automatic speech and lipreading application of Mandarin language.

5. References

A.W.C. Liew, S.H. Leung, and W.H. Lau, "Lip Contour Extraction from Color Images Using a Deformable Model", Pattern Recognition, Vol. 35(2), pp. 2949-2962, 2002.

H. Yehia, P. Rubin, and E. Vatikiotis-Bateson, "Quantitative association of vocal-tract and facial behavior," Speech Communication, vol. 26, pp. 23-43, 1998.

I. Almajai and E. Milner, "Maximising Audio-Visual Speech Correlation", proceedings of the Auditory-Visual Speech Processing 2007 (AVSP2007), Kasteel Groenendaal, Hilvarenbeek, The Netherlands, 2007.

J. P. Barker and F. Berthommier, "Evidence of Correlation Between Acoustic and Visual Feature of Speech," Proceedings of the ICPhS '99, San Francisco 1999.

A. V. Barbosa and H. C. Yehia, "Measuring the relation between speech acoustics and 2D facial motion", Proceedings of the IEEE Int. Conf. Acoustics, Speech, Signal Processing, Salt Lake City, Utah, USA, 2001.

J. Jiang, A. Alwan, L. E. Bernstein, P. Keating, and E. T. Auer, "On the correlation between face movements, tongue movements, and speech acoustics," EURASIP Journal on Applied Signal Processing, vol. 11, pp. 1174-1188, 2002.

汪高武, 鲍怀翘, and 孔江平, "从声道截面积推导普通话元音共振峰," 中国语音学, 商务印书馆, 北京, vol. 第一辑, 2008.

G. Wang, T. Kitamura, X. Lu, J. Dang, and J.P. Kong, "MRI-based Study on Morphological and Acoustic Properties of Mandarin Sustained Vowels ", Journal of Signal processing, 2008, in press.

G. Wang, "A Stuy of Mandarin Chinese Using X-ray and MRI," Journal of Chinese Phonetics (中国语音学报) , 2008.

喉头仪(Electroglottography)

尹基德、吉永郁代

1. 喉头仪的特点和原理

喉头仪（Electroglottography[1]，简称是EGG）的基本原理，把一双电子感应片分别固定在喉结两边，开电流从一个电子感应片发送，另一个接收，然后处理器把所测到的电流变化形状同时显示而纪录。喉头仪的特点就是对身体完全无害。在它的动作原理上只用小量的电流，被试者不会感觉到电流的刺激。在 1957 年Fabre起初使用喉头仪以来，喉头仪在桑音研究和医疗方面广泛使用。

图 1）喉头仪

图 2）中最上面的图指的是喉头部分，表面的两个黑色的是喉头仪的两个电子感应片（电极），两个电子感应片之间复数的细线指的是喉头仪的测定范围。为了避免伤害身体或者过敏的生理反应，需要保持低电流。在一般情况下，除非电流很强，感觉不到高频率。因此在喉头仪使用微弱的高频交流电，测量组织的阻抗。声门是空气充满的空间，空气是一种电流不容易通过的媒体。所以声门打开时两个电极之间的电气阻抗值上升，声门关闭时下降。电极流动不是一个方向，电流的流动难推测，即电流不仅经过声带，而它经过其他组织。结果声带开关时的阻抗值差异达到喉头整个电极的 1~2%。喉头仪把喉头周边的动态做最小化，把声门信号做最大化来显示。

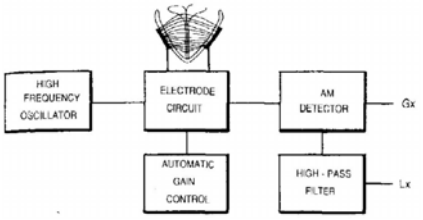


图 2）喉头仪装置的结构（Baken，1992）

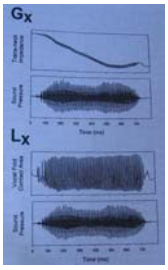


图 3）Gx 和 Lx(高频处理后)

图 4）表示喉头仪信号和声门开关的关系。喉头仪信号反映出声带接触情况。

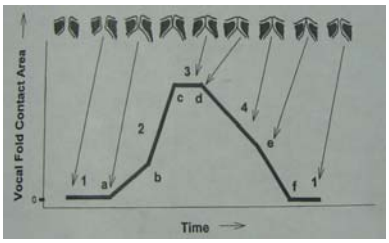


图 4）喉头仪信号和声门开关的关系(Baken,1992)

2. 喉头仪信号的可信度

喉头仪信号跟其它声门测量方法的对比研究不少。这些研究明确喉头仪信号和声带运动的关

[1]喉头仪测量的是声带的接触区间，所以严格来说 Laryngography 比 Electroglottography 更准确。

系。

下面的图 5)、图 6)、图 7）表示 stroboscopy 测的声带振动形状和同时喉头仪信号之间的对应。图 5）是正常嗓音的表现。喉头仪信号的曲线上突然上升表示声带的关闭。从喉头仪信号和声压来可以限定声带的接触段，加上能够分开声门上和声门下的共鸣。在连续照片的第 7、8 张上看到的粘膜波动是在正常嗓音常见的特点。图表 6 的气嗓音的特点之一是其长开相。长开相引起减少共鸣。气嗓音的关闭相没有正常嗓音的快速度。图表 7 的假声的特点是声带的伸长、粘膜波动的欠缺、喉头仪信号的正弦曲线。声带关闭速度比较慢，结果把高频减低，产生很简单的声波。

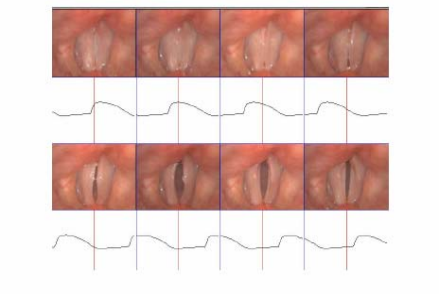


图 5）正常嗓音；男性 120Hz (Fourcin)

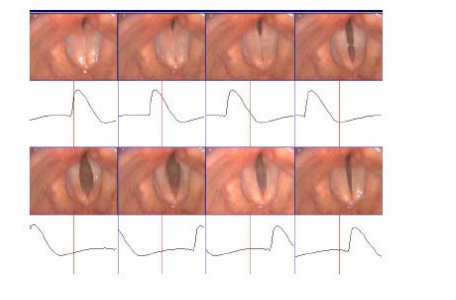


图 6）气嗓音；男性 105Hz (Fourcin)

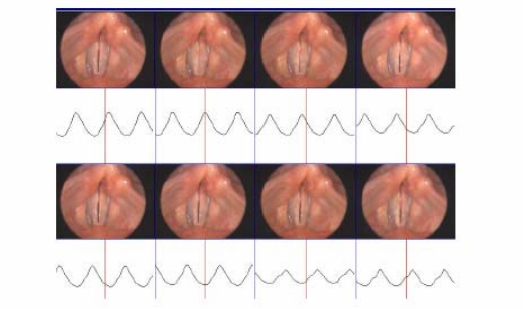
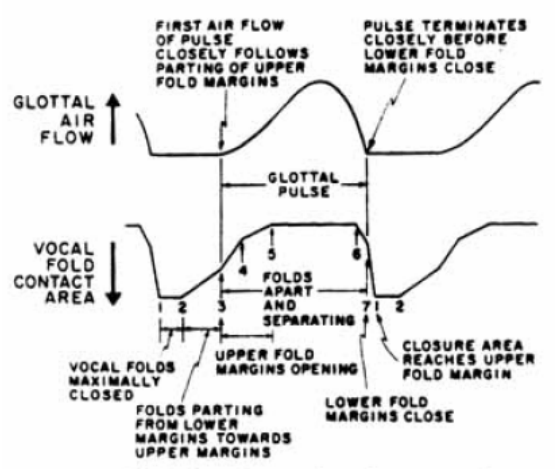


图 7）假声；男性 323Hz (Fourcin)

下面图 8）指的是喉头仪信号的倒谱和声门气流的关系。声门脉搏的结束部分对应于声带接触开始锐减的部分（图表 8 种第 7 地点），即可以说第 7 地点是声带接触的开始点。从第 2 点到第三点，声门气流是平的，一方喉头仪信号有变化。这阶段是声带从下部慢慢地开始打开的阶段或者可以说声带没有接触但非常接近状态。总之从第 3 点到第 7 点是声门打开的阶段，从第 7 点到第 3 点是声门关闭的阶段。



图表 8 喉头仪信号的倒谱和声门气流的关系 (M.Rothenberg,1979)

喉头仪信号和其它声门测量方法之间的对比研究结果支持喉头仪信号和声带的实际运动状况有密切的对应关系。不过，对猴头仪信号的可信度也有怀疑的看法。Baken(2000)指出喉头仪信号本身是声带接触的显示，缺乏声门开放以后情况的信息，不能被运用于声门宽度有关的研究。 开商，速度商等的参数，只有在正常嗓音的研究上确保一定的可信度。但其它非正常发声类型来说，不能同一个分析方法来对比。

3. 喉头仪信号分析方法

声门开关的一个周期可以分成闭相（Closed

Phase）和开相（Open Phase）的两个阶段。

闭相：从声门关闭点到声门开启点。

开相：从开启点到关闭点。

图表 9 指的是关闭点（contacting event）和开启点（de-contacting event）。

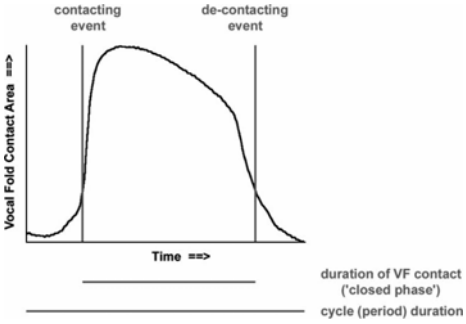


图 9）定接触点测接触商(Herbst,2004)

闭相可以分成关闭相（Closing Phase）和开启相（Opening Phase）。

关闭相：从关闭点到喉头仪信号最高点的阶段。

开启相：从喉头仪信号最高点到开启点。

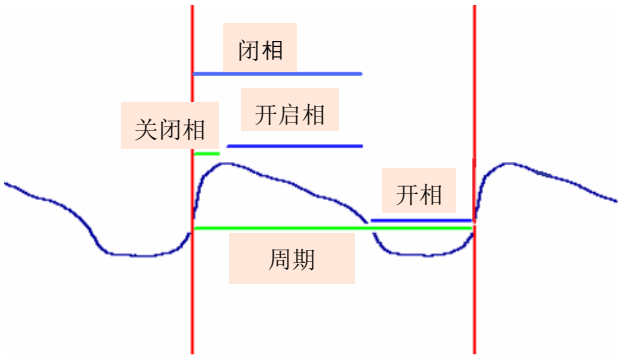


图 10）喉头仪信号主要参数

在语音研究当中主要使用的参数有接触商，接触商（Contact Quotient），开商（Open Quotient）、速度商（Speed Quotient）。

接触商 = 闭相 / 周期

开　商 = 开相 / 周期

速度商 = 开启相 / 关闭相

关于采取喉头仪信号参数的方法以往有不少讨论，为了测量开商或者接触商时的接触点和开启点的设定是其讨论之一。每个发声类型的声带关闭和开启的斜率趋势是不一样的，当设定接触・开启点时需要注意这一点。下面介绍一下三个主要分析方法。

3.1 尺度方法(criterion-level method)

Rothenberg 的‘尺度方法(criterion-level method)’是喉头仪信号的最大值和最小值为标准，设定 20~50%内一定的尺度，照此决定开启/关闭点。Kay Real-time EGG analysis 照 Rothenberg 的定义，算出的是‘相对的接触(relative vocal fold abduction)’。不同研究设定不同尺度，因而它们之间的直接对比没有意义。

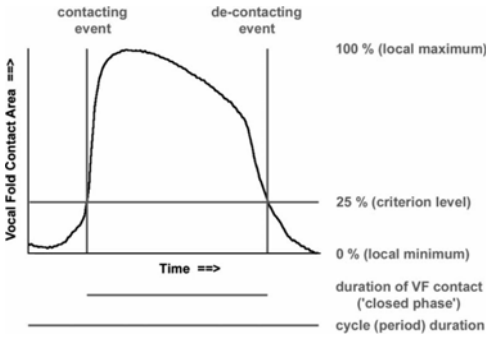
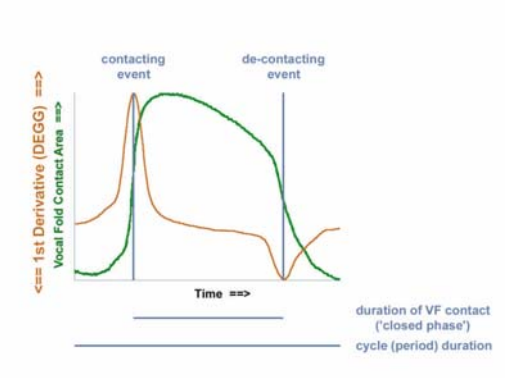


图 11）用尺度方法测接触点和开启点(Herbst,2004)

3.2 喉头仪信号微分方法

Childers et al.(1986)运用的是喉头仪信号微分（Derivative-EGG，简称是 DEGG)方法，就是用 EGG 的一介导数（微分）测出声门的关闭点和开启点的方法。Henrich et al.(2004)的 DECOM(Derivative-EGG Correlation-based method for Open quotient Measurement)是 微分信号上的上下峰(opening/closing peak)对应于声门的关闭点和开启点的方法。这些研究证明喉头仪微分信号的上下峰相当准确地反映声门的关闭点和开启点。



图表 12 微分信号上的接触点和开启点(Herbst,2004)

3.2.1 微分方法的问题

虽然以喉头仪微分信号的上下峰为关闭点和开启点是最理想的，但在主要三个情况下遇到问题；（1）下峰不明确、（2）上下峰的双峰情况、（3）高频噪音。参照图 13）。

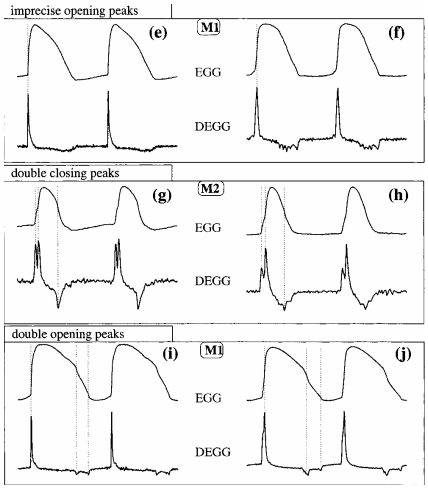


图 13）喉头仪信号和其微分信号（Henrich，2003）

3.3 混合方法(hybrid method)

Howard(1990，1995)的混合方法(hybrid method)是关闭点以喉头仪微分信号的上峰为标准，开启点以尺度方法为标准，测量声门的关闭点和开启点的方法。

4. EGG 有关代表性的研究

4.1 喉头仪的原理与使用方法

Childers and Krishnamurthy(1985)

I.R Titze(1990)

Baken(1992)

Baken and Orlikoff(2000)

4.2 喉头仪信号的可信度问题（喉头仪信号跟其它声门测量方法的对比研究）

Fourcin(1974)

Lecluse et al.(1975)

Pedersen(1977)

Teaney and Fourcin(1980)

Anastaplo and Karnell(1988); Karnell(1989)

Baer et al.(1983a)

Childers et al.(1983b,1984,1990); Childers and Krishnamurthy(1985); Childers and Larar(1984)

Baer et al.(1983a,b)

Berke et al.(1987)

Dejonckere(1981)

Gerrat et al.(1988)

Kitzing(1977, 1983); Kitzing et al.(1982)

Titze et al.(1984)

Fourcin(1981)

Rothenberg(1981); Rothenberg and Mahshie(1988)

4.3 喉头仪信号分析方法

Childers et al.(1986)

Howard(1990，1995)

Henrich et al.(2004)

C. Herbst(2006)

4.4 运用喉头仪的嗓音研究

孔江平(2001)

Askenfelt et al.(1980)

Kitzing(1982)

Lecluse(1977)

Lecluse and Brocaar (1977)

Roubeau (1993); Roubeau and Castellengo

(1993); Roubeau et al.(1987)

Henrich et al.(2004)

4.5 喉头仪的嗓音研究历史

Raymond H. Colton and Edward G. Conture (1990)

5. 喉头仪产品

目前已有几个研究单位和医疗仪器商推出的喉头仪设备，这里介绍 4 种仪器和它们的网站。

5.1 F-J Electronics Portable

(www.f-jelectronics.dk，丹麦)

5.2 Glottal Enterprises

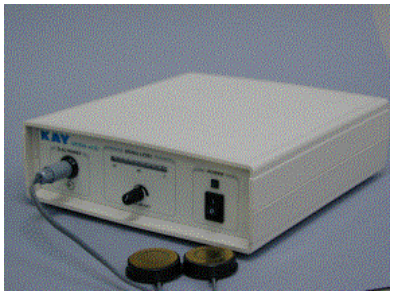
(www.glottal.com，美国)

5.3 Laryngograph Ltd.

(www.laryngograph.com, 英国)

5.4 KayPENTAX

(www.kayelemetrics.com，美国)



参考文献

孔江平, 论语言发声, 中央民族大学出版社, 2001.

Alan M.Smith, D.G. Childers, Laryngeal Evaluation Using Features from Speech and the Electroglottograph, IEEE Transactions on biomedical engineering, 1983;30;11:755-759.

Childers et al., A model for vocal fold vibratory motion, contact area, and the electroglottogram, Journal of Acoustic Society of America, 1986.

R.J.Baken, Robert F.Orlikoff, Clinical Measurement of Speech and Voice, 2000:413-427.

M.Rothenberg, Some Relations Between Glottal Air Flow and Vocal Fold Contact Area,Proceedings of the Conference on the Assessment of Vocal Pathology, ASHA Reports, 1979;11:88-96.

Henrich,N., On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation, Journal of Acoustic Society of America, 2004.

I.R Titze, Interpretation of the Electroglottographic Signal, Journal of Voice, 1990; 4;1:1-9.

Fourcin, A.J. Precision stroboscopy, voice quality and

Electrolaryngography.
David M.Howard, Geoffrey A.Lindsey, and Bridget Allen, Toward the Quantification of Vocal Efficiency, Journal of Voice, 1990; 4;3:205-212.
David M.Howard, Variation of Electrolaryngographically Derived Closed Quotient for Trained and Untrained Adult Female Singers, Journal of Voice, 1995; 9;2:163-172.
Henrich,N.,R,Bernard.,and Castellengo,M., On the electroglottography for characterization of the laryngeal mechanisms, Proceedings of the Stockholm Music Acoustics Conference, 2003.
M. Rothenberg, Some Relations Between Glottal Air Flow and Vocal Fold Contact Area, Proceedings of the Conference on the Assessment of Vocal Pathology, ASHA Reports, 1979;11:88-96.
M. Rothenberg and James J. Mahshie, Monitoring Vocal Fold Abduction Through Vocal Fold Contact Area, Journal of Speech and Hearing Research, 1988;31:338-351.
Ronald J. Baken, Electroglottography, Journal of Voice, 1992; 6;2:98-110.
R.J.Baken, Robert F.Orlikoff, Clinical Measurement of Speech and Voice, 2000:413-427.

动态电子腭位

李英浩

0 引言

动态电子腭位（简称 EPG）能够在连续语流中实时记录舌和硬腭接触过程中舌的空间位置和舌腭接触部位的变化。EPG 主要用于腭裂、耳聋、运动神经障碍等言语疾病的诊治和言语治疗。另外，EPG 也用于言语产生机制的研究。语言中的辅音发音动作的变化比元音的发音动作快的多，使用 EPG 就可以用较高的时间精度和空间精度来记录辅音发音动作的大部分过程，因此是研究舌协同发音（lingual coarticulation）必备的研究手段。

同其他的发音器官追踪仪器一样（如 EMA，X-ray 等），EPG 也有自身的优点和缺陷。EPG 是无侵入的仪器，使用相对简单，时间精度较高。缺陷是假腭只适用于特定的发音人；佩戴假腭后可能会影响发音人的发音质量；假腭向后无法达到软腭，因此无法记录软腭和舌的接触；最后电子假腭只能记录舌和硬腭接触过程的发音动作，而无法记录没有接触时候的动作（Byrd, 1995）。

本本首先介绍动 EPG 的发展历史和目前国内外使用的 EPG 的种类和性能，而后介绍假腭的制作过程和电极分布特点以及 WinEPG 的 62 点系统的构成和工作原理。接着介绍 EPG 数据缩减方法；最后简单介绍 62 点 EPG 在言语治疗和语音研究中的一些应用。

1 发展历史和主要类型

EPG 技术产生之前，研究者主要使用静态腭位照相技术（palatography）研究发音的部位。1930 年，德国人希凌（Schilling）提出了用电路进行腭位研究的构想。1964 年，美国华盛顿大学的基德（Kydd）和比尔特（Belt）设计了具有现代意义的电子假腭系统。他们使用直流电路，把银电极放置在被试者假牙托的表面上，这样当舌和硬腭接触的时候，就会产生微弱的电流。然而由于电流十分微弱，同时由于唾液的导电性会产生“假接触”，所以在实际应用中有很大的局限性。

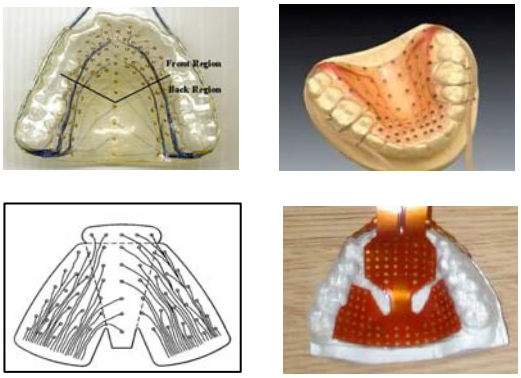


图 1 目前使用的 EPG 的类型（左上为 Kay 的 Palameter，左下为 Rio 系统，右上 2 为 EPG3 系统，右下为 LogoMetrix 系统）

上世纪 60 年代以后，交流电取代了直流电，柔性电路（flexible circurt）的设计使得使用更多电极来记录舌和硬腭的接触成为可能（Wrench, 2007）。电子腭位的研究和制作技术分别在日本，英国和美国得到发展。1968 年日本东京大学的柴田（Shibata）首先使用交流电路设计制作了电子腭位系统，但是电极数目很少。但到了 1978 年，柴田和他的合作者设计制作了 63 点的便携式电子腭位（Shibata 等，1978）。这就是后来的 Rio 电子腭位系统。也是在上世纪 60 年代，Hardcastle 和 Fletcher 改进了基德和比尔特的电路设计。Fletcher 教授在美国的亚拉巴马大学设计了 96 点假腭的电极排布方案，并委托 Kay 公司生产。这就是后来的 Kay 公司的电子假腭（Palatometer）。1974 年，Wilf Jones 在 Hardcastle 教授的指导下设计了 62 电极的雷丁电子腭位系统（Reading EPG system）。日本东京大学实验室采用了日本理光公司产生的 63 点电子腭位。2000 年，美国 LogoMetrix 公司生产了由 Fletcher 设计的 118 点的电子假腭。目前国外主要研究机构使用的电子腭位有四种，美国 Kay 公司的 Palatometer，英国雷丁大学的 EPG3 系统，日本的 Rio 系统以及美国的 LogoMetrix 系统（图 1）。英国 Articulate Instrument 公司根据 EPG3 系统设计制作了 WinEPG 系统。Kay 公司和日本理光公司已经停止生产电子腭位。

2 假腭的制作和电极的分布

制作假腭首先要获取发音人硬腭和上齿的模型。按照发音人硬腭的生理模型用丙烯酸基材制作假腭。在假腭表面安放电极。

电极间距、电极的直径以及电极的材质是设计假腭需要考虑的关键因素。由于唾液具有导电性，电极之间的距离应在一定的范围内。目前一致接受的距离在 1 毫米左右。EPG3 系统和 Palameter 系统电极在假腭上的分布类似，在齿龈区域电极分布比较密集，间距在 3 毫米左右，在龈腭区域电极分布比较稀疏，间距在 3 毫米以上。电极的大小和导电性相关，电极越大，导电性能越好。不同类型 EPG 系统的电极材料不同，Kay 和 Rio 系统用黄金或者白金材料，EPG3 和 LogoMetrix 系统采用银质电极。

导线从口腔引出的方式以及假腭、硬腭固定的方式以及假腭的厚度对语音的产生有较大的影响。EPG3、Palatometer 以及早期的 WinEPG 系统的电极导线都是分为两路，从两侧的臼齿由内向外引出，从嘴唇两角伸出口腔。导线的粗细直接影响口腔的完全咬合以及嘴唇的完全关闭。EPG3 和 WinEPG 使用牙箍把假腭固定在上齿上，而 Palatometer 覆盖了前齿和两侧的臼齿。这会对齿龈擦音和边音的气流产生一定的影响，特别是 Palatometer 系统。假腭的厚度对发音的影响最大，特别是阻塞音。

电极的分布有两种不同的分布理念（Wrench，2007），第一种是电极等距分布，如 Rio 系统和 LogoMetrix 系统。第二种是按照假腭的生理特征摆放电极，如 EPG3 系统和 Palatometer 系统。前者的优点在于能够比较精确地记录舌和硬腭接触的特点，缺点在于不便于人际之间的比较。后者的优点克服可前者的缺点，但是假腭生理生理特征并不是一致的，因此摆放的时候参考系是不同的。另外一个问题就是假腭电极的第一行和最后一行的位置。一般情况下，假腭只能覆盖硬腭区域，如果再向后延伸并且接触到软腭，则会有可能影响正常的发音。这是因为一旦软腭受到刺激，会使人产生呕吐的效果。因而，目前大多数的假腭的最后一行离软腭都有一定的距离。最前面一行的电极摆放在前齿背还是齿龈前，则必须根据不同语言中语音的发音特点。对于英语而言，第一排的电极就必须摆放在前齿背，因为齿间音（interdental）在连续语流中往往实现为[t̪]。汉语普通话中能够接触上齿背的语音只有齿龈音和舌尖前音。这些音是否接触上齿背只具有语音定位的作为，而没有音系学的价值。EPG3 的第一行电极摆放在前齿后 1 毫米左右的位置，而 Palatometer 的第一行电极放在前齿背。EPG3 的最后一行电极在硬腭和软腭分界线前 1-2 毫米左右（Hardcastle, 1991），而 Palatometer 最后一行电极在接近软腭的区域。因此 Palatometer 能够准确地记录齿间擦音/θ/和/ð/在语流中的表现，在记录软腭成阻方面也优于 EPG3。

下面主要介绍最新的 WinEPG 电极分布特点（Wrench, 2007）。WinEPG 使用 62 个电极，电极分布充分考虑了前面两种类型分布方式的特点。四个（或者三个）电极一组等距分布在承载电极的细条上。第一行电极比 EPG3 向前推进直到前齿跟附近，第七行电极处于软硬腭交界线。最后一行电极向后 7-12 毫米，与 Palatometer 类似。第四行电极处于齿龈桥附近。前四行电极和后四行电极分别等距排列（图 2）。

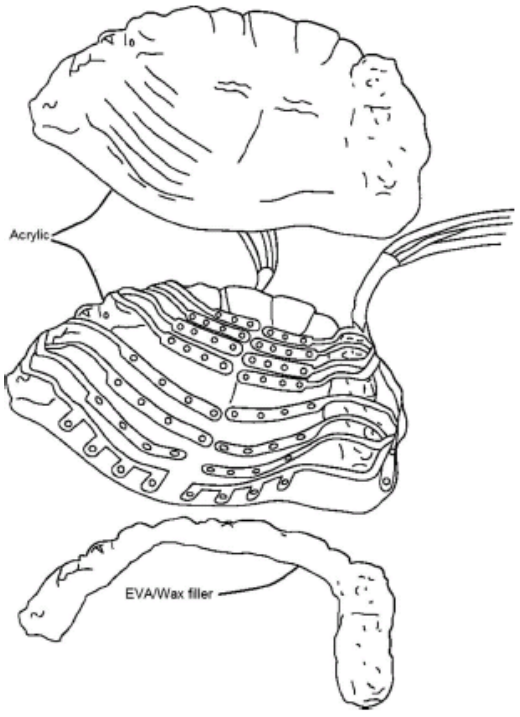


图 2 WinEPG 电极分布（引自 Wrench, 2007）

3 WinEPG 系统的构成和工作原理

WinEPG 硬件有串行通讯接口（Serial Port Interface），腭位扫描仪（EPG3 scanner），多路复用器（Multiplexer），电极手柄（Chrome Handgrip）和假腭（EPG palate）。

多路复用器连接三个部分，电极手柄，腭位扫描仪和假腭。电极手柄实际上就是一个电极，上面加有一定的电压，一旦假腭的某个电极产生接触，电流形成回路，信号就传送到多路复用器，再由后者对信号缓存放大后发送到腭位扫描仪进

行处理。多路复用器对电极扫描的时候，一次扫描 4 个电极，以此提高扫描速度。扫描 62 个电极需要 3.2 毫秒，所以 WinEPG 的理论采样频率可以达到 300Hz，然后一般采用 100Hz 或者 200Hz，因为这个采样频率足以记录舌腭接触的动态过程（Hardcastle, 1991）。

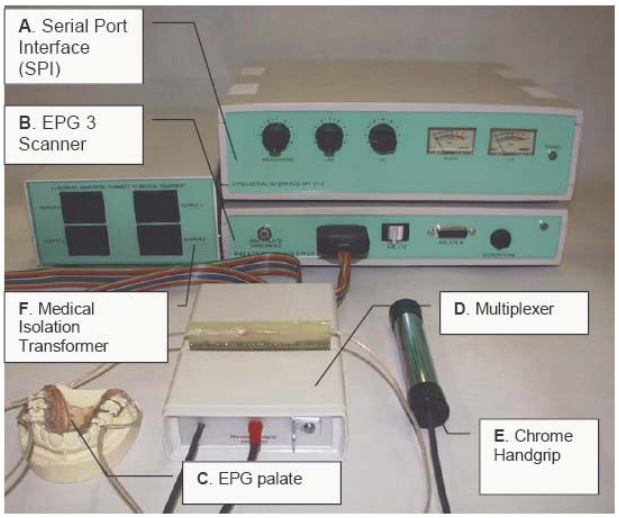


图 3 WinEPG 硬件（图片引自 Articulate Assistant User Guide）

腭位扫描仪有计时电路（timing circuit）和信号探测电路（signal detection circuit）。计时电路用来控制多路复用器对假腭电极进行扫描的时间，信号探测电路用于接触电极信号的获得、分析和传输。信号经过多路复用器放大后传送到腭位扫描仪，探测电路首先获得某时刻信号的峰值，然后与预先设定的阈值进行对比，而后把有效的信号进行数字编码传输到串行通讯接口。

串行通讯接口为腭位扫描仪提供电源并且根据电脑的指令控制腭位扫描仪的运行。此外，串行通讯接口还有两个输入口，分别是喉头仪信号输入口和一个线路电平（line-level）信号输入口，后者一般为语音信号。换句话说，WinEPG 可以同时采集三个信号，EPG 信号，语音信号和喉头仪信号。设备连接方法见图 4。最新型号的 WinEPG 还配备了 EMA 接口和超声波图像的接口，功能变得更加强大。

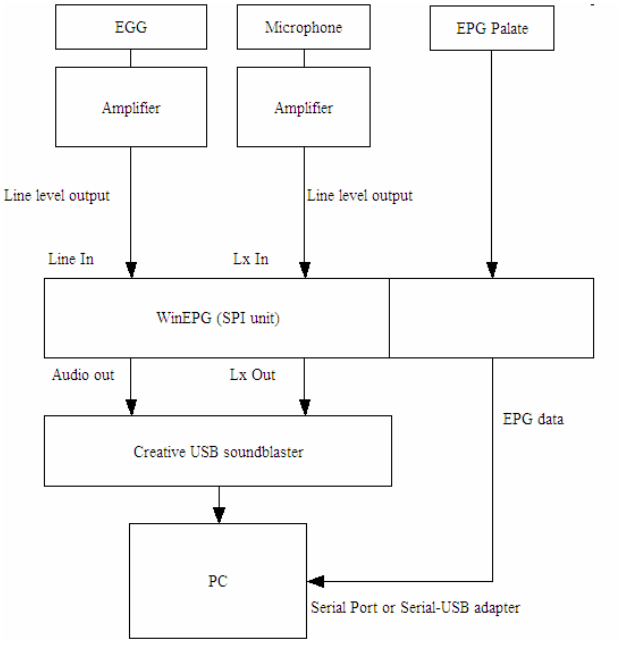


图 4 多通道信号采集连线图

4 EPG 数据缩减方法

EPG 数据缩减方法根据不同的研究目的而不同。下面主要介绍 62 点（EPG3 或者 WinEPG）的电子腭位的数据缩减方法。

由于不同发音人的发音习惯和策略不同，进行数据分析之前最好分析不同发音人的发音区域（Byrd,1995）。EPG3 或者 WinEPG 系统在设计的时候就考虑了不同发音人的硬腭生理特征，因此在许多研究针对不同的发音人采用相同的发音区域进行分析。而发音区域的划分却因不同的研究者以及研究目的而不同（图 5）。我们看到有的研究把假腭分为两个语音区域，有的把假腭细分为 5 个区域。实际上，对假腭的语音分区不仅与发音人硬腭的生理结构有关系，而且和特定语言语音的发音部位密切相关。进一步说，如果一种语言里面的辅音的部位比较多（发音部位是对立的），那么我们就可以把语音区域划分地细致一些。但是如果一种语言辅音发音较少，就可以划分地粗糙一些。然而对于言语工程的问题，我们就要对假腭的区域进行细致划分。这个时候，其工程意义大于其语言学意义。

Hardcastle等（1991，1999）把数据缩减技术分为三类：发音部位指数（Place of articulation measures），动态变化指数（Contact Profile Display）和接触指数（Contact indices）。第一类指数主要用于描写不同音类的舌腭接触区域并进行区域划分，与传统的语音学描写对应。一般采

用的方法如下[1]：

- 选取相同语音样本舌腭接触面积最大帧，而后进行叠加，计算每个电极的接触频率，最后按照研究的目的确定发音区域（Farnetani，1989）；
- 对于特定音类，如阻塞音（sibilant），可以采用槽宽，槽长以及阻塞的位置和 dimension；
- 选取关键帧界定语音样本的不同阶段。如对塞音而言，可以用第一个和最后一个完全阻塞帧（对 62 点假腭而言就是指任何一行全部接触）来界定塞音的持阻开始和结束；

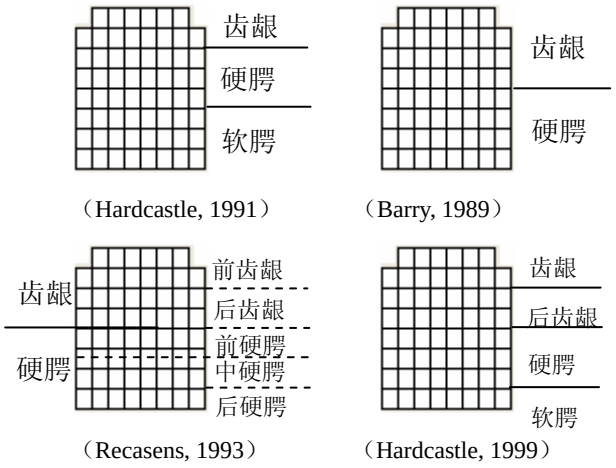


图 5 电子腭位电极和发音区域对应

第三类指数主要思路就是把一帧 62 点数据缩减成为一个数据。定义这类数据依赖于对腭位电极的分区。不同的研究目的就有不同的电极分区，因此此类指数为数众多。下面只介绍几个常用的指数。

最常用的指数包括面积重心指数（Center of Gravity，或者简称 CoG），趋前性，趋后性以及居中性指数。

面积重心指数（式（1），Hardcastle,1991）用来衡量接触电极在假腭前后维度上集中的程度。实际上，面积中心指数就是计算加权的平均数，电极的排数越靠前，权重就越大。

$$CoG=(R8\times0.5+R7\times1.5+R6\times2.5+R5\times3.5+R4\times4.5+R3\times5.5+R2\times6.5+R1\times7.5)/\sum_{i=1}^{8}Ri \quad (1)$$

Ri，i=1,2,3…8 — 每行接触电极的数目；

[1] 下面的分类和描写引自 Hardcastle 等（1999），236-242 页。

Hardcastle 等（1993）在研究/k/和/l/在不同元音环境中的发音部位的时候，修改了上述公式。首先，把和/l/和/k/的发音动作相关的区域分别确定在前四行和后四行中间四列。计算公式（2）与（1）类似。指数越高，表明发音部位受临近语音影响越大，位置越靠前。

$$后部\ CoG=(R8\times0.5+R7\times1.5+R6\times2.5+R5\times3.5)/\sum_{i=5}^{8}Ri \quad (2)$$

Ri，i=1,2,3…8 — 每行接触电极的数目；

Fontdevila 等（1994）认为 CoG 指数只能粗略地估计电极接触的分布。他提出使用接触趋前指数（CA），接触趋后指数（CP）和接触居中指数（CC）能够更加精确地表示接触的分布。其基本思想是前行（以 CA 为例）归一化后的接触点数的权重大于后面行所有可能接触点权重的总和。这些指数的优点是能够通过指数就能判断（以 CA 为例）接触电极能够能够达到的行数，在该行接触的点数以及其后面行接触的情况。比起 CoG，CA/CP/CC 更能够说明接触的方向性。下面给出 CA 和 CC 的公式，CP 的公式与 CA 类似。

$$CA=\log\left(\sum_{i=1}^{8}a_i\times R'_{9-i}+1\right)/\log\left(\sum_{i=1}^{8}a_i+1\right) \quad (3)$$

— a_i 为系数，$a_1=1$，$a_i=R_i\sum_{j=1}^{i-1}a_j+1$（$i>1$）

— Ri 为 i 行电极数目，R'_i 为 i 行接触电极比率；

$$CC=\log\left(\sum_{i=1}^{4}a_i\times C'_{5-i}+1\right)/\log\left(\sum_{i=1}^{4}a_i+1\right) \quad (4)$$

— a_i 为系数，$a_1=1$，$a_i=2C_i\sum_{j=1}^{i-1}a_j+1$（$i>1$）

CA，CP 和 CC 还可以应用在不同的发音区域里面。如 Fontdevila 把 62 点分为齿龈区和硬腭区，对硬腭区采用式（3）和（4）的指数，形成新的指数 CAp，CPp 和 CCp。采用后面的指数就能够很好地描写元音对前面辅音发音动作的影响。国内的学者如李俭等（2004），平悦铃（2005）就把 Fontdevila 的三个参数改造到 Kay96 点的电子腭位系统上。

动态变化指数就是计算假腭特定发音区域

接触电极的比率随时间的变化过程。指数的计算必须建立在假腭语音分区的基础上。动态变化指数能够展现假腭不同区域的电极的接触变化以及时间关系。图 6 汉语普通话“匝”和“渣”的齿龈、硬腭和软腭区域电极接触比率的动态变化指数。第三类指数随时间变化所形成的指数也可以是动态变化指数。

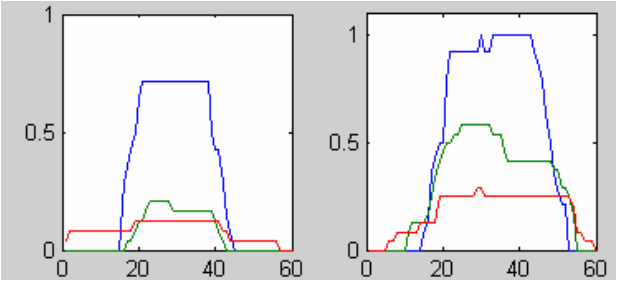


图 6 汉语普通话“匝”和“渣”齿龈、硬腭和软腭区域电极接触比率动态变化。蓝线为齿龈区域，绿线为硬腭区域，红线为软腭区域。

5 EPG 的应用

由于电子腭位能够反应舌和硬腭接触的过程，所以在言语疾病诊断和治疗中有很多用处。在言语疾病诊断中，医生往往需要判断病人的病症是语音知识层面的问题还是语音实现层面上的问题。如果是语音知识层面的问题，则说明病人无法正确感知自己母语的不同音位的区别。如果是语音实现层面的问题，则说明病人能够区分不同音位的差别，但是控制发音器官的神经或者肌肉发生了病变。言语治疗方面，EPG 能够通过各种视觉的手段对病人进行发音能力的恢复。

电子腭位在语言研究和言语工程也有重要的贡献。到目前位为止，EPG 是唯一能够以较高时间精度来对舌的运动方式进行记录的仪器，因而广泛用于研辅音的成音过程以及连续语流中舌头的运动的研究。最后，EPG 和 MRI 以及 X 光数据可以用于构建舌运动的三维模型，因而可以解释许多言语产生过程中的语音现象。

在中国，使用电子腭位对汉语进行研究以及开展言语疾病的诊断和治疗刚刚起步。中国社会科学院民族学和人类学研究所已经建立了一个汉语普通话的电子腭位数据库。北京大学也正在着手建立更加全面的汉语普通话的电子腭位数据库。

参考文献

[1] Byrd, D. et al. 1995. Using regions and indices in EPG data reduction. *Journal of Speech and Hearing Research* 38: 821-827

[2] Wrench. A. 2007. Advances in EPG palate design. Advances in Speech-Language Pathology. 9(1):3-12

[3] Shibata, et al.. 1978. A new portable type unit for electropalatography. Ann Bull RILP. 12: 5-10

[4] Hardcastle, W.J. et al. 1991. Visual display of tongue-palate contact: electropalatography in the assessment and remediation of speech disorders. International Journal of Language& Communication Disorders. 26: 41-74

[5] WinEPG Users Manual (Revision 1.9)

[6] Hardcastle, W.J. et al. 1999. Coarticulation: Theory, Data and Techniques. CUP.

[7] Farnetani, E. 1989. an arituclatory study of voicing in Italian by means of dynamic palatography. *Speech Research 89, Budapest*: 395-398

[8] Barrry, A. 1989. Measuring coarituclation and variability in tongue contact patterns. Clinical Linguistics & Phonetics. 3(1):39-47

[9] Recasens, D. et al. 1993. An electropalatographic study of alveolar and palatal consoants in Catalan and Italian. Language and Speech. 36(2/3): 213-234

[10] Fontdevila, J. et al. 1994. The contact index method of electropalatographic data reduction. Journal of Phonetics 22: 141-154.

[11] 李俭等. 2004. 汉语普通话动态腭位的数据缩减方法. 现代语言学与语音学研究, 天津社会科学出版社，主编：路继伦，王嘉龄，26—31

[12] 平悦玲. 2005. 上海方言语音动态腭位研究. 香港. 香港文汇出版社

肌电与语音研究

Electromyography and Phonetics Study

潘晓声

Pan Xiaosheng

摘要： 人体在运动和静止时都会产生生物电。使用肌电仪可以提取微弱的生物电用于分析人体行为。人本文介绍了肌电仪的工作原理。并对口轮匝肌的肌电信号进行采集并用于协同发音研究。

Abstract： Muscle cells of human being generate electrical potential all the time. We can use electromyograph to record the weak muscle signal for analysis the behavior of human being. This paper introduced the basic principle of electormyography. We record the electromyogram of orbicularis oris muscle for coarticulation research.

关键词： 肌电，肌电仪，肌电图

Key words： electromyography（EMG）, electromyography, electromyogram

0 引言

科学技术在过去的几十年里有了日新月异的飞速发展。而生命科学、计算机科学以及数字信号处理技术正是其是发展最快的几个学科之一。人们出于健康的要求，希望对人体本身有更加深入的了解。因此，螺旋 CT、MRI 等各种新技术、新设备都被应用到医疗行业。肌电仪也是其中的一种，人们用它来检测肌肉的状态。

1666 年，Francesco Redi的文章中首次提出了肌电的概念，但他所说的肌电是专指电鳗的肌肉产生的电。1792 年，Luigi Galvani[1]提出肌肉收缩时会产生电。1849 年，法国科学家Dubois-Reymond发现在肌肉收缩时可以用仪器把产生的肌电记录下来。1890 年，Marey第一次记录下了肌电图。1922 年, Gasser和Erlanger用示波器显示了采集到的肌电信号。在 20 世纪 60 年代，临床上开始使用表面肌电（surface EMG）。20 世纪 80 年代中期，电极的集成技术有了快速的发展，已经能批量生产足够小和足够轻的电极。近 15 年来，人们对皮肤表面肌电有了更加深入的研究，并对如何记录皮肤表面肌电的特性有了更好的了解。因此，近年来，在医疗上，人们更多的用表面电极来采集体表的肌电信号，对于深部肌肉更多的采用针式电极来采集肌电信号。

自从可以记录肌电信号开始，人们对肌电展开了研究。人们可以通过检测人体肌肉组织的肌电信号来来推断肌肉状态与肌肉相关的病情等。因此，对肌电的研究主要集中在医学方面。在语言学方面，主要是利用肌电信号进行协同发音的研究，也有人利用肌电信号来做手语识别、无声语言识别的研究。

早期的工作大量采用肌电信号进行协同发音研究。语言学家主要利用肌电信号来检测口轮匝肌的激活时间。Bell-Berti 和 Harris[2, 3]，Gay[4]，Gelfer，Bell-Berti，Harris[5]使用 EMG 为参数对英语进行了研究。他们的研究支持嘴唇的协同发音模型为固定时间模型。其中 Gelfer，Bell-Berti 和 Harris[5]用的是 EMG 和唇运动学这两种参数。另外，Daniloff 和 Moll[6]对英语；Lubker，McAllister 和 Carlson[7]，McAllister[8]，Lubker[9]对瑞典语；Benguerel 和 Cowan[10]对法语；Magno-Caldognetto 等[11]对意大利语进行了研究。这些研究用的是 EMG 和/或唇运动参数。他们认为在以上这些语言中，嘴唇的协同发音模型不是固定时间模型，而是向前看模型。

1989 年, Michael S. Morse[12]用肌电信号的能量，幅值，方差等做为特征来识别 10 个英文单词，识别率达到 60%。2002 年,Chan[13]发现在强噪声背景下，加上三路肌电信号可以把语音信号的识别率从 10%提高到 80%。2005 年,Nan Bu[14]用肌电信号识别五个日文元音和一个日文辅音/n/，达到 90%的识别率。国内也有人用肌电做唇读的研究。2005 年，戴立梅[15]等人对提口肌、颧肌部分、颈阔肌、压口板和二腹肌前腹采集发普通话 0 至 10 时的肌电信号。用短时傅立叶变换对肌电信号进行处理，之后利用 PCA 的方法降维得到特征用于识别。识别的错识率在 15%以下。2006 年，许佳佳[16]等人只测提口肌、颈阔肌和二腹肌前腹这三块肌肉的肌电信号。并用小波的方法来提取信号中的特征，将其用 BP 人工神经网络训练和识别。在识别汉语普通话 0 至 10 时，达到平均 95%的正确率。2006 年，王旭[17]等人通过检测颧肌和二腹肌前腹的肌电信号来识别汉语普通话 6 个元音。他用肌电信号的 AR 模型系

数、倒谱系数和美尔倒谱系数作为原始特征向量，使用遗传算法找出了原始特征的次优组合,并组成新的特征向量。将 GA 找出的次优特征向量向着 fisher 最佳可鉴别基投影得到最佳鉴别特征向量。最后用改进的 BP 神经网络作为分类器得到了较好的识别效果。

1 EMG 的工作原理

1.1 EMG 的工作方式

对于采集肌电信号，主要有两种方式：1、针式电极采集方式。2、表面电极采集方式。侵入式的针式电极，可以较准确的找到需要检测的运动单元的信号，因此能较精确的反应微小区域里单个运动单元的活动状态。但能否精准的找到某一根肌纤维的位置，对操作人员的要求非常高，就连专业的医生也不能保证一次成功。同时，因为侵入式的操作不可避免的要使被测的肌肉处于紧张状态,这对采集到的信号会有较大的不良影响。而非侵入式的表面电极表面面积比较大，贴在皮肤表面时，不会有疼痛感。但是得到的信号是整个覆盖区域内运动单元肌电信号的耦合，对于信号的后续处理和分析难度较大。

1.2 EMG 的工作原理

我们知道肌电是在肌肉收缩时产生的。动作单元（Motor Unit）是肌肉最小的收缩单位。它是由一个α运动神经元、运动终板、轴突及神经未梢支配的肌肉纤维组成。如图 1[18]所示：

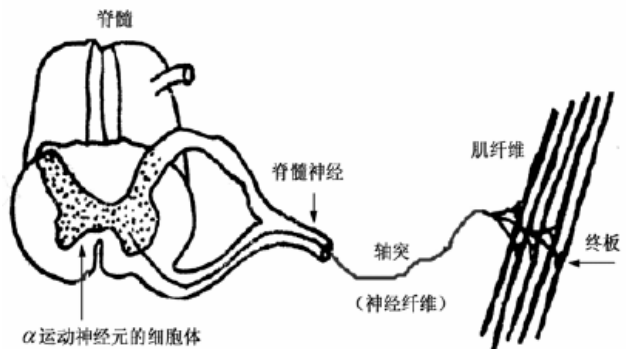


图 1 动作单元的解剖结构

肌细胞有四种肌电位，分别是静息电位、动作电位、终板电位和损伤电位。肌电仪所测得到肌电信号就是肌细胞的肌电位。

安静的肌细胞受刺激时，细胞外液中的大量正钠离子渗放到细胞体中。膜内电位从静息状态的-90mv迅速上升到 30mv。此过程被称为去极（在临界电位-65--70mv时，就产生动作电位）。到顶峰 30mv后，马上开始复极。此时，对正钾离子和负氯离子的通透性增大。主要是正钾离子外流和负氯离子内流。这样，肌细胞的电位重新下降到 0。动作电位的有以下特征：1、有清晰的阈值，一但产生，幅度和时程就不由刺激的幅度和时程决定。2、一个动作电位结束后，另一个动作电位才能产生。两个动作电位之间，必定有一个安静期。当某个区域产生动作电位后，会与相邻区域产生电位差，这样电流我们知道，当动作电位沿着肌肉纤维传导时，可以被针式电极检测到。这种应在单根肌纤维中动作电位传播的波形电位被称为MFAP(Muscle Fiber Action Potential)。动作电位示意见图 2[18]。

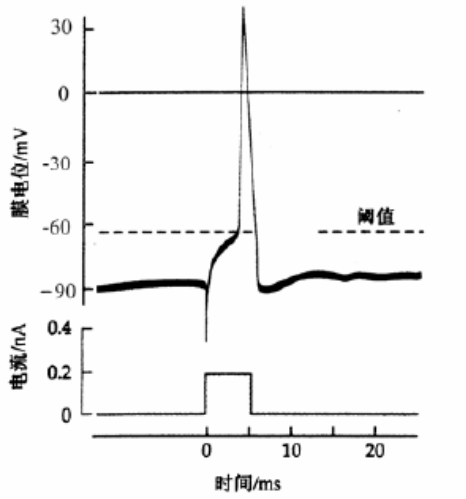


图 2 动作电位

一般情况下,总是一束肌纤维同时受到刺激，不会出现单根受刺激的情况，因此无法测到 MFAP。我们把由同一运动单元控制的一束肌纤维的 MFAP 之和称为 MUAP(Motor Unit Action Potential)。为维持一块肌肉的动作，α运动神经元不断对肌纤维束发来刺激。每次刺激产生一个 MUAP，由此可以生成 MUAPT(Motor Unit Action Potential Train)。同时由于人体中有体液，因此 MUAPT 会受到附近其它 MUAPT 产生的电场的影响。所以，我们采集到的表面肌电（sEMG）信号是多个 MUAPT 信号加上噪声的组合。

2 软硬件设备及其使用

肌电信号是一种非常微弱的生物电信号。因此想得到肌电信号,必须把采到的信号前置放大。我们实验室所使用的是澳大利亚的 GUGER TECHNOLOGES 公司生产的生物电放大器 g.BSamp。

它的前侧面板如图 3所示。

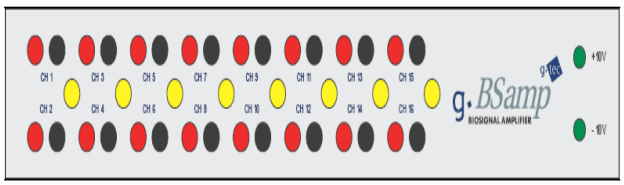


图 3 g.BSamp 前侧面板接口示意图

其中共有 16 组输入端口。这 16 组端口又按垂直方向分为 8 组模块。水平方向相邻的两个端口（一红一黑）为一组。每两组端口共用一个黄色的接地端口。每个端口都可以连接一个电极。其中正电极接到红色端口，负电极接到黑色端口。每组端口都被预先定义为检测某些特定信号。具体如**错误！未找到引用源。**所示：

1	2	3	4	5	6	7	8
EEG				EOG			
9	10	11	12	13	14	15	16
EMG				ECG			

表 1 端口使用分配

其中，EEG(electroencephalo-graph)表示脑电图，EOG(electro-oculogram)表示眼电图，EMG(Electromyographic)表示的是肌电图，ECG（Electrocardiograph）表示的是心电图。

它的背侧面板如图 4所示。

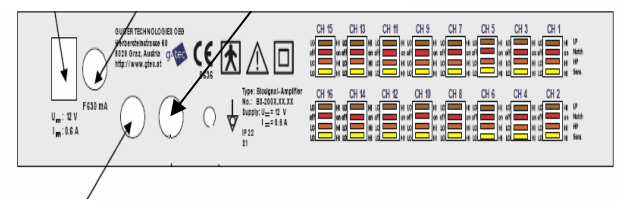


图 4 g.BSamp 背侧面板接口示意图

要注意的是，产家声明装有心脏起搏器或其它电子刺激产品的人不能使用此设备，因为此设备会给人体带来危险。

从电极采到的信号经过生物电放大器的前置放大，再输入到肌电仪中进行模拟信号和数字信号的转换及后续处理。常见的肌电仪的产家及型号有很多。主要的有日本光电工业株式会社生产的MEB-2200 系列，MEB-9204，MEB-9400 系列，芬兰的 Mega Electronics 有限公司生产的ME6000。

我们实验室购买的是新西兰ADInstuments公司生产的Powerlab ML880 16/30。它的前部面板如图 5所示。

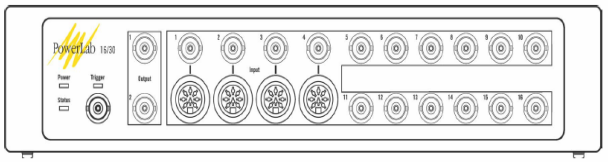


图 5 ML 880 16/30 前侧面板接口示意图

前侧面板上有如下端口：

1、BNC 输入端口：用于记录外部输入的模拟信号。每个 BNC 端口之间采集的信号都相互独立。我们可以用 Chart 或 Scope 对每一个 BNC 端口分别进行设置。输入信号电压的范围是-10V~+10V。如果输入的电压超过±15V 就会损坏电路。

2、BNC 输出端口：从 powerlab 输出电压信号。

3、BNC 触发端口：从外部引入一个同步信号。

4、DIN 输入端口：也用于记录外部输入的信号。要注意的是和 BNC 端口不能同时用于记录输入信号。但两种端口可以同时用于不同的通道，比如一路用于输入，另一路用于输出。

其背部面板如图 6所示:

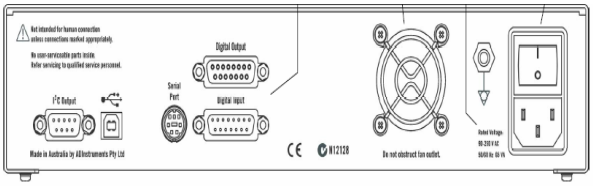


图 6 ML 880 16/30 背部面板接口示意图

背部面板上有如下接口：

1、 电源接口。
2、 保险更换接口。
3、 数字信号输入端口与数字信号输出端口：用于连接 TTL 设备。
4、 串行口：用于将来扩展，不可用于连接计算机。
5、 USB 接口：将 powerlab 连接至计算。
6、 I^2C端口：用于连接前端（front-end）设备。由于我们没有购买前端设备，所以这里不做深入介绍。

为配合PowerLab的使用，ADInstruments公司提供了软件Chart和Scope。本文只介绍实验中所用到的Chart，如图 7所示。

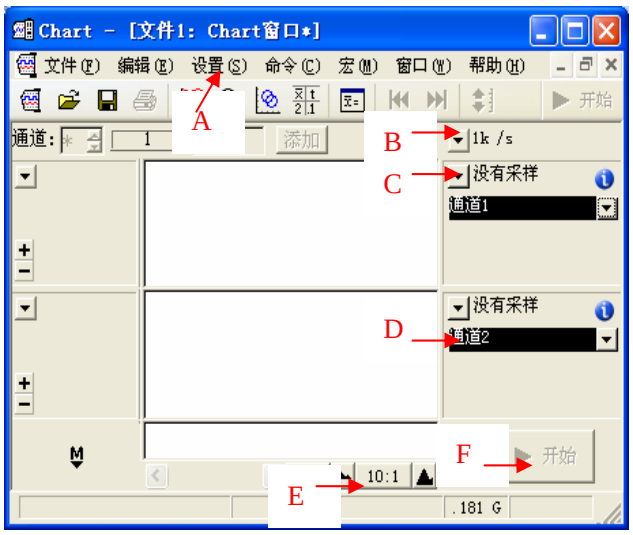


图 7 Chart 软件界面

Chart可以对每一个输入端口的信号分别进行设置，并实时显示。还可以对数据进行保存，统计分析。

根据具体使用情况，我们将要使用的功能分别以英文字母A—E标出。

A：设置菜单，其中可以设置每次录制几个信道。

B：分别对每个信道设置分辨率。

C：所录制的信号的电压范围。

D：对录制的信号进行处理，比如滤波。

E：对录制的信号水平方向的显示比例进行设置。

F：开始录制。

3 实验设计及结果

人在发音时，脸部有很多肌肉都会协同工作。口轮匝肌是最重要的参于发音的肌肉之一。以圆唇元音/u/为例，在发音时，口轮匝肌会有一个从松驰到紧张的过程。我们可以从肌电图中发现，肌肉在松驰状态和紧张状态时，肌电信号处于两种不同的模式。因此我们可以从肌电图中检测出口轮匝肌开始紧张的时刻，即口轮匝肌的激活时刻。早期的协同发音研究大多是研究声学信号和口轮匝肌的激活时间之间的关系，进而找出嘴唇协同发音的模式。我们的实验将检测人在发音时，口轮匝肌右侧嘴角处的肌电激活状态。

人体肌肉并不是一个点，因此表面电极贴在皮肤的哪个位置将极大的影响采到的肌电信号的质量。总体来说，我们可以把电极分为两类：参考电极和测量电极。参考电极应放在肌电信号弱的肌腱处，而测量电极应放在相关肌肉的肌腹处。另外，电极之间的距离不宜过近以防止电极之间的电流相互干扰，也不宜过远，使两电极之间所测得的信号没有相关性。一般说来，两电极之间的距离设到 2cm左右。另外，要注意的是，皮肤表面的油脂等污垢都会产生噪音，而肌电信号本身就非常微弱，极易受噪音干扰。所以在贴表面电极之前，要用酒精擦拭待测部位的皮肤去除污垢以减少干扰。同时要在表面电极的凹陷处填入导电膏以加强电极的导电性能。对于本实验中参考电极和测量电极的位置，见图 8[19]。

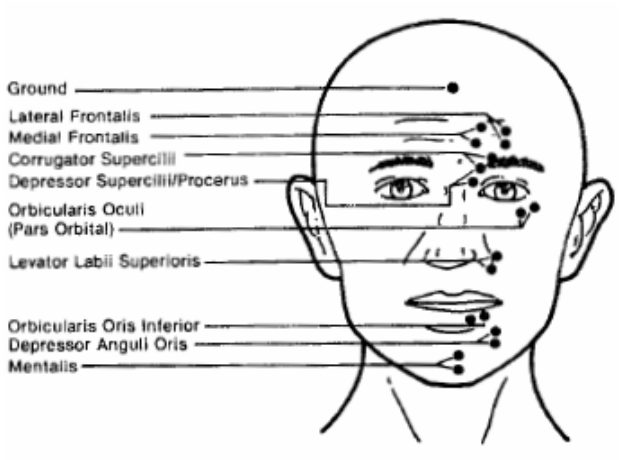


图 8 人脸头部测量不同肌肉电极位置

对于 Chart 软件，推荐设置为：采样率：1000k/s。带通滤波：低频的截止频率为 20HZ，高频的截止频率为 500HZ。量程：推荐为 100mv。图 9为发音人在发圆唇元音/u/时右侧嘴角口轮匝肌的肌电信号。我们可以从中找出口轮匝肌开始激活的时刻。

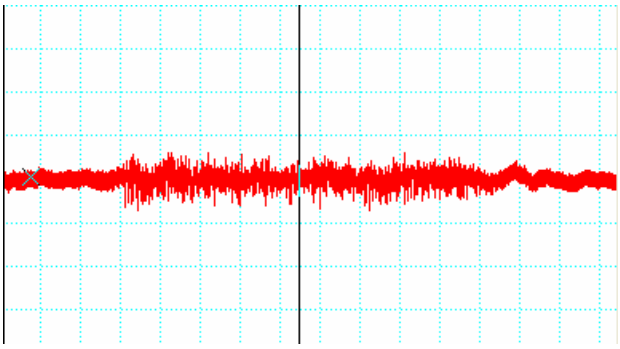


图 9 发圆唇元音/u/时右侧嘴角口轮匝肌的肌电信号

由于尚未设计完成所要采集的词表，因此我们将在将来的工作中结合语音信号进行汉语普通话嘴唇协同发音的研究。

参考文献

[1] L. Galvani and G. Aldini, *Aloysii Galvani de viribus electricitatis in motu musculari commentarius*: Apud Societatem

Typographicam, 1792.

[2] F. Bell-Berti and K. Harris, "Anticipatory coarticulation: Some implications from a study of lip rounding," *The Journal of the Acoustical Society of America*, vol. 65, p. 1268, 1979.

[3] F. Bell-Berti and K. Harris, "Temporal patterns of coarticulation: Lip rounding," *The Journal of the Acoustical Society of America*, vol. 71, p. 449, 1982.

[4] T. Gay, "Coarticulation in some consonant-vowel and consonant clustervowel syllables," *Frontiers of Speech Communication Research, Academic Press, London*, pp. 69-76, 1979.

[5] C. Gelfer, F. Bell‐Berti, and K. Harris, "Determining the extent of coarticulation: Effects of experimental design," *The Journal of the Acoustical Society of America*, vol. 86, p. 2443, 1989.

[6] R. Daniloff and K. Moll, "Coarticulation of lip rounding," *Journal of Speech, Language and Hearing Research*, vol. 11, p. 707, 1968.

[7] J. Lubker, R. McAllister, and J. Carlson, "Labial Co-Articulation in Swedish: A Preliminary Report," 1974, pp. 55-63.

[8] R. McAllister, S. universitet, and I. f. r. lingvistik, *Temporal asymmetry in labial co-articulation*: Institute of Linguistics, University of Stockholm, 1978.

[9] J. Lubker, "Temporal aspects of speech production: anticipatory labial coarticulation," *Phonetica*, vol. 38, p. 51, 1981.

[10] A. Benguerel and H. Cowan, "Coarticulation of upper lip protrusion in French," *Phonetica*, vol. 30, pp. 41-55, 1974.

[11] E. Caldognetto, K. Vagges, G. Ferrigno, and M. Busa, "Lip rounding coarticulation in Italian," 1992.

[12] M. Morse, S. Day, B. Trull, and H. Morse, "Use of myoelectric signals to recognize speech," 1989, pp. 1793-1794.

[13] A. Chan, K. Englehart, B. Hudgins, and D. Lovely, "Hidden Markov model classification of myoelectric signals in speech," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, pp. 143-146, 2002.

[14] N. Bu, T. Tsuji, J. Arita, and M. Ohga, "Phoneme Classification for Speech Synthesiser using Differential EMG Signals between Muscles," 2005, pp. 5962-5966.

[15] 戴立梅, 姚晓东, 王蓓, 邹俊忠, and 王行愚, "EMG 在语音信号识别中的应用," *电子技术应用*, vol. 31, pp. 5-7, 2005.

[16] 许佳佳 and 姚晓东, "基于 EMG 信号的无声语音识别应用及实现," *计算机与数字工程*, vol. 34, pp. 133-136, 2006.

[17] 王旭, 贾雪琴, 李景宏, and 杨丹, "基于优化肌电特征的无声语音信号识别," *东北大学学报: 自然科学版*, vol. 27, pp. 1095-1097, 2006.

[18] 颜志国, "基于粗糙集和支持向量机的表面肌电特征约简和分类研究," in *生命科学技术学院*. vol. 博士 上海: 上海交通大学, 2008, p. 138.

[19] A. Fridlund and J. Cacioppo, "Guidelines for human electromyographic research," *Psychophysiology*, vol. 23, pp. 567-589, 1986.

X-ray 语音研究报告

Report of Phonetic Study on X-ray

杨锋

Yang Feng

摘要：本文介绍了 X-ray、CR、DR 和 CT 成像技术的原理和发展，主要介绍了 X-ray 在语音研究中的应用。

Abstract：This paper introduces the principles and development of modern medical image technic such as X-ray, computed radiograph(CR), digital radiograph(DR), and computed tomography(CT). And a preliminary introduction to phonetic study on X-ray and medical image is given.

关键词：X-ray　医学成像　语音研究

Key words：X-ray, Medical Image, Phonetic Study

0 引言

X-ray 成像技术是医学成像的主流之一，近些年来成像技术日趋成熟，能够为医生提供更高分辨率的医学影像。同时这些影像也可用来研究发音器官，对研究舌位、小舌、唇等发音器官具有及其重要的意义。

1　X-ray 成像技术

1.1　X-ray 的发现

1895 年 11 月 8 日，德国物理学家威尔姆·康拉德·伦琴,在进行阴极射线的实验时发现了 X-ray，翻译为 X 射线，俗称 X 光，或伦琴射线。因为当时对于这种射线的本质和属性还了解得很少，所以称它为 X 射线，表示未知的意思。伦琴射线是人类发现的第一种所谓“穿透性射线”，它能穿透普通光线所不能穿透的某些材料。在初次发现时，伦琴就用这种射线拍摄了他夫人手的照片，显示出手骨的结构图像。1901 年伦琴获得了第一个诺贝尔物理学奖。此后有 13 位科学家在物理学、化学、医学、生物学、晶体学等领域利用 X 射线技术做出了开创性的工作和成果而荣获了诺贝尔奖。

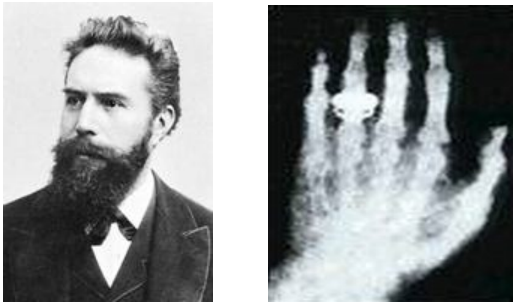


图 1 德国物理学家伦琴和第一张 X-ray 图像

X 射线是一种电磁辐射,是一种波长比紫外线更短的不可见射线,与红外线、紫外线及可见光一样,具有波动性的一切特点,即反射、折射、衍射、干涉等现象。

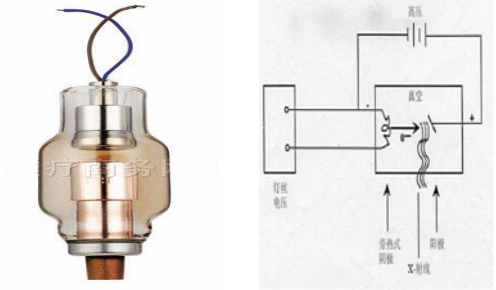


图 2　X 射线管及其电路图

今天常用的 X 射线通常由射线管产生。射线管是在真空玻璃管内安装灯丝,阴极和阳极靶,通电加热灯丝，使阴极发射出大量电子，在两极间加上几十千伏的高压，让这些电子快速飞向阳极高能电子与阳极靶碰撞的结果，射线就从靶上发射出来。

1.2　X 射线成像原理

用 X 射线照射人体,由于人体内不同的组织或器官拥有不同的密度与厚度，因此对 X 射线产生不同程度的衰减作用，从而形成不同组织或器官的灰阶影像对比分布图，进而以病灶部位的相对位置、形状和大小等改变来判断病情。

1.3　X 射线平面成像技术

医学成像技术多种多样，按成像的维数可分为二维和三维成像；按成像的实时性

可分为电影和图片成像；按成像机理又可分为 X 射线成像（X 射线平面成像、X 射线 CT）和非 X 射线成像(超声、磁共振和核医学成像）等。（苟量，2002）

常规 X 射线成像技术是将人置于 X 射线管与影像接收器之间，作为接受器的平板荧光屏,将接受的 X 射线转换为可见光。由于屏的亮度较低,只能在暗室中观察。为了解决荧光屏亮度低的问题,研究出了影像增强管。在影像增强管中,碘化铯等材料制成的荧光屏和光电阴极紧密相接。入射的 X 射线与荧光屏作用后产生可见光,可见光又使光电阴极产生电子,这些电子经过一个透镜系统加速并聚焦在输出荧光屏上。使带有影像增强管的 X 射线图像质量明显改善,可以在明室内观察,达到临床应用的要求。这就是我们常说的 X 光透视检查（如胸透）。

图 3　X 射线平面胶片成像设备

由于病人的检查结果需要备案,以便对病人的发病史和治疗过程进行跟踪,而使用涂上感光乳剂的胶片与荧光增强屏组成的屏—胶片系统,可以得到高分辨的 X 射线图像,胶片所记录的 X 射线图像可以长期保留（如对骨折部位的 X 光拍照）。

1.4　计算机射线成像技术（CR)

随着信息技术的发展，传统方法与数字技术相结合，已派生出一系列数字成像技术。1981 年日本富士公司推出数字化 X 射线成像技术(Computed Radiograph，即 CR)。CR 技术采用影像板代替传统的胶片来记录 X 射线，再用激光激励影像板,通过专用的读出设备读出影像板存储的数字信号,之后再用计算机进行处理和成像。（苟量，2002）

1.5　直接数字化 X 射线成像技术(DR)

1997 年，又出现了直接数字化 X 射线成像技术(Direct Radiography，即 DR),DR 技术的探测器可以迅速将探测到的 X 射线信号直接转化为数字信号输出，而不需要 CR 中的激光扫描和专用的读出设备，所以 DR 的实时性高于 CR，在动态成像领域颇具发展潜力。（苟量，2002）

目前平面成像呈现出数字技术与传统技术平分天下的情况,但是影像技术的数字化大趋势却不容置疑。数字图像在采集、显示、存储和传输方面的优点不言而喻，更为重要的是可以进行各图像后处理，如窗口调节、图像融合等等。此外,CR 和 DR 的探测器对 X 射线的量子检测率高达 60% 以上(传统仅为 20%～ 30%)，密度分辨率高达 210～ 14 灰阶(传统仅为 26 灰阶),大大降低了对病人的辐射剂量,但 CR 和 DR 设备昂贵。（苟量，2002）

图 4 采用 DR 技术的 X-ray 设备

2. 计算机断层扫描技术(Computed tomography，简称 CT)

X 射线平面成像把具有三维结构的人体拍摄成二维的平面图像，所以各种组织结构的影像必定相互重叠。若相邻的器官或组织之间如对 X 射线的吸收差别小，则不能形成对比。（苟量，2002）

1972 年英国工程师 Hounsfield 发明了 X 射线 CT，以 X 射线沿患者身体某一截面（某一选定的体层）的不同方向上进行扫描,探测器测定每条线上透过的 X 射线量，形成一个投影。这些直线投影的集合形成一个“投影截面”，每完成一次直线扫描，探测器旋转一定角度(旋转角的大小根据所需图像分辨率来定)，再扫描一次，取得另一个投影截面，如此反复，直到整个截面扫描结束。然后，用计算机处理这个截面数字化后的 X 射线信息，得出该截面组织各个单位容积的吸收系数，重建图像，是目前医学影像技术中体层摄影最为完善，应用最

多的技术。(王骏,2002)

图 5　X 射线计算机断层摄影设备(CT)

在 CT 扫描下，实时观察人体横断面的解剖结构的变化对介入治疗有很大帮助，这对现有数据采集和重建速度、减少患者的扫描剂量提出了新的要求。第一代 CT 扫描一帧图像要 100 s 以上，目前速度达 6～ 8 帧/s，图像重建速度达到亚秒的数量级。电子束 CT 每帧 20～ 50 m s，因而检查运动器官(如心脏大血管等) 能得到清晰的图像，实现了电影 CT。(苟量，2002)

3. 螺旋 CT

螺旋 CT 突破了传统 CT 的设计,采用滑环技术,将电源电缆和一些信号线与固定机架内不同金属环相连,运动的 X 射线管和探测器滑动电刷与金属环导联。球管和探测器不受电缆长度限制,沿人体长轴连续匀速旋转,扫描床同步匀速递进(传统 CT 扫描床在扫描时静止不动),扫描轨迹呈螺旋状前进,可快速、不间断地完成容积扫描。

多层螺旋 CT (multi-slice CT)是指扫描一圈所得到的图像数,例如，4 层 CT 就是扫描一圈出 4 层图像。多排螺旋 CT (multi-detector CT 或 multi-row CT)是指组成 CT 的探测器排数。二者的共同特点是 X 射线管与探测器阵列沿螺旋线轨迹围绕人体旋转,每旋转一周能同时获得多幅断面图像,大大提高了扫描速度。按照临床上使用较多的称法,统称之为多层螺旋 CT。多层螺旋CT 能高速地完成较大范围的容积扫描,图像质量好,成像速度快,具有很高的纵向分辨率和很好的时间分辨率。与单层螺旋 CT 相比,在不增加 X 射线剂量的情况下,每 15s 左右就能扫描一个部位;5s 内可完成层厚为 3mm 的整个胸部扫描；一次屏气 20s,可以完成整个体部扫描;病人接受的射线剂量明显减少。(苟量，2002)

4. X 光的危害性

X-ray 是一种电磁波，属于不可见光，但其辐射吸收率极高，它在穿透皮肤、肌肉等软组织的同时，被组织接受的辐射能量是很强的，足以将肌肉组织中的电子撞击出围绕原子核运动的轨道，产生不稳定的分子，进而生成对人体有危害的自由基。较强的辐射能直接引起染色体破裂或基因突变,甚至导致癌症的发生。(卓然，2008)

5. X-ray 在语音研究中的应用

5.1 研究目的

运用X-ray成像设备对发音器官进行拍摄可得到动态图像，对研究舌位、小舌等发音器官具有及其重要的意义。从声道的X-ray 录像中能够提取出声道形状，通过计算得到声道传递特性，运用声道传递特性进行语音合成实验。

5.2 研究步骤

第一步：建立 X-ray 录像数据库。由于X-ray 对人体具有一定的危害，同时设备昂贵，因此样本数据获得较为不易。目前较大的X-ray录像数据库主要有鲍怀翘先生建立的普通话单音节和双音节的 X-ray 录像库，还有 ATR 和 Strasbourg 大学所建的录像库。

第二步:对 X-ray 录像中的声道动态图像进行分析和标记，提取出声道形状的数据。标记大多使用手动和自动相结合的方法。

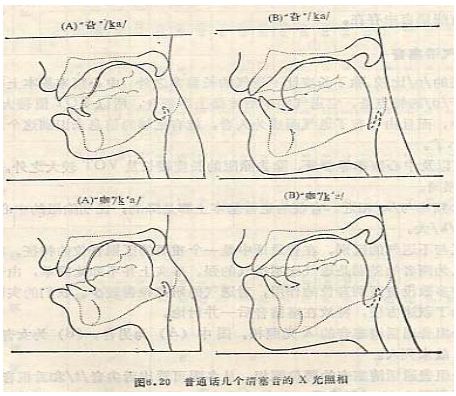


图 6 根据 X-ray 图像画出的发音器官图(鲍怀翘,1983)

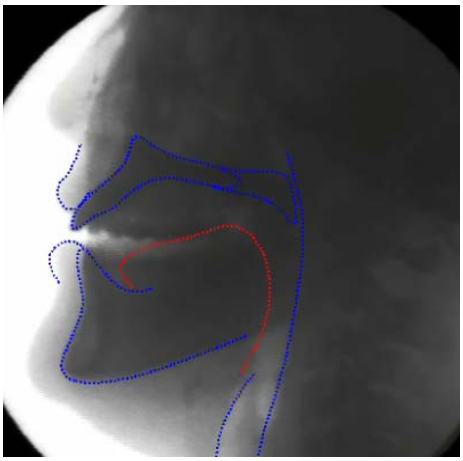


图 7 从-ray 录像中提取出声道形状(汪高武,2008)

第三步：通过计算得到声道传递特性，运用声道传递特性进行语音合成实验，归纳发音特点和规律。

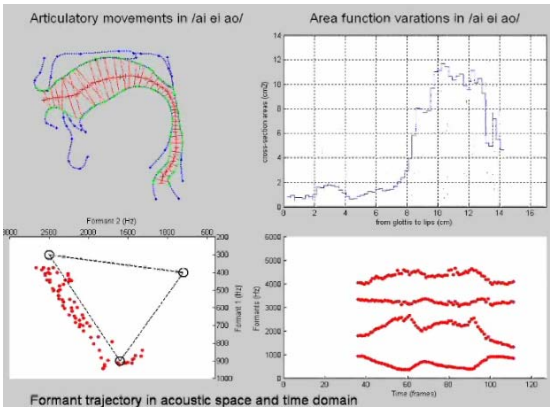


图 8 根据声道形状计算出的声道传递特性(汪高武,2008)

6. 总结

X-ray 成像技术是医学影像中使用最广泛，也是目前发展得最成熟的成像技术之一。运用 X-ray 技术进行语音研究能够得到更精确的动态图像，对研究舌位、小舌等发音器官运动变化的过程具有重要作用。而且可对图像进行参数提取，得到声道参数，通过计算声道面积得到声道传递特性，运用声道传递特性进行语音合成实验。这些对于语音学的深入研究具有及其重要的意义。

参考文献

[1] Fant,G.1960.Acoustic Theory of Speech Production: With Calculation Based on X-Ray Studies of Russian Articulation. Mouton, The Hague.

[2] Wood,S. 1979. A radiographic examination of constriction location for vowels. J. Phonetics 7:25-43.

[3] Yves Laprie and Marie-Odile Berger. Extraction of Tongue Contours in X-ray Images with Minimal User Interaction. Vandceuvre-lts-Nancy FRANCE.

[4] 鲍怀翘.1983.声道面积函数和共振峰频率的初步报告,语音研究室语音研究报告 1983-1984,中国社会科学院语言研究所

[5] 汪高武,孔江平,鲍怀翘. 从声道形状推导普通话元音共振峰,中国语音学学报,2008,(1)

[6] 汪高武,卢绪刚,党建武,孔江平,鲍怀翘.基于 X 光和 MRI 的汉语普通话声道研究,第八届中国语音学学术会议暨庆贺吴宗济先生百岁华诞语音科学前沿问题国际研讨会. 2008

[7] 苟量,王绪本,曹辉.X 射线成像技术的发展现状和趋势,成都理工学院学报,2002,(2):27-31

[8] 王骏.我国医学影像技术近十年发展概述，影像技术,2002,(3):6-11

[9] 卓然.X光检查应慎重,科学养生,2008,(1):31-32

（杨锋 北京大学中文系语音乐律实验室 100871 yangzihuai@163.com）

生理语音学仪器–气流气压计

Physiological Phonetic Instrument– Airflow & Air pressure Meter

吴韩娜

Oh Han-na

摘要： 随着现代科技的发展，出现了许多新的生理语音学仪器。近十年来，国外的语音生理实验使用的设备也纷纷进入中国，给予中国语音学的发展巨大支持。本文对生理语音学仪器之一的气流气压计的特征和使用方法作了简单介绍，并涉及国内该仪器的应用状况以及研究趋势。

Abstract： With the development of modern science and technology, there have been a lot of new physiologyical phonetic instruments. Instuments for physiological experime-nts have been come into China over the past decade, giving greater support on development of phonetics in China. This paper presents the characters of airflow & air pressure meter which is one of the physiological phonetics instruments , as well as deal with the application status of this instrument and current research in China.

关键词： 生理语音学 空气动力学 气流气压

Key words： physiological phonetics, aerodynamic study, airflow, air pressure

0 引言

“语音生理的研究一直是语言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（孔江平，2007）。” 随着科学的发展，能够采集生理信号的仪器日益增多，而给予中国语音学下一步发展的新空间。近十年来，进入中国语音生理实验使用的电子设备主要有动态电子腭位仪(EPG) 、电磁发音仪(EMA) 和口鼻气流气压计等。其中，本文将介绍的是“气流气压计”。

气息是语音产生的原动力，我们说话的时候都有气流出来也需要一定的压力，因此气流气压计可以反映出很多信息。其应用范围为各种言语和语音病变的各种研究。本文将依据北大语音乐律实验室设备的现状，主要介绍气流气压计和有关的研究成果。

1 仪器的简介

据调查，以气流气压计做的语音生理研究是从 1950 年代开始的。1959 年 Draper 首次做了对言语活动和呼吸中肺容积变化的研究（Shadle, 1997）。后来，由 Rothenberg, M.(Rothenberg, 1971)设计的Rothenburg mask[1] 出现，对用于气流气压计的生理语音研究的发展意义巨大。

目前气流气压计系统具有代表性的有：SCICON R&D 有限公司的PCquirer[2]，Glottal Enterprise的Aeroview Pro Glottal Flow Resistance System[3] 和KAY公司的Phonatory Aerodynamic System (PAS)和Nasometer[4]等。

北大语音乐律实验室有两个气流气压计系统：（1）SCICON R&D 有限公司的 PCquirer；（2）Glottal Enterprise 的 Aeroview Pro Glottal Flow Re-sistance System。

PCquirer 在 90 年代初由美国 UCLA 的 Henry Tehrani 开发的。在 1991 年 P.Ladefoged 首次使用 PCquirer 的第一代到东非做录音工作（P.Ladefog-ed, 2003）。该仪器的主要组件有：主系统（main system），口罩(oral mask)，鼻罩(nasal mask)，变换器(transducer interface)，迷你线(mini din cable)，传输线(USB cable)，电源线(power cord)，密狗(USB dongle)和该仪器的专用软件。参看下图。

图 1 PCquirer 的基本组件

[1] 参看 www.rothenberg.org
[2] 参看 www.sciconrd.com
[3] 参看 www.glottal.com
[4] 参看 www.kayelemetrics.com

PCquirer 的主要测量参数有：口腔的气流量和气压级、鼻腔的气流量和气压级和基频等。它除了用于腭裂、运动性言语障碍、听力障碍、腭修复、功能性的鼻音问题等嗓音病变和语音矫正外；还可以提取各种不同的参数用于言语产生的生理研究。如：引起声带振动必要的气流气压特性；气流相关的嗓音发声类型（汉语方言、少数民族语言）研究；鼻音和鼻化元音的实验语音学分析；呼吸和韵律特征研究；歌唱教学等（李永宏、孔江平、于洪志，2008）。

Glottal Enterprise 的气流气压计系统可以溯源到上文的 Rothenberg mask（Rothenberg, M., 1971）。多年来，对该仪器的设计和功能进行了几次的改进，现在该仪器的整个系统由主系统（Main System），口鼻罩（Airflow Masks），气流气压变换器（Airflow & Air pressure Transducer），口压适配器（Oral Adapter）和相关软件（Waveview 和 Aeroview）等组成，如图 2 所示。

图 2 Glottal Enterprise 的气流气压计系统

2 仪器的特征

PCquirer 的主要特征有以下几个方面：

（1）多通道的信号采集功能，它同时可以采集五个通道的信号，包括声音信号、口腔气流信号、口腔气压信号、鼻腔气流信号和鼻腔气压信号；

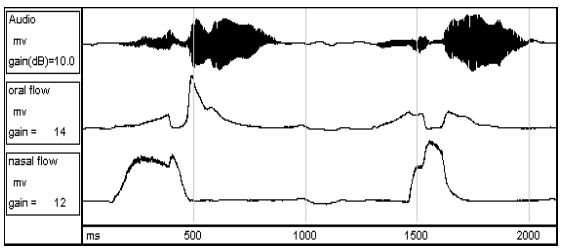


图 3 声音、口流和鼻流信号

对采集的不同信号进行信号处理，从而提取出不同的声学参数进行研究：除了声音、气流、气压、鼻流、鼻压信号的处理之外，还可以做 EMG（肌电仪）、EGG（后头仪）数据记录和处理，参看图 4 和图 5。

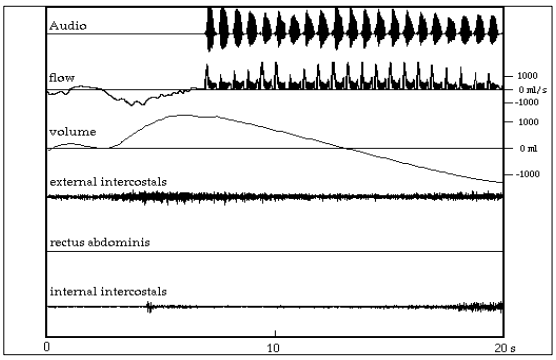


图 4 声音、气流和 EMG 信号

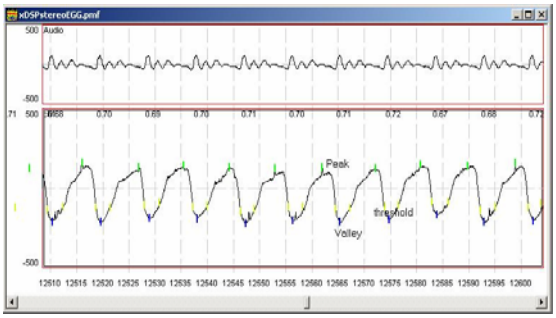


图 5 声音和 EGG 信号

（2）声学分析：

a. 高质量的语图分析，频谱、基频分析；

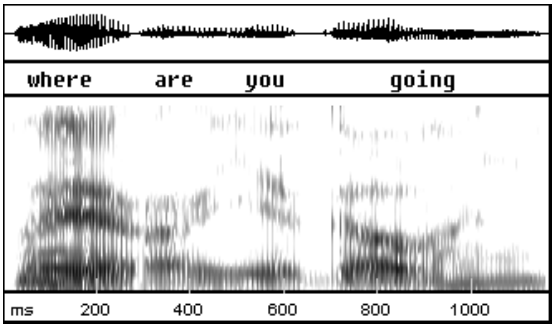


图 6 宽带语图

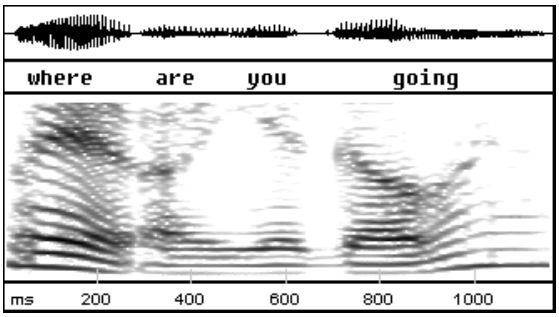


图 7 窄带语图

b. 快速博立叶变换（FFT），线性预测（LPC），音强分析；

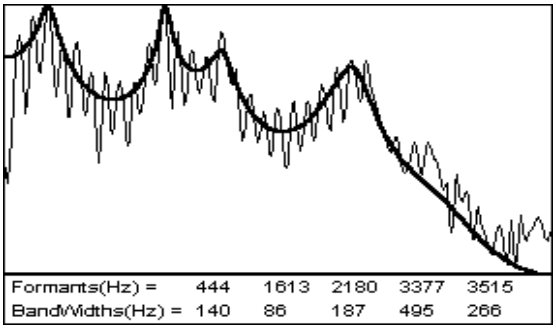


图 8 快速博立叶变换（FFT）

c. 完全标签系统在主界面、语图、音高图上。

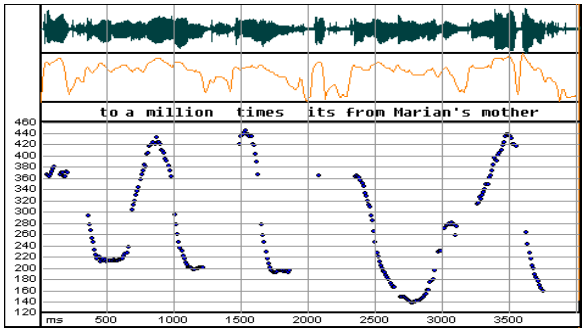


图 9 标签功能

另外，对单通道和双通道数据实行完全地波形编辑；直接读取 CSL,WAVES 格式的文件；自动日志记录在线编辑支持即时注释和其它试验记录；可以直接保存成位图格式；完全联机帮助系统等的特征。

3 仪器的操作方法

3.1 硬件的操作方法

首先看图 10 的口罩，基本上口罩和变换器连接使用。变换器上有四个通道，左边两个是测量口腔的通道，右边两个是鼻腔的；第一，三个通道是测气流的通道，第二，四是测气压的。

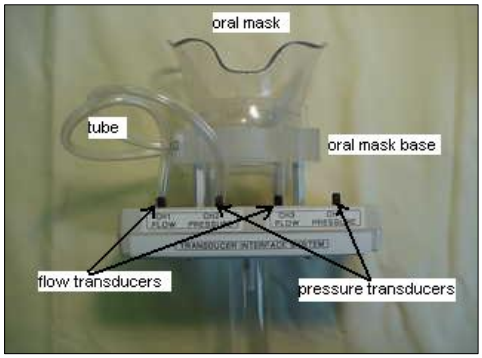


图 10 口罩

口罩和鼻罩也通过几根管子连在一起。在上文说明过，第三、第四个通道是测量鼻流的。但是，通常来说鼻压信号我们一般用不到，因为鼻压是非常小的，而且很多时候是没有。如果不用鼻压的话，把它塞住使得不要漏气。参看图 11。

图 11 鼻罩

下一个步骤是口罩和主系统的连接。参看下图。

图 12 口罩和主系统的连接

主系统是感应器，看变换器后边有四个小的输出口，中间还有一个大的输出口，如图 12 所示。按理说，四个通道分别输出，但是它用一根集成线，即四个通道集合到这一条迷你线。这条线插到变换器后边的大的输出口和感应器上。另外，感应器上的四个插口也不用，同一用一条线。但可以单独用，我们可以把 EGG 信号加进去使用。

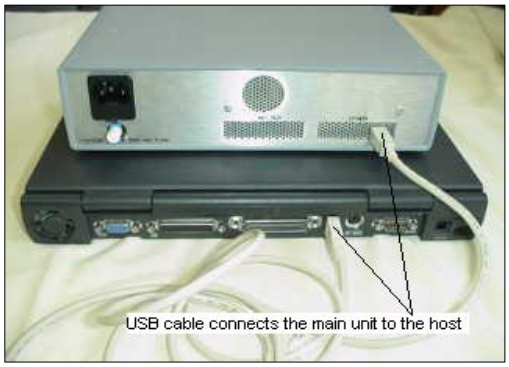


图 13 主系统和电脑的连接

最后，主系统的传输线连接电脑，这样组成了一个系统。要注意的是一定要先装好软件，然后连接硬件，使得电脑上不发生出错。现在发音人可以带上开始录音了。

图 14　仪器的使用方式

从上图 14 看出，我们实际使用的时候，口罩和鼻罩可以单独使用，也可以组合使用。我们要注意的是用鼻罩的时候，一定要把它拉紧，使得不要漏气，最好要有人帮忙。用口罩的时候也一定要贴近，不要让下巴那部分漏风。另外，口罩里边有一个管子，是要放在口里之后含住它。这个管子放在上齿龈，要靠上一点力求发 d、t、n 的时候能够保持那个阻塞。通常用这个口罩的时候有阻碍，因为这个管子很硬。所以这个管子越细越好，当然太细了气流不能通过也不行。

3.2 软件的操作方法

首先打开软件，然后按下列的顺序打开选择项：Options => Record/Play

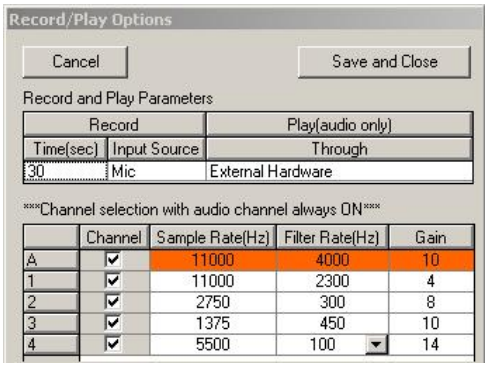


图15 录音/播放选择项

正式录音之前，我们要先设计这些选择项。Sample Rate(采样频率)一般设为11025（Hz），设计Filter Rate（滤波器）的时候，如果把声音也要录进去的话，选4000为合适。如果你想把嗓音信号去掉的话，可以把滤波器设的100，75（Ladefoged，P.(2003)比较推荐75Hz）等。可以其他通道都设4000。设了之后一定要保存（按上侧右边的save键）。

然后，软件有事先可以演示的功能。按照下面的顺序打开Scope Mode，最后确认各个通道有无信号：

Options => Record/Play => Scope Mode

没有出现问题的话，打开新的文件，接着按上边右侧的红色小圆头键，或者按下列的方式来开始录音。录完了一定要保存，参看下列顺序。

Options => Record/Play => Record

File => Save

其他使用方法我们可以参考使用手册，本文不涉及详细说明。

4 仪器的应用

用气流气压计研究的成果大体分成两大类：一类是语音学领域的，另一类是医学领域的研究。

整体来说，国内比较缺乏与嗓音和韵律有关的研究，相比方言研究呈发展趋势。下文主要介绍用于气流气压计研究汉语的实例。

4.1 语音学领域的研究

（1）普通话辅音研究

首先，我们看一下吴宗济（1987）先生的“汉语普通话辅音不送气/送气区别的实验研究”。该文章属于中国最早期的气流气压实验研究之一。国外其他辅音的气流气压特征，我们可以参看 Ladefoged.P.(1993)。

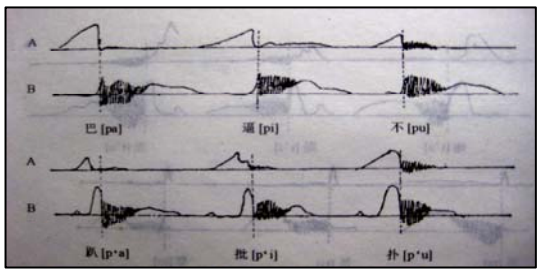


图 16 普通话塞音不送气/送气音的气流气压

图 16 是汉语普通话塞音[p]/[p‘]的声门上的气流气压值。 A 为声门上气压；B 为声门上气流。从图中得知，塞音不送气音和送气音在除阻之前都有一定的压力峰值，而不送气音的压力较大。我们一般认为可能送气的压力比较大而叫它

有气的“送气音”，但从实验结果看，其实不然。原理上说，持阻前声门上压力的大小持阻时间的长短成正比，所以，不送气音的气压更大。至于气流量不送气/送气辅音之间没有显著的区别。

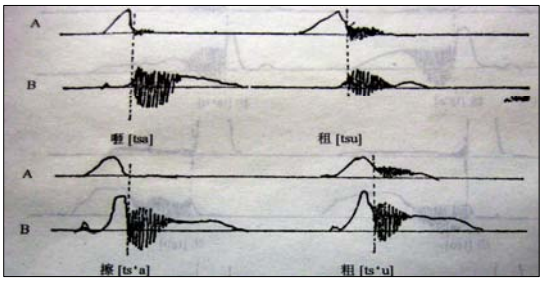


图 17 普通话塞擦音不送气/送气音的气流气压

上图 17 为普通话不送气、送气塞擦音的气压和气流。塞擦音在声学分析中较易观察到它是先塞后擦，然后不送气塞音的流量峰值减弱时就接上元音，而送气的塞擦音就在流量峰之后，还有比较长的一段微弱气流，才接上元音。送气的塞擦音还出现了两次流量峰。

这篇文章以气流气压计的实验揭示汉语普通话不送气/送气辅音真正的区别不在于气流量，而在于共振峰的不同。至于塞擦音，此实验再次证明理论上的生理机制。

（2）普通话儿化韵研究

下面的是LEE Wai-Sum（2001）研究的北京儿化韵实验分析，该研究主要观察口流和鼻流的变化而解释儿化之后的一些变化。

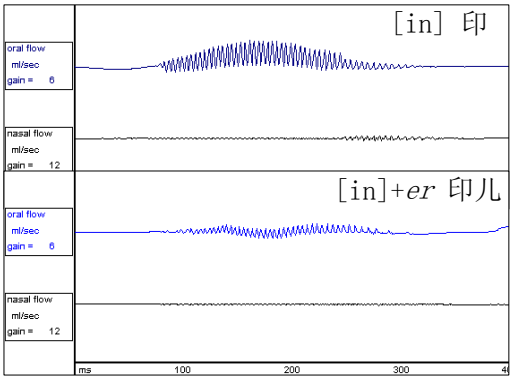


图18 “印”和“印儿”的口流鼻流比较

实验结果表明，发“印”时，气流减弱时，就有一点点的鼻流出现。相比，发印兒的时候，鼻流非常的弱，可以忽略掉。这说明变了儿化词，就取消了鼻韵母，以[i]来取代，如图18所示。

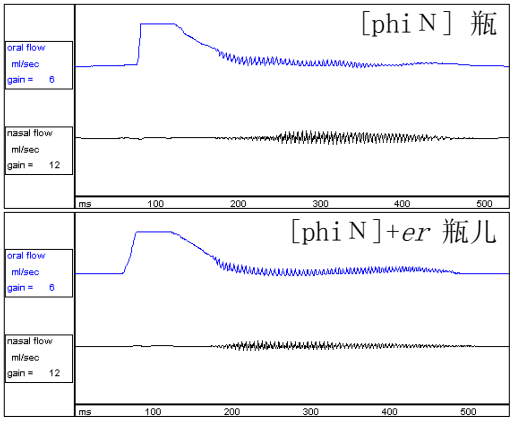


图19 “瓶”和“瓶儿”的口流鼻流比较

图19的情况有所不同。“瓶”变成“瓶儿”之后，口流和鼻流几乎同时出现，即由完全鼻化的i来代替鼻韵母。

（3）方言语音研究

最近方言语音生理研究正在变得越来越活跃，使得对方言语音的特征和难以解决的问题给予更准确的描写和解释。Cun Xi（2005）使用气流气压计对 Wuyang 和 Wenchang 的内爆音和浊爆音做了比较，观察口流和口压的变化，力图找出内暴音和浊爆音的区别特征。一般认为内爆音的口压应该是负值，但从下图 20 看出，Wuyang 方言的内爆音的口压值却时负时正。这是因为在 Wuyang 方言的内爆音没有和浊爆音对立，所以二者不必区分。国外已有内爆音呈正值的研究，参看 Ladefoged, P(1963)的文章。

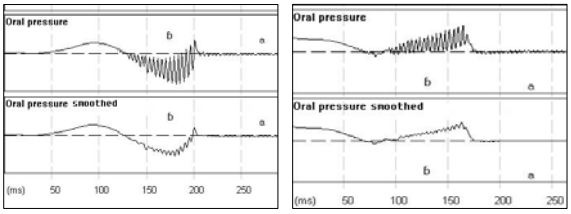


图 20 Wuyang 方言区内爆音的口流（上）口压（下）

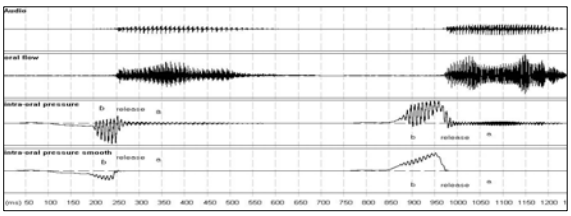


图 21 Wenchang 方言区的内暴音（左）和浊爆音（右）

相比，Wenchang 方言里有内爆音和浊爆音的对立，要严格区分。Wenchang 方言的内爆音口压呈负值，而浊爆音的口压呈正值，如图 21 所示。

从该实验结果提出区分内爆音和浊爆音的特

征不一定是口压的负值表现。最后以通过声门的气流量（transglottal airflow）的不同来解释二者的区别。虽然对通过声门的气流量的定义和测量方法不太明确，但是从方法上的突破给了方言语音研究非常大的启发。

应用气流气压计也可以考察方言语音的演变过程，我们可以参看曾婷（2004）的“湘乡方言[n]和[l]的气流与声学分析”。 湘方言大部分地区[l]和[n]既有分也有混，总的情况是洪音前[l]/[n]不分；而细音前则有[l]和[n]的区别。关于[l]和[n]分布的规律还有：[n]只出现在有鼻元音或者有鼻尾的韵母前，而[l]只出现在口韵母前。但是关于湘方言[n]和[l]的这些规律只是人们或者研究者的一种听觉感知，而没有相关实验语音学的研究和分析。

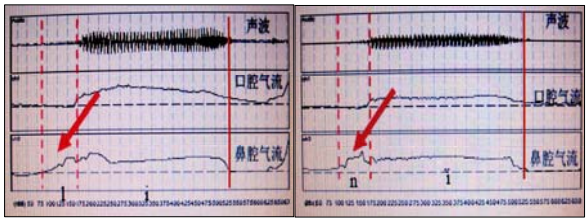


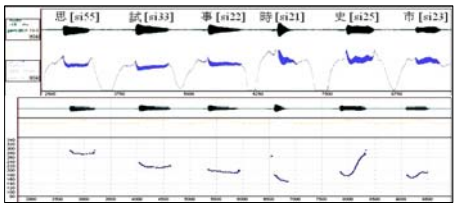
图 22 [li]（左）和[ni]（右）的声波、口腔及鼻腔气流

[l]和[n]的气流分析结果表明，两者在气流上没有区别，即它们在口腔阻塞阶段都表现出很强的鼻气流，如图 22 所示。

[i]位于非鼻音环境下（零声母），在整个元音阶段几乎没有鼻气流，但当它位于[l]后时却被完全鼻化了。这说明以前研究把这个音段描写成[l]是有问题的。结合气流分析的结果，可以看出湘乡方言中的[l]是被鼻化了的[l]---[ln]。基于以上的气流和声学研究结果，湘乡方言有一个[l]和[n]趋于合并的历史发展方向（[l]-》[n]）。

（4）韵律研究

气流气压计的应用范围可以扩展到韵律研究。Yuk-Man Cheung（2004）做了关于香港话韵律的实验分析。该研究的目的为揭示香港话的基频和口流的相关性并找出基本呼吸单位。



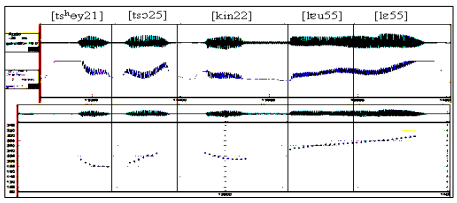


图 23 香港话[si]的六个声调的基频和口流变化（上），一段从句的基频和口流变化（下）

实验结果表明，无论是音节还是从句基频和口流变化都呈现出较大的相关度，如图 23 所示。但问题是句中出现鼻音音节的话，造成口流量的减少而其相关度也呈下降。参看图 24[pin55]的口流变化。除此之外，协同发音等其他因素也会影响二者的相关度。

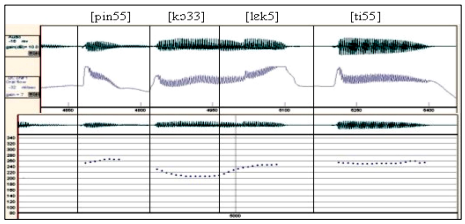


图 24 带鼻音音节的句子的基频和口流变化

下一步，使用气流气压计找出基本呼吸单位也并不简单。实验结果表明，由于个人差异较大而难以定义基本呼吸单位。如图 25 所示。

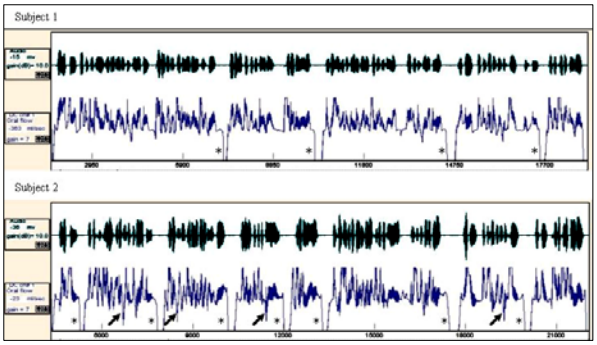


图 25 香港话语流中的呼吸单位。星号为大的呼吸单位，箭头为小的呼吸单位。

4.2 医学领域的研究

医学方面的研究主要是腭裂患者的语音特征，即腭裂语音。总体来说，在中国有关腭裂的文章不少，但大部分的研究都用语图分析或者医生的听辨分析来观察患者的情况。如，鼻漏气等。

相比，国外从1963年由美国Warran和Debois使用气流气压气压技术来评价腭咽闭合功能（刘辉，1990）开始，用于多种生理实验仪器来研究腭裂语

言，给与腭裂患着的语音治疗实际的帮助。国内也非常需要空气动力学方面的临床实验研究。

5 仪器的缺陷

笔者认为使用气流气压计的缺陷以下几点：发音人含住面罩的管子，说话时发音准确会受影响；发音人自己手持口罩进行录音，很容易漏风；有些音比较难以测量（例,舌根音）；卫生管理十分不便等。但是随着技术的发展，新出现的仪器有的已经克服以往的仪器具有的缺陷，如 Glottal Enterprise 解决了由口罩造成的声音变形问题。

6 结语

本文简单介绍生理语音学仪器之一的气流气压计的使用方法以及国内应用情况。使用气流气压计我们能够采集语音的生理信号，如，口流、口压、鼻流、鼻压等，而对各种言语和语音病变的各种研究。但是，与国外相比，目前国内气流气压计的应用范围还不够广泛，普及率也比较低。因此，国内用于气流气压计做的言语研究还处于初步阶段，至于语音病变的实验研究几乎没有。为了中国语音学的发展，我们应该充分引入国外先进技术和仪器，对语音生理的研究继续深入。

参考文献

[1] 吴宗济. 1987. 普通话辅音不送气/送气区别的实验研究. 吴宗济语音学论文集. 北京. 商务印书馆.

[2] 孔江平. 2007. 中国语音学研究的历史与现状. 北大语音乐律实验室论文集

[3] 曾婷. 2004. 湘乡方言[n]和[l]的气流与声学分析. 第七届中国语音学学术会议暨前沿问题国际论坛.

[4] 李永宏，孔江平，于洪志. 2008. 现代语音学仪器及生理语音学研究. 生命科学仪器. 第6卷. 第9期.

[5] 刘辉. 1990. 气流动力学技术在口腔颌面外科的应用-对腭裂患者语音及腭咽闭合功能的研究. 国外医学口腔医学分册. 第17卷. 第4期.

[6] Baken, R. J., and Orlikoff, R. F.2000. Clinical measurem -ent of speech and voice. (2nd edn.). chapter 8 and 9. San Diego: Singular Publishing.

[7] Cun Xi. 2005. The Phonetic Characteristics of Implosiv -es in Two Chinese Dialects.

[8] Dart, S. 1987. An aerodynamic study of Korean stop sonsonants: Measurements and modeling. Journal of the Acoustical Society of America, 81(1), 138-47.

[9] LEE Wai-Sum. 2001. A Phonetic analysis of the *er-hua yun* 儿化韵 in Beijing mandarine proceedings of the 5th national conference on modern phonetics. Beijing: Tsinghua University.

[10] Ladefoged, P., 1962. Subglottal activity during speech. Proceedings of 4th International Congress of Phonetic Sciences, 73-91.

[11] Ladefoged, P., 1963. Some physiological parameters in speech. Language and Speech 6, 109-119.

[12] Ladefoged, P. 1964. A phonetic study of West African languages, Cambridge: CambridgeUniversity Press.

[13] Ladefoged, P., 1967. Three areas of experimental phonet -ics. Oxford: Oxford University Press.

[14] Ladefoged, P. 2003. Aerodynamic Investigations. Phon -etic Data Analysis. Oxford: Blackwell Publishing.

[15] Nihalani, P. 1986. Phonetic implementation of implos -ives, Language and Speech, 29, 253-262

[16] Ohala, J. J. 1995) Experimental Phonology In A Handbook of Phonological Theory edited by Goldsmith J. A. Oxford: Blackwell 713-722

[17] Rothenberg, M. 1971. The Glottal Volume Velocity Wav -eform during Loose and Tight Voiced Glottal Adjustme -nts. Proceedings of the Seventh International Congress of Phonetic Sciences. pp.380-388.

[18] Rothenberg, M. 1973. A new inverse-filtering technique for deriving the glottal air flow wave-from during voicing. Journal of the Acoustical Society of America, 53, 1632-45.

[19] Rothenberg, M. 1977. Measurement of airflow in speech. Journal of Speech and Hearing Research, 20, 155-76.

[20] Shadle, C., and Scully, C. 1995. An articulatory-acoustic -aer9odynamic analysis of [s] in VCV sequences. Journal of Phonetics, 23, 53-66.

[21] Shadle, C. 1997. The aerodynamics of speech.. In Hardc -astle ,W. J. and Laver,J. (eds.) The Handbook of Phonetic Sciences. Oxford: Blackwell.

[22] Warren, D. W. 1996. Regulation of speech aerodynamics, in Lass, N. (ed.). Principles of experimental phonetics. St. Louis: Mosby, pp.46-92.

[23] Warren, D. W. , et al. 1964. Cleft Palate. J. 1.52-71.

[24] Yuk-Man Cheung. 2004. An Aerodynamic Analysis of Intonation in Hong Kong Cantonese. Speech Prosody 2004.

（吴韩娜 北京大学中文系语音音律实验室 100871 hanna.pku@gmail.com）

呼吸带说明

An Introduction to Respiratory Belt

吴君如

Wu Junru

摘要： 本文旨在介绍呼吸带的工作原理和使用方法，提供一种呼吸参数的提取方法。此外，本文还简要介绍了前人使用呼吸信号进行的语言学研究。

Abstract： This paper aims at introducing the working principle and usage of respiratory belt, providing one of the methods for the extraction of respiratory parameters. On the other way, some approaches of linguistic research with respiratory belt are provided.

关键词： 语音生理学 呼吸带 说明

Key words： physiological phonetics, respiratory belt, introduction

0 引言

呼吸带在呼吸生理研究领域已经有了广泛的利用，在语言学领域，呼吸带主要用于言语呼吸动力及朗读节奏的研究。

1 呼吸带简述

1.1 呼吸带原理

首先，让我们稍微了解一下呼吸运动的生理基础。呼吸运动是由肺容积的改变决定的，吸气时肺的容积变大，呼气时肺容积缩小。而肺容积是由胸腔和腹腔容积共同决定的，呼吸运动是胸腹联合运动的结果。胸腔体积可以通过胸廓的运动而增大也可以通过横隔膜的运动来增大。由于这两部分是共同作用的，所以肺部体积的变化应该是这两部分体积变化的和。如果胸腔扩大而横隔膜收紧，那么它们都会导致吸气。但是，如果胸腔体积缩小，而同时横隔膜收紧，那么是呼气还是吸气就取决于胸腔缩小的体积是否大于腹腔扩大的体积。因此我们必须明白不能单凭胸腔或腹腔的变化知道此时是吸气还是呼气，在判断是呼气还是吸气之前，我们必须首先了解这两个部位的变化以及它们对于胸腔体积变化的贡献度。现在利用呼吸带进行研究通常采集两条呼吸带的信号，一条采集胸部的呼吸信号，一条采集腹部的呼吸信号。这是因为肺容积是由胸腔和腹腔容积共同决定的，呼吸运动是胸腹联合运动的结果。(Konno and Mead 1967; Hixon 1973; Hixon, Mead et al. 1976; Baken and al. 1979)

呼吸带利用的正是呼吸时胸腔或腹腔周长的改变。呼吸带的核心是一个压电传感器，这个压电传感器能够用电压信号反映长度的线性改变，利用这个特性，我们可以用呼吸带来测量呼吸时胸腔或腹腔周长的改变。

1.2 呼吸带结构

一条呼吸带包括三个部分：压电传感器、松紧带、电缆。呼吸带产生的原始信号须经采集器处理，才能输入计算机，用于分析。

图 1(MLT1132 Piezo Respiratory Belt Transducer, ADInstruments)

1.3 呼吸带生产商

现在语言学研究采用的呼吸带主要有两种：一种是澳大利亚埃德仪器公司（ADInstrument）生产的 MLT1132 压电呼吸带传感器（MLT1132 Piezo Respiratory Belt Transducer），另一种是美国动态监测公司（Ambulatory Monitoring, Inc.）生产的 Respitrace system。前者生产的呼吸带须和该公司生产的生物电信号采集器（PowerLab）及 Chart 软件配合使用，后者生产的呼吸带须和该公司生产的 Inductotrace 接口及 Inductoview 软件配合使用。(ADINSTRUMENTS: Ambulatory Monitoring)

2 呼吸带的使用方法

2.1 实验设备

以下以 MLT1132 压电呼吸带传感器为例。

（1） MLT1132 压电呼吸带传感器（呼吸带）一条或两条；

（2）生物电放大器（如 ML880 PowerLab 16/30）一部；

（3）计算机一部，须装有 Windows 或苹果系统；

（4） Chart 软件（如 Chart5 for Windows）；

（5）（可选）话筒一个，用于同步采集语音信号，须为莲花插头或转接莲花插头；

（6）（可选）电子声门仪（EGG）一套，用于同步采集声门阻抗信号；

2.2 操作步骤

（1）启动计算机；

（2）初次使用应安装 Chart 软件（也可在 4. 打开生物电放大器后安装）；

（3）接好生物电放大器电源，并将呼吸带电缆连接到生物电放大器一个 BNC 输入口上，如果要同步采集语音信号，将话筒莲花头接到生物电放大器的另一个 BNC 输入口上，如果要同步采集 EGG 信号，将 EGG 的 BNC 输出口接到生物电放大器的另一个 BNC 输入口上；

（4）开启生物电放大器电源；

（5）运行 Chart 软件，新建文件，设置通道数（每条呼吸带、话筒、EGG 各占一个通道）、采样频率（如果和声音信号同步使用，建议使用 1000Hz 以上采样频率）、和信号量程（呼吸信号 10mV-20mV）；

（6）将第一条呼吸带环扎于被试胸部，上缘至腋下五厘米；若使用两条呼吸带，另一条呼吸带环扎于被试腹部，上缘在肚脐位置，不能接触下肋，以免胸腹呼吸信号混淆。实验者可以根据需要调整呼吸带的位置，原则是要将呼吸带系于呼吸时胸或腹扩张最大处，在吸气至最大时，呼吸带应接近最大延展极限，以保证呼吸带达到其最高灵敏度，并保证采集到的呼吸信号和胸腹周长线性相关。(Denny and Smith 2000; McFarland 2001; Connaghan, Higashakawa et al. 2004; Huber 2007; Huber 2008)

7.根据 Chart 软件显示的呼吸信号调整软件设置，既要充分利用量程，又要避免出现过载削波现象。

注意，计算机启动完毕再启动生物电放大器，以免出现设备冲突等问题。

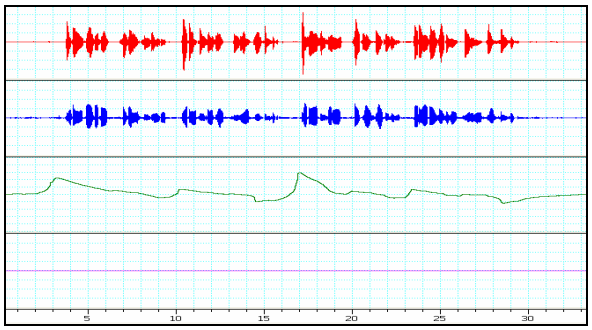


图 2 用 Chart5 采集的四通道信号（谭晶晶 2008）

3 呼吸参数的提取和分析

3.1 呼吸参数的定义

对呼吸信号进行分析，首先需要定义一些概念和参数。下面以(谭晶晶 2008; 谭晶晶 2008; 谭晶晶, 李永宏 et al. 2008)的研究为例说明一种呼吸信号参数的提取方法：

呼吸曲线：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这就是“呼吸曲线”，呼吸曲线上升表示吸气，下降表示呼气。

呼吸重置时长：从开始吸气到开始呼气之间经过的时间，即图中波谷和波峰之间的水平距离 T。

呼吸重置幅度：一次吸气过程中呼吸信号数值的变化幅度，即图中中波谷和波峰之间的垂直距离 A。

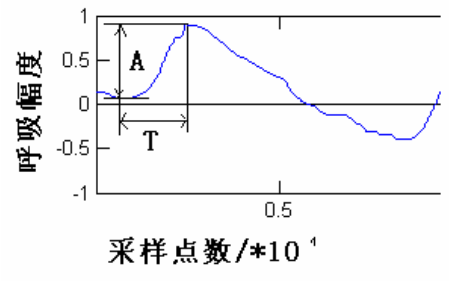


图 3 参数定义（谭晶晶 2008）

3.2 呼吸参数的提取

参数提取可以使用北京大学中文语音乐律平台呼吸信号处理软件（BreathBand），该软件须在 Matlab 环境下运行，可以对呼吸信号进行归一化和平滑滤波，对呼吸重置时长和呼吸重置幅度进行自动标记和手工标记，还可以批处理标记好的文件，自动标记数据到 EXCELL 表格中。

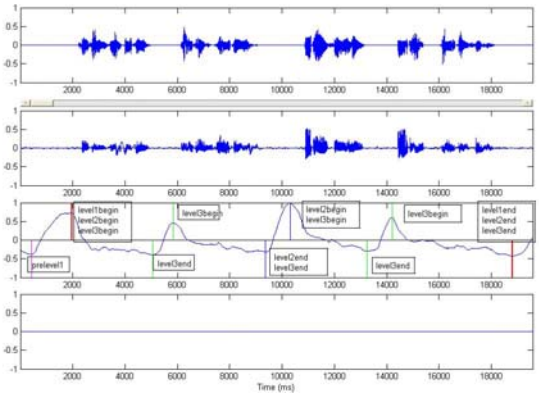


图 4 北京大学中文语音乐律平台呼吸信号处理软件（BreathBand）界面。从上往下第一格为语音信号，第二格为 EGG 信号，第三格为第一通道的呼吸信号

4 利用呼吸信号进行语言学研究

使用呼吸信号进行的语言学研究主要有以下几个方面：

（1）语言的呼吸动力学研究。研究表明，说话时呼吸动力模式和正常呼吸的呼吸动力模式是大为不同的(Hoit 2000; Connaghan, Higashakawa et al. 2004)。在这个基础上，还有呼吸动力参数和声强的关系的研究。声强指的是声波携带的按分贝计的总能量，语音的声强与发音动力息息相关，而呼吸带采集的呼吸信号可以反映发音动力的情况。虽然通常两者间是正相关关系，但这种关系显然是复杂的和非线性的。(Huber 2007)

（2）语言心理研究。一则谈话互动研究发现自然交谈中说话人转换话题时会伴随呼吸信号的改变，而且这时听话人的呼吸信号也会相应地发生改变。(McFarland 2001)

（3）语言病理研究。有研究者利用呼吸信号研究构音障碍的模式和变异(Hammen 1994)，还有研究者将呼吸信号和脑电（EMG）信号结合起来研究口吃者的呼吸神经控制问题。(Denny and Smith 2000)

（4）朗读呼吸节奏研究。早期研究即已发现，呼吸重置往往发生在语法边界上(Hixon 1973; Hixon, Mead et al. 1976; Baken and al. 1979)。近期研究发现呼吸重置可以依其幅度分为不同级别，呼吸重置的级别与呼吸边界前的高音点与低音点的音高有关，不同级别的呼吸重置的呼吸边界上无声段时长有显著差异，重置时长和重置幅度的相关。(谭晶晶 2008; 谭晶晶 2008; 谭晶晶，李永宏 et al. 2008)

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（谭晶晶 2008）

4 总结

以上介绍了呼吸带的工作原理和使用方法，并提供了一种呼吸参数的提取方法。呼吸带在语言学领域是一个比较新的研究手段，在言语呼吸动力研究和语言节奏研究方面日益展现出其迷人风采。

参考文献

[1] ADINSTRUMENTS. "MLT1132 Piezo Respiratory Belt Transducer." from www.ADINSTRUMENTS.com.

[2] Ambulatory Monitoring, I. "Battery Operated Inductotrace System." from www.ambulatory-monitoring.com.

[3] ADINSTRUMENTS. "MLT1132 Piezo Respiratory Belt Transducer." from www.ADINSTRUMENTS.com.

[4] Ambulatory Monitoring, I. "Battery Operated Inductotrace System." from www.ambulatory-monitoring.com.

[5] Baken, R. J. and e. al. (1979). "Chest wall movements prior to phonation." Journal of Speech and Hearing Research **22**: 862-872.

[6] Connaghan, K. P., M. Higashakawa, et al. (2004). "Respiratory kinematics during vocalization and nonspeech respiration in children from 9 to 48 months." Journal of Speech, Language, and Hearing Research **47**(1): 70.

[7] Denny, M. and A. Smith (2000). "Respiratory Control in Stuttering Speakers: Evidence From Respiratory High-Frequency Oscillations." Journal of Speech, Language, and Hearing Research **43**(4): 1024.

[8] Hammen, V. L., & Yorkston, K. M. (1994).

"Respiratory patterning and variability in dysarthric speech." Journal of Medical Speech-Language Pathology **2**: 253-262.

[9] Hixon, T. J., Goldman, M. H., & Mead, J. (1973). "Kinematics of the chest wall during speech production: Volume displacements of the rib cage, abdomen, and lung." Journal of Speech and Hearing Research **19**: 297-356.

[10] Hixon, T. J., J. Mead, et al. (1976). "Dynamics of the chest wall during speech production: Function of the thorax, rib cage, diaphragm, and abdomen." Journal of Speech and Hearing Research **19**: 297-356.

[11] Hoit, J. D. a. H. L. L. (2000). "Influence of Continuous Speaking on Ventilation." Journal of Speech, Language, and Hearing Research **43**(5): 1240.

[12] Huber, J. E. (2007). "Effect of cues to increase sound pressure level on respiratory kinematic patterns during connected speech." Journal of Speech, Language, and Hearing Research **50**(3): 621.

[13] Huber, J. E. a. J. S., III (2008). "Age-related changes to speech breathing with Increased vocal loudness." Journal of Speech, Language, and Hearing Research **51**(3): 651.

[14] Konno, K. and J. Mead (1967). "Measurement of the separate volume changes of the rib cage and the abdomen during breathing." Journal of Applied Physiology **22**: 407-422.

[15] McFarland, D. H. (2001). "Respiratory Markers of Conversational Interaction." Journal of Speech, Language, and Hearing Research **44**(1): 128.

[16] 谭晶晶 (2008). 普通话朗读时的呼吸节奏研究. 中文系, 北京大学. **硕士**.

[17] 谭晶晶 (2008). "新闻朗读的呼吸节奏研究." 中国语音学报 **1**.

[18] 谭晶晶, 李永宏, et al. (2008). 汉语普通话不同文体朗读时的呼吸重置研究. 清华大学学报——2007 年第九届全国人机语音通讯学术会议 NCMMSC 特刊. 北京, 清华大学出版社.

（姓名 吴君如 单位 北京大学中文系语音乐律实验室 邮编 100871 电子邮件 yuerr@163.com）

脑电帽原理和使用简介

杨若晓

YANG Ruoxiao

0 引言

人类语言是一套特殊的符号系统，具有十分复杂的心理加工机制。语言心理加工的生理基础就是人脑。人们对语言的脑机制研究由来已久，从初期通过解剖研究非正常语言能力者的大脑发现和语言行为息息相关的 broca 脑区和 wernickle 脑区，到现在技术进步背景下利用各种脑成像技术对人脑语言机制的各种探索。而在各种脑成像技术中，脑电技术（Electroencephalogram，简称 EEG）以其具有的高时间分辨率特点和性价比高而被广泛利用。而在脑电技术中，事件相关电位（Event related potential，简称 ERP）通过有意地赋予刺激以特殊的心理意义，利用多个或多样的刺激引起的脑的电位，成为一种特殊的脑诱发电位。由于它不仅能够反映大脑的单纯生理活动，还能反映认知过程中大脑的神经电生理的变化，因而被称为认知电位，在对语言加工的研究中广泛运用。本文将首先对脑电技术原理和 ERP 技术原理进行简单介绍，之后对目前为止发现的与语言加工过程有关的 ERP 成分进行简单说明，最后介绍 16 导脑电帽（MLAEC1/EC2 EEG Electro-Cap System）的基本使用方法。

1 脑电技术简介

脑电技术，或者脑电图（EEG）是通过置于头皮表面的电极记录的脑波图谱，是用神经电生理的方法检测而得到的脑神经细胞的活动。脑电图的最大的优越性在于时间分辨率相当高，大约在 1 毫秒左右，这就使得脑电能够相当好地记录到脑波的上升和下降。人类的自发 EEG 波幅（amplitude）约为 10-100μv（1 微伏=1 伏特的百万分之一）。而由心理活动引起的脑电要比自发脑电更弱，一般只有 2 到 10 微伏，通常淹没于自发电位中，所以 ERP 需要从 EEG 中提取。

2 事件相关电位（ERP）技术原理

事件相关电位（ERP）是经由将记录到的 EEG 脑部原始电生理信号进行再分析处理而得到的。它是经由内在事件或外在事件刺激所引发的脑电位波形变化，因而可以反应出人类生理或心理活动相关的脑电活动，一般用来研究大脑处理刺激至反应认知过程的活动过程。

2.1 事件相关电位 ERP 的基本定义

事件相关电位（ERP）的定义有广义和狭义之分。从广义上来说，凡是外加一种特定的刺激于有机体，在给予刺激或撤销刺激时，在神经系统任何部位引起的电位变化都可称为事件相关电位。从狭义来讲，事件相关电位是指凡是外加一种特定的刺激，作用于感觉系统或脑的某一部位，在脑区引起的电位变化，目前一般 ERP 仅指该狭义定义。有时为更清楚地专指脑产生的事件相关电位，有的场合也会使用“事件相关脑电位（event-related brain potentials）” 的说法 (魏景汉，罗跃嘉, 2002)。

2.2 提取 ERP 的基本原理

自发的脑电（EEG）成分复杂而不规则，而一次刺激所诱发的 ERP 波幅约为 2-10 微伏，比自发的 EEG 电位要小得多，淹没于 EEG 中，二者构成小信号和大噪音的关系，无法直接测量研究，所以 ERP 需要从 EEG 中提取。

2.2.1 ERP 的主要特点

ERP 的主要特点有以下三个方面：

2.2.1.1 ERP 需要开放电场

脑电（EEG）是由于皮质大量神经组织的突触后电位同步总和而成，而单个神经元电活动非常微小，不能在头皮记录到，只有神经元群的同步放电才能记录到。这种脑组织神经元排列方向一致的情况，构成所谓的开放电场（open field），反之则是方向不一致相互抵消的封闭电场（closed field）。因此，ERP 只能反映某些脑部的激活情况，而有些脑部即使处于激活状态，但由于其神经元没有能够形成开放电场，ERP 上也是无法反映的。

2.2.1.2 ERP 的潜伏期和波形

ERP 有两个重要特性：潜伏期恒定和波形恒定。潜伏期就是 ERP 波形与刺激间的时间间隔。与 ERP 的两个特性相对，自发脑电则是随机变化的。所以，所以利用这两个恒定就可以通过叠加从 EEG 中将 ERP 提取出来。

2.2.1.3 ERP 是平均诱发电位

ERP 是通过对原始 EEG 进行叠加得到的，也就是说叠加 n 次后的 ERP 波幅增大了 n 倍，因而需要再除以 n，使 ERP 恢复原形，即还原为一次刺激的 ERP 数值。所以 ERP 也被称为平均诱发电位，平均指的是叠加后的平均。

2.2.2 ERP 的提取基础和过程

下面简单介绍 ERP 的提取基础和过程。

2.2.2.1 ERP 的采集装置

首先介绍 ERP 的采集装置。

2.2.2.1.1 采集 ERP 时的头部定位系统

采集 ERP 时的头部定位系统是一个电极帽，上面有多个记录或吸收头皮放电情况的电极，这些电极在帽子上的位置是根据国际脑电学会在 1958 年制定的 10-20 国际脑电记录系统(Jasper, 1958)而设置的。

10-20 系统的原则是头皮电极点之间的相对距离以 10%和 20%来表示，采用 2 条标志线：1）矢状线：从鼻根至枕外隆凸的连线，又称中线。从前往后标出 5 个记录点——Fpz, Fz, Cz, Pz 和 Oz。Fpz 之前与 Oz 之后各占中线全长的 10%，其余点间距皆占 20%。2）两外耳道之间的连线。从左至右也记录 5 个点——T3, C3, Cz, C4 和 T4。T3 和 T4 外侧各占 10%，其余点间距皆占 20%。

经过以上两线的边缘 4 点，以 Cz 为圆心画圆，4 个点间各在圆周上等距地取 2 个点，并在 Fz、C3、Pz、C4 间各取一个点，这样 10-20 系统共由 21 枚有效电极组成。其他的 16 导、32 导、64 导、128 导和 256 导电极帽也是根据 10-20 系统扩展而成的。

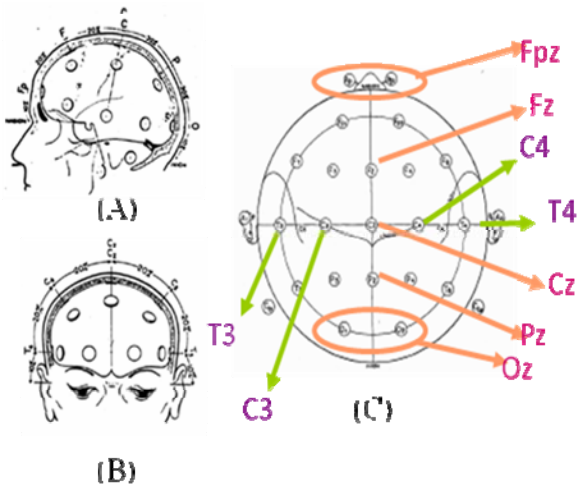


图 1 (A-C) 10-20 国际脑电系统示意图：(A) 矢状线；(B) 冠状线；(C) 10-20 国际脑电系统电极位置。

2.2.2.1.2 ERP 实验室的基本设置

ERP 实验室的基本设施主要包括：1）洗漱间，用于被试头皮处理，例如洗头等；2）主控间，用于监控试验程序呈现、脑波记录和被试实验；3）被试间，用于被试进行实验，应该具有较好的隔音和隔光效果。

2.2.2.2 采集 ERP 时的叠加技术原理

由于在实际采集数据过程中 ERP 是淹没在 EEG 中的，所以为了从 EEG 中提取出 ERP，需要对被试者施以多次重复刺激“S”，再将每次刺激产生的含有 ERP 的 EEG 加以叠加与平均，最后得到一次刺激的 ERP 值。ERP 可以通过叠加原始 EEG 数据获得的原因在于作为 ERP 背景的 EEG 波形与刺激间没有固定的关系，而其中所含的 ERP 波形在每次刺激后都是相同的（当然不是绝对的相同），并且 ERP 波形与刺激之间的时间间隔（即潜伏期）也是固定的，所以经过叠加，ERP 会随叠加次数成比例地增大，而 EEG 则按照随机噪音的方式加和。于是，经过叠加后的 ERP 就会从 EEG 的背景中浮现出来。

2.2.2.3 ERP 数据采集流程

下图 2 和图.3 分别给出了普通情况下 ERP 实验过程的示意图和 ERP 数据提取的主要过程示意图。

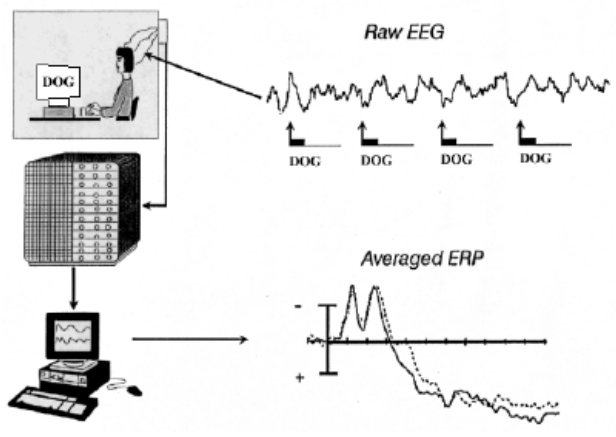


图 2 ERP 实验示意图（DOG，Data output gate，数据输出器）

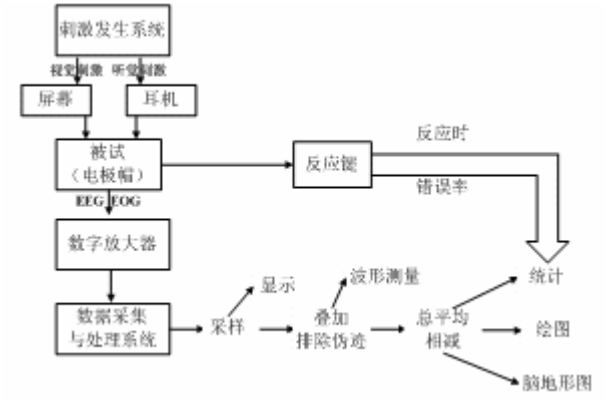


图 3 ERP 数据提取的主要过程示意图

2.2.3 ERP 信号的优点和缺点

ERP 的优点在于：1）无创性和时间分辨率（ms）高；2）便于与反应时配合进行认知过程（认知可分为认知过程和认知状态，过程指的就是时间过程）研究；3）设备相对简单，对环境的要求不高。

ERP 的主要缺点在于低空间分辨率，ERP 在空间上只能达到厘米级，主要的影响因素是容积导体效应与封闭电场问题。另外，ERP 只能采用数学推导来实现脑电的源定位，这种方法获得的脑内激活区域位置的可靠性也是有限的。

3 和语言加工有关的 ERP 成分

和语言加工有关的 ERP 成分主要包括 MMN、CPS、(E)LAN、P600/SPS 和 N400。

首先，在语音加工层面，MMN（mismatch negativity, 失匹配负波）反映了脑对听觉刺激的物理属性的变化进行的自动化的前注意加工过程，MMN 在 1978 年被首次报道 (R. Näätänen, Gaillard, & Mäntysalo, 1978)。随后的研究表明 MMN 可以反映出脑对语音类别的敏感性(Risto Näätänen et al., 1997)，因而它常常被用来作为语音范畴所对应的一种脑电成分。近来一些研究还试图在更大的语言单位和高级的认知过程中寻找 MMN 的痕迹(Pulvermuller & Shtyrov, 2006)。测量 MMN 的实验中一般给被试呈现一系列标准的听觉刺激，在这些标准刺激中随机插入与之不同的变异刺激，这时，如果变异刺激与标准刺激相比属于不同的类别，即跨越了类别的界限，就会看到非常明显的 MMN，而同一类别内的变异刺激则不会诱发出 MMN。另外一种和语音加工关系密切的 ERP 成分是新近发现的 CPS（closure positive shift），CPS 是在关于德语的韵律研究中被发现和命名的 (Steinhauer, Alter, & Friederici, 1999)，它和自然语言的韵律加工关系密切，CPS 的发现在一定程度上说明韵律因素对句法分析的作用。最近基于汉语的 CPS 研究也得到了开展 (Li & Yang)。

(E)LAN 和 P600/SPS 是主要反映句法加工的 ERP 成分。(E)LAN（（Early）Left anterior negativity，左前负成分）与句子的句法加工密切相关,如果某个句子环境要求出现某个词类，而实际上出现的不是这个词类的词，那么就会诱发出 (E)LAN。代表句法违反的 LAN 成分独立于与语义违反相关的 N400，因而被认为是与句法违反的直接的指标(刘燕妮 & 舒华, 2003)。P600/SPS（Syntactic Positive Shift，句法正漂移）是与句法违反有关的、出现在关键词之后 500-600ms 的正成分(Osterhout & Holcomb, 1992)，它可以在多种句法违反中出现，如表 3.1 所示。

表 3.1 P600/SPS 效应示例(刘燕妮 & 舒华, 2003)

短语结构违反（Phrase structure anomalies）
The scientist criticized Max’s of proof the theorem.
句法类别歧义（syntactic－category ambiguity ）
The mental became refining by the goldsmith who was honored.
句子成分位置违反（Sentence-constituent movement anomalies）
I wonder which dress the guests at the party were shocked when the bride wore.
动词时态违反（Verb tense anomalies）
The cats won’t eating the food that Mary leaves them.

而在语义加工层面，由 Kutas 和 Hillyard(Kutas & Hillyard, 1980)在 1980 年发现的 N400 被视为反映语义加工的重要脑电指标。在他们的实验中，他们给被试视觉呈现 7 个词组成的句子，每个词单独呈现 700ms,直到句子结束，其中 75％的句子是语义一致的，25％的句子是最后

一个词与整个句子不一致的。语义不一致的条件会诱发出一个偏后侧分布的负波。

4 实验室脑电帽使用简介

北京大学中文系语言学实验室所有的脑电帽系统为配合 ADInstruments 公司的多导生理记录仪 PowerLab 和 LabChart 及 Scope 软件一起使用的脑电帽系统，其型号为：MLAEC2 EEG Electro-Cap System 2 (large & medium cap)，为 Electro-Cap International 公司生产，即下文所说的 ECI 脑电帽。

4.1 实验室脑电帽系统的组成部分

下面将整个脑电帽系统软硬件的构成分“脑电帽系统”和“采集系统”进行介绍：

1. 脑电帽系统：
   a) 16导（共16个电极）EC I™ 大型脑电帽和中型脑电帽各一个。
   b) 电极适配器（Electrode Board Adapter）。【实验室还未到位】
   c) 身体绑带（Body Harness）。
   d) 用于临时快速插入的电极2个（quick insert electrodes）。
   e) 耳部电极2个（ear electrodes）。
   f) 用于固定额头处两个参照电极的一次性圆形海绵体100个（disposable sponge disks）。
   g) 用于注射导电膏的注射器2支（needle/syringe kits）。
   h) 导电膏2盒（electro-gel）。
   i) 头部测量皮尺（head measuring tape）。
   j) 脑电帽清洗溶剂（ivory cleaning liquid）。
   k) 一本纸本使用说明（a manual）。
   l) 一张关于使用说明的DVD光碟。

图 4 脑电帽系统组成（MLAEC2 EEG Electro-Cap System 2 (large & medium cap)）

2. 采集系统：
   a) 16通道生物电放大器（GT201 16 Channel Bio Amp），如图5。
   b) 5导生物电放大缆线（MLA2540 5-Lead Shielded Bio Amp Cable），如图6。
   c) LabChart及Scope采集信号软件，如图7和8。

图 5 16 通道生物电放大器（GT201 16 Channel Bio Amp）

图 6 导生物电放大缆线（MLA2540 5-Lead Shielded Bio Amp Cable）

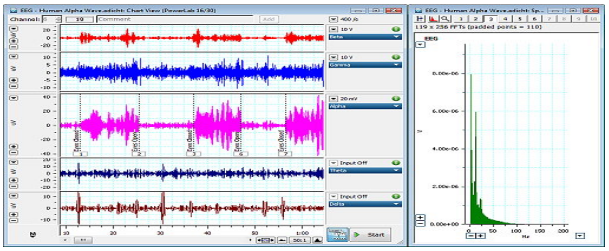


图 7 LabChart采集信号软件

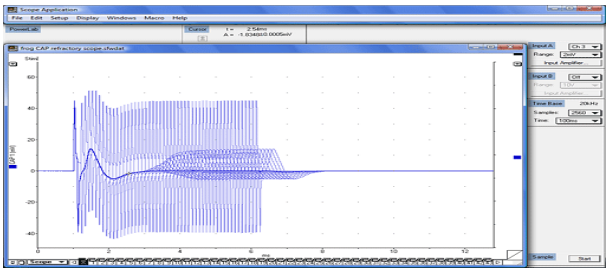


图 8 Scope采集信号软件

4.2 实验室脑电帽系统的使用简介

脑电帽系统的使用步骤由以下几个方面组

成，以下以成人被试为例进行介绍，如果以婴幼儿为被试，还有一些不同步骤和注意事项，但具体可以参看使用说明手册，在本文不做说明。使用步骤如下：

1. 安装电极适配器（Installing the Electrode Board Adapter）。使用时可以根据脑电帽电极连接线的颜色插入适配器上相应颜色孔中。

2. 为被试做好带脑电帽和记录脑电前的准备。具体的一些注意事项包括：1) 被试的头发必须是干的；2) 被试头发上的如发夹这样的饰物、帽子和耳环都必须在实验前摘除；3) 对于 ECI 脑电帽，实验前并不要求被试必须清洗头发；4) 在为被试戴脑电帽时，可请被试坐在一个椅背笔直的椅子上，在戴好脑电帽和测量过电阻抗后再请被试移到椅背有倾斜的椅子上或者躺到床上。

3. 为被试戴上身体绑带（Put the body harness on the patient）。可请被试伸直双臂和地面平行，将身体绑带绑在被试腋窝以下的胸前，使绑带上的按扣眼位于胸前中部，方便后面步骤的使用。身体绑带的松紧要求以被试感到舒服为准。

4. 测量被试头部周长（Measure the circumference of the head）以选择大小合适的脑电帽。可以利用和脑电帽系统配合的颜色皮尺或者普通皮尺测量，测量时将皮尺缠绕脑部一圈，缠绕时有两个参照点：鼻根点（nasion，即头盖骨中鼻骨和前额骨相连接的那点）向上 1 英寸的点和枕骨隆突（inion，即脑后枕骨外的突出部分）向上 1 英寸，这两个参照点决定了皮尺的测量范围。如果采用的脑电帽自带的有色彩的皮尺，可以根据实际测量结果看皮尺最后的测量点落于哪种颜色区域，就选择那种颜色的脑电帽，如果最后测量点落入两个颜色交界处，那么选择较大尺寸那个颜色的脑电帽；如采用普通皮尺，则可参考使用手册上给出的标准。

5. 佩戴耳部电极，并注入导电膏和使电极和皮肤紧密接触（Attach the ear electrodes: add gel and abrade the skin）。

6. 在脑电帽的 Fp1 和 Fp2 两个参照电极点套上一次性圆形海绵体（Place disposable sponge disks around Fp1 and Fp2）。Fp1 和 Fp2 是脑电帽位于前额的两个参照点。

7. 测量和标记 Fp 线（Measure and mark the Fp line）。在给被试戴脑电帽之前，测量从鼻根点到枕骨隆突的距离，即第 4 点所述测量头部周长，之后将测量结果的小数点向左移动一位，例如 38cm→3.8cm，之后由鼻根点向上测量头部周长左移一位小数点的距离，获得一个点，在前额处该点的平行位置取两点，使用可清洗笔标记，即为 Fp1 和 Fp2 点。

8. 根据以上测量结果为被试选择尺寸合适的脑电帽，将脑电帽前部的带有一次性海绵体的两点固定在前额已标记的 Fp1 和 Fp2 点，然后用流畅的动作向前往后拉脑电帽，帮助被试戴好脑电帽（Slip the proper size cap onto the patient's head）。

9. 将脑电帽绑带和身体绑带连接起来（Attach the cap straps to the body harness）。脑电帽绑带需紧扣身体绑带，否则可能导致后续脑电测量不准。

10. 将脑电帽的电极接线口和电极适配器的连接口相连（Connect the cap connector to the electrode board adapter）。

11. 用系统所带注射器为每个脑电帽上的电极内注入导电膏；并快速地在电极内前后摇动注射器针管（Fill each electrode cavity with electro-gel; rock the blunted needle/syringe rapidly back and forth）。注入导电膏时可用左手食指和中指夹住要注射电极，右手使用注射器注入导电膏；注射时可以轻微地提升注射器针管，使其不完全接触头皮，之后注入导电膏，注入量以导电膏刚刚溢出电极一点为准；之后保持针管在电极中，用右手前后快速地前后摇动注射器；当每个电极都注入导电膏后，用干净纱布擦去溢出的导电膏。

12. 检测电极阻抗，保证每个电极的电阻抗都等于或低于 3000 欧姆（Check the electrode impedance; equaling or being lower to 3K ohms are necessary）。必须保证每个电极的电阻抗都等于或低于 3000 欧姆，否则不能开始脑电记录。如果有电极的电阻抗高于 3000 欧姆，可采用以上 12 中所述方法在该电极内加入适量导电膏，再进行电阻抗测试；如电阻抗依然不符合要求，则可使用“快速插入电极（Quick Insert Electrode）”插入该电极孔内，并在外面贴上 Micropore 胶带固定该电极，之后需要在电极适配器（Electrode Board Adapter）上移除对应的脑电帽电极，而将快速插入电极的连接口连入电极适配器的该位置，注意记录使用快速插入电极的位置和过程，之后再次检测该电极的电阻抗；如果使用快速插入电极后电极电阻抗依然不符合要求，则需直接向 ECI 公司咨询。当所有电极的电阻抗测试均满足等于或低于 3000 欧姆，则被试可以进行脑电

测试。

13. 请被试移到有倾斜椅背的椅子上或者躺到床上，使他们的姿势更加舒适一些。（Move the patient to the bed or reclining chair for recording）。

14. 正式记录开始之前，在被试头部下、后脖颈下放置一块卷好的毛巾（Place a towel roll under the patient's head before starting the recording）。放置毛巾的缘故是为了防止脑电帽电极和床或椅子产生摩擦。放置毛巾的位置以被试感到舒适为佳，可根据被试需要灵活调整。

15. 在正式开始记录前需要调试每个电极是否通畅和是否正确地接入相应记录通道（Electro-Cap system check）。调试时按照电极顺序进行，主要方法是依次轻敲（tap）电极，观察记录显示屏上是否有相应的波动，如果有说明电极已通，可以进行测试下一个电极。如果没有波动显示或者记录通道错误，则需检查是否敲对电极和通道是否接入正确，如果问题不能解决，可联系 ECI 公司进行咨询。

16. 记录脑电实验结束后，为被试移除脑电帽、耳部电极和身体绑带；之后用干净毛巾擦除被试额头、耳朵和头发上的导电膏，经过这些处理后被试可以很容易地洗去头部的残留导电膏（Patient cleanup）。

17. 清洗脑电帽和对脑电帽的保养（Cap cleaning and preventive maintenance）。ECI 脑电帽必须在每次实验后进行清洗，否则会降低其使用寿命。清洗脑电帽时的注意事项包括以下一些：1) 清洗脑电帽时只能采用系统自带的脑电帽清洗溶剂（Ivory cleaning liquid）或者 Palmolive 清洗剂；2) 清洗时应该取下脑电帽与身体绑带连接的绑带，脑电帽绑带应单独清洗，由于其比较厚硬所以可以用刷子和肥皂水清洗，建议一周清洗一次；3) 不要将颜色不同的脑电帽混洗，在最初几次清洗时脑电帽可能会掉色；4) 清洗脑电帽时注意只将带有电极的脑电帽放入加入了专用清洗剂的水中清洗，而与适配器相连的电极连接头一定不能着水；5) 用清洗溶剂清洗后，用清水完整冲洗干净脑电帽，6）之后将脑电帽放置晾干，晾干时一定注意将脑电帽电极部分放置得低于与适配器相连的电极连接头部分，以防止后者着水；6) 每个月应该用棉签擦拭脑电帽电极防止氧化物在其中的逐渐堆积。

此外，实验中可能还会使用到“快速插入电极（Quick insert electrode）”，其使用方法可参见上文 13 中关于“检测电极电阻抗”的相关内容。

以上 17 点为使用实验室脑电帽系统时的基本步骤和注意事项，具体更为详细的内容可以参看纸本使用手册和观看 DVD 光碟的演示说明。

6 结语

脑电成分 ERP 波形是大脑对各种事物变化所引起的电活动，是人对事物认知，心理行为的一种客观的表现形式，是“观察脑功能的窗口”，在心理学、生理学、认知科学、神经科学、临床医学及其他生命科学相关领域具有很高的研究与应用价值。事件相关脑电位(ERP)采集和分析系统，凭借其精度高、抗干扰性强、操作简便、功能多样等优点，成为 ERP 分析的有效工具，是国际和国内科学研究和临床应用方面目前最为流行的研究技术之一；并已被应用到认知心理学、精神病学、运动医学、临床医学、人体工学等各个科学研究领域。

参考文献

[1] Jasper, H. (1958). The ten-twenty electrode system of the International Federation. Electroencephalography and clinical neurophysiology, 10, 371-375.

[2] Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. Science, 207(4427), 203-205.

[3] Li, W., & Yang, Y. Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. Neuroscience, In Press, Accepted Manuscript.

[4] Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. Acta Psychologica, 42(4), 313-329.

[5] Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature, 385, 432-434.

[6] Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. Journal of Memory and Language, 31(6), 785-806.

[7] Pulvermuller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The

mismatch negativity as a tool for studying higher cognitive processes. Progress in Neurobiology, 79(1), 49-71.

[8] Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat Neurosci, 2(2), 191-196.

[9] 魏景汉，罗跃嘉 (Ed.). (2002). 认知事件相关脑电位教程. 北京: 经济日报出版社

[10] 刘燕妮, & 舒华. (2003). ERP 与语言研究. 心理科学进展, 11(3), 296-302.

（杨若晓 北京大学中国语言文学系语言学实验室 100871 yangruoxiao@gmail.com）

录音室建设与数字录音设备

叶泽华

一. 录音室基本规格和检测参数

录音室的建设由北京科奥克声学技术有限公司完成。

1.1 录音室基本规格

北京大学中文系语音录音室建成于 2004 年，地点位于北京大学中文系五院一楼。主要用于语音科学研究的录音和录像。

录音室的外形尺寸为长 3.6m，宽 2.4m，高 2.68m，内部尺寸为长 3.4m，宽 2.2m，高 2.4m。录音室隔声门尺寸为高 2m，宽 0.9m。隔声窗尺寸为宽 0.9m，高 0.6m。

1.2 录音室组成

录音室是由 CA IA45 金属隔声-吸声模块组成，可拆卸后异地重新组装，而不会影响其声学特性。该金属模块一面为隔声、一面为吸声，兼有两种声学特性。模块厚度为 100mm，一面为钢板，一面为穿孔镀锌板，中间填充隔声、阻尼即吸声材料。

录音室由上述模块用专用连接件连接而成，包括四面墙及顶部，地面为专用的隔声地板，厚 100mm，两面均为钢板，内填隔声即阻尼材料，整个地面为声反射面。地板座在橡胶隔振器上，四座墙座在隔声地板上，顶座在四墙上，整个录音室形成一个与外围没有任何刚性连接的独立浮筑构造。

隔声门为 CA AD43 钢制隔声门，门体厚度为 64mm，密封方式为双道磁密封，采用旋升铰链，无门槛设计，门外表面做喷漆处理，内表面加 25mm 的吸声材料。

下面是隔声-吸声模块和隔声门的声学特性：

倍频程中心频率（Hz）	63	125	250	500
隔声量(dB)	27	30	32	41
吸声系数		0.94	1.19	1.11

CA IA45 金属隔声-吸声板声学特性

倍频程中心频率（Hz）	1k	2k	4k	8k
隔声量(dB)	50	59	67	71
吸声系数	1.06	1.03	1.03	1.04

CA IA45 金属隔声-吸声板声学特性

倍频程中心频率（Hz）	125	250	500	1k	2k
隔声量(dB)	30	39	39	45	44

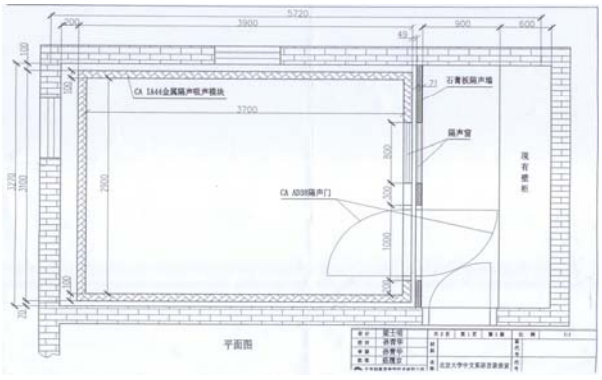
CA AD 钢制隔声门声学特性

下面是录音室基本构成单元的列表及面积：

序号	类型	面积
1	CA IA44 金属隔声吸声板	59.2
2	橡胶隔振垫层	12.1
3	CA AD 隔声门	2
4	隔声窗	
5	石膏板隔声墙	9
6	CA AD38 隔声门	2
7	地毯	12
8	送风消声器	2m 长
9	低噪音风机箱	1 台
10	回风消声器	2m 长
11	隔声窗	
12	室内普通照明	

北京大学中文系可拆卸式语言录音、录像室项目

下图是录音室的设计平面图纸：



录音室平面设计图纸

1.3 录音室声学检测

录音室声学检测工作由中国科学院声学计量测试站完成。

1.3.1 基础概念

（1）分贝计算

10*log(被测功率/参考功率)

20*log(被测声压/参考声压)

参考听觉分贝级：

150-160	靠近喷气式飞机
130-140	痛阈
120	机场跑道
120-110	雷声
100-90	地铁通道
90-80	重型车辆运输
80-70	一般工厂
70	繁华街道
60	一般谈话
50	一般办公室
40	低声谈话
30	安静办公室
10	寂静录音室
0	听阈

录音室声学检测采用的是 A 加权声级。A 声级是由声级计用 A 计权网络测得的声级。A 声级能够比较好的反映人耳对各种噪声的主观评价。下图是声级计权网络特性图，其中的 A 表示的就是 A 计权网络的特性：

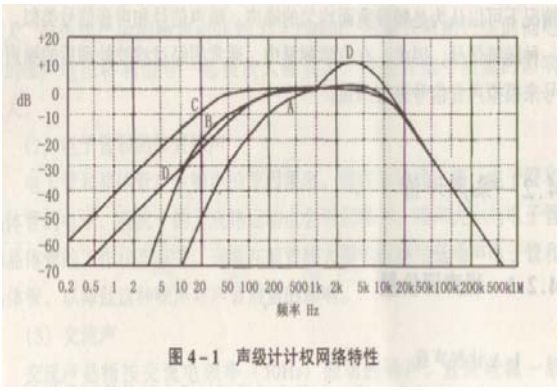


图 4－1 声级计计权网络特性

（2）混响时间概念及计算

声源在封闭空间不断辐射声能，停止发声后，室内声音并不会立刻消失，而是需要一个过程的。首先是直达声消失，反射声则将继续下去，每反射一次，声能被吸收一部分，室内声能密度将逐渐减弱，直至完全消失。这个逐渐衰减的过程就是“混响过程”。

混响时间可以赛宾公式计算得到。赛宾公式是 T60=K*V/A，其中 T60 表示混响时间，K 为常数（一般取 0.161），A 表示室内总吸声量，V 表示房间容积。实验室的声学测试报告中的中频混响时间就是用赛宾公式计算得到的。这个估算有可能不是很准确，实际上赛宾公式适用于室内平均吸声系数小于 0.2 的时候，而我们录音室的中频平均吸声系数是 0.8。用伊林公式进行计算可能得到更准确的结果。伊林公式是 T60=0.161*V/[-S*ln(1-a)+4mV]。其中 T60 表示混响时间。0.161 为常熟项。a 表示室内各表面的平均吸声系数。4m 表示的是和温度湿度相关的空气吸声系数。S 表示室内总的表面积。V 表示室内总体积。用伊林公式计算得到的录音室的混响时间结果是 0.448 秒，比用赛宾公式计算的结果小。

1.3.2 测试仪器及测试结论（检测报告内容）

测试仪器为 BK2033 精密声级计。

测试结论为录音录像室的本底噪声是 18dBA，估算中频混响时间为 0.09s，可保证在该录音室内录制的材料的质量。

1.3.3 具体测试内容即结果（检测报告内容）

（1）录音录像室内本底噪声：

送风机关、空调机关 18dBA

送风机开、空调机关 18dBA

送风机开、空调机开 19dBA

（一般地，录音室允许噪音建议值：20-25dBA。）

（2）录音录像室外本底噪声：

送风机关、空调机关：24-30dBA

送风机开、空调机关：47dBA

送风机开、空调机开：47dBA

（3）录音录像室的混响时间问题

录音录像室容积很小且做了强吸声处理，估计混响时间很短。无必要做混响时间的测量。但可根据它的容积、表面积和平均吸声系数，对混响时间进行估算，以判定混响是否对录音有影响。

录音录像室地面面积为 3.6*2.6m^2，高 2.2m，容积为 20.6m^3，表面积为 46.0 m^2；

中频吸声系数为 0.8。按赛宾公式计算算的时间为 0.09s。这么短的时间不会对录音产生影响。这从已有的录音中也得到了证明。

二. 数字录音设备简介

2.1 数字录音的理论基础

数字信号指的是时间和幅值都是离散值的信号，是相对于时间和幅值都是连续值的模拟信号而言的。我们实验室录音基本采用的数字录音，格式为.wav 格式。

时间和幅值都是离散值的信号能否还原为原始的连续信号，这是一个很重要的理论问题。奈奎斯特采样定理回答了这个问题。它奠定了数字通讯系统的理论基础，回答在什么样的条件下，可以无失真地从抽样信号中恢复出连续信号。采样定理告诉我们，当采样频率大于信号中最高频率的 2 倍时，则采样之后的数字信号完整地保留了原始信号中的信息，一般取 2.56-4 倍的信号最大频率。也就是说，如果你想还原的信号最高频率为 F_max 时，采样频率至少要为 2*F_max。应用到我们的语音数字录音上来，如果我们需要研究的是元音或者嗓音的性质的话，采样频率采用 11k 就可以了，因为元音的前三个共振峰信息基本都涵盖在 5k 频率段以下了。若是研究辅音的性质的话，采样频率一般采用 22k，因为辅音在高频仍具有比较高的能量，如汉语普通话中心频率最高的辅音是[s]，其中心频率 6k 左右。

2.2 录音室录音设备组成

北京大学语音实验室常用录音设备包括传声器话筒（麦克风），调音台，外置声卡，ibmR50 笔记本电脑。使用的录音软件是 Adobe 公司制作的软件 Adobe Audition。

设备基本连接是麦克风和 EGG 两路信号输入到调音台，再将信号接入外置声卡进行数字转换，最后接入笔记本电脑把数字信号采集下来，并存储文件。

2.3 录音室录音设备简介及使用

2.3.1 传声器性能简介及使用

北京大学中文系录音室采用的传声器是 Sony 公司生产的 Sony electret condenser microphone（驻极体电容传声器）。这种传声器的好处是由于驻极体材料的使用，使得传声器内部电路得到简化，使传声器小型化。

传声器的主要技术指标有输出阻抗、灵敏度、频率响应、瞬态响应、动态范围等。这里主要介绍一下录音室所用传声器的指向性、灵敏度和频率响应特性。

在介绍传声器指向性之前，我想先介绍一下语声指向性特性。语声的指向性指的是声源辐射的声压随方向变化的特性。一般而言，在 4000Hz 以下的频率，语声的指向性较弱，基本可以看成是全方位辐射。声强近似地与距离的平方成反比。在频率高于 4000Hz 时，语声表现出较强的指向性，在水平面左右偏离 40 度和在垂直面上下偏离 40 度的范围内，各频率的声压级基本上与轴向声压级相同。这个辐射范围内的频率特性与正前方轴向频率特性比较一致。在其他辐射角度语声的高频能量会急剧下降。所以在说话者背后听话，会由于高频能量的减弱而使清晰度降低。语声的这种属性可用指向性图表示。如下图所示：

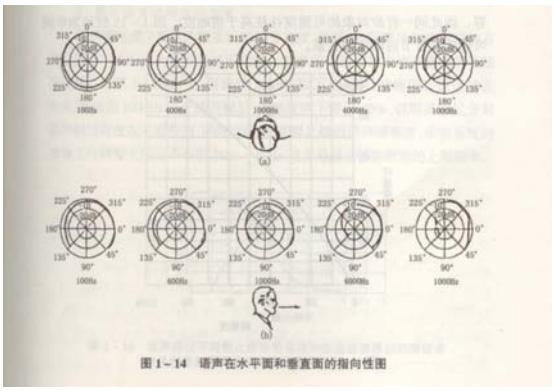
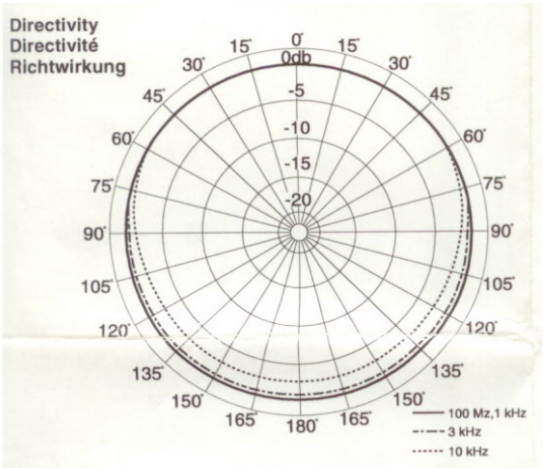


图 1－14 语声在水平面和垂直面的指向性图

语声指向性的极坐标表示

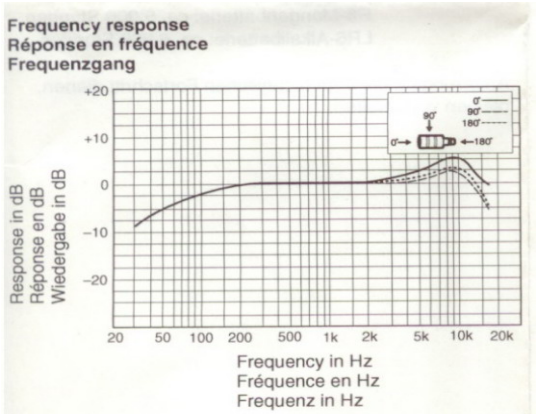
传声器的指向性是指传声器的灵敏度与声波入射角的关系，入射角是声波入射方向与传声器主轴的夹角。常见的传声器其指向性有全指向性、双指向性和单指向性。传声器的这种性质可以用极坐标形式表示。下图即是录音室所使用麦克风的指向性极坐标表示图，可以看出，它基本上是一个全指向性的传声器：



录音室麦克风指向性的极坐标表示

传声器的灵敏度指的是传声器的声电转换效率。它是指在自由声场中，当向传声器施加一个声压为 0.1pa 的声信号时，传声器的开路输出电压。灵敏度较高的传声器，可以在较低声压级的声音获得较高的信噪比。它有利于反映出声音的种种细节，但是也容易拾取到更多的噪声。

传声器的频率响应指的是传声器灵敏度随频率变化的特性。即对于不同频率的输入信号传声器输出电压的大小。其响应范围是指传声器正常工作的频带宽度，又叫带宽。传声器的频率响应与测量的角度有关。下图就是实验室所使用的麦克风的频率相应特性图：

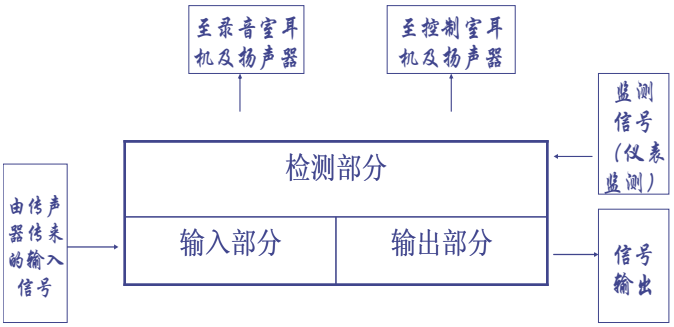


录音室麦克风的频率响应曲线

麦克风的使用比较简单。在实际的录音室录音过程中，一般我们都把麦克风夹在一侧的衣领上。

2.3.2 调音台简介

一般的调音台都是有三部分组成的，即输入部分、输出部分和监测部分。其基本结构可由下表说明：



调音台基本组成结构

调音台的功能有很多，我这里只简单列出在语音录音时我们使用最多的功能。

调音台可以实现录音通道的分配。一般我们录音时，都会一起录进 EGG 信号，即录入双通道信号。这时候，我们就需要调音台来实现通道的分配。一般把语音信号放在第一通道，EGG 信号放在第二通道。

调音台还可以用来对输入信号的大小进行调节。对于每一个通道的输入信号，调音台都有相应控制钮调节信号大小。录音时，我们可以需要观察录音软件中信号的大小，如果信号过载或者太小的话，我们可以可以在调音台上进行调节，使信号大小符合我们的要求。我们实验室常用调音台上面有一个控制钮可以控制所有通道信号的大小，另外，各通道分别也有控制钮来调节信号大小。

调音台还可以用来实现左右通道的平衡输入还有模拟信号的滤波以及分频处理。信号滤波可以实现对信号的低通、高通等效果。而分频处理则可以对信号的某个频段进行增强或者减弱等，已达到某种预期的效果。不过实际录音时，我们比较少用到这些功能，这里就不详细介绍了。

调音台还有一个很重要的功能就是对录入信号进行监听。这主要是通过在调音台的监听模块接入耳机或者扬声器来实现的。

2.3.3 外置声卡简介

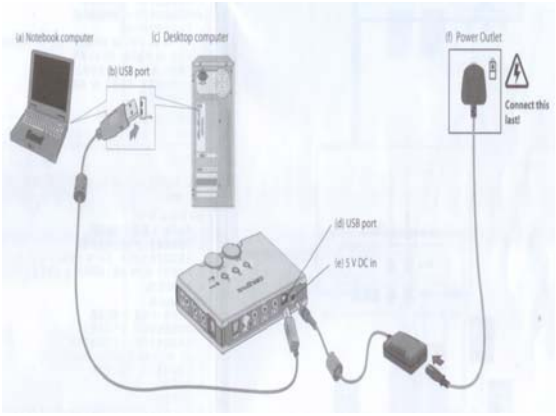
传声器和扬声器处理的都是电路模拟信号，而电脑所能处理的都是数字信号，两者不能混用，声卡的作用就是实现两者的转

换。从结构上分，声卡可分为模数转换电路和数模转换电路两部分，模数转换电路负责将传声器等声音输入设备采到的模拟声音信号转换为电脑能处理的数字信号；而数模转换电路负责将电脑使用的数字声音信号转换为扬声器等设备能使用的模拟信号。

所以实际上我们的数字录音就信号形式的转换而言经过了振动信号到电信号的转换（传声器），然后再是电信号到数字信号的转换（声卡），中间的调音台则起到对于电信号进行电信号处理的作用。

声卡发展至今，主要分为板卡式、集成式和外置式三种接口类型，分别是板卡式（独立显卡） 、集成式 、外置声卡式。外置式声卡是新加坡创新公司（creative）独家推出的一个新兴事物，它通过 USB 接口与 PC 连接，具有使用方便、便于移动等优势。这类产品主要应用于特殊环境，如连接笔记本实现更好的音质等。

外置声卡的连接比较简单，首先要接上电源。然后接入从调音台来的模拟信号，一边是输出信号输出到 PC 机。一般地，我们需要在是用外置声卡之前装上外置声卡的驱动，这样会有比较好的效果。下图是外置声卡的连接示意图：



外置声卡连接示意图

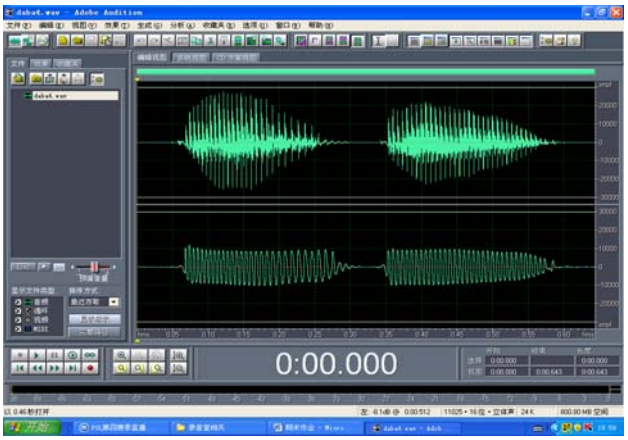
2.3.4 采集终端——PC 机和软件设置

录音室所使用的 PC 机是 IBM 公司生产的 R50 笔记本电脑。IBM 系列笔记本的性能是比较好的，这主要体现在 IBM 笔记本电脑的本底噪音即内部电路噪音比较小。

我们使用的信号采集软件是 Adobe 公司制作的 Adobe Audition 软件。打开软件以后，先新建文件，有三个选项，分别是采样速率、通道和解析度。采样速率的选择根据研究内容的不同可以选择不同，比较常用的采样速率为 11k（11025）和 22k（22050）。依据采样定理，可知采样率为 11k 和 22k 的信号可以还原的原始信号最高频率是 5.5k 和 11k。所有 11k 比较合适元音和嗓音研究，而 22k 比较合适辅音研究。没有特别要求的话，我们一般使用 22k 的采样速率。通道的选择一般都选立体声，即两个通道，第一通道录入声音信号，第二通道录入 EGG 信号。当然，也可以只录入声音信号，这样，就可以选择单通道了。解析度是有 8 位、16 位和 32 位三种，一般地，我们选择 16 位文件。这样，信号的幅值范围就是-32768—32767。

在录入信号之前，我们需要根据 Audition 软件上显示的信号波形情况，使用调音台对信号大小进行调节。首先需要注意的是不能让信号过载，不然就会出现削波的情况。经验上，我们一般都是用/da4/、/ba4/做测试文件。一般地，信号最大值为 2/3 量程比较合适。在 smpl 坐标上信号最大值不超过-20000 比较合适。信号的大小可以通过调音台来调节。

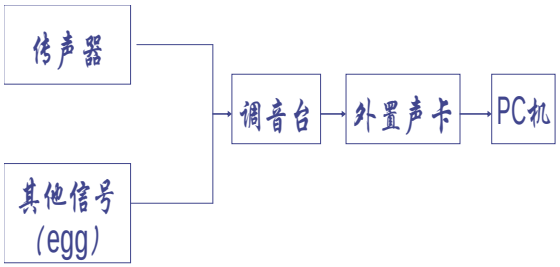
下图是用 Audition 软件录制双通道波形文件的一个示例：



双通道波形文件示例

2.3.5 数字录音设备连接示意图

最后，给出录音室常用录音设备连接顺序的示意图：



常用录音设备连接示意图

参考书目：

[1]《录音室手册》 [美]约翰·姆·白兰姆 著
中国电影出版社 1989
[2]《录音技术》
朱伟 编著 中国广播电视出版社 2003
[3]《录音手册》 [英] 约翰·奥尔德雷德
中国电影出版社 1980
[4]《录音播音建筑声学设计》
端瑞祈 著 北京大学出版社 1994
[5]《声音与人耳听觉》
陈小平 编著 中国广播电视出版社 2006
[6]《言语链》
[美]P·B·邓斯E·N·平森 著
中国社会科学出版社 1983
[7] 实验室麦克风、声卡说明书
[8] 百度百科

北京大学中文系 叶泽华
zehuay@gmail.com

Multi－Speech Model 3700 多功能语音分析系统简介

钱一凡

1. 概述

多功能语音分析系统（Multi－Speech Model 3700）是一套能够全方位分析语音的系统，包括语音分析、语音合成、实时声门阻抗分析、音域分析、多维嗓音分析等分析程序和视频语音学、病变嗓音和动态电子腭位等多个数据库。

2. 系统配置与安装共有以下 8 张光盘，安装时须按照由 1－8 的顺序安装。

1. Multi－speech（主程序）
2. MDVP（Multi－dimensional Voice Profile，多维语音分析软件）
3. Analysis Synthesis Laboratory（语音分析与合成实验室）
4. Real－time EGG Analysis（实时声门阻抗分析软件）
5. Voice Range Profile（音域分析软软件）
6. Palato meter Database（电子腭位分析软件和数据库）
7. Video phonetics（语音学视频数据库）
8. Voice Disorders Database（病变嗓音数据库）

3. 系统安装完成后，程序菜单如下面左图所示：

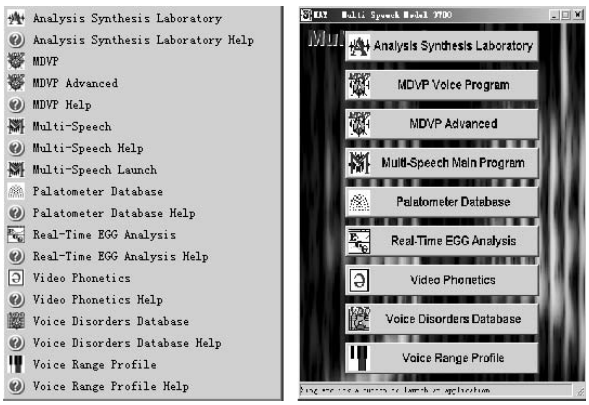


图1 程序菜单

点击“Multi－Speech Launch”可进入系统快捷菜单，将显示所有已经安装的程序，如上面右图所示

点击程序名称进入相应的程序，使用程序时，需要将密狗插入USB 口。

4. 系统功能简介

4.1 多功能语音分析（Multi－speech Main Program）

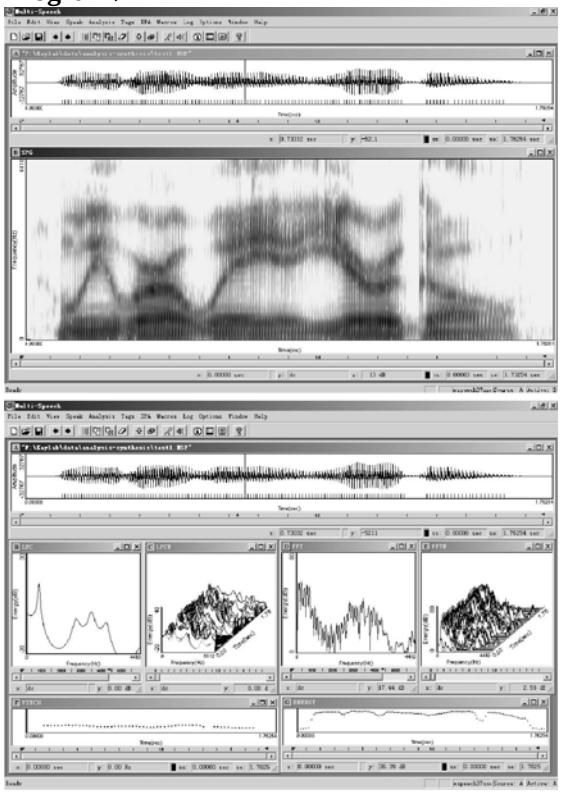


图2 主程序界面

主程序的分析功能包括：三维语图、功率谱、线性预测分析、共振峰提取、倒谱分析、基频频包络检测、能量包络分析等。

4.2 多维语音分析析软件

多维语音嗓音分析软件的主要功能是研究正常嗓音和病变嗓音的声学性质。能从语音信号中提取包括频率抖动、振幅抖动、基频、能量等32项声学参数。利用这些参数可以评价嗓音的特性，同时可以利用各参数的正常参考值对病变嗓音进行临床诊断。

4.3 语音分析与合成实验室

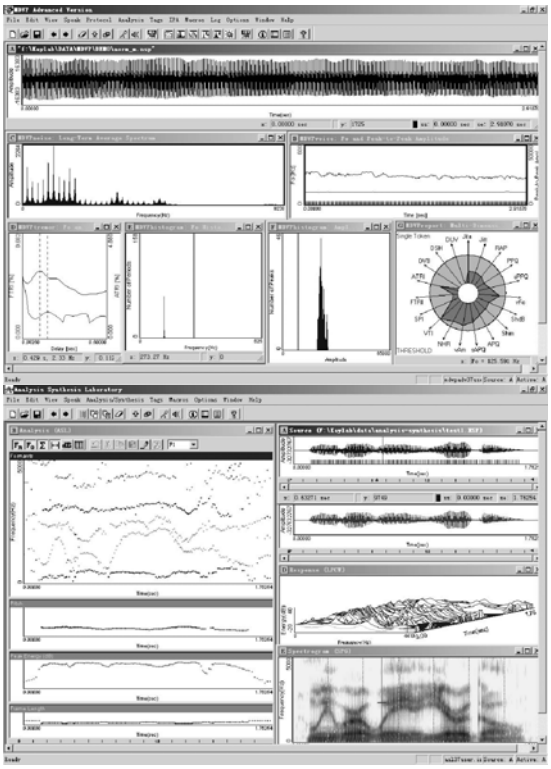


图3 语音分析与合成实验室界面

语音分析与合成实验室可以对语音进行声学分析，提取基频、能量、共振峰、带宽等声学参数，同时可以修改提出的声学参数，并进行语音合成，得到到新的样本。

4.4 实时声门阻抗分析软件

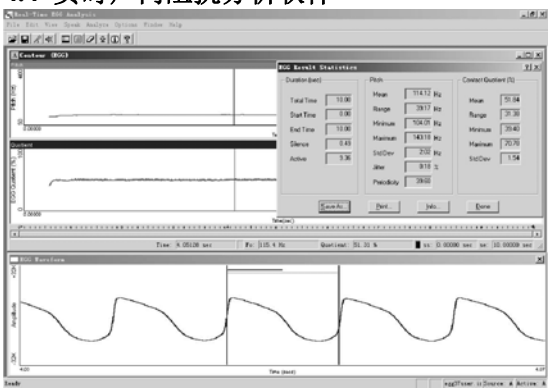


图4 实时声门阻抗分析界面

实时声门阻抗分析软件的主要功能是对声门阻抗信号（EGG）进行实时录音、分析和参数提取。提取的参数有基频、开商、速度商、接触商（可自定义）等。这些参数可以用于研究声带的振动情况，建立嗓音振动模型。

4.5 音域分析软件

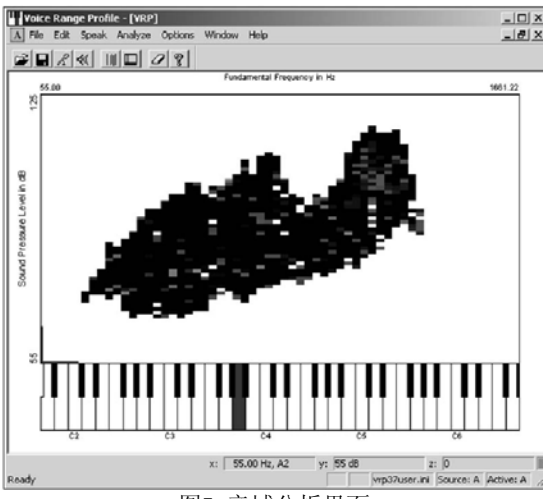


图5 音域分析界面

音域分析软件可以对发音人某个特定频率的相关能量进行分析，并根据全频率段的能量分布，判断发音人的音域。该软件不仅可以用于正常语音分析、声乐分析，还可以用于病变嗓音音的临床诊断。

4.6 电子腭位分析软件和数据库

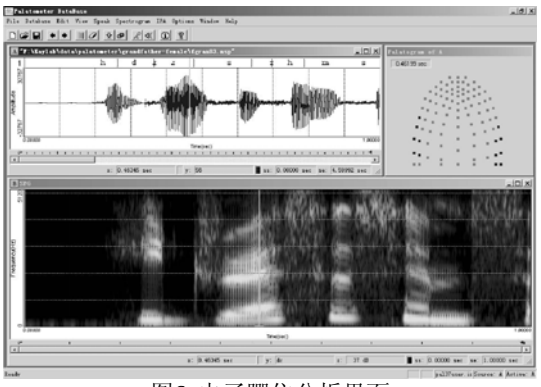


图6 电子腭位分析界面

电子腭位分析软件可用于分析动态电子腭位的参数与语音声学参数之间的关系。动态腭位数数据库中包括各种如IPA标准音素、单音节、多音节和语篇的语音、电子子腭位和共振峰数据。

4.7 语音学视频数据库

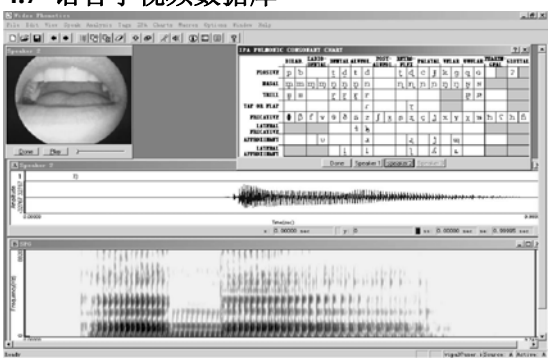


图7 语音学视频是数据库界面

语音学视频频数据库包括IPA标准元音音和辅音的发发音视频，共共有男女两位位发音人。

4.8 病变嗓音数据库

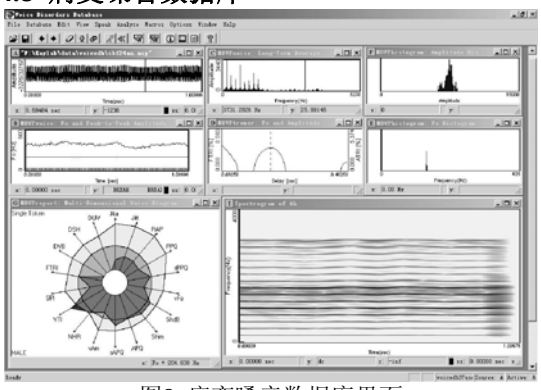


图8 病变嗓音数据库界面

病变嗓音数据库包括各种类型的临床病变嗓音，可以对病变嗓音进行声学分析，并分析声学学参数与病变类型的关系，为病变嗓音的临床诊断提供声学依据。

5．数据分析示例（以实时声门阻抗分析软件为例）

实时声门阻抗分析软件可以对EGG信号进行录音、计算和分析，可提取的参数包括基频、开商、速度商、接触商等。下面是对一个汉语单元音进行分析和参数提取的过程。

1. 选择菜单File - Open File and Analyze：打开一个语音文件，该语音文件的第二通道须为EGG信号。

2. 如下图所示，Contour(EGG)这个窗口下显示两条曲线，Pitch是基频曲线；Quotient默认是嗓音开商（open quotient）的曲线。

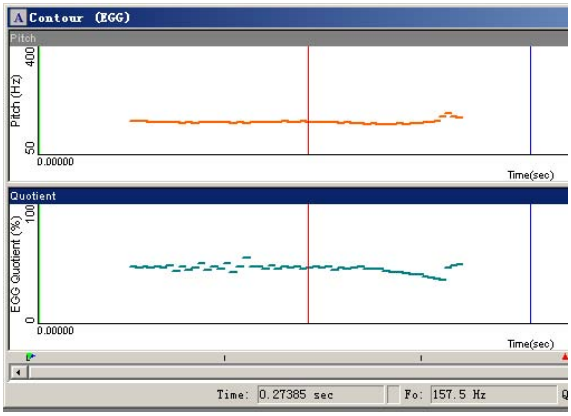


图 9 基频和接触商包络

可以在Options - Select Real - time EGG Options菜单栏中，自定义上图中Quotient窗口中的数值定义。系统默认的定义有两种：Contact Quotient（接触商）和Open Quotient（开商）；同时可以选取User defined来定义商的分子和分母。

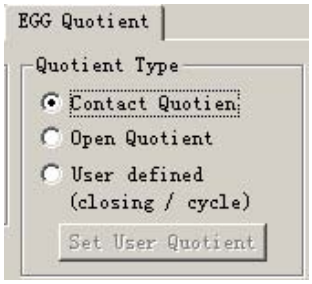


图10 自定义示例

3.选择菜单Analyze - Compute Result Statistics 可以显示各参数的统计数值。

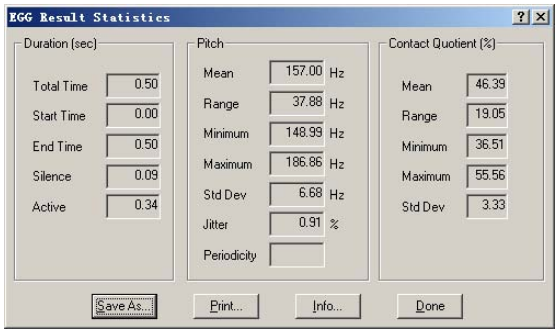


图11 输出的统计数值

在这个界面上，选择Save As..按钮可以将参数存成外部文件，方便后续分析使用。

（钱一凡 北京大学中文系 100871）