

# 普通话语音多模态研究与多媒体教学

## Multi-speech modality research in Putonghua and multimedia teaching

孔江平

香港大学香港普通话培训测试中心

北京大学中文系，汉语言研究中心

**摘要：**本文根据北京大学中文系对汉语普通话进行的语音多模态研究，讨论了汉语普通话语音在多媒体教学系统中的作用和实际应用的可能性。文章对香港人学习普通话的特点及多媒体教学系统的制作也进行了初步的介绍，并展望了汉语普通话多媒体教学的前景。

**Abstract:** This paper introduces the multi-speech modality research of Putonghua done in the department of Chinese language and literature, Peking University and discusses the role and possibility in establishing the multimedia teaching system of Putonghua. The paper also discusses the property in Putonghua learning by Cantonese and the multimedia teaching system which is being established in Hong Kong University. Finally the paper presents the prospect of Putonghua multimedia teaching.

**关键词** 普通话；语音多模态研究；多媒体教学

**Keywords:** Putonghua; multi-speech modality research; multimedia teaching

### 1. 引言

随着中国力的增强和文化科技水平的提高，中国无论在经济方面还是在文化方面和国际的交流都在迅速发展和扩大，这致使汉语普通话的使用范围也迅速扩大。汉语是联合国的工作语言，在国际政治经济活动中有着重要的地位，汉语也承载着中国五千年的文明所创造的灿烂文化。普通话是汉语口语的标准，因此，对汉语普通话的学习必然会成为世界各国语言教育中的一个重要的方面。

语言的学习与其他科学文化知识的学习不同，有其自身的特点，这些特点对语言教学的十分重要，因为它直接影响到语言学习的成效。众所周知，人们在习得母语的过程中，没有人会感到是一种负担

或者十分辛苦，无论学习者的智商如何，都能在母语的环境中自然掌握一种语言。然而，在第二语言或方言的学习方面，如果没有语言环境，学习过程都会感到十分辛苦，语言教学的这种特点只从经验很难解释。

在以往的教育中，对教学法的理解更多的偏重于经验，甚至认为是一种艺术，这些理念在教育中也确实起到了一定的作用。然而，随着科学的发展，特别是人类行为的科学研究和人类大脑科学研究的进展，越来越多的人已开始认识到教育的科学性质。同时，言语科学的发展，特别是言语脑科学的发展，也使人们看到了语言教学理论的方向和科学的曙光。

在语言科学研究中，语音学的地位十分重要和直接，因为它主要研究语言的物理特性和语言学意义之间的关系。语音学分为传统语音学和语音科学，虽然语音学已经完成了从传统到科学的转化，然而，传统的语音学然在语言学研究中广泛使用，只是研究的领域不同。本文将从语音科学的角度，探讨一下汉语普通话语音多模态研究和普通话语音教学的关系以及建立普通话语音教学多媒体系统科学基础。

### 2. 语音科学研究与教学

汉语普通话语音多模态研究主要涉及到语音声学、X光动态发音动作研究、核磁共振声道研究、电子腭位发音动作研究、嗓音发声类型研究、唇形唇读研究、动态声门研究及呼吸带语音韵律研究几个方面，下面将从这几个方面讨论一下这些基础科学研究和普通话教学的关系以及应用的可能性。

#### 2.1. X光动态发音动作研究与教学

在传统语音学研究中，X光的使用使语音学完

成了从传统到科学的第一次飞跃。因为在此之前，人们是靠观察和自我感觉确定发音的部位和动作，有了 X 光以后，人们可以从科学实证的角度研究发音器官、发音动作和音义的关系。在教学方面，由于 X 光的使用，人们更精确地了解了语音发音器官的运动方式（汪高武，鲍怀翘，孔江平，2008），大大地提高了语音教学的科学性，特别是推动了语音教学的方法和理论的产生，从此语音学理论进入了普通的语言教学。国内这方面的成果有《汉语普通话发音图谱》（周殿福，吴宗济，1963）、图谱教材等，都属于这一类。在我国，汉语普通话 X 光的动态资料只有鲍怀翘教授在文革后拍摄了 4 位发音人的资料，由于健康保证方面的原因，根据普通话声韵调的组合关系和轻声儿化等语音现象，每人拍摄了 200 多样本，其中大部分是单音节样本。由于信号处理技术条件的限制，直到最近这些样本才被北京大语言学实验室处理为可供研究和开发教学软件的数字化数据库。

## 2.2. 语音声学分析与教学

在二次大战后期，出现了频谱分析技术，这种技术可以实时实现语音信号时域到频率域的转换，对于语音学研究来说可谓又一次技术上的飞跃。这使人们对语音的性质有了质的认识，大大推动了语音学和语音教学的发展。在 19 世纪贝尔实验室出版了《可视语言》（Ralph K. Potter, 1947）一书，此书的出现极大地推动了声学语音学的发展，很快人们发现了语音声学特征和生理特征的关系，从而建立起了语音声学和生理学的基本关系以及基础理论。将语音变换成图像可以使听力有障碍者通过视觉读懂语音，为残疾人的语言教学提供了新的方法。在正常人语言的学习上，可以通过分析学习者的发音，将其转换为可以理解的发音形式，如，声调曲线可以用于教学中声调偏误纠正，通过提取共振峰画在元音图上，帮助纠正元音的学习等等（吴宗济 林茂灿主编，1989）。由于电脑速度和信号处理技术的进步，现在人们正在研究语音到生理发音器官动作的反推问题，这一问题如果能被彻底解决，自我纠正发音将会变得很容易，这将彻底改变语音学习和教学的基本模式。

## 2.3. 核磁共振声道研究与教学

核磁共振用于人体透视检查是近十年的事，一般认为少量拍摄对人体无害，主要用于医疗检测，核磁共振的原理与 X 光不同，对于语音学研究来说，

发音器官的成像更为清晰，而且可以拍摄动态二维、静态三维和动态 2.5 维的图像，非常适合语音学的研究。目前，语音学、言语生理学和言语工程都在利用核磁共振技术进行言语科学最前沿的研究。北京大学中文系语言学实验室近几年在日本国家实验室(ATR)拍摄了 9 个人普通话基本发音动作的资料，并且已经建立起了一个全数字普通话发音数据库。基于核磁共振的普通话声道和发音动作的研究使我们对汉语普通话有了更深入的认识，目前正在努力进行汉语普通话三维立体的数字模型研究。在国际上，语音教学领域一般认为最高水平的教学系统应该是三维立体的数字发音模型，目前这样的商用教学系统英语也还没有做出来，只有瑞典语的一个博士生做了一个演示系统。我们目前正在研究制作基于二维的核磁共振普通话教学系统，这种系统可以让学习者直接看到发音部位和发音动作，提高普通话语音的学习和教学效率，而且可以帮助聋儿童学习普通话的基本发音。

## 2.4. 电子腭位发音动作研究与教学

动态电子腭位技术是将一个假腭放在口腔中，在发音过程中，舌头和上颚的接触可以实时地被记录下来。通过对这些数据的分析，可以研究普通话的发音动作，特别是普通话的协同发音。动态电子腭位最初的目的是为了腭裂儿童术后学习发音而设计的，这是因为腭裂儿童常常使用代偿性发音，在手术后仍然不会正确的发音，电子腭位可以用于纠正腭裂儿童的发音动作，主要是辅音的发音动作。可以看出，电子腭位主要用于研究辅音的发音动作和辅音的协同发音，在普通话的教学显然有重要的应用前景。我们这几年建立了一个很大的普通话电子腭位数据库，正在做基础研究和腭裂儿童的语言康复的应用研究，如果要应用在普通话的教学方面，除了基础研究以外，其瓶颈主要在技术的应用转化和成本方面，但理论上已经成熟。

## 2.5. 唇形研究与教学

从语言交际的角度看，唇形和语言的学习及教学有着密切的关系，语言学习过程中的模仿始终都离不开唇形的信息。在聋哑儿童的唇读学习中，唇形的作用更是必不可少。从语言学的角度，唇形的研究主要注重唇位的研究，即唇形的语言学意义。在言语工程研究上，往往用图位来定义唇型，更注重唇形产生的模型和模拟。这两方面的成果为语言的教学提供了可能和坚实的基础。在唇位研究方面，

主要可以从视频中检测出唇形的基本信息和动态信息进行研究。近几年,由于技术的发展,唇型检测又有了专门的设备红外三维立体唇型信号采集系统。目前我们建立了一个基于红外三维立体信号的汉语普通话唇型基础研究和教学应用研究的大型数据库,希望建立一个用于汉语普通话教学系统的基础模型,基于这个模型,开发具有唇型视频的普通话语音多媒体语音教学系统。

## 2.6. 动态声门研究与教学

动态声门研究是基于高速数字成像技术对声带振动频率和振动方式的研究,具体是利用高速摄像机(4000 帧/秒)拍下声带的振动过程,然后用图像处理技术检测出相关参数对声带的语言学特性进行研究(Kong J.P., 2004, 2007)。如对汉语普通话声带振动的研究表明,汉语普通话的上声的低音部分是一种挤喉音的发声类型,这说明普通话上声不仅靠相对低的音调来体现声调的语言学意义,同时,特殊的发声类型是学习普通话上声的关键。许多外国人只是从音调或旋律上来模仿上声,没有意识到发声类型的不同,因此学习效果很差。在对汉语普通话发声类型的研究中发现,上声的发声在生理和声学上有五六种类型和表现形式。近期的研究也证明了汉语普通话发声类型对声调的感知有一定的贡献,特别在自然度方面更是如此。动态声门要真正用于普通话的教学就要进一步作发声类型视觉反馈的研究,即从语音中算出声带振动的方式,给学习者声带发声的视觉反馈,这样就能更有效地通过教学系统学习。

## 2.7. 呼吸带语音韵律研究与教学

汉语普通话的语音率是比较复杂的,这是因为普通话除了基本的语调外还有声调,这两种语音学现象在物理上主要通过基频变化来实现,因此这对普通话韵律的研究带来了很大的困难,以往的研究主要是通过语音来提取基频、时长和振幅等参数,这些参数对韵律和声调的性质不能进行充分的解释,因而长期制约了普通话韵律的研究。通过呼吸带同时采集语音和呼吸信号,可以很好地研究韵律以及韵律和声调的关系。目前我们已经建立了一个大型的汉语普通话呼吸-语音数据库,研究表明,呼吸在律诗、散文、新闻等不同文体中都有不同的表现和固定的模式(谭晶晶等, 2008)。而且,通过呼吸和语音声学的研究发现,呼吸信号更能表现出朗读的个人风格 and 特点。这就为汉语普通话学习系统

提供了更多的可应用功能。可以看出,普通话呼吸的研究可以为普通话朗读教学和普通话艺术教学提供广阔的应用前景。

## 2.8. 语音感知范畴研究与教学

我们都知道语音学研究对语音的教学很重要,但目前在教学中运用的语音学知识实际上是很片面和很陈旧的。如在普通话的语音教学中,通常只用到发音部位和发音方法一些基本的语音学知识,而且有些还不准确,这常常体现为一个很普通的语音学问题会有很多人在那里反复争论,不断强调自己的感觉,而不是去用科学的方法证明。然而,普通话语音研究已经从声学 and 生理开始进入语音感知领域,恐怕很少有普通话的老师了解普通话各个音位之间感知关系。如,在教“上声”发音时,只注重音调频率的变化,而对“上声”发声类型不了解,然而上声的感知和发声类型密切相关。最近的普通话声调感知研究表明,发声类型对汉语声调的感知范畴和自然度都有较大的贡献。因此,在普通话的教学中,将新的语言学知识特别是范畴感知和连续感知等基础知识引入教学是十分必要的。

## 2.9. 语音识别与测试

语音识别可以分为基础的研究和工程应用两种。基础的究主要注重声学语音学的研究,即研究语音的基本参数和语言学意义的关系,进而用于语音的判断和识别。而工程应用的主要在语音识别系统的技术实现上,在声学层面一般只用美倒谱系数,在识别层面主要用统计模型-隐马尔科夫模型和某种定义的语言学模型。在中国做基础研究识别的人很少,国家的投入和成果都较少,而做工程语音识别的人很多,国家投入很大,目前也有很多语音识别系统。在语音教学和测试方面,两种语音识别都能有广泛的应用前景。基础的识别研究元音共振峰参数和普通话音位系统以及韵母的关系、基频和声调语调的关系、辅音声学参数和辅音音位等,最终希望能从语音的声学参数识别出具体的语音音位,这些研究的成果在教学方面可以应用在语音声学视觉反推上,甚至能应用在语音生理的视觉法推上,如发一个元音就能在一个坐标上画出这个元音的相对位置,或者画出这个元音的声道形状,这样就可以在学习发音时进行自我矫正,从而达到高效自我学习的目的。聋哑人和言语障碍者也可以通过这种系统进行语言学习和语言矫正,其应用前景极其广阔。用于工程的识别在自然语言的识别上已经有了

很大进展，某些领域已尝试应用，但离实际的应用还有不少差距。在语音的测试方面，由于这种识别主要是统计模型，不太注重语音的语言学意义，因而在测试的应用上也还有一很大的困难。从目前的研究水平看，建立一个单音节的汉语普通话语音测试系统因该是可以做到的，这主要取决怎样建立训练识别系统的语音数据库和建立怎样的语音学模型。

### 3. 香港普通话语音多媒体教学

在第二语言或方言的学习中，语音的学习有着自身的特点，因为它涉及到两种语言音位感知系统的冲突。如果母语和所学语言或方言的音位感知系统冲突很大，学习困难就会很大，也会导致学习的效果很差，这和人的智力没有关系。例如，印度人学英语很难学会送气音，日本人学英语很难学会弱化音节，中国人学英语很难分辨清浊辅音，这些都是因为音位系统的冲突造成的。在方言的语音习得过程中，常常会受语音历史音变规律的影响，香港人在学习普通话过程中，最大的障碍就是会受到语音历史音变的影响，从而导致语音类推的失误。众所周知，汉语从中古音到各方言有一定的演变规律，可以进行语音的类推，但这种规律复杂到一定程度，类推就会常常出错。因此，在制作香港人学习普通话的教学系统时，一定要理论上考虑汉语历史音变的规律和音变规律在学习过程中的误用。

在对语言接触和语言融合的研究中发现，两种语言感知的冲突在自然语言接触和融合过程中是有规律的（陈保亚，1996）。这种规律反映了语言感知系统对冲突的自然处理方式，这种处理包括生理的、语言学的和大脑认知的，对其研究会为我们编写普通话语音教材和教学提供依据。在目前的汉语普通话对外教学和方言的教学中，人们都在研究偏误，无论经验的、声学的和生理的都对普通话的教学起了积极的作用。然而从语言学理论的角度看，偏误的内涵在大脑音位系统的冲突所反映出来的语音感知结构的差异，这方面的研究需要脑电和功能性核磁共振等高端的仪器和设备，目前国内研究的还不多，正在起步阶段。对于香港普通话语音多媒体教学系统的研制来说，首先要搞清楚香港话和普通话在语音感知体系上的差别，才能研制出适合香港人学习的汉语普通话语音多媒体教学系统。

### 4. 普通话多媒体教学的展望

任何一种语言的学习都需要进行大量的操练，这是语音学习的特点。一般情况下这些操练都是在课堂上进行的，然而，课堂操练的时间是相当有限的，这就形成了语言学习和教学上的一个矛盾。从现在的语音学研究看，语音多媒体教学系统可以较好地解决这个矛盾，因为，多媒体学习系统可以不受时间和地点的限制，同时能具有音频、视频甚至判错等功能，这就为语音教学的自我学习创造了条件，使多媒体教学具有很大的优势。

在国际上，各国研究和制作语音多媒体的教学软件已经成为语音学科学研究的一个重要的方面，而且已经有了较好的系统。在国内，基于生理和物理的基础研究也已经逐渐展开，我国的语音学研究和言语科学技术也已经十分先进，但在语音大脑认知方面的研究才刚刚起步，这就阻碍了高端多媒体系统的研发。另一方面，在语音多媒体教学系统的研制方面，大家的认识有很大的差别。目前国际国内都有两种方式，一种是从语音的发音发声原理和语言学理论方面来进行语言多媒体系统的研究和制作。另一种是从技术的角度研究和制作教学系统。香港普通话培训测试中心和北京大学中文系正在联合研制的属于第一种方法，试图将基础研究应用于教学系统的制作。

随着中国国力的不断提高，汉语普通话在国际政治、经济和日常生活中的地位会不断提高，汉语普通话的教学势必越来越重要。然而，语言是有民族性的，这就要求我们不仅要要对汉语进行大量深入的语音学和语言学的研究，而且也要对汉语普通话的教学进行大量的研究。为了能使汉语普通话走向世界，研究制作汉语普通话的语音多媒体教学系统是将汉语普通话推向世界的重要一环。随着汉语语音学、生理语音学、声学语音学、心里语音学、语音认知研究以及言语科学技术的发展，汉语普通话的多媒体教学系统也将会不断完善。语音脑科学的研究使我们正在不断接近普通话音位感知系统的本质，这将会对普通话的教学方式产生根本的改变。然而，我们也十分清楚，在普通话的教学中，我们只知道汉语普通话语音的认知体系是完全不够的，我们还必须研究学习者母语的语音认知体系以及生理和物理的特性，只有这样才能研发出真对世界不同语言的汉语普通话语音多媒体教学系统和汉语普通话水平测试系统。

## 5. 参考文献

1. Ralph K. Potter, George A. Kopp, Harriet C. Green, 1947, Visible Speech, D. Van Nostrand Co., New York
2. JP Kong (2004), Phonation Patterns of Tone and Diatone in Mandarin, From Traditional Phonology to Modern Speech Processing, Foreign Language Teaching and Research Press, edited by G. Fant et al, ISBN 7-5600-4075-6.
3. JP Kong (2004), Phonation Patterns of Tone and Diatone in Mandarin, From Traditional Phonology to Modern Speech Processing, Foreign Language Teaching & Research Press, edited by G. Fant et al.
4. JP Kong (2007), Laryngeal Dynamics and Physiological Model, Publishing house of Peking University
5. Gaowu WANG, Tatsuya KITAMURA, Xugang LU, Jianwu DANG, Jiangping KONG, 2008, MRI-based Study on Morphological and Acoustic Properties of Mandarin Sustained Vowels, Signal Processing
6. 周殿福, 吴宗济, 1963, 《普通话发音图谱》, 商务印书馆
7. 吴宗济 林茂灿主编, 1989, 《实验语音学概要》, 高等教育出版社, 北京
8. 理论陈保亚, 1996, 《论语言接触与语言联盟》, 语文出版社
9. 汪高武, 鲍怀翘, 孔江平 (2008), 从声道截面积推导普通话元音共振峰, 中国语音学报, 第一辑, 商务印书馆
10. 谭晶晶, 李永宏, 孔江平, 2008, 不同文体朗读时的呼吸重置特点, 清华大学学报, 第四期, 自然科学版