

汉语方言的实验语音学研究旨趣¹

胡方

中国社会科学院语言研究所

hufang@cass.org.cn

语言学属于人文科学 (Humanities) 范畴, 其研究旨趣以描写语言事实并进行总结见长, 是一种归纳推理导向的研究范式 (a generalization-based approach); 实验语音学, 顾名思义, 属于实验科学范畴, 其研究旨趣在于以生理、物理、心理实验为基础对语言中的语音现象进行描写或者解释, 是一种证据导向的研究范式 (an evidence-based approach)。无论采用哪种研究范式, 任何科学研究的核心任务都是为人类提供新知识, 也就是创新。那么, 在汉语方言的研究这个问题上, 实验语音学可以为方言语音研究提供什么样的新知识? 换句话说, 研究方言语音为什么需要实验语音学呢?

汉语方言的语音研究是在高本汉 (1915-1926)、赵元任 (1928) 开创的传统的基础上发展起来的。这个传统, 简要地说, 就是描写主义的传统, 科学与人文并重。研究者主要基于口耳之学的听音辨音, 记录汉语方言材料, 整理方言音系, 并结合汉语历史音韵, 厘清汉语方言的发展关系, 地理分布特点。这一传统的研究范式为我们了解汉语方言的基本面貌和特点提供了宝贵的第一手的资料, 新中国成立以来, 尤其是在改革开放之后取得了丰硕的研究成果, 比如:

(1) 北京大学中国语言文学系语言学教研室编的《汉语方音字汇》, (2) 中国社会科学院语言研究所李荣主编的 41 本全国各地方言词典, (3) 中国社会科学院语言研究所李荣、熊正辉、张振兴等主持的中国语言地图项目, (4) 北京语言大学曹志耘主编的《汉语方言地图集》等等。

语言学领域的实验语音学研究也有源远流长的口耳之学的传统, 但自 20 世纪以来, 其学科发展渐渐走上了一条实验科学的道路, 而且越来越呈现出跨学科的特点。不用说语音工程、人工智能等领域的研究, 就是偏语言学本身的语音学研究, 也普遍遵循实验科学的一般范式, 对语音进行采样测量, 对数据进行统计分析, 进而对语音现象进行描写与解释。实验科学在传统语音研究领域的发展为我们观察人类语言中的语音现象提供了新的理论视角与方法工具。

那么, 实验语音学究竟提供了怎么样的新方法呢? 首当其冲的便是语音数据的采样, 尤其是作为语言物理外壳的语音音频数据的采集, 现代科技的发展为我们提供了极大的方便, 以前需要专业录音设备才能完成的任务, 现在可以在便携电子产品上轻松实现。同时, 新技术不仅仅提供新的方法论, 而且, 在语言研究的哲学层面带来基础理论的变革。从结构主义以来的语言学传统区分“语言” (language) 与“言语” (speech) 这一对概念, 定义前者是抽象的语言能力 (即生成学派术语中的 linguistic competence), 后者则是具体的产出 (即生成学派术语中的 linguistic performance); 语言学研究的是“语言”, 而不是“言语”。吊诡的是, 语言学研究的材料都是“言语”; 可是, 从“言语”研究“语言”, 在语言学的研究中设定如是, 是不需要进行论证的。比如说, 作为语言基本单位的音位 (phoneme) 的概念, 语言学直接就认为人脑中有个抽象的“音位”或更抽象的一组“区别特征” (distinctive features), 而同时, “音位”的物理产出物“音素” (phone) 则只是一些对语言学核心任务来说不大重要的“变异” (variations)。更为关键的是, 在学科立论 (argumentation) 上, 也不需要从具体的言语产出 (speech production) 去对语言学上的音类 (即人类储存在大脑中的语音类别) 进行论证。而实验语音学则认为这是需要论证的, 抽象的音类是建立在对采样数据的科学论证基础上的; 因此, 丰富的实验语音细节 (fine-grained phonetic details) 改变我们的语音观、语言观。

因此, 实验语音学为方言语音研究带来的创新之处, 从表面上看, 最显著的就是对数据的

¹ 载《方言》2018年第4期, 385-400页。

测量、统计了。那么，我们的任务就是做测量，做统计，通过数据的可视化图解语音范畴或者语音现象么？比如经常见到这样的一类研究，甲方言声调的实验语音学研究，测量了若干个人的声调的基频曲线，通过某种采样、平均、归一的方法画出来，据说这样便和声调的五度值（赵元任，1930）对应了；然后便是结论，讨论一下和传统的口耳之学的方言学记音有什么异同。这是非常危险的研究，因为语音的声学参量与语言学意义上的语音范畴之间并不是一种简单对应的关系。更有一些复杂点的设计，增加发音人中男女、年龄、或者其他社会属性的不同，再复杂一些的，增加乙方言的类似情况，然后就开始谈语音的性别差异、年龄差异，甚至语音演变等，那就更加危险了。语音的物理参量的采样、测量、统计本身并没有错，数据可视化也是实验语音学研究中常见的手段；但实验语音学并不是为了测量而测量，为了统计而统计，实验语音学遵循实验科学的一般方法，最为关键的是，实验语音学的测量与统计必须有语言学意义上的考量。在下文第一节，我们结合研究实例来谈这个问题。

关于实验语音学对方言语音研究的用处，一个广泛被接受的说法是“补口耳之缺”；这是从方言学研究（*dialectology*）的角度说的，或者更宽泛一些，是从一般的语言学的立场说的，即前文所说的通过言语来研究语言而不需要进行论证的语言观。从事方言调查研究工作的学者对于实验语音学最大的需求可能就是这个“补口耳之缺”了，因为即便是调查经验再丰富的学者，终也会遇见自己不熟悉、很难记的音，或者不确定某个音的音值，这个时候，就非常需要借助实验语音学来确定或者证明一下这个音。非常遗憾，本文要说明的是，实验语音学往往不能直接回答这个问题，因为我们的语言观、语音观是不一样的：语音的本质不是抽象的音位、音位变体或者区别特征可以完全概括的，语音是具体的、变异的。本文的第一节具体谈语音观的问题，第三节结合例子专门说明语音的变异性。“补口耳之缺”这个说法的另一个问题是这句话的预设，即认为常见的音类是不需要实验语音学的，只有僻见的语音才有必要进行实验研究。本文的第二节重点谈这个问题，实验语音学并非只是研究一些僻见的语音现象，而是对一切语音现象系统、全面的研究，并通过举例说明，即使并不复杂的西南官话的元音，通过基本的声学实验分析，也可以丰富汉语元音研究的视野。

实验语音学对于语音研究的创新就在于直面通过言语（*speech*）来研究语言（*language*）这一语言学的二元论问题。在一般语言学的理论框架中，作为语言基本单位的音位及其区别性主要是从心理的角度，通过思辨推理来定义的，即音类在语言或大脑中的表征（*representation*）是通过逻辑推理，而不是实际的发音生理、感知心理实验来论证的。在这个框架下，语言具有自主性（*autonomy*），有一个著名的比喻，就像一副象棋，缺个子儿，无论是“车”“马”还是“将”“相”，拿块石头代替即可。也就是说，重要的是系统与功能，具体语言单位的确切物理值对语言系统来说一点儿都不重要。在这个语言观之下，实验语音学研究除了验证音类的音值之外，其实并没有太多其他的用武之地，而且，验证音值本身也并不是语言研究的重要核心问题。有意思的是，持这个语言观的学者往往觉得用实验的方法验证音值并不是件难事，因为他们往往认为语音区别与声学参数之间存在着简单对应关系。但事实上，在实验语音学的研究范式中，音位，或者更宽泛一点的音类，与研究中容易采样测量到的语音声学之间，并不存在着直接对应关系，因为二者之间隔着语音产生（*speech production*）与语音感知（*speech perception*）两大科学问题。而且，近些年的实验研究表明：大脑中处理语音产生与感知是由不同的区域与机理控制的（*Bouchard et al., 2013; Mesgarani et al., 2014*）。事实上，在实验语音学领域，一直不断地有理论试图建立语音产生与感知之间的关系，阐释二者之间的相关性甚至一致性，比如语音感知的运动神经理论（*The motor theory of speech perception*，参见：*Liberman 等, 1967; Liberman & Marttingly, 1985, 1989; Liberman & Whalen, 2000; Galantucci 等, 2006*）、语音产生的量子理论（*The quantal theory of speech production*，参见：*Stevens, 1972, 1989*）等，但是，语音产生与感知毕竟是由独立机制控制的，而二者之间的不一致性可能就是造成世界语言中的语音多样性的原因之一，比如 *Ohala* 就试图从感知误差去解释语言中的语音演变，

认为这是语音演变的主要原因 (Ohala, 1981, 1993)。

实验语音学的研究总体上就是用实验科学的方法描写人类语言中的语音现象, 作为一个多学科共同关心的研究领域, 其研究旨趣可以有不同的取向。比如可以是偏描写性的 (descriptive approach), 描写世界语言中的语音多样性与普遍性; 又比如可以是偏实验性的 (experimental approach), 也就是把言语作为实验材料, 测试大脑中与语音相关的语言能力。语音产生与语音感知是人类大脑的固有功能, 在大脑中由相应的模块与机制控制, 受语音产生与感知机制控制的音类在具体的语言中如何组织成系统, 成为语法的一部分, 便是语言的音系 (Phonology) 了。偏描写性的实验语音学的主要任务就是描写这个“从大脑至语言/言语”的过程; 而偏实验性的实验语音学则有“从言语/语言至大脑”的意味, 旨趣上更偏向解释性。

本文主要以汉语方言及相关少数民族语言中的语音现象为实例谈语音学, 即在研究对象上主要谈汉语方言及相关少数民族语言的语音研究, 在实验语音学中大体上属于偏描写的、偏语言学的研究取向。虽然偏描写, 但与传统的方言学描写不同; 我们提倡一种新描写主义的方法, 也就是实验描写主义, 在研究内容上大致包括语言的语音与音系。生成学派 (Chomsky and Halle, 1968) 区分语音 (phonetics) 与音系 (phonology), 并认为语音学是研究语音的, 主要是语言外部 (linguistic external) 的内容, 音系学是研究音系的, 是语言内部 (linguistic internal) 的内容。如文中所述, 我们的语言观与此不同, 本文的实验语音学研究包括语言的语音与音系, 即大体采用 Ohala (1991, 1995) 所倡导的取向。简要地说, 就是以现在的语音科学实验手段探寻方言语音中的音类的性质与规律: 这些音的物理特性是什么? 发音上是如何实现的? 感知上有什么特点? 但不是简单地将原来方言学研究的音类画一个图, 统计一下, 验证一下记音的准确性; 更不是为了寻找某些僻见的语音。汉语方言的实验语音学研究是语言学的一个独立的子学科, 兄弟学科会发现它的研究成果有用, 但它本身并不是为别的学科服务的。本书稿全面阐释这个问题, 书稿共有五个讲题, 本文是第一讲, 解释这个子学科的任务, 第二讲是方法论, 第三至五讲以不同的音类为例讲解具体的研究。所举的例子以我自己的研究为主, 而且, 很多例子都是正在进行之中的研究, 着重讲述研究的旨趣与思路, 并没有“标准”答案。

因此, 当我们在谈论实验测量、统计的时候, 我们还是在谈论语言学, 一种跟原来的非实验传统有些不一样的语言学。

1.1 发音人在干什么? ——从语音基本单位出发

实验语音学的终极目标是从科学上弄清楚: 发音人在干什么? 这是通俗的话语, 换成科学术语, 首先就是语言/语音的基本单位问题: 它是相对连续的某一个整体, 比如音节 (syllable)? 还是离散的、抽象的一个一个音段 (segment) /音位或者其他更基本、更抽象的组成成分? 同时就是语音单位的真实存在性问题, 即它们在大脑/语言中的表征 (representation) 问题。

这不是个容易的问题。比如教语言 (尤其是外语) 的课本上一般都有章节介绍如何发音, 但往往语焉不详, 语言教师一般也不管这些, 只是要求学生: 跟我念! 念熟练了, 你就会了。我们做田野语言学调查的, 也是这样的。你不能奢望让你的发音合作人跟你解释某个音是怎么发出来的, 因为他根本不知道自己在干什么 (比如实现哪个目标), 他只是这么着自然而然地就发出音来了。受过专业训练的田野语言学工作者, 只有通过反复模仿发音人的发音, 直至发音人完全认可, 这个发音完全正确, 没有异地口音了, 才算过关, 然后, 语言学家便可以根据自己模仿的没有异地口音的发音来正确记录这个音了。从科学上讲, 这里面牵涉到一个理论假设, 即上文提到的语音感知的运动神经理论, 简单地说, 就是只有当你可以正确地发某个音的时候, 你对这个音的感知才是正确的。但并非所有人都同意这个理论, 事实上这个理论在语音感知领域支持者不多, 因为论断太强了, 一般的语音感知其实并不需要同时唤醒发音器官的运动神经。而语言学家们反而喜欢这个理论, 因为传统口耳之学的语音学或者描写语言学、方言学强调的就是通过内省的发音训练正确感知陌生语言的语音, 以达到正确记录语音的目标。语言学家们

相信通过自己的口腔实践可以找到各类语音发音时的肌肉内省感 (proprioceptive sensation), 构音时发音器官的触碰感 (tactile sensation)。但是, 有一些语音学家是不大相信这些的, 因为他们认为声学-听感是更重要的相关物。比如, 当你训练正则元音 (Cardinal Vowels) 的发音, [i-e-ε-a] 等距离下降舌位时, 他们认为下降的其实并不是舌头的位置, 而是你的听感声学距离 (参见胡方, 2008 中的相关述评)。

所以, 问题的复杂性在于语言发音这件事情牵涉的不仅仅是生理发音 (发音器官的运动) 这么简单, 而是生理发音受运动神经控制 (speech motor control)、感知声学调节的综合结果; 而且, 语音还受到更高层级的语言学与认知层面的制约, 因为语音本身也是语言的语法的一部分。对于“发音人在干什么”这个问题, 广义的回答包括以上诸多层面, 在科学研究上, 这是需要厘清的, 即我们要知道我们是在哪个层面上回答这个问题。一方面, 我们可以从语音产生 (speech production) 的角度进行回答: 在运动神经控制层面, 就像每个人写字的笔迹不同, 发同一个音, 发音人可以有不同的运动神经控制模式; 运动神经控制的输出就是可以测量到的发音动作, 进而就是发音位置的不同, 这就是我们一般所说的狭义的发音的不同。另一方面, 我们也可以从语音感知 (speech perception) 的角度考虑, 发音与声学上的不同是如何被感知的。从目的论的角度看, 发音是为了实现感知声学目标; 可是, 目的论把问题过于简单化了。其一, 如前文所述, 大脑中控制语音的产生与感知的机制是不同的 (Bouchard et al., 2013; Mesgarani et al., 2014)。其二, 发音与声学之间的关系不是一对一的, 这种非线性关系, Stevens (1972, 1989) 将其理论化为发音与声学之间的量子特性 (quantal nature)。其三, 即便是量子特性, 发音与声学关系也存在着发音人之间 (inter-speaker) 与发音人内部 (intra-speaker) 的变异 (variations), 而且, 还可能跟特定发音所涉及的生理解剖相关 (Perkell, 1996); 近年的研究发现, 个体的不同是发音与声学关系、产生与感知关系的一个重要变异源 (Fuchs et al., 2015)。目的论在语言学中很流行, 因为语言学的论证一般是基于直观推理的, 而不是实验科学, 诚如雅可布森所言, “我们说话是为了让人听见, 需要让人听见是为了被人理解²” (Jakobson & Waugh, 1979), 因而, 从语言是交际工具这个角度出发, 目的是决定性的。在目的论的逻辑看来, 发音人自觉或不自觉地运用某种发音策略, 所产生的语音具有客观的声学特性、感知特性, 其目的是实现该语音的音系地位, 即该语音在目标语言中的区别性。也就是说, 语言学将整个发音过程简化为实现音位的区别对立性了。事实上, 整个过程是怎么样的? 这是需要研究的。而且, 更为复杂的是, 发音人并不是想做什么就能做到什么; 比如, 你会发现: 怎么有些平调 (level tone) 的基频曲线 (F_0 contour) 有点降呢? 怎么浊塞音在语图上不是 100% 有浊音杠呢? 怎么单元音的共振峰轨迹 (formant trajectory) 有时也会有动程呢?

简要地说, 非实验的语言学是建立在语音的区别性之上的, 具有区别性的音构成语言中不同的音类, 整个现代语言学大厦便建立在此基础之上; 而基于实验的语音学研究则要回答这些音类是怎么样的, 如何产生, 如何感知, 如何在语言中构成系统? 复杂的问题从简单处入手, 可以暂且先把发音策略、感知特性等科学问题放在一边, 从语音声学 (speech acoustics) 出发, 测量语音 (phones or sampled segments) 的声学特性, 观察语言中的音类是怎么样的? 这个过程, 从语言学的角度来看, 就是观察语音的区别特性如何实现³! 在具体操作上, 我们采样的是个体的语音, 但是, 我们要描写的其实是语音的类 (phonetic category), 音系上的自然类 (natural class)。因此, 从语言学的角度在检视所采样的语音的声学参数时, “发音人在干什么”这个问题便更加具体化了: 哪些声学特性在我们所研究的目标语言中是具有区别音类的作用的? 哪些又是冗余的?

² 原文: We speak in order to be heard and need to be heard in order to be understood.

³ 在语言学传统上, 语音之间的区别特性是从心理的角度根据直观推理定义的, 即由区别特征 (distinctive features) 构成, 但近些年来心理学领域的研究发现人类在类别感知上可能具有范例性特点 (exemplar model), 语言的音类也是如此 (参见: Johnson, 2007), 由于本文不具体讨论这个问题, 因此, 不细究。

我们来看一个声调的例子 (Zhang & Hu, 2015)。

据刘丹青的调查, 徽语祁门方言有 6 个声调 (见平田昌司主编, 1998, 126 页):

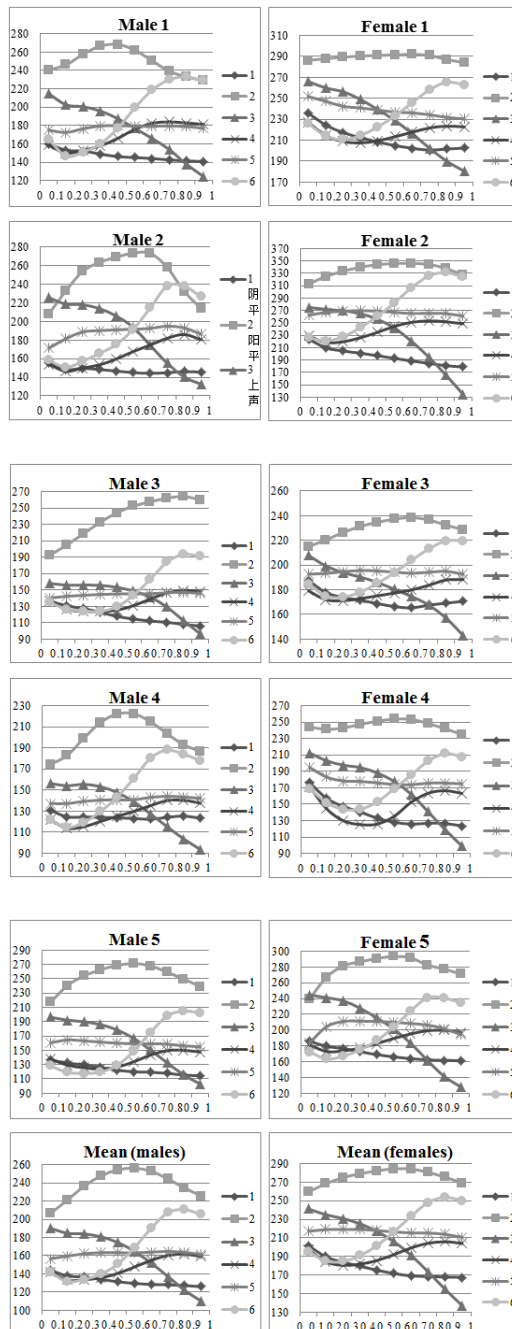
阴平 11 阳平 55 上声 42 阴去 213 阳去 33 阴入 435

中古平、去、入今分阴阳, 不过, 阳入并入阳去, 独立成调的阴入也不短促; 上声不分阴阳, 但有一部分古全浊上声字今读阳去。

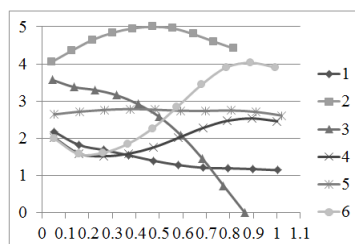
$$y_i = \lg x_i ; z_i = \frac{y_i - m_y}{s_y} \quad (1)$$

$$RD_i = \frac{z_i - z_{min}}{z_{max} - z_{min}} \quad (2)$$

$$DR_i = \frac{D_i}{M_i} \quad (3)$$



图一： 祁门声调的基频曲线，左：男发音人；右：女发音人。



图二： 祁门声调。

在核对了记音的基础上，我们对祁门声调进行了声学采样。每个声调用5个单音节例字，例字既单念也放在载体句中，录音重复5遍。5男5女共10位成年发音人参与了录音，他们均说地道祁门话，没有言语或听力障碍。录音在田野调查中进行，在一个安静的房间内通过TerraTec DMX 6Fire USB声卡和SHURE SM86麦克风直接录到笔记本电脑中，声音的采样率是11,025赫兹，16位。我们在praat 5.3.48 (Boersma & Weenink, 2014) 中将每个采样例字的韵母段标注为声调的承载段 (Tone Bearing Unit, TBU)。在检视每个采样之后我们将每个样本的基频曲线均分并提取10个点上的基频，我们用Logarithm Z-score (LZ-score)方法对基频进行归一。如公式(1)所示， x_i 代表采样点的基频值， y_i 是 x_i 的对数值，那么，LZ-score值 z_i 就可以表示为 y_i 与对数均值之差除以标准差 s_y 。然后，我们将LZ-score值进一步转换成以每个发音人为基础的相对值 (Relative Degree, RD)：如公式(2)所示， z_{max} 和 z_{min} 分别是某个发音人的LZ-score最大值和最小值；这样，计算的结果与转写声调的五度值 (Chao, 1930) 之间就有一个直观的参照。声调时长也用相对化 (Duration Relativization, DR) 进行归一，如公式(3)所示， D_i 代表测量的实际时长， M_i 代表6个声调的平均时长。

图一总结了每个发音人的声调基频曲线均值以及男女发音人均值，横坐标是采样的10个点，纵坐标是赫兹值。祁门的六个声调阴平、阳平、上声、阴去、阳去、阴入用数字依次表示。男女发音人的基频幅度均有160赫兹左右，其中，男发音人约在100-260赫兹之间，女发音人约在130-290赫兹之间。在检视了每个发音人的声调基频曲线之后，我们把按上述方法归一之后的基频曲线总结如图二。

先来看祁门方言的三个平调。只有记为33的阳去的基频曲线大致是平的，男发音人均值约在163赫兹，女发音人均值约在216赫兹；当然，在不同的个体发音人中，也有一些变异的情况，比如：在女发音人1中，它是下倾的，在女发音人4中，它有个下降的调头，在男发音人2、女发音人5中，它有个上升的调头等。记为11的阴平调在大部分发音人中都是下倾的，也就是说，从基频曲线看，它是个低降的调形：男发音人约从144赫兹下降至131赫兹，女发音人约从202赫兹下降至175赫兹。而记为55的阳平调从基频曲线上看则是个升降型曲折调形，根本不是平调。

那么，它们是平调么？我们认为的。而且，我们认为正是因为祁门有三个平调，所以，其中有两个平调在基频曲线上才不实现为平的调型。音系特征有所谓的偶值理论，声调也有所谓高低调域论；但祁门的平调是典型的三值。中平实现为一条平实的基频曲线 (plain level)；低平略带缓降；高平在基频上特别高，发音人往往会运用特别的发声态-假声 (falsetto) 来实现：男发音人的高平 (即阳平调) 最高基频均值是256赫兹，女发音人是284赫兹。需要注意的是，无论是假声，还是基频曲线上所见的曲折，都是发音人为了实现这一特高音调目标 (pitch target) 所使用的策略；而并不是假声这个发声态具有什么重要的语言学意义。也就是说，祁门方言的“高、中、低”这三个平调对立是通过发音人运用不同的发音策略来实现的。

当然，对这个问题的完全解释需要进一步的研究，这超出了本文的能力范围。我们这里强调的是：基频虽然是声调最直接的物理关联物，但却并不是直接等同于声调。我们检视一个语言或方言的基频样本时，应该思考的是它的声调对立是什么？它的声调对立是如何通过基频

来实现的？而不是简单地根据基频曲线所反应的调形折算出一个五度值了事。在文献中与祁门类似的有尼日利亚的Yoruba语（Hombert, 1976a、b）、泰语（Abramson, 1962）。Yoruba语仅有的三个声调从基频曲线上看也是低调降、中调平、高调升降（但基频没有祁门那么高），但它们都是音系上的平调；泰语的三个平调的调型与Yoruba类似，不过，泰语除了这三个平调之外，还有一个升调，一个降调，在声调音系上与祁门更接近，而且，Abramson（1972, 1975, 1976, 1978）有一系列的文章讨论泰语声调的产生与感知问题，可以参考（另参见Gandour, 1978对声调感知问题的述评）。

上声是祁门唯一的降调，其前半部分（约首45%时长）缓降，后半部分速降：男发音人约从190赫兹下降至110赫兹；女发音人约从241赫兹下降至137赫兹。而且，无论男女，上声的最低点就是声调的最低点，即调域的下限。因此，从这个意义上讲，上声记成41更妥一些。

祁门的另外两个声调阴去与阴入的调形是一样的，都是降升曲折型。而且，两个声调的调头几乎相同：在男发音人中，阴去的调头下降约9赫兹，阴入的调头下降约10赫兹；在女发音人中，阴去的调头下降约17赫兹，阴入的调头下降约10赫兹。两个声调的调形的区别仅在于上升的幅度，阴去上升不足30赫兹，而阴入则上升约79赫兹。也就是说，祁门的两个升调虽然调头相同，但高升调（阴入）升得高，低升调（阴去）升得低。祁门的升调明显不支持所谓的调域理论（Yip, 1980; Bao, 1999），因为正如Bao（1999）所明确指出的，调域理论认为一个语言的声调系统应该避免拥有相同的调头或调尾的声调，而祁门的两个升调恰恰如此，而且，阴入调贯穿低-高两个调域。刘丹青（1998）将这两个声调记为曲折调，其中阴去213，阴入435，保留了语音细节；我们则认为在音系上可以将这两个声调记为升调，阴去23，阴入25。这里有两个理由：首先是两个声调的调头相同，都是低降升，而且，低升调拥有一个略降的起始也是符合低升调的产生与感知特点的（Shen & Lin, 1991）；其次，如果这两个声调记成降升调，祁门方言便没有升调，不符合类型学的一般规律，因此，基于剃刀原则，音系上还是处理为两个升调为妥。另外需要指出的是，刘的记音明显受到调域观的影响，阴去213在低调域，阴入435在高调域，站在调域理论的角度看，很完美，但是，这并不符合事实。

基频（ F_0 ）曲线是声调的物理关联物，通过声学采样，我们可以看到这些基频曲线的形状，即调形：平的、升的、降的、升降曲折型的、降升曲折型的等等。而在谈论这些物理测量的时候，我们思考的是语言的声调：这个语言或方言中的声调之间的区别在哪里？声调是个语言学概念，与直观的物理关联物基频不同，语言学概念具有一定的抽象性，因此，在检视基频曲线的时候，我们事实上是在分析这些拱度（contour）究竟是语音的，还是音系的？要完整回答这个问题，需要考察声调产生与感知的相关层面。从现在的技术手段条件下最容易采样到的物理的基频出发，想要谈论抽象的语言学的声调，至少还需要仔细考虑声调的心理感知问题。基频的感知是音高（pitch）。虽然人耳对于基频的感知接近于线性关系，因为语言中声调的基频一般都在100到300赫兹范围内，也就是不到两个倍频程（octave）的距离，但是，人耳对于基频上升与下降的感知是不同的，在方言中我们常常发现：基频升高10赫兹很容易被感知为明显的升调，但基频降低10赫兹则基本还是会被感知为平调；而调头、调尾及相关的感知问题可能更加复杂（Hombert, 1978），在声学研究的基础上，这些声调感知方面的问题都是可以进一步探讨的。

1.2 实验语音学不是只研究疑难冷僻音类，而是研究所有语音问题

上面我们用了一个简单的声调的例子来说明，方言语音的实验语音学研究的目的并不是简单地将音类具体化，数字化，或者图形化，而是借助音类（上文即是声调）的物理关联物（上文即是基频曲线）更为直观地分析、寻找语音之间区分的语言学意义所在。同时，我们也看到，语音细节可以帮助我们更好地理解音类之间的区别，而不仅仅是补口耳之缺，验证一下记音。当然，实验语音细节可以修正口耳记音中的偏差。这里需要注意的是，有些偏差在语言学意义上无关紧要，比如祁门的上声是记成42调还是41调；但也有些偏差是具有一定的理论意义的，

比如上文对于调域的讨论。

本节将再讨论一个例子，元音共振峰的问题。不少人对实验语音学有一个误解，认为只有那些不常见的、奇怪的语音才需要进行实验语音学研究，而像元音这样最正常不过的语音是不需要实验语音学研究的，好像实验语音学是专门针对疑难杂症的。我们这里要强调的便是，实验语音学并不是只对僻见的语音现象感兴趣，不是说只有发现哪里有内爆音（implosive）、哪里有嘎列音（creaky voice）才有意义，相反，内爆、嘎列本身并没有什么意义，内爆了、嘎列了又如何才是有语言学意义的。比如上文谈到的祁门的高平调的假声就没有太多的语言学意义，它是说话人实现高平音高目标的发音策略之一，才是假声的语言学意义所在。

实验语音学研究所有的语音现象，汉语方言语音的实验语音学是阵地战，是在新的语音观、语言观的指导下，使用实验科学的方法，全面研究方言语音问题。元音问题后面有专门的章节讨论，这里简要地说一个单、双元音的问题。单、双元音都是汉语方言中常见的元音，双元音又可分为降峰双元音、升峰双元音等。我们不使用更常见的术语如前响双元音、后响双元音，是因为前响、后响等在字面上均涉及听感判断，不如降峰、升峰等根据纯声学参数定义的术语更为中性。

传统的汉语方言研究以声（母）、韵（母）、（声）调为纲，一般不进一步分析辅音音位、元音音位、声调音位等。声调的音位问题可能相对简单；由于汉语方言一般也没有复辅音，因此声母与辅音音位也基本类同；但是，韵母与元音音位则区别大了。除了普通话的元音音位问题有多篇文章进行讨论之外，方言语音的文章很少涉及元音音位问题，一般就是总结出多少韵母，类型学研究也是依靠单元音韵母、复合元音韵母、鼻音尾韵母、塞音尾韵母等简单的分类来进行。但是，这样的研究是很危险的，因为它会得出危险、甚至是荒唐的结论。比如说，梅县客家话只有 6 个单元音 [ɿ i e a o u]，因此，Zee & Lee (2007)、叶晓锋 (2011) 的类型学总结中都把客家话归类为元音最少的汉语方言。我们认为这种类型学总结具有很大的误导性，它的问题就在于机械地以单双元音分类为纲，割裂了元音音位对立的音系本质问题。同为客家方言，在江西的客家话中，有 11 个单元音的方言（张倩，2013），那么，江西客家话的元音音位的数量与梅县是否截然不同了呢？如果机械地以单元音论，答案是明显的。但在我们看来，不能这样简单机械地进行类型学归纳，因为梅县客家话还有 5 个降峰或平峰双元音 [ai ɔi ui au ɛu]，这些双元音也是元音音位，因此，与江西的客家话类似，梅县客家也是 11 个元音音位（Zhang & Hu, 2015）。事实上，由于汉语的音节结构简单，因此，汉语方言拥有 10 个或以上的元音音位，是一种常态现象；而割裂单双元音关系，机械地以单元音来进行类型比较研究，是非常危险的。最荒唐的一个例子来自著名的《科学》杂志。Atkinson (2011) 撰文，力图证明世界语言的多样性分布与人类的基因图谱是吻合的，因此支持人类的非洲起源说。他的证据就是语音复杂性的计算，理论假设便是越古老的语言其语音也越复杂，回到我们这里讨论的元音问题，那么，元音个数越多的语言便是越古老的语言，根据他的计算，全世界的语言刚好是吻合他的理论的。此文刊出之后招致诸多批评，其中，Wang 等 (2012) 的文章采用“以彼之道，还施彼身”的方法，提出一个完全相反的结论，即：按照 Atkinson 的逻辑，亚洲才是语言起源的地方；其中，一个重要的证据就是吴语元音的复杂性。当然，在他们的统计、计算中，元音均只包括单元音，这便是得出荒谬结论的一个根本原因之所在。

关于双元音的性质问题，文献中很早便有讨论（Pike, 1947; Lehiste & Peterson, 1961; Holbrook & Fairbanks, 1962）。大致说来，有两种观点，一种观点认为双元音是一个单独的元音，它的核心在语音上是复杂的（Malmberg, 1963; Abercombie, 1967; Catford, 1977）；另一种观点则认为双元音就是两个元音成分或者一个元音成分和另一个半元音成分的组合序列（Sweet 1877; Jones 1922）。也就是说，前者将双元音视为单一发音事件，只有一个动态的目标；而后者则将双元音看成是两个发音事件，从一个静态的目标过渡到另一个静态的目标。

另一方面，降峰双元音与升峰双元音在语音或音系上的表现往往是不同的，赵元任（1928：

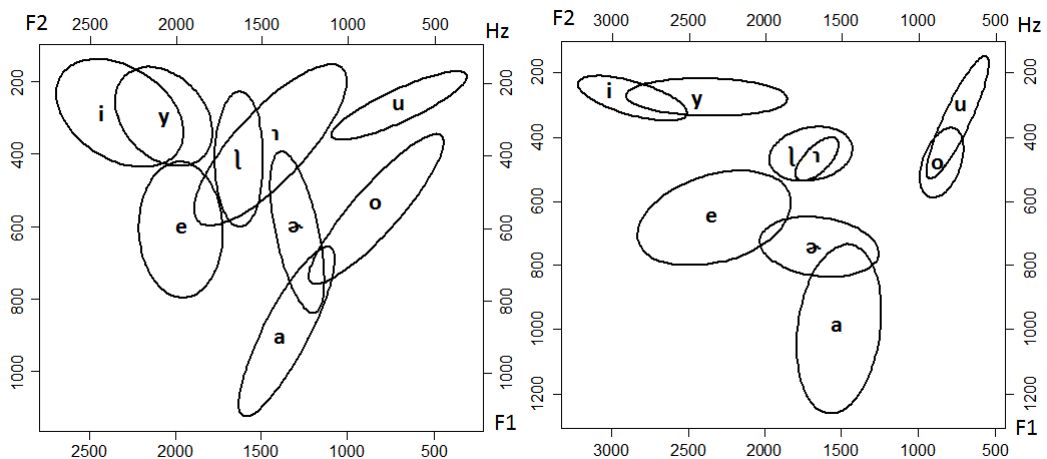
65-66) 早就指出：吴语中只有降峰双元音才是“真复合元音”，升峰双元音不是。事实上，无论在历时或共时层面，在语言中常常能发现，降峰双元音可以和相应的单元音形成交替 (alternation)，升峰双元音则不同。最常见的如/ai/与/ɛ/、/au/与/ɔ/之间的交替，在晋语、吴语等方言中都是很常见的。基于声学与发音的材料，我们最近的研究则明确指出：升峰双元音拥有两个目标，因此，[ia]应该视为[i]与[a]的序列；而降峰双元音则只有一个动态的目标，因此，[ai]并不是[a]与[i]的序列，而是一个整体的元音，与单元音[a]、[i]等构成音位对立 (胡方, 2013)。我们并不认为我们的这个结论一定就放之四海而皆准，也并不期待如此，相反，我们期待不同的方言可能会有不同的情况。比如说，我们认为普通话或者北京官话的双元音/ai/是一个动态目标，是一个整体的元音音位，而英语的双元音/ai/，可能就是[a]、[i]的序列；因为听北京人或者相关官话方言背景的人说英语，他们说不好英语的/ai/，而是用[ɛ]代替，比如前些年北京的公交地铁上的报站语音，英语的“arrive”就说成[ə'ɪɛv]。这是很有意思的一个语音现象，北京话或者普通话有双元音/ai/，但是，在说英语的双元音/ai/的时候，不用自己语音系统里面有的双元音/ai/，而是用了一个在北京话/普通话中并不符合 CV 音节拼合关系的单元音[ɛ]来代替；这便说明在北京人的大脑/语法中，他们认为普通话的/ai/与英语的/ai/是完全不同的东西。

我们这里举例的是如何通过语音观、语言观的革新，通过语音数据的采样，借助看似简单、枯燥的声学测量，将以前口耳之学中只能意会的一些内容展示出来，然后便可以比较各方言中的各类单元音、双元音，以及一切其他音类，厘清一些事实。这里，我们举一个西南官话的例子 (邱玥、胡方, 2013)。

隆昌话属西南官话江贡小片，有韵母 36 个，如表一所示。

表一：隆昌话韵母表

ɿ	ʅ	a	o	e	ə
ai	əi	au	əu		
an	ən	aŋ	oŋ		
i	ia	ie	io	iau	iəu
ien	in	iaŋ	ioŋ		
u	ua	ue	uai	uəi	
uan	uən	uaŋ			
y	ye	yen	yn		



图三：隆昌话单元音 (左：男；右：女)。

我们对所有的韵母进行了声学录音采样，选取含有隆昌话目标元音的单音节字，并尽量选取阴平调、零声母或唇音声母字，嵌入“__，读__三遍”的载体句中。录音在实地调查中进行，有效样本是五男四女九位发音人，均为 20-25 岁的青年，从小在隆昌县长大，母语为本地方言，也学过普通话和英语，属于典型的隆昌青年口音。录音使用 Sony D50 线性录音棒，采样率为 16,000 赫兹；录音重复五遍。由于没有显著差异，本文将“__，读__三遍”中 2 个位置的目标元音音段数据统计在一起分析，即每位发音人每个目标元音的有效样本数据为 10 个。这里，我们主要讨论隆昌话的 9 个单元音[ɿ ʅ i u y a e o ə]与 4 个降峰双元音[ai əi au əu]；在降峰双元音与升峰双元音的对比中，我们选取[ai au]与[ia ua]这两对语言中最常见的双元音。而且，讨论集中在共振峰模式，略去时间结构。

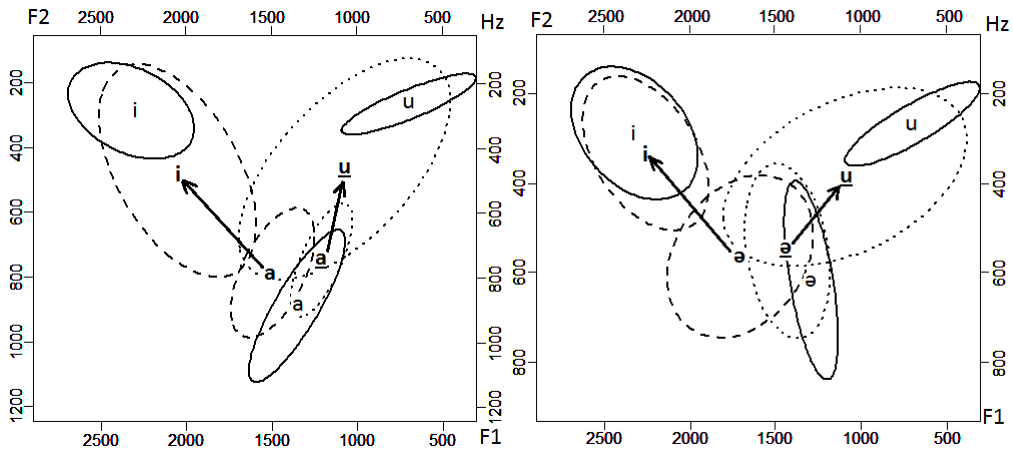
隆昌单元音在声学元音图中的分布见图三所示，图中纵轴是元音的第一共振峰（F1）、横轴是第二共振峰（F2），单位：赫兹，置信椭圆为两个标准差（关于声学元音图与置信椭圆的详细信息，我们会在元音的章节介绍，也可参见胡方，2014）。如上所述，男发音人的共振峰数据来自五位发音人，每个单元音各自共有 50 个样本；女发音人的共振峰数据来自四位发音人，每个单元音各自共有 40 个样本。

从图中可以看到，隆昌话的单元音构成一个典型的三角形分布，前高元音[i]、后高元音[u]、央低元音[a]分别处于顶点位置。隆昌话的元音高低分三个层级，除了高元音[i y u]、低元音[a]之外，还有半高元音[e o]。这里需要指出的是，虽然标为[e o]，但事实上与正则元音相比较，他们实际上位于半高、半低之间，即不区分半高、半低。隆昌话除了在前高位置有圆唇与否的对立，其他前元音都不圆唇，后元音都圆唇。与普通话类似，隆昌方言的两个舌尖元音[ɿ ʅ]在声学元音图中位于高、央的位置，但比高元音略低。另外，也是与普通话类似，隆昌有一个儿化的央元音[ə]。

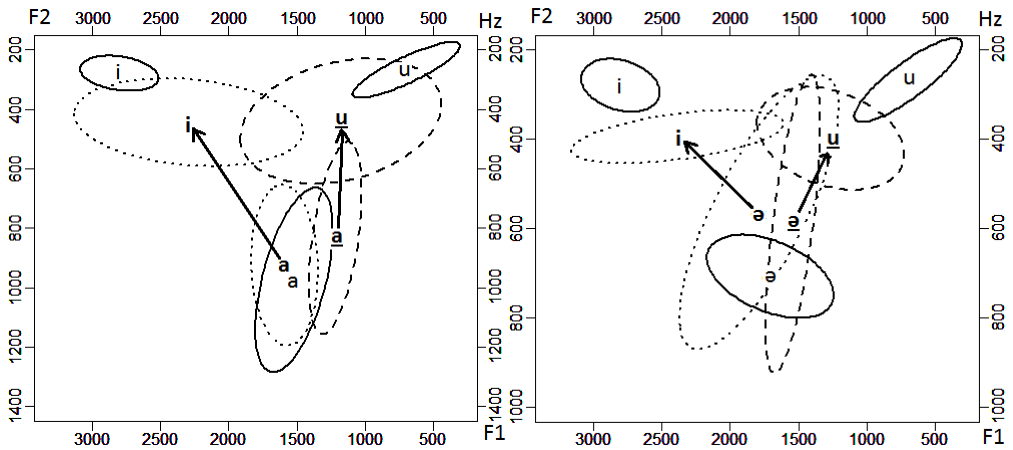
因此，如果只看单元音，隆昌话的元音系统非常简单、整齐。但事实上，隆昌还有四个降峰双元音[ai au əi əu]。图四和图五分别比较了男、女发音人的双元音首尾成分与相应的单元音在声学元音图中的分布；其中，实线椭圆是单元音，虚线或点线椭圆是双元音的首尾成分，箭头则简化示意双元音共振峰模式的变化方向。

之所以比较双元音首尾成分与对应的单元音，是想知道双元音如何实现其声学目标。虽然，不能说单元音就是双元音成分的声学目标，但是，两者之间的比较提供了一个很好的参照。从图中可以看到，双元音[ai au]中的首成分[a]与其相应地单元音[a]相比分别偏前和偏后，这应该是受到各自的尾成分的影响，显现出一定程度的逆协同发音，但是，他们的椭圆的大小与单元音[a]类似，而且重叠明显；也就是说，这说明双元音[ai au]的发音是从一个类似于单元音[a]的声学位置开始的，即其声学目标是比较清楚的。相比较而言，双元音[ai au]的尾成分[i u]和对应的单元音[i u]位置差距很大，椭圆几乎不重合，说明并未到达目标位置，而且，椭圆面积显著增大，说明数据分布的离散度增加、可变性增大；也就是说，双元音的[ai au]的发音似乎并没有一个明确的尾成分声学目标，其尾成分的位置是由双元音的动态特性所决定的。与双元音[ai au]类似，[əi əu]的尾成分也没有一个明确的声学目标。而且，[əi əu]的首成分[ə]的可变性也非常大，虽然[əi əu]中的[ə]的椭圆与相应的单元音[ə]的椭圆还是重叠的，但是双元音[əi əu]中的首成分[ə]的离散度显著增加，尤其在女发音人的数据中，这个特点更加明显。

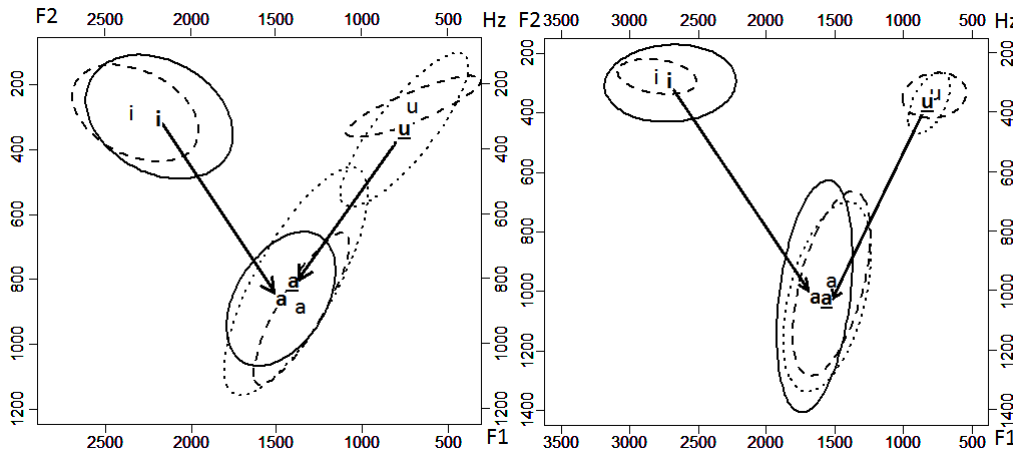
综上所述，隆昌话的降峰双元音[ai au əi əu]并不如其标音所示，是由两个目标组成的；相反，声学材料的分析表明，他们更像是一个动态的目标。[ai au]由一个接近于单元音[a]的声学位置开始，但不需要实现其尾目标[i]或[u]，相反，由于受到双元音自身的动态特性制约，在不到目标的位置就结束了；[əi əu]也类似，而且，从其变异性更大的首成分[ə]的分布来看，其动态特性更加重要。因此，我们认为降峰双元音应该如单元音一样，是隆昌话元音音位对立的组成部分；如果把这四个双元音也增加到如图三所示的隆昌元音图中，那么，隆昌元音的高低对立维度和/或圆唇对立层次就会复杂得多。



图四：隆昌话双元音[ai au]（左）、[əi əu]（右）与单元音[i u a ə]的比较（男）。



图五：隆昌话双元音[ai au]（左）、[əi əu]（右）与单元音[i u a ə]的比较（女）。



图六：隆昌升峰双元音[ia ua]与单元音[i u a] 的比较（左：男；右：女）。

升峰双元音则全然不同。如图六所示，虽然有不同程度的可变性，隆昌升峰双元音[ia ua]的首尾成分（实线或点线椭圆）与其对应的单元音[i u a]的椭圆（虚线）大部分重合，并且均值点的位置非常接近；因此，声学材料支持将隆昌话的升峰双元音看做是由两个相对稳定的声学目标组成的，其发音过程中就是从—个稳定的目标位置过渡到另一个稳定的目标位置。也就是

说，[ia ua]就是[i]与[a]、[u]与[a]的序列。

1.3 语音的变异性

在现在通行的普通语言学的理论框架中，一般把语音看做是均质的，即一个音就是百分之一百的一个音。如前文所述，普通语言学更偏重于抽象层面的分析，无论是形式学派的还是功能学派的，基本还是只关心语言中抽象的那部分，即将语言视为人类的一种或内在的（innate）或交际驱动的抽象的能力（competence），将语言的具体产出视为一种执行（performance）。因此，音类、音位都是从感知、心理的角度根据直观推理定义的，在语言学中强调其区别性。但事实上，语音是变异的（variant），一个音并不一定百分之一百是那个音。

语音学已经逐渐发展为实验科学，拥有完全不同的世界观（语言观），语音研究的核心任务当然还是人脑中抽象的语言/语音能力，但实验语音学在研究中直接面对大量具体的语音产出样本，主张从具体的采样中去研究抽象的语言/语音能力。也就是说，语音学关心的不只是语言的执行，而是“怎么执行”，根据执行来建立音类的模型。与传统的非实验的普通语言学相比较，实验语音学提供了基于言语产生（speech production）的视角去观察语言学音类问题：可以根据“具体”的产出去思考“抽象”的音类，当然，其中要受到感知以及语音产生与感知之间的关系的制约。因此，在实验语音学的视角下，语音是变异的。比如上文的语音分析，我们根据每个元音 40-50 次采样用置信椭圆来建立其分布的模型，这就是基于语音产生视角的音类概念，因为人不能两次发出完全一样的同一个音。

有些语音之间存在着范畴边界，有些并不存在范畴边界，比如一般认为元音的感知是连续的，而大部分辅音的感知则是范畴的。无论有没有感知上的范畴边界，语音的类别化在不同的语言或方言中是带有个性化的（language-specific）。以连续感知的元音为例，从高元音[i]至低元音[a]之间，汉语普通话在高低维度上只区分/i/与/a/，其他的对立是通过其他语音手段引入的，比如通过圆唇引入/ɥ/，通过后响双元音引入/ɛɪ/，通过前响双元音引入/ai/、/ei/；日语则区分/i/、/e/、/a/；上文讨论的西南官话隆昌方言则在区分/i/、/e/、/a/的基础上通过元音动态化手段引入/ai/、/ei/；北部吴语方言则大多区分/i/、/e/、/ɛ/、/a/，并往往在高、半高位置进一步区分圆唇与否；欧洲的语言，比如德语则在区分/i/、/e/、/ɛ/、/a/的基础上通过时长手段再进一步引入长短元音对立。

我们这里以范畴感知的辅音为例，讲一个跨感知边界的例子。世界语言中的塞音、塞擦音以清浊、送气与否可以分为四类：清不送气、清送气、浊不送气、浊送气，印度的一些语言保持这四种对立（Ladefoged, 2006: 146-151），但在一般的语言中，塞音、塞擦音通常只有二类或者三类。汉语也是如此：中古汉语有三类（高本汉, 1915-26），在典型的吴语中得到了保留，即塞音、塞擦音今读清不送气、清送气、浊（不送气）三类（赵元任, 1928）；在大部分汉语方言中，中古浊塞音、塞擦音按照一定的条件并入了清不送气、清送气，即塞音、塞擦音今读清不送气、清送气两类。

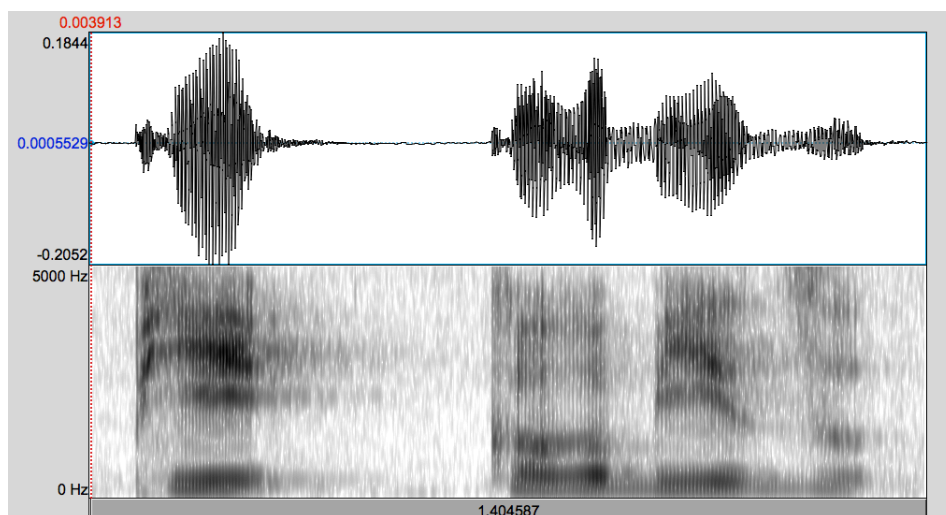
那么，浊塞音、塞擦音消失了吗？在音系的层面，答案是肯定的：浊塞音、塞擦音作为音类已经消失了。大致地说，在闽语中，中古浊塞音、塞擦音今读不送气清音；在客赣方言中，今读送气清音；官话方言如普通话，复杂一些，但也大致遵循“平送仄不送”之类的规则。而且，在保留浊塞音、塞擦音的方言如吴语、老湘语中，浊音也似乎正在消失，因为一般只能在音节间的位置，才能观察到浊塞音、塞擦音的浊音杠，而在单念这些所谓的浊塞音、塞擦音时，发音人的声带并不振动。在曹剑芬（1982, 1983）之后，一般大家都知道，除了伴随的低声调之外，吴语方言的浊塞音的浊感可以主要归因于后接元音的气声化。

那么，浊的塞音、塞擦音是被说这些方言的说话人从人脑中删除了么？还是被类别化到其他相关的音类中去了？以汉语官话方言为例，汉语官话方言将塞音、塞擦音类别化为清送气、清不送气二类。那么，浊塞音、塞擦音哪里去了？我们认为并不是从说话人的脑中删除了，而

是被类别化入清不送气音类了。理由有二：其一，清不送气塞音、塞擦音在语流中不重读时容易实现为浊；其二，在学习外语时，浊塞音、塞擦音这个似乎被屏蔽的语音范畴就会浮现出来。同样是将塞音、塞擦音类别化为二类，英语与汉语官话方言是不同的：英语区分清、浊。因此，汉语官话区的人学习英语的一个重要任务就是练习浊塞音、塞擦音的发音；不练习或者练习不好的话，他们就会拿清不送气音去代替浊音。此外还有一个挑战，英语的清塞音、塞擦音一般实现为清送气，只有在 *s* 后才实现为清不送气。在英语母语者看来，清送气、不送气是同一个音类，它们根据语音条件自然出现；但是汉语官话区的英语学习者是不同的，清送气、不送气是泾渭分明的二类，因此，他们在 *s* 之后也容易发成清送气音。怎么办？*s* 之后浊化的观点在国内英语教育界占有很大的市场，即英语教师们要求学生们在 *s* 之后清塞音、塞擦音要浊化，很大一部分汉语母语者的英语专业的学生就是这么学习英语的，而且，效果也还不错(梁波, 2017)。这种语音事实与教学法的错位，从另一个角度就折射出跨范畴语音在不同语言里类别化过程中的语言个性问题。

因此，在大部分汉语方言中，塞音、塞擦音今读清不送气、清送气二类，浊音作为音类已经消失了，而且，浊塞音、塞擦音作为语音范畴一般并不在自然语音产生中出现。如上文，我们只有通过语流弱化、外语学习等特殊情况，从逻辑推理层面去论证浊塞音、塞擦音却并没有从说话人的脑子里删除，而是隐含在相关的音类之中。幸运的是，在还有一些汉语方言中，虽然塞音、塞擦音也被类别化为清不送气、清送气二类，但是作为语音范畴的浊塞音、塞擦音在自然语音产生中还是继续出现的，因此，这里可以通过实验证据来讨论塞音、塞擦音在语言或方言中类别化的问题。

这些方言大致就是被视为有“送气分调”现象的赣西北方言(李如龙、张双庆, 1992; 刘纶鑫, 1999)。赣语的塞音、塞擦音今两分，即一般认为中古的浊塞音、浊塞擦音作为音类已经消失，演变成了现在的送气清音。我们通过对南昌、新建、安义、湖口、星子、都昌、修水、德安、武宁、永修等十个县市共计 30 位发音人的采样调查发现，在赣西北方言中，虽然塞音、塞擦音也是两分，但是，浊塞音、浊塞擦音作为语音范畴却并没有消失，而是与送气清音合并了。也就是说，在发这个塞音、塞擦音的音类的时候，这些方言的说话人并不是在发一个送气清音，或者在发一个浊音；而是他们并不区分这两个语音范畴，在他们的发音目标中，这两个语音范畴是同一类，因此，你会观察到，浊塞音、塞擦音一会儿实现为清送气，一会儿实现为浊。



鼻， 这 个 鼻 字

图七：永修方言“鼻”/p^hi?ɔ/的发音：单念与载体句。

我们这里用永修这个典型的例子来说这个问题。如图七所示，把目标词“鼻”放在载体句“X, 这个 X 字”中进行音频数据采样，永修方言的一位男性发音人，把第一个 X 位置单念的“鼻”实现为清送气，但是却把第二个 X 位置载体句中的“鼻”实现为浊不送气塞音。这说明永修并不区分 $[p^h]$ 与 $[b]$ ，它们都属于同一个音类 $/p^h/$ ，只是在单念的时候容易实现为 $[p^h]$ ，在语流中容易实现为 $[b]$ 。为了更全面地了解 $/p^h/$ 的变异，接下来，我们进一步分析了永修方言双唇塞音声母拼-i 韵母的情况。我们统计了三位永修发音人的声学样本：测量了清不送气声母、清送气声母的例字，其中，清送气声母的例字包括来源中古次清声母字与中古浊声母字。

清不送气声母 $/pi/$ 共四个例字，每个例字重复三遍，除去三个错漏样本，共得到 33 个有效样本。在单念的时候，永修的 $/pi/$ 中的塞音声母 $[p]$ 嗓音开始时间 (VOT) 均值是 16 毫秒，标准差是 8 毫秒；韵母时长均值 208 毫秒，标准差 65 毫秒。对于清不送气双唇塞音来说，这个 VOT 是偏大的，因为我们在测量其他的方言或语言时，这个数值一般是小于 10 毫秒的。当目标例字处于载体句中时，除了 2 个例子，其他 31 个样本均不发生塞音浊化现象，即他们在音节间时，还是清不送气塞音。我们分别测量了这 31 个样本的闭塞段时长均值为 97 毫秒，标准差 13 毫秒；VOT 为 11 毫秒，标准差 6 毫秒；韵母时长为 185 毫秒，标准差 43 毫秒。这里有意思的是，在音节间时，清不送气塞音 $[p]$ 的 VOT 均值缩短为 11 毫秒，接近其他方言或语言的数值。另一点是，不管处于什么声调、舒声音节还是促声音节，从较小的标准差值可以看到，闭塞段的时长相当稳定。而且，将近 100 毫秒的闭塞段从另一个侧面反应了，较长的闭塞段时间可能是保持音节间清不送气塞音的一个特征条件；因为在清不送气塞音浊化的两个特例中，音节间的闭塞段全部浊化，同时，时长只有 50 毫秒左右，音节结构与下面要讨论的送气塞音类似。

来自中古的次清声母与全浊声母字的例字也各有四个，不过，并没有发现在永修的发音人中有任何差别，因此，我们将这两类数据放在一起讨论；但是，我们倒是发现 3 个发音人中，有 1 个是与另 2 个不同的，因此，我们分开来讨论。其中 2 个发音人的情况是这样的，除去错漏，共有 40 个有效样本：在单念时，均实现为清不送气塞音 $[p^h]$ 。 $[p^h]$ 的 VOT 均值是 42 毫秒，标准差 16 毫秒；韵母时长均值是 210 毫秒，标准差 65 毫秒。在载体句中时，均实现为浊塞音 $[b]$ 。 $[b]$ 的 VOT 均值是 -57 毫秒，标准差 12 毫秒；韵母时长是 212 毫秒，标准差 42 毫秒。也就是说，在单念时， $/p^h/$ 实现为 $[p^h]$ ，用较大的 VOT 值与不送气的 $[p]$ 相区别；而在音节间时， $/p^h/$ 实现为 $[b]$ ，用 VOT 负值与不送气清塞音 $[p]$ 相区别。另外一点与清不送气塞音不同的是，当 $/p^h/$ 处于音节间时，闭塞段时长较小，均值 57 毫秒，即在音节间的闭塞段，声带是始终保持振动的。

另一个发音人的 $/p^h i/$ 在单念时 $[p^h]$ 的 VOT 明显大于前面两个发音人，24 个样本的均值为 68 毫秒，标准差 26 毫秒，显示更强的送气特征；相应的韵母时长均值为 253 毫秒，标准差 67 毫秒。而且，这种强送气特征在音节间得到体现：只有在 3 个样本中， $/p^h/$ 实现为 $[b]$ ；而在另外 21 个样本中， $/p^h/$ 还是实现为 $[p^h]$ ，其中，闭塞段时长均值为 63 毫秒（标准差：14 毫秒），VOT 均值为 64 毫秒（标准差：21 毫秒），韵母时长均值为 219 毫秒（标准差：43 毫秒）。

从上面的讨论中，我们可以看到，除了说话人内部的差异 (intra-speaker variations)，说话人之间的差异 (inter-speaker variations) 也是很重要的一个变异来源。事实上，方言间的变异情况会更加复杂，以这里讨论的赣西北方言塞音、塞擦音为例，我们相信中古次清声母与全浊声母合流为一个音位是这片方言共同创新 (shared innovation) 的起点，但是，合流之后的这个音类经过长时间的演变，现在不同方言、不同说话人中是如何实现的，会呈现出复杂的多样性，尤其是当它与声调发展问题交织在一起时，更是添加了复杂性。因为不是本文的主题，此处不赘。这里强调的是，语音是变异的，不仅仅连续感知的语音范畴具有变异性，各个语言或者方言对于跨范畴的语音的类别化也是带有个体性的。以此处讨论的塞音、塞擦音类来说：吴语类别化为清不送气、清送气、浊三类；英语为清（含送气、不送气）、浊两类；官话方言为

清不送气（隐含浊）、清送气两类；赣西北方言则为清不送气、清送气（含浊）。

1.4 结语：实验证据为基础的语音研究，但还是语音学研究

实验语音学测量所提供的语音细节为我们观察语音区别特性的具体内容提供了可能性，但是同时，语音也具备了具体、变异的特性，不再像以往概念化似的那般抽象。一个语音目标，无论其是一个语音单位还是一个区别性的特征，它可能并不是绝对抽象、均质的，而是相对具体、变异的，都可以从产生与感知的角度对他们进行研究。一方面，我们强调，无论何种测量手段，我们最终关心的是语音的语言学意义；而另一方面，我们也知道，即使是描写汉语方言中的语音现象，我们也需要全新的方法论与研究范式。因此，虽然我们在谈论实验语音学测量的时候还是在谈论语言学问题，但却已经是一种全新的语言学了。

新在什么地方呢？最核心的就是这是一种以实验证据为基础的研究（an evidence-based approach），而不是纯粹基于归纳推理的范式（a generalization-based approach）。语音学的学科发展可以 1886 年国际语音学会（International Phonetic Association）的成立为起始标志，但是经过了 130 多年，语音科学并没有发展成一个界限清楚的学科，相反，随着近几十年科技的迅猛发展，语音研究事实上形成了一个众多学科交叉的局面，因此，国际语音学会常设理事会举办的四年一届的语音科学大会（International Congress of Phonetic Sciences）将语音科学定义为复数形式。语音研究是语言学、语言教学、言语病理学、言语信息工程、人工智能等诸多科学、医学、工程学、人文科学领域共同关心的问题。仅就语音的本体研究而言，其核心科学问题是解释人类所拥有的语音能力，按照现在的理解，就是人体所具有的语音产生与语音感知这两项生物功能，在方法上可以有偏生理、偏心理、偏工程模拟的种种研究旨趣。本文提倡的是偏语言学本体意义的研究旨趣，也就是关心人类的语音能力在具体的语言或者方言中是怎么实现的，而不是将语音语料视为探索人脑语音能力的刺激项；本文以汉语方言为主要研究对象，就是偏重描写语音能力在具体方言中的实现，即种种汉语语音现象，其实是继承描写主义的传统，实践一种新的实验描写主义。

参考文献：

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Abramson, A. S. (1962). The vowels and tones of Standard Thai: Acoustic measurements and experiments. *International Journal of American Linguistics*, 28(2), Part II; Also *Publication Twenty of the Indiana University Research Center in Anthropology, Folklore, and Linguistics*, Bloomington.
- Abramson, A. S. (1972). Tonal experiments with whispered Thai. In Valdamm, A. (ed.), *Papers on Linguistics and Phonetics to the Memory of Pierre Delattre*, pp. 31-44. The Hague: Mouton.
- Abramson, A. S. (1975). The tones of central Thai: some perceptual experiments. In Harris, J. G. & Chamberlain, J. (eds.), *Studies in Tai Linguistics*, pp. 1-16. Bangkok: Central Institute of English Language.
- Abramson, A. S. (1976). Thai tones as a reference system. In Gething, T. W., Harris, J. G. & Kullavanijaya, P. (eds.), *Tai Linguistics in Honor of Fang-Kuei Li*, pp. 1-12. Bangkok: Chulalongkorn University Press.
- Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, 21(4), 319-325.
- Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332: 346.
- Bao, Z. (1999). *The Structure of Tone*. Oxford: Oxford University Press.

- Boersma, P., Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 5.3.68, retrieved 20 March 2014 from <http://www.praat.org/>.
- Bouchard, K. E.; Mesgarani, N.; Johnson, K.; Chang, E. F. (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495 (7441), 327-32.
- Catford, I. (1977). *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.
- Chao, Y.-R. (1928). *Studies in the Modern Wu Dialects*. Peking: Tsinghua University Research Institute Monograph, 4.
- Chao, Y.-R. (1930). A System of Tone-Letters, *La Maitre Phonetique* 45, 24-47. Reprinted in *Fangyan*, 2, 81-82, 1980.
- Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.
- Fuchs, S., Pape, D., Petrone, C. & Perrier, P. (eds.) (2015). *Individual differences in speech production and perception*. Peterlang.
- 胡方 (2008) 《论元音产生中的舌运动机制-以宁波方言为例》, 载《中国语音学报》第 1 辑, 148-155 页, 北京: 商务印书馆。
- 胡方 (2013) 《降峰双元音是一个动态目标而升峰双元音是两个目标: 宁波方言双元音的声学 & 发音运动学特性》, 《语言研究集刊》第十辑, 12-37 页, 上海辞书出版社。
- 胡方 (2014) 《宁波话元音的语音学研究》, 北京: 中国社会科学出版社。
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review* 13 (3): 361–377.
- Gandour, J. T. (1978). The perception of tone. In Fromkin, V. A. (ed.), *Tone: A Linguistic Survey*, pp. 41-76. New York: Academic Press.
- Holbrook, A. and Fairbanks, G. (1962). Diphthong formants and their movements. *Journal of Speech and Hearing Research*, 5: 38-58.
- Hombert, J.-M. (1976a). Consonant types, vowel height, and tone in Yoruba. *UCLA Working Papers in Phonetics*, 33, 40-54.
- Hombert, J.-M. (1976b). Perception of tones of bisyllabic nouns in Yoruba. *Studies in African Linguistics, Supplement 6*, 109-121.
- Hombert, J.-M. (1978). Consonant types, vowel quality, and tone. In Fromkin, V. A. (ed.), *Tone: A Linguistic Survey*, pp. 77-111. New York: Academic Press.
- Jakobson, R. & Waugh, L. (1979). *The Sound Shape of Language*. Harvester Press. 96-97.
- Johnson, K. (2007). Decisions and Mechanisms in Exemplar-based Phonology. In Solé, M.J., Beddor, P. & Ohala, M. (eds) *Experimental Approaches to Phonology. In Honor of John Ohala*, pp. 25-40. Oxford University Press.
- Jones, D. (1922). *Outline of English phonetics (2nd Edition)*. New York: E. P. Dutton.
- Ladefoged, P. (2006). *A Course in Phonetics (5th ed.)*. Boston, M.A.: Thomson Wadsworth.
- Lehiste, I. & Peterson, G. E. (1961). Transitions, glides, and diphthongs. *Journal of the Acoustical Society of America*, 33: 268-277.
- 李如龙、张双庆 (主编) (1992) 《客赣方言调查报告》, 厦门大学出版社。
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review* 74 (6): 431–461.
- Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21 (1): 1–36.
- Liberman, A. M. & Mattingly, I. G. (1989). A specialization for speech perception. *Science* 243 (4890): 489–494.

- Liberman, A. M. & Whalen, D. H. (2000). On the relation of speech to language. *Trends in cognitive sciences* 4 (5): 187-196.
- 梁波 (2017) 《跨语言音姿对比的二语语音教学原则研究》，北京大学外国语学院博士论文。
- 刘纶鑫 (主编) (1999) 《客赣方言比较研究》，中国社会科学出版社。
- Malmberg, B. (1963). *Structural linguistics and human communication*. Berlin: Springer-Verlag.
- Mesgarani, Nima; Cheung, Connie; Johnson, Keith; Chang, Edward F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 28, 343 (6174), 1006-10.
- Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior*, pp. 178-203. Chicago: Chicago Linguistic Society.
- Ohala, J. J. (1991). The integration of phonetics and phonology. *Proceedings of the XIIth International Congress of Phonetic Sciences*, Vol. 1, pp. 1-16, Aix-en-Provence, 19-24 Aug 1991.
- Ohala, J. J. (1993). Sound change as nature's speech perception experiment. *Speech Communication*, 13, 155-161. [Also reprinted in: G. Fant, K. Hirose, and S. Kiritani (eds.) *Analysis, perception and processing of spoken language. Festschrift for Hiroya Fujisaki*, pp. 155-161. Amsterdam: Elsevier, 1996.]
- Ohala, J. J. (1995). Experimental phonology. In John A. Goldsmith (ed.), *A Handbook of Phonological Theory*, pp. 713-722. Oxford: Blackwell.
- Perkell, J. S. (1996). Properties of the tongue help to define vowel categories: hypotheses based on physiologically-oriented modeling. *Journal of Phonetics*, 24, 3-22.
- Pike, K. L. (1947). On the phonemic status of English diphthongs. *Language*, 23: 151-159.
- 平田昌司 (主编) (1998) 《徽州方言研究》，中国语学研究《开篇》单刊No.9，东京：好文出版。
- 邱玥、胡方 (2013) 《隆昌话的元音》，《语言研究集刊》第十辑，38-51页，上海辞书出版社。
- Shen, X-N. S., Lin, M. (1991). A Perceptual Study of Mandarin Tone 2 and 3. *Language and Speech*, 34 (2), 145-156.
- Sweet, H. (1877). *A handbook of phonetics including a popular exposition of the principles of spelling reform*. Oxford: Clarendon Press.
- Wang, C., Ding, Q., Tao, H. and Li, H. (2012). Comment On “Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa”, *Science* 335: 657.
- 叶晓锋 (2011) 《汉语方言语音的类型学研究》，复旦大学博士论文，44-47页。
- Yip, M. (1980). *The Tonal Phonology of Chinese*. Ph. D. dissertation. Massachusetts Institute of Technology.
- Zee, E. & Lee, W. (2007). Vowel Typology in Chinese. *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 6-10, Saarbrücken, Germany.
- Zhang, M. & Hu, F. (2015). Tone features in Qimen Hui Chinese dialect. *Proceedings of ICPHS 2015*, Glasgow, Scotland.
- 张倩 (2013) 《信丰（铁石口）客家方言的元音格局》，《语言研究集刊》第十辑，52-63页，上海辞书出版社。

【中文摘要】 汉语方言的语音研究中经常用到实验语音学，但却常常对实验语音学存在两大误解。其一，觉得实验语音学就是做测量，做统计分析的，将语言学范畴，比如声调，图解一下，好像就是实验语音学了；其二，觉得实验语音学就是研究偏僻音类的，好像只有偏僻、稀奇的音才需要实验研究，以补“口耳之缺”。

本文以汉语方言的语音研究为例讨论这个问题。我们认为，实验语音学并不是音类的简单图解，语音学测量与统计必须有语言学意义上的考量。简要地说，汉语方言的实验语音学研究就是以现在的语音科学实验手段探寻方言语音中的音类、音值区别的性质与规律：这些音的物理特性是什么？发音上是如何实现的？感知上有什么特点？而不是简单地将原来方言研究的音类画一个图，统计一下，验证一下记音的准确性；更不是为了寻找某些僻见的语音。我们认为汉语方言的实验语音学研究是语言学的一个独立的子学科，实验语音学测量所提供的语音细节为我们观察语音现象提供了新的方法论上的可能性，语音作为人类语言的物理载体具备了具体、变异的特性，不再像以往概念化似的那般抽象。语音的物理特性与语言学范畴之间并不存在简单对应的关系，而是交融着语音产生机制与语音感知特性的复杂映射关系。

【英文标题与提要】

What we talk about when we talk about phonetic measurements: towards an experimental approach to the phonetics and phonology in Chinese dialects.

【Abstract】 Speech research is now overwhelmingly experimental, while mainstream linguistics is either functional or formal in general. What is explicit is that people do physical or psychological measurements in experimental approaches; but what is implicit is that the philosophical base and methodology changed as well in experimental approaches. In this connection, the fundamental unit of speech/language, i.e. phoneme or distinctive features in traditional terminology, is not an abstract assumption any more as in generalization-based approaches, but an object that can be studied from perspectives of production and/or perception. This paper demonstrates the difference between an experimental approach and a generalization-based (either functional or formal) approach through the discussion on case studies on the phonetics and phonology of Chinese dialects.