对语调研究历史的回顾与综述

许 毅
（伦敦大学学院 言语、听觉及语音学系，伦敦 WC1N 1PF）

摘要： 本文通过对语调研究史的回顾来追寻我们对语调认知的发展。从这个回顾中，我们可以看到两个大的趋势，即从形式为先到功能为先的理论思路的转变和从抽象描述到参数生成的研究方法的演进。以这两个大趋势为主线，我们可以从海量的文献中理出头绪，使今后的研究能集中精力探索真正未知的领域，避免重复性的工作。在第一个大趋势里，以发现交际功能的编码方式为目标的研究已经揭示了诸如焦点、陈述 / 疑问、分组 / 标界、情感等功能的韵律编码规律。而从抽象描述到参数生成语调的发展趋势只是刚刚起步。学界对通过数字建模研究语调的投入还是很少，一是因为难度较大，更重要的是一般人觉得数字建模不是基础研究的主流。而且，由于参数型模型不可避免地要模拟表层音高曲线，容易让人担心它会牺牲认知理解所需要的抽象。本文通过对 PENTAtrainer 开发过程的回顾，显示了数字建模跟理论建模不仅不矛盾，反而是检验语调理论的有效手段。真正有效的理论抽象必须是能够复现表层细节的抽象。

关键词： 语调；声调；词重音；焦点；功能为先；形式为先；分组；标界
中图分类号： H014 **文献标志码：** A **文章编号：** 2097-7093(2026)04-0036-25

人类的话语不仅仅用构词造句来表达语义，还通过改变句子的音高模式来传达更多的意义，这些超词汇的音高模式就是狭义上的语调。但由于词的编码也常常用到音高，超词汇性的音高模式就不可避免地会跟词汇性音高模式交互融合，形成表层的语调；所以，语调研究很难脱离词调而单独进行。另外，除了音高，广义上的语调还可以包括音长、音强和嗓音型的变化。所以，本文综述的语调研究虽然是以讨论超词汇的音高变化模式为主，但也会涉及声调、词重音和边界标注等多个与韵律相关但不一定全靠音高编码的功能。语调研究已有极为丰富的历史，不可能在一篇文章里充分详述，而且中外都已有很多综合性论著(Botinis, Granström & Möbius, 2001[1]; Cruttenden, 1997[2], 2001[3]; Barnes & Shattuck-Hufnagel, 2022[4]; Xu, 2015[5], 2019[6])。所以，本文意在用有限的篇幅，通过追踪语调研究的历史走向，探讨几个重要的发展趋势，使今后的研究有所参照。

语调研究的发展趋势之一，是从形式为先逐步走向功能为先的理论思路；发展趋势之二，是从描述型理论逐步走向参数型理论。这两个发展趋势并不完全独立，而是在多个节点上有交集。另外，在这两个大趋势之下，还有许多具体方面的考虑，如生理、物理因素对语调的影响，语调的感知、习得、教学、病理，等等。这些语调研究的发展趋势并不完全孤立于语言学的其他方面，而是反映了整个语言科学发展的大趋势，即逐步从一个文、哲学科走向严格意义上的科学。下面每一节的讨论，都将首先从国际上的语调研究的发展说起，并以之为理论背景，再延伸到汉语语调研究的相应发展。

收稿日期： 2026-01-29
作者简介： 许 毅(1956—)，男，江苏南京人，教授，博士，主要研究方向为连续语流中的发音和感知的基本机制、声调、语调、韵律以及情感的语音表达等。

一、从形式为先到功能为先

所谓形式为先，就是直接用观察到的音高模式定义语调的成分和结构，然后再考虑这些成分和结构可能包含哪些意义。功能为先则是一开始就以交际功能为起点，逐个探讨每个潜在的功能是否和如何用音高模式来编码，以及多个功能如何同时编码，并且最终仍以功能来定义语调的范畴和成分。回看语调研究的历史，早期的理论几乎无一例外都是以形式为先的，这主要是因为语调研究的最大难点是缺乏参照系(Pierrehumbert 1980[7]，2000[8]；Xu，2015[5])，即无论普通人还是语言学家，都很难像对待词汇一样，对语调的基本单元作出肯定的判断。所以，早期的语调研究不得不首先通过听辨和基频的声学显示来观察各种音高模式，然后给他们各自命名、归类，如高、低、升、降，等等。随后就把这些模式认定为语调的单元，并把它们贯穿起来形成语调体系。在这样的语调体系的基础之上，再进一步探讨语调单元与交际意义的关联，这种看起来好像不可避免的思路，不仅贯穿了早期的主要语调研究，而且是到目前仍具影响力的几个最大语调学派的基础。

（一）形式为先的理论

1. 调核分析(nuclear tone analysis)

调核分析始于 Palmer（1922）[9]，后来经过 Kingdon（1958）[10]、O' Conner & Arnold（1961）[11]、Halliday（1967）[12]、Crystal（1969）[13]、Brazil（1980）[14]、Cruttenden（1997）[2]、Wells（2006）[15] 等人的探索，得到进一步发展，并被广泛应用于语言和外语教学(Nolan, 2022[16]; Romero-Trillo, 2012[17])。该学派认定，英语的语调首先是分为一个个语调短语(intonational phrase)，其中，每个语调短语都必须且只能有一个调核(nucleus)，即该语调短语里的最后一个带音高重音的音节(accented syllable)。调核必须落在重读音节(stressed syllable)上，并且音高有大幅度变化，所以要用像蝌蚪一样的符号来标注，蝌蚪的尾巴代表音高变化的方向。如图 1 所示，调核之前可以有调头(head)，涵盖第一个重读音节到调核前的所有音节；调头之前的非重读音节属于调前头(prehead)（Kingdon, 1958）[10]；调核之后的音高曲线则都属于调尾(tail)。

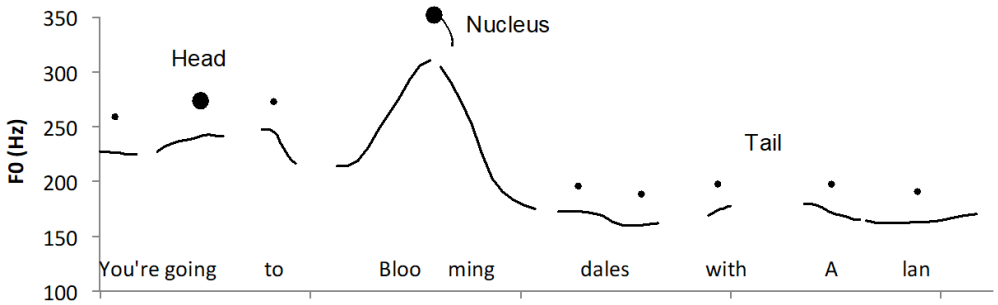

图 1　传统英国学派的调核分析法模型

2.AM 理论 / 语调音系学 /Pierrehumbert 模型 /ToBI

AM 理论的名称来自 Ladd（1996）[18]。在此之前，它主要以 Pierrehumbert 模型而知名，这是基于 Pierrehumbert（1980）的博士论文而形成的。这一学派最早是基于美国语调研究的传统。美国学派成长于音位学理论快速发展的时代，所以，其主流的语调理论从一开始就着力于建立一个音高音位(pitch phonemes)体系。其早期的系统定义了四个音高等级(Pike, 1945[19]; Trager and Smith, 1951[20])，由此建立的调位系统相当复杂。Pierrehumbert 则效法后来盛行的两分法，在博士论文里把美国英语的调位简化为只有高(H)和低(L)的两级对立。

虽然 Pierrehumbert 在博士论文里强调她的系统跟英国学派很不相同，但在比较了许多语调理论之后，我们不难看出这两个体系在许多方面很相像。首先，两者都认定每个语调短语都必有一个调核，或核心音高重音(nuclear accent)；不同的只是 AM 理论在调核之外还加了短语调(phrase accent)、边界调(boundary tone)等单元。其次，AM 理论用一个有限状态语法来限制音高重音的数量。最后，Pierrehumbert 还提出了用直线插值、下垂线插值以及调域调节、语音实现等一系列跟表层基频（f_0）曲线对接的机制。

AM 是 autosegmental–metrical(自主音段 – 节律) 的缩写，其中的节律部分是用来代表重音的，而重音决定了音高重音落在哪个音节上。不过，正如 Pierrehumbert 在其博士论文中所承认的，她对 AM 理论的节律部分如何跟音高部分结合以及 “metrical tree of the text” 是否绝对必要并不是很清楚。对这个问题的疑惑实际上已经表明，以形式为先的思路难以解释形式的来源。而来源问题不解决，理论的预测力就有限了。

Pierrehumbert(1980) 在博士论文里非常清晰地阐述了她如何实施形式为先的原则：“在已有文献里，判断哪些语调模式在语言学上相互对立，哪些是同一模式的变体，有两种方法。一种是根据观察到的 f_0 曲线特征，归纳出一个音系学表达体系；该体系归纳出来之后，下一步再比较这些音系学区别模式的用途。与之相反的方法是，首先辨识不同语调模式是转达相同还是不同的意义；然后再建立一个音系学体系，其中给相同的意义赋予同一底层表达形式，给不同的意义赋予不同的表达形式。……本论文采取的是第一种方法，并且只限于第一种方法的第一步。”[7]59 直到 10 年之后，Pierrehumbert & Hirschberg(1990)[21] 才提出一套方案，给音高重音赋予意义，使之变成语调语素。不过，这个方案并没有给已经定了型的 AM 体系带来任何结构上的变动；所以，AM 理论是典型的形式上一锤定音、意义上事后补救。

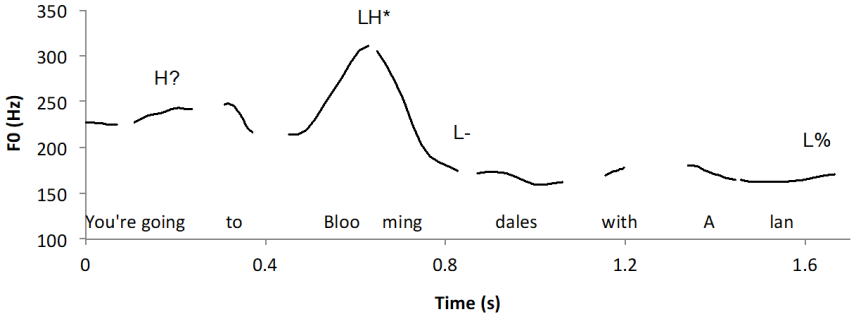

图 2　AM 理论对图 1 语例的语调标注

3. 其他符合形式为先原则的模型

另外还有许多模型，它们虽然没有像 AM 理论和调核分析法那样形成一整套有广泛影响力的音系系统，或在二语教学中得到广泛应用，但其基本思路也是形式为先的，那就是一开始先建立某种音高模式的描述系统，然后再试图把该描述系统跟功能和语义联系起来。它们描述的对象有的是音高转折点(Hirst, 2005[22], 2022[23])，有的是音高的峰(Taylor, 2000[24])，有的是转折点之间的连线('t Hart et al., 1990[25])，有的是音高峰的上限和音高谷的下限(Gårding, 1987[26])。其共同特点是，它们对描述对象的设定首先是基于描述的方便，其次才考虑背后可能的发音或感知原理。IPO 学派('t Hart et al., 1990[25]) 的模型是先把表层音高曲线简化为感知上可以接受的一段段直线，把这些直线连接起来就形成了格局模式，然后就可以按格局模式形成一套语调语法。Hirst et al.(2000)[27] 的语调系统首先用 MOMEL 模型来简化音高曲线，并通过二次条样函数(quadratic spline) 来平滑地衔接相邻的音高转折点(target point)；然后再用 INTSINT 来代表表层音系学层次，其里面的符号描述每个音高目标的位置相对于前一个音高目标的走向，或上升，或下降，或持平；最后，再把 INTSINT 层次的单元跟语调音系层次的单元对接，而语调音系层次又由 H、L、intonational unit(IU) 等一系列单元构成，只有这些 “底层” 的音系学单元才最后跟诸如强调、疑问、节奏等功能对接起来。Taylor(2000)[24] 的 Tilt 模型假设语调是由一个个音高峰链接而成，而该模型的任务就是把这些音高峰参数化。但是，Taylor 的研究最后并未完成如何将参数跟功能对应的工作。

4. 形式为先原则在汉语语调研究中的应用

由于汉语是声调语言，我们很容易在概念上把声调和语调分开，而且这种分离首先是基于功能而不是形式。但是，在研究具体问题时，我们有时还是躲不开形式为先的诱惑，其最主要的表现是容易集中注意于从形式上可直接观察到的现象。不过，常被忽略的不光是功能，还有发音机制，后者在下面会有更深入的讨论。其次就是常常受到国外主流理论的影响。例如沈炯(1994)[28] 就把英国学派的调核分析理论的调冠、调头、调核、调尾概念以及 Gårding(1987)[26] 的高、低音线概念都引入了汉语语调研究。

从上面的讨论可以看到，典型的形式为先的理论都是把表层形式直接定义为音系学单元，如调冠、调头、调核、调尾、音高重音、调核重音、边界调、音高上线、音高下线，等等。这些单元有确切的形式定义，但功能定义却相对模糊。换言之，在这个思路里，形式语调单元是第一性的实体，而与之相对应的交际功能则是第二性的；而且两者之间的对应常常是不直接的，例如焦点要通过一套复杂的投射规则才能跟音高重音联系起来（Ladd, 2008）[29]。

（二）功能为先的理论

功能为先的前提假设是，语言是一个传达意义的信号系统，而意义主要是通过多个交际功能来表达的，语音则是交际功能的编码；所以，语调是某些交际功能在音高维度上的编码。功能为先的思路就是以探索具体的交际功能为起点，逐个考察各种功能如何通过语调来编码；并在编码方式确定之后仍以交际功能来命名和定义语调的编码，而不是把编码用的音高模式组成一个自主的、可以脱离功能而存在的系统。这种功能为先的思路，实际上是符合语言学上已被普遍接受的音位原则的。那就是，考察一个语音系统，首先要确立哪些语音差别是能分词辨义的，哪些是无关紧要的，而那些能分词辨义的差别就叫音位差别。例如，在英语里，[f]/[v] 之差可以将 ferry 和 very 这两个词区别开来，所以，这两个辅音在英语里就被认为是不同的音位；相反，英语里的 /r/ 有许多变体，如 [ɹ][r] 甚至 [R]，它们在形式上的差别非常明显，但因为不能区别词，就被认为是同一音位的不同变体。所以，一个音段的音位归属取决于它能否用来分词，而不取决于它的形式。换言之，分词是交际功能，音位只是用来分词的语音编码而已。如果要把音位原则用于语调，就首先要问，有哪些跟分词类似的交际功能是用音高来编码的？不过，这并不是一个容易回答的问题。

1. 功能为先之难

许多有关语调的论述和研究其实都带有功能为先的色彩，如焦点语调、疑问语调、边界调，等等。但是，在研究过程中或构建理论时，它们总是难以摆脱形式为先的压力。回看语调研究的历史，可以发现做到功能为先的确有很多难处。

首先，在研究过程中，人们一旦观察到某种似乎有规律的形式，就会觉得有必要马上给它命名，以方便进一步的讨论。例如，调核分析和 AM 理论都是基于对一些功能语调的观察，尤其是焦点语调，但由于过于关注形式，一些研究者在观察到焦点处的一些显著的音高表现后，就立刻将之定为独立的语调单元，而不是顺藤摸瓜、全面考察焦点语调的整体表现，因此未能发现后面将要讨论的焦点后压缩现象。

其次，对表层音高曲线的视觉审视在相当程度上使语调研究更为困难，因为表层音高曲线跟底层交际功能的关系太不直接了（Xu, 2004a）[30]，所以很多探讨只好知难而退。例如前面谈到的 IPO 学派（'t Hart et al., 1990）[25]，虽然他们已经找到了许多感知上完全相似的直线音高模式，但是难以把它们跟具体的交际功能对应起来。Pierrehumbert（1980）[7] 干脆就把她建立的语调音系系统跟意义的关联推给未来的研究。Hirst（2005[22], 2022[23]）、Kohler（1991[31], 2005[32]）等体系一直非常强调交际功能的重要性，却无法建立一套能把底层功能跟表层音高曲线链接起来的系统。Taylor（2000）[24]、Fujisaki（1983）[33]、van Santen（2005）[34] 都建立了可以相当精确地生成表层音高曲线的模型，却也无法把生成曲线的参数跟具体的交际功能挂上钩。

再次，难以厘清词重音和语调的关系。一个经典的例子是，Fry（1958）[35] 用严格控制的实验发现，英语的音高是比时长和音强更有效的词重音感知音征，5 Hz 的音高差别就足以在感知上造成轻重音的清晰对立。然而，这篇文章后来多遭诟病，因为它未能解决词重音跟语调的关系。同样，由于没有把词重音、焦点和陈述 / 疑问的交互作用分开，Kochanski et al.（2005）[36] 居然得出结论说，重音与音高无关。

2. 考虑功能的理论

有一些语调模型或理论对交际功能有认真的考虑，并有不同程度的强调。Fujisaki 模型分了音高重音（pitch accent）和短语重音（phrase accent）两个层次，其中前者在日语里属于分词功能，但后者却被用来涵盖几乎所有超词汇的功能，所以它不能算完全功能为先的模型。还有 Hirst（2005[22], 2022[23]）的语调理论对凸显、分界等交际功能颇有强调，但在建立语调模型时，却无法把功能跟具体的音高曲线关联起来，他的 MOMEL 模型只管如何代表基频曲线。而 INTSINT 则是一个通用语调标注系统；然后，INTSINT 还要跟音系学单元

对应;最后,音系学单元才能跟交际功能对应。

Kohler(1991[31], 2005[32])提出的 Kiel model 在理论上非常强调交际功能,他把该模型的最终目标定为确定德语韵律的所有交际功能,并把它们与涵盖上下文及个人变体的区别性韵律特征联系起来。但是,该模型的核心部分是用对音高曲线的描述来搭建一个语调音系系统,其中的关键单元是音高重音(pitch accent),而与之对应的交际功能是凸显(prominence)而不是焦点(focus),这就涉及一个至今悬而未决的问题,即凸显跟焦点是什么关系?以及哪一个是真正的交际功能?或是两者都是交际功能。这在后面回顾研究具体功能的历史时会展开讨论。

3. 全面功能为先的理论

(1)SFC 模型

SFC(superposition of functional contours)是 Bailly & Holm(2005)[37] 提出的模型。其基本假设是,表层语调同时传达多层次的元语言学(metalinguistic)功能,其中包括:分段、分层、强调和态度。其每个功能都对应一条底层音高曲线,这些底层音高曲线叠加之后就形成了表层语调曲线;这些曲线的长度不一,但都以音节为单位;每个音节的曲线都由前、中、后三个音高值定义。但这些音高值都是通过机器学习从实际语音里提取的,而不是在模型里被赋予诸如高、低、升、降这样的形式定义。所以, SFC 模型里的韵律单元都是由功能定义的,而其形式值则来自实际的语音数据。

(2)PENTA 模型

PENTA 的全称是“平行编码及目标趋近”(Parallel Encoding and Target Approximation)(Xu, 2004b[38], 2004c[39], 2005[40]),与 SFC 相似, PENTA 也是一个全面功能为先的模型;但不同的是, PENTA 中还包含了目标趋近的发音机制。图 3 是 PENTA 的概念流程图。左起的第一个模块代表由语音传递的各种交际功能,这些交际功能之间的关系是平行的,即相互之间没有统领关系;第二个模块代表与各种交际功能对应的编码方案;这些编码方案规定了第三个模块里的发音参数,即音高目标参数;这些参数则控制着第四个模块所代表的目标趋近的生物力学发音过程,而该过程最终生成表层的连续音高曲线。可以看到,音高目标的各个参数都可以分别调节,而且每种调节对音高曲线都会有显著影响。但是,交际功能并不直接控制表层的声学特征,而是借助各自的编码方案来参与对音高目标的控制,并最终实现为表层音高曲线。音高目标的参数也是要通过机器学习从实际语音里提取,而不是在模型内部被赋予任何形式定义。所以,跟 SFC 模型一样, PENTA 里的韵律单元也都是由功能定义的,而其形式赋值则来自实际的语音数据。

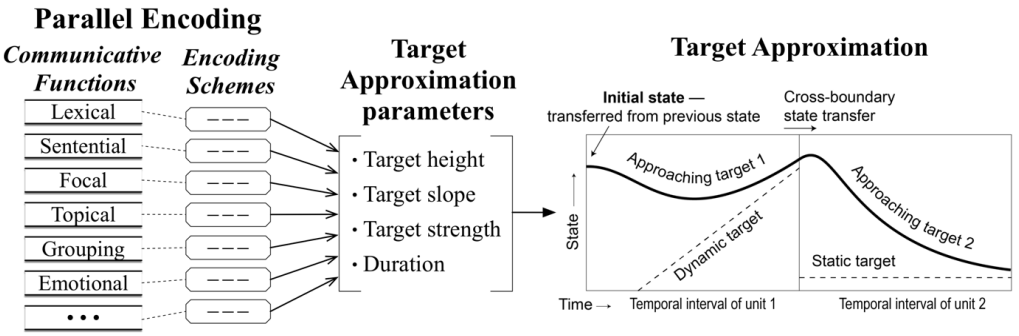

图 3　PENTA 模型

PENTA 模型提出后受到 AM 学派的多次批评。其中, Pierrehumbert(2022)[41] 在对 PENTA 的评价里提出了最本质上的质疑,即 PENTA 模型里没有包含一个类似于音系的层次。而音系层次的最大特点是:(1)意义跟形式相联的任意性;(2)形式单元可以重复使用。在第一点上, PENTA 并没有假设交际功能必须根据其内在的意义决定编码方式,所以两者之间的关联完全是任意的。对于第二点,关键的问题是,语言里究竟有没有可以重复使用的语调单元,如高、低、升、降,等等。Pierrehumbert 认为是必须有的,可以以音段音位为证。因为任何语言里用元音、辅音(或声调)区分的单词有成千上万,甚至数十万,所以只有把元、辅音范畴化才能把分词的编码限制在几十个,这样才易于儿童习得。但是,这个论点刚好说明了语调和词汇的

不同。首先，虽然音段音位要区分的词有数十万之多，但它们都同属于一个单一的功能，即分词。同一原理如果扩展到语调，譬如焦点这个功能，它里头又有多少个性质等同的单元呢？顶多有焦点前、焦点处、焦点后三个区段，每处的音高模式又各不一样（详见后述）。更为重要的是，用于焦点编码的音高模式是不能重复用在其他功能上的。例如，我们不可能把焦点处的高 H 重复用在分词上去代表声调里的 H，因为两者不仅功能性质不同，实际的音高值也很不一样。到目前为止，我们还没有发现任何一个音高单元是可以跨语调功能重复使用的，也没有发现在同一语调功能内可以用同一值来代表的不同单元。

Arvaniti & Ladd（2009）[42]、Arvaniti（2022）[43] 则一再指责 PENTA 的致命错误是把交际功能跟表层音高曲线直接挂钩。图 3 就明确显示，交际功能跟表层音高曲线的联系是非常不直接的，因为它至少要跨过平行编码、编码方案、目标趋近三个环节（Xu, 2004a）[30]；所以，任何表层曲线都不可能跟某一特定的交际功能或音系学单元直接对应。相反，在 AM 理论里，音高曲线的峰和谷反而是跟有表层值的音系单元 H 和 L 直接对应的（Pierrehumbert, 1980）[7]。

4. 对汉语语调功能的探讨

由于汉语是声调语言，几乎所有的汉语语调模型都必须考虑如何处理声调跟语调的关系。这种前端分离，正如前边已说过的，本身就已从概念上把分词功能跟超词汇的语调分离开来了。而基于非声调语言的模型，包括调核分析和 AM 理论，大多是把分词功能跟语调功能混在一起的，所以才需要一些特殊规定，如音高重音只能落在重读音节上。其后果是，调核分析和 AM 理论一类的模型几乎无法用于声调语言。因为像 H、L 这样的音系学单元，一旦用来代表声调，就很难再代表音高重音。由于调核和音高重音在这类理论里被等同于焦点，它们也就难以用于描述声调语言里的焦点语调。

所以，几乎所有的汉语语调模型都至少是部分功能为先的，即首先把声调和语调这两个功能分开，认为语调是在基本保证声调曲线的前提下对音高作进一步调整的产物。如 Chang（1958）[44] 就提出成都方言的语调凌驾于整个句子之上，并对声调做出调节。Chao（1968）[45] 还考虑到了表情语调、句终语调。沈炯（1985[46], 1994[28]）则从功能上把句重音分成语势重音、强调重音和对比重音，把句型语调分为陈述、疑问、感叹、祈使等类型。石锋（2021）[47] 的韵律格局理论则用焦点、句型等功能为语调分类。

虽然语调研究已广泛意识到功能的重要性，但很少有像 SFC 和 PENTA 模型那样系统地用功能为语调模式分类，并有一套算法跟表层音高曲线最终联系起来的。究其原因，主要是缺乏参数化的语调模型。而语调的参数化则是语调研究史上的另一条主线，它更代表未来的发展趋势。

二、从抽象描述到参数生成

所谓描述型理论，就是以标注的方式对音高模式做分类、命名；而参数型理论则是通过数字建模，用连续性参数导出实际的音高曲线。跟其他许多科学领域一样，早期的语调研究由于技术手段有限，不可避免地要先从描述性工作开始，而且只能是口耳之学。所以，给观察到的语调现象命名几乎是必不可少的步骤。后来随着技术的发展，可以用仪器直接测量音高，并能把连续的音高曲线做成图来仔细考察。尽管如此，许多理论还是倾向于把测量到的音高转记为有命名的标注，如高、低、中，或 5 度值，等等。有的讨论甚至把有无标注系统作为衡量一个理论优劣的重要尺度，以至试图将参数模拟音高曲线的模型作为理论低劣的范例。如 Arvaniti（2022）[43] 就对 Fujisaki 模型和 PENTA 模型有这样的评价："那些模型的重点是忠实地再现音高曲线，例如 Xu 的 PENTA（Xu 2005[40]）和 Fujisaki 模型（Fujisaki 2004[48]）。这样的模型在 f_0 建模上可以非常成功，但在语调建模上就不一定很成功，因为它们不能提供对语调的原则性描述，只能对它们作事后模拟。" 尤其是语调音系学的兴盛导致在相当长的一段时间内形成了一种视参数化为有害的学术气氛，以至于 Taylor（2000）[24]1710 在发表 Tilt 模型时不得不为自己辩护："我们觉得用连续参数音系代表音系单元是完全合适的。"

实际上，任何一个学科在走向成熟时都会变得更加数字化，语调研究也不会例外，关键是如何实现有理论意义的数字化。回看已有的各种语调参数化方案，发现其难点在于如何做到既保证对音高曲线的精确模拟，又让模型具有预测能力。在这个大方向下，我们可以总结出以下三个具体发展趋势：一是从概念参数式

到数字参数式模型；二是从模拟个例音高曲线到预测性语调的生成；三是从直接模拟音高曲线到首先模拟发音机制。

（一）从概念参数到数字参数

很多语调模型都在概念上试图将语调参数化，但是并没有让这些概念形成能够生成具体音高曲线的数字系统。赵元任（1933）[49]就提出，句子中的声调是两个因素的代数和，一个因素是本字调，一个是句调本身。他（Chao，1968）[45]提出“可将音节的声调跟句子的语调比作小波浪跨在大波浪上面。实际结果是两种波浪的代数和。正加正数值增大，正加负则减少”。他在《国语语调》（1935）[50]中还提出用“橡皮带效应”解释句子调域和情感对声调的影响。Gårding（1993）[51]的上、下线和调域转折点的概念也是试图描述音高曲线的具体高度，而不是把它们转注为范畴型单元。

另外，Thorsen（1980）[52]的丹麦语语调模型也是一个在概念上把语调参数化的模型。她把语调分为词调和句调两个层次，词调的音高取决于所在音节是重音、非重音，还是非重音单音节词，句调分为无标记疑问句、倒装词序疑问句、陈述句，表层音高曲线是由词调和句调叠加而成的。

沈炯（1994）[28]提出了“声调音域”方案：“最好把声调音域分成两个音区，即高音区和低音区。用这两个音区来说明的声调特征就是高音特征和低音特征。它们的实际音高用高音点和低音点的实测数据来表示。声调音域的变化可以从这些对偶关系里反映出来。在许多音节组成的语句里，把这些高音点和低音点分别连接起来，就得到全句的高音线和低音线。这两条音高极限构成了语句的橡皮带，汉语语调就是借助它们表现出来的音高形式。音高边缘线的错位，使相邻声调音域之间出现音高断层。高音线的突变跟语势重音有关，节奏重音则主要靠低音线下移来形成。语势重音和节奏重音对边缘线的调节，是相邻音节之间的事。语调是在它们基础上的全句性调节。”

吴宗济（1997）[53]从声调和乐律的关系提出用移调规则来处理普通话语调，即“可以先根据一位发音人的平叙句的平均基调（根据此发音人平叙句语调统计定值，相当于作曲的旋律基调），设定一系列短语变调模式。然后以此作为基础，对语句中某些需要着重或减轻的短语调群，增减其调值，使其达到目标语调的要求。这样，一切语调就都可以从一个平叙句的基础调型来‘生成’了。在一般语境下，说话人的语调上限可由意识的语气强弱来定，而下限则是由不自觉的本语系的音系规则来安排的。这是以调阈上限为基调的说法；如果以下限为基调，则上限也可由音系规则来安排。换言之，也就是在通常情况下，调域高度由语气强弱决定，而调域宽度则由音系规则安排。”

以上这些模型虽然提出了各种方案，但并没有把它们付诸实施，即形成一套数字操作系统，并从语料中提取参数，然后用提取的参数通过模型生成表层音高曲线，再把生成的曲线或者跟实际曲线相比，得出吻合度，并用于语音合成，从而可以让人来判断其清晰度和自然度。

（二）从模拟个例音高曲线到预测性语调的生成

说到语调的数字模型，就不可避免地涉及是个例模拟，还是预测性语调生成的问题。所谓个例模拟，就是用某种参数系统去优化匹配一个个具体的音高曲线，也叫再合成（resynthesis）；而预测性语调生成则是用事先训练出的参数预测新的音高曲线，不再经过重新优化匹配。当然，任何参数语调模型都至少要能模拟个例音高曲线，因为如果连这个都做不到，预测性语调的生成也就无从谈起了。

模拟个例音高曲线的最简单的方法，就是减少原始音高曲线里的点数，然后把剩余的点用直线连起来。前面已经介绍过的IPO模型用的就是这种方法。这种方法也以Stylize pitch为名内装在Praat软件（Boersma，2001）[54]里，作为一种音高分析的方法。由于只是对原始音高曲线的精简再现，这样的模拟只有最初级的预测力。预测力略高一步的是多项式模拟。在数学上，任何一条复杂的曲线都可以由一个足够阶数的多项式函数完美地表达，该多项式的参数就可以视为那条音高曲线的预测参数。除此之外，我们还可以用较低阶数的分段多项式来模拟音高曲线，这已用于很多语调或声调研究（Andruski & Costello，2004[55]；Gandour，Tumtavitikul & Sattharnnuwong，1999[56]；Grabe，Kochanski & Coleman，2007[57]；Hirst，2005[22]；Liu，Surendran & Xu，2006[58]）。但跟Stylize pitch一样，从模拟个例音高曲线得到的多项式参数，很难用来模拟其他句子的

音高曲线，所以其预测力也很有限。

另外，Pierrehumbert（1981）[59] 提出可以用直线或悬垂抛物线（sagging parabolic line）来连接标注为 H 或 L 的音高峰或谷，这样就可以用很少的具有音系学意义的参数来模拟个例的音高曲线。但在这种模拟中，唯一的预测成分是连接的直线或抛物线，因为峰和谷的值都直接来自原始音高曲线，所以没有任何预测，这恰恰符合 Arvaniti（2022）[43] 所批评的“事后模拟”。可惜的是，在 AM 理论学派里，没有人继续研究如何增进该模型的预测力。反倒是 Lee & Xu（2014）[60] 和 Lau & Xu（2019）[61] 根据 Pierrehumbert（1981）[59] 用 Praat script 为 AM 理论建模，发现它可以达到一定的预测能力，但其精确性随着预测范围的增加而降低，而且总精确性低于 PENTAtrainer（Lau & Xu, 2019）[61]。Tilt 模型（Taylor, 2000）[24] 也试图用很少几个参数模拟音高曲线上的尖峰，然后用直线连接相邻的峰。但 Sun（2002）[62] 通过测试发现，虽然 Tilt 模型可以做大规模的预测，但其精确度和感知自然度都低于更简单的三点模型（Black & Hunt, 1996）[63]。

（三）线性模型与非线性模型

以上的直接模拟音高曲线的模型有一个共同特点，就是都属于线性模型。所谓线性模型，就是假定每个时间段都只有一个语调单元，所以模型只需做到直接分段模拟音高曲线。与线性模型不同的是叠加式模型，它们假设语调是由不同层次的单元叠加而成的。赵元任的大波浪加小波浪说就是一个概念上的叠加模型。另外，Thorsen（1980）[52] 的丹麦语语调模型也是一个概念式的叠加模型。而最广为人知的数字叠加模型则是 Fujisaki（藤崎博也）模型（Fujisaki, 1983[33], 1988[64]）。它把语调分为音高重音命令（accent command）和短语命令（phrase command）两个层次，它们各自生成一条连续的音高曲线，然后两者相加形成最表层的音高曲线。其他的叠加式参数模型还有 SFC 模型（Bailly & Holm, 2005[37]）和多层次单元序列模型（multi–level unit sequences）（van Santen et al., 2005[34]）。

一般说来，线性模型用来体现形式为先的理论比较容易，因为其基本前提也是表层音高曲线跟语调单元的直接对应。但是，用线性模型来体现功能为先的理论却很难。例如，迄今为止没有一个线性模型能够很好地模拟声调和语调对表层音高曲线的共同影响。即使对于非声调语言，多重交际功能的影响也只能用模型之外的机制来模拟。例如，AM 模型虽然也认为语调音系单元要受调域、基线、重音层级结构等功能因素的多重调节，但并没有把这些因素融入参数模型的核心部分。

叠加模型的前提是认定表层音高曲线来源于共时存在的多层次的音高曲线，所以容易在理论上用不同层次分别代表不同的功能。不过在实际建模时并不那么简单，其中一个重要分歧是要不要在建模时模拟发音机制以及如何模拟。

（四）对发音机制的模拟

语调建模是否要模拟发音机制首先取决于它对音高曲线的影响到底有多大。如果影响不大，就可以忽略不计，但如果影响很大就有必要认真考虑。已有的研究发现，相对于音节，发音时改变音高的速度是很慢的（Ohala & Ewan, 1973[65]; Sundberg, 1979[66]; Xu & Sun, 2002[67]），即使最小的音高变化也需要约 100 毫秒的时间（Xu & Sun, 2002[67]），而音节在语流中的平均时长为 143–200 毫秒（即每秒 5–7 个音节）。所以，即使两个音高单元的高度之差非常小，也至少需要半个音节的时长才能实现两者之间的过渡，而更大的幅度需要的过渡段就更长了。因此，表层音高曲线其实大都是由音高单元之间的过渡构成的。这意味着发音机制不仅不能忽略不计，反而是语调建模的一个关键。

不过，模拟发音机制绝非容易，认真尝试的努力非常之少，到目前为止提出的可运行的模型只有 Fujisaki（1983）[33]、Stem–ML（Kochanski & Shih, 2003）[68] 和 qTA/PENTA（Prom–on et al., 2009[69]; Xu & Prom–on, 2014[70]）。模拟发音机制之难，从 Fujisaki 模型的测试经历上可略见一斑。Tilt 模型（Taylor, 2000）[24] 和 SFC 模型（Bailly & Holm, 2005）[37] 的作者，都是在对 Fujisaki 模型做实际测试时遇到了难以克服的困难（Bailly, 1989[71]; Taylor, 1994[72]），才决定自己另建不包含发音机制的模型。而 Fujisaki 对自己模型的测试，也一直是通过手工调试个例句子的命令参数（Fujisaki, 1983[33], 1988[64]），所以效率极低。Mixdorff（2000）[73] 提出了一种自动提取 Fujisaki 模型参数的算法，但 Raidt et al.（2004）[74] 发现，用它合成语调还不如没有发

音机制的 SFC 模型好。由此可见，开发有充分预测能力的发音机制模型并非易事。

模拟发音机制之难在于需要考虑音高控制的几个关键方面，缺一不可：（1）基本发音原理；（2）底层单元；（3）基本生理特征；（4）控制的时间范围；（5）非核心的机制。

关于基本发音原理，Fujisaki、Stem–ML 和 qTA 有相似的假设，即由于音高的改变不可能在瞬间实现，所以，表层音高曲线不可能是底层音高目标，而是在接近底层目标过程中生成的音高变化轨迹。因此，模型设计中既要有底层音高目标，也要有接近目标的机制；而且，对目标的形式和时域、接近目标的方式和方向也要有规定，表 1 列出了三个模型在这些方面的同异。

表 1　Fujisaki、qTA、Stem–ML 模型的比较

	Fujisak	qTA	Stem–ML
目标形式	水平音高	带斜度直线音高	复杂音高曲线
目标时域	半音节、重音、短语	音节	短语
目标层次	双层次	单层次	单层次
接近机制	趋近 + 离去	趋近	最大平滑
接近方向	从左至右	从左至右	左右双向

由表 1 可以看出，Fujisaki 模型只有水平音高目标，因此，如果声调是动态的，如普通话的阳平或去声，就需要由两个高低不同的水平目标来构成（见图 4，Fujisaki et al., 2005）[75]；qTA 模型的目标形式是带有不同斜度的单一线性目标，其依据是，动态声调最稳定的特征为调尾速率(Xu, 1998[76], 2001[77])，跟双元音的动态特征相类似(Gay, 1968[78]; Xu A 2025[79])；而 Stem–ML 模型的目标形式是复杂音高曲线。

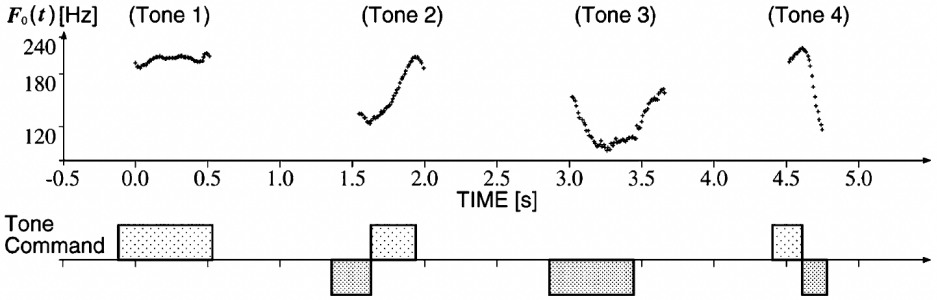

图 4　Fujisaki 模型如何用音高重音命令模拟普通话声调（略去图例说明）

在目标时域上，Fujisaki 和 Stem–ML 模型都允许灵活的时域。Fujisaki 模型在“小波浪”层次里的每个目标，或对应于整个音高重音，或对应于比音节还小的单元，所以，一个声调可以有 1–2 个命令，而且各命令所占时间也是灵活的；在“大波浪”层次里的时域也完全是灵活的，只是每个时域都很长。同样，Stem–ML 模型里的复杂音高曲线的时域也是完全灵活的。目标时域灵活的好处是在个例曲线匹配时容易达到最优化，但弱点是自由度太高，增大了实现预测的难度。在 qTA 里，目标时域则是固定的，它是包括声母和韵尾的整体音节。以音节为目标时域的依据，一方面来自对声调的系统考察(Xu, 1997[80], 1999[81])；另一方面来自音节的声、韵、调同步假说(Xu & Liu, 2006[82]; Xu, 2025[83])，即音节的本质就是声、韵、调同时开始向各自目标趋近的动作，以使大脑对发音动作的控制成为可能。所以，每个音节都必须有自己的音高目标，无论该语言是否有声调。

在目标层次上，Fujisaki 模型分音高重音和短语两个层次，而 qTA 和 Stem–ML 都只有一个层次。双层次的好处在于可以同时表达两个功能，但缺点是自由度太大，而且与模拟发音机制的本意相矛盾。目标趋近之所以需要时间，是因为在发音时要克服物理和生理的惯性。但惯性不只是作用于产生音高的喉头，也作用于产生元、辅音的声道，所以不可能先发一个没有音高的元、辅音音节，再先后克服两种不同的喉头惯性生成两条底层曲线，然后再把两者相加，并把叠加好的音高曲线附加在音节上。而 qTA 模型的假设是，每个音节必须且只能有一个音高目标，而这个目标的趋近跟整个音节的发音完全同步。

在目标接近机制上，Fujisaki 模型的每一个命令既包括趋近目标的动作，也包括离目标而去的动作。这等于假定目标趋近往往是达标的，所以还有剩余的时间离开目标。但已有的研究表明，目标趋近经常因为时间不够而不能达标(Xu & Prom-on, 2019[84]; Xu & Sun, 2002[67])，所以没有剩余时间做离开目标的动作。而且，即使有充分的时间，发音动作也会主动放慢，以防止超过目标(Xu, 1998[85], 2025[83]; Xu & Prom-on, 2019[84])。所以，qTA 的假设是，在目标时域内，只有趋近的动作，并且常常不达标。Stem-ML 模型的基本机制是目标曲线上的每一个点都受到相邻目标的同化作用，其作用的大小取决于该点跟相邻目标的距离，距离越近，同化越重(Kochanski & Shih, 2000)[86]。

最后，在目标接近的方向上，Fujisaki 和 qTA 模型里都只有从左至右的目标趋近，而 Stem-ML 模型则假定有双向的同化。这种双向同化的发音机制并不清晰，因为很难想象喉头肌肉可以同时朝两个相邻音高目标分别发力，也难以想象大脑每时每刻都在计算目标里的每一个点离相邻目标里的各个点有多远。

从以上介绍可以看到，在这三个模型里，qTA 的原理相对简单，自由度也较低：目标时域固定于音节，每个音节只有一个带斜度的直线音高目标，目标的实现只靠单向的趋近，目标之间既无缝隙，也无交错，而且只有一个层次的音高目标，所以，qTA 的数字模型(qTAtrainer)很容易就实现了大规模自动提取个例音高的目标参数(Prom-on et al., 2009)[69]。不过，qTA 只解决了个例音高曲线的匹配，而更重要的是解决音高曲线如何同时表达多个交际功能的问题。PENTA(平行编码目标趋近)模型就是为了解决这个问题而提出的。

(五)PENTA：含发音机制的多功能平行编码模型

前面已经初步介绍过的 PENTA(图 3)，最早是以概念式模型提出的(Xu, 2004b[38], 2004c[39], 2005[40])。它以每个音节在发音时只能有单一音高目标为前提，试图解决如何用单一音高目标同时代表多个交际功能；其不同交际功能分别对应一套编码方案，每个方案都对以音节为单位的音高目标施加影响；最后，受到多功能影响的音高目标由目标趋近模型生成表层音高曲线。由于 qTA 模型的提出(Prom-on, et al., 2009)[69]，PENTA 得以形成参数化的模型(Xu & Prom-on, 2014)[70]，并且发展出以 Praat 脚本为形式的 PENTAtrainer 软件①。

参数化的 PENTAtrainer 最重要的是解决了一个概念版模型没有厘清的要点，即如何实现序列式编码。首先，跟 SFC 相似，每个交际功能都有自己明确的管辖时域和内部分区。但跟 SFC 不同的是，不管该功能外部时域的大小以及内部如何分区，其影响力最后都要分别落实到一个个以音节为单位的音高目标上。所以，如图 5 所示，我们需要从语料中提取的只是一系列多功能的音节音高目标(图 5 底部)。[70] 在合成语调时，我们只需知道每个音节载带的功能是什么，就可以启用该多功能音高目标为合成参数，实现预测性语调合成。PENTAtrainer 目前已用于英语(Liu et al., 2013)[87]、日语(Lee & Xu, 2015[88]; Lee, Xu & Prom-on, 2014[60])、波斯语(Taheri-Ardali & Xu, 2015)[89]、阿拉伯语(Alzaidi et al., 2023)[90] 和萨沃萨沃语(Savosavo)(Lee et al., 2023)[91]。

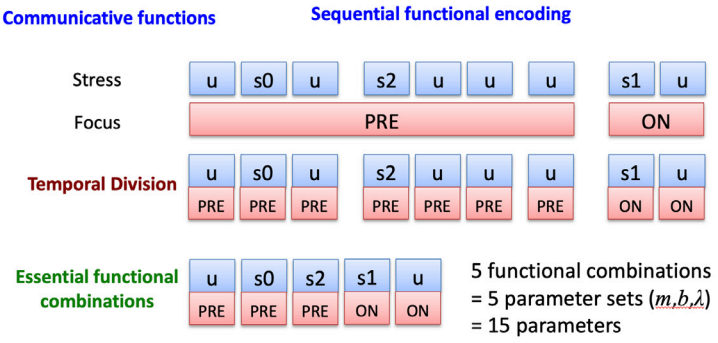

图 5　PENTAtrainer 如何把多层次的交际功能落实在以音节为单位的音高目标上

虽然 PENTAtrainer 已可以用有功能标注的语料库学习和预测语调，但还无法做到没有功能标注也能学

① http://www.homepages.ucl.ac.uk/~uclyyix/PENTAtrainer/.

习功能语调参数。其关键的难点还是本文开始就提到的缺乏参照系的问题。如果语料库的文本标注不包括语调功能，就无法用语料库来训练语调模型。所以，如何实现自动标注语调功能将会是使语调建模更上一层楼的一个主要课题。

三、几个主要交际功能的语调编码

以上回顾的是语调研究史上的几个大趋势，其中一个主要趋势就是从形式为先走向功能为先。本节将总结两个已被确认是以音高模式为主要编码手段的交际功能，即焦点和陈述 / 疑问；以及两个不主要用音高编码的交际功能，即分组 / 标界和情感。之所以只总结这些功能，是因为学界对它们已有较多研究，并有较可靠的发现。还有很多交际功能的编码研究还不够多，这里只能简单提一下，如话题、插入信息以及英语的反驳语调，等等。

（一）焦点

1. 焦点的定义

焦点研究的历史进展十分曲折。首先，在定义问题上至今尚无共识。从功能上看，跟焦点对应的概念是强调(emphasis)，即说话人在语句中刻意彰显的某个部分；但是，焦点常常会跟另一个被广为接受的概念混淆，即凸显(prominent)，因为后者能同时涵盖词重音，并外加其他一些诸如非重读音节等现象。但是，凸显会把焦点的功能模糊化，如词重音的功能是辨词，跟强调很不一样，而其他音节弱读的现象也大多跟强调无关。有一种流行的概念是，焦点的功能是标记新信息，这就意味着所有载带新信息的成分都应该是焦点，这当然是讲不通的。更不用说，正如 Krifka(2008)[92]、Ladd(2008)[29] 和 Terken(1984)[93] 已指出的，旧信息也一样可以成为焦点。例如，“是张三还是李四得了第一？”“张三得了第一。”这里的“张三”在问话中已经被提到了，所以不是新信息，但仍会被自然地强调。还有一种颇为流行的基于语义学的定义是，焦点是为了排除在理解上可能会出现的其他选项(Krifka, 2008[92]; Rooth, 1992[94])。例如在说“张三得了第一”时，如果强调“张三”就是为了排除“李四、王五”，如果强调“第一”则是为了排除“第二、第三”。这种定义实际上是把焦点跟辨词的功能混为一谈了。例如在“张三得了第一”中，张三作为一个独特的词，本身就已经排除了其他可能的词：张三 ≠ 李四、王五、赵六，等等，所以，强调某个词的真正功能不可能是为了再进一步排除句中同一位置上的其他选项。更可能的是，强调是为了让听者把注意力放在“张三”上，而不是放在“得了”或“第一”上。所以，排除的对象是同一句中其他位置上的词。对此已有一些实验的证据。Chen et al.(2012)[95]、Chen et al.(2014)[96] 和 Kristensen et al.(2013)[97] 通过 fMRI、ERP 和眼动实验发现，信息的新旧跟记忆提取有关，而焦点的作用则是在理解语句时对注意力的分配。所以，焦点的功能很可能是为了让听者额外注意被强调的信息。

一个完善的焦点理论应该能够预测什么时候、什么信息是有必要强调的，即焦点的确切落点在哪里。对此，到目前为止还没有理论能够做到。不过，实证研究已掌握了一些保证可以引导确切焦点的语境，最常见的就是上面例子中的特殊疑问句。除此之外，还可以用前文给出错误信息，然后引出后文用焦点否定。如 A：“张三去北京了。”B：“不对，李四去北京了。”有的语法结构或特定的词，如“是”“只有”“仅仅”等，也往往会引出焦点，但并不能决定焦点的精确落点。如“张三是跟李四去的北京”，焦点既可以落在“是”上，也可以落在“李四”上，甚至可以落在“跟”上。其真正的决定因素尚不清楚，所以仍然是焦点研究的一个具体难题。

还有一种颇有影响力的理论是把焦点作为信息结构 (information structure) 的一部分，该结构还包括已知信息(givenness) 和话题(topic)。信息结构的概念来自 Halliday(1967)[98]，后来又有 Chafe(1976)[99] 等的进一步发展。前面讲到的选项理论(Krifka, 2008[92]; Rooth, 1992[94]) 也和信息结构密切相关，所以它们也有共同的问题，即把焦点载带的新旧信息跟词汇载带的分词信息混淆了。另外，由于信息结构号称是一个自成一体的结构，这就意味着要把焦点这个本来就足够复杂的概念再用一个更加模糊的概念涵盖起来，这在理论上显不出明确的优越性，对实证研究也看不出会有多大的帮助。

2. 焦点的语调编码

焦点语调的研究历程可以说更为曲折。有意思的是，最传统的英国学派的调核分析理论从一开始就是围绕着焦点语调建立的。Palmer（2022）[9] 开篇就指出，英语中的强调主要是用音高模式来表达的，而调核就是一个语调群里最为凸显的重读音节，即一般所说的句重音。但由于该理论以形式为先，调核这个概念同时涵盖了词重音、句型等功能，所以，调核并不完全与焦点等同。同理，AM 理论的核心音高重音（nuclear accent）也跟焦点部分地对应，但也因为又跟词重音混在一起，所以也不能完全等同于焦点。

直接从功能出发的焦点语调研究，最经典的应该是 Cooper 和 Eady 在 20 世纪 80 年代中期三连发的文章（Cooper et al., 1985[100]; Eady & Cooper, 1986[101]; Eady et al., 1986[102]）。他们通过用特殊疑问句引导焦点产出的实验发现，在美国英语里，焦点在陈述句里不仅抬高了焦点处的音高峰值，而且压低了焦点后的音高峰值，但对焦点前的音高峰值没有明确影响。其中的第二条，后来被称为焦点后压缩（post-focus compression—PFC）①，是最重要的音征。Xu & Xu（2005）[103] 在对美国英语焦点的音高曲线做了详细的分析后，核实了 Cooper 和 Eady 描述的三段式焦点编码，并进一步揭示了焦点和词重音各自用不同的音征独立编码。Xu（1999）[81] 发现，焦点在普通话里用的也是和英语一样的三段式编码模式，即焦点前调域不变，焦点处调域加宽，焦点后调域下压。在此之前，沈炯（1985[46], 1994[28]）的研究已指出焦点处有高音线骤落，但没有明确提出焦点后压缩这个音征。随后的多项研究进一步核实了焦点在普通话里的三段式编码（S. Chen et al., 2009[104]; Y. Chen et al., 2014[105]; Wang & Xu, 2011[106]; Wang et al., 2018[107]）。

后来的研究进一步发现，焦点后压缩不只发生在英语和普通话里，也不只是散落在互不相关的不同语言里，而是跨语族的。其所涉及的语言包括印欧、阿尔泰、乌拉尔、达罗毗荼等横跨欧亚大陆的大多数语系（Xu, 2019）[108]。但是，焦点后压缩在汉藏和亚非语系里是两分的：偏北方的语言有，偏南方的语言没有。如在汉语里，多数北方方言都有，但南方方言，如粤语（Wu & Xu, 2010）[109]、闽南话（S. Chen et al., 2009）[104]、重庆话（Qin & Xu, 2020）[110] 就没有。在亚非语系里，阿拉伯语（Alzaidi et al., 2019[111], 2023[90]）、希伯来语（Amir et al., 2023）[112] 里有，但北非地区偏南的语言（Zerbian et al., 2010）[113] 就没有。有意思的是，焦点后压缩的这种跨语族的分布跟 Nostratic Macrofamily 假说（Bomhard, 2008）[114] 和语言随农业扩散假说（Diamond & Bellwood, 2003）[115] 有所相似。虽然汉藏语系没有被包括在 Nostratic 超语系里，但是中国北部的大片区域在历史上是属于阿尔泰语系所在地的。所以，Xu（2011[116], 2019[108]）、Xu et al.（2012）[117] 提出，焦点后压缩源自 Nostratic 超语系假说里的原始语所在地，即新月沃地——大、小麦农业发源地，至少有 1 万 3000 年的历史。而北方汉语和藏语里之所以有焦点后压缩，是因为他们跟印欧语系有共同祖先，因为该音征只有语言内部的跨代习得才可能保持，它是无法跨语言传播的（Chen, 2015[118]; Chen et al., 2014[119]; Xu et al., 2012[117]; Wu & Chung, 2011[120]）。而南方汉语方言没有焦点后压缩，可能是由于当地汉语最初来自北方派来的官员和军队造成的语言替换，当地人是作为二语来学习汉语的，因此无法学会焦点后压缩（Chen, 2015[118]; Chen et al., 2014[119]）。不过，最新的一项研究又发现了焦点后压缩在特定条件下跨语言传播的可能性。Yang（2022）[121] 发现，一些从内地移居香港多年的人说的粤语表现出了焦点后压缩，于是提出了一种新的可能，即焦点后压缩的起源仍然只有一个，但传播的方式也不排除特定条件下的跨语言传播。焦点后压缩的单一起源假说看起来有些过于离谱，因为历史语言学很少能把语言间的关联追踪到 1 万年前（Renfrew et al., 2000）[122]。但是，到目前为止尚未有足以证伪的反例，相反却有一再出现的新例证（刘璐、王蓓、李雪巧，2020[123]；王玲 et al., 2011[124]; Alzaidi, Xu & Xu, 2019[111]; Amir et al., 2023[112]; Ipek, 2011[125]; Lee & Xu, 2010[126]; Shen & Xu, 2016[127]; Syed et al., 2022[128]; Taheri-Ardali & Xu, 2015[129]; Wang, Qadir & Xu, 2013[130]）。由于针对这个假

① 最早使用焦点后压缩这个概念的是 Sugahara 的博士论文（2003）。Xu（1999）曾用焦点后下压（post-focus suppression）来描述北京话的焦点语调。Chen et al.（2009）开始用焦点后压缩（PFC）概念，以便涵盖美国英语疑问句里焦点后既压缩又抬高的现象（当时还不知道 Sugahara 的博士论文）。但焦点后压缩在调核分析理论里被归为焦点之外的低调尾，在 AM 理论里被描述为去音高重音（de-accenting），跟焦点处的调核重音分属不同的音系单元。

说的研究都是实证性的，所以一旦有例外可以很容易核实。这就使得今后很容易进行更多的研究。

3. 焦点和音高重音、声调、词重音的关系

从前面的回顾可以看到，焦点研究中最为纠结的一个问题就是焦点跟重音的关系。这里的重音指的是一种可以涵盖许多韵律现象的特征，该特征在语音学和心理语音学上常常表述为凸显(prominence)，而在语调音系学里则往往对应于音高重音。这种用某个单一特征涵盖许多韵律现象的思路碰到声调语言就产生了困难。例如，虽然普通话的一、二、四声在焦点下都会显现出明显的音高峰(Xu, 1999)[81]，但是，由于声调的功能显然不同于重音，把这些音高峰视为凸显或音高重音的直接声学表现就比较牵强了。更为合理的结论是，在焦点下，声调的调域被加宽了(Xu, 1999)[81]。Chen & Xu(2006)[131]又进一步显示，即使普通话里的焦点后压缩也跟声调的轻声一样分别有各自的声学表现，也不应该用低凸显度来同时涵盖。这种分离的原理一旦在声调语言里成立，自然也可以用于其他语言。譬如从功能上来看，英语的词重音其实跟汉语的声调相类似，因为它可以决定像 subject 这样的词是名词还是动词：前重是名词，后重是动词。早在 1958 年，Fry[35] 就通过感知实验发现，这类英语词的动、名之分在陈述句里只需两音节之间的音高差达到 5 Hz 就可以在感知上没有任何混淆了：前高的是名词，后高的是动词。并且，即使音高差增大到 90 Hz，分辨率也没有任何提高。由此可见，英语里用于分词的词重音只需高 5 Hz 就足够了，再多出来的，即使明明是发生在重读音节上，也是额外的，跟词重音本身无关。Xu & Xu(2005)[103]通过对美国英语的声学分析发现，无论是在焦点前或焦点后，重读音节的音高也只比相邻音节平均高出约 5 Hz。这么小的峰在大起大伏的语调曲线上虽显得微不足道，但也足以实现 Fry(1958)[35]发现的表达词重音所需要的音高差别。

焦点与凸显的本质差别还可以从它们的感知模式上看出来。Rump & Collier(1996)[132]发现，如果把荷兰语句子 Amanda is going to Malta 的句首词和句尾词的音高峰高度分别调节为连续统后，再让受试判断它是句首焦点、句尾焦点、双焦点，还是无焦点，得出的结果呈范畴式分布。Mixdorff(2004)[133]在芬兰语里也发现了类似的现象。相反，对于同样的连续统，如果受试的任务是判断凸显度，得到的结果则是非范畴的连续性感知(Gussenhoven et al., 1997)[134]。所以，Terken & Hermes(2000)[135]在总结了他们多年来所做的一系列凸显感知实验后坦承，弄清凸显的感知机制需要充分考虑它跟口语里的交际信息是如何对应的。时至今日，已有充分的证据表明，无论是声调或非声调语言，都没有必要再继续固守凸显这个笼统而模糊的概念，完全可以直接考察焦点和分词(用声调或词重音)各自的音高编码。

对于汉语普通话，词重音本身就是一个悬而未决的问题。尽管众说纷纭，看起来难以达成共识，但如果将之分为几个问题，就可以看到很多问题已经有答案了：(1)普通话有无类似于英语的区别性词重音？(2)普通话有没有非区别性词重音？(3)非区别性词重音的功能是什么？(4)重音的本质是什么？对于第一个问题，已有很多证据表明普通话里的轻声跟英语的弱读音节非常相似，见表 2：

表 2　英语弱读音节和普通话轻声共享的声学特征

	英语	汉语(普通话)
元音央化	Ladefoged(1982)[136]	Chao(1968)[45]
音高目标改变	Fry(1958)[35]	Chen & Xu(2006)[131]
时长缩短	Crystal & House(1988)[137], Nakatani et al.(1981)[139]	林焘(1985)[138]
发音弱化	Xu & Xu(2005)[103]	Chen & Xu(2006)[131]

另外的一系列研究发现，母语为法语、韩语等没有词重音语言的人对词重音不敏感，他们被称为重音失聪(stress deafness)；而母语为汉语者能够很好地感知英语重音(Lin et al., 2014[140]; Chrabaszcz et al., 2014[141])，说明汉语里的轻声使感知系统能够处理一个以轻重为标度的编码维度。在这个轻重维度里，关键的是轻而不是重。而轻，如表 2 所示，是一个多维度的特征。其中尤其值得一提的是发音弱化，因为它指的是目标趋近模型里的目标力度参数。声学分析和数字模拟的证据都表明，汉语轻声的力度远弱于非轻声的声调，这导致目标趋近严重不到位(Chen & Xu, 2006[131]; Xu & Prom-on, 2014[70])。相反，没有证据表明重读音节是靠增强发音力度来显示其重的，反而是音节越长，发音力度(stiffness)越弱(Ostry et al., 1983[142]; Xu &

Prom-on, 2019[84])。Xu & Prom-on(2019)[84] 在对英语的共振峰曲线做了仔细分析和模型推导后提出，正常话语中的发音力度已经达到饱和，没有进一步增强的余地。所以，表达重读所需要的目标充分趋近，靠的是增加时长，而不是增强力度。由此进一步提出，口语中最宝贵的资源是时间而不是能量。其原因是我们在单位时间内要传达尽可能多的信息，这就迫使目标趋近的力度常规地达到饱和状态。好在发音肌肉所需的能量是微乎其微的，所以聊几个小时的天都不会觉得累。这个发现对“省力说”(Zipf, 1949)[143] 这样的基本假说提出了挑战。至少在人类语言上，许多一般理解为省力的现象，其实很可能是省时现象，即尽可能在最短的时间内传达最多的信息(Xu & Prom-on, 2019)[84]。

有了以上的回顾，就可以进一步讨论非区别性词重音了。许多学者都提出，普通话在轻声之外还有轻重音节之分(Chao, 1968[45]；颜景助、林茂灿，1988[144]； Duanmu, 1993[145]；冯胜利、林晓燕，2023[146])。但是，正如端木三(2022)[147] 提出的长重短轻原则，这些轻重之分往往跟音节时长有关。例如 Chao(1968)[45] 认为，三音节词的轻重模式是 231(1 代表最重)，这正符合 Xu & Wang(2009)[148] 发现的三音节词时长模式，即末音节最长，中间音节最短。但是， Xu & Wang(2009)[148] 同时还提供了量化证据，表明这种时长之差并不反映发音力度，而是直接用来标注语流中音节组合的边界。Nakatani et al.(1981)[139] 发现，在英语这个典型的词重音语言里，词重音和边界对时长都有显著的影响，但两者的影响是相互平行的，极少交互作用。而边界标注的规律是，边界前音节会显著延长，并且边界强度越高，延长越多。所以，在普通话三字词里，最后一个音节最长是为了标注边界，而不是为了加重；而中间音节最短是由于对组内音节的压缩，而不是为了把它变成弱读。当然，短音节听起来轻、长音节听起来重也是有原因的。如表 2 所示，时长是区别性重音的特征之一。Fry(1958)[35] 也发现时长是比音强更重要的重音感知特征(不过远弱于音高)。但综合以上的回顾就可以明白，以往对普通话非区分性重音的判读，往往是把用于标界、分组的时长差别听成了轻重音之别。

对音节时长的研究还增进了我们对语言韵律的另一个方面的认识，也就是节奏。很长时间以来，已经被普遍接受的一种观念是，世界上的语言可以分为三种节奏类型：重音等时(stress-timed)、音节等时(syllable-timed)、莫拉等时(mora-timed) (Abercrombie, 1967[149]; Bloch, 1950[150]; Pike, 1945[151])。但是，从来就没有研究找到过这些单位在任何一个语言里实际时长相等的证据(Ramu et al., 1999)[152]。于是， Ramu et al.(1999)[152] 又提出一套仍然可以把语言按传统节奏类型分开的节奏指标(rhythm metrics)，并由此引领了新一轮的节奏研究热，研究者们纷纷提出多套节奏指标，为各种语言做分类。根据这些指标，汉语普通话总是被归类为音节等时的语言(Grabe & Low, 2002[153]; Lin & Wang, 2007[154]; Mok & Dellwo, 2008[155]; Nolan & Asu, 2009[156])。但问题是，这些节奏指标测量的并不是重音组、音节或莫拉的时长，而是元、辅音的时长分布和偏差，所以并未能证明传统的节奏类型假说，只是间接反映了一个语言里的音节复杂程度。为了给节奏类型假说找出哪怕一点蛛丝马迹， Wang et al.(2023)[157] 通过语料库对美国英语和汉语普通话作了细致的统计分析，结果发现英语没有任何等时的倾向，重音组的时长几乎完全等于组内音节时长的和；相反，普通话反而有短语等长的趋向，即短语总时长的增加慢于短语音节数量的增加。有意思的是，其主要原因之一正是前面提到的三音节词呈 231 的时长模式，而四字组的时长模式则为 2431；而英语里的三、四音节词呈现的是 221 和 2221 的模式(与 Nakatani et al.(1981)[139] 的发现相似)，即没有组内音节压缩。所以， Wang et al.(2023)[157] 的发现足以完全否定节奏类型假说，因为英语作为最典型的重音等时语言，却连一点等时的倾向都没有。

(二) 句调 / 句型 / 疑问 / 陈述

对疑问句语调的探讨很早就开始了，研究者们首先注意到的是疑问和陈述的语调差别，主要表现为句尾音高的升或降(Bolinger, 1978[158]; Greenberg, 1963[159]; Uldall, 1960[160])。后来进一步的研究发现，差别并不只是升或降那么简单。首先，一般理解的升降，是句末最后一个音节里有类似于声调一样的局部上升或下降。但是，很多语言的疑问句并没有句末音节的升调，例如那不勒斯意大利语的疑问和陈述句末的音节都是降调，只是疑问句句末音节的音高峰更偏后(D' Imperio, 2002)[161]。疑问句尾音节为降调还发生在西班牙语(Prieto, 2009)[162]、鲁尔蒙德荷兰语(Gussenhoven, 2000)[163]、希腊语(Grice, Ladd & Arvaniti, 2000)[164]、

匈牙利语(Gósy & Terken, 1994)[165]等语言里。其次,一般理解的疑问/陈述之差都只发生在句末音节上,以至于AM理论把句末音高模式称为边界调(boundary tone)。但有学者很早就提出,疑问句的音高上升可以发生在整句上。Thorsen(1978[166], 1980[52])发现在丹麦语里,陈述句、含疑问句法的以及不含疑问句法的疑问句可以由整句的音高倾斜度来区分。最后,句法上的疑问句并不一定表现为语调上的疑问句。Fries(1964)[167]在一个地图任务语料库里发现,在自然交流中的美国英语里,有高于60%的语法疑问句是用降调说的。Hirschberg(1995)[168]也报告说在一个自然语料库里有43.3%的是非问句呈降调。不过,Hedberg et al.(2017)[169]在标注了两个电话会话语料库后发现,绝大多数(90%)一般疑问句都是用升调的。所以,在这个问题上的早期发现有可能夸大了降调疑问句的比例。

即使整句音高上升或下降也不足以概括疑问/陈述语调之对立,因为这个功能跟分词用的音高模式以及焦点语调也有多重交互作用。Eady & Cooper(1986)[101]发现,在美国英语的疑问句里,音高峰的高度和形状都受制于句型和焦点位置。在无焦点和句末焦点句里,只有最后一个词的音高是疑问高于陈述;当焦点在句首时,焦点处的音高没有疑问/陈述之差,但焦点后所有词的音高都被抬高了。同时,无论焦点位置在哪里,焦点所在词内部的音高曲线都是疑问句里上升、陈述句里下降。Liu et al.(2013)[87]在对美国英语音节内音高作曲线分析和数字模拟后发现,陈述/疑问在音节内的升降之差不仅发生在焦点处,而且发生在所有重读音节的底层目标上。也就是说,美国英语里的疑问、陈述之差在第一个词重音上就体现出来了,而且这种差别使人在听到句子的第一个词时,就可以辨别出该句是疑问句还是陈述句(Saindon et al., 2017)[170],这跟其他一些印欧语疑问句给人的感知相似(Face, 2005[171]; Thorsen, 1980[52]; van Heuven & Haan, 2000[172])。有意思的是,Pierrehumbert在她的博士论文(1980)[7]里就已经注意到句尾前词内的音高曲线跟句末音节的音高曲线相似,但她将之称为回响音高重音(echo accent),意思是其曲线形状是对句末边界调的回响。更有意思的是,在随后被广为接受的AM理论里,包括ToBI语调标注系统(Silverman et al., 1992)[173],对此再也不提,形式为先的思路对发现功能性语调编码的阻碍由此可见一斑。

Liu et al.(2013)[87]还发现,普通话疑问语调的编码跟英语很不一样。首先,疑问与陈述之差是从句中第一个焦点处才开始显现的,而且完全没有像英语疑问句那样在焦点后大幅抬高音高,而是像陈述句一样在焦点后音高骤降,只是降的程度小于陈述句。其次,疑问语调即使在句末也不改变声调的基本形状,只是抬高声调的整体调域。在这一点上,多项研究的发现都很相似(劲松,1992[174];林茂灿,2004[175])。不过,有的汉语方言里的疑问句对声调影响很大,如粤语疑问句句末音节全都变为升调,造成了严重的声调混淆(Fok-Chan, 1974[176]; Ma, Ciocca & Whitehill, 2006[177])。这种语调与声调之间的关系因方言而异的现象是否有底层机制的因素还有待今后进一步探讨。还有一个更为奇特的跨语言差别是,在非洲一个地区的多种语言里,疑问语气的表达完全不用句尾升调,既没有像英语和粤语那样的音节内的上升,也没有像普通话那样的只是全句调域斜线上升,而是用句尾气声、元音加长或整句音高抬高等手段(Connell, 2017[178]; Rialland, 2009[179]; Salffner, 2010[180])。结合前面讨论过的焦点后压缩语调的历史来源来看,疑问语调的历史渊源可能更加久远(有可能追溯到智人走出非洲之前),非常值得进一步探讨。

(三)不主要靠语调编码的交际功能

本文回顾的研究主要是针对狭义上的语调,即只包括语句中音高变化模式的研究。但在语调研究史上,有一些交际功能长期以来一直被认为与音高模式有密切关系,但后来的研究却发现它们并不主要靠音高来编码,其中最主要的有两个,一个是分组/标界,一个是情感。

1. 分组/标界

分组和标界两个说法同时用是因为始终找不到一个最合适的术语。这个功能指的是用韵律手段把语流中的音节串分成或大或小的群组,这些群组可以是词,也可以是短语,还可以是其他单元。凭直觉,如何分组应该首先取决于语法结构,但很早就有人指出,韵律分组并不是完全由语法决定的(Selkirk, 1986[181]; Shih, 1986[182]),而是自成一个韵律层级体系,其组成单元从小到大为莫拉、韵律词、音步、短语、语调短语、音系短语、语句,等等。不过,这一类韵律层级体系里单元的定义严重依赖重音。正如前面已经指出的,脱离了具体

功能的重音把焦点、区别性词重音、非区别性词重音都混为一谈了，结果就造成了层级结构内部的定义多为循环性的(Wang et al., 2018)[107]。例如 AM 理论规定每个语调短语里只能有一个音高重音，所以，如果有两个音高重音就必须有两个语调短语。在这样的循环式定义下，对边界的语音标记的识别就无法保持客观了。前面已经提到，Nakatani et al.(1981)[139] 发现，英语的边界类型(词或短语) 对音节时长的影响独立于词重音对时长的影响。这就跟其他许多研究的发现一致，即标界的主要特征是时长，而不是音高或其他语音特征(English: Lehiste, Olive & Streeter, 1976[183]; Lehiste, 1996[184]; Korean: Jeon & Nolan, 2013[185]; Mandarin: Xu and Wang, 2009[148])。所以，对分组 / 标界功能的考察可以完全摆脱重音的束缚。

时长标注边界的最主要的特征是边界前的音节加长(Beckman & Edward, 1990[186]; Klatt, 1976[187]; Lehiste et al., 1976[183]; Nakatani et al., 1981[139]) 和边界处的无声停顿(Jeon & Nolan, 2013[185]; Yang & Wang, 2002[188])。除此之外，Xu & Wang(2009)[148] 还发现普通话里有组中缩短，即前面提到的 231 和 2431 时长模式；以及非单字组组末缩短，即非单音节组末音节短于单音节组的现象。主要是由于这两个外加的时长变化手段，普通话可以呈现英语里没有的短语等时的倾向(Wang et al., 2023)[157]。另外一个出人意料的发现是，普通话里的边界前延长到了类似于短语的单元之上就不再继续增加了，而是外加无声停顿；而英语里的边界前延长则跟随边界强度不断增加，只有到边界强度相当高时，才开始有无声停顿(Wang et al., 2023)[157]。有意思的是，这个发现跟说普通话的人的听觉感知是一样的(Kuang, Chan & Rhee, 2023)[189]。

对分组的语音标记的研究还会对一些广为接受的韵律组单元提出疑问。例如，一般都认为普通话里的二字词都自成一个音步(冯胜利，1998[190]; Duanmu, 1996[191], 2022[192])。但是，已有不止一项研究发现二字词并不是总有边界前延长(王晶，王理嘉，1993[193]; Wu, 2017[194]; Xu & Wang, 2009[148]; Zhang et al., 2026[195])，这就意味着二字词常常不被处理为带边界标记的韵律单元。如果非要强调音步的核心特征是轻重交错的模式，前面已经讲到，三音节、四音节字组里的所谓轻重交错实际上是对音长的感知效应，已有研究发现较短音节的发音力度反倒可能更大(Xu & Wang, 2009)[148]。王洪君(2004)[196] 曾提出，汉语的节奏类型应为松紧型，而不是像英语那样的轻重型。不过，正如本节内容所显示的，用于标界的时长变化在英语中同样很大，所以，按这个标准英语也是松紧型的。从这类争论中可以看到，不仅音步，其他被广泛接受的韵律单元都有功能和定义不清的问题；而且也不光是汉语，其他语言也没有真正解决这个问题。今后，从功能、实证、数字模型化着手的研究也许会有新的突破。

2. 情感

在谈到语调的功能时，人们常常会把语气、态度、情感都包括在内，由此引发了各种寻找与情感对应的语调模式的努力。但是，这一思路的研究收获甚少，因为很难找到与情感对应的特定语调模式(Bänziger & Scherer, 2005[197]; Bolinger, 1989[198]; Gussenhoven, 2002[199]; Halliday, 1985[200]; Mozziconacci, 2002[201]; Pierrehumbert & Hirschberg, 1990[21])。问题的关键在于这种研究思路没有考虑到人类情感的表达跟其他动物有很多共通之处，其来源很可能比语言本身还要古老。关于动物对情感的表达，Morton(1977)[202] 在考察了大量哺乳类和禽类动物的鸣叫后发现，与进攻性行为相伴的鸣叫声，音高偏低，音质偏粗糙；而与恐惧、顺从行为相伴的鸣叫声，音高偏高，音质偏柔润。他由此提出，前者的功能是使鸣叫者显得巨大，可以吓退听者；而后者是为了使鸣叫者显得矮小，可以吸引听者。Ohala(1984)[203] 又将该假设推广至人类语言，并进一步提出，说话时改变声道长度也可产生类似的效果，即拉长声道可以让自己显得块头大，缩短声道可以让自己显得块头小，而且微笑的表情也是为了缩短声道而把嘴角后拉。后来的一系列实验研究发现，这个可称为 Morton–Ohala 假说(Xu, 2023)[204] 的理论是有道理的(Chuenwattanapranithi et al., 2008[205]; Xu et al., 2013[206]; Xu, Kelly & Smillie, 2013[207])。最重要的是，话语中的情感表达并不需要特定的语调，而是可以通过改变任何语句的总体音高、共振峰离散度(formant dispersion) 和嗓音型来表达不同的情感；其中，嗓音型的作用尤其显著(Cai & Xu, 2023[208]; Xu et al., 2013[206])，如愤怒的语气用的是粗燥的嗓音、拉长的声道、较低的基频，而高兴的语气用的是气声的嗓音、缩短的声道、较高的音高。不过，在这些声学特征里，音高是最不可靠的，因为愤怒的语气往往会由于音量的提高而上升，以至和高兴不相上下(Cai & Xu, 2023)[208]。这就解

释了为什么早期的研究总是发现无法通过音高和音强来区别高兴和生气(Xu, 2023)[204]。

四、结语

本文通过对语调研究历史的回顾,着重追寻人们对语调认知的发展,而不只是简单历数谁做了什么。从这个回顾中,我们看到了两个大趋势,即从形式为先到功能为先的理论思路的转变,以及从抽象描述到参数生成的研究方法的发展。以这两个大趋势为主线,就比较容易从令人眼花缭乱的海量文献里理出头绪;还可以帮助今后的研究尽量避免重复性的工作,以便探索真正未知的领域。在第一个大趋势里,以发现交际功能的编码为目标的研究,已经揭示了诸如焦点、陈述 / 疑问、分组 / 标界、情感等功能的韵律编码,但其背后的交际功能的底层机理仍不甚清晰。比如对焦点语调,现有的研究只是确立了焦点的语音表现,但焦点会在何时何处出现还很不清楚。现有的种种理论明显偏向于预测过多的焦点,但目前也没有更好的替代方案。所以,今后一个可能的研究方向是用已知的焦点语调为检验标准,反过来弄清焦点功能的本质。这种用已知的编码方案为检验手段来探讨交际功能本身机理的研究,可能会让语调研究进入一个新的高度。另外,还有许多交际功能的语调编码也需要弄清更多的细节,其中一些是某语言特有的,如北京话的句末语气调(Chao, 1968[45]; Li et al., 2012[209], 2013[210]) 和英语的反驳语调(contradiction intonation) (Nolan, 2020)[211];还有一些是跨语言的,如话题(topic) 语调(Wang & Xu, 2011)[106]、插入句语调(parenthetical intonation) (Levis & Wichmann, 2015)[212],等等。

参考文献:

[1] Botinis A, Granström B, Möbius B. Developments and paradigms in intonation research[J]. Speech Communication, 2001(33): 263–296.

[2] Cruttenden A. Intonation[M]. Cambridge: Cambridge University Press, 1997.

[3] Cruttenden A. Mancunian Intonation and Intonational Representation[J]. Phonetica, 2001(58): 53–80.

[4] Barnes J, Shattuck–Hufnagel S. Prosodic Theory and Practice[M]. Cambridge: MIT Press, 2022.

[5] Xu Y. Speech Prosody: Theories, models and analysis[M]//Meireles A R. Courses on Speech Prosody. Cambridge: Cambridge Scholars Publishing, 2015: 146–177.

[6] Xu Y. Prosody, tone and intonation[M]//Katz W F, Assmann P F. The Routledge Handbook of Phonetics. Philadelphia: Routledge, 2019: 314–356.

[7] Pierrehumbert J. The Phonology and Phonetics of English Intonation[D]. Cambridge, USA: Massachusetts Institute of Technology, 1980.

[8] Pierrehumbert J. Tonal elements and their alignment[M]//Horne M, editor. Prosody: Theory and Experiment: Studies Presented to Gösta Bruce. London: Kluwer Academic Publishers, 2000: 11–36.

[9] Palmer H E. English Intonation, with Systematic Exercises[M]. Cambridge: Heffer, 1922.

[10] Kingdon R. The groundwork of English intonation[M]. London: Longman, 1958.

[11] O' Connor J D, Arnold G F. Intonation of Colloquial English[M]. London: Longmans, 1961.

[12] Halliday Mak. Intonation and Grammar in British English[M]. The Hague: Mouton, 1967.

[13] Crystal D. Prosodic Systems and Intonation in English[M]. London: Cambridge University Press, 1969.

[14] Brazil D M, Coulthard M, Johns C. Discourse Intonation and Language Teaching[M]. London: Longman, 1980.

[15] Wells J C. English intonation: an introduction[M]. Cambridge: Cambridge University Press, 2006.

[16] Nolan F. The rise and fall of the British School of intonation analysis[M]//Barnes J, Shattuck–Hufnagel S, editors. Prosodic theory and practice. Cambridge: The MIT Press, 2022: 319–349.

[17] Romero–Trillo J. Pragmatics and prosody in English language teaching[M]. New York: Springer Science & Business Media, 2012.

[18] Ladd D R. Intonational phonology[M]. Cambridge: Cambridge University Press, 1996.
[19] Pike K L. The Intonation of American English[M]. Ann Arbor: University of Michigan Press, 1945.
[20] Trager G L, SMITH H L. An outline of English structure[M]. Washington: Battenburg Press, 1951.
[21] Pierrehumbert J, Hirschberg J. The meaning of intonational contours in the interpretation of discourse[M]// Cohen P R, Morgan J, Pollack M E, editors. Intentions in Communication. Cambridge, Massachusetts: MIT Press, 1990: 271–311.
[22] Hirst D J. Form and function in the representation of speech prosody[J]. Speech Communication, 2005(46): 334–347.
[23] Hirst D J. A multi–level, multilingual approach to the annotation and representation of speech prosody[M]. Cambridge: MIT Press, 2022.
[24] TAYLOR P. Analysis and synthesis of intonation using the Tilt model[J]. Journal of the Acoustical Society of America, 2000(107): 1697–714.
[25] 't Hart J, Collier R, Cohen A. A perceptual Study of Intonation: An experimental–phonetic approach to speech melody[M]. Cambridge: Cambridge University Press, 1990.
[26] Gårding E. Speech act and tonal pattern in Standard Chinese[J]. Phonetica, 1987(44): 13–29.
[27] Hirst D J, Di Cristo A, Espesser R. Levels of representation and levels of analysis for intonation[M]//Horne M, editors. Prosody: Theory and Experiment. Dordrecht: Kluwer, 2000.
[28] 沈炯 . 汉语语调构造和语调类型 [J]. 方言 ,1994(3):221–228.
[29] Ladd D R. Intonational phonology[M]. Cambridge: Cambridge University Press, 2008.
[30] Xu Y. Separation of functional components of tone and intonation from observed F0 patterns[M]//Fant G, Fujisaki H, Cao J, Xu Y, editors. From Traditional Phonology to Modern Speech Processing: Festschrift for Professor Wu Zongji's 95th Birthday. Beijing: Foreign Language Teaching and Research Press, 2004: 483–505.
[31] Kohler K J. Prosody in speech synthesis: the interplay between basic research and TTS application[J]. Journal of Phonetics, 1991(19): 121–138.
[32] Kohler K. Timing and Communicative Functions of Pitch Contours[J]. Phonetica, 2005(62): 88–105.
[33] Fujisaki H. Dynamic characteristics of voice fundamental frequency in speech and singing[M]// Macneilage P F, editor. The Production of Speech. New York: Springer–Verlag, 1983: 39–55.
[34] Van Santen J, Kain A, Klabbers E, Mishra T. Synthesis of prosody using multi–level unit sequences[J]. Speech Communication, 2005(46): 365–375.
[35] Fry D B. Experiments in the perception of stress[J]. Language and Speech, 1958(1): 126–152.
[36] Kochanski G, Grabe E, Coleman J, Rosner B. Loudness predicts prominence: Fundamental frequency lends little[J]. Journal of the Acoustical Society of America, 2005(118): 1038–1054.
[37] Bailly G, Holm B. SfC: a trainable prosodic model[J]. Speech Communication, 2005(46): 348–364.
[38] Xu Y. The PENTA model of speech melody: Transmitting multiple communicative functions in parallel[C]//Proceedings of From Sound to Sense: 50+ years of discoveries in speech communication. Cambridge, MA, 2004: C–91–96.
[39] Xu Y. Transmitting Tone and Intonation Simultaneously: The Parallel Encoding and Target Approximation (PENTA) Model[C]//Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages. Beijing: International Symposium on Tonal Aspects of Languages, 2004: 215–220.
[40] Xu Y. Speech melody as articulatorily implemented communicative functions[J]. Speech Communication, 2005(46): 220–251.
[41] Pierrehumbert J B. Comparing PENTA to Autosegmental–Metrical Phonology[M]//Shattuck–Hufnagel S, Barnes J, editors. Prosodic Theory and Practice. Cambridge: The MIT Press, 2022: 408–424.
[42] Arvaniti A, Ladd D R. Greek wh–questions and the phonology of intonation[J]. Phonology, 2009(26): 43–74.
[43] Arvaniti A. The Autosegmental–Metrical model of intonational phonology[M]//Shattuck–Hufnagel S, Barnes J, editors. Prosodic Theory and Practice. Cambridge: The MIT Press, 2022: 25–63.

[44] Chang N T. Tones and intonation in the Chengtu dialect (Szechuan, China)[J]. Phonetica, 1958(2): 59–85.
[45] Chao Y R. A Grammar of Spoken Chinese[M]. Berkeley, CA: University of California Press, 1968.
[46] 沈炯 . 北京话声调的音域和语调 [M]// 林焘,王理嘉 . 北京语音实验录 . 北京:北京大学出版社,1985.
[47] 石锋 . 韵律格局:语音和语义、语法、语用的结合 [M]. 北京:商务印书馆,2021.
[48] Fujisaki H. Information, Prosody and Modeling with Emphasis on Tonal Features of Speech[C]//Proceedings of Speech Prosody 2004. Nara, Japan, 2004.
[49] Chao Y R. Tone and intonation in Chinese[J]. Bulletin of the Institute of History and Philology, 1933(4):121–134.
[50] 赵元任 . 国语语调 [M]// 吴宗济,赵辛那,编 . 赵元任语言学论文集 . 北京:商务印书馆,2002:406–425.
[51] Gårding E. On parameters and principles in intonation analysis[J]. Working papers/Lund University, Department of Linguistics and Phonetics, 1993(40): 25–47.
[52] Thorsen N G. A study of the perception of sentence intonation: Evidence from Danish[J]. Journal of the Acoustical Society of America, 1980(67): 1014–1030.
[53] 吴宗济 . 从声调与乐律的关系提出普通话语调处理的新方法 [G]// 中国语文编辑部,编 . 庆祝中国社会科学院语言研究所建所 45 周年学术论文集 . 北京:商务印书馆,1997:243–258.
[54] Boersma P. Praat, a system for doing phonetics by computer[J]. Glot International, 2001(5): 341–345.
[55] Andruski J, Costello J. Using polynomial equations to model pitch contour shape in lexical tones: An example from Green Mong[J]. Journal of the International Phonetic Association, 2004(34): 125–140.
[56] Gandour J, Tumtavitikul A, Satthamnuwong N. Effects of speaking rate on Thai tones[J]. Phonetica, 1999(56): 123–134.
[57] Grabe E, Kochanski G, Coleman J. Connecting intonation labels to mathematical descriptions of fundamental frequency[J]. Language and Speech, 2007(50): 281–310.
[58] Liu F, Surendran D, Xu Y. Classification of statement and question intonations in Mandarin[C]//Proceedings of Speech Prosody 2006. Dresden, Germany, 2006: PS5–25_0232.
[59] Pierrehumbert J. Synthesizing intonation[J]. Journal of the Acoustical Society of America, 1981(70): 985–995.
[60] Lee A, Xu Y, Prom–On S. Modeling Japanese F0 contours using the PENTAtrainers and AMtrainer[C]//The 4th International Symposium on Tonal Aspects of Languages, Nijmegen, Netherlands, 2014: 164–167.
[61] Lau E, Xu Y. Toward predictive modelling for AM theory of intonation[C]//Proceedings of The 19th International Congress of Phonetic Sciences. Melbourne, Australia, 2019: 2976–2980.
[62] Sun X. The determination, analysis, and synthesis of fundamental frequency[D]. Evanston, USA: Northwestern University, 2002.
[63] Black A, Hunt A. Generating F0 contours from ToBI labels using linear regression[C]//Proceedings of International Conference on Spoken Language Processing. Philadelphia, 1996: 1385–1388.
[64] Fujisaki H. A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour[M]//Fujimura O, editor. Vocal Physiology: Voice Production. New York: Raven Press, Ltd., 1988: 347–355.
[65] Ohala J J, Ewan W G. Speed of pitch change[J]. Journal of the Acoustical Society of America, 1973(53): 345–345.
[66] Sundberg J. Maximum speed of pitch changes in singers and untrained subjects[J]. Journal of Phonetics, 1979(7): 71–79.
[67] Xu Y, Sun X. Maximum speed of pitch change and how it may relate to speech[J]. Journal of the Acoustical Society of America, 2002(111): 1399–1413.
[68] Kochanski G, Shih C. Prosody modeling with soft templates[J]. Speech Communication, 2003(39): 311–352.
[69] Prom–On S, Xu Y, Thipakorn B. Modeling tone and intonation in Mandarin and English as a process of target approximation[J]. Journal of the Acoustical Society of America, 2009(125): 405–424.
[70] Xu Y, Prom–On S. Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model–based stochastic learning[J]. Speech Communication, 2014(57): 181–208.

[71] Bailly G. Integration of rhythmic and syntactic constraints in a model of generation of French prosody[J]. Speech Communication, 1989(8): 137–146.
[72] Taylor P. The rise/fall/connection model of intonation[J]. Speech Communication, 1994(15): 169–186.
[73] Mixdorff H. A Novel Approach to the Fully Automatic Extraction of Fujisaki Model Parameters[C]//Proceedings of ICASSP 2000. Istanbul, Turkey, 2000: 1281–1284.
[74] Raidt S, Bailly G, Holm B, Mixdorff H. Automatic generation of prosody: Comparing two superpositional systems[C]//Proceedings of Speech Prosody 2004. Nara, Japan, 2004: 417–420.
[75] Fujisaki H, Wang C, Ohno S, Gu W. Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command–response model[J]. Speech communication, 2005(47): 59–70.
[76] Xu Y. Consistency of tone–syllable alignment across different syllable structures and speaking rates[J]. Phonetica, 1998(55): 179–203.
[77] Xu Y. Fundamental frequency peak delay in Mandarin[J]. Phonetica, 2001(58): 26–52.
[78] Gay T J. Effect of speaking rate on diphthong formant movements[J]. Journal of the Acoustical Society of America, 1968(44): 1570–1573.
[79] Xu A, Van Niekerk DR, Gerazov B, Krug P K, Prom–On S, Birkholz P, Xu Y. Learnability of English diphthongs: One dynamic target vs. two static targets[J]. Speech Communication, 2025(170): 103225.
[80] Xu Y. Contextual tonal variations in Mandarin[J]. Journal of Phonetics, 1997(25): 61–83.
[81] Xu Y. Effects of tone and focus on the formation and alignment of F0 contours[J]. Journal of Phonetics, 1999(27): 55–105.
[82] Xu Y, Liu F. Tonal alignment, syllable structure and coarticulation: Toward an integrated model[J]. Italian Journal of Linguistics, 2006(18): 125–159.
[83] Xu Y. Syllable as a synchronization mechanism that makes human speech possible[J/OL]. Brain Sciences, 2025(15). https://doi.org/10.3390/brainsci15010033.
[84] Xu Y, Prom–On S. Economy of Effort or Maximum Rate of Information? Exploring Basic Principles of Articulatory Dynamics[J/OL]. Frontiers in Psychology, 2019(10). https://doi.org/10.3389/fpsyg.2019.02469.
[85] Xu Y. Consistency of tone–syllable alignment across different syllable structures and speaking rates[J]. Phonetica, 1998(55): 179–203.
[86] Kochanski G P, Shih C. Stem–ML: language–independent prosody description[C]//Proceedings of The 6th International Conference on Spoken Language Processing. Beijing, China, 2000: 239–42.
[87] Liu F, Xu Y, Prom–On S, Yu A C L. Morpheme–like prosodic functions: Evidence from acoustic analysis and computational modeling[J]. Journal of Speech Sciences, 2013(3): 85–140.
[88] Lee A, Xu Y. Modelling Japanese intonation using pentatrainer2[C]//Proceedings of The 18th International Congress of Phonetic Sciences. Glasgow, UK, 2015: 86–90.
[89] Taheri–Ardali M, Xu Y. An articulatory–functional approach to modeling Persian focus prosody[C]//Proceedings of The 18th International Congress of Phonetic Sciences. Glasgow, UK, 2015: 0799.
[90] Alzaidi M S A, Xu Y, Xu A, Szreder M. Analysis and computational modelling of Emirati Arabic intonation: A preliminary study[J]. Journal of Phonetics, 2023(98): 101236.
[91] Lee A, Simard C, Tamata A, Xu Y, Prom–On S, Sun J. Modelling Fijian focus prosody using PENTAtrainer: A pilot study[C]//Proceedings of The 2nd International Conference on Tone and Intonation (TAI 2023). Singapore, 2023: 9–10.
[92] Krifka M. Basic notions of information structure[J]. Acta Linguistica Hungarica, 2008(55): 243–276.
[93] Terken J M B. The distribution of pitch accents in instructions as a function of discourse structure[J]. Language and Speech, 1984(27): 53–73.
[94] Rooth M. A theory of focus interpretation[J]. Natural Language Semantics, 1992(1): 75–116.

[95] Chen L, Li X, Yang Y. Focus, newness and their combination: Processing of information structure in discourse[J]. PLoS One, 2012(7): e42533. DOI: 10.1371/journal.pone.0042533.
[96] Chen L, Wang L, Yang Y. Distinguish between focus and newness: An ERP study[J]. Journal of Neurolinguistics, 2014(31): 28–41.
[97] Kristensen L B, Wang L, Petersson K M, Hagoort P. The interface between language and attention: prosodic focus marking recruits a general attention network in spoken language comprehension[J]. Cerebral Cortex, 2013(23): 1836–1848.
[98] Halliday M A K. Intonation and Grammar in British English[M]. The Hague: Mouton, 1967.
[99] Chaffe W. Givenness, contrastiveness, definiteness, subjects, topics and point of view[M]//Li C, editor. Subject and Topic. New York: Academic Press, 1976: 25–55.
[100] Cooper W E, Eady S J, Mueller P R. Acoustical aspects of contrastive stress in question–answer contexts[J]. Journal of the Acoustical Society of America, 1985(77): 2142–2156.
[101] Eady S J, Cooper W E. Speech intonation and focus location in matched statements and questions[J]. Journal of the Acoustical Society of America, 1986(80): 402–416.
[102] Eady S J, Cooper W E, Klouda G V, Mueller P R, Lotts D W. Acoustic characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments[J]. Language and Speech, 1986(29): 233–251.
[103] Xu Y, Xu C X. Phonetic realization of focus in English declarative intonation[J]. Journal of Phonetics, 2005(33): 159–297.
[104] Chen S–W, Wang B, Xu Y. Closely related languages, different ways of realizing focus[C]//Proceedings of Interspeech 2009. Brighton, UK, 2009: 1007–1010.
[105] Chen Y, Xu Y, Guion–Anderson S. Prosodic realization of focus in bilingual production of Southern Min and Mandarin[J]. Phonetica, 2014(71): 249–270.
[106] Wang B, Xu Y. Differential prosodic encoding of topic and focus at sentence initial position in Mandarin Chinese[J]. Journal of Phonetics, 2011(39): 595–611.
[107] Wang B, Xu Y, Ding Q. Interactive prosodic marking of focus, boundary and newness in Mandarin[J]. Phonetica, 2018(75): 24–56.
[108] Xu Y. Prosody, tone and intonation[M]// Katz W F, Assmann P F, editors. The Routledge Handbook of Phonetics. Philadelphia: Routledge, 2019: 314–356.
[109] Wu W L, Xu Y. Prosodic Focus in Hong Kong Cantonese without Post–focus Compression[C]//Proceedings of Speech Prosody 2010. Chicago, 2010.
[110] Qin Z, Xu Y. Lack of prosodic focus in Chongqing dialect and possible historical sources[C]//Proceedings of Speech Prosody 2020. Tokuo, Japan, 2020: 630–634.
[111] Alzaidi M S, Xu Y, Xu A. Prosodic encoding of focus in Hijazi Arabic[J]. Speech Communication, 2019(106): 127–149.
[112] Amir N, Silber–Varod V, Anavi O, Deouell E, Mixdorff H. The effect of f0 and post–focal compression on the perception of narrow focus in Hebrew[C]//Proceedings of The 20th International Congress of Phonetic Sciences. 2023: 1391–1394.
[113] Zerbian S, Genzel S, Kügler F. Experimental work on prosodically–marked information structure in selected African languages (Afroasiatic and Niger–Congo)[C]//Proceedings of Speech Prosody 2010. Chicago, 2010: 100976:1–4.
[114] Bomhard A R. Reconstructing Proto–Nostratic: Comparative Phonology, Morphology, and Vocabulary[M]. Leiden: Brill, 2008.
[115] Diamond J, Bellwood P. Farmers and their languages: the first expansions[J]. Science, 2003(300): 597–603.
[116] Xu Y. Post–focus compression: Cross–linguistic distribution and historical origin[C]//Proceedings of The 17th International Congress of Phonetic Sciences. Hong Kong, 2011: 152–155.
[117] Xu Y, Chen S–W, Wang B. Prosodic focus with and without post–focus compression (PFC): A typological divide within the same language family?[J]. The Linguistic Review, 2012(29): 131–147.

[118] Chen Y. Post–focus compression in English by Mandarin learners[C]//The 18th International Congress of Phonetic Sciences, Glasgow, UK, 2015.
[119] Chen Y, Xu Y, Guion–Anderson S. Prosodic realization of focus in bilingual production of Southern Min and Mandarin[J]. Phonetica, 2014(71): 249–270.
[120] Wu W L, Chung L. Post–focus compression in English–Cantonese bilingual speakers[C]//Proceedings of The 17th International Congress of Phonetic Sciences. Hong Kong, 2011: 148–151.
[121] Yang Y. First language attrition and second language attainment of Mandarin–speaking immigrants in Hong Kong: evidence from prosodic focus[D]. Hong Kong: Hong Kong Polytechnic University, 2022.
[122] Renfrew C, Mcmahon A M, Trask L, Trask R L. Time depth in historical linguistics[M]. Mcmahon A M S, Renfrew C, editors. Cambridge, England: McDonald institute for archaeological research, 2000.
[123] 刘璐,王蓓,李雪巧 . 白语焦点和边界的韵律编码方式 [J]. 民族语文,2020(6):71–79.
[124] 王玲,王蓓,尹巧云,等 . 德昂语布雷方言中焦点的韵律编码方式 [J]. 中央民族大学学报(哲学社会科学版),2011(2):129–133.
[125] Ipek C. Phonetic realization of focus with no on–focus pitch range expansion in Turkish[C]//Proceedings of The 17th International Congress of Phonetic Sciences. Hong Kong, 2011: 140–143.
[126] Lee Y–C, Xu Y. Phonetic realization of contrastive focus in Korean[C]//Proceedings of Speech Prosody 2010. Chicago, USA, 2010.
[127] Shen C, Xu Y. Prosodic Focus with Post–focus Compression in Lan–Yin Mandarin[C]//Proceedings of Speech Prosody 2016. Boston, USA, 2016.
[128] Syed N A, Shah A W, Xu A, Xu Y. Post–focus compression in Brahvi and Balochi[J]. Phonetica, 2022(79): 189–218.
[129] Taheri–Ardali M, Xu Y. An articulatory–functional approach to modeling Persian focus prosody[C]//Proceedings of The 18th International Congress of Phonetic Sciences. Glasgow, UK, 2015: 0799.1–5.
[130] 王蓓,吐尔逊·卡得,许毅 . 维吾尔语焦点的韵律实现及感知 [J]. 声学学报,2013(1):92–98.
[131] Chen Y, Xu Y. Production of weak elements in speech: Evidence from f0 patterns of neutral tone in standard Chinese[J]. Phonetica, 2006(63): 47–75.
[132] Rump H H, Collier R. Focus conditions and the prominence of pitch–accented syllables[J]. Language and Speech, 1996(39): 1–17.
[133] Mixdorff H. Quantitative tone and intonation modeling across languages[C]//Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages. Beijing, 2004: 137–142.
[134] Gussenhoven C, Repp B H, Rietveld A, Rump H H, Terken J. The perceptual prominence of fundamental frequency peaks[J]. Journal of the Acoustical Society of America, 1997(102): 3009–3022.
[135] Terken J M B, Hermes D J. The perception of prosodic prominence[M]//Horne M, editor. Prosody: Theory and experiment: Studies presented to Gösta Bruce. Dordrecht: Kluwer, 2000: 89–127.
[136] Ladefoged P. A Course in Phonetics[M]. New York: Hartcourt Brace Jovanovich, 1982.
[137] Crystal T H, House A S. Segmental durations in connected–speech signals: Syllabic stress[J]. Journal of the Acoustical Society of America, 1988(83): 1574–1585.
[138] 林焘 . 探讨北京话轻音性质的初步实验 [M]// 林焘,王理嘉 . 北京语音实验录 . 北京:北京大学出版社,1985:1–26.
[139] Nakatani L H, O' Connor K D, Aston C H. Prosodic aspects of American English speech rhythm[J]. Phonetica, 1981(38): 84–106.
[140] Lin C Y, Wang M, Idsardi W J, Xu Y. Stress processing in Mandarin and Korean second language learners of English[M]. Bilingualism: Language and Cognition, 2014(17): 316–346.
[141] Chrabaszcz A, Winn M, Lin C Y, Idsardi W J. Acoustic cues to perception of word stress by English, Mandarin, and Russian

speakers[J]. Journal of speech, language, and hearing research, 2014(57): 1468–1479.
[142] Ostry D, Keller E, Parush A. Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movement in speech[J]. Journal of Experimental Psychology, 1983(9): 622–636.
[143] Zipf G K. Human Behavior and the Principle of Least Effort[M]. Cambridge, Massachusetts: Addison–Wesley Press, 1949.
[144] 颜景助,林茂灿 . 北京话三字组重音的声学表现 [J]. 方言,1988(3):227–237.
[145] Duanmu S. Rime length, stress, and association domains[J]. Journal of East Asian Linguistics, 1993(2): 1–44.
[146] 冯胜利,林晓燕 . 汉语的词重音:原理、结构与表征 [J]. 长江学术,2023(4):63–77.
[147] Duanmu S. Evidence for Stress and Metrical Structure in Chinese[M]//Huang C–R, Chen I H, Lin Y–H, Hsu Y–Y, editors. The Cambridge Handbook of Chinese Linguistics. Cambridge: Cambridge University Press, 2022: 361–382.
[148] Xu Y, Wang M. Organizing syllables into groups: Evidence from F0 and duration patterns in Mandarin[J]. Journal of Phonetics, 2009(37): 502–520.
[149] Abercrombie D. Elements of general phonetics[M]. Edinburgh: Edinburgh University Press, 1967.
[150] Bloch B. Studies in colloquial Japanese IV: phonemics[J]. Language, 1950(26): 86–125.
[151] Pike K L. The Intonation of American English[M]. Ann Arbor: University of Michigan Press, 1945.
[152] Ramus F, Nesporb M, Mehlera J. Correlates of linguistic rhythm in the speech signal[J]. Cognition, 1999(73): 265–292.
[153] Grabe E, Low E L. Durational Variability in Speech and the Rhythm Class Hypothesis[M]//Gussenhoven C, Warner N, editors. Papers in Laboratory Phonology: 7. The Hague: Mouton de Gruyter, 2002: 515–546.
[154] Lin H, Wang Q. Mandarin rhythm: an acoustic study[J]. Journal of Chinese Language and Computing, 2007(17): 127–140.
[155] Mok P P, Dellwo V. Comparing native and non–native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English[C]//Proceedings of Speech Prosody 2008. Campinas, Brazil, 2008: 423–426.
[156] Nolan F, Asu E L. The Pairwise Variability Index and Coexisting Rhythms in Language[J]. Phonetica, 2009(66): 64–77.
[157] Wang C, Xu Y, Zhang J. Functional timing or rhythmical timing, or both? A corpus study of English and Mandarin duration[J]. Frontiers in Psychology, 2023(13): 869049
[158] Bolinger D. Intonation across languages[M]//Greenberg J H, editor. Universals of human language. Stanford, California: Stanford University Press, 1978: 471–523.
[159] Greenberg J H. Some universals of grammar with particular reference to the order of meaningful elements[J]. Universals of language, 1963(2): 73–113.
[160] Uldall E. Attitudinal meanings conveyed by intonation contours[J]. Language and Speech, 1960(3): 223–234.
[161] D'Imperio M. Language–Specific and Universal Constraints on Tonal Alignment: The Nature of Targets and "Anchors"[C]// Proceedings of The 1st International Conference on Speech Prosody. Aix–en–Provence, France, 2002: 101–106.
[162] Prieto P. Tonal alignment patterns in Catalan nuclear falls[J]. Lingua, 2009(119): 865–880.
[163] Gussenhoven C. The boundary tones are coming: on the nonperipheral realization of boundary tones[M]//Broe M B, Pierrehumbert J B, editors. Papers in Laboratory Phonology V: Acquisition and the Lexicon. Cambridge: Cambridge University Press, 2000: 132–151.
[164] Grice M, Ladd D R, Arvaniti A. On the place of phrase accents in intonational phonology[J]. Phonology, 2000(17): 143–185.
[165] Gósy M, Terken J. Question marking in Hungarian: timing and height of pitch peaks[J]. Journal of Phonetics, 1994(22): 269–281.
[166] Thorsen N. An acoustical investigation of Danish intonation[J]. Journal of Phonetics, 1978(6): 151–175.
[167] Fries C C. On the intonation of "yes/no" questions in English[M]// Abercrombie D, Fry D B, Maccarthy P A D, Scott N C, Trim J L M, editors. In Honour of Daniel Jones. London: Longmans, 1964: 242–254.
[168] Hirschberg J. Prosodic and other acoustic cues to speaking style in spontaneous and read speech [C]//Proceedings of International Congress of Phonetic Sciences. Stockholm, Sweden, 1995: 36–43.

[169] Hedberg N, Sosa J M, Görgülü E. The meaning of intonation in yes–no questions in American English: A corpus study[J]. Corpus Linguistics and Linguistic Theory, 2017(13): 321–368.
[170] Saindon M R, Trehub S E, Schellenberg E G, Van Lieshout P H. When is a Question a Question for Children and Adults?[J]. Language, Learning and Development, 2017: 1–12.
[171] Face T L. F0 peak height and the perception of sentence type in Castilian Spanish[J]. Revista internacional de lingüística iberoamericana, 2005(3): 49–65.
[172] Van Heuven V J, Haan J. Phonetic correlates of statement versus question intonation in Dutch[M]//Antonis Botinis, editor. Intonation: Analysis, modelling and technology. Dordrecht, Netherlands: Springer, 2000: 119–143.
[173] Silverman K, Beckman M, Pitrelli J, Ostendorf M, Wightman C, Price P, Pierrehumbert J, Hirschberg J. ToBI: A standard for labeling English prosody[C]//Proceedings of The 1992 International Conference on Spoken Language Processing. Banff, 1992: 867–870.
[174] 劲松 . 北京话语气和语调 [J]. 中国语文,1992(2):113–123.
[175] 林茂灿 . 汉语语调与声调 [J]. 语言文字应用,2004(3):57–67.
[176] Fok–Chan Y Y. A perceptual study of tones in Cantonese[M]. Hong Kong: University of Hong Kong Press, 1974.
[177] Ma J K, Ciocca V, Whitehill T L. Effect of intonation on Cantonese lexical tones[J]. The Journal of the Acoustical Society of America, 2006(120): 3978–3987.
[178] Connell B. Tone and intonation in Mambila[J]. Intonation in African tone languages, 2017(24): 131.
[179] Rialland A. African "lax" question prosody: its realisations and its geographical distribution[J]. Lingua, 2009(119): 928–949.
[180] Salffner S. Tone in the phonology, lexicon and grammar of Ikaan[D]. London: SOAS, University of London, 2010.
[181] Selkirk E. On derived domains in sentence phonology[J]. Phonology Yearbook, 1986(3): 371–405.
[182] Shih C. The Prosodic Domain of Tone Sandhi in Chinese[D]. San Diego: University of California, San Diego, 1986.
[183] Lehiste I, Olive J P, Streeter L A. Role of duration in disambiguating syntactically ambiguous sentences[J]. Journal of the Acoustical Society of America, 1976(60): 1199–1202.
[184] Lehiste I. Suprasegmental features of speech[M]//Lass N J, editor. Principles of Experimental Phonetics. Boston: Mosby, 1996: 226–244.
[185] Jeon H–S, Nolan F. The role of pitch and timing cues in the perception of phrasal grouping in Seoul Korean[J]. The Journal of the Acoustical Society of America, 2013(133): 3039–3049.
[186] Beckman M E, Edwards J. Lengthenings and shortenings and the nature of prosodic constituency[M]//Kingston J, Beckman M E, editors. Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech. Cambridge: Cambridge University Press, 1990: 152–178.
[187] Klatt D H. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence[J]. Journal of the Acoustical Society of America, 1976(59): 1208–1221.
[188] Yang Y, Wang B. Acoustic correlates of hierarchical prosodic boundary in Mandarin[C]//Proceedings of Speech Prosody 2002. 2002: 707–710.
[189] Kuang J, Chan M P, Rhee N. Boundary perception as interplay by acoustic cues, syntactic constituency, and prominence[C]// Proceedings of The 20th International Congress of Phonetic Sciences. Prague, Czech, 2023.
[190] 冯胜利 . 论汉语的 "自然音步" [J]. 中国语文,1998(1):40–47.
[191] Duanmu S. Pre–juncture lengthening and foot binarity[J]. Studies in the Linguistic Sciences, 1996(26): 95–115.
[192] Duanmu S. Evidence for Stress and Metrical Structure in Chinese[M]//Huang C–R, Chen I H, Lin Y–H, Hsu Y–Y, editors. The Cambridge Handbook of Chinese Linguistics. Cambridge: Cambridge University Press, 2022: 361–382.
[193] 王晶,王理嘉 . 普通话多音节词音节时长分布模式 [J]. 中国语文,1993(2):112–116.
[194] Wu D. Cross–regional word duration patterns in Mandarin[D]. Urbana–Champaign: University of Illinois at Urbana–Cham-

paign, 2017.
[195] Zhang J, Feng Z, Xu Y. Dynamic Prosodic Grouping: Evidence from Recitation of Classical Chinese Poetry[C]//Proceedings of Speech Prosody 2026. Philadelphia, USA, 2026.
[196] 王洪君 . 试论汉语的节奏类型——松紧型 [J]. 语言科学,2004(10):21–28.
[197] Bänziger T, Scherer KR. The role of intonation in emotional expressions[J]. Speech Communication, 2005(46): 252–267.
[198] Bolinger D. Intonation and Its Uses: Melody in Grammar and Discourse[M]. Stanford, California: Stanford University Press, 1989.
[199] Gussenhoven C. Intonation and interpretation: Phonetics and Phonology[C]//Proceedings of The 1st International Conference on Speech Prosody. Aix–en–Provence, France, 2002: 47–57.
[200] Halliday M A K. Introduction to functional grammar[M]. London: Edward Arnold, 1985.
[201] Mozziconacci S. Prosody and Emotions[C]//Proceedings of The 1st International Conference on Speech Prosody. Aix–en–Provence, France, 2002: 1–9.
[202] Morton E W. On the occurrence and significance of motivation–structural rules in some bird and mammal sounds[J]. American Naturalist, 1977(111): 855–869.
[203] Ohala J J. An ethological perspective on common cross–language utilization of F0 of voice[J]. Phonetica, 1984(41): 1–16.
[204] Xu Y. Phonetics of emotion[K]// Mark Aronoff, editor. Oxford Research Encyclopedia of Linguistics. New York: Oxford University Press, 2023. DOI: 10.1093/acrefore/9780199384655.013.743.
[205] Chuenwattanapranithi S, Xu Y, Thipakorn B, Maneewongvatana S. Encoding emotions in speech with the size code: A perceptual investigation[J]. Phonetica, 2008(65): 210–230.
[206] Xu Y, Lee A, Wu W–L, Liu X, Birkholz P. Human vocal attractiveness as signaled by body size projection[J]. PLoS ONE, 2013(8): doi.org/10.1371/journal.pone.0062397.
[207] Xu Y, Kelly A, Smillie C. Emotional expressions as communicative signals[M]//Hancil S, Hirst D, editors. Prosody and Iconicity. Philadelphia: John Benjamins Publishing Co., 2013: 33–60.
[208] Cai Z, Xu Y. An acoustic analysis of Berlin database of emotional speech based on bio–informational dimensions[C]//Proceedings of The 20th International Congress of Phonetic Sciences. Prague, Czech, 2023: 4091–4095.
[209] Li A, Fang Q, Jia Y, Dang J. More targets? Simulating emotional intonation of mandarin with PENTA[C]//Proceedings of Chinese Spoken Language Processing (ISCSLP), 2012 8th International Symposium on. Hong Kong, 2012. DOI: 10.1109/ISCSLP.2012.6423546.
[210] Li A–J, Jia Y, Fang Q, Dang J–W. Emotional intonation modeling: A cross–language study on Chinese and Japanese[C]// Proceedings of Asia–Pacific Signal and Information Processing Association Annual Summit and Conference. Shangri–La Singapore, 2013: 1–6.
[211] Nolan F. Intonation[M]//The handbook of English linguistics. Chichester, West Sussex, UK: John Wiley & Sons, 2020: 385–405.
[212] Levis J M, Wichmann A. English intonation – Form and meaning[M]// Reed M, Levis J M, editors. The handbook of English pronunciation. Malden, Massachusetts, USA: Blackwell, 2015: 139–155.
[213] Prom–On S, Xu Y, Gu W, Arvaniti A, Nam H, Whalen D H. The Common Prosody Platform (CPP): where Theories of Prosody can be Directly Compared[C]//Proceedings of Speech Prosody 2016. Boston, USA, 2016.

（下转第 109 页）

［32］李友梅 . 中国社会治理转型［M］. 北京：社会科学文献出版社，2018：172.

Comparative Study on the Realization Path of Common Prosperity in Jiangsu, Zhejiang and Shanghai: From the Perspective of Second-best Choice

MA Yiqun, FENG Yinqin

(National Audit Institute, Nanjing Audit University, Nanjing 211815, China)

Abstract: Common prosperity constitutes an essential requirement of socialism with Chinese characteristics. This paper systematically compares the disparities in common prosperity development across Jiangsu, Zhejiang and Shanghai under diverse development models, and explores the path selection for advancing common prosperity in the three regions from the perspective of the second-best choice. The research concludes that the level of common prosperity varies significantly among Jiangsu, Zhejiang and Shanghai, and different development models exert differentiated influences on the realization of common prosperity. The Southern Jiangsu Model, rooted in township enterprises, achieves outstanding performance in resident affluence, ecological livability and public service provision. The Wenzhou Model, dominated by the private economy, presents prominent advantages in residents' spiritual fulfillment and social harmony. The Pudong Model, relying on state-owned enterprises, shows distinct strengths in improving people' s living standards. Furthermore, this paper proposes practical implications for promoting common prosperity by adopting localized development strategies tailored to the characteristics of different regional development models.

Key words: common prosperity; development model; second-best choice

（上接第 60 页）

Review and Synthesis of the History of Intonation Research

XU Yi

(Department of Speech, Hearing and Phonetic Sciences, University College London, London, WC1N 1PF)

Abstract: This paper traces the development of our understanding of intonation through a review of the history of intonation research. From this review, two major trends emerge: a shift in theoretical orientation from form-first to function-first approaches, and an evolution of research methodology from abstract description to parametric generation. Using these two trends as guiding threads allows us to bring order to a vast body of literature, enabling future research to focus on genuinely unexplored areas while avoiding redundant efforts. Within the first trend, researches aimed at identifying the encoding of communicative functions have revealed prosodic patterns associated with functions such as focus, statement vs. question, grouping/boundary marking, and emotion. In contrast, the trend from abstract description to parametric generation of intonation is still at a very early stage. Investment in the study of intonation through computational modeling remains limited, partly due to its perceived difficulty, but more importantly because it is often not regarded as part of mainstream foundational research. Because parametric models inevitably aim to simulate surface pitch contours, there are concerns that they may sacrifice the level of abstraction required for cognitive understanding. However, a review of the development of PENTAtrainer shows that computational modeling is not only compatible with theoretical modeling, but can actually provide an effective means of testing intonation theories. A truly valid theoretical abstraction must be one that is capable of reproducing surface details.

Key words: intonation; tone; lexical tone; focus; function-first; form-first; grouping; boundary marking