

平行编码及目标趋近模型：原理、运用和意义 (I) *

许毅 Santitham Prom-on 刘芳 (著)
(伦敦大学学院;雷丁大学)

郭嘉 昂秋香 (译)
(南开大学外国语学院)

一、引言

语言是一个把信息从一个人传给另一个人的交际系统。这些信息丰富而且多层次，但经过发音系统的编码后，听者能够轻松的解码。这些看起来似乎显而易见，甚至不值一提的事实，却正是言语发音功能观的前提，而言语发音功能观又是平行编码及目标趋近 (PENTA) 模型 (Xu, 2005) 的基础。因此，PENTA 是一个研究韵律是如何用一个由神经控制的生物力学系统有效地传递多层次信息的理论。换言之，PENTA 是关于韵律如何作为交际系统来操作，如何可以习得，并如何会随着时间的推移而演变的理论模型。因此，PENTA 的使命不同于那些致力于直接诠释韵律形式的理论。它首先关注的是运作过程本身，而把韵律形式当作操作过程的副产品，而非最终目标。

二、概念框架

PENTA 模型远远不止关注上述的基本事实，还提出了若干假设。第一个假设是，音节同步、循序目标趋近 (PENTA 的 TA 部分) 是言语韵律的基本机制，通过该机制，信息编码

才能得以实施。第二个假设是，通过相互平行的编码方案 (encoding scheme)，而不是通过一个等级结构，韵律能够同时传递多层信息 (PENTA 的 PE 部分)。第三个假设是，编码方案的语音部分是由参数控制的，而不是由抽象符号来规定的。鉴于目前这些都是假说，每一个都需要独立论证。我们将在下面对 PENTA 模型作简要介绍后一一展开论述。

图 1 是 PENTA 模型的通用图式。该模型不仅适用于韵律，也适用于言语的其它方面 (Xu & Liu, 2012)。左起的第一个模块代表由言语传递的交际功能。图中这些交际功能以非等级性排列，由此表明它们之间的关系是平行的。第二个模块代表与交际功能对应的编码方案。该图式清晰地表现出，交际功能并不直接控制表层的声学特征，而是要借助于一系列特定的编码方案来实现。这些编码方案可以是高度程式化并随语言而异的，也可以是逐步变化和普遍通用的。第三个模块代表了由编码方案所决定的发音参数，这些参数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。该过程最终生成表层声学参数。

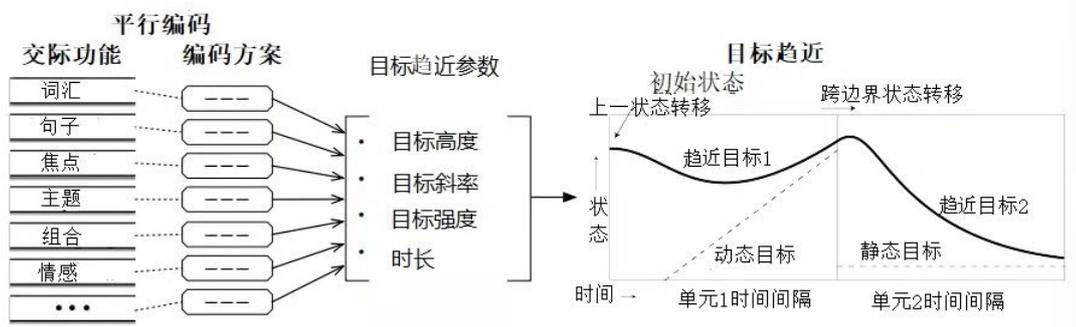


图 1. PENTA 模型图式(Xu, 2005; Xu & Liu, 2012; Xu & Wang, 2001)

图 1 最右边模块所描述的 TA 模型，假定

每个音节都被赋予有音高和斜率值的底层音

* 本译文是 “The PENTA model: Concepts, use and implications” 的第 1 和第 2 部分。原文见 Xu, Y., Prom-on, S. and Liu, F. (2022). The PENTA model: Concepts, use and implications. In *Prosodic Theory and Practice*. S. Shattuck-Hufnagel and J. Barnes. Cambridge: The MIT Press pp. 377-407. 译文经过原作者许毅老师审阅校改，在此深表感谢！
本文内容与 Xu, et al. (2015) 有重合之处。

高目标。因此，表层曲线体现了循序趋近一系列目标的结果，每个目标与一个音节同步。在相邻音节的边界点，前一音节的最终发音状态会传递到下一音节。这种传递常常会导致某一转折点处表层对齐的延迟。因此，此模型并不对表层转折点的时间对齐作明确规定。

2.1 发音机制

和整个理论一样，PENTA 的发声部分也是基于一些显而易见的事实。最基本的一个事实是，人们需要用不同的发音状态来代表不同的韵律成分（因为信息必须用不同的表征来编码），这样必然会导致不同发音状态之间需要有过渡动作，而每一个过渡动作都需要时间。因此，针对这些过渡动作，需要回答两个实证性的问题：A) 每一个过渡动作需要多长时间？若所需时间短到可以忽略不计，从理论上讲，就没必要太在意。B) 过渡动作的方式是什么样的？对这些问题了解有助于解释很多表层声学特征的细节。第二个基本事实是，生成 F_0 曲线的喉部运动必须和生成音段的喉部以上发音器官的动作同时发生，并且彼此之间必须及时协调。基于此，另外两个需要回答的实证性问题是：C) 时间协调的基本机制是什么？D) 在协调过程中，说话者有多少自由度？

关于问题 A，即 F_0 运动需要花多长时间，实验结果（Sundberg 1979, Xu & Sun 2002）表明，由于喉部状态改变而导致 F_0 运动所需的时间不容忽略。实验发现 F_0 运动的范围和完成该

运动所需要的平均最短时间有以下类线性关系：

$$\text{上升: } t_r = 89.6 + 8.7 d \quad (1)$$

$$\text{下降: } t_f = 100.04 + 5.8 d \quad (2)$$

其中， d 是 F_0 （半音）的变化幅度， t 是时长，以毫秒（ms）为单位。这两个公式表明，无论升调还是降调，即使最微小的音高运动也需要大概 100 ms。假定平均语速为每秒 5-7 个音节，那就意味着每个音节时长大概为 143-200 ms，所以每个音节时长的一大部分都必须用于音高过渡。

至于过渡动作的方式（问题 B），对汉语连续语流中字调的系统研究为其提供了关键的线索。如图 2 所示，当一个声调前面有不同声调时，其 F_0 曲线会呈现逐渐趋近一个符合其底层特性的线性曲线：阴平（H）为高平，阳平（R）为上升，去声（F）为下降。因此，声调之间的语音过渡动作可以描述为向底层目标的趋近。（Xu & Wang, 2001）

关于喉部和喉部以上发音器官动作时间协调的基本机制（问题 C），图 2 提供了相关线索。我们可以看到同一声调在不同声调后的差异在音节首最大，音节末最小，表明朝向本声调目标的动作起始于音节首，结束于音节末。因此，声调的目标趋近动作可视为与音节同步。换句话说，每一个目标趋近动作都发生于携带该调的音节。这些观察形成了 PENTA 模型关于生成 F_0 的基本发声机制的核心假设，即音节同步、循序目标趋近。（Xu, 2005）

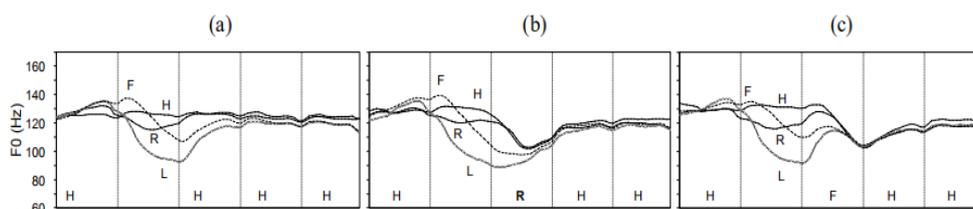


图 2: 不同声调环境下的汉语声调。基于平均时长归一化的 F_0 曲线。句子为“猫咪/迷/米/蜜 摸 / 拿 / 卖猫咪”。

每一个曲线都是四位北京普通话男性发音人所读的共 40 次发音的 F_0 均值（每一个重复读 5 次）。图中竖线代表音节边界。改编自 Xu (1999)。调名 H, R, L, F 分别代表高调、升调、低调（上声）和降调（去声）

该基本发音机制事实上可以分为两个部分：1. 每个音节都有音高目标，2. 音节同步、循序目标趋近。每个音节都有音高目标意味着，即使看起来好像有一条能够涵盖不只一个音节的整体曲线，每一个音节还是需要被赋予一个音高目标。这一论点是基于发声学的考量。如图 3 所示，首先，先生成一个没有音高的音节（左下图）和一个没有载体的 F_0 曲线（左上图），然后把两者结合在一起形成一个完整声学表层波形（右图），这在物理上是无法实现的。

虽然也可以想像左上图的 F_0 曲线和左下图的共振峰轨迹先在大脑中形成，然后在发音过程中如实地实现，但是如上所述，在音高变化最大速度的限制下，理想音高目标之间的 F_0 过渡速度相对于音节较为缓慢。共振峰的变化也呈现出同样的趋势（Cheng & Xu, 2013）。由此，图 3 中 F_0 和共振峰的曲线很大程度上是物理惯性的产物。然而，假如这种语音过渡是大脑发出的生成 F_0 和共振峰的指令，并传递给发声肌肉，那么惯性效应就会发生两次！

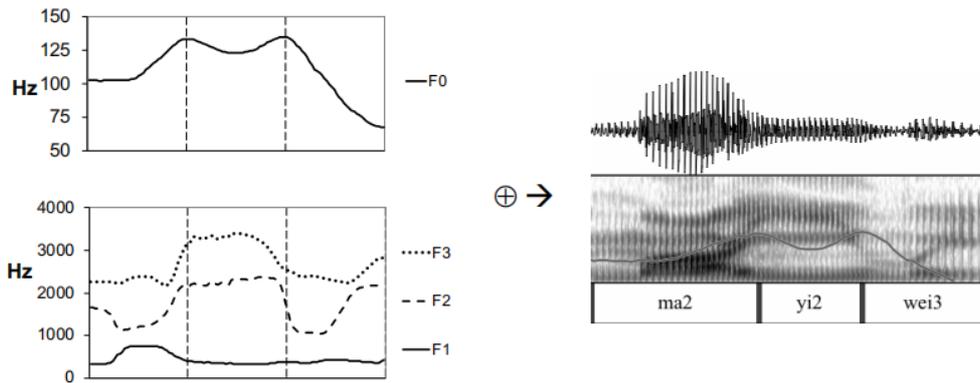


图 3. 左：汉语句子“(bi3) ma2 yi2 wei3 (shan4)” [More hypocritical than Aunt Ma] 的连续 F₀ 曲线(上)和共振峰(下)轨迹. 右：同一句子的波形、语谱和 F₀ 轨迹. 原始数据来自 Xu (2007)

因此，语音表层连续 F₀ 曲线的生成不太可能独立于音段而只是在发音过程时加入到音段丛中。相反，更有可能的是，在操控层面，每个音节都已被赋予所有相关的底层发音目标，包括音段目标、音高目标、甚至发声型目标（即嗓音型）。图 4（左）展示了音高和共振峰模式。这里的共振峰模式代表底层目标的声道形状。然后，发音过程同时通过目标趋近（右

上）循序产出所有这些目标。最后，通过生物力学过程生成连续的表层 F₀ 和共振峰轨迹（右下），这两者构成了绝大多数朝各自对应目标的过渡。因此，每一个音节在发音之前，都必须被赋予作为发音系统的控制信号的音段和超音段目标。重要的是，惯性作用只会发生一次，即在言语产生链的最后一个阶段——发音实施阶段。

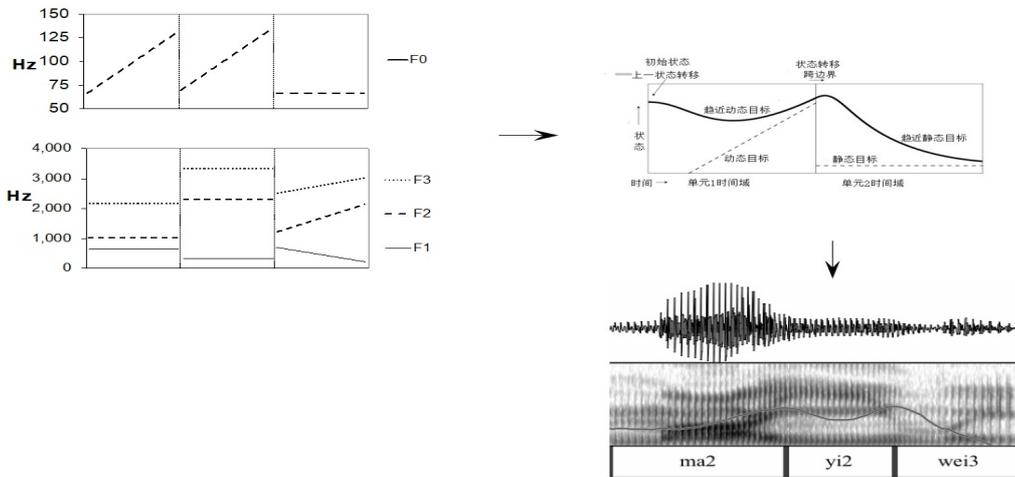


图 4. 左：图 3 汉语句子的假定底层音高(上)和共振峰(下)目标. 右（上）：目标趋近(TA)模型 (Xu & Wang, 2001). 右（下）：通过与 TA 模式相似的发音过程，利用底层目标产生的同一句子的波形、语谱和 F₀ 曲线

对每个音节都有音高目标的论点还有进一步的支持，即英语中不仅仅是重读音节，还有非重读音节，以及汉语普通话中的轻声都具有功能对立的音高目标。例如，在英语中，当一个单词成为句子焦点时，词内重读音节后的非重读音节主动降低音高，让它们属于焦点后的时域。但是，音高降低的速度没有在焦点后重读音节中那么快 (Xu & Xu, 2005)。这表明非重读音节被赋予了焦点后目标，但该目标的趋近力度较弱。这与声学分析 (Chen & Xu, 2006; Xu & Xu, 2005) 和计算机建模 (Liu et al,

2013; Xu & Prom-On, 2014) 中的发现一致。这种力度较弱的目标趋近，可以解释英语中非重读音节、以及汉语普通话轻声中音高的明显多变性（以致看起来好像没有目标）。

音高产生机制的另一个假设是目标趋近动作的音节同步。也就是说，喉部和喉部以上发音器官运动在发出音节时是同步的 (Xu & Liu, 2006, 2012)。该假设不仅受到如图 2 所观察到的结果的启发，同时还受到肌动系统只有通过同步功能才能快速协调多个动作这一发现的影响 (Kelso, 1984; Kelso, Southard &

Goodman, 1979; Mechsner et al., 2001)。这种同步制约可进一步归因于肌动控制的一个基本难题,即如何应对肌动系统的多维性(Bernstein, 1967; Latash, 2012)。Bernstein (1967) 提出,这个难题可以通过在动作习得的过程中,对自由度(DOF)进行功能性限制来缓解。对自由度(DOF)的限制类似于在一个方向盘的控制下,所有的车轮都围绕特定轴线旋转,使得车轮的运动完全同步,自由度融为一体。就语音而言,音节实际上是一种生物进化形成的肌动同步机制,用于实现多个动作同时编码。根据音节的时间结构模型(Xu & Liu, 2006, 2012),所有发音目标的趋近动作在音节首同步起始,只不过首辅音的动作结束于元音和喉音动作完成之前。

当前版本的 TA 模型仍存在一些问题,特别是通过 qTA (Prom-On et al., 2009) 进行的量化建模有一些问题至今尚未完全解决。第一个问题是目标趋近在动态目标上的延迟现象。研究发现,在汉语普通话中,如音节时长过长,阳平和去声的最终音高斜率仍会保持不变(Xu, 1998, 2001)。英语双元音中的共振峰轨迹的最后斜率也有类似的现象(Gay, 1968)。一种可能的发音机制是,发音人有能力随着时间的推移调整他们的发音力度以保证在音节结束时达到一致的运动速率。但这一机制还没有在当前版本的量化目标趋近模型(qTA)中得到运用。

第二个问题是同一音节内有可能有两个连续音高目标。例如,普通话中的上声表现为一开始是低降,随后上升。虽然整个曲线看起来像阳平(也是先有小幅下降然后上升),但是当上声后有其它声调时,其上升的调尾会完全消失。因此,上声的调尾是可有可无的,其底层目标不太可能与阳平相似。相反,单念的上声可能是在低目标后跟着一个中高目标或上升目标。一个音节中这样的连续目标仍将遵守同步制约,因为第一个目标的起始音和第二个目标的结束音分别与音节的起点和终点保持一致。与延迟的目标趋近一样,连续目标也尚未在 qTA 中实施。

第三个问题是低调后反弹现象。也就是说,在一个音高非常低的声调后面的(若干)音节,其 F_0 有时会先“弹起”然后再回归正常。如果低调后的声调调值较弱,比如轻声(Chen & Xu, 2006),这种反弹现象会很容易观察到。如果随后的声调不是弱值,这种反弹效应只有在低调位于焦点时才能观察到。我们推测这种反弹是由于喉外肌(包括主要在降低音高时起作用的胸舌肌、胸甲肌、甲舌肌)突然停止收缩时,

环甲肌(唯一能拉长声带的肌肉)(Zemlin, 1988)仍在收缩,导致声带紧张度突然增加,从而使肌肉对抗暂时失去平衡。通过增加低调向后面声调的状态转移值里的 F_0 二阶导数,我们已经能够模拟低调后的反弹。这一模仿机制已能良好运行(Prom-on et al., 2012),但还尚未被纳入 qTAtrainer 或 PENTAtainer 中。

与低调后反弹密切相关的是第四个问题,低调前抬高(pre-low raising, 或 anticipatory raising、anticipatory dissimilation、H raising),即带有低音成分的声调或音高重音导致前面音节的音高提升(Connel & Ladd, 1990; Hyman, 1993; Gandour, Potisuk & Dechongkit, 1994; Laniran, 1992; Laniran & Gerfen, 1997; Lee, Xu & Prom-on, 2017; Xu 1997)。鉴于在许多彼此毫无关联的语言中都发现了这种现象,这可能是一个普遍的发音机制,尽管其确切性质尚不明晰。我们目前认为,低调前抬高是在为 F_0 的急剧下降做准备,类似于人们在远距离投掷时,要先将手臂向后拉。低调前抬高也没有直接在 qTAtrainer 模型或 PENTAtainer 中模拟。不过可以通过让声调的音高目标以下一个声调为条件来部分模拟该效果。

最后,众所周知,在阻塞辅音后, F_0 在嗓音起始时会有暂时性的波动(大部分是向上波动)(Hombert, 1977; Silverman, 1986)。波动包括非常短暂(持续约 30 毫秒)的空气动力效应(Löfqvist & McGowan, 1992; Xu & Xu, 2003)和较长时间的声带紧张度效应(Stevens, 1998; Xu & Wallace, 2004)。同样,这些机制尚未纳入我们的计算机模型。然而, qTAtrainer 和 PENTAtainer 都允许(实际上是建议)用户将整个音节,包括所有的清辅音,设置为目标趋近域。与把目标趋近域局限于韵母相比,运用此策略生成的 F_0 曲线更接近原始曲线(Xu & Prom-on, 2014)。

2.2 交际功能

同样,遵循从基本事实出发的原则,我们首先认识到语言中的许多意义是通过韵律表达出来的。但这些意义到底是什么,以及它们是如何通过韵律编码的,这些问题还悬而未决。其中一个关键的问题是,意义是直接编码的,还是通过韵律音系这个中介编码的。对此我们有必要追溯到索绪尔关于语言符号的观点(de Saussure, 1916)。索绪尔的观点强调了需要把意义和意义的载体(这些载体本身没有意义),即所指和能指区分开来。因此,言语中的语言单位是能指与所指或功能与形式的结合。尽管

这种观点已经深植于现代语言学理论的很多方面，但韵律研究碰到的困难表明，理论上还需要进一步清晰化。首先，在包含所指—能指这一硬币的两面中，所指—交际功能，或能指—语音形式，谁该为韵律范畴提供基本定义？对于词的定义，这并不是难题，因为分别词项比较容易，其功能与形式的结合一目了然。然而，在韵律中，功能与形式的结合难以清晰界定。在这种情况下，在定义韵律单位时，是应该首要考虑功能还是形式？PENTA 和其他一些以功能为导向的模型（如 Bailly & Holm, 2005）一样，都认为当拿不定时，应该以功能而不是以形式为界定韵律单位的最后准则。

韵律研究的另一个复杂性在于它所传达的意义的多样性，这也是经典的索绪尔观点无法轻易解释清楚的。例如，在类似于汉语普通话或瑞典语等声调语言中，或像日语那样的音高重音语言中，音高用于形成声调或音高重音，从而区分音段相同的单词。但是，音高同时也用于编码焦点，强调话语中的特定成分（Bolinger, 1972），以及编码句子类型，表达陈述与疑问之别（Eady & Cooper, 1986）。因此，很难以形式优先的方法通过直接观察基频来确定哪些部分各自对应这三个意义的能指。

基于我们对这些固有困难的认知，PENTA 模型的研发采取了双管齐下的策略。首先，我们一直致力于一次只确立一个功能—形式的结合：从合理清晰地定义每个功能开始，然后通过实验发现（或排除）语言编码每个功能的韵律手段。正如我们接下来将看到的，与词项的语音编码不同，韵律手段并不一定是各自占据一个独立的时段，而常常是修改某些已被其它功能规定了形式的单元。例如，声调已经规定了音节音高曲线的基本形状。因此，焦点和句子类型只能通过用各种方式修改这些基本形状来编码。

其次，由于韵律研究中功能与形式关系的模糊性，我们一直致力于通过实验来发现每一个交际功能在韵律中编码的确切方式，而不是通过纯理论的推论或概念假设。对于每一个功能，都首先确立可靠的方法来诱导功能的发生

并控制其范畴化的对立。并在做好必要的准备工作后（包括最小对比对，以及重视细节等），开展实验，进行录音，并分析功能对立的语音实现（Xu, 2011a）。

我们尤其是采用了一种能够详细且系统地比较连续的 F_0 曲线的方法（Xu 2011a, 2013）。该方法的关键在于提取时长归一化以后的连续 F_0 曲线。时长归一化是在每个单元内等时取数量相等的 F_0 值。例如，在提取音节的 F_0 时，对同一音节的多次发音取时长归一化的 F_0 ，这样甚至可以将不同发音人的时长归一化 F_0 进行平均，以消除可能掩盖共同特征的个体差异。例如，图 2 中，因为把相同声调在不同语境中产生的变体的完整轨迹放在一起直接比较，图里的 F_0 均值曲线可以清楚地呈现出声调实现的目标趋近特征。^①

这些策略还有助于观察 F_0 如何能够同时编码多种交际功能。如图 5 和图 6 所示，在汉语普通话和英语中，焦点和句型可以通过修改局部 F_0 曲线的高度和形状在声调和词重音的辨词功能之上实现。这些句子，无论是陈述句还是疑问句，都以焦点位于句首（只在图 5），句中和句末的形式说出。如图所示，焦点和句型的 F_0 模式体现在他们彼此之间以及他们和声调或词重音之间的相互作用之中：

1. 焦点的一大特性就是有明显的焦点后音高位移，位移的方向与句型和语言均相关。在汉语和英语的陈述句中，焦点后调域都是下压。在疑问句中，英语上升，汉语下压，但汉语疑问句中的下压趋势比陈述句稍缓，因此，我们依然能根据焦点的声学表现区分出汉语的陈述和疑问句。

2. 在英语中，焦点、句型与词重音相互作用而决定音节音高目标的微观特性。词末焦点重读音节的目标斜率在陈述句中为降，疑问句中为升（job in b, d）；非词末焦点重读音节的目标斜率在陈述句和疑问句中均为升。

3. 在汉语中，句型的作用从焦点处开始显现。但在句末，即使是无焦点句也能观察到句型的差异。

^① 图 2 是利用我们开发的软件工具 ProsodyPro 绘制而成。ProsodyPro 是一个基于 Praat 的脚本，可以在 www.homepages.ucl.ac.uk/~uclyyix/ProsodyPro/ (Xu 2013) 网站获取。同样，FormantPro 也是一个小工具，可以用于音段分析 (www.homepages.ucl.ac.uk/~uclyyix/FormantPro/)，有助于进行连续共振峰轨迹的系统比较，这种比较在音段分析中还比较少见。

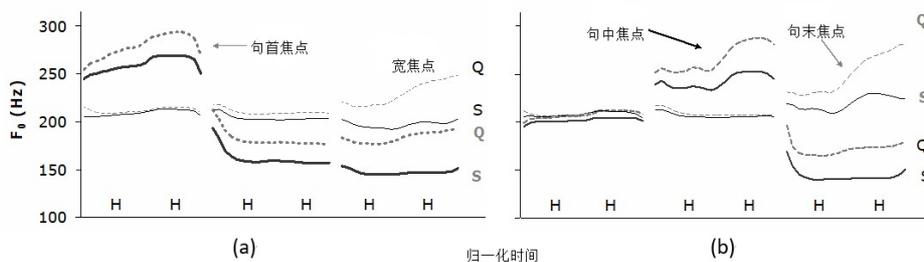


图 5. 汉语句子“张威担心肖英开车发晕”读作陈述句和疑问句时的 F₀ 均值曲线

左图中，焦点位于句首（粗线），或者没有窄焦点（细线）。

右图中，焦点位于句中（粗线），或者位于句末（细线）。实线代表陈述句，虚线代表疑问句。

数据来自 Liu and Xu (2005)

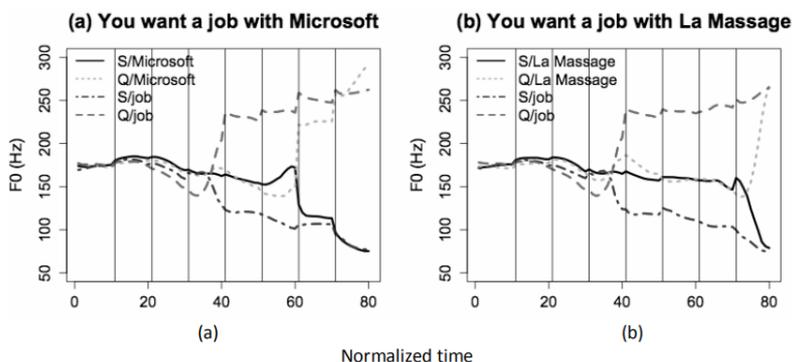


图 6. 美式英语中陈述句(S)和疑问句(Q)的 F₀ 均值曲线。“/”后面的单词是焦点词。数据来自 Liu et al. (2013)

用语音实验确立编码方案的原则还让我们了解到，并不是所有的韵律功能都必须主要由 F₀ 编码。例如，边界标记、分组或短语划分的功能主要由时长来标记（音节和停顿的时长）。一般来说，边界前时长增加意味着两个相邻组间相隔更大（Wagner, 2005; Xu & Wang, 2009）。在句末时，单词时长呈指数性增长（Yuan, Liberman & Cieri, 2006）。进一步的假设认为，音节时长与无声停顿二者结合，成为用时间距离来编码关系疏密的指标（Xu, 2009）。同时，还有一系列的实验研究表明，情感、态度和嗓音吸引力，由音高、频谱密度、嗓音型和语速等通过一套生物信息维度与其它纯语言的功能平行编码（Chuenwattanapranithi et al., 2008; Hsu & Xu, 2014; Liu & Xu, 2014; Noble & Xu, 2011; Xu et al., 2013a; Xu et al., 2013b）。

还有很多其它交际功能也可能用韵律编码，但正如 Xu (2011) 指出的那样，概念上合理的交际功能远远多于能达到一致性的编码方案。具体的编码方案只能通过一个个的实证研究逐渐形成。该策略的普遍性正是 PENTA 模型与 AM 理论最不相同的地方。Pierrehumbert（在本书中）对本书的评论里指出，AM 理论内在地用“少量的语调元素”“界定了单词的语音实现与其意义之间的瓶颈”。但

这些语调元素都是从对 F₀ 曲线的直接观测中得到的（Pierrehumbert, 1980）。相比之下，PENTA 只是默认会有一些声学、感知和历史层面的瓶颈阻止了许多交际功能韵律标记的形成。但这些阻碍是否存在、以及它们的特性如何只能通过受控的实验才能发现。因此，它们不应该先验性地被规定为 PENTA 模型核心的一部分。从这个意义上来说，PENTA 模型中的编码方案不是自由的，而是受限于实证研究发现的语言事实。（未完待续）

参考文献

- Bailly, G. and Holm, B. (2005). SFC: a trainable prosodic model. *Speech Communication*. 46: 348-364.
- Bernstein, N. A. (1967). *The co-ordination and regulation of movements*. Oxford: Pergamon Press.
- Bolinger, D. L. (1972). Accent is predictable (if you're a mind reader). *Language* 48: 633-644.
- Chen, Yiya. and Xu, Y. (2006). Production of weak elements in speech -- Evidence from f₀ patterns of neutral tone in standard Chinese. *Phonetica* 63: 47-75.
- Cheng, C. and Xu, Y. (2013). Articulatory limit and extreme segmental reduction in Taiwan Mandarin. *Journal of the Acoustical Society of America* 134(6): 4481-4495.
- Chuenwattanapranithi, S., Xu, Y., Thipakorn, B. and Maneewongvatana, S. (2008). Encoding emotions in speech with the size code — A perceptual investigation.

- Phonetica 65: 210-230.
- Connell, B. and Ladd, D. R. (1990). Aspects of pitch realization in Yoruba. *Phonology* 7: 1-29.
- de Saussure, F. (1916). Nature of the Linguistics Sign. In *Cours de linguistique générale*.
- Eady, S. J. and Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America* 80: 402-416.
- Gandour, J., Potisuk, S. and Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics* 22: 477-492.
- Gay, T. J. (1968). Effect of speaking rate on diphthong formant movements. *Journal of the Acoustical Society of America* 44: 1570-1573.
- Hombert, J.-M. (1977). Consonant types, vowel quality, and tone. *Studies in African Linguistics* 8: 173-190.
- Hsu, C. and Xu, Y. (2014). Can adolescents with autism perceive emotional prosody? In *Proceedings of Interspeech 2014, Singapore*
- Hyman, L. M. (1993). Register tones and tonal geometry. In *The Phonology of Tone*. H. v. d. Hulst and K. Snider. New York: Mouton de Gruyter pp. 75-108.
- Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative* 246: R1000-R1004.
- Kelso, J. A. S., Southard, D. L. and Goodman, D. (1979). On the nature of human interlimb coordination. *Science* 203: 1029-1031.
- Laniran, Y. (1992). Intonation in Tone Languages: The phonetic Implementation of Tones in Yorùbá. Ph.D. Dissertation, Cornell University.
- Laniran, Y. and Gerfen, C. (1997). High raising, downstep and downdrift in Igbo. In *Proceedings of The 71st Annual Meeting of the Linguistic Society of America, Chicago*: p. 59.
- Latash, M. L. (2012). *Fundamentals of motor control: Access Online via Elsevier*.
- Lee, A., Prom-on, S. and Xu, Y. (2017). Pre-low raising in Japanese pitch accent. *Phonetica* 74(4): 231-246.
- Liu, F. and Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 62: 70-87.
- Liu, F., Xu, Y., Prom-on, S. and Yu, A. C. L. (2013). Morpheme-like prosodic functions: Evidence from acoustic analysis and computational modeling. *Journal of Speech Sciences* 3(1): 85-140.
- Liu, X. and Xu, Y. (2014). Body size projection and its relation to emotional speech—Evidence from Mandarin Chinese. In *Proceedings of Speech Prosody 2014, Dublin*: 974-977.
- Löfqvist, A. and McGowan, R. S. (1992). Influence of consonantal environment on voice source aerodynamics. *Journal of Phonetics* 20: 93-110.
- Mechsner, F., Kerzel, D., Knoblich, G. and Prinz, W. (2001). Perceptual basis of bimanual coordination. *Nature* 414: 69-73.
- Noble, L. and Xu, Y. (2011). Friendly Speech and Happy Speech – Are they the same? In *Proceedings of The 17th International Congress of Phonetic Sciences, Hong Kong*: 1502-1505.
- Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT, Cambridge, MA. [Published in 1987 by Indiana University Linguistics Club, Bloomington.]
- Prom-on, S., Xu, Y. and Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *Journal of the Acoustical Society of America* 125: 405-424.
- Prom-on, S., Liu, F. and Xu, Y. (2012). Post-low bouncing in Mandarin Chinese: Acoustic analysis and computational modeling. *Journal of the Acoustical Society of America*.132: 421-432.
- Silverman, K. (1986). F0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica* 43: 76-91.
- Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, MA: The MIT Press.
- Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7: 71-79.
- Wagner, M. (2005). *Prosody and Recursion*. Ph.D. Dissertation, Massachusetts Institute of Technology.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61-83.
- Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179-203.
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27: 55-105.
- Xu, Y. (2001). Fundamental frequency peak delay in Mandarin. *Phonetica* 58: 26-52.
- Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46: 220-251.
- Xu, Y. (2007). Speech as articulatory encoding of communicative functions. In *Proceedings of The 16th International Congress of Phonetic Sciences, Saarbrücken*: 25-30.
- Xu, Y. (2009). Timing and coordination in tone and intonation--An articulatory functional perspective. *Lingua* 119(6): 906-927.
- Xu, Y. (2011a). Speech prosody: A methodological review. *Journal of Speech Sciences* 1: 85-115.
- Xu, Y. (2013). ProsodyPro — A tool for large-scale systematic prosody analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France*: 7-10.
- Xu, Y., Kelly, A. and Smillie, C. (2013). Emotional expressions as communicative signals. In *Prosody and Iconicity*. S. Hancil and D. Hirst. Philadelphia: John Benjamins Publishing Co. pp. 33-60.
- Xu, Y. and Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics* 18: 125-159.
- Xu, Y. and Liu, F. (2012). Intrinsic coherence of prosodic and segmental aspects of speech. In *Understanding Prosody – The Role of Context, Function, and Communication*. O. Niebuhr. New York: Walter de Gruyter pp. 1-26.
- Xu, Y., Lee, A., Wu, W.-L., Liu, X. and Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. *PLoS ONE* 8(4): e62397.

(下转第 23 页)

现。句首、句中的调群调域会因个性差异而有所浮动。由于入声调较为短促,入声疑问句的句末调域扩展不如舒声疑问句明显。3) 疑问句句首、句中的调群音高比陈述句低,而句末调群音高比陈述句高。

台山话无标记陈述句与疑问句的主要停顿率特征为: 1) 疑问句最后一个音节与陈述句最后一个音节相比,出现了明显的延长。2) 入声疑问句句末音节时长增加比舒声句更为明显。3) 韵律边界前的音节基本都发生了延长。

台山话无标记陈述句与疑问句的主要停顿率特征为: 1) 疑问句的句首的音量比略小于陈述句,而句末最后一个音节的音量比均大于陈述句。2) 音强最高点往往出现在句首音节,韵律边界的音强往往比中间音节高。3) 舒声陈述句音强逐渐减弱趋势比入声句明显。

参考文献

贝先明(2011)广州话陈述句语调的起伏度[J].武陵学刊, 36(03):121-127.
韩维新 王萍(2012)香港粤语无标记疑问句和陈述句语

调的对比分析[J].南开语言学刊,(01):8-19+184-185.

邵慧君 甘于恩(1999)广东四邑方言语音特点[J].方言, 05-25(2):128-135.

石锋 王萍(2017)汉语功能语调研究[M].北京:北京语言大学出版社.

石锋(2013)语调格局:实验语言学的奠基石[M].北京:商务印书馆.

谭洁莹(2016)粤方言台山话语音特征的实验语音学研究[D].暨南大学.

汤翠兰(1997)四邑方言声韵初探[D].国立台湾大学中国文学研究所硕士论文.

王力 钱淞生(1950)台山方音.岭南学报第十卷第二期.67-104.

王萍 石锋(2010)汉语北京话疑问句语调的起伏度[J].南开语言学刊,(02):14-22+185.

向柠(2013)穗、港、澳三地粤语的语调格局——以陈述句的起伏度为例[J].南开语言学刊,(01):65-71.

詹伯慧(主编)(1987)珠江三角洲方言字音对照[M].广东:广东人民出版社,21.

赵元任(1951)台山语料.台北史语所集刊.277-328.

McCoy John(1969) The phonology of Toi-Shan city: a Chinese dialect of KwangTung province[J]. *Bulletin international de Documentation Linguistique* 18: 108-122.

(上接第 104 页)

Xu, Y. and Wallace, A. (2004). Multiple effects of consonant manner of articulation and intonation type on F0 in English. *Journal of the Acoustical Society of America* 115, Pt. 2: 2397.

Xu, Y. and Wang, M. (2009). Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin. *Journal of Phonetics* 37: 502-520.

Xu, Y. and Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33: 319-337.

Xu, Y. and Prom-on, S. (2014). Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning. *Speech Communication* 57: 181-208.

Xu, Y. and Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111: 1399-1413.

Xu, Y. and Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33: 159-197.

Xu, C. X. and Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association* 33: 165-181.

Yuan, J., Liberman, M. and Cieri, C. (2006). Towards an Integrated Understanding of Speaking Rate in Conversation. In *Proceedings of Interspeech 2006*: 541-544.

Zemlin, W. R. (1988). *Speech and Hearing Science — Anatomy and Physiology*. Englewood Cliffs, New Jersey: Prentice Hall.