

Advances in prosody research and how they are related to other areas

许毅、刘芳
Yi Xu, Fang Liu

伦敦大学学院
University College London

[内容提要]

韵律研究近三十年来，尤其是在最近的十几年里得以长足进展。它从一个语音学里的边缘领域逐渐发展为其中的一个重要学科。在这个过程中，声调语言，尤其是汉语的韵律研究做出了关键性的贡献。其主要的发现是，要理解韵律，必须弄清韵律形式如何编码交际功能，而要了解编码的机制，又必须澄清表层韵律形式跟底层发音机制的关系。韵律研究还进一步发现，韵律在语音中几乎是无所不在的，而且音段跟韵律的基本机制并无本质上的差别。所以对韵律的理解有助于解决语音学里的一些悬而未决的关键问题，如音节和协同发音的本质。韵律研究的最新发现还与句法学、语义学、语用学、社会语言学以及历史语言学有密切的关系。最后，韵律研究所采用的实证方法有可能推广至其他语言学领域。

Abstract

Prosody research has seen vast progress in the past 30 years, especially in the recent decade or so. It has evolved from a marginal area to a major discipline in phonetics. In this process, the study of prosody in tone languages, particularly in Mandarin Chinese, has made critical contributions. The major findings indicate that, to understand prosody, it is necessary to explicate how prosodic forms encode communicative functions, and that the understanding of the encoding mechanisms requires clear knowledge about the relation between surface prosodic forms and underlying articulatory mechanisms. It has been further found that prosody is ubiquitous in speech, and that the core mechanisms of prosody are no different from those of segments. Therefore, the understanding of prosody is instrumental in solving a number of key conundrums in phonetic research, such as the nature of the syllable and the essence of coarticulation. The new findings also

demonstrate a close relation between prosody and other areas of linguistics, including syntax, semantics, pragmatics, sociolinguistics and historical linguistics. Finally, the basic strategy of experimental hypothesis testing extensively used in studying prosody should also be applicable to other areas of linguistic research.

[关键词]

交际功能 (communicative function), 功能优先 (function first), 形式优先 (form first), 发音机制 (articulatory mechanism), 音高目标 (pitch target), 目标接近模型 (target approximation model), 音节时间结构模型 (time structure model of the syllable), 功能编码 (functional encoding), 分词 (lexical contrast), 焦点 (focus), 焦点后压缩(PFC) (post-focus compression (PFC)), 话题 (topic), 话轮 (turn taking), 分界 (demarcation), 组合 (grouping), 句型 (sentence modality), 语气 (tone of voice), 态度 (attitude), 情感 (emotion), 实证方法 (experimental hypothesis testing)

一、 概论

(一) 韵律研究的对象和范围

韵律并不是一个有严格定义的概念。一般说来,它泛指一切超音段(suprasegmental)的语音现象,包括音高、音强、音长以及嗓音型。但实际上,韵律跟音段所涵盖的范围有很多相同之处。一方面,跟元、辅音一样,声调是用来区别词义的,但声调常常被归到韵律里。另一方面,许多所谓的音段现象,实质上是韵律现象,如许多元、辅音频谱特性的变异,共振峰走向的动态特征,如下面所介绍的,离开韵律不可能得到充分理解。所以,从研究的角度来看,几乎任何超出用于分词的音段对立的现象都可以是韵律研究的对象。

(二) 功能优先与形式优先

在最近30年的韵律研究研究中,争论最激烈的一个理论问题就是形式与功能谁优先。形式与功能之争本是语言学里至今悬而未决的一个问题。在语法领域里,自从1957年Chomsky出版《句法结构》一书以来,占主导地位的一直是以他为代表的形式学派。该派认为语言的核心是一套形式规则,而交际功能则是语言核心之外的。与之相对的是功能学派,他们认为交际功能是语言的核心,语言的形式来自于功能(Bresnan, 2001; Firth, 1957; Foley & Van Valin, 1984; Halliday, 1985)。形式学派在语言学里的主导地位也给现代音系学

带来很大优势。从生成音系学，到自主音段音系学，到优选论，音系形式都被视为语音的核心。交际意义与语音实现都被认为对音系形式和规则没有决定性的影响。同样，这种认识也主导了有关语言韵律的理论。例如Goldsmith (1976) 提出的自主音段音系学，由Pierrehumbert (1980)、Ladd (1996)、Guessenhoven (2004) 等人发展的语调音系学(intonational phonology), 都是以形式为核心的。Pierrehumbert (1980: p. 59) 便明确说明,“在已有文献里, 判断哪些语调模式在语言学上相互区别, 哪些是同一模式的变体, 有两种方法。一种是根据观察到的F0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。”(我的加重)¹ 这大概是形式优先最明确的表述了。重要的是, 整个语调音系学以及由其衍生的, 现已被广泛应用的ToBI标注体系, 都是建立在由这种形式归纳法得出的描述性语调成分之上的。

很少有人注意到, 语调音系学的思路实际上已经偏离了语言学上普遍接受的音位原则。那就是, 考察一个语言的语音系统, 要先确立哪些语音差别是能分词辨义的, 哪些差别无关紧要。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也是很明显的, 但因为不区别词, 就被认为是同一音位的不同变体。在这两个例子里, 一个音的音位归属决定于它是否分词辨义, 所以是由功能而不只是全靠形式来决定的。而且因为功能在形式之外, 所以这样的定义是非循环性的。音位是现代语言学最重要的发现之一。这个发现, 给语言学带来了长足进步。但是, 现代音系学对形式的强调使人们产生这样的印象: 语音单元的确立和变换可以完全独立于意义。所以语调音系学可以几乎完全基于形式就建立起一套完整的语调体系。而且, 即使后来曾试图考察这个已经成型的体系跟意义如何对应 (Pierrehumbert & Hirschberg, 1990), 也没有给这个已经确立好的体系带来任何结构上的变动。

与形式优先相对应的是功能优先或至少是注重功能的理论。在韵律研究里, 对交际功能的考虑一直占有一定的地位, 多种注重功能的理论也一再出现。这些理论大多是自成体系, 而且每一个都似乎是有关韵律的整体理论。但实际上, 每一个理论都只涵盖了韵律功能的一小部分, 而且各个理论所涉及的功能有许多重叠之处。言语行为 (Speech act) 理

¹ “In the literature, one can distinguish two approaches towards the problem of establishing which intonation patterns are linguistically distinct and which count as variants of the same pattern. One approach attacks the problem by attempting to deduce a system of phonological representation for intonation from observed features of F0 contours. After constructing such a system, the next step is to compare the usage of F0 patterns which are phonologically distinct. The contrasting approach is to begin by identifying intonation patterns which seem to convey the same or different nuances. The second step is to construct a phonology which gives the same underlying representation to contours with the same meaning, and different representations to contours with different meanings. ... The work presented here takes the first approach, in fact, it stops at the first step in the first approach.”

论只考虑以言为行的功能，如询问、命令、警告、许诺等等 (Austin, 1962)。信息结构 (information structure) 理论只考虑话题、焦点等等 (Büring, 2007; Halliday, 1967)。进一步考察这些理论的细节还会发现，它们跟语调音系学的做法实际上有一个本质的相通之处，那就是，跟 Pierrehumbert (1980) 所描述的方法相似，也总是先建立一套独立于意义的基本语调单元，如 Bolinger 的 "accents" (Bolinger, 1986, 1989)，Halliday 的 "tones" (Halliday, 1967)。这种做法实际上是基于一个根本的假设，那就是，语调单元是可以从语音信号里直接观察到的。或者说，可以直接观察到的韵律模式就是语调基本单元本身，语调研究的目的不过是确立那些模式是最基本的，以及他们如何跟功能、意义对应。所以，多数重功能的理论也没有真正做到功能优先。

（三）功能优先的必要性、困难性和可行性

关于语言的一个不争的基本事实是，语言是用来传达交际意义的。所以弄清楚这种意义传达如何实现应该是语言学的根本目标之一。但是当代语言学对此并不是特别关心。相反，主流的理论主要致力于解释为什么有些形式被人判断为合语法的，有些被判断为不合语法的。很少有人想到，语法判断能力很可能是一种派生现象 (epiphenomenon)，而其原生现象则是意义的传达。如果真是如此，只有弄清楚意义传达的机制，才有可能真正理解语法现象本身。

同样，由于韵律传达交际意义也是不言而喻的事实，我们就有必要首先弄清楚韵律是如何实现传达交际意义的功能的，而不是首先致力于弄清在某个语言里，哪些韵律模式是合法的，哪些是非合法的。但是考察韵律的功能有两大困难。第一，跟音段音位相比，韵律的功能很难确定。音段音位的主要功能似乎很清楚，就是分词：音位不同，词就不同。而且音段音位多在文字中有直接反映。所以，在多数情况下，判断音段音位是很容易的 (当然也有许多例外，见 Chao (1934))。韵律的功能远远不象分词那么清楚，除了标点符号，很难在书面语言里找到直接的标志。第二，韵律中的功能成分跟表层的语音形式并不直接对应。两者之间至少有三重分离 (Xu, 2004a)：(1) 目标实施 (target implementation)、(2) 目标设定 (target assignment)、(3) 平行编码 (parallel encoding)。目标实施指的是实现底层形式的发音过程。由于这个过程种种生理、物理局限，表层形式不可能等同于理想的底层目标。尤其是在不同的语音环境里，同一发音目标的表层形式会很不一样。目标设定指的是交际功能与底层发音目标的对应也常常是一对多的，这种非单一对应跟发音机制并不直接相关，而是随语言而变的。平行编码指的是，由于有多种交际功能需要通过有限的声学参数来传达，语流中的任何一个时段都同时载带着几种功能的信息。所以，任何时段的声学模式都不可能直接对应于单一的交际功能。

这样的多重分离显然给我们寻找功能与形式的对应带来很多困难。但是，这些都是不可回避的困难，如果不解决，对韵律的认识就不可能有真正的进步，因为正是这些困难阻碍了以往许多企图建立完整韵律学理论的努力。换一个角度来看，认清这些困难的来源，反倒有可能帮助我们找到理解韵律的突破口。从近十几年研究进展来看，一套行之有效的策略已经逐渐成形。策略之一是先考察低层次的机制，从而弄清第一重分离的原理。策略之二是从显而易见的功能入手，逐个考察，而不是试图一次就穷尽所有的功能。策略

之三是采用严格控制的实验，而不只是凭非系统的观察。下面将介绍实施这些策略带来的一些对韵律的新认识。

二、从发音机制和交际功能角度理解韵律

(一) 发音机制

从交际功能到表层形式，即语音信号，必然有一个编码过程。这个过程应有一套编码机制。不清楚这套机制，就无法理解交际功能与表层形式的关系。编码离不开具体的设备。说话的设备是由神经系统(包括脑和外围神经)、肌肉、以及发音器官的骨骼和软组织构成的。编码就是操作这些设备的过程。虽然我们目前对这些设备的认识很有限，但至少不难想像，编码机制跟编码设备的特性有密切的关系。不理解这些特性，就不可能理解编码的机制。总的说来，我们对语言的外围设备，即发音器官，比对中央设备了解得更多一些。所以语音学可以为我们的认识编码机制提供一个起点。

图1显示的是发音动作的最基本模式，它不仅适用于声调和语调，也适用于语音的其他方面，包括音段 (Xu & Liu, 2006, 2007)。每个动作都是从发音器官的某个初始状态开始，然后在给定时段内，即所分配的时间里向一个理想状态不断接近。理想状态就是当前音的底层标准状态。初始状态来自于前一个动作的结束状态，或者是发音的起始状态。如果给定时长不够，理想状态就会达不到，如图中所示。这种基本的动作模式很可能不是语音特有的，而是人或者动物的任何动作的基本模式。从这一点上讲，发音动作其实没什么特殊之处。不同的只是使用的器官、理想状态的特性、动作的力度、以及给定的时长。

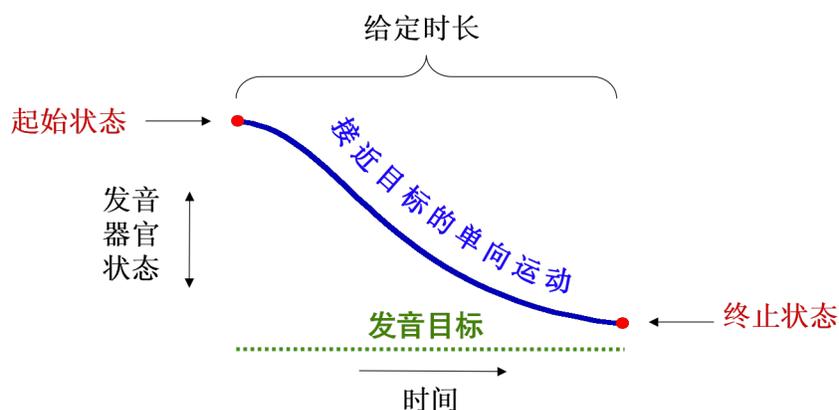


图1：发音动作的基本模式。

图2显示的是建立在发音基本动作模式上的适用于声调和语调的目标接近模型 (Xu & Wang, 2001)。跟发音动作基本模式相比，目标接近模型添加了如下成分：

- 一、音高目标 (即模型里的 pitch target) 既可以是静态的，也可以是动态的。
- 二、初始状态不仅包含定点状态，还包含从前一动作继承下来的速率 (velocity) 和加速度 (acceleration)。
- 三、音高目标所分配的时间段与音节所分配的时间段等同。

其中一、二两项其实已经隐含在图1所示的发音动作基本模式里，目标接近模型只是把它们明确化。

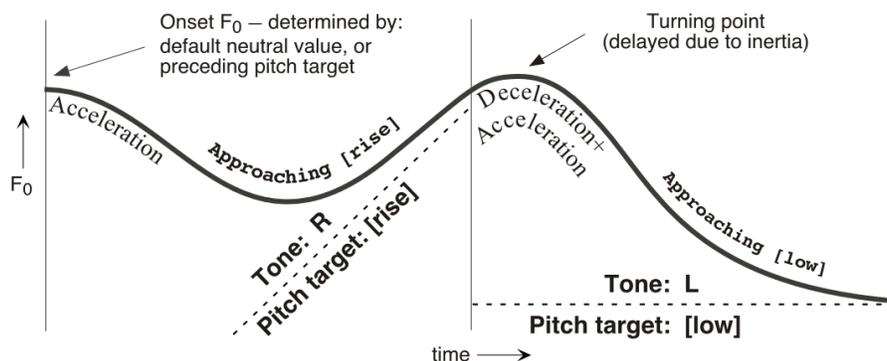


图2：目标接近(TA)模型。图中的竖线代表音节边界，虚线代表底层音高目标，连续的粗实线代表表层基频曲线 (Xu & Wang, 2001)。

目标接近模型可以有效地解释连续语言里的基频曲线。图3显示的是北京话里声调连读的平均基频曲线。在每个图里，只有第二个音节的声调是不一样的，而其他音节的声调保持不变。我们可以看到，第二音节的终点基频由于声调不同而差别很大，结果第三音节的起点基频也大不相同。但是，在第三音节里，所有的基频曲线都随时间推移逐渐向一个直线靠拢 (如箭头所示)。而这个直线的值跟该音节声调的理想特征相符：第一声为高平，第二声为上升，第四声为下降。所以，图3里的基频曲线跟目标接近模型是一致的。唯一不太一致的地方是第二音节的声调对第一音节也有所影响。但这种影响是逆化的，即第二音节的基频越低，第一音节的基频越高。这种逆化影响的机理还不清楚，有待于今后进一步研究 (初步探讨见 Gandour et al., 1994; Xu, 1997)。

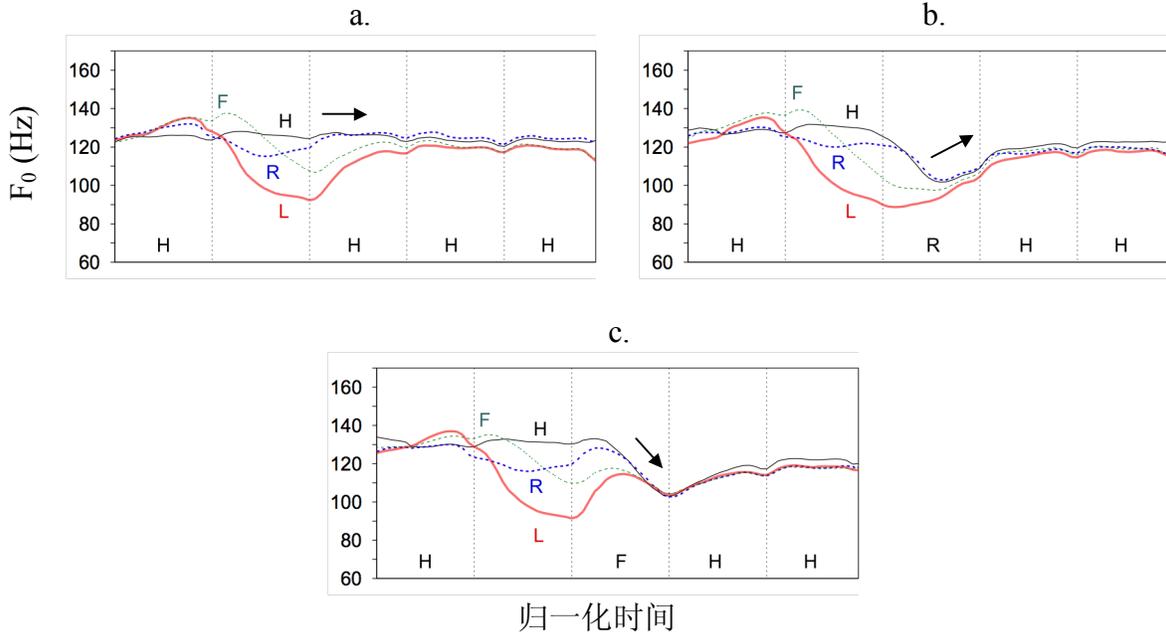


图3: 北京话“猫咪 / 迷 / 米 / 蜜 摸 / 拿 / 卖猫咪”的平均基频曲线。图中 H、R、L、F 分别代表第一、二、三、四声。数据来自 Xu (1999)。

通过目标接近模型我们不仅可以看到实现单个声调的机理，而且可以看出，通过控制该过程的各个方面就有可能实现多层次的编码。如图4所示，不光是音高目标 (如 [high], [rise], [low], [fall])，调域、力度、时长，都可以分别控制。在图4a里，只是音高目标有变化。在图4b里，第一、二音节的调域被抬高并加宽，第三、四音节的调域被降低并压缩。但每个音节基频曲线的基本形状跟图4a里仍大体一致。在图4c里，第一、四音节的力度加强，导致目标实现更加充分；第二、三音节的力度减弱，致使目标实现很不充分。在图4d里，第一、四音节的时长增加，第二、三音节的时长缩短，其结果跟图4c里加强或减弱力度的效果相似。在下一节里我们可以看到，图4所显示的所有控制手段都在交际功能的编码中得到充分应用。

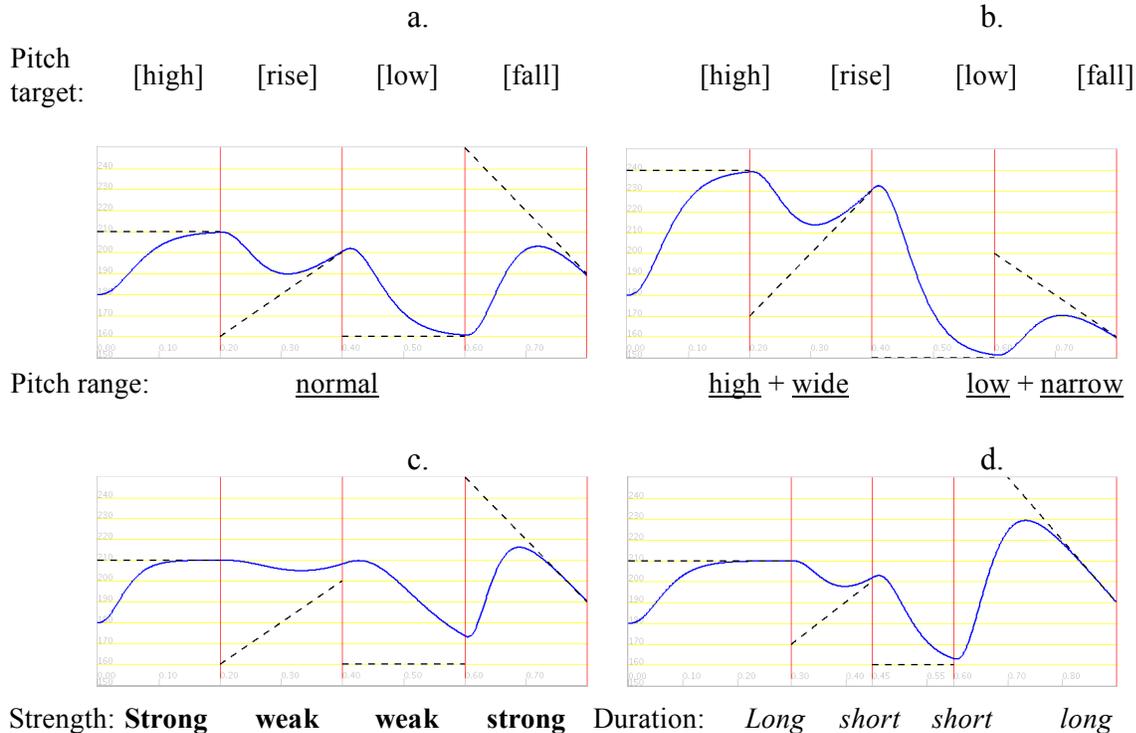


图4: 音高目标、调域、力度和时长对表层基频曲线的影响。图像生成于 <http://www.phon.ucl.ac.uk/home/yi/qTA/>

(二) 交际功能及其编码

前面已经谈到, 从功能的角度来看, 说话的目的不是为了表现自己的语法是正确的 (除非是应付口语考试), 而是要转达各种信息。这些信息也并不只是局限于词义和由句法结构体现的命题意义(propositional meaning), 还包括很多其他层次的意义。这些意义可以统称为交际功能。其中跟基频有关的就由多重功能, 包括:

1. 分词 (Lexical contrast)
2. 焦点 (Focus)
3. 句型 (sentence type)
4. 话题 / 话轮 (Topic / turn taking)
5. 分界 / 组合 (Demarcation / grouping)
6. 情感 / 态度 (Emotion / Attitude)

基本上是一维的基频怎么能参与转达这么多层次的功能呢? 图4已经为我们提供了一个初步的线索。从那里我们看到, 目标接近模型的各个控制参数都可以分别调节, 而且每种调节对基频曲线都有明显的影响。所以每个交际功能都有可能用某个或某套目标接近模型参数来编码。图5是“平行编码及目标接近模型”(Parallel Encoding and Target Approximation model), 简称 PENTA, 用来模拟从交际功能到表层语音的编码流程 (Xu, 2004b, 2005)。图左的多个小方框代表的是各个具体的交际功能, 它们每个都有一个对应

的编码模式。每个模式是由一套参数构成。这些参数控制图2里显示的目标接近模型 (Target Approximation)。该模型象图4显示的那样通过调节各个音高目标来生成表层的基频曲线。下面的讨论说明分词、焦点、句型、话题、分界 / 组合以及情感、态度和语气如何通过目标接近实现语音编码。

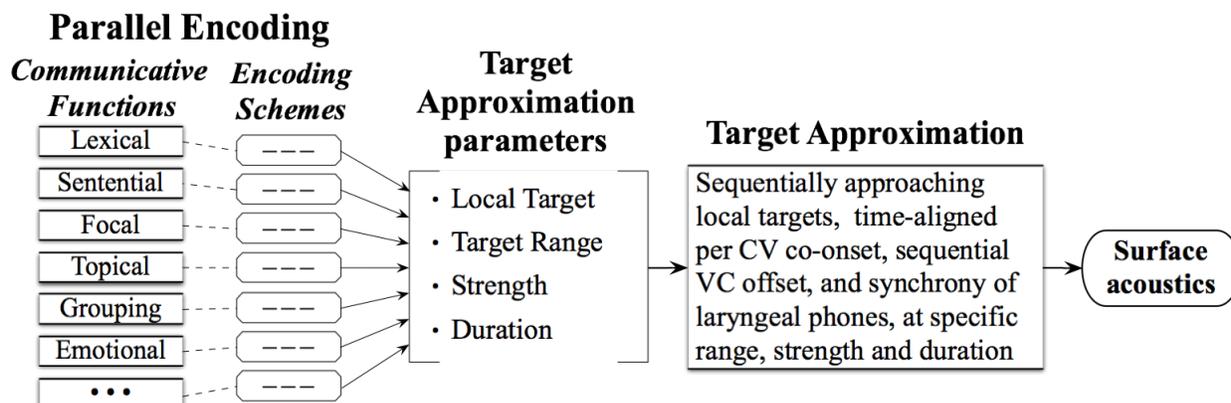


图5: 平行编码及目标接近模型 (Parallel Encoding and Target Approximation model — PENTA, Xu, 2004b, 2005)

分词 (Lexical contrast)

词是语义和语音的综合体，也是语音和语义的联结点。分词就是将不同的词区分开来。这是语音的最基本功能之一。分词的手段非常多：元音、辅音、声调、发声型、词重音、时长，等等。这其中元音和辅音大概是世界上所有语言都用的。声调、发声型、词重音、时长都只是一部分语言使用。以声调为例，在汉语这样的声调语言里，声调的不同可以使一个声、韵母都相同的音节代表不同的语素。在北京话里，一个音节可能有五个不同的声调：阴、阳、上、去、轻。早期的研究就已经发现每个声调在单念时都有特定的音高模式 (白涤洲, 1934; 赵元任, 1922)。这些模式的独特性足以使它们在感知上有效地相互区别开来 (Gandour, 1983)。不过在连续语流里，如图3所示，声调的基频曲线会大大偏离单念的形式。但是正如目标接近模型所模拟的，尽管有这些偏离，到了每个音节尾部，基频曲线会接近或达到理想的目标。从感知试验的结果上看，声调的这种实现方式是有效的。即使在连续语流中的声调识别率也可以接近90% (Prom-on, Xu & Thipakorn, 2009)。在最近的神经网络模拟实验中发现，一旦把表层基频曲线转换成速率曲线(velocity profile)，一个无监督学习系统 (unsupervised learning system) 就可以通过处理这些曲线自动学会识别连续语流中的声调 (Gauthier et al., 2007a, 2007b)。

再来看词重音。在英语里，象 *subject*, *contract* 这样的词，当第一个音节重读时是名词，第二个音节重读时是动词。早在1958年，Fry 就通过感知实验发现这样的动、名词之分可以靠基频、时长和音强来辨别，但是辨别效果最好的是基频。² 只要两音节的基频

² Fry (1958) 的发现常常被误解为是关于句重音 (即焦点) 或一般重音 (prominence) 的感知。实际上在他的实验里，受试只是判断听到的是动词还是名词，并没有作重音的判断。

之差为5 Hz，感知上便没有任何混淆。这跟我们考察英语里实际的基频曲线时得出的结论是一致的 (Xu & Xu 2005)，也就是，英语里非焦点处的轻、重音节之间的基频之差很虽然很小，但总是存在。所以从功能的角度来看，这些起伏虽然很小，却载带着清晰的分词信息，因为它们的幅度已足以区别象 'subject/sub'ject 这样的名词和动词。以往的很多理论认为焦点后的基频曲线完全是平的，因为能从基频曲线上直接看到的上下起伏是很小的 (Bolinger, 1986; Pierrehumbert, 1980)。所以，脱离功能的观察会导致对表层形式里重要细节的忽视。

如果对功能和发音同时考察，我们可以看到编码机制里更多的细节。例如，在最近的研究里我们发现，从目标接近模型的角度看，英语象汉语一样每一个音节都有一个音高目标，但这个目标会随音节在词里的位置、句中焦点的位置以及句型而变 (Liu & Xu, 2007; Xu & Xu, 2005)。这样看英语和汉语在语调上的差别比一般认为的要小。但跟汉语不同的是，英语音节的音高目标不是完全由分词功能，即词重音决定的。这在下面会进一步讨论。

还有一个跟词重音有关的问题也是从功能的角度更容易看清。那就是，有分词功能的词重音只在某些语言里有，如英语、德语、俄语。另外有许多语言不用重音分词，如法语、朝鲜语、波兰语。从分词角度上看，北京话里的轻声也相当于词重音 (Chen & Xu, 2006)。在不用重音分词的语言里，也常常可以观察到一些貌似词重音的语音现象。人们往往把它跟词重音相提并论，并试图找到一种跨越不同语言的、以形式为定义的抽象的重音 (prominence)。从功能的角度看，这种抽象的重音没有多大意义。因为不同功能的编码必须是不一样的，不然它们就会相互混淆。这在下面关于焦点的讨论里可以清楚地看到。

焦点 (Focus)

我们说一句话，有可能特别强调其中某一部分。这种强调现在一般叫做焦点 (focus)。例如在“是我请客”这个句子里，强调的是“我”。在这里强调的手段是用“是... (的)”这种句型。但是也可以说“**我**请客”，只是在语气上额外加重“我”。所以焦点既可以用句法也可以用语音来实现。那么语音上如何实现焦点呢？以往的研究发现，音高变化是焦点的主要特征之一。可是在汉语里，音高已经用于声调，怎么能又用来表现焦点呢？在图4b里我们已经看到，在音高目标基本不变的条件下，调域是可以作调整的。根据系统实验的发现 (Xu, 1999; Liu & Xu, 2005)，在北京话里，跟焦点有关的调域控制分三个区域：焦点前、焦点处、焦点后。焦点前的调域不变，也就是跟没有焦点的句子相似，焦点处的调域加宽，焦点后的调域降低并压缩。图6显示的是一个全是阴平字构成的句子如何通过三区段调域调整实现焦点。

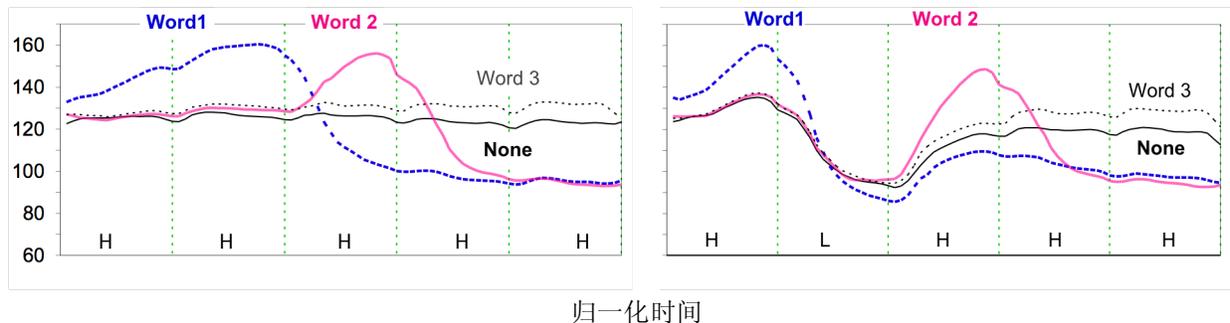


图6：北京话“猫咪/米摸猫咪”四种焦点的平均基频曲线。数据来自Xu（1999）。

进一步的研究发现，英语里焦点的实现方式跟汉语的很相像，也用三区段调域调整，而且这些调整并不抵消词重音对音高影响，就跟北京话里焦点并不抵消声调对音高的影响一样(Xu & Xu, 2005)。但英语有一点跟北京话不同的是，焦点处的词如果最后一个音节是重读，该音节的音高目标会由 [high] 变成 [fall]。这种从功能角度看英语语调的观点跟当前流行的观点是不一样的。按照流行的语调音系学或AM理论，词重音只是在载带 pitch accent 时才会有音高上的表现。但正如在前面提到过的，他们确定 pitch accent 的出现与否完全取决于基频曲线有无一目了然的运动 (Pierrehumbert, 1980)。由于焦点后的基频运动幅度很小，于是就认定不值得考虑 (Beckman & Pierrehumbert, 1986)。所以，完全从形式出发，有可能连形式本身的描述都会不准确。因为描述必然要抽象，抽象就免不了取舍。只有考虑了功能之后才能知道那些哪些该取哪些该舍。

从其他许多研究的报告来看，焦点的三区段调域调整的编码方式也存在于别的语言里，包括希腊语、瑞典语、德语、日语、朝鲜语、芬兰语。但是最新的一些研究发现，并不是所有语言都用三段式的焦点编码，很多语言都缺乏其中的焦点后调域压缩 (PFC)。有趣的是，虽然同属于汉语，台湾的闽南话，香港的广东话都没有 PFC (Chen, Xu & Wang, 2009)。尤其令人惊讶的是，跟闽南话一样，台湾国语也没有 PFC (Chen et al., 2009)。根据 Zerbian, Genzel & Kügler (2010) 的总结，许多非洲语言很可能也没有 PFC。更重要的是，从目前的文献来看，PFC 只存在于印欧语系，阿尔泰语系的许多语言以及中国的北方话。这就引出一种可能性，即 PFC 只有一个单一的历史起源，其后通过语言接触逐步传播，到目前为止只扩散到世界上的一部分语言里 (Chen et al., 2009; Xu, Chen & Wang, forthcoming)。还有一种更为极端的可能，那就是 PFC 完全是继承的，并无传播。当然这些假说都必须通过大规模的，跨语言的实验研究才能得以核实。

句型 (sentence type)：疑问 / 陈述之别

区别陈述句与疑问句历来被认为是韵律和语调的主要功能之一，有时甚至被作为语调的代名词。很早的研究就注意到，陈述句多以低调结束，疑问句多以高调结束。正因为如此，大多数对这种句型语调的研究只关注句尾的语音表现。这种做法在韵律音系学里发展到了极端。它假定句尾基频的高低之分来自于两个不同的边界调 (boundary tone)：%H 和 L% (Pierrehumbert, 1980; Pierrehumber & Hirschberg, 1990)。但是其他研究早就注意到

疑问、陈述之差并不局限于句尾。Thorsen (1980) 通过切音感知实验发现，在丹麦语里，陈述、疑问之差从句中第一个重读词就开始了。O'Shaughnessy & Allen (1983) 和 Eady & Cooper (1986) 发现，在英语里，从焦点开始，每个重读词的基频都随陈述、疑问而变。我们也同样在北京话里看到类似的现象 (Liu & Xu, 2005)，见图7。但是通过更详尽的分析发现，疑问和陈述的音高差别从焦点开始到句尾以双指数形式加速增大。所以，虽然句尾的音高模式受句型影响最大，疑问、陈述之差的语音编码不是只局限于句尾。同时，加速度式的音高差别又能解释为什么早期的研究只注意到句尾的基频变化。

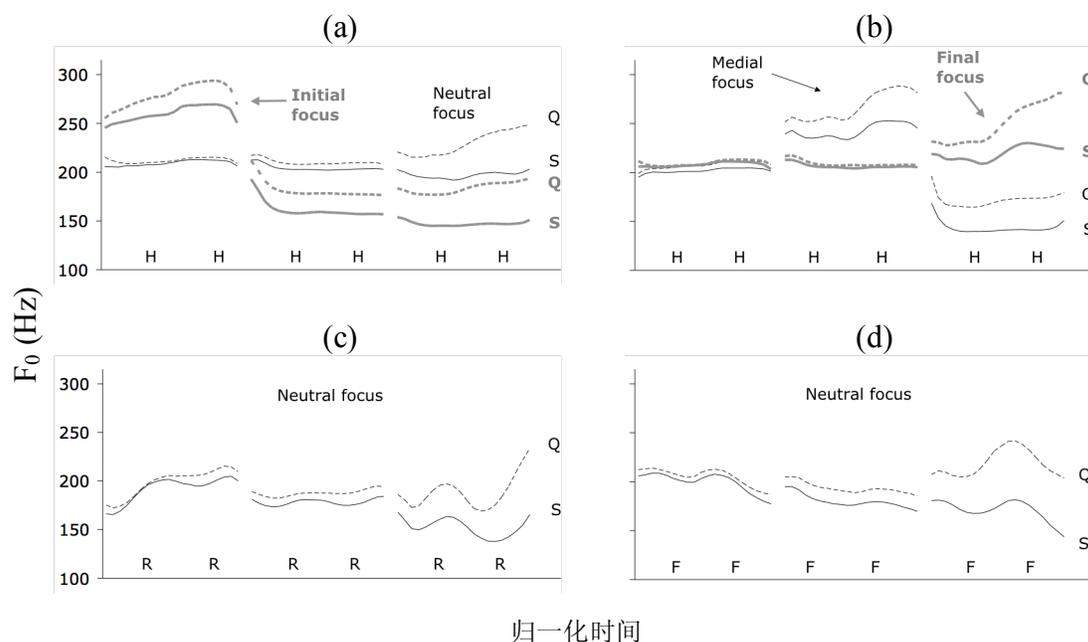


图7：北京话陈述、疑问句的平均基频曲线。(a)、(b)：“张威担心肖英开车发晕”，(c)：“王梅怀疑刘宁划船着迷”，(d)：“叶亮害怕赵丽睡觉做梦”。在(a)、(b)里各自有两种不同的焦点，在(c)、(d)里焦点均为中性。数据来自 Liu & Xu (2005)。

Liu & Xu (2005) 所发现的双指数型音高变化还进一步表明，陈述、疑问之差不仅来自于疑问句语调的加速上升，而且还来自于陈述句语调的加速下降。这种加速下降可以解释早已报道过的句尾下降 (final lowering, Liberman and Pierrehumbert, 1984)，但不是象早期所说的是基于生理机制，而是源于功能性编码。

跟北京话相比，英语里的句型编码有更进一步的复杂性。O'Shaughnessy (1979) 和 Eady & Cooper (1986) 都注意到，在美国英语里，疑问句不仅句尾基频抬高，而且处于焦点处的词的基频呈上升状。Liu & Xu (2007) 通过更系统的研究发现，英语疑问句里每一个重读音节的基频都呈上升状，随后的非重读音节跟随抬高。而且焦点后的基频大大抬高，但调域显著压缩，如图8所示。这里需要指出的是，以往的语调理论大都把这些模式理解为各自独立的现象。尤其是语调音系学的处理最为系统，把焦点处基频曲线定为核心调 (H* 或 L*)，把句尾的基频模式归结于边界调 (H% 或 L%)，把焦点到句尾之间的基频归结于短语调 (phrase tone)，把疑问句尾基频的进一步升高归结为一个叫做 upstep 的语音

实施规则。Pierrehumbert & Hirschberg (1990) 则更进一步试图为每一个成分找到具体的意义。但是，到目前为止，并没有实验证明这些成分可以单独运作。相反，系统的实验一再发现，只要句型和焦点已定，说话人总是相当一致地使用上面所说的所有的句型标记。这说明，疑问、陈述之差才是实际的功能之别，也就是实际的音系之别，而所有具体的韵律标记都只不过是句型编码的组成成分。

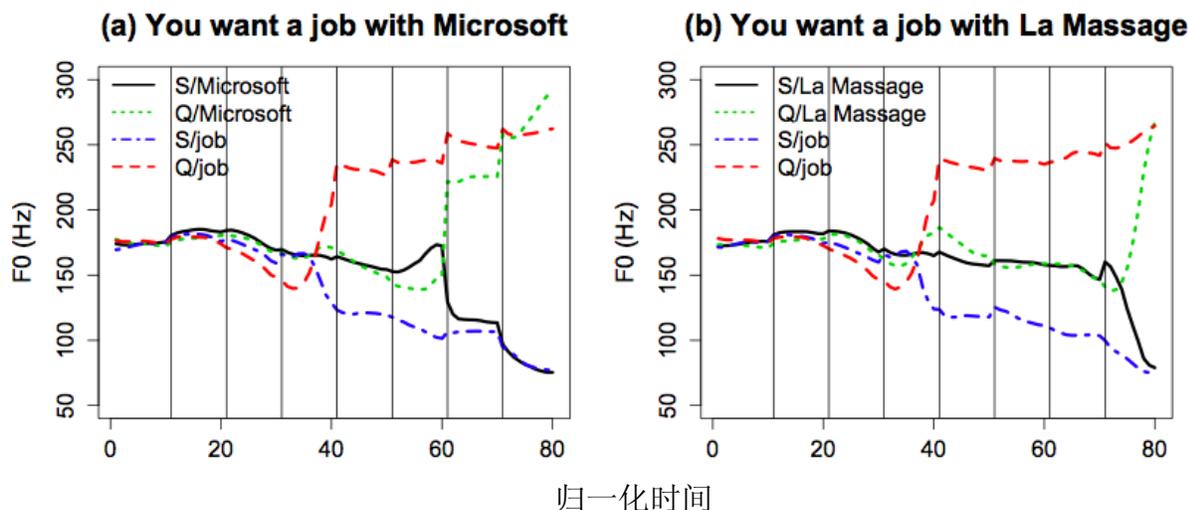


图8：美国英语陈述、疑问句的平均基频曲线。S = 陈述，Q = 疑问，“/”之后的词为焦点所在。数据来自 Liu & Xu (2007)。

最新的一些研究发现，跟焦点一样，上面所描述的通过加速抬高基频来表达疑问并不是普遍用于世界各语言的。在非洲中部的一群语言里，疑问句是靠句尾音节加长，气声，以及抬高整句的基频来标志的 (Rialland, 2009)。听觉上这些标志一点也不象其他许多语言里的疑问句。有趣的是，这些语言都集中于一个地理区域，互相之间有各种各样紧密的联系。所以，跟焦点一样，疑问句的编码模式也很可能各自有单一的历史起源。而且从编码模式的地理分布来看，这些起源要比焦点起源的历史更为久远。

话题 (Topic)

话题指的是当前讨论的议题，它是信息结构 (information structure) 理论的主要组成部分之一 (Büring, 2007; Halliday, 1967)。韵律对话题的表现，主要是显示当前的议题是新的还是旧的。多项研究发现，一段话的第一个句子，也就是载带新话题的句子，往往基频较高，调域较宽 (Lehiste, 1975; Nakajima & Allen, 1993; Sluijter & Terken, 1993; Terken, 1984; Tseng, et. al., 2005; Umeda, 1982)。以往有关话题的理论，有一些悬而未决的问题。一是对比话题 (contrastive topic) 是否有特定韵律特征，二是话题的新旧 (newness) 是否取决于单词在语境中的新旧。Wang & Xu (2006) 发现，话题和焦点的表现可以清楚地分开。在句首的焦点不仅加宽焦点处的调域，而且还压缩并降低焦点后的调域。新话题只是提高句首的调域，并不改变焦点后的调域。由此可见，所谓对比话题的韵律表现很可能跟焦点有关，而不一定是话题本身的表现。当然这还需要今后通过专门的实验进一步探讨。

Wang & Xu (2006) 还发现，一旦话题和焦点的条件同时得到控制，一个词是否在前文里已经出现过并没有基频上的差别，只是已出现过的词时长略短。这表明，新话题带来的句首调域抬高并不是由于某个词在语境中是刚出现的。Wang & Xu (2006) 还发现，句首的基频渐降 (declination) 相当一部分来自于由声调中[低]的成分造成的降阶 (downstep)，其降幅不小于由句首焦点带来的降幅。所以，任何一个单一成分的渐降模型，如 Cohen, Collier & 't Hart, (1982), Liberman & Pierrehumbert (1984), Shih (2000)，都没有指出基频下降趋势的全部来源。

分界 / 组合 (Demarcation / grouping)

在当前流行的语调音韵学 (intonational phonology) 里一个重要的观念是，韵律结构比句法结构要简单，因此很多复杂的句法结构没有韵律单元与之相对应 (Ladd, 2008; Selkirk, 1984)。我们先抛开句法为什么要跟韵律对应这个基本的问题，这种认识的一个关键的前提是，语调韵律学里的韵律单元都是由音高模式代表的 (Pierrehumbert, 1980)。在本章一开头就说到，韵律的语音特征并不限于基频，而是涉及所有非音段的因素，包括音高、时长、音强、频谱和嗓音型。Wagner (2005) 通过实验考察了句法中多种递归结构的韵律特征，结果发现诸如 “A and B and C” 和 “(A or B) and C” 这样多层递归的结构都有韵律表现，但是表现的方式是时长而不是基频：边界强度 (boundary strength) 高的地方时长就长，反之亦然。其实这种跟边界强度相关的时长变化早已有报道 (Klatt, 1975; Lehiste, 1973; Oller, 1973; Turk & Shattuck-Hufnagel, 2000; Wightman et al., 1992)。Xu & Wang (2009) 考察了北京话里跟分界、组合相关的声学特征，发现在二、三、四字词或词组里，处于边界的音节长，靠中间的音节短。该研究还发现，声调基频模式的变化跟音节时长的变化完全一致，较长音节里的基频起伏幅度大于较短音节，如图9所示。但是这种变化可以由图4d所示的发音机制解释。同时，Xu & Wang (2009) 还发现全阴平词里的基频完全不随音节长度变化。所以，跟英语相似，北京话里在词和短语层次的组合、分界完全是靠时长而不是基频来标记的。

Xu (2008, 2009) 进一步提出，由于更大的边界，如句子、段落，常常伴有停顿，边界处音节和停顿的长度相加后就成为 Inter Onset Interval (IOI)。IOI 实际上是相邻音节之间的时间距离，而所谓边界实际上是单元之间的关系距离。所以，从编码角度看，IOI 是作为疏密标记 (affinity index) 而起作用的。

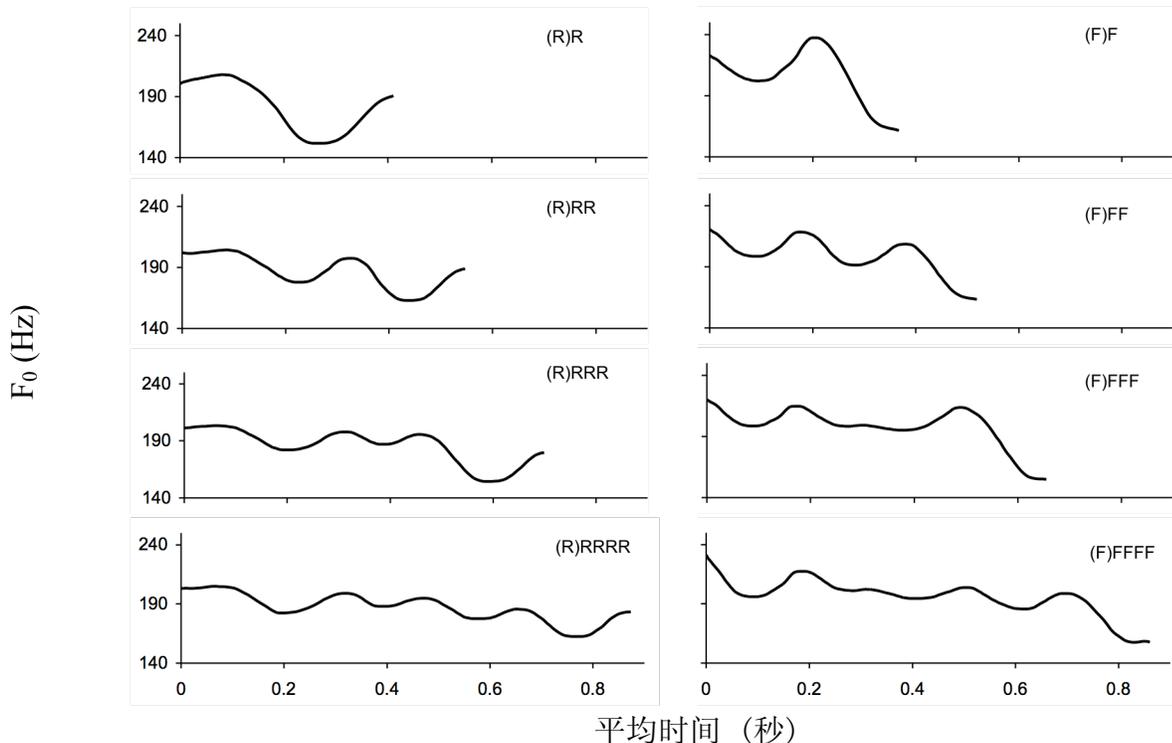


图9：北京话一、二、三、四字词在句中的平均基频曲线。图左从上至下：南、云南、云南人、云南人民，图右：用、外用、外用药、外用药量。数据来自 Xu & Wang (2009)。

语气、态度、情感 (Tone of voice, Attitude and Emotion)

现在一种普遍的看法是韵律、语调主要是用来表达语气、态度和情感的。许多韵律和情感的理论因此力图找到直接对应于情感的语调模式，但至今没有明确的结论 (Bolinger, 1989; Gussenhoven, 2002; Halliday, 1985; Kohler, 2005; Mozziconacci, 2002; Pierrehumbert & Hirschberg, 1990)。最新的一些研究采用了不同于传统的思路。Chuenwattanapranithi et al. (2008) 发现，体积码 (size code) 的假说有助于发现区别喜、怒的区别特征。该假说源于Morton (1977) 对哺乳和禽类动物鸣叫的研究。他发现，与进攻性行为相伴的鸣叫基频偏低，音质粗糙，与恐惧、顺从行为相伴的鸣叫基频偏高，音质柔润。他的假设是，前者是为了显得体积庞大，以吓退听者，后者是为了显得体积娇小，以吸引听者。Ohala (1984)将该假设推广至人类。他提出，改变声道长度亦可达到相同效果，比如微笑时嘴角后拉就是为了缩短声道。Chuenwattanapranithi et al. (2008) 请泰国的受试听一套由三维发音合成器 (articulatory synthesizer, Birkholz, Jackel & Kröger, 2006) 生成的泰语单元音。正如体积码的预测，用较长声道和较低基频合成的元音被判断为来自块头大的人，并且听起来象生气。而用较短声道和较高基频合成的元音被判断为来自块头较小的人，并且听起来象高兴。后来又有一系列实验，让受试听通过声学修改的自然单词和句子，得到了相似的结果 (Xu & Kelly, 2010; Xu, Kelly & Smillie, forthcoming)。这些实验告诉我们，情感、态度、语气很可能自有一套编码，而且是凌驾在其他韵律成分之上，而不是取而代之的。这样就完全符合PENTA模型里平行编码的假设。当然，这一全新方式的情感研究刚刚开始，还需今后大量的后续工作。

（三）功能编码小结

以上对功能编码的讨论远远没有涵盖韵律的全部，但已经让我们看到了跟传统思维很不一样的思路。首先，从功能角度看，分词只是无数交际功能中的一个，其他功能也一样重要。所以，以往认为音段音位是语音的核心在相当程度上阻碍了我们对语言的深入理解。第二，韵律并不只是语调，语调也不只是局限于某几种。所以，信息结构 (information structure) 理论、句尾语调 (sentence final intonation) 理论、言语行为 (speech act) 理论、语气 (tone of voice) 理论、节奏 (speech rhythm) 理论，等等，都各自只涉及了韵律的一小部分，不可能是全面的韵律理论。第三，从功能的角度来看，每个功能都是独立、自成一体的，有自己的一套内部机制。功能之间没有管辖关系，也没必要有一个大一统的韵律层级结构 (hierarchical structure)。由于韵律编码的语音手段有限，各功能之间免不了有竞争和相互影响。其结果是各自形成一套独特的编码机制：分词主要靠局部发音目标，焦点靠三段式调域调节，话题靠调节句首调域，句型靠调节句子后部的调域，分界、组合靠调节时长，情感、态度、语气靠调节全局的基频、频谱、嗓音型、速率以及音强。最后，如果某一假设的功能完全没有独特的编码标记，不管该功能以往被描述得多么有道理，它的存在与否应该重新考虑。

三、韵律研究对音段研究的贡献

在上面的讨论中我们看到，声调的研究发现声调发音的基本机制是在给定时段内向理想目标过渡，而对这个机制各个方面的控制可以用于多种韵律功能的编码。所以声调的研究深化了语调和韵律的研究。但是声调发音机制的发现还有助于我们对许多音段问题的理解。下面只从两个方面略加介绍。

（一）对音节和协同发音的新认识

前面刚讲到，分词功能的主要编码手段是规定局部发音目标，其中包括元音、辅音和声调，所以声调和音段的发音机制应该有相通之处。由此推知，二/（一）节里讨论的目标接近 (TA) 模型也因该适用于音段。但问题是，音段的具体目标是什么，对应的时间段又是什么？Xu & Liu (2006, 2007) 从北京话和英语里的滑音 [w, j, ɹ] 入手探讨了这个问题。该研究利用近年发现的基频转折点与音段之间精确的时间对应 (Arvaniti, Ladd & Mennen, 1998; Ladd et al., 1999; Xu, 1998, 1999, 2001; Xu & Xu, 2005)，以基频转折点为时间参照，以鼻辅音为频谱参照，考察滑音跟鼻辅音相对应的时间段应该是什么。结果发现，滑音里相当于鼻辅音起始点的位置大大早于滑音里共振峰最极端之处。以图10为例，10b里第二共振峰最低之处往往被认为是 [w] 的起点。但根据基频转折点 (竖实线处) 与10a

里鼻辅音的对应，[w] 的起点应为竖虚线处。但这一点正处于第二共振峰从前边的 [aɪ] 向后边的 [w] 的过渡之中。根据目标接近模型，向一个音的理想目标接近的过程，就是这个音的发音过程。所以 [w] 应该是从第二共振峰的高转折点开始，到低转折点结束。如果继续这样推论，[n] 也应该从第二共振峰的高转折点，而不是从鼻音共振峰出现处开始。更进一步来看，早已有证据表明，朝向下一音节里起首元音的发音动作，是从上一音节的最后音渡时开始的 (Öhman, 1966)。由此，第二音节里的元音 [a] 也应该从那里开始。也就是说，后音渡并不是来自于逆向协同发音 (anticipatory coarticulation)，而是本身就是下一音节里元、辅音的共同起点。

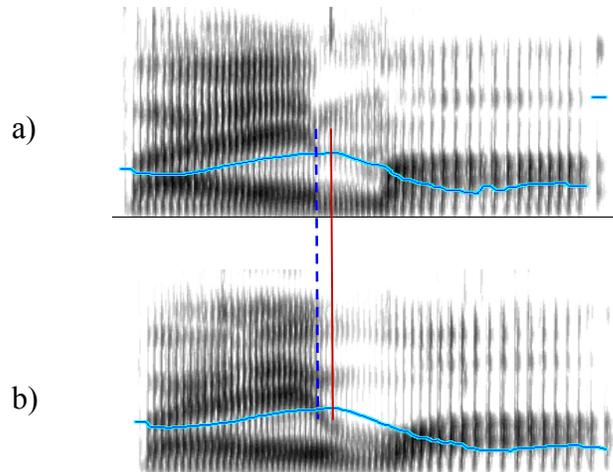


图10：以基频转折点（竖实线）为时间参照决定滑音的起点。a) 白马，b) 百瓦。数据来自 Xu & Liu (2006)。

基于这个新的思路，Xu & Liu (2006) 提出了音节时间结构模型 (time structure model of the syllable)，如图11所示。根据这个模型，在音节起始处辅音、元音、声调以及发声型同时开始向各自的目标接近。声调和发声型用整个音节完成目标接近动作，辅音在音节前半就完成发音，元音在没有韵尾辅音时用全部音节，在有韵尾辅音时则提前结束。按照该模型里的定义，一个音的时段，就是发音动作朝向其目标渐进的时段。一旦发音动作开始转向下一个目标，当前音便已结束，后一音便已开始。按此定义，不存在任何顺向协同发音，也不存在严格意义上的逆向协同发音，只有起首辅音跟元音的前半部是同时协同发音。这样一来，协同发音的范围就大大减小。传统上，协同发音是用音与音之间的相互影响来定义的 (Kühnert & Nolan, 1999)。因为需要解释的正是音与音之间的相互影响，所以传统的定义是循环性的，没有解释力。因此，对音节机制的深入理解，导致了对协同发音这个悬而未决的问题的新见解。

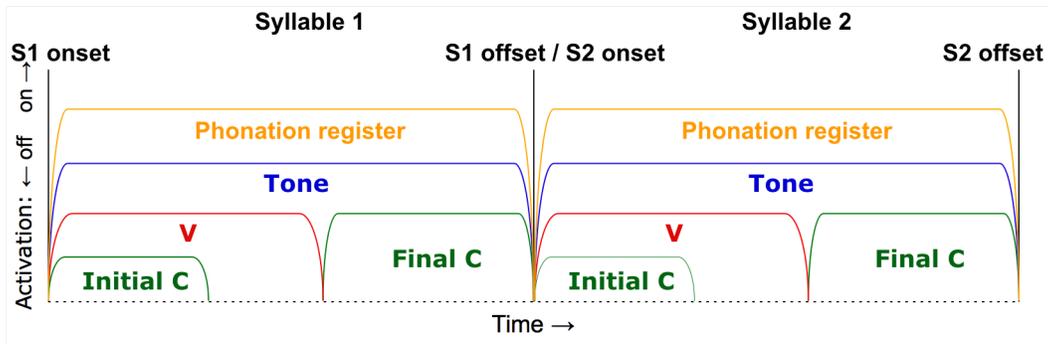


图11: 音节时间结构模型 (Xu & Liu, 2006)

音节时间结构模型不仅给了音节新的定义，而且解释了为什么 CV 结构比 VC 结构更为普遍。这是因为元、辅音在音节前部是同时发音，而在音节后部是顺序发音的，即收尾辅音必须在元音动作结束之后才能开始 (见图11)。由于时间压力的缘故，CV 自然要比 VC 容易实现。更有意义的是，由于音节时间结构模型解决了协同发音这个难题，音段和韵律可以统一于同样的发音机制。

(二) 弱化和缩读的机制

我们在前面讨论分界和组合的韵律标记时谈到 (见 二/ (二))，在词和词组层次上的基频模式变化可以由时长变化跟发音速度极限的交互作用来解释。该原理也可以用来解释音段里至今难解的弱化 (reduction) 和缩读 (contraction) 现象 (Johnson, 2004; Kohler, 1990; Myers & Li, 2009; Tseng et al., 2005)。Cheng & Xu (2009) 考察了台湾国语里的缩读现象。在台湾国语里，多音节词常常被缩读为单音节。如把“知道”说成“赵”，把“这样”说成“降”，把“不知到”说成“抱”。跟以往直接考察自然语言的方法不同，该实验让受试在实验室里念事先编好的句子，其中对各种可能的因素都严加控制，而且关键词都是无意义的。结果发现，尽管是在实验室条件下念含无意义词的句子，只要语速较快，缩读现象就频繁出现。进一步的声学分析表明，只要音段的缩短超过一定域限 (双音节 < 200 ms)，缩读就会发生。由此看来，词频、力度、语境等这些通常被认为的重要因素很可能都是间接因素。也就是说，它们导致的缩读，是通过时间压力这个直接因素起作用的。当然这只是初步结果，还有待于今后更大量实验的进一步验证。

四、韵律研究跟其他语言学领域的关系

韵律跟语言学许多其他领域都有密切的关系。上面介绍的有关韵律的新发现引出了一些新的可能性。由于篇幅有限，下面只从三个方面简略讨论。

（一）焦点的韵律编码跟句法结构的关系

英语、汉语以及其他用 PFC 编码焦点的语言里有一个重要特点，就是句末焦点对调域的扩展远远小于非句末焦点 (Eady & Cooper, 1986; Xu, 1999; Xu & Xu, 2005)，而且句末焦点不可能有 PFC。结果，句末焦点的识别率大大低于非句末焦点 (Botinis, Fourakis & Gawronska, 1999; Chen et al., 2009; Liu & Xu, 2005)。这一现象有可能是一些句法结构的历史来源。例如英语、汉语里的强调句型都是把被强调成分挪到靠近句首的位置，而那样正好可以充分实现焦点后的调域压缩，从而保证焦点的识别。同样，在汉语里特殊疑问词一般是焦点所在。而英语里的特殊疑问句里则把疑问词挪到句首，也就是实现焦点的有效位置。这种韵律影响句法的可能性，跟冯胜利提倡的韵律句法学所探索的机制有相似之处 (冯, 2000)。

（二）韵律为什么要跟句法对应？

在二/（二）节“分界 / 组合”里我们谈到了句法里的递归结构在韵律上有明确的表现。虽然这是很重要的发现，但是它还引出一个很少有人问的问题，即，为什么韵律跟句法要对应？从形式优先的角度看，这种对应是不言而喻的 (Beckman, 1996; Selkirk, 1984)。从功能优先的角度看，这种对应就有些费解了。在多数情况下句法结构通过词、句的构成已经很清楚，为什么在韵律还要有表现呢？一种可能的解释是，语言转达信息的基本策略是高冗余度的重复编码。也就是说，同一信息用多种方式编码，以保证信息的准确传递。这种重复编码可以是在领域内的，对此语音学研究已有大量证据 (Lisker, 1986; Stevens, 2003)，也可以是跨领域的，但对此现在少有讨论，因为几乎没人从这个角度考虑问题。随着功能优先的重要性得到越来越多的认识，对跨领域重复编码的研究也应该大力展开。这一新的研究方向很可能会大大深化我们对语言的理解。

（三）韵律焦点的跨语言差别与其他语言学领域的关系

在二/（二）节“焦点”里我们谈到了焦点韵律编码跨语言的差别。这个新发现，不仅有语音学上的意义，而且对句法学、语用学、社会语言学、历史语言学、语言类型学都可能会有影响。在句法学上，如果前面刚讲到的焦点韵律导致句法结构变化是事实，那么一个语言是否有焦点后压缩 (PFC) 也可能影响它的句法结构。对语用学和语义学来说，焦点本来就是一个重要概念 (Pierrehumbert & Hirschberg, 1990; Rooth, 1992)。但是到目前为止的理论探讨大都是建立在概念推论之上的，极少有引用系统实验的发现，或直接通过实验验证其理论。所以这方面将来大有文章可做。对历史语言学来说，PFC 的跨语言分布很可能为我们提供研究语言演化和语言分类的一个新的工具 (Xu, Chen & Wang, forthcoming)。

五、 结语

语言是用来交流信息的，语言里的信息是通过语言设施，即神经系统和发音器官来编码的，所以弄清交际功能如何通过语言设施编码应该是理解语言的关键。以往的语言学理论很少从这个角度考虑问题，而是多致力于描述、解释直接观察到的种种语言现象，包括直观的语法判断 (grammaticality judgment)。很少有人指出，具体的语言现象很可能是功能编码过程的衍生物，所以只有理解了功能编码的机制，才可能真正理解这些语言现象。本章讨论的韵律研究正是从这个角度入手，先以声调的分词功能为起点，发现其借助发音器官编码的基本机制。进一步的研究发现，声调编码的机制可以延伸到语调和音段。另一方面，从功能角度研究韵律，发现多重交际功能可以各自用独特的编码方式同时编码，而且它们相互独立，之间并无管辖关系。有关韵律功能编码的发现还可能有助于理解其他语言学领域里的许多问题，如焦点的本质、韵律跟句法的对应、以及跨语言的韵律差别跟历史语言学和语言类型学的关系。

韵律研究的方法对其他语言学领域也可以有所启发。首先，从功能编码入手的策略也应适用于其他领域。当然在其他层次上编码的主要设施不是发音器官，而是中央神经系统，即大脑。许多功能的编码很可能受到大脑的各种特性的制约，例如短时记忆的容量 (见陆丙甫的初步探讨：陆，1983)。从探索神经系统的编码机制入手，有可能作出意想不到的发现。第二，本章讨论的韵律方面的理论发展都是基于严格控制的实验研究，并且不断受到新实验结果的检验。近期进一步发展到由理论引出量化的预测式模型，可以产生能用自然语言直接检验的表层形式 (Prom-on et al., 2009; Xu & Prom-on, 2010)。这一套实证性的研究方法也应该适用于其他语言学领域。可以想像，在不久的将来，句法学、语义学、语用学也会象语音学一样逐步变成全面的实证科学。

参考文献

- Arvaniti, A., Ladd, D. R. and Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* **36**: 3-25.
- Austin, J. L. (1962). *How to Do Things with Words: The William James Lectures delivered in Harvard University in 1955*. Oxford: Clarendon.
- 白澍洲 (1934) : 关中声调实验录。《史语所集刊》4: 355-361
- Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes* **11**: 17-67.
- Beckman, M. E. and Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook* **3**: 255-309.
- Birkholz, P., Jackèl, D. and Kröger, B. J. (2006). Construction and control of a three-dimensional vocal tract model. In *Proceedings of The 31st International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France: I-873-876.

- Bolinger, D. (1986). *Intonation and its parts: melody in spoken English*. Palo Alto: Stanford University Press.
- Bolinger, D. (1989). *Intonation and Its Uses -- Melody in Grammar and Discourse*. Stanford, California: Stanford University Press.
- Botinis, A., Fourakis, M. and Gawronska, B. (1999). Focus identification in English, Greek and Swedish. In *Proceedings of The 14th International Congress of Phonetic Sciences*, San Francisco: 1557-1560.
- Bresnan, J. (2001). *Lexical-Functional Syntax*. Oxford: Blackwell.
- Büring, D. (2007). Semantics, Intonation and Information Structure. In *The Oxford Handbook of Linguistic Interfaces*. G. Ramchand and C. Reiss: Oxford University Press pp.
- 赵元任 (1922) : 中国言语字调底实验研究法。《科学》 7: 871-882.
- Chao, Y.-R. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica*. 4.4: 363-397. Reprinted in *Readings in Linguistics*, ed. Martin Joos, pp. 38-54. Washington D.C.: American Council of Learned Societies.
- Chen, S.-w., Wang, B. and Xu, Y. (2009). Closely related languages, different ways of realizing focus. In *Proceedings of Interspeech 2009*, Brighton, UK: 1007-1010.
- Chen, Y. 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. (2009). Extreme reductions: Contraction of disyllables into monosyllables in Taiwan Mandarin. In *Proceedings of Interspeech 2009*, Brighton, UK: 456-459.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- 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.
- Cohen, A., Collier, R. and 't Hart, J. (1982). Declination: Construct or intrinsic feature of speech pitch. *Phonetica* 39: 254-273.
- Dik, S. C. (1997). *The theory of functional grammar, 2nd Edition*. Berlin: Mouton de Gruyter.
- 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.
- 冯胜利 (2000): 《韵律语法学》。上海教育出版社。
- Firth, J. R. (1957). *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Foley, W. A. and Robert D. Van Valin, J. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech* 1: 126-152.
- Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics* 11: 149-175.
- Gandour, J., Potisuk, S. and Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics* 22: 477-492.
- Gauthier, B., Shi, R. and Xu, Y. (2007a). Learning phonetic categories by tracking movements. *Cognition* 103: 80-106.
- Gauthier, B., Shi, R. and Xu, Y. (2007b). Simulating the acquisition of lexical tones from continuous dynamic input. *JASA Express Letters* 121: EL190-195.
- Goldsmith, J. A. (1976). *Autosegmental Phonology*, Ph. D. dissertation, MIT. [Published in 1979 by Garland Press in New York].
- Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and Phonology. In *Proceedings of The 1st International Conference on Speech Prosody*, Aix-en-Provence, France: 47-57.

- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*: Cambridge University Press.
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English, Part II. *Journal of Linguistics* **3**: 199-244.
- Halliday, M. A. K. (1985). *Introduction to functional grammar*. London: Edward Arnold.
- Johnson, K. (2004). Massive reduction in conversational American English. In *Proceedings of 10th international symposium: Spontaneous speech: Data and analysis*, Tokyo, Japan: 29-54.
- Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics* **3**: 129-140.
- Kohler, K. J. (1990). Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In *Speech production and speech modelling*. Hardcastle and Marchal. Dordrecht: Kluwer pp. 69-92.
- Kohler, K. (2005). Timing and Communicative Functions of Pitch Contours. *Phonetica* **62**: 88-105.
- Kühnert, B. and Nolan, F. (1999). The origin of coarticulation. In *Coarticulation: Theory, Data and Techniques*. W. J. Hardcastle and N. Newlett. Cambridge: Cambridge University Press pp. 7-30.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Ladd, D. R. (2008). *Intonational phonology*. Second Edition. Cambridge: Cambridge University Press.
- Ladd, D. R., Faulkner, D., Faulkner, H. and Schepman, A. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* **106**: 1543-1554.
- Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America* **54**: 1228-1234.
- Lehiste, I. (1975). The phonetic structure of paragraphs. In *Structure and process in speech perception*. A. Cohen and S. E. G. Nooteboom. New York: Springer-Verlag pp. 195-206.
- Lieberman, M. and Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In *Language Sound Structure*. M. Aronoff and R. Oehrle. Cambridge, Massachusetts: M.I.T. Press pp. 157-233.
- Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* **29**: 3-11.
- Liu, F. and Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* **62**: 70-87.
- Liu, F. and Xu, Y. (2007). Question intonation as affected by word stress and focus in English. In *Proceedings of The 16th International Congress of Phonetic Sciences*, Saarbrücken: 1189-1192.
- 陆丙甫 (1983): 人脑短时记忆机制和人类语言结构的关系。《世界科学》1983.9: 21-30。
- Morton, E. W. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist* **111**: 855-869.
- Mozziconacci, S. (2002). Prosody and Emotions. In *Proceedings of The 1st International Conference on Speech Prosody*, Aix-en-Provence, France: 1-9.
- Myers, J. and Li, Y. (2009). Lexical frequency effects in Taiwan Southern Min syllable contraction. *Journal of Phonetics* **37**(2): 212-230.
- Nakajima, S. and Allen, J. F. (1993). A study on prosody and discourse structure in cooperative dialogues. *Phonetica* **50**: 197-210.

- Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* **41**: 1-16.
- Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* **39**: 151-168.
- Oliveira, M. and Freitas, T. (2008). Intonation as a cue to turn management in telephone and face-to-face interactions. In *Proceedings of Speech Prosody 2008*, Campiñas: 485-488.
- Oller, D. K. (1973). The effects of position in utterance on speech segment duration. *Journal of the Acoustical Society of America* **54**: 1235-1246.
- O'Shaughnessy, D. and Allen, J. (1983). Linguistic modality effects on fundamental frequency in speech. *Journal of the Acoustical Society of America* **74**: 1155-1171.
- 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].
- Pierrehumbert, J. and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In *Intentions in Communication*. P. R. Cohen, J. Morgan and M. E. Pollack. Cambridge, Massachusetts: MIT Press pp. 271-311.
- 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.
- Rialland, A. (2009). African "lax" question prosody: its realisations and its geographical distribution. *Lingua* **119**: 928-949.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics* **1**(1): 75-116.
- Selkirk, E. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass: MIT Press.
- Shih, C. (2000). A declination model of Mandarin Chinese. In *Intonation: Analysis, Modelling and Technology*. A. Botinis: Kluwer Academic Publishers pp. 243-268.
- Sluijter, A. M. C. and Terken, J. M. B. (1993). Beyond sentence prosody: Paragraph intonation in Dutch. *Phonetica* **50**: 180-188.
- Stevens, K. N. (2003). Acoustic and Perceptual Evidence for Universal Phonological Features. In *Proceedings of The 15th International Congress of Phonetic Sciences*, Barcelona: 33-38.
- Terken, J. M. B. (1984). The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech* **27**: 53-73.
- Thorsen, N. G. (1980). A study of the perception of sentence intonation — Evidence from Danish. *Journal of the Acoustical Society of America* **67**: 1014-1030.
- Tseng, C. Y., Pin, S. H., Lee, Y. H., Wang, H.-m. and Chen, Y.-c. (2005). Fluent speech prosody: Framework and modeling. *Speech Communication* **46**: 284-309.
- Tseng, S. C. (2005). Contracted syllables in Mandarin: evidence from spontaneous conversations. *Language and Linguistics* **6**: 153-180.
- Turk, A. E. and Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics* **28**: 397-440.
- Umeda, N. (1982). "F0 declination" is situation dependent. *Journal of Phonetics* **10**: 279-290.
- Wagner, M. (2005). *Prosody and Recursion*. Ph.D. Dissertation, Massachusetts Institute of Technology.
- Wang, B. and Xu, Y. (2006). Prosodic encoding of topic and focus in Mandarin. In *Proceedings of Speech Prosody 2006*, Dresden, Germany: PS3-12_0172.

- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* **91**(3): 1707-1717.
- 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. (2004a). Separation of functional components of tone and intonation from observed F0 patterns. In *From Traditional Phonology to Modern Speech Processing: Festschrift for Professor Wu Zongji's 95th Birthday*. G. Fant, H. Fujisaki, J. Cao and Y. Xu. Beijing: Foreign Language Teaching and Research Press pp. 483-505.
- Xu, Y. (2004b). Transmitting Tone and Intonation Simultaneously — The Parallel Encoding and Target Approximation (PENTA) Model. In *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, Beijing: 215-220.
- Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication* **46**: 220-251.
- Xu, Y. (2008). Multi-dimensional information coding in speech. In *Proceedings of Speech Prosody 2008*, Campinas, Brazil: 17-26.
- Xu, Y. (2009). Timing and coordination in tone and intonation--An articulatory-functional perspective. *Lingua* **119**(6): 906-927.
- Xu, Y., Chen, S.-w. and Wang, B. (forthcoming). Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family?
- Xu, Y. and Kelly, A. (2010). Perception of anger and happiness from resynthesized speech with size-related manipulations. In *Proceedings of Speech Prosody 2010*, Chicago
- Xu, Y., Kelly, A. and Smillie, C. (forthcoming). Emotional expressions as communicative signals. In *Prosody and Iconicity*. S. Hancil and D. Hirst pp.
- 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. (2007). Determining the temporal interval of segments with the help of F0 contours. *Journal of Phonetics* **35**: 398-420.
- Xu, Y. and Prom-on, S. (2010) PENTATrainer.praat, downloadable at <http://www.phon.ucl.ac.uk/home/yi/PENTATrainer/>
- 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 Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* **33**: 159-197.
- Zerbian, S., Genzel, S. and Kügler, F. (2010). Experimental work on prosodically-marked information structure in selected African languages (Afroasiatic and Niger-Congo). In *Proceedings of Speech Prosody 2010*, Chicago: 100976:1-4.

作者简介

许毅1993年获美国康涅狄格大学语言学博士。毕业后先后在麻省理工学院从事博士后研究，在美国西北大学任教，以及在芝加哥大学和哈斯金斯实验室从事研究工作。从2004年起他在伦敦大学学院任教，现任言语、听觉及语音学系的副教授。他早期的研究集中探讨北京话连续语流中声调的产生与感知，在此研究基础上提出了目标接近模型。他后来将此模型扩展至语调和音段，分别提出了PENTA和音节时间结构模型。近年来他又着手研究情感的语音表达以及焦点的跨语言分布。

刘芳在北京大学中国语言文学系获得汉语语言学学士和语言学及应用语言学硕士学位。之后她又在 University of Chicago 获得统计硕士和语言学博士学位。她的博士论文探讨了英汉语的韵律和语调。2008-09年，她在Goldsmiths, University of London心理学系担任助理研究员，研究失乐症对语言能力的影响。2009-10年，她获得英国 Economic and Social Research Council 的博士后基金，先后在Goldsmiths, University of London和 University College London进行语言和音乐关系的研究和教学工作。她将于2011年赴 Stanford University 与Professor Patrick Suppes合作研究英汉语的音位和区别特征的脑波表现。