

汉语韵律层级结构边界的声学分析*

王 蓓 杨玉芳 吕士楠¹

(中国科学院心理研究所 北京 100101)

(1 中国科学院声学研究所 北京 100080)

2002 年 2 月 8 日收到

2002 年 4 月 17 日定稿

摘要 基于大规模语料库, 对比了较慢和较快两种语速的语料, 研究了韵律层级结构边界的声学表现。研究主要得到以下结果: (1) 汉语语句音高的下倾和重置是由音域下限的移动实现的。(2) 韵律词边界的声学线索是低音线的不连续性和边界前音节的延长, 一般没有无声段。(3) 韵律短语和语调短语边界的声学线索是低音线重置和无声段。并且, 边界等级越高, 低音线重置程度越大, 无声段的长度也越长。(4) 知觉等级与无声段长度成对数增长关系。
PACS 数: 43.70

Acoustic analysis on prosodic hierarchical boundaries of Chinese

WANG Bei YANG Yufang LU Shinan¹

(The Institute of Psychology, The Chinese Academy of Sciences Beijing 100101)

(1 Institute of Acoustics, The Chinese Academy of Sciences Beijing 100080)

Received Feb. 8, 2002

Revised Apr. 17, 2002

Abstract Based on large corpus with both low and fast speed, we investigated acoustic correlates to prosodic hierarchical boundaries in Chinese. The main results are: (1) The declination of intonation in Chinese is realized by the bottom line which is composed of low points of the pitch. (2) The acoustics correlates of prosodic word boundary are pre-boundary lengthening and discontinuation of the bottom line of intonation. (3) The acoustics correlates of prosodic phrase and intonational phrase are pitch reset of bottom line of intonation and insertion of silence. Moreover, the higher the prosodic boundary is, the higher the extent of pitch reset is and the longer the silence is. There is no significant difference on pre-boundary lengthening between syllables on these two prosodic boundaries. (4) There is logarithm correlates between perceived boundary strength and length of silence.

引言

随着波形拼接和 PSOLA 算法的不断改进以及大规模语料库录制的不断成熟, 汉语文语转换系统已经有了很高的可懂度, 但自然度不令人满意。提高自然度一个最重要的问题就是韵律的控制和调节, 其中韵律层级边界的设置是一个关键。

研究者比较一致地认为句子内部的韵律结构包括音节、音步、韵律词、重音短语、中间短语和语调短语^[1-5]。Booij 指出韵律词可以对应几个词典词, 也可以小于一个词典词^[6]。汉语中的韵律词是反映

汉语节奏的两音节或三音节组^[7]。中间短语将词组成短语, 并且每个短语中至少有一个重读音节^[3]。由于重音短语和中间短语很难界定, 本研究中都归为韵律短语。语调短语由若干个中间短语构成, 在语调短语末用高或低的边界声调标志^[8,9]。

研究韵律边界等级及其声学线索时, 一个最大的障碍是如何判定韵律边界的位置及其强度。韵律首先是一个感知概念, 因此依赖人的感知得到的韵律结构才是最可靠和最客观的。心理实验中的等级评定法较好地解决了这个问题^[4,10]。De Pijper 等的研究证明了没有经过训练的听者可以可靠地判断

* 国家自然科学基金 (39970254) 和 Motorola 公司资助项目

出知觉到的边界强度, 并且知觉判断的边界强度和音系上假设的韵律结构有很高的一致性^[4]。这种方法的好处是: 可以直接构建韵律层级结构; 避免请专家判断时, 他们的语音学和语言学知识的介入; 避免了循环论证的危机。

现有的研究一致认为, 韵律边界层级结构的声学表现是, 边界前音节的延长、音高重置和加入无声段等。Streeter的研究发现在标志主要短语的边界时, 语调和时长起主要作用, 而音强的作用很弱^[11]。关于声学线索对句法结构的预测, Streeter提出了三个概率模型^[11]: 第一个模型假设音高和时长线索由独立的通道或加工器加工, 每个加工器都输出一个正确的句法分析的概率值。然后对这两个值再做一次概率判断, 给出最终的句法分析的结果。第二个模型是句法决策可能用音高, 或用时长信息, 两者取其一。听者需要在音高和时长之间做出选择。第三个模型是音高和时长两个加工器各输出一个值, 表示每种句法分析的可能性, 然后再将这些独立估计的值平均。Streeter的研究结果支持后两个模型。

韵律边界处音高的变化有两种表现方式: 一是音高曲线的不连续性, 二是音高下倾曲线的重置^[4]。Swerts对标志句子的语调线索进行了研究, 认为除音高重置外, 还有语调曲线的差异和音域^[10,12]。由于汉语是声调语言, 对汉语语调曲线的描述必须考虑音节固有声调的音高曲线。赵元任和沈炯用音域的概念较好地解决了这个问题^[13-15]。声调音域是指同一条件下各种声调占用的音高范围, 或者说是同一特定条件下各种声调聚合在一起的音高分布范围, 声调音域可以用低音点和高音点两个变量描述^[15]。音域下限在语流中的延伸反映节奏单元结构, 而音域上限跟语义的加强有关^[14,15]。

对于区分不同等级边界, 时长有三方面的作用: 音步延长、边界前音节的延长和停顿。Wightman指出, 边界前音节延长指最后一个音节, 而不是音步或最后一个重读音节^[5]。通常认为无声段越长, 边界的知觉等级也越高, Swerts对荷兰语的研究中两者的相关系数为0.63^[16]。关于无声段与音节延长, 有研究指出边界前音节延长程度不是很大时, 不出现无声段^[4]。杨玉芳研究了不同等级的句法边界的韵律表现, 指出边界前音节的时长和停顿之和随边界等级提高几乎是线性增长^[17]。还有研究指出无声段长度与前音节的时长是相互补偿的, 而无声段长度与后音节音高上、下线的重置程度是相互增益的^[18]。

上述对韵律层级边界的声学表现的研究中, 多是用少量实验语句。而目前已有的自然语言和新闻语言

语料库从结构到内容都大大超出了实验语料, 因此大规模语料库的研究越来越受到研究者的重视。另外, 对韵律边界处各种声学线索的相对重要性及每一线索如何单独作用还没有一致的看法^[5]。特别是不同韵律层级边界的声学表现没有系统的研究。虽然英语研究中得到的结果有重要的参考价值, 但是汉语韵律研究必须要考虑汉语的特点。考虑到言语工程技术的实际应用, 韵律词、韵律短语和语调短语是本文考察的基本韵律层级结构。文章首先讨论知觉的韵律层级结构与音系上的韵律结构之间的关系。然后集中考察边界前音节的延长、音高和无声段三个声学特征在不同等级韵律边界发生的变化。最后报告语速对韵律层级结构表现的影响。

1 韵律边界知觉等级标定实验

语句的韵律边界标定通过知觉实验, 采用等级评定的方法。

1.1 材料

本研究的语料库第一部分为500个单句(语料A), 共7988个音节。发音人为一名女性专业播音员, 年龄24岁。以自然的方式朗读句子, 没有特别的强调和感情色彩, 语速较慢。句子的平均长度为16个音节。

语料库第二部分为10个朗读语篇, 发音人为另一名女性专业播音员, 年龄24岁。以自然的方式, 流畅地朗读语篇, 语速较快。由于语篇中各自然段段首的句子与单句的朗读没有太大差别, 可以与前面500个单句对比, 选择段首的句子为第二组研究材料(语料B), 共有37句, 827个音节。句子平均长度为22个音节。

语料A与语料B的音节平均时长分别为286ms和189ms, 无声段所占的比例分别为14%和9%。

1.2 被试

分别有24名和19名北京籍大学生参加了语料A和语料B的韵律边界标定实验。被试年龄20-23岁, 听力正常。

1.3 实验任务

韵律边界等级对应的知觉概念是语流间断。知觉实验要求被试分三级判断每个音节后的间断等级, 即明显、较明显和不明显的间断, 其余为无间断。从预实验中发现, 被试对间断等级做三级判断比较容易, 结果可靠。另外, 韵律层级结构主要包括韵律词、韵律短语和语调短语。结果发现三级知觉判断与3个韵律成分有很好的对应关系。但实验时不告诉被试

分三级判断的意义, 被试也不具备相应的语音学和音系学知识, 他们仅凭知觉做判断。这样, 就可以实现由知觉结果确定韵律结构。

1.4 实验程序

实验中, 每句话播放 2 遍, 间隔 800 ms。要求被试对间断的标注完全基于听到的声音。语料 A 分为 5 组, 完成每组实验需要 1 h 左右。所有 24 名被试完成全部 5 组实验, 每两组实验之间休息时间不少于 2 h。语料 B 自成一组, 另外 19 名被试完成语料 B 的全部标注。实验前被试先进行一个练习, 以形成统一的判断标准。

1.5 声学参数测量

声学参数测量包括每个音节音高曲线的高音点、低音点、平均音高、时长和音节后无声段长度。声学测量用 Multispeech 语音分析软件完成。

其中, 音节的高音点、低音点和平均音高分别指这个音节音高曲线的最大值、最小值和平均值。阴平声只取平均音高, 阳平、上声和去声取其高点、低音点和平均音高。

2 结果与讨论

2.1 韵律层级结构

语料 A 的知觉标定实验中, 将被试的标注结果两两作相关分析, 发现相关系数均在 0.8 以上, 并且在 0.001 水平显著相关。语料 B 的知觉标定实验, 被试间的一致性也在 0.8 以上。这说明听者倾向于用同样的方式界定韵律边界。数据统计时语流的明显间断点计为 3 分; 较明显为 2 分; 不明显为 1 分, 然后取所有被试的平均值。如果某个音节后的边界等级只有一名被试标为 1, 则认为这个值是由被试的偶然误差造成的, 这时边界值小于 0.06。由此, 知觉等级界定为: 平均值小于 0.06 时知觉等级为 0; 0.06~1 为 1 级; 1~2 的为 2 级, 2~3 的 3 级。

语料 A 除去句末的音节共 7488 个, 其中只有 27.8% 的音节后感知到有间断。知觉为无边界的音节中, 大部分为词内的音节, 另外一部分为词边界音节。

句内主要的韵律成分有韵律词、韵律短语和语调短语。我们将知觉等级为 1 对应韵律词边界; 2 对应韵律短语边界; 3 对应语调短语边界。

知觉等级与语料库文本语法分析结果见表 1, 其中列出各语法成分对应的知觉等级判断的次数。语法成份包括词、短语和小句。小句指小于句子, 但大于短语、词或语素的语法结构单位^[20]。

表 1 语法成分对应的知觉等级次数

知觉等级	语法词	语法短语	小句
1	685	—	—
2	239	415	—
3	113	103	344

表 1 中容易发现对角线上的数字显著地大于其它位置的。这说明, 三级知觉等级的划分与语法单元基本上是对应的。知觉等级越高, 所对应的句法成分也越高。这也验证了 De Pijper 等人的研究中指出知觉判断的边界强度和音系上假设的韵律结构有很高的一致性的结论^[4]。语法词后有边界的情况中, 知觉等级为 1 的有 77.4%, 但还有 22.6% 知觉等级为 2 或 3。同样, 语法短语后知觉等级为 3 的情况也有 19.9%。这说明语法单位与韵律成分有紧密的联系, 但不是一一对应。某些情况下, 一个语调短语可能是一个语法词或语法短语。

图 1 是从语料库中选出的一个句子的示例。从中可以看出语法上的小句“该行业潜在着巨大的回报价值”和“被许多有识之士看好”分别构成整个句子的两个语调短语。语调短语又由若干韵律词和韵律短语组成, 韵律短语中包含着韵律词。从这个意义上讲, 韵律结构是层级的。并且, 韵律成份与句法成份基本上是对应的。因此, 韵律结构从低到高分别是韵律词、韵律短语和语调短语, 它们所对应的知觉等级分别为 1、2 和 3。

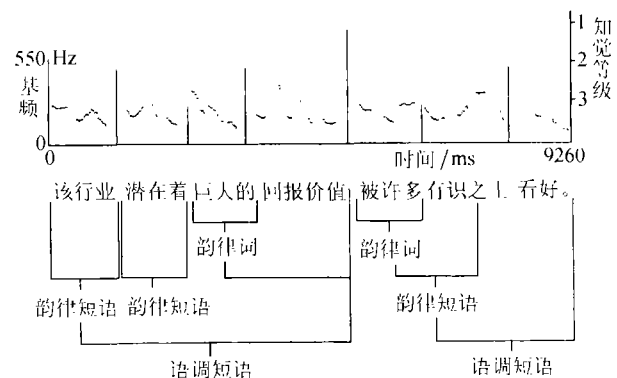


图 1 句法结构、韵律结构与知觉等级示意图

注: 图中虚线为基频曲线, 竖的线段对应知觉等级

2.2 语速较慢时韵律边界的声学分析

2.2.1 韵律层级边界的时长分析

韵律边界处边界前音节出现延长现象, 但边界后音节没有系统的变化规律^[17]。因此, 本研究只分析韵律边界前音节的时长及其延长量, 结果见表 2。其中, 时长的延长量 = (音节的时长 - 该音节的平均时长) ÷ 该音节的平均时长。

表 2 不同韵律层级边界前音节时长及时长延长量

韵律边界	平均值 /ms	时长延长量 /%
无边界	273	-7
韵律词	325	11
韵律短语	392	29
语调短语	388	21

从表 2 发现, 随着韵律等级的提高, 时长显著加长 ($F(3, 2160) = 375.12^{**2}$)。韵律词边界和无边界相比, 时长 ($F(1, 1080) = 143.08^{**}$) 和时长延长量 ($F(1, 1080) = 133.24^{**}$) 都显著加大。韵律短语边界和韵律词边界相比, 时长 ($F(1, 1080) = 258.71^{**}$) 和时长延长量 ($F(1, 1080) = 105.21^{**}$) 上都有显著差异。韵律短语和语调短语边界处音节的时长差异不显著 ($F(1, 1080) = 0.813, P = 0.368$)。虽然两者在延长量上有显著差异 ($F(1, 1080) = 25.45^{**}$), 但是在韵律短语边界处边界前音节的延长反而大。这个结果和杨玉芳^[17]关于短语边界处音节时长达到最大的结果是一致的。也就是说, 边界前音节的延长是有限的, 到韵律短语边界处达到最大, 它只是较小边界的声学线索, 不具有区别更高边界的功能。

不同韵律成分对应不同句法成分时的时长延长量见表 3。

表 3 对应不同句法成分时韵律边界前音节时长延长量 /%

韵律边界	语法词	语法短语	小句
无边界	0	—	—
韵律词	11	--	—
韵律短语	31	26	—
语调短语	32	21	16

从表 3 中可以看出, 对于语法词来说, 韵律等级越高音节的时长延长量越大。但语法词对应韵律短语与对应语调短语相比, 边界前音节的延长量上没有显著差异 ($F(1, 337) = 0.31, P = 0.578$)。另外, 同样的韵律边界, 语法成分越低边界前音节的延长反而越大。这两点都证明了边界前音节的延长不是区分较高等级韵律边界的充分条件。句内较高的韵律边界处边界前音节的值更小是因为有其它区别性线索, 如音高重置或无声段; 而较低的韵律边界处边界前音节的延长是区别性线索, 在缺少其它线索时边界前音节的延长起标志韵律边界的作用, 因此延长程度更大。

2.2.2 韵律层级边界的音高分析

从音域的观点看, 低音点与节奏群有关, 而高音点与重音有关。与低音点的变化相比, 不同重读程度

² 注: * 表示 $P < 0.05$, ** 表示 $P < 0.001$

音节的高音点的变化要自由的多^[20], 从而也影响到平均音高和音域的变化。因此对韵律层级边界的音高分析主要考察低音点的变化。由于阴平声没有低音点, 在此不参与分析。另外轻声和处于句末的音节也不参与分析。表 4 中分别对阳平声、上声和去声在不同韵律层级边界做了分析。

表 4 不同韵律层级边界前 3 种声调音节的低音点 /Hz

韵律边界	阳平	上声	去声	平均值
无边界	163	146	180	167
韵律词	165	139	157	155
韵律短语	151	111	125	130
语调短语	140	106	119	122

表 4 中, 从 3 个声调的平均值看, 随着韵律边界等级的增加, 低音点呈逐渐下降的趋势, 并且不同等级间低音点的差异是显著的 ($F(3, 175) = 156.66^{**}$)。阳平声 ($F(3, 511) = 20.92^{**}$)、上声 ($F(3, 407) = 39.6^{**}$) 和去声 ($F(3, 830) = 143.92^{**}$) 都表现这种趋势, 上声和去声的这种变化尤其明显。用聚合声调的方法估计音域下限时, 上声和去声的低音点基本上达到了音域下限。因此, 低音点与韵律边界等级之间的关系是从音域的概念理解的。

不同等级韵律边界处音高重置程度可以通过比较不同等级边界前、后音节的低音点, 其结果见表 5。

表 5 不同韵律边界前、后音节的低音点及其差值 /Hz

韵律边界	边界前	边界后	后 - 前
无边界	167	156	-11
韵律词	155	167	12
韵律短语	130	179	49
语调短语	122	187	65

从表 5 中可以发现, 无边界时, 后一音节的低音点低于前一音节的, 这反映了低音线的下倾趋势。另外, 韵律等级越高, 边界后音节与前音节的差值越大。在韵律词边界处, 边界后音节略高于边界前音节, 仅相差 12 Hz。这时出现低音线的不连续性。韵律词边界通常嵌在韵律短语或语调短语中, 在更高级的韵律单元内音高下线保持着连续下倾的趋势。如果韵律短语或语调短语中有可察觉到的韵律词边界, 这条下倾曲线就可能出现局部的小变化, 但整体的下倾趋势没有改变。这时出现的音高曲线的不连续性不同于音高重置。在韵律短语和语调短语边界处, 边界后音节则显著地高于边界前音节, 出现低音

线重置。另外, 语调短语边界处的低音线重置程度显著地大于韵律短语边界处的 ($F(1, 895) = 4.78^*$)。

从高音点来看, 阴平、阳平和去声在不同知觉等级边界前后高音点的平均值见表 6。

表 6 韵律边界前、后音节的高音点及其差值 /Hz

韵律边界	边界前	边界后	后-前
无边界	262	250	-18
韵律词	244	266	22
韵律短语	232	270	38
语调短语	232	269	37

表 6 中, 3 种韵律边界后高音点的差异不显著 ($F(2, 1481) = 0.63, P = 0.53$), 韵律短语边界和语调短语边界前音节的高音点也没有显著差异 ($F(1, 895) = 0.021, P = 0.884$)。这个结果与低音点的结果不一致。

表 5 中, 3 种韵律边界后低音点差异显著 ($F(2, 1481) = 32.227^{**}$); 韵律短语和语调短语边界前音节的低音点差异也显著 ($F(1, 895) = 13.489^{**}$)。高音点的结果之所以不显著, 或者说高音点不具有标志韵律层级结构的作用, 正是因为高音点受重音的影响, 而低音点受重音的影响不大。如果一个韵律词被重读, 那么它的高音点提高, 因此从平均的结果来看韵律词边界处高音点的值与语调短语边界也就不会有太大差异了。既然高音点不能标志不同等级的韵律边界, 音域也就不是一个可靠的变量了。理论上, 边界末音域应该比边界前窄^[12], 但是边界末出现重音, 高音点提高后, 音域反而更大, 举例见图 2。通过重音知觉标注, 图 2 中, “改变”和“电影”都被重读。“变”的高音点显著地高于前面的音节, 而低音点与前面的音节相比保持一个较好的下倾趋势。句末“电影”被重读, “电”的高音点值较大, 以致句末的音域也不缩小。图 2 中还可以看出, 在韵律短语和语调短语后, 低音线的重置是系统变化的, 而高音点的变化受重音的影响。汉语中, 用低音线研究汉语的韵律结构是合理的, 而不是用高音点、平均音高或音域。

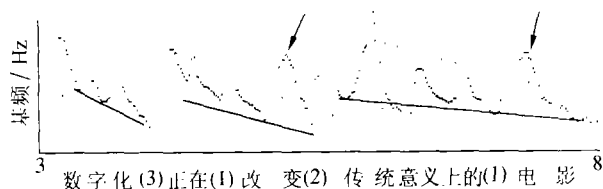


图 2 高音点受重音影响示例图

不同韵律边界对应不同语法成分时的低音点见表 7。

表 7 中, 同样是语法词, 对应韵律短语和语调短语不同韵律等级时, 其边界前音节低音点的值差

异不显著 ($F(2) = 9.74, P = 0.28$)。语法短语对应不同韵律等级时, 低音点差异也不显著 ($F(1) = 0.17, P = 0.676$)。这可能是因为, 语法词和语法短语的音节数比较少 (一般少于 5—6 个), 音高曲线下倾的时间有限, 下降的量也有限, 不足以区分不同的韵律等级。而语调短语对应不同的语法成分时, 其低音点的差异显著 ($F(2) = 11.81^{**}$)。语调短语对应的句法单元越大, 音高曲线的下降有更多的时间, 下降量也越大。

表 7 对应不同语法成分时韵律边界音节低音点 /Hz

韵律边界	语法词	语法短语	小句
韵律词	154	—	—
韵律短语	141	123	—
语调短语	135	122	118

总的来说, 边界前音节低音点的值以及音高重置程度可以区分不同等级的韵律边界, 特别是较高的韵律边界。边界等级越高, 边界前音节的低音点越低, 边界重置程度也越高。音域下限的估计可以参考^[15]给出的 4 个声调的相对关系图。但对于语音工程, 如何确定语句的音高下线还有待进一步研究。

2.2.3 韵律层级边界的无声段

无声段的长度与知觉等级之间的关系见图 3。

由图 3 可见, 知觉等级基本上随无声段长度呈对数增长。无声段与知觉等级的相关系数为 0.793^{**}。

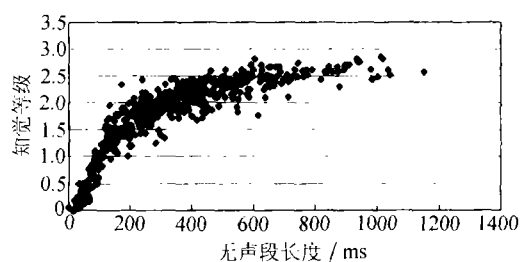


图 3 无声段长度与知觉等级的关系

51% 的韵律词边界的无声段小于 10 ms。即使韵律词后有无声段, 无声段也较短 (Mean = 63.3 ms, SD = 18.94)。

韵律短语和语调短语边界处无声段长度的分布见图 4。从图 4 中可见, 韵律短语边界的无声段比较集中 (Mean = 147, SD = 101.93), 而语调短语边界的无声段比较离散 (Mean = 408, SD = 194.29)。无声段长度在 300 ms 以上, 绝大部分情况下就是语调短语边界。而无声段在 300 ms 以下时, 更多的情况是韵律短语边界。

不同韵律成分对应不同语法成分时无声段的长度见表 8。

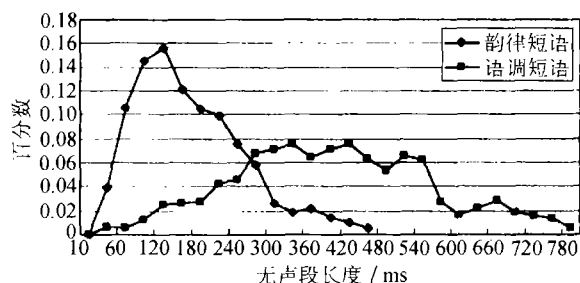


图 4 不同韵律边界无声段长度次数分布图

表 8 韵律边界对应不同句法成分时无声段长度 /ms

韵律边界	语法词	语法短语	小句
韵律词	26	—	—
韵律短语	112	165	—
语调短语	271	372	465

从表 8 中可见, 韵律等级越高, 无声段的长度越大; 相同的韵律边界对应的语法成分越高, 无声段的长度也越大。对照表 7 和表 8 可以发现, 语法词对应不同韵律边界时, 其低音点的值没有显著差异, 但无声段的差异显著。

不同边界间边界前音节的延长量的变化范围是 11% ~ 32%(见表 3), 而边界处无声段长度的变化范围是 20~780 ms(见图 4)。可见, 无声段的差异对边界等级带来的影响远大于边界前音节的延长。因此, 无声段的作用和边界前音节的延长的作用是不可以在一个层面上对比的, 很难说两者是增益的或互补的。边界前音节的延长是有限的, 而无声段的长度是自由的。两者对于区分不同的韵律边界有不同的功能。

以上是对一个发音人的语料进行的分析, 这些结果对其它发音人是否具有推广性? 另外, 这部分语料语速较慢, 语速加快后不同韵律边界的声学线索有什么新的特点? 针对这两个问题, 以下将对另一位发音人, 较快语速的语料作对比分析。

2.3 语速较快的语料库分析

2.3.1 韵律层级边界的时长分析

两种语速, 不同韵律边界前音节时长延长量的比较见图 5。

从图 5 中可见, 两种语速共同的结果是, 随着韵律边界等级的提高, 时长延长量增大, 但是时长的延长是有限的(不超过 30%)。较快语速的情况下, 韵律词与韵律短语边界相比, 边界前音节的时长差异显著 ($F(1, 227) = 6^*$)。而韵律短语和语调短语边界相比, 边界前音节的时长差异不显著 ($F(1, 83) = 1.51, P = 1.22$)。这些结果都与语速较慢的情况是一致

的。由此可见, 两种语速下, 边界前音节的延长都是区分较小韵律边界的线索。

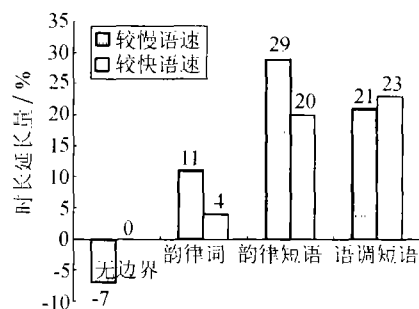


图 5 两种语速不同韵律边界处音节延长量的比较

2.3.2 韵律层级边界的音高分析

由表 4 的结果发现阳平、上声和去声的低音点在标志韵律层级结构时的作用基本是一致的, 因此这里只取 3 个声调低音点的平均值进行分析。较慢和较快两种语料, 在不同韵律边界前音节低音点比较见图 6。

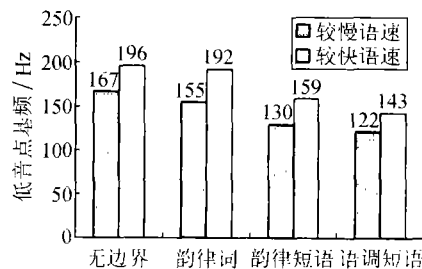


图 6 两种语速不同韵律边界处前音节低音点的比较

从图 6 中可见, 虽然两位发音人的绝对音高上有差别, 但共同的结果是, 随着边界知觉等级的升高, 低音点降低。语速较快的情况下, 不同韵律边界处低音点的差异也显著 ($F(3, 582) = 10.61^{**}$)。

较快与较慢两种语料的韵律边界前后低音点差值的比较见图 7。

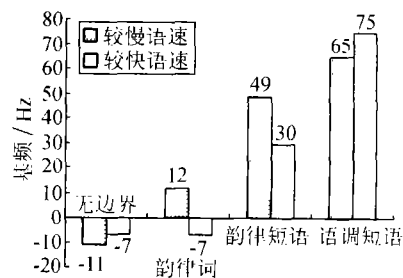


图 7 两种语速不同韵律边界前后低音点差值

从图 7 中发现, 较快的语速, 不出现韵律边界时, 低音线也有下降趋势。另外, 边界等级越高, 边界前后处两音节的差值越大。两种语速相比不同的是, 韵律词边界处, 较快的语速没有出现音高曲线的

不连续。这可能是因为在语速较快的情况下, 韵律词边界处没有足够的时间使音高有显著的下降。

2.3.3 韵律层级边界的无声段

较快的语速中无声段与韵律边界知觉等级的关系见图 8。

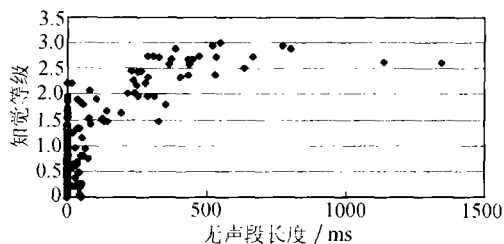


图 8 无声段长度与知觉等级

图 3 和图 8 相比可以看出, 两种语速韵律边界知觉等级与无声段长度都基本呈对数关系。在有无声段的情况下, 语速较快时, 无声段与知觉等级的相关为 0.849^{**}。这个值略大于语速较慢时的 ($r = 0.793^{**}$)。这是因为, 语速较慢时, 音高突变等其它线索更突出。而语速快时, 音高下降和边界前音节的延长程度都较小, 无声段对知觉等级的贡献就更大。

3 讨论

语料库分析越来越多地受到研究者的重视。本研究对语速不同的两个语料库进行了分析, 系统地讨论了边界前音节的延长、音高变化以及无声段这三种声学线索对不同韵律边界的贡献。

韵律边界处有边界前音节延长、音高不连续、音高重置和无声段 4 个声学表现, 按实际出现的情况, 有表 9 的 9 种组合。音节的延长量大于 0 则标记为边界前音节延长, 用“+”表示, 否则认为没有延长, 用“-”表示。与此类似, 边界前后低音点的差值在 10~20 Hz 之间则标记为低音线的不连续(+); 大于 20 Hz 标记为低音线重置(+)。无声段长度大于 80 ms 则认为有明显的无声段(+), 其余情况均注为“-”。两种语速的语料, 6 种声学线索组合条件与知觉等级的关系见图 9。

表 9 与韵律边界有关的声学特征有/无组合

声学特征	1	2	3	4	5	6
边界前音节延长	-	+	+	+	+	+
音高不连续	-	-	+	+	+	+
音高重置	-	-	-	+	-	+
无声段	-	-	-	-	+	+

图 9 中可以看出, 虽然语料来自两个发音人, 并且语速有很大差别, 但声学线索与知觉等级间的关

系是一致的, 即声学线索越多时, 知觉等级越高。这个结果和文献 4 的非常一致, 但 De Pijper 只分析了音高和停顿叠加作用的结果。

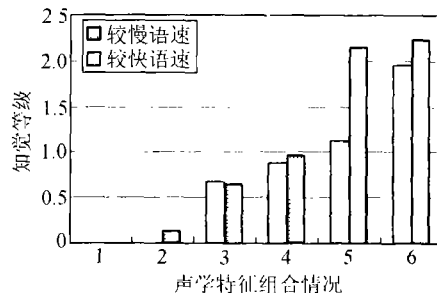


图 9 两种语速, 三种声学线索与知觉等级关系的比较

韵律边界处三种声学线索, 边界前音节的延长、音高变化和无声段, 对于韵律边界的知觉的贡献是一级级深入的。只有边界前音节的延长时, 边界前音节的延长可以区分无边界和弱边界。出现了音高变化后, 音高变化程度是决定知觉等级的主要声学线索, 虽然这时也伴随有边界前音节的延长。出现了无声段后, 无声段的对边界知觉等级的作用最大, 其次是音高重置, 而最终的结果是无声段长度与音高重置两种作用的和。这个结果支持 Streeter(1978) 提出的模型 3, 即对音高和时长两个加工器所输出的对边界判断的值进行平均。可能的情况是在计算平均值时, 无声段所占的权重大于音高的, 边界前音节的延长权重最小。本研究结果将对文献 21 中语音合成的韵律模型提供理论的依据和参数调节的指导。

4 结论

本研究主要得到以下结果:

- (1) 汉语语句音高的下倾和重置是由音域下限的移动实现的。
- (2) 韵律词边界的声学线索是低音线的不连续性和边界前音节的延长, 一般没有无声段。
- (3) 韵律短语和语调短语边界的声学线索是低音线重置和无声段。并且, 边界等级越高, 低音线重置程度越大, 无声段的长度也越长。
- (4) 知觉等级与无声段长度成对数增长关系。

两种语速相比, 边界前音节的延长、音高重置程度和无声段长度这些声学线索的相对变化都是一致的。只是语速较慢时, 音节的时长和无声段长度的绝对值更大。

综上所述, 边界前音节的延长和音高的不连续是弱韵律边界的声学线索, 而音高重置和无声段是强韵律边界的声学线索。并且, 弱边界上的声学线索

在强韵律边界上也有所表现。总体来说, 韵律边界等级是三种声学线索综合作用的结果, 即边界前音节的时长、音高变化(音高不连续和音高重置)和无声段。线索越多, 韵律边界等级越高。

致谢

社科院语言所曹剑芬研究员和 Motorola 公司祖漪清研究员在本文撰写过程中提供了宝贵意见。中科院心理所郑波和周迅溢两位同学在语料库的声学分析和知觉标定实验中投入了大量辛勤劳动。声学所贺琳女士为语料库的录制和整理付出了艰辛的工作。首都师范大学教育系郭春彦教授为知觉实验的顺利完成提供了极大的便利条件。作者在此深表感谢!

参 考 文 献

- 1 Selkirk E. The role of prosodic categories in English word stress. *Linguistic Inquiry*, 1980; 11: 563—605
- 2 Nespor M, Vogel I. Prosodic phonology. Foris, Dordrecht, Holland: 1983
- 3 Beckman M, Pierrehumbert J. Intonational structure in Japanese and English. *Phonology Yearbook 3*, edited by J. Ohala, 1986: 255—309
- 4 De Pijper J R, Sandemrman A A. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal Acoustic of Society of America*, 1994; 96(4): 2037—2047
- 5 Wightman C W. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal Acoustic of Society of America*. 1992; 91(3): 1707—1717
- 6 Booij G. Principles and parameters in prosodic phonology. *Linguistics*, 1983; 21: 249—280
- 7 贺琳, 初敏, 吕士楠等. 汉语合成语料库的韵律层级标注研究. 五届全国现代语音学学术会议论文集. 北京: 清华大学出版社, 2001: 323—326
- 8 Pierrehumbert J. The phonology and phonetics of English intonation. Ph.d. thesis, Massachusetts Institute of Technology, Cambridge: 1980
- 9 Ladd R. Intonational phonology. Cambridge University Press: 1996: 7—29
- 10 Swerts M. Prosodic features at discourse boundaries of different strength. *Journal Acoustic of Society of America*, 1996; 101(1): 514—521
- 11 Streeter L A. Acoustic determinants of phrase boundary perception. *Journal Acoustic of Society of America*, 1978; 64(6): 1582—1592
- 12 Marc Swerts, Don G. Bouwhuis, Rene Collier. Melodic cues to the perceived “finality” of utterances. *Journal Acoustic of Society of America*, 1994; 96(4): 2064—2075
- 13 Chao Y R. A preliminary study of English intonation (with American variants) and its Chinese equivalents, The TS'ai Yuan P'ei Anniversary Volume 1932
- 14 沈炯. 北京话声调的音域和语调. 北京语音实验录, 北京: 北京大学出版社, 1985: 75—107
- 15 沈炯. 汉语语调模型刍议. 语文研究, 1992(4): 16—24
- 16 Marc Swerts. Prosodic features at discourse boundaries of different strength. *Journal Acoustic of Society of America*, 1997; 101(1): 514—521
- 17 杨玉芳. 句法边界的韵律学表现. 声学学报, 1997; 22(5): 414—421
- 18 钱瑶, 初敏, 潘悟云. 普通话韵律单元边界的声学分析. 五届全国现代语音学学术会议论文集, 北京: 清华大学出版社, 2001: 70—74
- 19 David Crystal 编, 沈家煊译. 现代语言学词典(第4版). 北京: 商务印书馆, 2001: 58—321
- 20 于倍, 吕士楠, 杨玉芳. 汉语语句重读音节音高变化模式研究. 声学学报, 2002; 27(3)
- 21 陶建华, 蔡莲红. 基于音节韵律特征分类的汉语语音合成中韵律模型的研究. 声学学报, 2003; 28(5): 395—402