

# 面向语音合成的新闻播报语音库构建及其特殊韵律结构\*

邹煜<sup>1</sup> 何伟<sup>1</sup> 侯敏<sup>1</sup> 滕永林<sup>1</sup> 朱维彬<sup>2</sup>

1. 中国传媒大学国家语言资源监测与研究中心有声媒体语言分中心 北京 100024

2. 北京交通大学信息科学研究所 北京 100044

E-mail: {zouiy, hewei, byhoumin, tengyonglin}@cuc.edu.cn; wbzhu@bjtu.edu.cn

**摘要:** 韵律结构及其与句法语义的关系对于语音合成、语音识别以及自然口语的理解来说已经变得越来越重要了。本文是在构建基于语义的语音合成韵律分析数据库基础上,发现并初步分析了广播新闻播音中具有嵌套结构的复合韵律短语。经分析,我们发现复合韵律短语内部除了具有语音声学上的音高下倾趋势外,还与句法结构和语义表达具有颇强的关联性,它在播音员的创作中也发挥着十分重要的作用。基于这种特殊的韵律结构进行语音韵律建模,有利于提升语音合成的表现力。

**关键词:** 广播新闻 韵律结构 复合韵律短语

## A Broadcasting Speech Corpus for Speech Synthesis: Annotation and Its Special Prosodic Phrasing

ZOU Yu<sup>1</sup>, HE Wei<sup>1</sup>, HOU Min<sup>1</sup>, TENG Yonglin<sup>1</sup>, ZHU Weibin<sup>2</sup>

1. Broadcast Media Language Branch, National Language Resources Monitoring and Research Center, Communication University of China, Beijing 100024

2. Institute of Information Science at Beijing Jiaotong University, Beijing 100044

E-mail: {zouiy, hewei, byhoumin, tengyonglin}@cuc.edu.cn; wbzhu@bjtu.edu.cn

**Abstract:** Prosodic information becomes more and more important in speech recognition, speech synthesis and understanding, especially in the spontaneous speech. But there is some difference of prosodic structure in broadcasting speech, our earlier investigations of presenter's spoken language revealed that the prosodic chunk of the talking style uses the bigger unit and a more diversified syllable density change. This paper is intended to analyze the special prosodic phrase in broadcasting news program, namely complex prosodic phrase (C-PP). In our study, the C-PP, which there is the pitch declination trend within it, a stable syntactic structure and whole semantic chunk, has a very important role for creation of announcers.

**Keywords:** Broadcasting news, prosodic structure, complex prosodic phrase (C-PP)

\* 本研究得到了教育部科学技术研究重点项目(项目批准号:107118)和国家863项目(项目批准号:2007AA01Z198)的资助。

## 1 前言

众所周知，韵律在言语交际当中具有相当重要的作用。关于韵律及韵律结构，许多研究表明，在韵律结构中，韵律短语和语调短语内部低音线的音高重置和边界处的无声间断是其重要的声学相关物（Yang Yufang *et al.* 2002, Wang Bei *et al.* 2004, 熊子瑜 2003）。也有研究表明，由于韵律短语边界与句法短语的边界之间有着很强的相关关系，因此，对于韵律短语边界的考察，有助于我们把握上下文中韵律与语义表达的关系（Price *et al.* 1991）。也就是说，韵律短语的边界在语言理解中能有效帮助消除语义上的歧义，在语言理解中具有重要作用。

这同时说明，韵律结构与语义的关联性，有助于提升语音合成系统的表现力。然而，大部分用于语音合成的语音数据库只是要求发音人按照规定的语速和语调读一些孤立的句子（A. Batliner *et al.* 1995），这就很难真实地反映上下文中韵律与语义的相关关系。

我们前期的研究发现，基于真实语言环境的广播电视节目，能够很好地弥补这一不足。通过分析，广播新闻播音的韵律表现有一些独特之处，比如主持人（或播音员）谈话体在韵律调节方式上多使用大韵律单元，音节的疏密变化也更为丰富（邹煜，2007）。

基于此，我们在构建面向汉语语音合成的新闻播报数据库的过程中，发现在广播新闻播音中有一种特殊的韵律短语结构——即复合韵律短语，它不光结构复杂，而且包含的音节数量也比较的多。因此，本研究还对这种复合韵律短语进行了初步考察和分析。

## 2 语料

本研究的语料主要来源于中国传媒大学国家语言资源监测与研究中心有声媒体语言分中心的动态流通语言语料库。该语料库包括两个密切相关的子语料库，即文本语料库和语音语料库，它既包括了不同时期的广播电视音视频语料。本研究从中选取了部分语音语料进行韵律标注，旨在构建面向汉语语音合成的新闻播报语音数据库并对其韵律特点进行初步分析。

### 2.1. 语音数据的选取

经过分析，我们将中央人民广播电台实际播出的广播新闻节目《新闻和报纸摘要》作为面向汉语语音合成的新闻播报数据库的数据来源。该节目的内容和形式比较丰富，其主要内容包括：国内外要闻和中央报纸的言论。该节目总时长为 30 分钟，每期节目由一男一女两名播音员轮流播报，节目构成要素有开始曲、节目正文（含呼台号、开始语、内容提要 and 结束语）、间奏曲和广告等。其中节目正文平均时长为 28 分钟左右。我们最终将 2006 年 1 月 1 日至 2008 年 6 月 30 日间共计 908 期节目（约合 26020 分钟）作为数据选取的对象。

为什么要选择《新闻和报纸摘要》节目呢？理由如下：

- 1) 在以受众为对象实际播出的广播节目中，播音员处于真实的语言语音环境。
- 2) 播报文体规整，利于语言学分析；发音风格内敛稳定，便于语音学分析及建模。
- 3) 该节目是中央人民广播电台的王牌新闻节目，它的收听率 50 年来始终名列中央人民广

播电台各类节目榜首，拥有广泛、长期而固定的听众群，播音员精神状态饱满，语音表现力丰富。

## 2.2. 语音数据的预处理

在确定的 908 期节目中，我们以一条（则）新闻为单位对每期节目进行了切分，并与其文本进行了对齐。由于新闻文体的不同，所切分出来的语音片段的长度也不同。在这 908 期节目中，有 4 男 4 女比较固定的 8 位播音员。

我们根据其发音风格类型，最终选取了音色宽厚、声音弹性大、播音处理比较细腻、丰富的两位女播音员的数据作为新闻播报数据库内容进行进一步处理（其中一位播音员的语音数据作为主要的数据库，另一位播音员的数据作为对比数据库）。这样，我们就形成了一个待标注的、约 13 小时的面向语音合成的新闻播报语音库，其数据的具体分布详见表 1 和表 2。

**表 1 待标子库及其对齐文本的分布情况**

	文件数	文本大小 (byte)	语音文件 长度 (秒)
主数据	1601	428680	44363
对比数据	110	59220	6382
<b>总计</b>	<b>1711</b>	<b>487900</b>	<b>50745</b>

**表 2 待标子库播音文体风格的分布（单位：秒）**

播音文体风格	主数据	对比数据	总计
新闻播报体	39635	3222	<b>42857</b>
评论播音	1086	1108	<b>2194</b>
通讯播音[A]**	2009	281	<b>2290</b>
通讯播音[B]	1117	1557	<b>2674</b>
讣告播音	169	82	<b>251</b>
其他	347	132	<b>479</b>

\*\* 根据通讯的构成要素，我们把通讯分为两类，一类是由单一要素构成的，即由一名播音员单纯用自己的声音来表现，称之为 A 类通讯；另一类是由两种或两种以上要素（如播音员的声音、记者的声音、被采访者的声音、现场同期声和背景音乐等）构成的通讯，称之为 B 类通讯。

## 2.3. 标注方法

经过上述预处理后，6 位标注者对所有的语音数据进行了音段切分和韵律标注，其中韵律标注主要标注了韵律边界、重音等级和篇章基调。其中，韵律边界主要标记言语的韵律结构。中断指数（break index）代表感知到的音节之间以及音节和无声段（silence）之间的音联程度，每一个音节之后都必须有一个中断指数标记。不同的韵律边界就形成了韵律层级组织结构。根据广播新闻播音的特点，我们定义的韵律层级结构从小到大可以分为：韵律词（prosodic word）、韵律短语（prosodic phrase）、语调短语（intonational phrase）、语调组群（intonational group）和意合群（meaning expression cluster）：

- /1 用以标记韵律词的边界。
- /2 表示韵律短语的边界。
- /2\* 用以标记复合韵律短语（complex prosodic phrase，即 C-PP）——含有嵌套结构的韵律短语的边界，同时在其左边界处标记“^”。
- /3 用以标记语调短语的边界。
- /4 用以标记语调组群的边界。

- /5 表示意合群的边界。一个意合群通常由一个或多个语调组群组成。在播音创作中，播音员透过意合群可以“提纲挈领地掌握作者的思路和文章的结构，即文脉”，“播音时才可能表现出明晰的条理性和有机的完整性”（张颂，2003）。一般情况下，通讯报道可以根据情节的发展、矛盾展开的不同阶段来确定意合群的边界；评论则可以按照“提出问题、分析问题、解决问题”的脉络来划分意合群的边界；一则篇幅较长的新闻可以按照“标题”、“导语”、“主体”、“背景”和“结尾”分为五个意合群；短新闻或简讯一般由一个意合群组成。

我们的韵律层级结构和标注实例如图 1 和图 2 所示：

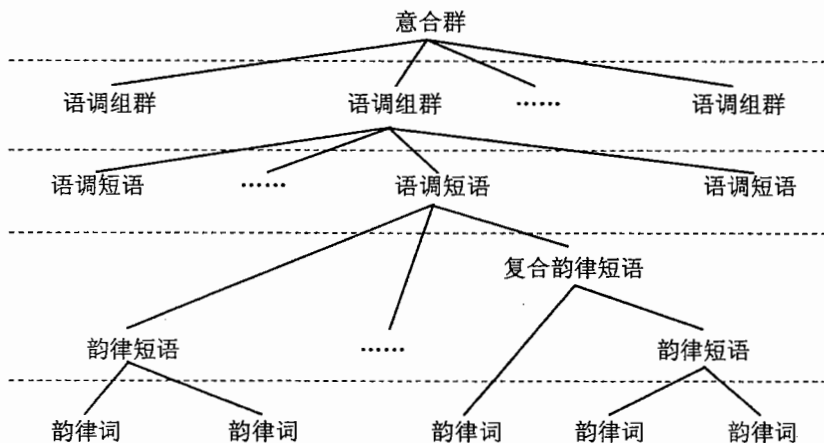


图 1 新闻广播节目中韵律层级结构示意图

#20070912\_08\_zl:  
 中共中央/1 政治局/1 委员/2 国务/1 委员/2 ^公安部/1 部长/2 周永康/3 ^在/1+ 昨天/1 召  
 开的/2\* 全国/1 公安机关/2 ^电视电话/1 会议上/2 强调/4 公安机关/1 要/2 紧紧/1 围绕/2  
 ^做好/1+ 党的/1 十七大/3 [安全]5/1 保卫/1 工作/2 这一/1 中心任务/4 [积极主动]4/1 [扎  
 扎实实]4 地/2 开展/1 工作/3 ^[确保]4/1+ 社会/1 和谐稳定/4 ^为/1+ 党的/1 十七大/2\* 胜  
 利/1 召开/2 ^创造/1+ 良好的/1 社会环境/5

图 2 标注示例<sup>①</sup>

### 3 初步结果分析与讨论

通过对标注语料的分析，我们发现每一种韵律边界的数量并不相等，韵律词的边界约占所有边界的 53%，韵律短语、语调短语、语调组群和意合群的边界则分别是 35%、8.2%、2.1%和 1.5%

<sup>①</sup> 示例中标记为“/1+”的地方表明此处的韵律词边界与一般的韵律词边界略有不同。

(详见表3)。

表3 不同韵律边界的分布情况

边界类型	频率(次)	百分比(%)
韵律词边界	54062	53.2%
韵律短语边界	35542	35.0%
语调短语边界	8296	8.2%
语调组群边界	2091	2.1%
意合群边界	1561	1.5%

根据表3的数据显示,除韵律词的边界外,韵律短语边界的频率是最高的。通过分析韵律短语边界的组成类型,其中复合韵律短语边界占整个韵律短语边界的16.8%。这说明在广播新闻播音中,尤其是在《新闻和报纸摘要》节目中,复合韵律短语在播音员的播音创作具有重要的作用。

新闻广播播音中为什么会存在复合韵律短语(C-PP)?它是怎样形成的呢?曹剑芬(2003)指出,韵律短语作为介于韵律词和语调短语之间的中等韵律组块,一般由7±2个音节构成。另一方面,韵律短语一般由一个以上的韵律词构成,如果一个韵律短语包含有两个以上的韵律词,其间的韵律词边界必定会有所不同,这样就会形成嵌套结构,比如A+(B+C)、(A+B)+C、A+(B+C)+D、(A+B+C)+D、A+(B+C+D)、A+B+(C+D)、(A+B)+(C+D)……。因此,我们将凡是具有嵌套结构的韵律短语成为复合韵律短语,没有嵌套结构的即为简单韵律短语。复合韵律短语的标注如下:

- ^A/1+B/1 C/2\*, 例如“^取得/1+ 新/1 突破/2\*”、“^接受/1+ 群众/1 监督/2\*”
- ^A/1 B/2 C/2\*, 例如“^又好/1 又快/2 发展/2\*”
- ^A/1+B/1 C/2 D/2\*, 例如“^作为/1+ 改进/1 服务的/2 方向/2\*”
- ^A/1 B/1 C/2 D/2\*, 例如“^中华/1 人民/1 共和国/2 政府/2\*”
- ^A/1+B/1 C/1 D/2\*, 例如“^坚持/1+ 文明/1 规范/1 服务/2\*”
- ^A/1+B/1+C/1 D/2\*, 例如“^维护/1+ 西藏/1+ 社会/1 稳定/2\*”
- ^A/1 B/2 C/1 D/2\*, 例如“^今年/1 60岁的/2 济南市民/1 郑承镇/2\*”
- .....

那么,复合韵律短语具有哪些声学语音学特征呢?有研究表明,韵律短语具有相对稳定的短语调模式,即音高音阶的逐步下倾(declination)走势,具有相对稳定的短语重音配置模式,即与句法结构相关的常规重音模式(曹剑芬,2003)。也有研究结果显示,广播风格的韵律短语的低音线具有音高下倾的趋势(Wang Anhong *et al.* 2002)。

图3是语料中关于复合韵律短语的示例。依据图3我们可以看出,复合韵律短语的边界特征和内部的音高模式与韵律短语没有区别,如“中国/1+ 全国/1 人大/2\*”和“同/1+ 希腊/1 议会的/2 关系/2\*”是两个复合韵律短语,其内部没有明显的音高重置,有明显的音高逐步下倾走势,“/2\*”处的边界特征与独立的“/2”是一致的,因此,其在韵律层级上仍属于韵律短语级,只是其内部嵌套有韵律短语结构。另一方面,复合韵律短语还与句法结构和语义有着很强的关联性。这表明,语义信息是广播新闻播音中播音员播音创作中依据的重要因素。

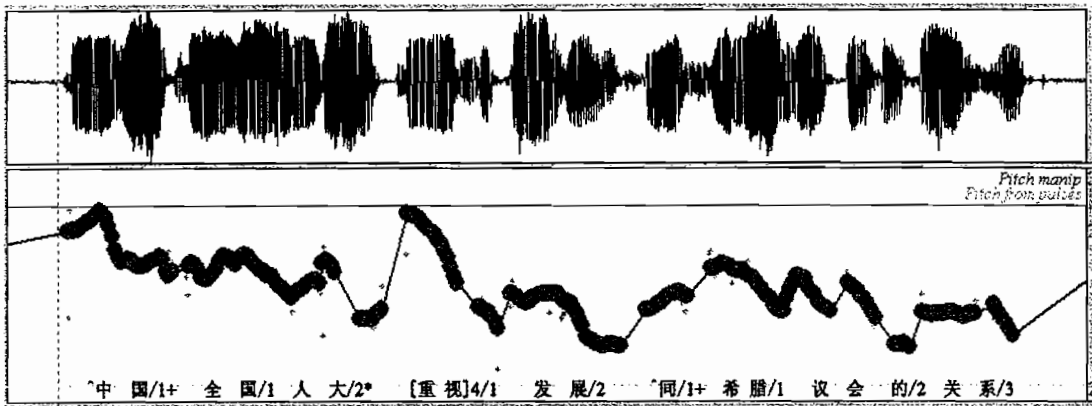


图3 复合韵律短语示例

#### 4 结论及下一步工作

通过前面的分析我们初步发现,复合韵律短语作为广播新闻播音中具有嵌套结构的特殊韵律短语,其内部不止具有语音声学上的音高下倾趋势,而且与句法结构和语义关联性强。同时,它在播音员的创作中也发挥着十分重要的作用。

在以受众为对象实际播出的广播节目中,播音员处于真实的语言语音环境,其语音表现力丰富。基于这种特殊的韵律结构进行建模,有利于提升语音合成的表现力。这有待于进一步研究。

#### 参 考 文 献

- [1] A. Batliner, R. Kompe, A. Kießling, E. Nöth, H. Niemann and U. Kilian, "The Prosodic Marking of Phrase Boundaries: Expectations and Results", *Speech Recognition and Coding – New Advances and Trends*, Rubio-Ayuso and Lopez-Soler (eds.), Springer, 1995, pp. 89-92.
- [2] P. Price, M. Ostendorf, S. Shattuck-Hufnagel and C. Fong, "The use of prosody in syntactic disambiguation", *Journal of the Acoustic Society of American*, 90, 1991, pp. 2956-2970.
- [3] Wang Bei, Yang Yufang, Lv Shinan, "Acoustic analysis on prosodic hierarchical boundaries of Chinese", *Acta Acustica*, Vol. 29, No.1, 2004, pp. 29-36.
- [4] Wang Anhong, Lu Shinan and Chen Ming, "Pitch Declination in the Statement Sentences in Mandarin", *Proceedings of ISCSLP 2002*, Taipei China, August. 2002.
- [5] Yang Yufang and Wang Bei, "Acoustic Correlates of Hierarchical Prosodic Boundary in Mandarin", *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, 2002, pp. 707-710.
- [6] 曹剑芬. 基于语法信息的汉语韵律结构预测. 中文信息学报, 2003, 17(3), 41-46.
- [7] 熊子瑜. 韵律单元边界特征的声学语音学研究. 语言文字应用, 2003(2), 116-121.
- [8] 张颂. 中国播音学(修订版). 北京: 中国传媒大学出版社, 2003.
- [9] 邹煜. 基于传媒语音语料库的主持人口语韵律研究. 博士学位论文, 中国传媒大学, 2007.