

新闻播报言语数据库构建及韵律标注*

李晓红¹, 朱维彬¹, 何伟², 邹煜²

(1. 北京交通大学 信息科学研究所, 北京 100044;

2. 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)

文 摘: 将篇章层面的文本转换为准确、生动的合成语音, 涉及了语言计算、韵律建模、语音合成等一系列技术问题。而这些问题的解决, 一方面需要制定一个更具科学性的韵律描述方案, 用以做为语言学与声学间的信息交互接口, 进而构建语音合成系统的语言学、语音学计算模块; 另一方面还将依赖具有丰富标注信息的言语数据库, 由此提供定性分析与定量计算的实际样本。因而, 言语数据库及其韵律标注, 构成了合成系统研究的基础。本文选取新闻播报语料为研究/加工素材, 综合考虑计算建模需求及样本的特点, 制定了一套新闻播报言语数据库的韵律标注内容及其标注规范。本文对语篇的停顿、重音、语气、语调的标注规范进行了具体的阐述; 通过对言语数据库构建及韵律标注的具体实施, 证实了所制定的韵律标注规范的科学性。

关键词: 言语数据库; 韵律标注; 新闻播报

中图分类号: TP391

1 引言

近些年, 语音合成技术虽然已经在计算方法和实现等方面取得了显著进步, 但合成语音的质量仍存在着缺陷^[1]。几乎所有的文语转换系统计算所涉及的最大语言单位都是单句, 对应的最大语音单位为语句, 即便输入的文本可能会是篇章或段落。由于文本分析计算所利用的皆为语句之下的分词及 POS 信息, 所分析的结果也只能是语句及之下的韵律层级结构。具有更准确、生动语义表现能力的合成语音的实现, 依赖于更精准的韵律描述, 除了韵律层级结构之外, 还需要体现语义焦点的重音, 实现特定语用功能的语调, 表达情感立场的语气, 与之相对应的是文本分析计算需要涉及合成文本的篇章结构、话题演变、主题与焦点、内在立场等更高层面的语言学信息。为此, 语音合成所处理的文本对象将上升到篇章层面, 相关的言语数据库构建、韵律描述内容及标注规范, 都将在篇章层面进行。

本文选取新闻播报语料作为合成语音

的参照目标和言语数据库的基础数据, 主要有以下考虑:

(1) 语音与文稿类型涵盖语音合成系统相当大比例的应用情形。

(2) 播音文稿内容广泛, 但文体规整, 有利于更深入的、准确的语言学分析。

(3) 言语表达方式丰富多样, 具有丰富的表现能力; 同时发音风格内敛稳定, 便于语音学分析及建模。

(4) 数据来源及品质稳定, 素材丰富, 可保证数据库的规模。

以下, 我们将针对新闻播报言语数据库的构建, 着重介绍篇章层面的韵律标注内容及标注实施规范。

2 韵律标注内容制定

2.1 韵律标注内容

韵律标注的实质就是对语音信号中具有语言学功能的韵律特征进行定性描写^[2]。而汉语的韵律标注系统已有相当多的工作。陈玉东^[8]将语段标注为六层: 拼音层、语调构造层、小句类型层、重音等级层、语段单

* 基金项目: 863 计划(2006AA010104、2007AA01Z198)

作者简介: 李晓红 (1984 年-), 女 (汉族), 河北省, 硕士生。

通讯联系人: 朱维彬, 副教授, E-mail: wzbzhu@bjtu.edu.cn

元层和单元边界层；李爱军^[10]所指的韵律标注系统是将每个句子分五层进行标注：正则发音层（*orthographic tier*）、声调和语调层（*tone and intonation tier*）、语句功能层（*sentence function tier*）、停顿指数层（*break index tier*）和重音/突显层（*stress/prominence tier*）。

由于新闻播报语料全是基于语篇的，所以本文标注的是语篇之下的停顿（韵律边界）、重音、语调和语气四个方面的内容。其中，停顿和陈玉东^[8]所指的单元边界层相似，和李爱军^[10]所指的停顿指数层相似，只是具体的层级结构不同；重音和陈玉东^[8]所指的重音等级层相似，和李爱军^[10]所指的重音/突显层相似；语调和陈玉东^[8]所指的语调构造层及李爱军^[10]所指的语句功能层相似；而语气的标注规范则是基于语篇的需要，采取的一个有别于其它标注系统的标注内容。

2.2 停顿（韵律边界）的标注

韵律边界主要标记言语的韵律结构。韵律结构是层级组织结构，从小到大可以分为：音节、音步、韵律词、韵律短语、语调短语、语调组群和意合群等。

已有研究中的停顿指数层和本文所指的韵律层级结构相似，如：李爱军^[10]所指的停顿指数层分为：*minor phrase*、*major phrase* 和 *sentence break* 三层；朱维彬^[9]所指的层级结构分为：*prosodic word*（*the minor rhythmic unit*）、*prosodic phrase*（*the intermediate rhythmic chunk*）和 *intonation phrase*（*the major rhythmic group*）三层；郑秋豫^[12]所指的停顿指数层分为：*syllables*（*SYL*）、*prosodic word*（*PW*）、*prosodic phrase*（*PPh*）、*breath group*（*BG*）和 *prosodic phrase groups*（*PG*）五层。这些分法都是基于一定的语料的，并且各有不同和相似之处。陈玉东^[8]对韵律单元的切分也依据了话语中有无明显无声段和听觉上有无明显的停顿，但是还参考音高的变化。单元边界和韵律单元的切分基本对应。韵律单元确定之后，各韵律单元之间的边界要依据韵律单元的大小和层级确定。单元边界一般有四种类型：语节级边界、跨语节级边界、小句级边界、语段级边界和无声段。语段是有声语言的一个重要层级^[8]，多

个语段构成语篇，语篇同样是有声语言的重要层级，本文采用的言语数据库的文本是基于篇章的，我们将其标注为六个层级：韵律词、复合韵律词、韵律短语、语调短语、语句和段落，具体的界定和标注如下：

段落（对应于“意合群”）：用以标注语势连贯、语义关联的语句组块；段落之间有大的停顿；一个段落通常由一个或多个语句组成；段落是语义相对完整、语音处理相对独立的有声语言段落，这里强调语义的完整性和语音的凝聚力^[8]；标注时采用“//+双回车（双空行）”。

语句（对应于“语调组群”）：用以标注语调连续的语调短语组块；语句之间有较大停顿；一个语句通常由若干个基频曲线依次递降的语调短语组成；标注时采用“//+单回车（换行）”。

语调短语：用以标注具有完整的语调，听感上可独立成句的韵律短语组块；语调短语间有明显停顿，边界处感知到的停顿较韵律短语边界处的停顿长；边界前后的连接较松散，而且有基频重设等现象；标注时采用“//”。

韵律短语：用以标注由若干韵律词构成、之间没有明显停顿的组块；韵律短语间有停顿；韵律短语是介乎韵律词和语调短语之间的中等节奏组块，（韵律短语一般由 7 ± 2 个音节构成^[11]，但在新闻播报语料中，韵律短语的作用范围有时候会超过 $7+2$ 个音节的限制）；韵律短语分为简单韵律短语和复合韵律短语，后者专指含有嵌套结构的韵律短语；标注时采用“/”。

复合韵律词：用以标注超过一个音步、结合紧密的韵律词组合；韵律词间没有明显的停顿；标注时采用“*（词间）”，标注明显的节奏边界。在这个层级的标注中，复合韵律词的类型较为广泛，现举一例。

例 1 由于 主题 鲜明/、时效性* 强//，社区 文化* 活动/ 得到 居民的 积极 参与，

韵律词：用以标注节奏的基本单位，以两音节构成的音步为主，跨度在 $1\sim 4$ 音节；韵律词内没有明显停顿；标注时采用“空格”。

2.2.1 层级结构标注规范

按下列顺序实施标注：

（1）根据听到的读音、读法、断句、语

调来分析句子,给出相应的分界点。

(2) 根据词的重音进行分界。

(3) 观察句子的长度平衡、语法完整性、名意等辅助项来判断边界。

特例:

(1) 遇到“在”、“和”、“反倒”、“或”、“与”等词语的时候,大多是强调前面半句,因此在这类词语前给一个边界,此词归于后面短语。

例 2 促进影响社会主义新农村建设/实际问题的解决/和整改任务的落实。

例 3 中国广播网银河台/与全球知名饮料厂商可口可乐联合主办/“我的舞台、我最精彩”活动。

(2) 由于读音轻重引起的像是边界的,一般不在此处标记边界,如人名。

例 4 胡锦涛三月三十一号在北京/会见了日本日中友好/七团体负责人。

(3) 不能确定是否加边界时,要根据句子长短、意思完整性、语法完整性等因素来综合考虑。

(4) 此外,我们选择的资深播音员在读到并列句时,习惯在倒数第二个词与后面的并列关系的词处停顿稍大,在判断边界时应根据具体句子读法来划分语调短语的边界。

例 5 此外他们还将/承诺人/、承诺事项/、承诺措施//、承诺实现等/列入承诺备案表中//,指定监督人/督促落实。

例 6 未来六周/,银河台/将进驻校园//,开展银河台/主持人选拔/、主持人访谈//、网络歌手见面会等/上百场活动。

2.2.2 划分韵律短语边界可依据线索:

在标注韵律短语的边界时,需要根据以下线索:

- 句中停顿
- 语调上扬
- 语调下倾
- 末音节拖长
- 仿当事人口气说话

例 7 中国证监会昨天宣布//已于近日批准/(末尾音节拖长)上海期货交易所上市黄金期货。

这是继锌/(并列结构)、菜籽油/(句中停顿)、现行低密度聚乙烯期货成功推出之后//,今年证监会批准的/(简短停顿)第四个期货新品种。

例 8 即使在为儿子治病/(重音,表强调)很需要钱的情况下//,他仍然坚持理想/(并列结构),接受学校返聘/(并列结构),奉献教育事业,

在回答妻子/(仿当事人口气说话)“辛辛苦苦一辈子你得到了什么/”这个问题时//,甘兰佑动情地说/(仿当事人口气说话),能为孩子们做点事/(句中停顿)我很快乐。

2.3 重音的标注

重音是一个非常重要的韵律特征。王蓓^[4]是从重音音节和时长的关系的角度来考察重音的,并得出音节重音级别是随重读程度的提高、时长延长量增加而增加的结论;重音的层级问题是实现标注的关键,朱维彬^[5]在语调短语和韵律词这两个层级上设定重音指数,每个层级上又分三个级别:重、中、轻,并且在实施重音级别标注时,综合利用了听觉和视觉信息;王韞佳^[6]是从语句中韵律词的重音和词所在的韵律边界的关系、词的重音感知与音节之间的高音点差和时长差的关系、双音节词的调类组合和重音感知的关系三个方面来考察重音的;从重音的表达功能出发,一般都应将汉语的语句重音分为节奏重音和语义重音两类,王韞佳^[7]分别从语句和短语两个层次对普通话语句中的语义重音分布进行了考察;陈玉东^[8]考察重音主要着眼于音高和音长特征方面,并将其分为语节级、跨语节级、小句级和语段级四个级别,按照级别进行标注。

针对基于语篇的新闻播报语料,我们将重音分为三个级别:语句重音、段落重音和篇章重音,分别用于标注句子级重音、段落级重音和语篇级重音,且在实施标注时,主要利用听觉上的突显度,并结合语篇的语义理解。

语句重音是语句中的突显成分,用“(…)”标定听感上的突显音节;段落重音是段落的语义焦点,用“[…]"标定听感上的突显音节;篇章重音是篇章的文眼,用“{…}"标定听感上的突显音节。

2.4 语气(基调)的标注

本文将情感语音的标注列入了标注内容,主要通过语气标注来体现。

张颂^[3]讲到,基调是指作品的基本基调,即作品的总的态度感情,总的色彩和分量。而基调的色彩、分量,既包括态度分寸,

又包括感情色彩。其中的态度分寸包括：肯定和否定、严肃和亲切、祈求和命令、客观和直露、坚定和犹豫五个方面；感情色彩包括：挚爱和憎恨、悲哀和喜悦、惊惧和欲求、焦急和冷漠、愤怒和疑惑五个方面。张颂^[3]同时又指出语气的色彩包括：爱和憎、悲和喜、惧和欲、急和冷、怒和疑五个方面，这是和感情色彩相对应的；语气的分量包括：重度分量、中度分量和轻度分量三个方面。

本文主要从语用功能角度来进行基调标注。语用功能角度包括态度、感受和性质三个维度，其中的态度和张颂^[3]所指的态度分寸是相似的，感受和张颂^[3]所指的感情色彩或语气色彩相似，但是张颂^[3]所指的某些态度分寸和感情色彩是有交集的，例如犹豫和疑惑，并且考虑到本文采用的言语数据库是基于新闻播报的，某些语气极少出现，同时为了实现标注实施的易操作性，本文将其精简分类如下：态度由赞扬、中性和贬斥构成；感受由高兴、中性和悲伤构成；性质由严肃、中性和轻松构成。

语气的标注主要是在篇章（discourse）、段落（paragraph）和语句（sentence）这三个层级上考察，标注符号是分别在每个层级的开头标“\D”、“\P”和“\S”，感受（feeling）标为“F”，态度（attitude）标为“A”，性质（property）标为“P”，这三个方面均分为三个维度，所有的三个维度均按顺序标为“+n,...,+1,0,-1,...,-n”（n=2 or 3），同一个类别的不同取值是和语气的分量相似的，代表了每个类别的程度（具体标注见例句9）。

2.5 语调的标注

语调的标注主要是在语调短语这一层级上，对四种句子类型（陈述句、祈使句、疑问句、感叹句）进行标注，标注符号分别是：“//、//#、//?、//!”，这和李爱军^[10]所指的语句功能层相似。例9说明了本文制定的韵律标注规范的具体标注情况：

例9 \D\F2 中共* 中央 政治局* 常委/、全国* 政协* 主席（贾庆林）/ 昨天 在 人民* 大会堂/ 会见了 前来 北京 参加/ “王宽诚先生诞辰（一百周年）/ 暨 中国科学院与王宽诚教育基金会合作二十周年纪念大会”的/ 香港 著名 实业家/、爱国 人士/（王宽诚）先生的 家属 一行。

\P\A2 贾庆林 说，王宽诚先生/ 一生 心系 民族 存亡//，关心 国家 发展//，为 中国 革命的（胜利）/ 和 新中国的（建设）//，为 香港的 平稳 过渡/ 和 顺利 回归// 做出了 [重要 贡献]。

\P\A1 他 希望 王宽诚 先生的 亲属/ 继承 先生（遗志）//#，发扬 “[宽以 济事，诚以 育人”的 宽城 精神]//#，继续 发展（壮大）/ 王宽诚 教育 基金会//#，一如既往的 关心 和 支持/ 祖国的 教育/ 和 人才 培养 事业//#，\S\A1 为 香港 地区的（繁荣 稳定）/、为 祖国 改革* 开放/ 和 现代化 建设* 事业// 做出 {新的 更大的 贡献} //#。

说明：语气的标注是针对语用功能的三个维度的，但是为了标注的简洁，当其中某一个维度的类别是中性时，省略其标记，默认为F0、A0或P0。

3 新闻播报言语数据库的构建

3.1 新闻播报言语数据库语料的选取

本文搜集了中央人民广播电台《全国新闻和报纸摘要》中的新闻通稿播报录音的2649段语音及其文稿，并将其做为初始语料来进行新闻播报言语数据库的设计与构建。

3.2 新闻播报言语数据库的初加工

根据对言语数据库的设计要求对新闻通稿播报录音的文本进行初加工，分别从播音人、播音文体、播音时长、文本长度（短、中、长）、发音方式（速度、力度、亮度、对比度）、语用功能（性质、态度、重要性）、篇章分类等方面进行了标记。

这里的发音方式和语用功能方面的标记是基于文本初加工的，有别于韵律标注规范中基于韵律标注的界定。发音方式的标注采用4维度10分制，以每条新闻作为篇章，在4个维度上主观给出0至10分值。速度、亮度、力度依据篇章主基调打分，0至10分值对应速度由慢至快，亮度由暗至亮，力度由轻至重，对比度依据篇章内速度、亮度、力度的波动幅度打分，0至10分值对应对比度由小至大。语用功能的3个维度是态度、性质、重要性，“态度”是由3维因素构成，

即喜悦和悲伤、赞扬和贬斥、中；“性质”是由3维因素构成，即好、坏、中；“重要性”是由3维因素构成，即大、小、中。这里的态度、性质和重要性是针对数据库初加工的，和前文讲到的语用功能的三个维度（针对标注的）是有区别的。播音文体是从播报、通讯（A、B）、评论、宣告、讣告等5个方面来分类的。篇章分类是从人文艺术、新闻与媒体、商业与经济、娱乐与休闲、计算机与因特网、教育、区域、自然科学、政府与政治、社会科学、医疗与健康、社会与文化等12个方面来分类的。

根据以上各个方面的分类，本文将各个文本所属类别一一标示出来，列于一个表中，这就完成了言语数据库的初步加工工作。

3.3 新闻播报言语数据库的设计

语料库本是为非限定领域的语音合成系统设计的，则在言语数据库的设计时要求尽量均衡不同长度、不同语用功能和篇章类型的文本的分布，本文根据发音风格（语用功能）、发音内容（篇章分类）、文本长度这三个角度，从2649个文本中挑选出600个文本，使得这600个篇章的文本长度适中，不同的发音风格和篇章类型的文本都得到了均衡的分布。这就完成了我们所需要的言语数据库的设计工作。

4 新闻播报言语数据库的韵律标注

4.1 校对文本

我们所设计的语料库包括600个篇章，在语料选取和初加工时，已尽量使其达到发音风格、发音内容和文本长度三方面的均衡分布，接下来就是结合已有的中央人民广播电台《新闻和报纸摘要》中的新闻通稿播报的录音和文本以及网络上的新闻文本，通过对文本中文字的比对，完成对语料库中错别字、错误的标点符号的校对。

4.2 录音

在录音过程中涉及到2个问题，第一，说话人的发音状态保持稳定，以排除由于生理变化引入的音色的变化；第二，发音风格与文稿内容保持一致，要求根据语篇需要通过适当的语气恰当地表达情感，这有助于我们对情感语音的考察和标注。

为此，我们选择一名资深专业的播音

员。她的发音错误很少，并且有很强的自我调节能力，因而提高了录音效率。即便如此，录音监控也是必需的。为保证每次录音的设备设置（比如音量）及播音员状态的一致性，我们预先录制了几个标准语句/篇章作为参考语音。在每次录音开始时，播音员都会根据这些标准语音的重放而调整自己^[9]。

满足以上条件后，就可以进行录音工作了。完成第一次录音之后，还需要边听录音边根据校对过的文本进行识别，及时发现错误或不当的录音，并进行补录，直至所有的语音文件都达到要求为止。

4.3 韵律标注的实施

人工韵律标注的一个关键问题是提高不同人在同一时间以及同一个人在不同时间标注的一致性，下面是我们的实施安排^[9]：

- 50个篇章被用来做训练篇章，按照标注规范实施：
 - a. 两个人进行标注训练
 - b. 分析对比标注结果以便完善我们的标注规范
 - c. 对比结果用来训练语料
 - d. 对比结果用来达成不同标注人之间的一致
- 剩下的550个篇章做如下标注：
 - a. 一个人完成所有的标注
 - b. 另一个人同时标注10%的语料
 - c. 对比结果用来修正2个人的标注规范
- 将训练用的50个篇章重新标注，以评估同一个人不同时间标注的一致性。

实验结果表明：整个标注的平均一致性超过了90%，这说明我们的标注规范是科学可取的。

5 总结

本文介绍了基于语篇的新闻播报言语数据库韵律标注内容及标注规范的制定，介绍了所完成的新闻播报言语数据库内容，以及对其实施的韵律层级结构、重音、语气和语调的人工标注。通过对韵律标注规范的实施，揭示了2个事实：第一，用于基于语篇的语料库建设的这个韵律标注规范是合理的；第二，使用这个标注规范达到的一致性是可以接受的。

基于篇章的韵律标注规范及新闻言语语料库的构建,为基于数据驱动方法建立语言学、语音学计算模型,实现由篇章级的文本分析计算得到更为丰富的韵律描述,进而转化为具有准确、生动表现能力的合成语音的研究奠定了理论及资源基础。

参 考 文 献:

- [1] 朱维彬, 语音合成技术现状, 创新与和谐——中国声学进展, 科学出版社, 2008.8, 662-669
- [2] 邹煜, 胡凤国, 传媒语音语料库的预料选取和标注, 第八届中国人机语音通讯学术会议论文集, 2005
- [3] 张颂, 朗读学, 中国传媒大学出版社, 1998
- [4] 王 蓓, 杨玉芳, 汉语重读音节时长变化模式研究, 中国声学学会 2002 年全国声学学术会议论文集, 2002.
- [5] 朱维彬, 支持重音合成的汉语语音合成系统, 中文信息学报, 2007, 3
- [6] 王韞佳, 初敏, 贺琳, 冯勇强, 汉语中双音节韵律词重音感知的初步研究, 新世纪的现代语音学——第五届全国现代语音学学术会议论文集, 2001
- [7] 王韞佳, 初敏, 贺琳, 汉语语义重音分布的初步研究, 第六届全国现代语音学学术会议论文集(上), 2003
- [8] 陈玉东, 传媒有声语言语段的构造和调节, 北京大学博士论文, 2004
- [9] Zhu Weibin, Shi Qin, et al. Corpus Building for Data-Driven TTS Systems, IEEE TTS Workshop 2002, Santa Monica, USA, 2002
- [10] Li Aijun et.al. Speech Corpus Collection and the Annotation, Report of Phonetics Research, 2000
- [11] Cao Jianfen, The Rhythm of Mandarin Chinese, Report of Phonetics Research, 2002
- [12] Tseng Chiu-yu, Higher Level Organization and Discourse Prosody, The Second International Symposium on Tonal Aspects of Languages, 2006, 23-24