

# 基于新闻言语数据库的语气标注及其韵律特征分析

高莹莹、朱维彬

**摘要:** 本文提出了一套基于新闻言语数据库的语篇下的语气标注方案,并分别对大小两个规模的新闻言语数据库进行了语气标注,对标注过程中存在的一些疑点进行了列举和分析,最后以两种语气为例做了韵律特征分析。

**关键词:** 新闻言语数据库、语气标注

## 1. 引言

将文本转换为准确、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,用以作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[4]。

在语音合成的研究中,人们关注可懂度和清晰度,追求自然度。目前,相对于基于语料库的语音合成系统,依赖于对语音先验知识的基于规则的语音合成系统在自然度方面还远达不到让人满意的程度。分析其原因,人们认为主要是由于我们现阶段对音段和超音段信息的提取和利用还很少[9],在韵律信息描述方面还不够精确。更进一步说,目前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等[4]。为此,我们希望构建一个篇章层面的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语言学信息的言语数据库,具体到韵律标注主要包括语篇之下的停顿(韵律边界)、重音、语调和语气等四个方面的内容。本文将对其中的语气标注部分进行标注规范和实施细则等方面的介绍。

关于语音的情感、语气标注,国内外学者已有不同程度的研究。目前两种主要的情感表示方法是离散表示法和维度表示法。现阶段的情感语音合成,用到的多是离散表示法。比较完整的情感类型划分是 Plutchik 等人提出的“情感轮(emotion wheel)”分类方法,提出包括高兴、期望、愤怒、厌恶、悲伤、惊奇、恐惧、赞同在内的八种情感[2],而用到最多的是喜、怒、惊、悲四种情感。我们比较关注的是连续的维度表示。其中受到广泛认可的有 H.Schloberg 的“快乐维—注意维—强度维”的三维表示,和 R.Plutchik 的“激发维—评价维—强度维”的三维表示[3]。李爱军等提出了积极度(arousal 或 activation)、肯定度(pleasure 或 evaluation)、确信度(power)三个评价维度[6]。蔡莲红从三个方面研究语音表现:情感、风格和个性;指出情感是由心理—生理变化引起的情绪反应,除了激发度(arousal)和评价度(evaluation)两种情感维度外,还借鉴 Schlosberg 的情绪模式,把风格看成是由愉快-愤怒、紧张-松弛和激活水平这三个维度构成,分析了松弛、中性和活跃三种风格[7]。

不同于以往情感语音合成的概念[5]，我们不是基于特别为情感语音合成录制的感情色彩浓重的情感语料库来合成语音，而是基于大部分属于中性色彩的新闻播报言语数据库来进行情感语音合成的，参考张颂在《朗读学》[8]一书中对语气、基调的色彩和分量的阐述，并结合新闻播报稿的性质和特色，最终概括出了态度、性质、感受三个划分维度，其中态度维正向是赞扬，反向是贬斥，中间即为中性；与此类似，感受维正向分别为高兴和悲伤；性质维正反向是轻松和严肃。

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[4]：

- (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，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。

## 2. 方案设计

为了保证标注工作的准确客观，标注之前，我们需要制定一套完整的标注方案。首先，在原理上，应保证该方案的科学性。我们采用连续的维度表示，将语气空间划分为态度、性质、感受三个维度，用维度空间上的点表示测听语料的语气（如图 1 所示），相对于离散的语气表示，能更全面、细致、准确地表示出测听语料的所有语气成分。其次，在考虑这三个划分维度时，既保证尽可能涵盖新闻播报稿中的所有语气，同时每个维度又能从不同角度表现语音的语气，做到既全面又相互独立，即所谓的完备性。另外还对每个维度进行了种类细分和程度界定，进一步保证能包含不同种类和不同程度的语气。最后，在具体实施时，我们将标注内容符号化，标注程度离散化，保证方案的可实施性。

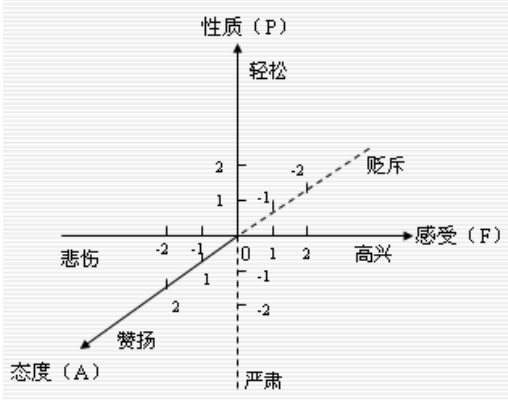


图 1：语气维度划分

### 2.1 标注方案

标注方案具体如下：

标注主要在篇章（discourse）、段落（paragraph）和语句（sentence）三个层级上进行，标注符号分别是“\D”、“\P”和“\S”，三个维度表示为：感受（feeling）标为“F”，态度（attitude）标为“A”，性质（property）标为“P”。

本文中我们将每种语气划分为三个程度，即主基调很明显的定为 2（或-2），程度稍弱但含语气成分的定为 1（或-1），语气平淡、听不出任何感情或态度的定为 0，其中严肃、悲伤和贬斥取负值，轻松、高兴和赞扬取正值。

为了尽可能保证标注的规范性、客观性，我们选用 2 名具有一定语音学背景的人进行标注工作。

### 2.2 注意问题

标注时需要注意以下问题：

- (1) 依据语音标注：由于最终我们要处理的是语音信号，所以标注时要尽可能的排除语义内容以及听话人本身立场等因素的影响，以听感为主要依据标注。
- (2) 标注不宜过细：由于我们现阶段要做的只是抓住整个篇章的主要基调即可，所以在单独语句的标注以及每种情感的程度区分上不需要分得过细。

(3) 一致性问题：为排除同一人不同时间的听感会有不同的问题，尽量做到一次性连续标注尽可能多的篇章；同时，两个标注人员遇到问题要及时讨论，即使对标注准则做出修正。

## 2.3 语气定义

另外，本文参考张颂在《朗读学》[8]里对一些态度分寸和感情色彩的定义，并结合新闻稿的播报内容，对每种语气做了如下定义：

(1) 严肃：指庄重、重视、不苟且、不随便，一般出现于讣告、国际事务等较庄重的场合。

(2) 轻松：指轻快、活泼、不重视、比较随便，一般用于日常生活讯息的播报。

(3) 赞扬：歌颂人物、英雄事迹，对好的行为、政策予以肯定等。

(4) 贬斥：批判某种违法乱纪或有违道德风尚的恶劣行径。

(5) 悲伤：是一种伤心、痛惜的感情，像沉痛、悲恸、哀婉、凄切等都是，一般用于发生了某些人员伤亡、财产损失、重要人物去世等。

(6) 高兴：是一种喜悦、快乐的感情，像兴奋、欣喜、欢快、愉悦等都是，一般用于取得了某种成绩、成就、成功等。

(7) 中性：并不见任何感情在里面，语气平淡。“无意肯定，无意否定”，“未见支持，未见反对”。

## 3. 标注实施

### 3.1 数据选取

本文选用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 600 篇语音及其文稿，并在语料选取和初加工时，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。

### 3.2 标注流程

为了保证不同人对同一文本以及同一个人不同时间标注的一致性，我们进行了如下几步标注过程：

(1) 首先由一人从 600 篇语音中选出 10 篇涵盖所有语气、且程度较为明显的典型范例，两人再分别按照最初设定的语气定义和程度的界定进行独立标注，然后对照标注结果，商讨出尽可能一致的标注准则；

(2) 再随机抽出 10 篇语料，两人仍然进行独立标注，比较标注结果，一致率达到 90% 以上则可以进行以下操作，否则则需对标注准则做进一步修正；

(3) 两人分别选取 300 篇语料进行独立标注，其中不确定的部分由两人共同讨论后再做出标注；

(4) 交换两人标注的 300 篇语料，分别将对方的 300 篇再标注一遍；

(5) 比较整个 600 篇的标注结果，计算出以句为单位的的不一致率为 53.88%，对不一致的语句进行商定，商定后得到最终两人都认同的标注结果。

### 3.3 小规模语料库的构建

为了方便以后的研究和学习，我们从 600 篇语料中又提取出了一个规模更小的语料库，提取办法如下：

(1) 找出 600 篇语料中每种语气程度最强的语句，统计出数目，确定出小语料库的规模为 60 篇语料，其中除包括七种只含一种语气的语料外，还包括三种含一种以上语气的混合语料，合起来共十种，所以暂定每种语气选 6 篇左右，之所以加入混合语气，主要是由于这些语料本身的语气色彩很明显，具有很强的代表性；

(2) 有些语气在整个 600 篇语料库中存在就比较少（如：轻松），所以选取数目少于

平均的 6 句，相反有些比较多，程度也比较明显（如：赞扬），则可以增加相应的数目；

初步确定的每种语气的篇章数目为：

表 1

语气	数目
轻松	5
严肃	4
高兴	8
悲伤	0
赞扬	8
贬斥	7
严肃+贬斥	7
严肃+悲伤	9
赞扬+悲伤	6
中性	6

注意，虽然表中没有单独含悲伤语气的语料，但在混合语气里悲伤的比例并不算少（合计 15 篇）；另外可以看出，整个小语料库中，含严肃语气的篇章最多（共 4+7+9=20 篇），悲伤（9+6=15 篇）次之，贬斥和赞扬均为 14 篇，相比之下高兴和轻松就少些，其中轻松最少。

（3）对一些数目比较多的语气进一步做程度的区分，即使得每种语气既包含程度为 2 的，也包含程度为 1 的，程度为 1 的占少数；

最终每种语气的篇章数目为：

表 2

语气	轻松		严肃		高兴		悲伤	赞扬	
程度	2	1	-2	-1	2	1	--	2	1
数目	1	4	3	1	4	4	0	5	3
语气	贬斥		严肃+ 贬斥		严肃+ 悲伤		赞扬+ 悲伤	中性	
程度	-2	-1	-2/-2		-2/-1/-2	-1/-1	2/-2		
数目	5	2	7		6	3	6		6

注意，由于语气为轻松的语料共有 5 篇，程度为 2 的只有 1 篇，所以比例为

1:4；悲伤的程度区分在混合语气“严肃+悲伤”中体现；另外就是高兴、赞扬和贬斥三种语气，尽量将比例定为 2 的偏多，1 的偏少，其中高兴是由于程度为 2 的一共有 4 篇，所以比例为 1: 1。在找不同程度的语料时，我们将每种语气的各个程度的语料分别找出来，再分别进行测听，筛选出相应数目的较典型的篇章。

（4）分类测听：对每种语气的语料先进行分别测听，感受程度的区分是否明显，不明显的进行调整；

（5）整体测听：再对筛选出的所有 60 篇语料进行重新测听，对标注不恰当的语料进行修改或更换；

（6）统计修改后的每种语气的语料数目，看比例是否有很大变动。

统计结果如下：

表 3

语气	轻松		严肃		高兴		悲伤	赞扬	
程度	2	1	-2	-1	2	1	--	2	1
数目	1	2	3	1	3	5	0	4	4
语气	贬斥		严肃+ 贬斥		严肃+ 悲伤		赞扬+ 悲伤	中性	
程度	-2	-1	-2/-2	-1/-1	-2/-2	-1/-1	2	1	--
数目	5	2	6	1	6	3	5	1	8

比较表 2 和表 3 可以看出：轻松的语料数目有所减少，原来标为轻松的标成了中性，相应中性的数目有所增加；高兴和赞扬的两种程度的比例发生了变化，程度弱的都有所增加；严肃+贬斥和赞扬+悲伤都有篇章从原来的 2 级程度降为 1 级。总体来说，这次标注后语气的程度都有减弱的趋势，但每种语气的比例浮动不大。

4. 疑问点分析

(1) 语句或段落的基调与篇章主基调矛盾，或者段落间基调矛盾的，如何归纳篇章基调：

根据具体情况，若只是个别句情感发生变化，则篇章主基调不受影响；若是文章内容是先抑后扬或先扬后抑的则根据文章主旨以后面那个基调定篇章主基调；如果看不出前后关联或分不清主次，我们暂时根据正负相抵原则将篇章主基调定为中性。

(2) 中性与轻松两种语气的区分：

我们理解的轻松即节奏轻快，朗读的比较随便，带有不重视的语气，通常用在一些日常生活讯息的播报上，如天气预报、粮油副食品价格变动、演出表演、天文观测等。而通常这类新闻的语气比较平淡或篇幅很短，听感上很容易忽略，所以容易划分到中性上，这也是造成最终含轻松语气的篇章数目偏少的原因之一，对于这个问题，我们考虑将部分这类新闻的语气调为轻松，但程度定为 1。

(3) 口语化表述的语气界定：

一些文章中出现很口语化的表述，如果单纯地按朗读的不重视程度应归为轻松，但根据语境和语义内容，以及朗读时的情感倾向，归为轻松又不恰当，对于这种情况暂定归为中性。

(4) 语气种类的完整性问题：

个别篇章或语句感觉朗读者朗读是带有明显的感情色彩，但划分到本文中提到的的几种语气里都不合适，对此也暂时归为中性。

5.小规模语料库的韵律特征分析

以贬斥和严肃+贬斥两种语气为例：

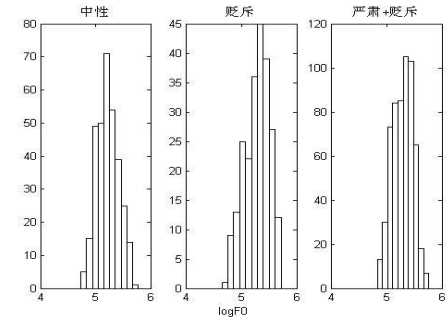


图 2：三种语气韵律词为单位基频均值统计

从均值直方图统计（图 2）可以看出，贬斥和严肃+贬斥的直方图中轴线比中性语气的中轴线偏向较大值，即这两种语气的基频均值比中性语气大，表 4 是三种语气韵律词均值的计算结果，也能说明这点，其中严肃+贬斥的均值最大；还可以看出贬斥和严肃+贬斥两种语气均值较多分布在较低的区间，结合基频曲线，可以看出这两种语气相对中性语气来说均值线下倾更明显；另外，三种语气中，贬斥的音域分布最广，严肃+贬斥的音域最窄，但比较集中在高音段。

表 4

语气	中性	贬斥	严肃+贬斥
均值	5.2234	5.2846	5.2941

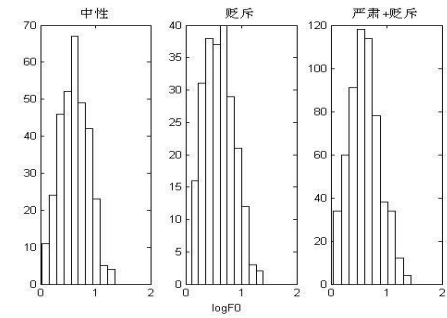


图 3：三种语气韵律词为单位基频音域变化统计

从音域变化直方图（图 3）可以看出，贬斥的音域变化范围较小的数目偏多，相反，严肃+贬斥的音域变化范围较大的数目多，反映出严肃+贬斥这种语气的音域变化更剧烈，即词与词之间的基频变化幅度更大。

综上，我们可以得出贬斥和严肃+贬斥两种语气的一些整体上的韵律特征，即我们可以通过将基频均值线整体向高频搬移，并增加倾斜度来逼近这两种语气，对于贬斥还需要扩大音域范围，而严肃+贬斥则要增大词与词之间的基频变化幅度。

## 6. 总结与展望

本文介绍了一套基于新闻言语数据库的语篇下的语气标注方案，据此分别对大小两个规模的新闻言语数据库进行了语气标注，并对标注过程中存在的一些疑点进行了列举和分析，最后利用小规模语料库对贬斥和严肃+贬斥两种语气进行了韵律特征分析，得出这两种语气在超音段层面相比中性语料的一些整体上的差别。

基于篇章的韵律标注规范及新闻言语数据库的构建，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。我们下一步工作就是基于该语料库，进行语音合成技术的改进，探索不同发音风格下的音色转换的实现。初步设想参考文献[1]中的方案，构建一个三级声学模型，即在语音信号和声学特征参数间增加一个感知层，由一些人类听感产生的形容词构成，作为语音信号和声学参数间的桥梁，然后分别建立语音信号与感知层、感知层与声学特征参数间的联系，希望通过这种建模方案更直观、更准确的描述不同发音风格下的语音信号。

## 7. 致谢

本研究得到国家 863 计划项目（编号：2007AA01Z198）的经费支持。

## 8. 参考文献

- [1] Masato AKAGI, Multi-layer model for expressive speech perception and its application to expressive speech synthesis,

JAIST, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan

- [2] 国辛纯 郭继昌 窦修全（2006）基于 HMM 的语音信号情感识别研究。《电子测量技术》，29 卷第 5 期，69-70,80 页。
- [3] 韩纪庆 邵艳秋（2006）基于语音信号的情感处理研究进展。《电声技术》，第 5 期：58-62,67 页。
- [4] 李晓红 朱维彬 何伟 等（2009）新闻播报言语数据库构建及韵律标注。《第十届全国人机语音通信学术会议论文集》
- [5] 陶建华 许晓颖（2007）面向情感的语音合成系统。中文信息学报。
- [6] 王海波 李爱军（2003）普通话情绪语音库的建立及听辨实验。《第六届全国现代语音学学术会议论文集》。
- [7] 杨鸿武 蔡莲红 蒋丹宁（2005）语音表现力的感知分析研究。《第一届建立和谐人机环境联合学术会议论文集》。
- [8] 张颂（1998）《朗读学》北京：中国传媒大学出版社。
- [9] 朱维彬（2009）语音合成中的语言学计算模型：现状及展望。《当代语言学》。

高莹莹 北京交通大学信息科学研究所 100044

朱维彬 北京交通大学信息科学研究所 100044