

# 关于新闻语料语气标注的初步研究

高莹莹、朱维彬

**摘要:**

**关键词:**

## 1. 引言

新闻播报的语音中,除了传递新闻稿件的字面信息之外,播音员还利用语气传递所代言对象关于新闻内容的感受、态度与立场。从语音学角度,语气是指:通过发音方式的调节,传递特定的立场、态度、感受等副语言学信息,主要由韵律声学特征所负载[1]。从语言学角度,语气是思想感情运动状态支配下的、语言的声音形式,传递了说话人对某一行或事情的看法和态度[10]。张颂[10]认为,语气包含三个要点:其一,语气以内心感情的色彩和分量为灵魂、为神;其二,语气以具体的声音形式为躯体、为形;其三,语气存在于一个个有具体语境的语句当中。简明地阐述了语气的形式与内涵。

关于“语气”的含义及表达方式很多，术语使用没有统一。为方便计，在

本文研究中,我们将“语气”界定为:负载特定情感的言语特征。

通过对新闻语料语气的标注,一方面可以作为转写文字的补充,更完备地记录言语信号所负载的信息;更为重要的,以语气标注为接口,可以就播音员的朗读技巧乃至艺术风格、对应文本的内在语义及总体意图、其语音实体的声学特征等诸多方面与角度,展开关联性、区分性分析,进而推进言语艺术、科学、技术等多方面的研究。为了达到上述目的,首先要形成一套语气标注体系。这套体系除了能够完备、准确地描述语气所负载的特定情感,还应包括其声音形式的特征描述;而且在情感描述与声音特征之间,应存在着显性的、可解释的投射关系。

文章第二部分主要分析了情感描述方面已有的工作；第三部分介绍了本文提出的语气标注体系的具体内容；第四部分是针对此标注体系的心理感知验证实验，以及由发音描述刻画语气声音特征的听觉感知实验；最后是实验结果分析和总结。

## 2. 言语的情感描述

关于情感的描述,很多学者从心理学或行为学等角度给出各种表示方法。目前,情感的表示方法主要分为离散表示和维度表示两种。离散表示法中,Plutchik[5]等人提出的“情感轮(emotion wheel)”分类方法,提出了包括高兴、期望、愤怒、厌恶、悲伤、惊奇、恐惧、赞同在内的八种情感;Descarte 将情感分为基本类和扩展类两种,

基本类对情感的描述起到重要作用，针对情感基本类的常见定义有：害怕、发怒、高兴、悲伤、惊奇、厌恶等六种[7]。连续维度表示方法中，受到广泛认可的是 VAD（Valence, Activation and Dominance）三维模型，即评价维、激发维和强度维，评价维表示情感的正负属性，激发维表示生理的激活水平，强度维表示对他人或情境的控制[2]。三维的名称在他人的研究中会稍有不同，但意思基本一致均表示情感的正负、激活和力度三个基本属性，如 Mehrabian[4]的 PAD（Pleasure, Arousal and Dominance）模型将评价维命名为 Pleasure，激发维命名为 Arousal，国内的学者一般将其翻译为愉悦度、激发度和优势度。另外 Valence 还可以被写作 Appraisal 或 Evaluation；Activation 可以写为 Excitation；Dominance 也可以表示为 Power。三维中起主要作用的是评价维和激发维，第三维主要用于区分生气和恐惧[3]。Cowie[6]的“Feeltrace”即是只应用了评价维和激发维的二维子空间，并增加了圆圈大小及颜色的变化以进行情感的实时捕捉。根据 Wichmann 的说法，激发维在表示内容上倾向于情绪，评价维倾向于态度[8]。

新闻播报中的情感是播音员在特定的立场条件下，由播音内容所触发的情绪感受和态度倾向。心理学认为，情绪是以生理唤起、主观感觉、面部表情和适应行为为特征的状态；感受被定义为个体对某种情绪的主观体验；态度是人们在自身道德观和价值观基础上对事物的评价和行为倾向[9]。顺应这种思路，结合新闻播报中情感的产生特点，我们提出新的情感描述方案：言语中的情感包含发音人当前所处的情绪状态、心理感受和态度倾向三方面内容。其中，情绪状态表示发音时生理状态的激活水平和主观控制；心理感受表示由发音内容触发的某种主观体验；态度倾向表明发音人对发音内容的评价和行为倾向。

### 3.语气描述体系

本文提出的语气描述体系，除了包含对言语中所蕴含情感的刻画，还包括其声音形

式的特征描述。我们希望通过这两部分的描述，找出两者之间显性的投射关系。图 1 给出了该体系的整体架构，下面将就每部分的具体内容及由来给出进一步说明：

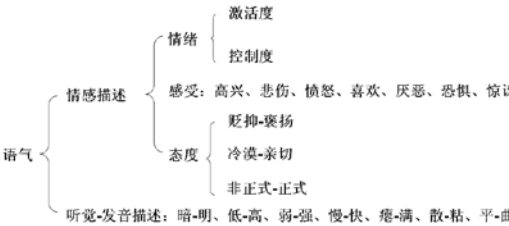


图 1：语气描述体系结构图。

### 3.1 情感描述

我们从发音人当前所处的情绪状态、心理感受和态度倾向三方面描述言语情感。

情绪状态，表示发音时生理状态的激活水平和主观控制。关于这部分描述方案的提出，经历了一个由繁到简的筛选过程。首先，列举出一系列常见的不带明显感情色彩、不带明显态度倾向的情绪词（如：兴奋、激动、活跃、放纵、紧张、低沉、抑制、松弛、轻松、懈怠等）；然后，通过列举其近反义词对这些词进行归类；最后，对每个类别的特征进行归纳，发现这些情绪词主要受两种主要因素控制，即激活度和控制度，激活度体现生理状态的激活水平，控制度体现主体对这种激活的主观控制。最终情绪状态部分就由这两个维度刻画。

心理感受，表示由发音内容触发的某种主观体验。这部分的描述选择借鉴情感的离散表示方法，最常用的基本感受是高兴、悲伤、愤怒、恐惧、惊讶、厌恶六种，其中，正向的感受只有高兴一种，我们认为这还不足以全面刻画感受的所有成分，于是根据厌恶的对立面，增加了喜欢作为第七种基本感受，通过它们程度上的变化及组合作用，可以衍生出其他更多的感受。

态度倾向，表示发音人的对发音内容的评价和行为倾向。这部分的提出过程与情绪状态类似，也是首先列举出一系列常见的态度类型；然后发现态度大多可以成对出现，因此将它们归成态度词语对；再对其进行相

关性分析和归并操作；最终提炼出三组最不相关、表意最独立的态度对，将它们作为三个基本维度构成一个立体空间来表征态度倾向。

每个维度或类型均分为“无、弱、中、强”四个级别，分别用 0、1、2、3 表示。

3.2 听觉-发音描述

为避免声学参数解释性较弱的问题，关于语气的声音形式特征刻画，我们采用了基于感知的发音描述，包括以下几个方面：

- 暗-明：声音的明亮程度
- 低-高：声音的高低
- 弱-强：声音的强度
- 慢-快：语速
- 瘪-满：发音是否到位，音节饱满程度
- 散-粘：音节间的离散或粘连程度
- 平-曲：语调的变化是否丰富

4. 实验

4.1 心理感知实验

为了验证上述标注体系对言语情感的刻画能力，分别为每个部分选取了一组测试词语进行心理感知实验。

测试人员：40 位在校大学生。

4.1.1 情绪状态

被试人员根据对词语意思的理解，对每个待测试词语的激活度和控制度按照 0、1、2、3 四个程度打分。

测试词语：平和、激动、紧张、焦躁、烦闷、抑制、轻松、放纵、活跃、郁闷、压抑、失落、慌乱、惊慌失措、镇定。

4.1.2 心理感受

被试人员根据各自的理解，选出测试词语包含的基本感受成分，并给出程度上（0、1、2、3）的打分。

测试词语：满意、惊愕、寂寞、欢愉、羞愧、悲哀、自豪、妒忌、憎恶、愤恨、惊恐、懊悔、忧愁。

4.1.3 态度倾向

被试人员将根据各自的理解，分别给态度空间的三个基本维度打分（-3、-2、-1、0、1、2、3）。

测试词语：尊敬、轻蔑、冷峻、坚决、犹豫、无礼、傲慢、恭敬、谦卑、和气、庄重、戏谑、夸奖、贬斥、指示、乞求、鼓励、压制、警告、恐吓。

4.2 听觉感知实验

标注对象：稿件为中央人民广播电台“全国新闻和报纸摘要”中播报的新闻通稿，从中筛选出 60 段情感丰富的语料；发音人为广播电台专业播音员，女性；采样率为 22050Hz，精度 16bit，单声道。

标注人员：两位具有专业背景的研究人员和 20 名具有正常听力的在校大学生

标注内容：上述小规模新闻语料库的情感描述和听觉-发音描述

标注方法是先由两位研究人员对语料的情感描述进行标注，标注分为标注训练、正式标注和结果商定几个步骤：通过标注训练阶段的独立标注-共同讨论-再独立标注的几次反复过程，商定出比较一致、稳定的标注准则；再进行独立的大规模正式标注，经过前一阶段的训练，正式标注的一致率已经达到 90%以上；最后针对不一致的结果进行的共同讨论，敲定最后的标注结果。

然后由 20 名被试人员根据听到的声音对这些语料的发音描述特征进行评价，由于所招募被试人员都没有专业背景，所以仅需选出每一项的特征类型即可，如“暗-明”项，测试人员只需选出声音听上去是明亮还是暗沉即可。对于最终该语料的明暗程度确定则通过统计计算得出，以“暗-明”项为例，统计公式如下：

$$d_i(D/B) = \frac{\sum_{n=0}^N B_i}{\sum_{n=0}^M D_i + \sum_{n=0}^N B_i} \tag{1}$$

分子表示选“明”的人数，分母表示选“明”和选“暗”的总人数和，这里去除了因为不确定而没有做出选择的情况，只对

做出确定选择的人数进行统计。可以看出该计算值越接近 1，该语料听起来越明亮，反之越暗沉。

5. 结果分析

5.1 心理感知实验结果分析

5.1.1 情绪状态

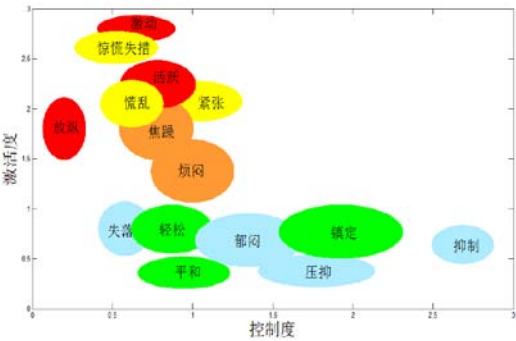


图 2：情绪状态测试词在情绪描述空间的分布。

由情绪状态的测试词语在情绪描述空间的分布（图 2）可以看出：

（1）平和的激活度和控制度均较低，激活度越高越偏向激动，控制度越高越偏向抑制；

（2）激活度较高、控制度较低情况下是放纵、活跃，激活度较低、控制度较高情况下是郁闷、压抑，失落相比这两种情绪控制度较低；

（3）烦闷-焦躁-慌乱-惊慌失措是一组激活度逐渐递增，控制度逐渐递减的情绪，紧张与慌乱的激活度相当，控制度比慌乱高；

（4）镇定是一种激活度略高于平和，控制度又很高的情绪；轻松的激活度也略高于平和，但控制度与平和相当。

综合来看，整个情绪描述空间可以分为激活度高-控制度高、激活度高-控制度低、激活度低-控制度高和激活度低-控制度低四大类，其中，目前这些测试词语在激活度和控制度都很高的区域没有分布，这与人们对测试词语的理解和测试词语的选取都有关

系。通过其他三大类以及类别内部程度的差别，可以对测试词语进行刻画和区分，从而验证了这个描述空间对常见情绪状态的刻画能力。

5.1.2 心理感受

表 1：心理感受测试词的基本组成成分（程度±均方差）

	高兴	悲伤	恐惧	惊讶	愤怒	厌恶	喜爱
满意	2.7±0.5	--	--	--	--	--	2.6±0.5
惊愕	--	--	1.2±1	2.7±0.5	0.7±0.8	0.6±0.8	--
寂寞	--	2.2±0.7	0.6±0.7	--	--	1.2±1	--
欢愉	2.7±0.5	--	--	--	--	--	2.5±0.7
羞愧	--	1.1±0.7	--	--	0.7±0.8	1.5±0.9	--
悲哀	--	2.8±0.4	--	--	--	1.2±1	--
自豪	2.6±0.5	--	--	--	--	--	2.4±0.7
妒忌	--	0.7±0.7	--	--	1.3±0.9	2.3±0.7	--
憎恶	--	--	--	--	1.9±1	2.7±0.6	--
愤恨	--	0.9±1.1	--	--	2.7±0.5	2.6±0.6	--
惊恐	--	--	2.7±0.5	2.4±0.7	--	0.9±0.8	--
懊悔	--	2.2±0.7	--	--	--	1.3±1.1	--
忧愁	--	2.3±0.7	--	--	--	0.8±0.8	--

由心理感受测试词的基本组成成分及其程度（表 1）可以看出：

（1）愤恨、憎恶和妒忌的主要成分都是愤怒和厌恶，只是愤怒的程度不同，妒忌的愤怒程度最低，愤恨的愤怒程度最高；

（2）懊悔、忧愁、寂寞、羞愧、悲哀的主要成分都是悲伤和厌恶，其中悲哀的悲伤成分最重，忧愁的厌恶成分最轻，羞愧的悲伤成分最轻；

（3）满意、欢愉和自豪的主要成分都是高兴和喜爱，且程度相当，分析可能由于被试对词语理解的偏差，导致这三种情感没有区分出来；

（4）惊恐和惊愕的主要成分都是惊讶和恐惧，惊恐的恐惧成分更高。

需要注意一点，为了找到每种情感的主要成分，相对程度较弱的成分我们认为这是由于个体理解差异所致，暂且不予以讨论。

另外，悲伤和厌恶、愤怒和厌恶、高兴和喜爱经常同时出现，组成新的复合感受，但不能因此就认为这些基本感受可以合并，该实验的目的是为了验证这个描述体系对不同心理感受的刻画能力，所以特意选择了一些复合程度较高的测试词语，而没有侧重对

这些基本感受的相关性分析，这里我们默认它们是表意独立的基本感受，单从程度上的变化还会产生新的感受。总体来看，通过这些基本感受的组合和程度变化可以衍生出更多的感受，即验证了这个描述方案对心理感受的刻画能力。

5.1.3 态度倾向

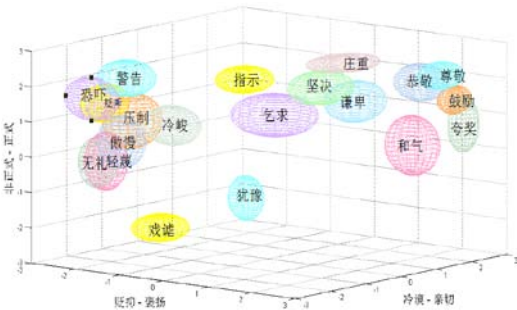


图 3: 态度倾向测试词在态度描述空间的分布。

通过图 3 以及其在不同平面的映射图可以看出，

(1) 分布在不同区域的词语有：

- 犹豫：贬抑、非正式；
- 戏谑：贬抑、非正式、冷漠；
- 乞求：正式、亲切；
- 指示：正式、冷漠；
- 庄重：褒扬、正式；
- 坚决：褒扬、冷漠、正式；
- 和气：褒扬、亲切；

(2) 有些词语分布在相同区域，但通过程度上的差异也可以进行区分：

无礼、轻蔑：均为冷漠、贬抑，但无礼贬抑程度更重；轻蔑冷漠程度更重；

谦卑、尊敬、恭敬、鼓励、夸奖：均分布在褒扬、正式、亲切域，但谦卑的褒扬和亲切程度最低；尊敬较恭敬褒扬成分稍重；鼓励较夸奖稍正式；

警告、恐吓、贬斥、压制、傲慢、冷峻：均分布在贬抑、正式、冷漠域，但傲慢的正式程度最低；冷峻的贬抑程度最低；警告较恐吓更正式；恐吓较警告更冷漠；贬斥较压制更贬抑。

总体上，分布在正式区域的词语偏多，这与词语的选取有一定关系；且多集中在贬抑-冷漠和褒扬-亲切两个区域，一定程度上表明褒扬和亲切、贬抑和冷漠两对态度经常同时出现，但仅从这些测试词语不足以断定这两维可以合并，不能排除有既褒扬又冷漠的态度存在，如本实验中“坚决”即是如此。从上述（1）（2）两方面的分析可以看出，该三维模型对常见的态度倾向词语有较好的刻画能力。

5.2 听觉感知实验结果分析

将情感描述与听觉-发音描述结果进行回归分析，从回归系数中可以看出二者间一些显著的投射关系。初步采用线性分析，公式如下：

$$Y = A * X + C \tag{2}$$

其中 X 为情感描述结果，Y 为听觉-发音描述特征，A 相当于一个转换矩阵，由回归系数组成，C 为常数。

表 2: 情感描述与声音描述线性回归分析结果

	暗-明	低-高	弱-强	慢-快	瘪-满	散-粘	平-曲
激活度	0.07	***0.197	***0.183	0.129	*0.058	0.077	0.094
控制度	-0.09	-0.101	-0.068	*-0.212	-0.029	-0.118	-0.004
高兴	0.07	0.094	0.085	0.129	0.024	*0.129	-0.035
悲伤	** -0.131	** -0.122	-0.052	* -0.142	* -0.045	-0.067	-0.031
恐惧	0.027	0.176	0.19	0.213	0.031	0.17	-0.117
惊讶	-0.016	-0.134	-0.095	-0.176	-0.039	-0.041	0.142
生气	0.138	*0.232	0.126	0.049	0.098	0.21	0.059
厌恶	-0.057	-0.025	-0.03	-0.014	0.03	-0.094	0.029
喜欢	0.037	-0.023	0.072	-0.085	-0.046	0.001	0.052
贬抑-褒扬	0.006	0.015	-0.024	-0.078	0.019	-0.061	0.058
冷漠-亲切	-0.012	0.069	-0.037	0.094	**0.076	0.044	0.055
非正式-正式	-0.026	0.034	0.011	-0.023	0.009	-0.019	-0.031
常数	**0.812	**0.434	***0.495	**0.779	***0.804	**0.558	*0.268
标准差	0.14	0.153	0.152	0.206	0.074	0.18	0.154
模型显著程度	0	0	0.001	0	0	0	0

\*p<0.05; \*\*p<0.01; \*\*\*p<0.001

从回归分析的结果（表 2）可以看出：

(1) 激活度与声音的高低、强弱和音节的饱满程度（即发音是否到位）都有显著关系，激活度越高，声音越高、越强，音节越饱满，其中与满瘪度的相关系数相对较小；

(2) 控制度主要体现在控制语速上，控制度越高，说话速度越慢；

(3) 高兴与粘散度显著相关,越高兴音节间的粘连性越低,即音节越离散;

(4) 悲伤与声音的明暗、高低、快慢以及发音的饱满程度都有关,越悲伤,声音越暗沉,也越低沉,语速越慢,音节越干瘪,但这个相关系数比较小;

(5) 愤怒主要体现在音高的抬升上,越愤怒音高越高;

(6) 态度的冷漠与亲切与发音的饱满程度成一定程度的正相关,但相关系数较小;

(7) 恐惧和惊讶与发音描述间没有出现显著地关系,这主要是因为实验语料中包含恐惧和惊讶的段落过少,其中标有恐惧的段落只有4个,惊讶的有3个,仅用这些数据还不足以发现显著地投射关系;

(8) 厌恶、喜爱、贬义-褒扬、非正式-正式也没有显著关系得出,分析可能这几类情感的感知不仅与声音形式有关,还可能受发音内容、上下文语境等文本信息的影响,因而仅由听觉上加以感知和描述是不够的。

## 6. 总结

本文提出的语气标注体系,同时包含对言语中所蕴含的情感信息的描述,和其声音形式的听觉-发音特征描述。

通过对一些常见情感词语的心理感知实验,验证了该体系能够完备、准确地描述语气所负载的特定情感。一个小规模新闻语料库标注数据线性回归分析表明,语气标注体系中的情感描述、听觉-发音特征描述存在着显性的、可解释的投射关系。这些关系的揭示,对播音员的朗读技巧乃至艺术风格的培养和训练,对语音情感识别、丰富语气语音合成等言语工程应用,具有重要的指导意义。

更多可靠的结论及定量分析,还须更大规模的实验支持。如在当前情况下,仍需考虑的问题有:(1)情感描述体系中各基本项间的相关性分析问题,如七种基本感受间是否完全独立,三个态度维的刻画能力是否均衡等;(2)语料库中各情感类型的分布均衡问题,如最后回归分析中恐惧和惊讶与发音描述间没有出现显著地关系,主要就是

由实验语料中包含恐惧和惊讶的段落过少造成;(3)一些情感类型能否仅用声音就能描述,是否还和发音内容、上下文语境及发音人自身所处立场等文本信息有关,这可能也是造成诸如:厌恶、喜爱、贬义-褒扬、非正式-正式等情感没能得出与听觉发音描述显著关系的原因。

## 7. 致谢

本文研究得到了国家863项目(编号:2007AA01Z198)的经费支持。

## 8. 参考文献

- [1] Fujisaki, H. 1997. Prosody, models, and spontaneous speech. *Computing Prosody New York*, Springer: 27-42.
- [2] Kehrein, R. 2002. The prosody of authentic emotions. *Speech Prosody*: 423-426.
- [3] M Grimm, K. K., E Mower, et al. 2007. Primitives-based evaluation and estimation of emotions in speech. *Speech Communication Elsevier*. **49**: 787-800.
- [4] Mehrabian A., 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology* **144**: 261.
- [5] Plutchik, R. 1984. Emotions: A general psychoevolutionary theory. *Emotion Theory research and experience* **1**: 197-219.
- [6] R Cowie, E. D.-C. 2000. 'FEELTRACE': An instrument for recording perceived emotion in real time. *ISCA, Citeseer*: 19-24.
- [7] Schröder M, C. R., Douglas-Cowie E, et al. 2001. Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis. *Eurospeech, Citeseer*. **1**: 1-4.
- [8] Wichmann, A. 2000. The attitudinal effects of prosody and how they relate to emotion. *ISCA Workshop on Speech and Emotion*. Newcastle, North Ireland: 143-148.
- [9] D Coon, J. M. 郑. 2004. 《心理学导论——思想与行为的认识之路》. 北京, 中国轻工业出版社.
- [10] 张颂. 1999. 《朗读学》. 北京, 中国传媒大学出版社: 226-228.

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