

言语情感描述体系的试验性研究

高莹莹 朱维彬

摘要 基于新闻有稿播音,采用多视角分析方式,探究言语情感生成及其衍化的可能机制。我们选取了与情感联系最为直接的三方面——认知评价、心理感受和生理状态,来构建言语情感描述体系,以期能够更为精准地刻画情感类型,反映情感的生成过程。在实验环节,组建了全部由播音专业高年级学生参与的测试队伍,通过对其心理感知实验数据的聚类、相关计算分析,形成了情感描述体系中的分级结构。

关键词 言语情感,情感产生,情感描述

A STUDY OF A TRANSCRIPTION SYSTEM FOR SPEECH EMOTION

GAO Yingying, ZHU Weibin

Abstract The generation and evolution mechanisms of speech emotion are explored based on a news broadcast corpus using a multi-perspective analysis approach. The most related three perspectives: cognitive evaluation, psychological feeling, and physiological state, are adopted to compose a transcription system, with which the refinement of speech emotion can be described more accurately while the process of speech emotion generation can be revealed clearly. Based on psychological tests, whose subjects are all senior broadcasting students, a hierarchical structure of the system is obtained via clustering and correlation analysis.

Key words Speech emotion, Emotion generation, Emotion transcription

1. 引言

随着言语技术的应用越来越广泛,人们已不再满足于从文字到声音的单纯转换,而是希望合成语音更为逼真,表现力更加丰富,除了字面意思,还能表现字面背后所蕴含的情感信息。

关于言语情感的研究,大多从几个离散的主要(primitive)情感或基本(basic)情感入手[1-6],希望找到这些情感与言语信息或语音信号的映射关系,更为细致的情感类型则可以通过基本情感的衍生得到[7]。这种方法有其合理性,也得到了很多有意义的结果,但是由于方案众多,且各自的研究平台不同,很难将它们统一起来运用于工程。随着研究的深入,人们发现直接寻找某一情感与言语或语音

的对应关系并不容易,即便有,也可能只是管中窥豹,并不完备。近年来兴起的维度学说[8-12]一定程度上弥补了离散观的不足,它通过将某一情感映射到一个多维空间来增加可区分性。这个空间由多个基本维度支撑起来,如PAD的“愉悦维—激发维—控制维”[12]等。维度观的提出开拓了研究视野,同时也为研究结果的统一奠定了基础。但具体到言语中的情感研究,完全的拿来主义并不适用。一方面,它对于说话时情感是怎样产生的并没有一个清晰的解释;另一方面,更细致的言语情感如何刻画仍有待解决;同时,它也并没有明显缩小情感信息与语音信号之间的鸿沟。

研究言语情感,首先需要解决情感的描述或表示问题,接下来才是探寻情感与语音或言语的映射关系问题。本文工作主

要围绕第一个问题进行。关于情感的描述, 因其本身的模糊性、复杂性及跨领域性, 不同的学者对其有不同的阐述和理解 [13-15], Cornelius 综合了前人的观点, 汇总出四大学说: 达尔文学说、詹姆士学说、认知学说和社会生成学说 [16]。我们认为, 它们分别基于四个不同的视角来解释或定义情感, 即生理学视角、心理学视角、认知学视角和社会学视角。几种视角都有其存在的合理性和独特性, 同时它们之间又有所交叠或关联, 因此我们提出了一种综合多种视角的情感描述方案 [18], 以期能够更为精确地刻画情感表达中的细节及其衍化机制。

本文的工作基于广播中的新闻(有稿)播音开展。之所以选择新闻播音作为研究对象, 我们有以下几点考虑: 首先, 就其应用性来讲, 工程应用中机器将文字转换成声音的过程可视为对播音员将稿件播报出来的一种模拟; 其次, 在科学性方面, 新闻播音领域有很多成熟的实践经验或播音理论可供借鉴; 最后, 在可靠性方面, 新闻播音相对于自然语音是一种更加工整更加稳定的语音, 尤其经专业训练后里面的情感表达是可以重复的, 也就是说每种情感表达都有其内在机制; 除此之外, 播音前及播音过程中播音员会收集多方面其他领域信息以求更为准确地理解、表达稿件内容, 这些辅助信息都为我们研究言语中情感的产生因素提供了参考。

文章第二部分将介绍融合了多个视角的基于新闻播音的言语情感生成机制方案; 第三部分在前期工作 [18] 的基础上, 针对其中与情感联系最紧密的几个层面的内容进行了进一步实验探索, 通过对心理测试数据处理分析, 形成了分级结构描述方案; 最后对全文的工作进行总结和展望。

2. 言语情感生成机制

言语技术的发展需要对言语情感的产生因素及其衍化过程有更加详尽且系统化的描述, 由此, 我们提出了一种融合了多视角的言语情感生成机制来解释言语情感

的产生。我们认为, 在言语情感产生过程中, 不同视角或领域间的因素会相互影响, 且呈一种自上而下的层级关系, 不同层级的效用长短和作用大小也不同, 越上层的因素影响可能越深远, 跨度越大, 与底层的关系也越间接。目前该机制主要模仿有稿播音的过程建立, 一共涉及六个不同领域或视角的因素影响: (1) 最上层来自社会文化背景差异对发音人价值观、自身立场、代言角色等社会属性的影响, 是一种最深远、最固有的影响因素; (2) 接下来进入播音的正式创作阶段, 首先是发音人接受稿件内容的刺激, 对稿件的语法、语义及逻辑关系作出语言学分析; (3) 然后评估所要播报内容的整体倾向性, 并根据文体对相应要采取的话语样式等给出预判 [17]; (4) 某些相应的心理感受随之被调动, 有别于心理学中自发的无意识情感, 这种感受是可以受发音人主观控制的表演情感; (5) 最后, 一定程度的生理特征被激活, 并决定是抑制还是推进这种激活, 它的影响是即时性的, 与语音学的关系也更紧密; (6) 至此, 某些特定的情感已经形成, 发音人据此调整发音器官, 产生具有相应感情色彩的语音信号。图1给出这一衍化机制的流程图, 上层因素可以直接作用于发音器官, 也可以通过影响下层因素间接影响发音, 虚线框中的几项被认为与情感产生直接相关, 加入它们将有助于建立语言学和语音学间更为清晰、明显的映射关系。

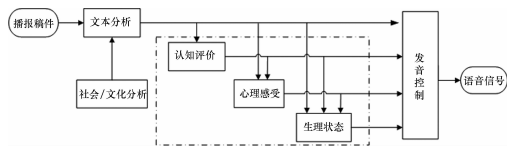


图1 多视角言语情感生成机制流程

3. 情感描述体系

如前所述, 情感的产生受认知评价、心理感受、生理状态三方面因素直接影响, 因此我们将基于这三个层面来构建情感描述体系。其中, 生理状态可通过激活

度(高一低)、控制度(抑制—推进)两个维度来描述,这在我们的前期工作中已得到验证[18],本次试验不作调整,而另外两个层面的内容还有待实验进一步确认与调整。

3.1 心理感受

关于心理感受的研究心理学界有其悠久的历史 and 比较成熟的经验可供借鉴。针对究竟有多少种基本情感或主要情感分别是什么的问题,不同的学者有不同的研究方法,得到的结果却很大程度上存在一致性,如:(1) Ekman 及其同事找到相当数量的证据证明一小部分情感的面部表情具有普遍性[15],即 Cornelius 所谓的“big six”——高兴、悲伤、恐惧、厌恶、愤怒和惊讶[16];(2) Shaver 用层级聚类的办法找到六种跨语言认可度也比较高的情感——喜爱、愉悦、惊讶、愤怒、悲伤和恐惧[13];(3) Laros 和 Steenkamp 汇总了心理学文献中不同学者提出的基本情感[19],如果保留其中五位以上学者都提到的基本情感,则只剩下以下七种——愤怒、厌恶、恐惧、高兴、喜爱、悲伤和惊讶。在最初的研究中[18],我们直接选用了这七种情感作为基本情感,并列举了一些常见的情感词对这些基本情感的刻画能力进行验证,实验发现它们之中有几对存在高度相关性,如高兴和喜爱,恐惧和惊讶,厌恶和愤怒等,这对于这几种基本情感的必要性和独立性问题提出质疑。起初我们希望通过一定规模的情感词聚类的

办法,重新探寻汉语中的基本情感有哪些,但在研究过程中发现类别数目可以随着聚类粗细程度的变化而调整,探寻类别结构的这种层级性变化似乎比只寻找哪些情感最基本更有实用价值,因为即便找到了基本情感有哪些,还要对基本情感和其他情感的衍生关系、成分比例等作进一步分析,这又对完备性、科学性提出了新的挑战;而直接采用层级结构聚类树的办法,主干与枝叶的生成关系或包含关系在聚类之初已经清晰可见,研究者可根据需要自主选择停留在哪一层级上进行研究。

以下是心理感受聚类实验的主要步骤:

● 步骤一: 情感词表汇总

关于情感词的汇总及划分,已有一些成果可供借鉴,如吉大的李轶博士将情感词根据心理感受分为愉快—中性—不愉快三大类,共32个单一小类和3个复合类[20](见表1第二行);北师大的许晓颖教授将现代汉语中基于心理感受的390个情感词划分为23个小类和1个其他类[21](见表一第四行),二人的研究成果在很大程度上存在共通性。这些小类既在一定程度上保证了感受类型的多样性和详尽性,又比直接应用数目庞大的情感词进行分类要省工省力的多,因此我们选择对二人的研究成果进行取并集操作,汇总后共得到44类情感作为即将要聚类的情感词表(其中“信”、“疑”两类情感被认为主要与认知评价相关,所以这里没有采用)。

表1

情感词汇总

李轶											
宁静类	舒适类	满足类	快乐类	激动类	喜爱类	希望类	荣耀类	尊敬类	骄傲类	平静类	清闲类
怜悯类	紧张类	惊奇类	寂寞类	倦怠类	沮丧类	颓废类	耻辱类	羞愧类	悲哀类	烦闷类	焦躁类
怨恨类	忧愁类	轻蔑类	懊悔类	嫉妒类	憎恶类	愤怒类	恐惧类	复合类			
许晓颖、陶建华											
喜乐	爱	愁闷	悲	慌	敬	激动	羞疚	烦	急	傲	吃惊
怒	失望	安心	恨(恶)	嫉	蔑视	悔	委屈	谅	信	疑	其他

汇总

宁静	舒适	体谅	失望	恐惧	愉悦	悲哀	委屈	喜爱	安心	着急	自豪
惊奇	快乐	怨恨	尊敬	吃惊	懊悔	耻辱	希望	嫉妒	羞愧	怜悯	清闲
愁闷	满足	紧张	愤怒	蔑视	激动	沮丧	骄傲	寂寞	忧愁	烦闷	憎恶
颓废	轻蔑	惊慌	平静	焦躁	愧疚	荣耀	倦怠				

●步骤二：测试队伍组建

所有被试者均来自中国传媒大学播音主持专业的高年级学生，由于他们经过几年的专业训练，对词语理解和心理感受的把握要比其他人敏锐而且准确，因此基于他们得到的分类结果用于描述播音过程中的心理感受也更合理。被试规模为 49 人，男女比例是 19 : 30。

●步骤三：人工分类

请被试人员对情感词表进行人工分类，分类规则基于被试者自身对于这些词语所属心理感受类型的理解，为了避免分类过粗或者过细，对分类数目进行限定，最少不少于 4 类，最多不多于 10 类，分类完成后给每一类别进行命名以备我们后续参考。经过筛选最终使用的分类结果是 44 份，男女比例为 17 : 27。

●步骤四：数据处理，算法聚类

得到人工分类的结果后，首先构建一个可以表示数据集中两两情感间相似性的特征矩阵，一般是一个对称方阵，每一行或每一列的数据表示聚类词表中的一种情感，构建方法是：

(1) 初始设定一个 $m \times m$ 的零值方阵， m 为分类词语的总个数，此处为 44；

(2) 在每一份分类结果中，如果词语 A 和 B 被分作一类，则方阵 (i_A, j_B) 和 (i_B, j_A) 位置加 1， i_A, i_B, j_A, j_B 分别表示 A、B 所在行数和列数，同时 A、B 分别与自己也成一类，因此在对角线位置 $(i_A,$

$j_A)$ 和 (i_B, j_B) 也要加 1；

(3) 44 份结果全部统计完之后除以总人数，将数值作归一化处理。

在分类方法的选取上，对 K-means 和层次聚类两种算法的聚类效果作了对比：

K-means 算法

(1) K-means 算法中，初始聚类中心的好坏会影响聚类效果，初始中心的随机设定也会使聚类结果具有随机性，因为本实验中样本集不大，所以这种影响会比较明显，因此为了更合理地设定初始中心，我们使用文献 [22] 中用前 k 大特征值对应的特征向量乘以样本集的办法对其进行近似估计；

(2) 关于聚类个数的问题，K-means 算法需要预先设定聚类个数，因此我们对所感兴趣的 4~10 类都进行了尝试，发现分成 4 类的结果已具有一定合理性，如表 2 所示，但随着聚类个数的增加，有些不希望再拆分的类别越拆越细，如表 2 中第三类在整体聚类个数为 10 的时候就被分成了 3 类，而有些类别的混合度仍然较高。针对这个问题，我们采取了阶段性局部聚类的办法，第一轮全部数据聚成四类，将第三类作为最终结果保留不参与之后的聚类；第二轮其他三类类别内部再各自聚类，聚类个数分别试验了 2、3、4 类，最终得到的聚类结果为八类，即保留表 2 第三类，第一类拆分为 3 类，第二、四类各自拆为 2 类，如表 3 所示。

表 2 K-means 算法聚为 4 类结果

1	失望	悲哀	委屈	沮丧	怜悯	倦怠	忧愁	烦闷	颓废	懊悔	羞愧	愁闷	愧疚	寂寞
2	愉悦	满足	激动	骄傲	荣耀	希望	自豪	快乐	尊敬	喜爱				

续表

3	舒适	体谅	清闲	平静	宁静	安心								
4	恐惧	紧张	愤怒	蔑视	轻蔑	惊慌	吃惊	惊奇	憎恶	耻辱	嫉妒	着急	怨恨	焦躁

表3 K-means 算法分阶段聚类结果

1	舒适	体谅	清闲	平静	宁静	安心		
2	怜悯	懊悔	羞愧	愧疚				
3	悲哀	倦怠	烦闷	寂寞				
4	失望	委屈	沮丧	忧愁	颓废	愁闷		
5	愉悦	荣耀	自豪	快乐				
6	满足	激动	骄傲	希望	尊敬	喜爱		
7	恐惧	紧张	惊慌	吃惊	惊奇	着急	焦躁	
8	愤怒	蔑视	轻蔑	憎恶	耻辱	嫉妒	怨恨	

分析 K-means 的聚类结果, 我们发现一些地方和人们一般的理解不一致, 如“愁闷”和“烦闷”被分成了两类, 而“激动”跟“喜爱”被分成了一类; 另外, 在这个方法的改进过程中, 我们意识到聚类个数是一个很难掌握的问题, 很难界定分成几类最合适、最科学。因此作为对照, 我们又试验了层次聚类的处理算法。我们发现, 利用该算法产生的聚类树可以更直观地展示聚类粗细变化的分级结构, 研究者可以根据需要选择采用不同的聚类级别, 从而可以绕开聚类个数的问题, 而且各级之间的包含关系是显性的, 可方便今后的原理分析及工程应用。

层次聚类算法

采用二叉树聚类策略, 通过对聚类结果的进一步分析、处理得到心理感受的分级结构。具体实施是通过调用 Matlab 工具包中的 4 个聚类分析相关函数实现的。

(1) 根据特征矩阵计算两两变量之间的距离;

调用向量距离函数 `pdist()`。令 $Y = \text{pdist}(A, 'distance')$, A 为特征矩阵, 'distance' 是距离的类型, 有表 4 第一行中几种类型可供选择, Y 为返回的距离矩阵。

(2) 根据距离远近定义变量之间的连

接关系;

调用连接函数 `linkage()`。令 $Z = \text{linkage}(Y)$, Y 为距离矩阵, Z 为连接关系矩阵, 前两列是索引下标列, 用以索引聚类节点, 最后一列表示它们之间的距离。经由 `linkage` 函数计算之后, 实际上二叉树式的聚类已经完成了, 只是 Z 数组可视化不强, 还需下面的步骤生成可视化聚类树。

(3) 可视化聚类树绘制;

调用树图绘制函数 `dendrogram()`。键入 `dendrogram(Z)` 即可生成二叉树的 figure 文件, Z 为上面得到的连接关系矩阵。

(4) 评价聚类信息;

调用相似性分析函数 `cophenetic()`, 用以评价二叉聚类树与实际距离的相符程度。令 $c = \text{cophenetic}(Z, Y)$, c 为相关系数。表 4 是 (1) 中各类型的距离与聚类树的相关性评价对比, 可以看出相关性最高的是 hamming 距离, 其次是标准化欧式距离 `seuclidean`, 因为 hamming 距离统计的是两长度相等的字符串在相同位置上不同字符的个数, 而我们不仅要关注不同字符的个数, 还要关注这种差异的值的的大小, 因此最终选用标准化欧式距离, 即欧式距离用标准差归一化之后的值。

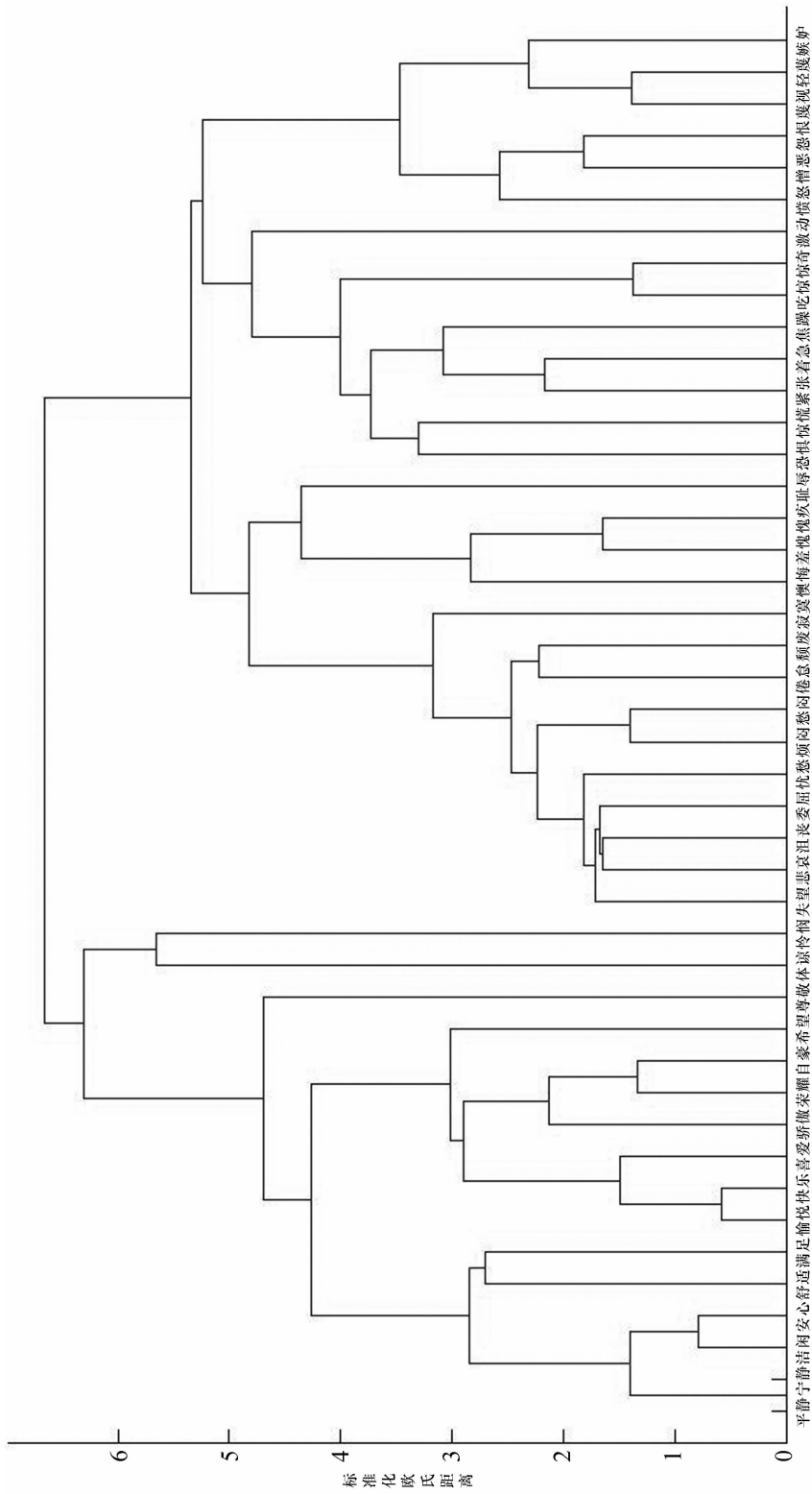


图2 心理感受层次聚类二叉树

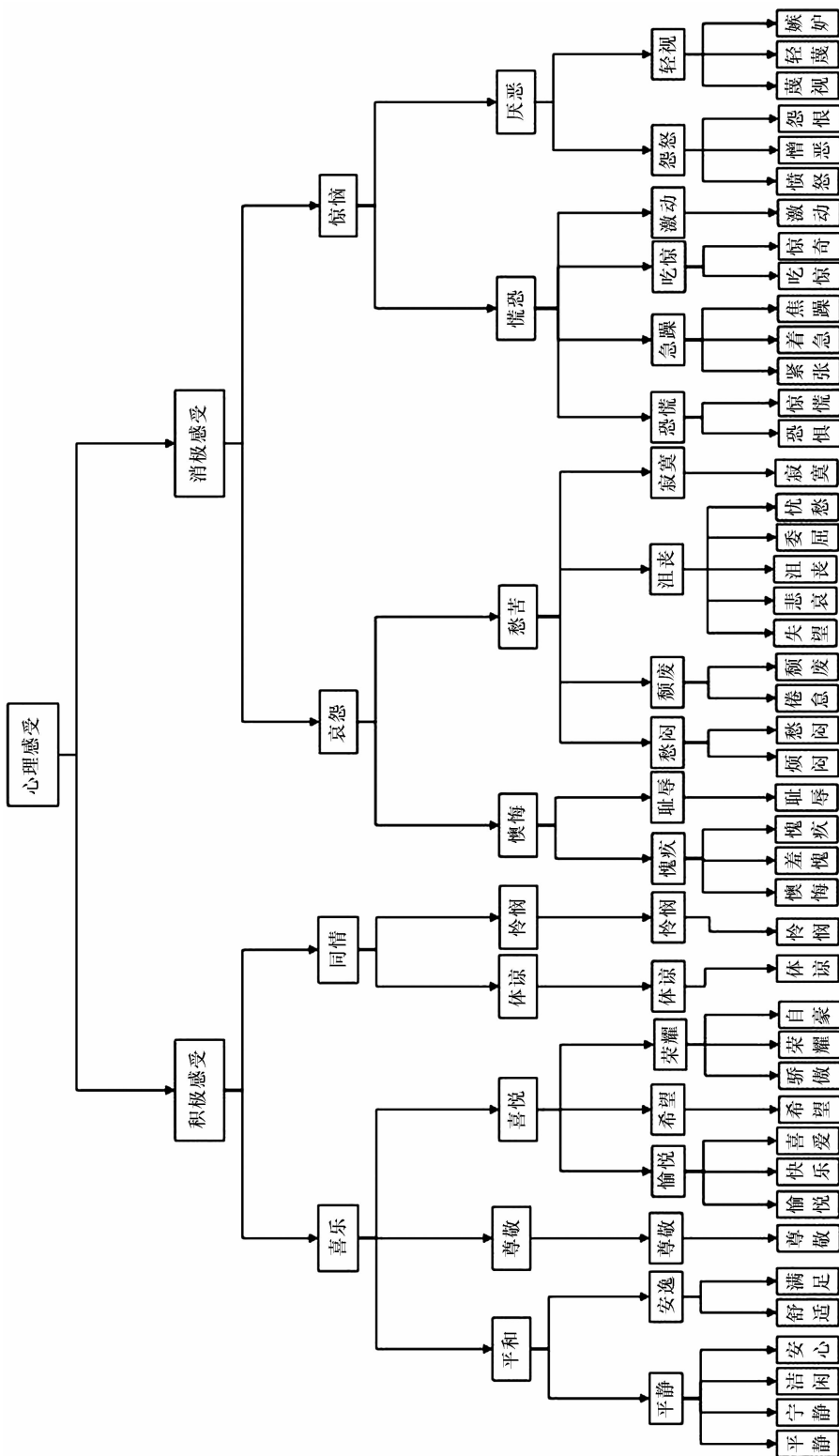


图3 心理感受聚类层级结构拓扑

表 4

各类型距离与聚类树的相关性评价

距离	euclidean	seuclidean	cityblock	minkowski	cosine	correlation	hamming	jaccard	chebychev
相关系数	0.8547	0.8898	0.8631	0.8547	0.8652	0.8374	0.8950	0.7671	0.8496

图 2 即层次聚类算法生成的二叉聚类树, 横轴即 44 种不同情感, 纵轴是它们之间的标准化欧式距离。从图中可以看出, 根据距离尺度由大到小的变化, 类别呈一种由粗到细的层级分布, 最底层划分最细致, 同一类别内的成员关系也最紧密, 越往上成员间距离越大, 上层将下层的类别根据距离的远近进行合并, 最终汇集全部待归类心理感受。为了方便观察这种层级性的结构变化, 我们将处于同一层级的类别作了距离归一化处理, 即忽略同一层级内部各类别间距离的数值差异, 只关注层级间的纵向变化, 据此给出层级结构的拓扑图, 如图 3 所示, 并对每一层上的类别名称进行了尝试性归纳。

至此, 我们可以得到心理感受的分级结构。

由图 3 可以看出, 整个心理感受自上而下被分为五个层级, 最底层即汇总出的用来人工归类的 44 个小类; 越往上划分越粗, 最粗被分为积极感受和消极感受两类, 二者叶子成员比例是 16 : 28, 消极感受比例偏大; 第二层分为 4 类, 积极感受、消极感受各自又细分为两类, 分别是喜乐、同情和哀怨、惊恼; 由于第二层的类别复合程度还是较高, 如惊恼类就可能包含惊讶、恼怒、厌恶等多种成分, 因此又对每个类别进行分裂得到第三层, 共分为 9 类, 同情、哀怨和惊恼各一分为二, 喜乐一分为三, 经分裂处理这一层的大部分类别成分已趋于单一化; 第四层在此基础上进一步细分为 20 类, 属于积极感受的 8 类, 消极感受的 12 类, 这一层已基本能保证所有类别成分单一。通常情况下, 第三层即划为 9 类已能保证工程应用的需求。

3.2 认知评价

认知评价是播音主体根据自身所扮演的角色、受众的身份以及播报所要达到的

目的等, 对稿件内容的整体倾向性作出评估, 并根据文体对相应要采取的话语样式等作出预判。在前期工作中 [18], 我们使用一个三维空间: “贬斥—赞扬”, “非正式—正式”和“冷漠—亲切”三维来对认知评价的结果进行刻画; 并利用若干常见的表示态度倾向性的词语在这三维空间的投射分解, 验证该方案的刻画能力。实验结果显示: 三维空间对不同态度有一定的区分刻画能力, 但所设置的三个维度的完备性和独立性仍存在疑问。

为解决上述问题, 我们将扩充刻画的维度, 以解决完备性问题, 并利用态度词汇在各维度分解向量的相关性分析, 测试各维度间的独立性。具体做法如下:

●步骤一: 完善测试空间

首先, 对原有空间的三个维度进行了扩充和修正, 以保证更多更详尽的细节能得到刻画, 本实验待考察维度有五个, (1) 正负倾向性: 否定—肯定; (2) 正式程度: 非正式—正式; (3) 直露程度: 委婉—直接; (4) 话语温度: 冷漠—热情; (5) 话语硬度: 柔和—刚硬。

●步骤二: 测试词语构成

然后, 列举出若干常见的态度词语作为测试词, 每个维度上的词语也包含在其中, 作为测试其相关性的最直接体现。测试词语如下:

反对	非正式	否定	强硬	坚定
肯定	冷漠	命令	批评	乞求
热情				
柔和	随和	随意	委婉	严肃
犹豫	赞扬	正式	支持	直接
				庄重

●步骤三: 测试队伍组建

同上一实验, 来自中国传媒大学播音主持专业的高年级学生 49 人, 男女比例为 19 : 30。

●步骤四: 被试打分

要求被试者对每个测试词语在各维度上打分, 采用五级打分制, 除了正负两极

的区分外, 还给出了强弱两种级别的区分, 每个维度从左至右依次分为“-2”、“-1”、“0”、“1”、“2”五个分值, 以“否定—肯定”维为例, 分别表示“强否定”、“弱否定”、“中性”、“弱肯定”和“强肯定”。

●步骤五: 数据处理及分析

通过对49份打分结果进行求和、取平均以及相关分析, 可以得出这几个维度间的相互关系。图4是相关性分析结果, 可以看出“否定—肯定”与“冷漠—热情”的相关性较高, “非正式—正式”、“委婉—直接”和“柔和—强硬”三者的相关性较高。通过分析, 我们认为:

(1) “否定—肯定”与“冷漠—热情”主要与话语内容的倾向性判断有关, 其中“否定—肯定”又是倾向性的主要决定因素, 因此设定“否定—肯定”为倾向性主导维度, “冷漠—热情”为辅助维度, 加入对话话人所持态度的把握。

(2) “非正式—正式”、“委婉—直接”和“柔和—强硬”主要与话语样式相关, 新闻播音中基于文体和节目形式的不同主要分为三种话语样式: “宣读式”、“播报式”和“谈话式”, 表达技巧上主要体现于语言规整性、咬字力度、气息控制、语流速度、声音形式等特征的不同[18], 其中语言规整性是区分这三种话语样式的主要特征, 它与“非正式—正式”维所要表达的特征不谋而合, 因此设定“非正式—正式”为刻画话语样式的主导维度, “委婉—直接”和“柔和—强硬”为辅助维度, 帮助刻画话语样式的更多细节。

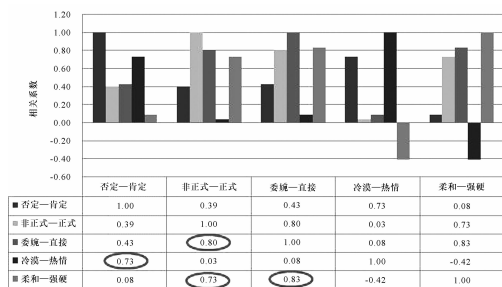


图4 认知评价描述空间几个基本维度的相关性分析结果

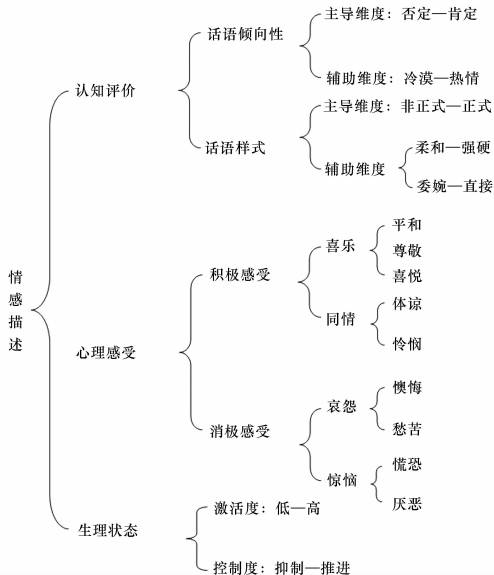


图5 多视角情感描述体系框架图

4. 总结与展望

基于两个实验的结果分析, 多视角情感描述体系的各部分内容得到调整和修正, 形成了一个分级结构的描述方案。最终的描述体系框架如图5所示:

(1) 整个描述体系分别从认知、心理和生理三个视角审视情感, 并采用分级结构以细致刻画言语情感。

(2) 认知评价主要负责对话语整体倾向性和话语样式的评估判断。话语倾向性以“否定—肯定”为倾向性主导维度, “冷漠—热情”为辅助维度。其中“否定—肯定”主导话语内容的正负倾向性评价, “冷漠—热情”辅助刻画说话人态度。话语样式以“非正式—正式”为主导维度, “委婉—直接”和“柔和—强硬”为辅助维度。在新闻有稿播音中, “非正式—正式”主要负责话语样式的语言规整性相关特征描述, 其他两维辅助刻画话语样式的更多细节, 如吐字力度、气息控制等, 至于这些维度与发音特征的具体关系将在下一步工作中进行探索。

(3) 心理感受如何描述才能准确又不失细致是本研究的难点和重点, 这里采用分级结构, 既能满足不同粗细程度的聚类

划分要求, 又可避开基本情感确定以及其他情感衍生等问题。本文将心理感受由粗到细依次分为2类、4类、9类、20类和44类, 上层是对底层的合并。图5只列举到第三层, 是因为该层大部分成员成分已经趋于单一化, 应基本能满足工程应用的需求。

(4) 生理状态仍然由激活度和控制度两个维度刻画, 前者表示个体生理水平的激活程度, 后者表示是抑制还是促进这种激活。

基于这个情感描述体系, 我们接下来会进行文本分析与情感描述、情感描述与发音方式、发音与声学特征等多方面关系的研究和探讨, 希望加入这个多视角多层面的桥梁能缩小语言信息和声学表达间的距离, 并能对情感表达中的更多细节进行刻画。

此外, 基于新闻有稿播音实践及经验的分析, 结合多视角情感描述的思想所形成的言语情感生成机制, 仍属于假设, 其有效性还有待于后期应用的检验, 包括言语工程和播音实践。

5. 致谢

感谢中国传媒大学何伟、邹煜和陈卓老师及播音主持专业2010级同学对我们实验的支持与配合。

参考文献

- [1] 王海波、李爱军:《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会议, 2003年。
- [2] 赵力、蒋春辉、邹采荣、吴镇扬:《语音信号中的情感特征分析和识别的研究》,《电子学报》2004年第4期, 第606—609页。
- [3] 周洁、赵力、邹采荣:《情感语音合成的研究》,《电声技术》2005年第10期, 第57—59页。
- [4] 徐俊、蔡莲红:《面向情感转换的层次化韵律分析与建模》,《清华大学学报》(自然科学版)2009年第49卷第S1期, 第1274—1277页。
- [5] 徐露、徐明星:《面向情感变化检测的汉语情感语音数据库》, 第十届全国人机语音通讯学术会议暨国际语音语言处理研讨会, 2009年。
- [6] 李爱军、邵鹏飞、党建武:《情感表达的跨文化多模态感知研究》,《清华大学学报》(自然科学版)2009年第S1期。
- [7] 韩纪庆、邵艳秋:《基于语音信号的情感处理研究进展》,《电声技术》2006年第5期, 第58—62页。
- [8] Mehrabian, “Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament”, *Current Psychology*, Vol. 14 (4): 261, 1996.
- [9] Douglas-Cowie, E., Cowie, R., “‘FEEL-TRACE’: An instrument for recording perceived emotion in real time”, ISCA, Citeseer: 19—24, 2000.
- [10] Kehrein, R., “The prosody of authentic emotions”, *Speech Prosody*: 423—426, 2002.
- [11] Grimm, M., Mower, E., Kroschel, K., Narayanan, S., “Primitives-based evaluation and estimation of emotions in speech”, *Speech Communication*, Vol. 49: 787—800, 2007.
- [12] 刘焯、傅小兰、陶霖密:《基于PAD三维空间的情感测量》,《中国计算机学会通讯》2010年第6卷第5期, 第9—14页。
- [13] Shaver, P., Schwartz, J., Kirson, D., O'Connor, C., “Emotion knowledge: Further exploration of a prototype approach”, *Journal of Personality and Social Psychology*, Vol. 52: 1061—1086, 1987.
- [14] Scherer, K. R., “Psychological models of emotion”, In: *The neuropsychology of emotion*, ed. by Borod, J., 2000, pp. 137—162.
- [15] Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K. R., Tomita, M., Tzavaras, A., “Universals and cultural differences in the judgments of facial expressions of emotion”, *Journal of Personality and Social Psychology*, Vol. 53 (4): 712—717, 1987.
- [16] Cornelius, R. R., “Theoretical approaches to emotion”, *ISCA Workshop on Speech and Emotion*, Newcastle, North Ireland: 3—10, 2000.
- [17] 张颂:《中国播音学》, 中国传媒大学出版社2003年版。
- [18] Gao, Y. Y., Zhu, W. B., “How to describe speech emotion more completely—An inves-

tigation on Chinese broadcast news speech”, ISCSLP, 2012, Hong Kong.

- [19] Steenkamp, J. E. M., Laros, F. J. M., “Emotions in consumer behavior: A hierarchical approach”, *Journal of Business Research*, Vol. 58(10):1437 - 1445, 2005.
- [20] 李轶: 《论情感类义位的范围及其分类》, 华夏文化论坛, 2010。
- [21] 许晓颖、陶建华: 《汉语情感系统中情感划

分的研究》, 中国情感计算及智能交互学术会议, 2003年, 第199—205页。

- [22] 钱线、黄莹菁、吴立德: 《初始化 K-means 的谱方法》, 《自动化学报》2007年第33卷第4期, 第342—346页。

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