

# 司法实践中情绪语音的话者识别初探

张子扬

上海师范大学语言研究所

DOI:10.32629/er.v3i9.3160

**[摘要]** 在实际的司法话者识别的过程中,有时候拿到的语音材料往往携带有说话人的情绪状态,这就需要我们情感语音中识别出说话人的情绪。在本文中,以元音/ei/作为词表,让被试者用生气、开心、平静三种情绪发音。本文利用Voice sauce软件对19名被试者的情感语音的基本声学参数进行提取,最后利用SPSS软件进行统计分析,来探究这些声学参数对话者情绪的区分能力。

**[关键词]** 情感语音; 主成分分析法; 话者情绪区分能力; 特征权重

**中图分类号:** G642 **文献标识码:** A

## 引言

近几年来,语音情感识别已成为国内外信号处理及人工智能多个领域的研究热点,越来越受到研究者的广泛关注。言语表情是人类特有的表达情绪的重要手段,它与面部表情和肢体表情一起,成为辅助言语交际的工具。因此,话语中不仅包含了语义信息,也包含着人们特定的情绪心理特征。目前对情绪的语音信号进行分析时,通常会用到四类参数,第一类:与时间相关;第二类:与强度相关;第三类:与频率相关;第四类:时间-频率-能量的复合参数。但是,发音是一个动态的过程,它的各种参数变化很大,通常只是测量到语言的一些片断,现在研究所得到的情感语音声学参数还相对比较单一。

对于情感语音的研究,主要集中在心理学方面的研究,多进行的是感知方面的研究。心理学家认为,人类进行情绪表达主要通过身体语言、声音和语言3种方法。而且,声音在其中所占的比例为38%,对情绪的表达做出了很大贡献,我们有必要对其进行深入地探讨。

## 1 实验过程

### 1.1 语料来源

本文所使用的语音,是对19个人使用专业的声卡和话筒在录音室中录制,语音格式为WAV格式(采样率44100字节/秒)。

### 1.2 实验目的

该实验的目的是探索同一个人在不同的情绪状态下说出的语音,我们通过Voice sauce所提取出的声学参数,来对情绪语音进行识别,看这些声学参数对话者的情绪区分能力如何。

### 1.3 实验方法

我们对19名发音人进行了不同情绪状态下的录音,主要有三种情绪,分别为:开心、生气、平静。其中有9名男性发音人,10名女性发音人。录音内容主要是元音的孤立词表(ei),每一个音录制三次,录制完成之后,我们对所有的语音进行了整理和重新命名,然后用软件Praat对语音文件进行打标,去除静音段,标注了语音的起始点与结束点。之后将所有的数据导入到Matlab软件中,启用Voicesauce程序,对语音文件进行声学参数的提取。

### 1.4 数据处理

我们使用的Voicesauce所提取出的语音音质特征共有45种声学参数,但是经过数据的整理与筛选,最终我们采用了一共32种声学参数,分别是:A1、A2、A3(共振峰的能量强度)、CPP(倒谱峰)、Energy(能量)、F2K(2000赫兹左右的共振峰)、H1(第一谐波)、H1A1c(第一谐波和第一共振峰的差值)、H1A2c、H1A3c、H1H2c、H2、H2H4c、H2K、H2KH5Kc、H4、H42Kc、H5K、HNR05、HNR15、HNR25、HNR35、

sB1、sB2、sB3、sB4、sF0、sF1、sF2、sF3、sF4、shrF0。然后我们将提取出来的数据选取开始一帧的数值、中间一帧的数值、最后一帧的数值以及归一化的均值。然后利用SPSS 23.0软件对数据进行主成分分析。

## 2 实验结果

通过SPSS 23.0得到对情绪区分的主成分分析结果,主成分1到主成分8的累积贡献值已经达到了81.614%,且特征值均大于1。所以我们选取前9个主成分来解释原有32个特征,得到十分有价值的分析结果。第1主成分的贡献率最高,所以这个主成分在选取的9个主成分中处于举足轻重的地位,它能够帮助我们认识情感能够被区分的主要原因。

在第1主成分中,依绝对值的大小选出对应的特征名称,依次是(1)A2 (2)A3 (3)H2K (4)A1 (5)H2 (6)sF3,显然,若将在第1主成分中,A2这一特征占据了主要的地位,这也是区分情感的最主要特征。其次是A3、H2K等特征。虽然第2第3主成分没有第1主成分显得重要,但是在所有主成分中,也占有较高的地位。在第2主成分中,HNR15、HNR25、HNR35在其中具有较大的意义。所有的声学特征在9个主成分中占有不同的地位。现根据32个声学特征参数在9个主成分中的比重,选取表2中每一行的绝对值的最大值及该最大值所在主成分贡献率的乘积

作为该特征的值。例如: A1是属于第1主成分中的特征, 所以其特征比例为0.7870.20218=0.159。分别计算所有特征的比例数值, 并求和。

如表1所示, 按32个特征权重从高到低排列顺序, 权重最高的特征是A2, 其次是A3, 接着是H2K。之前的研究中, 研究者通过单一特征的分声学特征进行相互比较, 来选取重要的语音特征。本文将特征数据进行量化, 科学地提取影响情感识别的主要特征, 通过对所有语音特征赋予不同的权值, 获得各个特征的影响因子。

### 3 总结

通过此次实验我们发现, A2、A3、H2K这三种参数能够区分出情绪, 换句话说, 这三种参数受人的情绪影响较大, 相反, 特征权数较小的特征如sB4、sB3、sB2等参数, 不会轻易受到人的情绪影响, 相对比较稳定, 很可能在话者识别的过程中有着很重要的作用, 还需进一步研究和探索。

### [参考文献]

[1] Ververidis D, Kotropoulos C. Emotional speech recognition: Resources, features, and methods[J]. Speech communication, 2006, 48(9): 1162-1181.

表1 32个特征的权重

特征	A2	A3	H2K	A1	H2	sF3	H4
比例	0.184	0.174	0.168	0.159	0.132	0.122	0.119
权重	6.85	6.47	6.25	5.91	4.91	4.54	4.43
特征	HNR15	HNR35	H1A2c	HNR05	HNR25	sF4	H5K
比例	0.094	0.094	0.093	0.091	0.090	0.090	0.089
权重	3.49	3.49	3.46	3.38	3.35	3.35	3.31
特征	H1A1c	H1A3c	sF2	SHRF0	sF0	sB1	CPP
比例	0.082	0.082	0.079	0.074	0.071	0.067	0.065
权重	3.05	3.05	2.94	2.75	2.64	2.49	2.42
特征	H2H4c	H1	sF1	H2KH5Kc	H1H2c	H42Kc	Energy
比例	0.062	0.059	0.079	0.052	0.045	0.039	0.038
权重	2.31	2.19	2.12	1.94	1.67	1.45	1.41
特征	F2K	sB3	sB2	sB4			
比例	0.037	0.032	0.027	0.019			
权重	1.38	1.19	1.01	0.71			

2006, 48(9): 1162-1181.

[2] Lee C M, Narayanan S, Pieraccini R. Recognition of negative emotions from the speech signal[C]//IEEE Workshop on Automatic Speech Recognition and Understanding, 2001. ASRU 01. IEEE, 2001: 240-243.

[3] Yang B, Lugger M. Emotion recognition from speech signals using new harmony features[J]. Signal processing, 2010, 90(5): 1415-1423.

1415-1423.

[4] 张梦翰, 李晨雨. 壮语分区的特征选取和权重量化[J]. 广西民族大学学报: 哲学社会科学版, 2013, 35(6): 7-11.

[5] 杨俊杰. 司法话者时别[M]. 北京: 中国人民公安大学出版社, 2009.

### 作者简介:

张子扬(1996--), 女, 汉族, 山西人, 在读硕士研究生, 上海师范大学语言研究所, 研究方向: 实验语音学。