****

音频信号情感识别及受众脑电信号的关联分析 东北电力大学硕士学位论文

**硕士学位论文**

**（学术学位）**

音频信号情感识别及受众脑电信号的关联分析

**THE CORRELATION ANALYSIS OF AUDIO SIGNAL EMOTION RECOGNITION AND THE AUDIENCE'S EEG SIGNAL**

**陈帅旗**

**2018年3月**

中图分类号： 学校代码：10188

UDC： 密级：公开

音频信号情感识别及受众脑电信号的关联分析

|  |  |  |
| --- | --- | --- |
| 硕士研究生 | ： | 陈帅旗 |
| 导 师 | ： | 侯一民 |
| 学位类别 | ： | 学术硕士 |
| 学科专业  或专业学位领域 | ： | 控制科学与工程 |
| 所 在 单 位 | ： | 自动化工程学院 |
| 答 辩 日 期 | ： | 2018年5月 |
| 授予学位单位 | ： | 东北电力大学 |

Classified Index:

U.D.C:

**THE CORRELATION ANALYSIS OF AUDIO SIGNAL EMOTION RECOGNITION AND THE AUDIENCE'S EEG SIGNAL**

|  |  |
| --- | --- |
| **Candidate：** | Chen Shuaiqi |
| **Supervisor：** | Prof. Hou Yimin |
| **Academic Degree Applied for：** | Master of Engineering |
| **Speciality：** | Control science and Engineering |
| **Affiliation：** | School of Automation Engineering |
| **Date of Defence：** | Mar.2018 |
| **Degree-Conferring-Institution：** | Northeast Electric Power University |

摘　　要

关键词：

# **Abstract**

**Key words:**

目　　录

[摘　　要 I](#_Toc508047657)

[**Abstract** II](#_Toc508047658)

[第1章　绪　　论 - 1 -](#_Toc508047659)

[1.1 课题背景及研究的目的和意义 - 1 -](#_Toc508047660)

[1.1.1 声音信号情感识别 - 1 -](#_Toc508047661)

[1.1.2 脑电信号情感识别 - 2 -](#_Toc508047662)

[1.1.3 音频信号引起的脑电信号分析 - 2 -](#_Toc508047663)

[1.2 国内外研究现状及发展趋势 - 3 -](#_Toc508047664)

[1.3 本文的主要研究内容 - 3 -](#_Toc508047665)

[第2章　语音情感识别基本原理 - 4 -](#_Toc508047666)

[第3章 脑电信号识别原理 - 6 -](#_Toc508047667)

[第4章 音频信号引起的脑电信号分析 - 7 -](#_Toc508047668)

[**1方法设计** - 7 -](#_Toc508047669)

[实验步骤如图1所示： - 7 -](#_Toc508047670)

[**2数据分析** - 8 -](#_Toc508047671)

[**3特征分析** - 11 -](#_Toc508047672)

[**4结论** - 14 -](#_Toc508047673)

[第5章　音频信号和脑电信号关联分析 - 16 -](#_Toc508047674)

[第6章　基于LabVIEW的语音和脑电情感识别系统 17](#_Toc508047675)

[结　　论 - 24 -](#_Toc508047676)

[参考文献 - 25 -](#_Toc508047677)

[攻读硕士学位期间发表的论文及其它成果 - 27 -](#_Toc508047678)

[东北电力大学学位论文原创性声明和使用权限 - 28 -](#_Toc508047679)

[《中国优秀博硕士学位论文全文数据库》和《中国学位论文全文数据库》投稿声明 - 29 -](#_Toc508047680)

[致　　谢 - 30 -](#_Toc508047681)

# 第1章　绪　　论

1.1 课题背景及研究的目的和意义

从音频信号中提取特征，用计算机识别说话人的情感状态，这已经是情感识别领域的一大分支。而基于脑电信号(electroencephalograph,EEG)的情绪识别与分析，因为其分辨率高，结果真实可靠引发了广泛的研究。音频，作为自然条件下可容易获得的多媒体材料，其与脑电信号之间的关系引起了学者的广泛关注。在实验条件下，音乐刺激或用来诱发实验对象情绪；或用于研究治疗抑郁或者阿尔茨海默症等神经精神性疾病。

作为人类情绪主要表现形式之一，声音在产生和传播过程中，什么样的信号会给人带来不同的感受？不同文化和背景下的人对于同样的音频信号会产生怎么样的情绪？不同情绪的音频信号会引起人类大脑中脑电信号怎么样的变化？音频信号概念在人脑中是具体的，特定的，还是抽象的？这不仅是唱片公司所关心的问题，也是在生物医学领域人类关心和探寻的问题。因此，本课题旨在探寻声音信号和人类脑电波之间的关系。研究被广泛标注情感的音频信号会引起人脑哪些区域产生哪种波形的变化，当人脑脑电波明显变化时对应播放着哪些声音以及声音特征和人类脑电特征有何种关系。课题采用NCERP-P型号脑电仪采集脑电，电极排列位置按照10～20国际脑电系统安放，双侧乳突平均参考，头皮电极电阻均调至5K欧以下，脑电采集频率 DC～50 Hz，采样率为256 Hz。

1.2 国内外研究现状及发展趋势

以声音信号来研究情感信息，首先要对情感进行有效的分类。语音情感识别属于模式识别范畴。目前为止，典型的模式识别算法都已经被应用其中。比较流行的算法主要有基于概率模型的识别算法和基于判别模型的方法，前者主要包括高斯混合模型(Gaussian Mixture Model,GMM)[1][2]和隐马尔科夫模型(Hidden Markov Model,HMM)[3][4]；后者主要包括支持向量机(Support Vector Machine,SVM)[1][5]和人工神经网络(Artificial Neural Network,ANN)[6][7]。这些识别模型大都单独使用时会有缺陷，例如陷入局部最优值和维数灾难，于是大量的混合模型被提出[1]。

语音情感特征的识别需要提取大量语音情感特征，研究学者在这方面做了大量研究[8]。例如Luengo选取84个特征组成的韵律特征集。riglia等人使用基频和能量相关的最大值、最小值、均值、标准差组成了一个31维的韵律特征集。Sanchez等人将基频、能量、共振峰、谱倾斜(spectral tilt)的90维全局统计特征用于WCGS数据库中沮丧情绪的检测;Schuller等人将过零率、能量、基频、声音质量、谐波噪声比、0~15阶MFCC等特征的5967维相关统计量用于特征的统计值用于电影维度情感的跟踪等[8][9]。为了提高语音识别的准确率和识别效率，对得到的特征向量进行维数约减成为了一个重要的问题。

虽然国内外研究学者对于语音情感识别做了很多的努力，但仍存在一些问题，例如，识别特征选择韵律学特征较多，对其他特征研究不够，不同的学者对于不同特征区分情感的性能得到的结论不一致[8][10]。韵律特征区的情感区分能力有限等[8]。可见，语音情感识别还有很长的一段路要走。

可以应用多种途径识别人类的情感，但是，由于人类的其他特征例如体态特征，面部表情等可以通过故意作伪使得识别的精确度下降。而由于脑电信号，在人思考时便已经产生，不易伪装或者被操纵，故通过脑电信号来识别人类的情感更加真实和便于分析。

对音乐刺激与脑电信号之间的关系，中外学者做了大量的研究并取得了丰硕的研究成果。为检验熟悉音乐和不熟悉音乐对受众大脑的影响，NattapongThammasan提取了DEAP数据库数据的功率谱密度（power spectra density,PSD）以及分形维数（fractal dimension,FD）用于分类，实验证明不论是支持向量机，多层感知机还是C4.5分类器，低熟悉度的音乐都能提高识别的正确率[3]。Yuiko Kumagai等用钢琴声音作为刺激信号，研究自相关函数平均值，分析认为对不熟悉音乐人的反应比熟悉的音乐强。Marcelo Bigliass等利用音乐刺激干扰受众的动作，用来检验同时进行两个任务时大脑控制动作的相关位置[4]。Joel A. Lopata等通过研究有无音乐即兴表演经验的两组受众，对比其聆听、学习后想象音乐的大脑的阿尔法频段强度，证明在进行创造性活动时前额部分的阿尔法频带活动性更强，创造性作为精神状态的一种概念，是可以训练和提高的[5]。PHNEAH SWEE WU 利用被试者最喜爱的音乐和阿尔法波双耳节拍的放松音乐做对比实验。结果表明，放松音乐对心理和生理上的参与者有较好的舒缓效果，放松音乐对参与者有长期的心理和生理影响。David A. Bridwell等通过对比有规律的吉他和弦旋律和随意的吉他旋律引发的脑电活动，证明在4HZ左右的音乐和弦模式与传统的听觉oddball范式有所不同[6]。Arturo Martínez-Rodrigo 等人为了建立短语节奏与脑神经连接之间的关系，设计了一个音乐实验来诱发与节奏有关的音乐诱发刺激。大脑活动通过脑电图（EEG）通过脑-计算机接口进行监测，通过估计每个脑电信道的功率谱值，得到功率方差随频率的分布实验结果显示，在两个古典奏鸣曲的短语节奏变化的θ和α波段的统计差异，一个在偶形式和回旋曲式等。Linshu Zhou等通过音高变化不同表示不同意义的音乐和自然音乐对先天性失乐症的比较，实验比较对12对先天性失乐症和普通人语义一致性任务的脑电图比较，发现失乐症的人对由音高变化引起的音乐含义变化没有N400效应，但对自然音乐的N400效应和控制组一致，结论证明失乐症可以通过其他线索区分自然音乐，但不能通过音高区分。尧德中利用方差检验和t检验根据情绪的二维模型选取了9首音乐作为刺激信号分为三组，进行脑电分析，提取了PSD和不同的节律信号特征，通过PCA降维，进行分类器训练分类，结果表明，相对于其它节律， beta 节律和 gamma 节律下的平均功率信息用于情绪音乐诱发脑电分类时的准确率较高，[支持向量机](http://baike.baidu.com/item/%E6%94%AF%E6%8C%81%E5%90%91%E9%87%8F%E6%9C%BA)(Support Vector Machine,SVM)的识别正确率较好[1]。为了能实时监测听众的情绪状态变化并据此调整音乐播放列表，马勇等通过标记播放某一类型歌曲脑电不同频段的频率谱和提取对应它的梅尔倒频系数来个性化情绪音乐推荐，并在Android平台上进行了初步开发实现[7]。

在根据不同音乐刺激产生脑电信号进行机器学习时，大多数集中提取脑电信号的一个或几个线性特征或非线性动力学特征，为了弥补单一特征识别的不足，广大学者做了不懈的努力。例如李小伟等为研究和预防轻度抑郁，提取了16个电极3个频带的17个特征，共816个特征，并结合CFS降维和机器学习方法提高了识别率，且认为FP1, FP2, F3, O2, T3是与轻度抑郁相关的电极[8]。徐佳琳，左国坤等为了弥补主成分分析对特征之间非线性关系衡量不足，提出一种基于互信息与主成分分析的算法，对2005国际BCI赛数据集，提取联合功率谱估计、连续小波变换、小波包分解、Hjorth参数特征，采用所提出的算法进行特征选择并与主成分分析算法对比，实验结果表明，所提出算法的降维效果更好，以支持向量机为分类器，相同维数的主成分，所得分类正确率更高[9]。

以上的研究深刻且有意义，在上述研究中，大部分学者分析了不同的脑电特征与不同的音乐情绪之间的相互关系，但是由于脑电特征比较繁杂，而以往研究多集中于分析某一类特征，对于主要特征对音乐情绪引起的脑电信号的识别分类贡献大小并没有明确的研究，对提取各个特征的有效性，分辨率的高低研究不足，且对音乐刺激引发的脑电信号特征的分析尚不充分。本文针对上述不足，重点通过研究对比不同情绪音乐引发的脑电特征，分析寻找对于情绪识别的贡献最大的脑电特征，不同情绪音乐主要引起哪些脑区发生明显的变化，这种变化有何明显的脑电特征，这对减少冗余脑电特征，提高分辨率具有重要的意义。

1.3 本文的主要研究内容

基于上文，本课题的研究主要分为两个部分：

1）对声音情感识别的研究

以往的研究中，国内外学者对音频情感识别做了很多的努力，但仍存在一些问题，例如，识别特征选择韵律学特征较多而韵律特征区的情感区分能力有限；对其他特征研究不够；不同的学者对于不同特征区分情感的性能得到的结论不一致等。故本文拟利用音频信号的多种融合特征分类识别音频情感，分析不同的音频特征与音频情感之间的关系。在以上研究的基础上，本文拟建立一个基于LabVIEW的声音情感识别系统，这对计算机了解人类心情，实现人工智能具有重要意义。

2）不同情绪类型音乐引起的听众脑电信号分析

在根据不同音乐刺激产生脑电信号进行机器学习时，大多数学者集中提取脑电信号的一个或几个线性特征或非线性动力学特征；各个特征的有效性，分辨率的高低研究不足；且对音乐刺激引发的脑电信号特征的分析尚不充分。故本课题的研究重点是音乐引发的情感主要引起哪些脑区发生明显的变化，这种变化有何明显的脑电特征，这些脑电脑电特征在表达不同的情绪时有何不同的表现，线性特征和非线性特征哪一个能更好的区分脑电情感。这对减少脑电特征的提取，提高脑电情感分辨率，探究音频信号和脑电之间的关系具有重要的意义。(需补充具体章节介绍)

# 第2章　音频信号和脑电信号情感识别基本原理

2.1 情感定义及划分

情感是人类的一种主观体验，能使人产生消极或者积极的心理反应。目前，有关情感的定义并没有达成一致，但无可置疑的是，情感是人类一种独特的能力，与人的性格，气质等相关。目前，对于情感的划分主要分为两类：离散情感模型和维度情感模型。

2.2.1离散情感模型

离散情感模型把情感定义为单独的，具有明确含义的几种情感。这几种标签被贴有形容词词性的标签，例如，高兴、悲伤等。其中，能跨越不同人类文化，为人类和具有社会性的哺乳动物所共有的情感类别为基本情感，Ekman提出的情感划分较为著名，他把情感划分为6种基本情感，即高兴、悲伤、厌恶、愤怒、恐惧和惊奇。【引用，情感综述，ekman】

2.2.2维度情感模型

早在20世纪中期由Wundt等提出的的环形模型( Circumplex Model )。这个模型的激励(Activation)和诱力(Valence)两个维度是正交，任何情绪可以放置在两维激励和诱力空间，如负价和适度低的激励认为是悲伤。情感基本都分布在这个环上，可以捕捉情感的相似程度。hayer二维情感模型更是广泛应用到音频情绪识别领域[38]。Thayer情感模型是二维情感连续型心理表示模型，其纵坐标表示的是激励维度（Arousal），从“平静的”到“活力的”变化，反应的是主体的情感活跃程度；横坐标表示的是诱力维度(Valence)，从“消极的”到“积极的”变化，反应的是主体的主观感受【引用】。从而将情感划分为具有代表性的四个区域。Thayer的情感模型可以使语音标签和该坐标空间进行相互转化，通过对情感状态语言描述的理解和估计,就可以找到它在情感空间中的映射位置。此外，还有三维的激励-评估-控制空间理论【引用】。

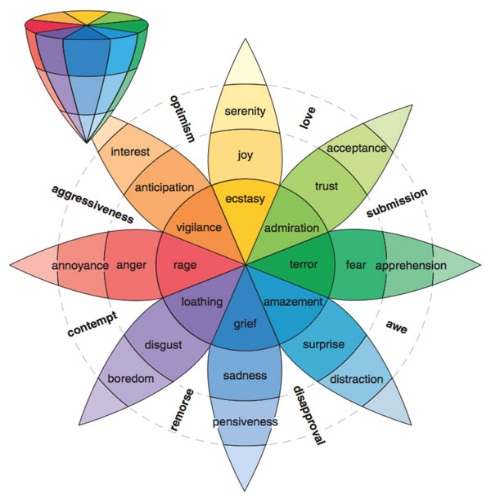
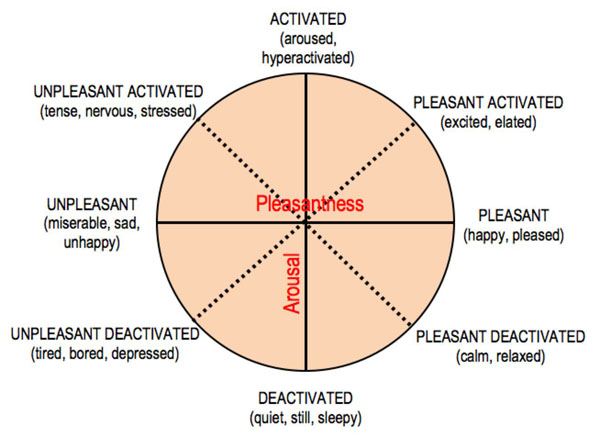


图 二维情感模型（左）三维情感模型（右）

2.2 音频信号情感识别基本原理

对于语音情感特征选择的方式，主要要动态特征和全局统计特征两种，全局统计特征包括韵律特征，谱特征，和其他特征[11]，大部分学者都同意韵律特征能表达较多的情感特征[12]。

短时能量：可用来检测语音信号的端点，主要用来区分浊音段与清音段，声母与韵母的分界，无话段和有话段的分界。

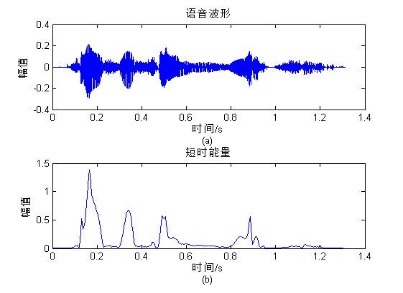


图1 平静状态下“就是下雨也去”语音信号及其短时能量图

基音频率：人在发浊音时声带振动的频率称为基音频率。在不同的情感下，语速，基音均值，基音变化范围等都会不同，例如在愤怒状态下，基音往往在重音出发生突变，在悲伤状态下，基音的趋势是向下弯曲[13]。

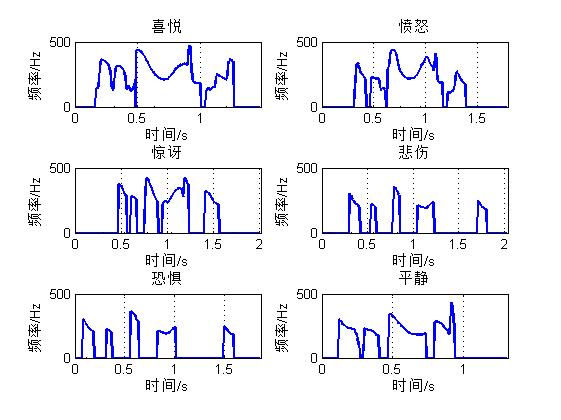


图2 六种情感“就是下雨也去”有音段基频比较

表1基音特征与语音情感关系

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 参数  情感 | 喜 | 怒 | 惊 | 悲 | 恐 | 平静 |
| 基音均值 | 很高 | 较高 | 高 | 高 | 略低 | 高 |
| 基音范围 | 很宽 | 很窄 | 正常 | 略窄 | 略窄 | 正常 |
| 基音变化 | 平滑向上 | 重音突变 | 尾端上翘 | 向下弯曲 | 正常 | 正常 |

共振峰：声音中的元音激励进入声道引起的共振频率叫做共振峰。情感变化的变化反应在第一，第二，第三共振峰[14]。

不同感情的共振峰，最值，峰值和范围都不一样，变化率也不一样。恐惧的第一共振峰变化率最小，喜悦的变化率最大[2]。

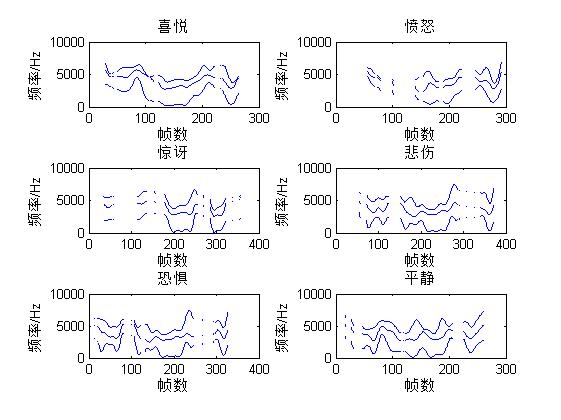


图3 六种情感“就是下雨也去”有音段共振峰比较

长久以来，学者们应用韵律学特征区分情感，取得了广泛的认可[15]。

此外，谱特征，发音帧数特征也得了广泛研究[16][17][18]。

综上，本文选取包括短时能量，基频，发音帧数，共振峰，谱能量特征共64维度的特征向量。

特征1-10：短时能量，及其一阶差分的最大值，最小值，均值，中值，方差；

特征11-26：基频，及其一二阶差分的最大值，最小值，均值，中值，方差，基频范围。

特征27-31：发音帧数，不发音帧数，不发音帧数与发音帧数之比，发音帧数和总帧数之比，发音区域数。

特征32-61：第一，二，三共振峰及一阶差分的最大值，最小值，均值，中值，方差。

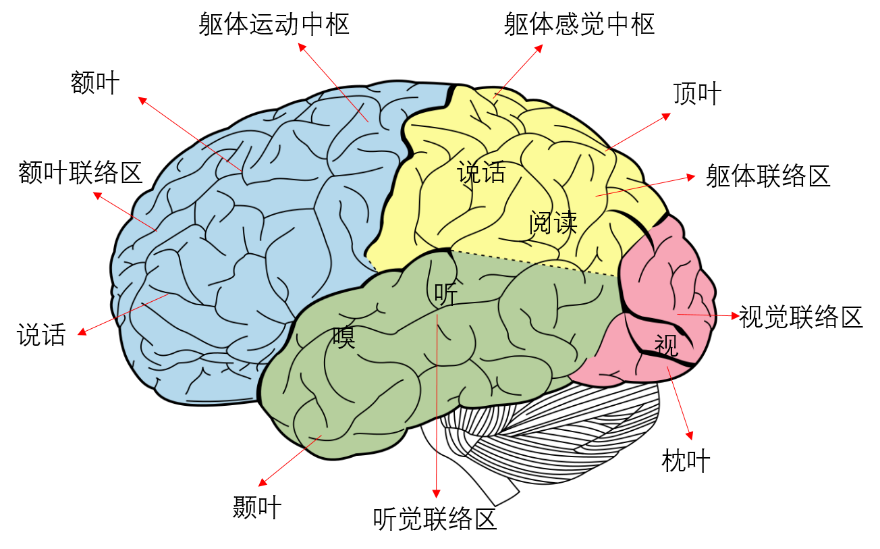
特征62-64：250HZ以下谱能量百分比，650HZ以上谱能量百分比，4K HZ谱能量百分比。

本文采用CASIA汉语情感语料库，CASIA汉语情感语料库由中国科学院自动化研究所录制，由4位录音人(2男2女)在纯净录音环境下(信噪比约为35db)分别在6类不同情感下(高兴、悲哀、生气、惊吓、难过、中性)对500句文本进行的演绎得到[13]。

2.3 脑电情感识别基本原理

2.3.1情感大脑

大脑是人类身体上最重要的器官之一，是人类进行各种活动的指挥中心和司令部。同样，大脑是人类的情感中心，人类的各种的情感全部来源于大脑。人的大脑分为左右两个部分，两个半球只见通过胼胝体连接，许多沟、裂、回之类的结构遍布其上，大脑皮层的表层灰质称之为大脑皮层，大脑的不同分区对应着不同的功能，其具体分布如下图所示。



2.3.2EEG信号特点

EEG信号是用专用的电极帽从头皮采集来的微弱的脑电信号，EEG电极是一种电压传感器，用来记录头皮上的电位变化。【书EEG脑电信号】 大脑的神经系统石油几十亿个神经元组成，在脑电信号传输的过程中，通过钠钾离子的离子通道的开闭传递电信号。脑电信号蕴含着大量人体生理信息的EEG与人类情感有着密切关系。根据EEG频率范围功能的不同，学术界统一将它们划分为5种基本节律：δ节律、θ节律、α节律、β节律和γ节律。不同的节律反映了人不同的生理和心理状态，例如，δ节律的变化可用来检测被试是否处在深度睡眠状态。（1）δ节律：频率在0.1～4Hz之间，振幅约为0μV-200μV。δ节律的出现一般发生在深度睡眠、脑部出现病变或是供氧不足。而正常成年人意识清晰时很少观察到δ节律。（2）θ节律，频率在4～8Hz之间，振幅约为100μV-150μV。θ节律往往出现于在人的中枢神经受抑制时，此时人脑易于接受外界的暗示作用。而意识清醒的正常成年人会呈现出少量且低幅度的此节律。（3）α节律：频率在8～13Hz之间，振幅约为20μV-100μV。α 节律又被分为 α18Hz～10Hz）和 α210Hz～13Hz）。梭状模型是α节律显著外在特征。当人清醒、安静闭目时会有α节律出现；而当人睁眼、思索问题或是有外界干扰时，则无α节律出现。（4）β节律：频率在13～30Hz之间，振幅约为5μV-20μV。β节律同样也被分为β1（13Hz～20Hz）和 2（20Hz～30Hz）。正常人大脑皮层呈现兴奋状态时，脑区会有β节律的伴随出现。因此，β节律可用作检测大脑皮层的兴奋状态。（5）γ节律：频率在31～100Hz 之间。γ波与注意、物体识别及一些条件下的感知绑定有密切的关系。当人进行某些精神活动时，如进行感知行为，γ节律则会呈现增强趋势。除了上述几种主要的节律外，EEG其本身也有着一些其它特点：

（1）微弱性。从上面对几种节律的介绍来看，EEG的振幅非常微弱，很容易就淹没在各种强背景噪声中。因此，需要在信号采集过程中严格控制各种干扰，使影响降到最低。（2）非平稳性。其产生是由人体的不断调节造成的。由于人体始终处在一个不断变化的动态调节过程中，作为反应控制中枢的大脑皮层就要求其自身有很强的随机性，这也是至今我们对EEG规律还没有一个清晰认识以及只能从一些统计学特征对其进行分析的原因。（3

）频域特性突出。此特点对于采用频域或时频域分析来处理EEG尤为重要。（4）多导信号信息相关联。EEG是通过贴在被试头皮表层的多个电极采集而来的多导信号，各导EEG信号间往往隐藏着一些关联信息，找出这些关联信息对信号分析有着极大帮助。

2.3.3EEG信号情感特征

研究表明EEG信号和人类情感之间存在有相关关系。目前表征情感特征的EEG特征包括：

1. 线性特征

Hjorth参数：Hjorth参数作为一种时域分析方法，由Hjorth在１９７０年提出，该分析方法最初是用于不同的在线脑电信号分析系统，例如脑机接口，睡眠分析，以及癫痫发作的偏侧化等。Hjorth法为表征脑电的ｎ阶距，定义了3个时域参数：活动性、移动性Ｗ及复杂性。送Ｈ个参数分别对应着脑电信号在时域上的三个特性：幅度，斜率以及斜率变化率。脑电ｎ阶距的计算公式如下；

脑电的Hjorth参数算法简单，时间复杂度较低，便于自动化和实时实现。

功率谱分析： 主要是EEG的频率谱估计，即把时域的脑电信号变换为脑电的频谱图，直观的展示α 等脑电节律的变化。经典谱估计法，包括自相关法和基于快速傅里叶变换（ＦＦＴ）的周期图法；现代谱估计法，主要是参数模型法；小波变换法。下刚我们简单介绍周期图法及现代谱估计法。

周期图法：若已知一个随机信号的自相关函数那么功率谱密度定义为：

式中，表示数学期望 \*表示复共轭。功率谱密度函数的另一种定义是：

当信号序列是有限长度时，忽略式1求期望和取极限运算，就得到周期图谱估计：

当数据长度足够长时，周期图法的分辨率高，但估计性能较差，方差不会随数据的增长而减小。

AR、MA( moving average)、自回归移动平移( autoregressive moving average,ARMA)及谐波信号模型等都是现代谱估计常用的模型。在EEG信号的谱估计中,使用较多的参数模型是AR模型,其优点是在高信噪比条件下频率分辨率高,特别适用于要求对短数据处理的场合。但是这种方法的缺点在于对被处理信号的线性、平稳性和信噪比要求比较高,而EEG又是非平稳比较突出的信号,因此,使用参数模型估计EEG的功率谱时般要分段处理。由于谱估计分析的是平均谱特征,对检测EEG中的瞬态信号,如癫痫EEG中的棘波和尖波一般无能为力。

1. 非线性特征

# 第3章 音频信号引起的脑电信号分析

# **1方法设计**

# 实验步骤如图1所示：

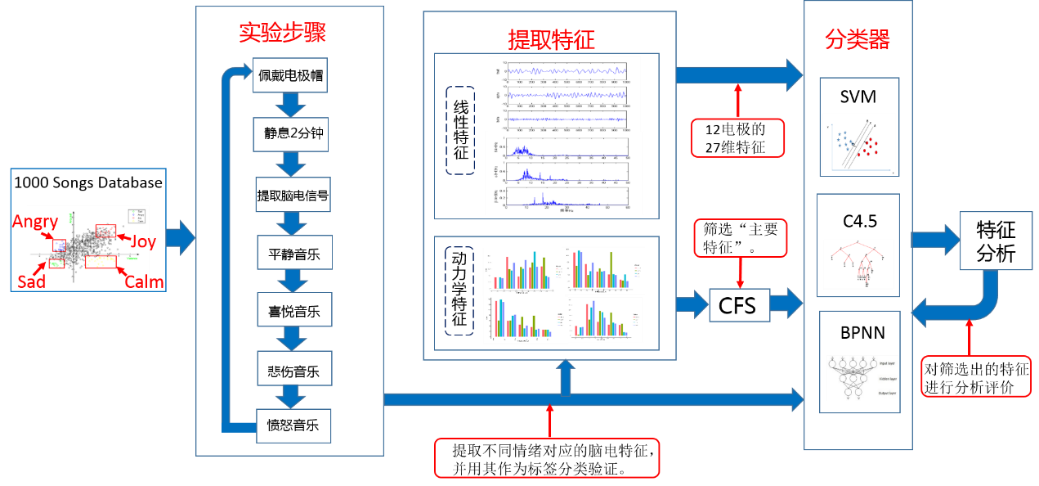


图1 实验流程图

如图所示，实验总体分为五个步骤，即收集实验数据；进行脑电实验；提取脑电特征；分类器分类验证；进行特征分析。

1.1音乐刺激来源

音乐刺激来源于1000 Songs Database，1000 Songs Database是由日内瓦大学的Mohammad Soleymani等人在Free Music Archive网站选择出来的用于情感分析的数据库，数据库内每首歌曲的采样率均为44100HZ，且每首音乐的时间长度为45秒，每首歌曲标注了效价维和激活维均值以及方差，可以用二维情感模型对其进行分类[10]。为了更好的对音乐情感进行分类，本文筛选激活维和效价维均值得分最高且方差比较小的四组，每组22个。如图2所示，组内得分没有显著差异，利用配对t检验，检验组间无关性，如表1所示，sig均为0，小于显著性水平0.05，说明组间数据有显著差异。根据二维情感模型，可以对坐标空间区域和情感状态形容词标签进行映射转化[11]。试验中，本文取效价维和激活维最高的一组定义为高兴；效价维和激活维最低的一组定义为悲伤，效价维最高且激活维最低的一组定义为平静；效价维最低且激活维最高的一组定义为愤怒。

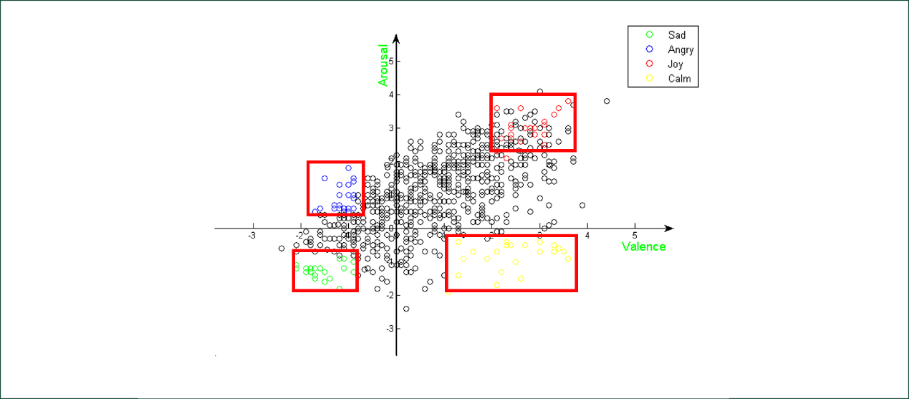


图2音乐分类

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **表1成对样本检验** | | | | | | | | | |
|  | | 成对差分 | | | | | t | df | Sig.(双侧) |
| 均值 | 标准差 | 均值的标准误 | 差分的 95% 置信区间 | |
| 下限 | 上限 |
| 对1 | sad\_arousal - joy\_arousal | -4.26818 | .58504 | .12473 | -4.52757 | -4.00879 | -34.219 | 21 | .000 |
| 对2 | sad\_arousal - angry\_arousal | -3.82727 | .81777 | .17435 | -4.18985 | -3.46469 | -21.952 | 21 | .000 |
| 对3 | clam\_arousal - joy\_arousal | -3.88182 | .50675 | .10804 | -4.10650 | -3.65714 | -35.930 | 21 | .000 |
| 对4 | clam\_arousal - angry\_arousal | -3.44091 | .74748 | .15936 | -3.77232 | -3.10950 | -21.592 | 21 | .000 |
| 对5 | sad\_valence - clam\_valence | -2.18182 | .50675 | .10804 | -2.40650 | -1.95714 | -20.195 | 21 | .000 |
| 对6 | sad\_valence - angry\_valence | -.41364 | .39677 | .08459 | -.58955 | -.23772 | -4.890 | 21 | .000 |
| 对7 | clam\_valence - angry\_valence | 1.76818 | .58748 | .12525 | 1.50771 | 2.02865 | 14.117 | 21 | .000 |
| 对8 | joy\_valence - angry\_valence | 3.86818 | .63650 | .13570 | 3.58597 | 4.15039 | 28.505 | 21 | .000 |

1.2脑电采集步骤

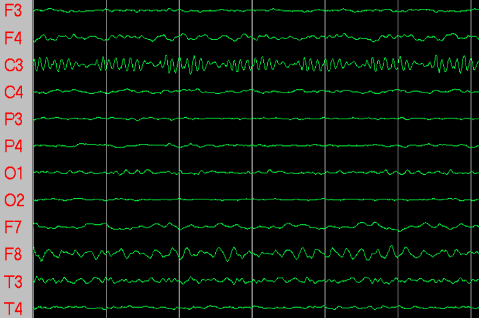
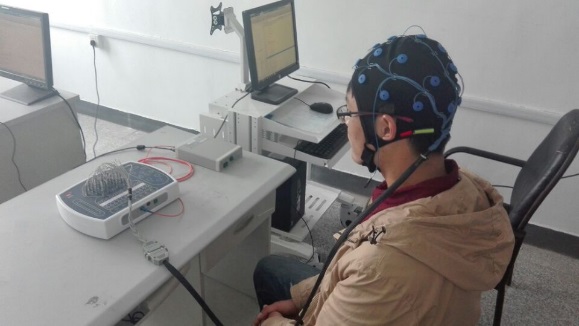


图3脑电采集设备及采集的信号

实验开始前，要求被试受众不能饱食，情绪相对稳定，实验首先要求填写记录表，记录个人信息。实验开始，首先佩带电极帽，闭眼，期间不播放任何音乐，使被试者放松2分钟左右。实验期间，为避免愤怒，焦躁声音引发的情感影响后续试验，音乐播放按照平静，喜悦，悲伤，愤怒的顺序依次进行，在播放音乐的同时记录脑电信号。每首音乐45s，每类感情播放20首音乐。在前一类型的音乐播放完毕后，休息10~15分钟后开始后一类型音乐的播放。最后，为保证实验的准确性，每位被试在第一次实验结束后2-3天，要求重复一次实验。实验拟采集8个人共16组数据，后因某次实验不成功人工排除一组数据，得到共15组数据。8人均为大学生，平均年龄为：23.113.14岁，包括6名男性以及两名女性。

采用NCERP-P型号脑电仪采集脑电，电极排列位置按照10～20国际脑电系统安放，双侧乳突平均参考，头皮电极电阻均调至5K欧以下，脑电采集频率 DC～50 Hz，采样率为256 Hz。

# **2数据分析**

2.1特征提取

在以往的研究中，线性特征常常被认为是与人类情感相关紧密的特征[12~14]。随着非线性动力学的发展，人们逐渐视大脑为一个高维、复杂的混沌系统，非线性动力学特征也广泛应用与脑电研究[15~17]。如图4所示：实验将获得的脑电数据首先经过预处理去噪，自适应滤波滤除50HZ工频信号，经过ICA滤除眼电信号，提取每首歌曲对应的第10~15s脑电信号，最后提取与情感相关的FP1、FP2、F3、F4、F7、F8、Fz、C3、C4、T3、T4、Pz电极，提取相关的18维线性特征，包括thalt，alpha，belta峰值，均值，方差，中心频率，最大功率，功率和特征；以及9维非线性动力学特征，奇异谱熵，LZ复杂度，谱熵，C0复杂度，最大李雅普诺夫指数，样本熵，近似熵，K熵，关联维。

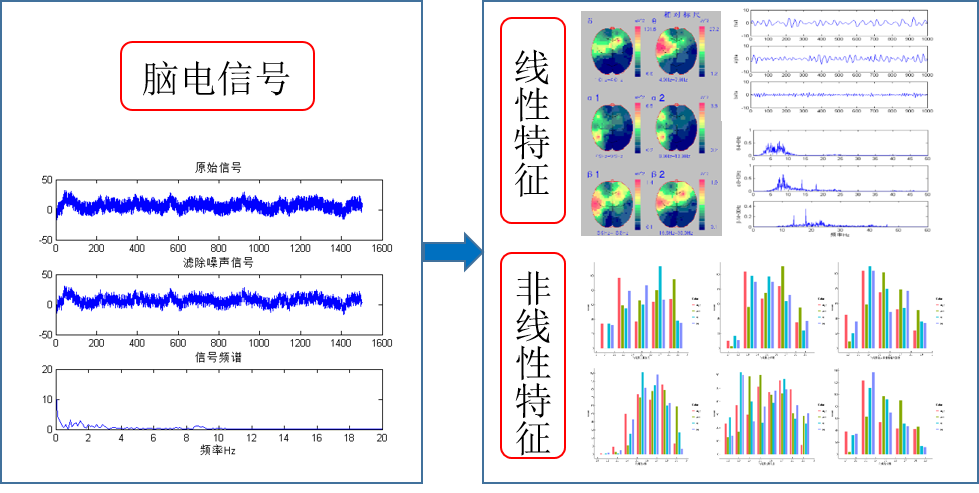


图4脑电信号提取的特征

2.2特征选择

CFS方法综合评价了特征与分类之间的相关性和特征之间的冗余性，广泛用于特征选择，数据清洗，是一种有效的特征选择方法[8,18~19]。它的核心思想是从原始特征集中，采用相关性测度寻找出特征之间的相关性较低但与类标记高度相关的特征构成的特征子集，达到去除冗余属性和与类无关属性的目的[20]。CFS对特征集中的特征与特征、特征与类别之间的相关度进行计算，计算公式如下所示：

（1）

式中的对特征子集的一个评价，其中中含有个特征；表示特征与类别之间的平均相关性；表示特征与特征之间的平均相关度。由式(1)可知，该方法计算得到的特征子集中，每个特征与类标记的关联度越大且特征之间的冗余度越小，则的值越大，当前的特征子集是优良的特征子集。

特征之间相关性的计算可以用信息增益的方法求得。假设属性为Y，y为Y的一个可能的取值，那么Y的熵的计算方法为：

（2）

若已知某一属性X，X条件下计算Y熵的方法为：

（3）

特征X对Y提供的附加信息称为信息增益。信息增益与两特征的相关性成正相关的关系。信息增益定义为：

（4）

由于信息增益是一种对称性的测量方法，需对其进行归一化，其方法如下：

（5）

由于相关系数描述两个变量间的相关的强弱程度，相关系数的取值在0和1之间，当相关系数越接近1，则表示两个变量间的相关度越强；当相关系数越接近于0，表示两个变量的相关性越弱。

贪婪逐步搜索算法用来从特征集中产生备选特征子集。即特征选择产生一个用关联度大小排序的一个特征序列，f1,f2,f3···f27，随后将特征子集（f1）,（f1,f2）,（f1,f2,f3）,（f1,f2,f3,f4），······(f1,f2,f3,f4···f26,f27)进行分类器识别验证。

2.3分类器验证

支持向量机，决策树，神经网络是机器学习中典型用到的分类方法，本文利用SVM，C4.5，BP神经网络验证降维后的识别率。实验采用九折交叉验证的方法，且取运行100次识别正确率的均值作为识别率。最后取CFS取的特征子集识别率最大的值作为最佳识别率。



图5 SVM，C4.5，BP验证降维前后的识别率

由分类器结果来看，CFS选择的特征均大于未降维前分类器的识别率；其中，Pz，T4电极的识别率明显大于其他电极，T3电极经CFS提取后的识别率也比较显著；根据以往的研究，在不同分类器分类验证中，SVM对于情绪分类具有更好的鲁棒性[21]。在本实验中SVM和C4.5都取得了比较好的分类效果。

由统计特征分析，本文对不同电极不同情绪特征之间进行了两两配对t检验，假设两样本无显著差异，取显著性水平p=0.05，对每个电极的每个特征进行t检验，最后统计原假设被拒绝的特征个数。由表2的t检验结果看，T4和Pz电极t检验不同电极间不相关特征数也要多于其他电极。

表2t检验12电极特征不相关个数

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table—t检验12电极特征不相关个数（p>0.05） | | | | | | | | | | | | |
| 电极 | FP1 | FP2 | F3 | F4 | F7 | F8 | Fz | C3 | C4 | T3 | T4 | Pz |
| 怒-静 | 146 | 135 | 139 | 144 | 136 | 112 | 138 | 130 | 125 | 94 | 168 | 152 |
| 怒-喜 | 132 | 116 | 107 | 130 | 109 | 105 | 115 | 126 | 76 | 92 | 157 | 142 |
| 怒-悲 | 99 | 126 | 83 | 99 | 100 | 97 | 100 | 117 | 66 | 67 | 120 | 127 |
| 静-喜 | 126 | 124 | 101 | 116 | 97 | 119 | 121 | 109 | 99 | 80 | 124 | 129 |
| 静-悲 | 161 | 151 | 127 | 138 | 151 | 133 | 137 | 149 | 126 | 107 | 179 | 164 |
| 喜-悲 | 112 | 127 | 94 | 111 | 116 | 110 | 103 | 127 | 89 | 103 | 162 | 145 |
| 总计 | 776 | 779 | 651 | 738 | 709 | 676 | 714 | 758 | 581 | 543 | 910 | 859 |

2.4特征检验

由以上分析可知，CFS选择出的特征比起原始特征更加有效，为研究上述原始特征哪个特征最有效，本文对最优特征子集中选择出的特征进行了分析。由于每个人选择的最优特征子集都是有所不同的，所以需要综合考虑所有被试的最优识别率，于是实验提取了所有受众所有特征子集的识别率，并考证对于大多数实验对象较优的特征子集。由以上分析可知，T4，Pz电极对于脑电情绪的识别比较有效，T3电极经CFS提取后识别正确率较高，故实验只对以上三个电极进行分析，如图6所示。



图6T3，T4，Pz不同特征子集的识别率

由以上不同分类器对不同电极的识别分析，不论是哪一种识别算法，哪一电极，所选的特征子集超过10维时，识别率皆没有明显提升，甚至会出现下降的趋势。故综合来看，CFS选择的前10维特征对于分类器的识别率相对最高。而这10维特征中，如果某一特征被选择的频数越高，则这一特征就越可能越有效。10维特征中包括的不同特征频数此时如图7所示。



图7T3，T4，Pz不同特征的特征频数

如图7所示，CFS选择的特征频数较高的有4（alpha波中心频率）、8（thalt波均值）、10（thalt波中心频率）、14（dalt波均值）、16（dalt波中心频率）、17（dalt波最大功率）、20（K熵）、21（近似熵）、22（最大李雅普诺夫指数）、24（C0复杂度）、25（谱熵）、26（lz复杂度）。选择出的特征包括6维线性特征，和6维非线性特征，这12维特征就是对于与大多数受众来说比较有效的特征组合。

# **3特征分析**

3.1线性特征

在经CFS选择出的线性特征中，Belta波占比较主要的比重，这与以往研究，在情绪的识别中Belta波特征常用来作为实验特征相符。在线性特征中，三个波形频段的中心频率显然也是比较重要的特征之一。这可能说明，不同波形的中心频率可以比较好的代表这个波形。

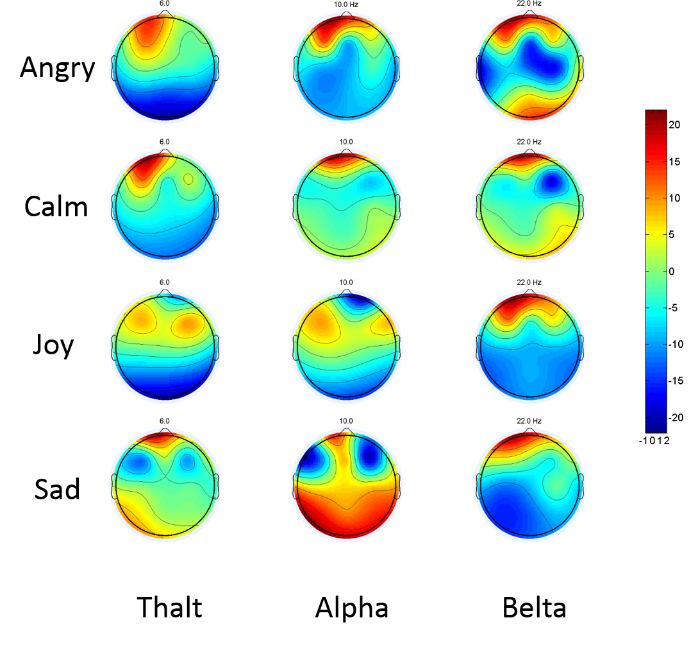


图8不同波段四种感情音乐受众脑地形图

为对线性特征进行分析，本文用EEGlAB对数据进行了分析。步骤如下：首先，利用小波对15组数据进行分频段（Thalt，Alpha，Belta）处理；接着对不同情绪不同频段的数据进行叠加平均；最后导入EEGLAB进行处理并画出脑地形图。实验用了以上所提到的与情绪相关的十二个电极。

如图8所示，受众在听愤怒和平静感情的音乐时，额区能量比较强。在听喜悦感情的音乐时，颞区的Thalt，Alpha波能量比较强，但前额的Belta波比较强。对于悲伤感情的音乐，Alpha波的活动范围比较广泛。从不同频段来看，受众听不同感情的音乐，Alpha频段变化较大。这可能说明，在听不同情感的音乐时，Alpha频段活动较强。

3.2非线性特征

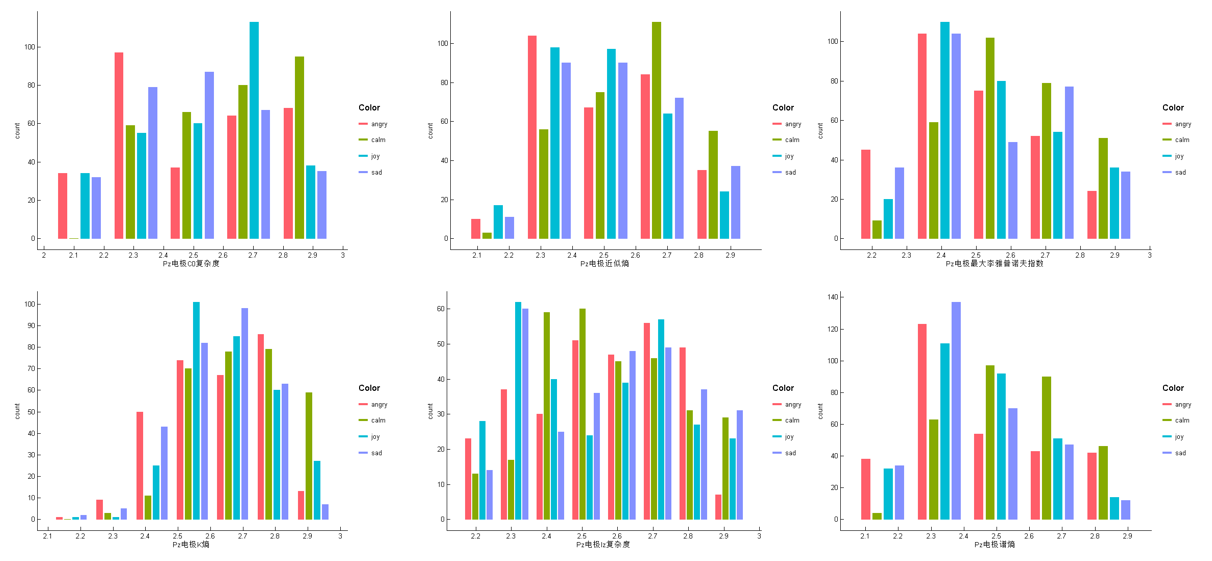
为了对选择的非线性特征进行分析，本文对选择出的特征做频率分布直方图，以直观的分析研究不同的情绪同非线性特征之间的关系。

图9Pz电极非线性特征频数分布

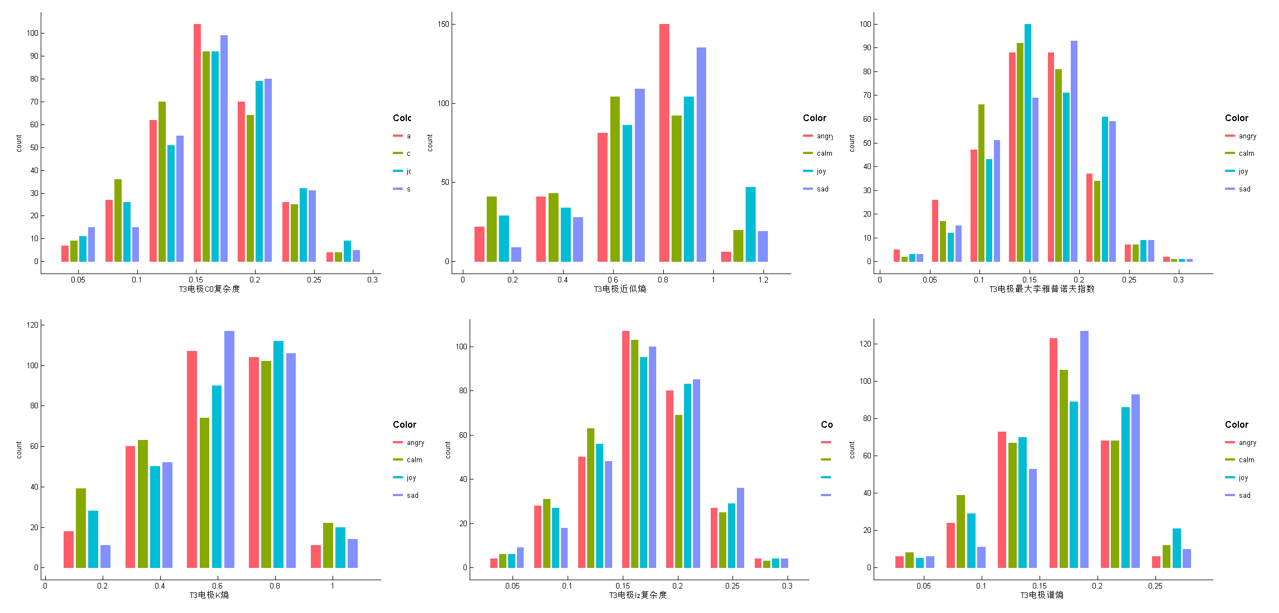
对于Pz电极，其C0复杂度对于二维情绪模型中效价维辨别度比较明显，愤怒和悲伤的效价维得分比较低，在C0复杂度多分布于较低数值段，喜悦和评价效价维得分比较高，在C0复杂度多分布于较高数值段； k熵和LZ复杂度在激活维得分较低情感（平静和悲伤）的数值多集中在较低数值，激活维得分较高情感（愤怒和喜悦）的数值多集中在较高数值；谱熵，近似熵和最大李雅普诺夫指数平静多分布在较高数值。

图10T3电极非线性特征频数分布

对于T3电极，近似熵，k熵，谱熵对愤怒及悲伤的识别率较好，但总体不如Pz和T4电极特征与情感对应明显。

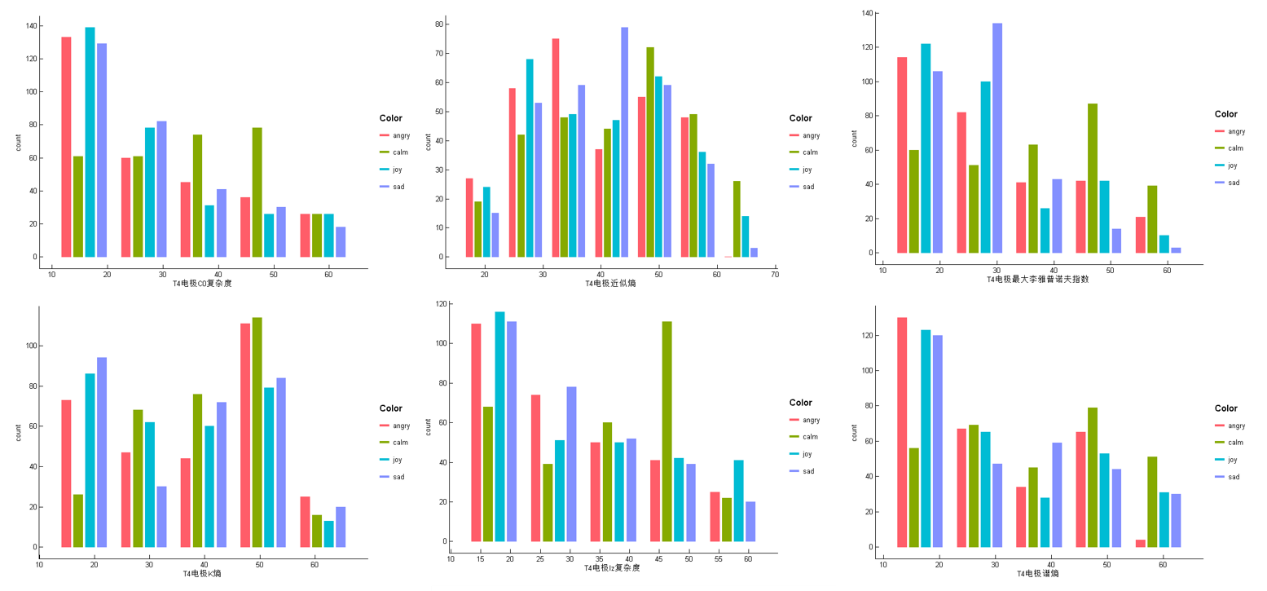


图11T4电极非线性特征频数分布

对于T4电极，其C0复杂度，最大李雅普诺夫指数，LZ复杂度，愤怒，喜悦，悲伤分布范围集中于较小数值段，平静情感的范围集中于较高数值段，这表明这三个特征对平静的分辨度较好；近似熵对于二维情感理论激活维分辨度较好，愤怒和喜悦情感激活维得分较高，分布于近似熵较小数值段，平静和悲伤激活维得分较低，分布于较大数值段；k熵的喜悦和悲伤情感大多分布在较低数值部分，愤怒和平静大多分布在较高数值部分。

3.3检验

为了检验选出的特征是否有效，比较线性特征和非线性特征在识别上有何不同，需要对选择出的特征进行验证，实验通过不同算法比较了以上通过CFS选择出的6维线性特征，6维非线性特征以及选择12维总特征，27维原特征的识别率。

图12分类器验证不同组合的特征子集

实验结果证明，对不同的算法，非线性特征的识别率都要高于线性特征的识别率，综合特征与原27维特征识别率在同等水平，就本实验而言，非线性特征的特征可能更能代表情绪的不同特征，但对于总体识别率而言，选择出的综合特征能代表总体特征。而非线性特征和综合特征与原27维对比，没有明显高于原特征，这可能是因为，选择出的这些特征是根据全部数据对比选出的，具体到每个人时并不是最优的特征子集组合。

# **4结论**

本文针对不同情绪音乐刺激引发的受众脑电信号，分析研究了脑电信号的线性特征和非线性特征，力图寻找出对情绪识别最有效的特征数据集。

实验中，利用80首标注好不同情绪（愤怒，喜悦，悲伤，平静）的音乐刺激，采集了8位被试受众在听不同情绪音乐时的脑电信号，每位被试隔一段时间重复同样的实验。人工排除一组数据后，共计15组脑电数据。利用得到的脑电信号提取12个与情绪有关的典型脑电电极特征，包括线性特征和非线性特征，最后每个电极得到共27维特征。为了找出哪个特征对于脑电情绪的识别比较有效，减少冗余特征，随后利用CFS的方法选择出了与原特征集相关性较大冗余较小的特征集。为检验选择出特征集是否比原特征数据更加有效，本文利用不同的分类器区分对比了原特征集和选择出的最优特征集在100次有监督学习的识别正确率，并用t检验的方法选择出了与音乐刺激最有可能有关联的T3，T4，Pz三个电极，结果表明：

1）对于分类器而言，SVM和C4.5算法对于脑电情绪分类都比较有效。

2）本文利用脑地形图和频率分布直方图对这三个电极线性特征和非线性动力学特征的最优特征集进行了分析，最终得出6维线性特征包括：4（alpha波中心频率）、8（thalt波均值）、10（thalt波中心频率）、14（dalt波均值）、16（dalt波中心频率）、17（dalt波最大功率）；6维非线性特征，包括：20（K熵）、21（近似熵）、22（最大李雅普诺夫指数）、24（C0复杂度）、25（谱熵）、26（lz复杂度）可能对音乐情绪的分类最有代表性。

3）本文最后通过分类器对比验证不同的线性特征和非线性特征组合，结果证明：对于本实验，非线性特征比线性特征能更好的区分不同的情绪特征，选择出的12维特征能较好的代表原27维的特征。

# 第4章　音频信号和脑电信号关联分析

# 第6章　基于LabVIEW的语音和脑电情感识别系统

本文利用LabVIEW和Matlab混合编程搭建了语音情感识别系统，其中特征的提取，训练和识别主要利用Matlab强大的计算能力完成，数据的可视化显示由LabVIEW完成。总体设计步骤如图1所示。首先利用MIRtoolbox工具包提取典型的音频统计特征，通过经过CFS方法改进的GA算法筛选出参与建模的主要特征。为证实筛选出的特征的有效性，再利用BP神经网络（Back Propagation Neural Network， BPNN）和支持向量机（Support Vector Machine， SVM）检验主要特征对于音乐情感的识别作用，实验采用十折交叉验证的方法，统计100次实验的识别率取其平均数作为最后的识别结果。

图1 总体设计流程图

1)特征提取。MIRtoolbox工具包是基于Matlab平台的一种声音特征工具包。Mirtoolbox包括50种音频提取特征和统计特征[11][12]。包括时域特征，波形特征，节律，频谱，音质，声调等。实验提取了音频信号的时域的均方差以及波形，谱特征，音色，语调等特征的均值，方差，斜率，周期频率，周期振幅，周期熵，以及峰值，质心等共138个特征（见表1）。对138个特征进行预处理，删除含有NAN的特征。经处理后的特征用于识别验证。

表1 MIRtoolbox工具包提取的特征

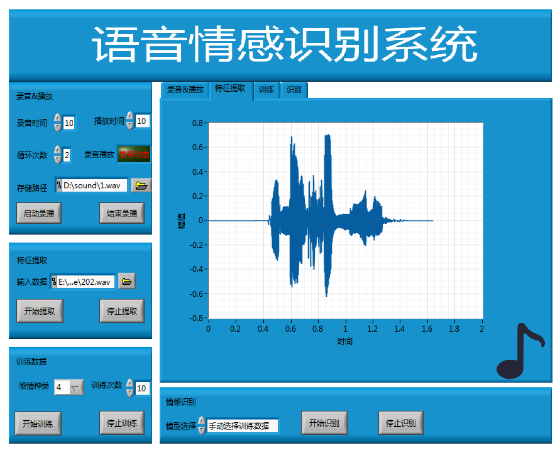
Table 1 Features extracted from the MIRtoolbox Toolbox

|  |  |  |
| --- | --- | --- |
| 特征 | 分类特征 | 特征提取值 |
| 时域 | 均方根 | 均值，方差，斜率，周期频率，周期幅度，周期熵 |
| 波形 | 峰值、质心 | 峰值位置，峰值最大值  质心均值 |
| 节奏 | 节拍、激起 | 节拍均值、激起均值，斜率 |
| 谱特征 | 质心、亮度、延展、偏斜、峰态、滚降（阈值95%）、滚降（阈值85%）、谱熵、谱平坦度、粗糙度、不规则度 | 均值，方差，斜率，周期频率，周期幅度，周期熵 |
| 音色 | 过零率、最低能量、谱通量 | 均值，方差，斜率，周期频率，周期幅度，周期熵 |
| 音调 | 色谱峰值位置、峰值最大值、色谱质心、关键清晰度、模式、谐波变化 | 质心等的均值，方差，斜率，周期频率，周期幅度，周期熵 |

2）遗传算法。遗传算法是一种模拟自然界中生物进化过程，按照优胜劣汰，适者生存的法则，搜索最优解的方法。遗传算法首先将问题的可能潜在解进行编码，然后对通过模拟自然界中生物进化发生的繁殖，交叉和基因突变现象，逐代演化产生跟更好的近似解，在每一代的选择时，通过个体适应度大小从解群中选取较优的个体，利用遗传算子( 选择，交叉和变异) 对这些个体进行组合，产生新一代的候选解群，重复此过程，直到满足某种收敛指标为止[13]。简单遗传算法具有早熟收敛和局部搜索能力弱的缺陷，故本文用CFS方法作为适应度函数来评价选择建模特征。

3） CFS方法。综合评价了特征与分类之间的相关性和特征之间的冗余性，广泛用于特征选择，数据清洗，是一种有效的特征选择方法[14~15]。它的核心思想是从原始特征集中，采用相关性测度寻找出特征之间的相关性较低但与类标记高度相关的特征构成的特征子集，达到了去除冗余属性和与类无关属性的目的[16]。CFS对特征集中的特征与特征、特征与类别之间的相关度进行计算，计算公式如下所示：

(1)

式中的对特征子集的一个评价，其中中含有个特征；表示特征与类别之间的平均相关性；表示特征与特征之间的平均相关度。由式(1)可知，该方法计算得到的特征子集中，每个特征与类标记的关联度越大且特征之间的冗余度越小，则的值越大，当前的特征子集是优良的特征子集。

CFS方法选择出的特征子集中，每个特征与类标记之间及特征之间的相关性计算应用了Pearson线性相关系数，其计算公式如公式(2)所示：

(2)

其中，表示特征与类标记之间的线性相关系数，是样本个数。而特征与特征之间的相关系数同上（公式2）。由于Pearson相关系数描述两个变量间的线性相关的强弱程度，相关系数的取值在-1和1之间，当相关系数越接近+1或-1，则表示两个变量间的相关度越强；当相关系数越接近于0，表示两个变量的相关性越弱。

4) GA+CFS降维。系统提取的特征使用GA+CFS降维，降维步骤具体如下：

(1) 编码。将编码长度设计为预处理后的特征数，每个编码对应问题的一个解，即染色体。当染色体的值为“1”时，表示对应的特征参与建模; 为“0”时，表示该特征未被选用，不参与最终建模。

(2) 设定初始群体和终止条件。本文种群大小设定为50，利用随机函数产生50个染色体组成初始群体，进化代数设置为 200。

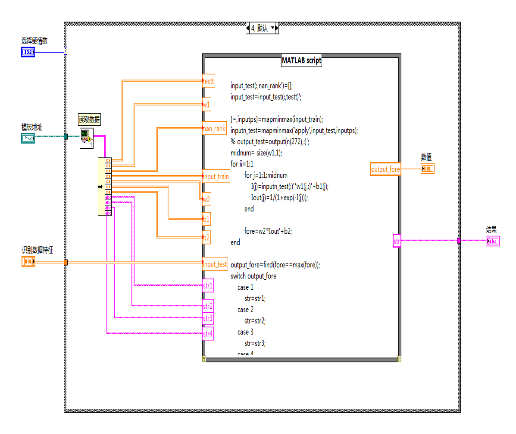
(3) 设定适应度函数。适应度用来度量群体中各个个体在优化计算中可能达到或者接近最优解的程度。适应度高则遗传到下一代的概率就大。度量个体适应度的函数就是适应度函数，本文将CFS方法作为遗传算法的适应度函数。

(4) 遗传算子的确定。通过轮盘赌方法、 单点随机交叉的方法和基因位变异的方法，将交叉概率设定为 0. 8，变异概率取 0. 2 来确定遗传算子。  
 (5) 种群经过选择、交叉、变异运算之后得到下一代种群。  
 (6) 当循环结束时终止运算。最终得到最大适应度函数以及最优特征子集。

2系统设计

LabVIEW是一种图形化编程软件，在其中可以直接调用Matlab程序。本实验采用Matlab2016a和LabVIEW2014混合编程，设计了一个语音情感识别系统。如图2所示，本系统共包含录音＆播放，特征提取，训练，识别四个模块。因为在本文第三部分已经验证BP神经网络的识别结果优于SVM的识别结果，故本语音识别系统训练和识别方法的内核利用BP神经网络。

图2（a） 语音情感识别系统前面板

Fig. 2 (a) front panel of speech emotion recognition system

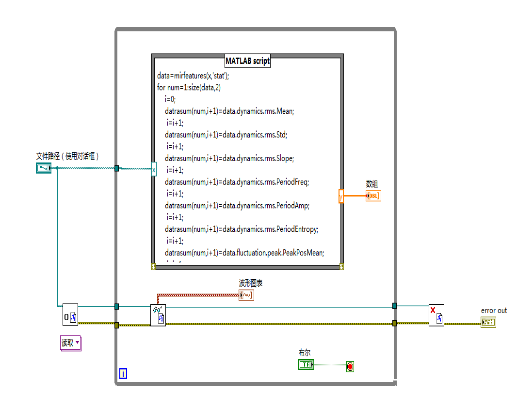
图2 （b） 特征提取子VI

Figure 2 (b) Feature extraction sub – VI

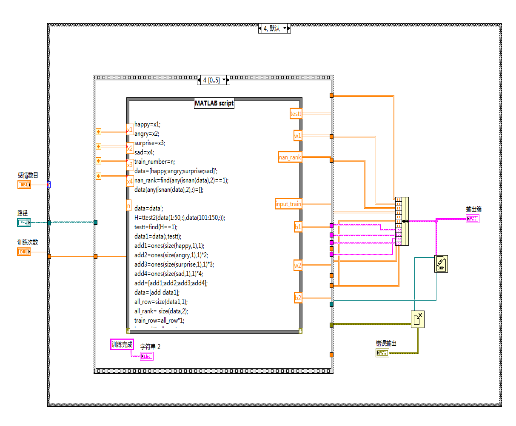
图2（c） 训练子VI程序框图

Fig. 2 (c) Program block diagram of training sub VI

图2（d） 识别子VI程序框图

Fig. 2 (d) Block diagram of identifying sub VI

录音＆播放模块可以录制播放音频，指定录制的文件位置和录制时间，录制的文件可以用来判别语音情感。

如图2（b）所示，特征提取模块主要是调用Matlab节点，并利用MIRtoolbox提取输入数据音频信号的统计特征，提取的特征以二进制的形式保存在同一文件目录之下。以备下一步语音情感使用。训练模块程序框图如图2(c)所示。训练模块在使用时要将已录制的经标注好的同一感情语音放在同一文件夹下，选择要训练的感情种类和训练次数，点击开始训练后会依次提取并训练选择文件夹下的语音，提取文件夹下的数据特征同样用到了Matlab节点，利用Matlab优秀的计算能力批量提取音频数据特征。最后训练的特征等会以二进制文件的形式保存在指定的文件夹下，训练完成后的文件用于识别时选择是否使用本次训练数据。

识别模块程序框图如图2(d)所示。识别模块可选择使用内置的训练数据或者手动选择使用哪一次的训练数据，内置训练数据是以CASIA数据库六种基本情感为模板训练的数据，手动选择数据可用来选择自己训练的数据，以达到针对不同人的情感识别的目的。

3实验

3.1数据库准备

语音情感识别研究的关键在于是否有好的情感语音数据库，情感定义准确，发音准确的语音数据库通常会有准确的分类识别结果，有利于进行情感的深入研究。目前在语音情感识别领域，如何使用数据库并没有统一的标准，按照激发情感的类型可分为表演型、引导型、自然型这3个类别；按照应用目标可分为识别型和合成型两个类别；按照语种不同可分为英语、德语、汉语等[1]。为避免争议，试验采数CASIA据库[1][17]。CASIA汉语情感语料库由中国科学院自动化研究所录制，由4位录音人(2男2女)在纯净录音环境下(信噪比约为35db)分别在6类不同情感下(喜悦、生气、惊讶、悲伤、恐惧、平静)对500句文本进行的演绎得到。

3.2 GA+CFS降维

实验通过MIRtoolbox提取特征，经过预处理后作为原始数据特征。为克服简单遗传算法的不足，本文以CFS方法作为适应度函数来评价选择建模特征。按照以上所述参数，其中一次实验适应度进化曲线如图3所示。

图3 CFS作为评价准则适应度函数进化曲线

Fig. 3 Fitness function evolution curve when CFS as an evaluation criterion

由图3可以看出，进化代数较小时，平均适应度值不能很好的接近最佳适应度值，当进化代数超过20代后，平均适应度值接近最佳适应度值附近，说明种群中每个个体都在最优解周围，也就是选择的特征近似最优。

3.3识别验证

将GA+CFS筛选的特征作为输入特征，由此检验此种方法的准确性。实验分别采用BP神经网络网络和SVM算法验证降维后特征的识别率，为了减少随机参数引起的误差，其中SVM的惩罚参数和核函数参数由GA算法优化。

分类算法采用具有代表性的SVM和BP神经网络算法进行验证比较。其中SVM算法调用台湾林智仁教授编写的libSVM包。SVM的核心思想是建立一个分类超平面作为决策曲面，使得不同类别之间的隔离边缘最大化。BP神经网络模拟人的思维方式，通过网络学习，调节权值和阈值，训练网络，使其减少下次犯错误的可能性。实验采用100次10折交叉验证方法进行分类，即对CASIA数据库样本中四个人每人的300句语音库中随机取270句提取特征进行建模，剩下的30句话作为测试集验证分类效果。验证结果如表2，表3所示；分类效果如图4，图5所示。

表2 SVM验证未降维和GA+CFS降维后识别结果

Table 2 Recognition results which compared original features with GA+CFS dimensionality reduction features verified by SVM

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 算法 | SVM识别结果 | | | | | | | |
| 喜 | 怒 | 惊 | 悲 | 恐 | 平静 | 平均 | 时间（s） |
| LC未降维 | 100% | 100% | 60% | 71.43% | 60% | 100% | 81.90% | 5.146 |
| LC GA+CFS降维 | 82.49% | 77.15% | 88.92% | 77.86% | 77.86% | 79.97% | 78.30% | 2.515 |
| WZ未降维 | 83.97% | 91.70% | 91.30% | 91.54% | 97.80% | 981% | 91.19% | 5.427 |
| WZ GA+CFS降维 | 72.25% | 81.29% | 76.89% | 74.42% | 79.90% | 86.16% | 78.48% | 3.245 |
| ZQY未降维 | 91.83% | 93.05% | 81.71% | 91.33% | 89.22% | 996% | 89.68% | 9.338 |
| ZQY GA+CFS降维 | 96.56% | 86.18% | 91.47% | 94.21% | 94.04% | 72.20% | 87.17% | 5.535 |
| ZZX未降维 | 85.79% | 93.98% | 91.85% | 84.11% | 82.32% | 88.88% | 87.82% | 7.146 |
| ZZX GA+CFS降维 | 70.31% | 84.61% | 91.47% | 75.63% | 70.21% | 86.75% | 79.83% | 4.020 |



图4 SVM验证降维前后测试样本

Fig.4 Classification figure of test samples between before and after dimensionality reduction verified by SVM

如表2所示，对于SVM来说，降维前后的总识别率略有下降，但是降维后建模时间减少了3s左右，建模速度提升了。由图4的某次单人实验降维前后测试样本分类图可见，降维前的错误分类数为6个，降维后错误分类数为7个，没有太大差别。

表3 BP验证未降维和GA+CFS降维后识别结果

Table 3 Recognition results which compared original features with GA+CFS dimensionality reduction features verified by BP

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 算法 | BP识别结果 | | | | | | | |
| 喜 | 怒 | 惊 | 悲 | 恐 | 平静 | 平均 | 时间（s） |
| LC未降维 | 80.44% | 84.39% | 76.93% | 77.91% | 55.72% | 52.77% | 71.36% | 80.433 |
| LC GA+CFS降维 | 99.80% | 99.58% | 97.96% | 93.48% | 80.38% | 83.74% | 92.49% | 6.128 |
| WZ未降维 | 86.42% | 90.45% | 85.48% | 66.57% | 66.01% | 69.91% | 77.47% | 79.465 |
| WZ GA+CFS降维 | 99.92% | 99.80% | 96.78% | 97.80% | 94.04% | 93.40% | 96.96% | 5.938 |
| ZQY未降维 | 89.84% | 87.82% | 79.99% | 68.40% | 40.34% | 54.79% | 70.20% | 83.696 |
| ZQY GA+CFS降维 | 96.56% | 86.18% | 91.47% | 94.21% | 94.04% | 72.20% | 87.17% | 13.465 |
| ZZX未降维 | 90.89% | 86.09% | 84.56% | 69.12% | 68.25% | 59.49% | 76.40% | 87.276 |
| ZZX GA+CFS降维 | 97.61% | 88.83% | 96.31% | 94.68% | 87.30% | 93.34% | 93.01% | 14.331 |



图5 BP验证降维前后测试样本分类图

Fig.4 Classification figure of test samples between before and after dimensionality reduction verified by BP

如表3所示，对于BP来说，降维后的识别率比降维前均高出20%左右，且降维后建模时间减少70s左右。由图5的某次单人实验降维前后测试样本分类图可见，降维前的错误分类数为8个，降维后错误分类数为4个，显然降维后比降维前的错误分类率大大减少了。

4结论

语音情感识别是现在研究的一个热点，本文利用LabVIEW平台结合Matlab设计了一个语音情感识别系统，该系统可用于语音情感识别，同样也可用于音乐情感的识别。语音信号特征的提取利用MIRtoolbox工具箱提取了音频信号的138维特征，并用经CFS改进的遗传算法对其进行降维处理，最终用SVM和BP神经网络对降维前后的特征进行识别验证。结果利用SVM，降维前后的总识别率略有下降，但是降维后建模时间减少了3秒左右；利用BP神经网络识别率平均提高了20%左右，建模时间均减少了70秒左右。由此可见，经CFS改进的GA算法筛选出了关键特征，提高了建模速度和识别正确率。

结　　论

参考文献

# 攻读硕士学位期间发表的论文及其它成果

1. **发表的学术论文**

侯一民, 周慧琼. 深度学习在语音识别中的研究进展综述[J]. 计算机应用研究, 2017,34(8).

# 东北电力大学学位论文原创性声明和使用权限

学位论文原创性声明

本人郑重声明：此处所提交的学位论文《基于深度学习的孤立词语音识别系统设计》，是本人在导师指导下，在东北电力大学攻读学位期间独立进行研究工作所取得的成果，且学位论文中除已标注引用文献的部分外不包含他人完成或已发表的研究成果。对本学位论文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。

作者签名： 日期： 年 月 日

学位论文使用权限

学位论文是研究生在东北电力大学攻读学位期间完成的成果，知识产权归属东北电力大学。学位论文的使用权限如下：

（1）学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文，并向国家图书馆报送学位论文；（2）学校可以将学位论文部分或全部内容编入有关数据库进行检索和提供相应阅览服务；（3）研究生毕业后发表与此学位论文研究成果相关的学术论文和其他成果时，应征得导师同意，且第一署名单位为东北电力大学。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。

本人知悉学位论文的使用权限，并将遵守有关规定。

作者签名： 日期： 年 月 日

导师签名： 日期： 年 月 日

# 《中国优秀博硕士学位论文全文数据库》和《中国学位论文全文数据库》投稿声明

东北电力大学：

本人同意《中国优秀博硕士学位论文全文数据库》和《中国学位论文全文数据库》出版章程的内容，愿意将本人的学位论文委托东北电力大学向中国学术期刊（光盘版）电子杂志社的《中国优秀博硕士学位论文全文数据库》和中国科技信息研究所的《中国学位论文全文数据库》投稿，希望《中国优秀博硕士学位论文全文数据库》和《中国学位论文全文数据库》给予出版，并同意在《中国优秀博硕士学位论文全文数据库》和CNKI系列数据库以及《中国学位论文全文数据库》中使用，同意按章程规定享受相关权益。

论文级别： □硕士 □博士

作者签名： 指导教师签名：

日 期： 年 月 日

作者联系地址（邮编）：

作者联系电话：

致　　谢

学位论文的完成意味着硕士生活即将接近尾声。回首这一段求学之路，感触很多，也收获也很多。

首先，我要衷心的感谢我的导师侯一民教授对本人研究生期间的帮助！感谢他对我的悉心教导与照顾。在我的课题上，侯老师给了我很多的帮助，每当遇到问题，老师都会给我提许多宝贵的意见，帮助我克服困难，是我得以顺利的进行课题研究。再次感谢老师对学业上指导，生活中的关切！在此，祝老师身体健康，家庭幸福、工作顺利！

在此，我还要感谢答辩组每位老师对我研究课题的指导与帮助！同时，还要感谢实验室的每一位同学，在平时的生活学习中对我的帮助。感谢大家营造的和谐，快乐的实验室环境。分别在即，在此祝愿大家前程似锦，心想事成！

另外，还要感谢我的闺蜜们，生活、学习上都少不了你们的帮助。是你们让我的研究生生活更精彩，这短短的两年半内，很开心有你们的陪伴，就算分离你们也是我一辈子的闺蜜。祝我的闺蜜们幸福、健康、快乐！

最后，最应该感谢的是我的父母，一直以来都非常尊重和支持我的任何决定，还有这么多年对我的陪伴和细心照顾，没有你们的不辞辛苦的付出，就没有今天的我。你们是我不断奋勇前进的动力。在此，想对我的父母说一声，你们辛苦了！谢谢你们一直以来对我的默默付出！愿我的父母可以身体健健康康！

再次感谢我身边所有的家人与朋友，你们是我生活中不可缺少的一部分。谢谢你们一直以来的陪伴！