****

**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

在人与自然，人与人的交流中，音频信号起了传递信号，表达信息的作用。不同情感的音频信号可能引发不同受众的不同体验，人类情感和不同情感的音频之间有什么量化关系，音乐厂商，唱片公司极其关注这个问题，以期获得唱片发行的成功或者对音乐，歌曲的发行作提前的预期避免损失；同时，探究人类情感，进行情感计算，也是进行人机交互，实现人工智能的一个热点。脑电信号，作为人类的一个典型生物信号，包含有大量的情感信息，能更客观的表示人类的情绪变化。近年来，随着对脑科学研究的不断深入，脑电信号的提取更加方便快捷，对脑电信号情绪特征的研究也引起了广大学者的兴趣。

在此背景下，本研究广泛提取了不同情感音频信号及由其引发的脑电信号的典型特征，并利用特征提取，数据分析，机器学习的等理论和方法，对不同情感的音频信号，以及由其引起的不同受众的脑电信号进行了量化分析，并对两者的关联性作了评价。具体内容如下：

（1）对语音情感识别，使用中科院的CASIA数据库；在音乐方面，利用T检验标注80首不同情绪（愤怒，喜悦，悲伤，平静）的音乐。提取语音信号的典型64维特征，同时利用Mirtoolbox工具箱提取音乐信号时域的均方差，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、峰值、质心等138个统计特征。利用多种降维方法对音频原数据降维，利用多种分类器识别验证上述降维前后的特征，结果证明了提取的语音信号的10维特征向量以及音乐信号的8维特征向量具有一定的普适性。

（2）采集了8位大学生受众在听不同情绪音乐时的脑电信号，得到15组实验数据，提取12个电极脑电信号的线性特征和非线性特征，得到共27维特征，然后，基于相关的特征选择（CFS）选择与原始特征最密切相关但最小冗余的特征集。利用不同分类器检验原始特征集和所选特征集的识别率和召回率。最后，针对实验结果进行了详细的分析，得到对音乐情绪最具代表性的脑电特征并加以验证。

（3）结合音频信号和脑电信号选择的最优特征，联合分析音乐信号特征和由其引起的脑电信号特征。最后选择20维的原音乐-脑电联合特征，GA+CFS降维后的6维特征，分别用BP，SVM，C4.5分类器对其进行识别验证。实验结果证明，经过GA+CFS降维的特征可以很好的对脑电情绪进行分类。在分类器性能上，BP分类器或C4.5分类器比较优秀。其中BP分类器识别率皆在80%以上，但BP分类器在实际应用中受网络训练次数限制，建模速度较慢。

（4）利用LabVIEW平台强大的可视化图形处理功能同时利用Matlab节点进行数据计算，设计了一个情感识别系统，用于语音和音乐情感识别。

关键词：情感计算；音频信号提取；脑电信号提取；特征选择；分类器

# **Abstract**

In the communication between man and nature, include man and others, audio signals transmit signals and express the function of information. The music makers and record company is extremely concerned about the problem that what kind of different emotions could lead to what different audio experience, in order to obtain the success or improve the anticipation of the release of the record and music and to avoid the loss. At the same time, that emotion calculation is also a hotspot of researching human-computer interaction and exploring human emotions. EEG signal is a typical signal of human biology, containing generous emotional information and more emotional changes. In recent years, with the progress of science and technology, the extraction of EEG characteristics of emotion has aroused extensive research because of extraction of EEG signals is more convenient.

Under this background, this study extracted the typical characteristics of widely different emotional audio signals and the corresponding EEG signals, then we quantified analysis and evaluate between different emotional audio signals and the EEG of different audiences by using data analysis. The specific contents are as follows:

(1) Firstly, using the CASIA database in voice signal and 80 different emotions be marked label (anger, joy, sadness, calm) by using T test. Extraction of typical 64 dimensional features present as voice signal feature. Then 138 statistical features has been extraction by using the Mirtoolbox, including the waveform and spectrum characteristics, tone, intonation features mean and variance, slope, periodic the frequency, amplitude, periodic entropy, peak centroid and so on. Last, a variety of dimensionality reduction methods are used to reduce the dimension of the original data, and the that features are verified by multiple classifiers. That results show that the 10 dimensional feature of extracted voice signals and the 8 dimensional features of the music signals are preeminent.

(2) This paper extracted 27 dimensional features for each 12 electrodes which are related with human emotions. Then, Correlation-Based Feature Selection (CFS) was employed to select the feature set which is most closely related to original features but with smallest redundancy. Some classifiers were used to test the recognition rates of the original feature set and the selected feature set. Finally, the experiment results were analyzed in details to get the effect from the features in the selected feature set to different human emotions.

(3) Combined with the above experimental results, the characteristics of the music signal and the EEG features caused by it are analyzed jointly. The experiment selects the original joint features of 20 dimension and the 6 dimension features through the GA+CFS method reduce dimension, then classifies it with a variety of classifiers respectively. The experimental results show that the above 6 dimensional original features can be used to classify the electroencephalogram. For classifier performance, the BP classifier or the C4.5 classifier is better. The recognition rate of BP classifier is more than 80%, but the BP classifier is limited by the number of network training times in practical application, because of the modeling speed is slow.

(4) a speech emotion recognition system is designed with the LabVIEW platform and Matlab, which is used for speech and music emotion recognition.

**Keywords:** **Emotion calculation; Audio signal extraction; EEG signal extraction; Feature selection; classifier**

目　　录

[摘　　要 II](#_Toc509257480)

[**Abstract** III](#_Toc509257481)

[第1章绪论 - 1 -](#_Toc509257482)

[1.1 课题背景及研究的目的和意义 - 1 -](#_Toc509257483)

[1.2 国内外研究现状及发展趋势 - 1 -](#_Toc509257484)

[1.3 本文的主要研究内容 - 3 -](#_Toc509257485)

[第2章　音频信号和脑电信号情感识别基本原理 - 5 -](#_Toc509257486)

[2.1 情感定义及划分 - 5 -](#_Toc509257487)

[2.1.1离散情感模型 - 5 -](#_Toc509257488)

[2.1.2维度情感模型 - 5 -](#_Toc509257489)

[2.2 音频信号情感识别基本原理 - 6 -](#_Toc509257492)

[2.2.1时域特征 - 6 -](#_Toc509257493)

[2.2.2频域特征 - 6 -](#_Toc509257494)

[2.2.3谱特征 - 8 -](#_Toc509257495)

[2.2.4调式 - 9 -](#_Toc509257496)

[2.3 脑电情感识别基本原理 - 9 -](#_Toc509257497)

[2.3.1情感大脑 - 9 -](#_Toc509257498)

[2.3.2EEG信号特点 - 9 -](#_Toc509257499)

[2.3.3 EEG信号预处理 - 10 -](#_Toc509257500)

[2.3.4 EEG信号情感特征 - 11 -](#_Toc509257501)

[2.4降维方法 - 13 -](#_Toc509257502)

[2.4.1线性判别分析 - 13 -](#_Toc509257503)

[2.4.2 多类簇特征选择 - 14 -](#_Toc509257504)

[2.4.3 平均影响值算法 - 14 -](#_Toc509257505)

[2.4.4遗传算法 - 14 -](#_Toc509257506)

[2.4.5特征选择算法 - 14 -](#_Toc509257507)

[2.5识别方法 - 15 -](#_Toc509257508)

[2.5.1 BP神经网络 - 15 -](#_Toc509257509)

[2.5.2 支持向量机 - 16 -](#_Toc509257510)

[2.5.3 决策树 - 16 -](#_Toc509257511)

[第3章　语音信号情感识别 - 17 -](#_Toc509257512)

[3.1 实验设计 - 16 -](#_Toc509257513)

[3.2 选取特征 - 16 -](#_Toc509257514)

[3.3 实验结果分析 - 16 -](#_Toc509257515)

[3.3.1 MIV降维 - 16 -](#_Toc509257516)

[3.3.2 LDA以及MCFS的方法降维 - 16 -](#_Toc509257517)

[3.3.4 用CFS方法降维 - 16 -](#_Toc509257518)

[3.4 特征分析 - 16 -](#_Toc509257519)

[3.5 推广验证 - 16 -](#_Toc509257520)

[3.6 本章小结 - 16 -](#_Toc509257521)

[4音乐情感识别 - 17 -](#_Toc509257522)

[4.1数据库 - 25 -](#_Toc509257523)

[4.2 音乐信号特征选取 - 26 -](#_Toc509257524)

[4.3 实验设计 - 26 -](#_Toc509257525)

[4.4 识别验证 - 27 -](#_Toc509257526)

[4.5 特征分析 - 29 -](#_Toc509257527)

[第5章 音频信号引起的脑电信号分析 - 31 -](#_Toc509257528)

[5.1方法设计 - 31 -](#_Toc509257529)

[5.2 脑电采集步骤 - 31 -](#_Toc509257530)

[5.3数据分析 - 32 -](#_Toc509257531)

[5.3.1 特征提取 - 32 -](#_Toc509257532)

[5.3.2 分类器验证 - 32 -](#_Toc509257533)

[5.3.3 特征检验 - 34 -](#_Toc509257534)

[5.4特征分析 - 35 -](#_Toc509257535)

[5.4.1线性特征 - 35 -](#_Toc509257536)

[5.4.2非线性特征 - 36 -](#_Toc509257537)

[5.4.3 检验 - 38 -](#_Toc509257538)

[5.5本章小结 - 38 -](#_Toc509257539)

[第6章　音频信号和脑电信号关联分析 - 40 -](#_Toc509257540)

[6.1 方法设计 - 40 -](#_Toc509257541)

[6.2 特征选择 - 40 -](#_Toc509257542)

[6.3 识别验证 - 41 -](#_Toc509257543)

[6.4 本章小结 - 43 -](#_Toc509257544)

[第7章　基于LabVIEW的语音和脑电情感识别系统 - 44 -](#_Toc509257545)

[7.1 系统流程 - 44 -](#_Toc509257546)

[7.2 程序设计 - 44 -](#_Toc509257547)

[7.2.1 录音与播放 - 45 -](#_Toc509257548)

[7.2.2特征提取 - 45 -](#_Toc509257549)

[7.2.3模型训练 - 46 -](#_Toc509257550)

[7.2.4 识别 - 47 -](#_Toc509257551)

[7.3本章小结 - 47 -](#_Toc509257552)

[结　　论 - 48 -](#_Toc509257553)

[参考文献 - 50 -](#_Toc509257554)

[攻读硕士学位期间发表的论文及其它成果 - 55 -](#_Toc509257555)

[东北电力大学学位论文原创性声明和使用权限 - 56 -](#_Toc509257556)

[致　　谢 - 57 -](#_Toc509257557)

# 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，因为其辨识度高，结果客观真实常作为情绪识别的评判标准。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，经常利用音乐刺激受众，从而引发受众的情绪变化；同时音乐刺激对于抑郁症的治疗或者阿尔茨海默症的预防也有相当明显的作用[1]。

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[2]？不同文化背景下的受众在接受到同样的音频信息后会产生怎么样的情绪[3]？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[3]？音频信号概念在人脑中是具体的，特定的，还是抽象的[4]？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的最优特征子集，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，电机的安置位置按照国际通用的10-20标准，左右乳突的物理连线作为参考信号，头皮电极电阻都置于5千欧以下，采集频率为直流电50赫兹，采样率为256赫兹。

1.2 国内外研究现状及发展趋势

音频信号主要包括语音和音乐两种信号。迄今为止，情感识别领域已经成功的运用了多种典型的模式识别算法。比如，比较典型的基于概率的识别算法，像高斯混合模型[5][6]和隐马尔科夫模型[7][8]；基于判别的模式识别算法，像支持向量机[5][9]和神经网络[10][11]。这些识别模型在单独使用时大多会有缺陷，例如陷入局部最优值或者维数灾难，于是在实际识别时经常混合使用各个模型[12]。

音频信号情感识别需要提取音频信号底层特征，专家学者在这方面投入了许多精力，做了许多研究[13]。例如，Luengo等研究了多维韵律特征组成的特征集。Riglia等人运用多个韵律统计特征组成的特征集[13][14]。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[13][15]，韵律特征对情感的辨识能力不足等[13]。可见，音频情感识别的发展仍不完善，还有很长的一段路要走。

在音乐情感识别方面，Wieczorkowska A 利用K近邻算法进行音乐情感识别，对不同的歌曲多标签分类[16]。Shi Y Y 等构建了一个双层音乐情感分类系统[17]。Cyril Laurier等人利用支持向量机设计了一种音乐情感自动标注系统[18]。由最近的研究来看，国际上各种研究所和学校提取了声音的各种特征并用各种分类方法（神经网络,支持向量机，决策树等）对音频情感进行了研究，取得了很好的效果[19~21]。浙江大学张克俊对生理信号标记的音乐信号库采用修正基因表达编程算法对其小波特征、梅尔倒谱系数、短时傅里叶谱特征进行在线情感识别研究[22]。上海同济大学的尤鸣宇等用多标记数据特征选择结合机器学习方法进行声音情感分类研究[23]。

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。而音乐刺激与脑电信号之间的关系，中外学者做了对应的研究并取得了丰硕的成果。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以及分形维数用于分类，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[24]。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[25]。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[26]。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[27]。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[28]。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[29]。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波段的有统计性的差异[30]。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[31]。谢康利用方差检验和t检验选取了9首音乐作为刺激信号并分为三组，提取了PSD和不同的节律信号特征，进行脑电分析[32]。马勇等通过脑电分析，在Android平台上初步实现了个性化情绪音乐推荐系统，可以根据受众情绪的变化调整播放列表[33]。

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，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[34]。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，提取多个参数特征，并比较了所提出的算法与主成分分析法的优劣[35]。

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；研究不同情绪的音频主要引起哪些大脑脑区发生对应的变化，以及这种变化如何进行定量分析。这对脑电信号采集时减少冗余特征，提高分辨率，更好的识别人类情绪具有重要的意义。

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；对音色特征和旋律特征研究较少；不同的学者对于哪种特征更能有效的代表情感认识不一致等。故本文拟提取音频信号的多种融合特征，进行分类识别，最后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，本文建立了一个基于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和部分非线性动力学特征，对各个特征的有效性、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的哪些特征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。

本文的具体组织和安排如下：

第一章讲了音频信号和脑电信号联合分析的背景和意义，现在国内外相关研究的趋势和不足，以及介绍整篇论文的组织和安排。

第二章讲音频信号（包括语音信号和音乐信号）和脑电信号情感识别的基本原理以及特征提取、特征降维、识别验证的常用手段。

第三章主要提取了CASIA数据库语音信号常用的64维特征，并用MIV结合CFS方法选择除了最具有影响性的13维特征向量，证明语音信号的基频特征和共振峰特征较其他特征更能代表说话人的情绪。

第四章标注了1000 songs database的4种情感，利用MIRtoolbox工具包提取了不同情感音乐的138维特征，利用CFS方法改进GA方法进行特征降维，最终得到了8维最能代表全局变量的特征向量。最后利用SVM和BP神经网络进行分类验证证明选取的特征向量可以很好的代表不同情感的音乐，并对选取的8维特征进行了分析。

第五章设计了一系列实验，目的是探究音乐引发的情感有何明显的脑电特征，以及对于不同的感情脑电特征有何不同。利用不同情绪的音乐作为刺激材料，提取与情绪相关的12个电极每个电极的27维特；体用脑地形图和定量分析的方法选择出了T3，T4，PZ三个与情感识别相关的电极。利用互相关特征选择 (Correlation-Based Feature Selection，CFS)方法的方法选择出这个三个电极与原特征集相关性较大冗余较小的12维最优特征集；利用不同的分类器分类验证了原特征集和最优特征集的识别率。最后，针对实验结果进行了详细的分析，得到对音乐情绪最具代表性的脑电特征并加以验证。

第六章利用第四章和第五章的实验结果，设计了一系列实验，探究音频信号同其由其刺激引发的脑电信号之间的关联。联合分析上述选取的8维音乐特征和12维脑电特征，利用GA+CFS的方法对其进行降维，并用SVM、C4.5、BP神经网络的方法对其进行识别验证，最终得到最具代表性的6维特征向量，实验证明这6维特征可以比较好的代表音乐刺激和其引起的脑电变化。

第七章基于以上研究，设计了基于LabVIEW的语音情感识别系统，该系统利用互相关特征选择经过CFS改进的遗传算法进行特征提取，对音频信号特征降维，并训练BP神经网络作伪识别的模型，提高了识别正确率，具有一定的应用价值。实验证明：该系统可以对音频信号进行有效的情感识别。

# 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

情感是人类独特的一种主观体验，是对外界刺激产生的消极或者积极的应对反应。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，情感模型主要被划分两个类别：离散情感模型和维度情感模型。

### 2.1.1离散情感模型

离散情感模型认为情感是单独的、具有明确含义的几种标签化的情绪。这几种情感被贴有形容词词性的标签，例如，高兴、悲伤等。其中，能被不同区域社会文化共同认可，除人类外，其他社会性的哺乳动物也具有的普遍性情感，被称为基本情感，其中，Ekman提出的情感划分理论较为著名，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[13]。

### 2.1.2维度情感模型

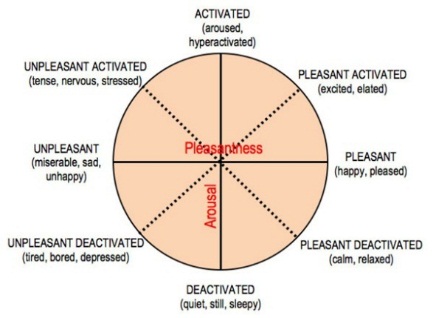
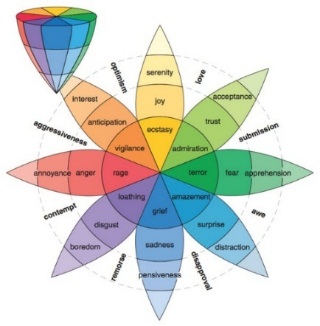
20世纪中期，Wundt等提出了环形维度情感模型。环形模型包括两个维度，即激励(Activation)和诱力(Valence)两个正交维，任何情绪可以放置在两维激励和诱力空间上，如得分低的诱力维和的激励区域被认为是悲伤。在这个环上散布着能大部分的人类情感，可以量化表示不同情感的距离。而Thayer二维情感模型更是被应用甚广的维度情感模型[36]。Thayer情感模型是二维情感连续型心理表示模型，其纵坐标表示的是激励维度（Arousal），从“平静的”到“活力的”变化，反应的是主体的情感活跃程度；横坐标表示的是诱力维度(Valence)，从“消极的”到“积极的”变化，反应的是主体的主观感受[13]。从而将情感划分为具有代表性的四个区域。Thayer的情感模型是联系音频情感标签和空间坐标的桥梁，使两者可以相互映射，通过对两个维度不同的量化，可以找到定义的情感标签在情感坐标空间的相对位置。此外，还有三维的激励-评估-控制空间理论[37]。

图2-1 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，主要有语音学特征和全局统计特征两种。语音学特征研究一般都只是定性研究而不定量分析，包括语音的元辅音、节奏等。而全局统计特征按信号参数的不同分类包括时域特征、频域特征、谱特征和其他特征[38]，按照音乐本身特征分为音色特征、韵律特征和其他特征[39]。时域特征包括波形特征，短时能量，均方根等，频域特征包括基音频率、共振峰等，谱特征主要包括短时频谱和美尔频率倒谱系数等。按照音乐理论划分，韵律特征包括基音频率，共振峰，发音时长，节奏特征等；音色特征包括美尔频率倒谱系数特征和倒谱特征和其他谱特征，此外音乐情感特征还包括调式特征。大部分学者都同意韵律特征能表达较多的情感特征[40][41]。下面主要介绍全局统计特征。

### 2.2.1时域特征

短时能量：表示语音信号随时间变化能量的变化，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，无话段和有话段。

C:\Users\Administrator\Desktop\提取\eng-pict\1.3.tif1.3

图2-2 平静状态下CASIA数据库“就是下雨也去”语音信号及其短时能量图

均方根（root-mean-square, RMS）：即对信号波形先取平方、再平均、最后开方。对于信号，其均方根定义如下：

(2-1)

### 2.2.2频域特征

基音频率：人在发浊音时声带振动的频率称为基音频率。在表达不同的情感时，基音频率的均值，变化快慢，覆盖范围等都会不同，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[42]。

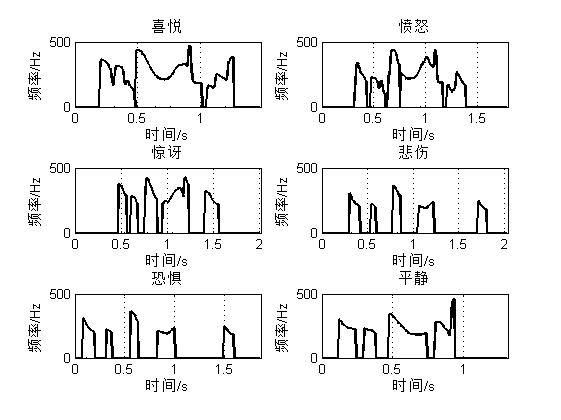


图2-3 六种情感“就是下雨也去”有音段基频比较

表1-1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：声音中的元音引发声道的共振，产生的频率叫做共振峰。情感变化的变化反应在第一，第二，第三共振峰[43]。

不同感情引发的共振峰其最大最小值、峰值和频率范围不一样，变化率也不相同。例如，恐惧的第一共振峰变化率最小，而喜悦的变化率最大[7]。

C:\Users\Administrator\Desktop\提取\eng-pict\2.8.tif2.8

图2-4 六种情感“就是下雨也去”有音段共振峰比较

### 2.2.3谱特征

短时频谱：由于语音、音乐，信号可以视为短时平稳的，故可以对其分帧处理，计算每一帧的傅里叶变换就是短时频谱，公式如下：

(2-2)



图2-5 语音信号语谱图

语谱图反映音频信号与短时频谱随在时间上的对应关系，它是短时频谱的集合，上图的灰色条纹即代表着此时音频信号的短时频谱。短时频谱的质心，带宽，滚降（95%或85%的截止频率）常用作音频信号分析[39]。

美尔频率倒谱系数：美尔频率倒谱系数（Mel Frequency Cepstrum Coefficient，MFCC）是科学家根据人耳听觉特征设计，用来分析音频信号的倒谱特征。由于人耳对语音的感知能力随频率的变化而不同，在0-1000赫兹，人耳的听觉是与此频域的音频成近似线性关系；而在1000赫兹以上，听觉则与此频域的音频成对数关系变化。根据这种特性，提出了美尔频率倒数用来模拟人耳的听觉特性，其公式为：

 (2-3)

式中 ——为Mel频率

——为实际语音频率

MFCC参数计算过程如下：



图2-6 MFCC的计算过程

### 2.2.4调式

在一首有意义的音乐中，一个孤立的音符或者和旋无法表征音乐的思想，只有一定的有关联的音符按照一定规则组合在一起的音符才能形成旋律。调式是围绕一个具有比较稳定感的中心音，按照一定的音程关系组合起来的音乐体系。不同的调式表现不同的思想，或阴郁或高亢，或慷慨或悲愤[44]。

此外，语音信号的发音帧数特征也得了广泛研究[45] [46]

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，大脑作为人类的情感中心，在表达不同思想感情时大脑会出现对应的脑电波。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。

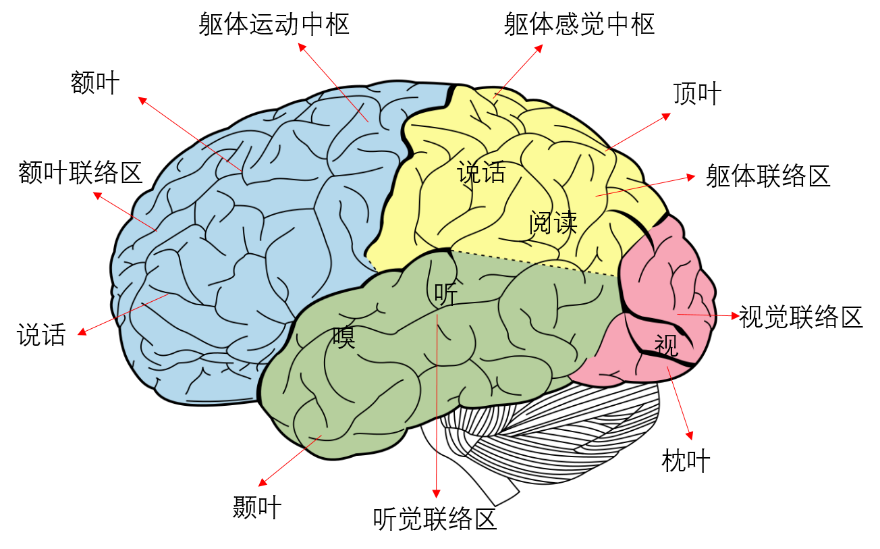


图2-7 大脑的不同功能分区

### 2.3.2EEG信号特点

大脑产生的电信号极其微弱，需要用特殊的电极帽采集，EEG电极帽是一种电压传感器，用来记录头皮上潜在的电位变化[47]。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，与人类情感有着密切关系。在频域，根据EEG信号频率范围的不同，划分了5种基本节律，即：δ节律、θ节律、α节律、β节律和γ节律。研究表明，不同的节律对应着人的心理状态或所处的生理情况不同，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：频率处于0.1～4Hz之间，幅度为0μV-200μV。δ节律多发生于婴儿期，成年人在意识清晰时难以记录到δ节律，但智力发育不成熟、或者极度疲劳或麻醉的情况下可以观察到成年人出现这种波段。

（2）θ节律，频率处于4～8Hz之间，幅度为100μV-150μV。θ节律多在少年人脑电频段中见到，在成年人情感受抑制或收到挫折时会出现这种波段，精神病患者的脑电波此种波段出现较多。

（3）α节律：频率处于8～13Hz之间，幅度为20μV-100μV。α节律是正常人的基本脑电节律，当人清醒、安静并且闭目时此节律比较明显，而当人睁眼、或受到其他刺激时，α节律就会消失。

（4）β节律：频率处于13～30Hz之间，幅度为5μV-20μV。当人的精神状态处于兴奋状态或者高度紧张时，就会出现β节律。

（5）γ节律：频率处于31～100Hz 之间。γ波与人类对于一些物体的识别和感知相关。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。由以上对于脑电节律的介绍可知，EEG的震动幅度十分微弱，极其容易夹杂在各种各样的背景噪音中，从而难以检测。所以，EEG信号的采集需要十分严格，需要采集过程中尽量避免混入各种噪声，使噪声的影响降到最低。

（2）非平稳性。EEG信号的产生是由大脑不断的活动、钠钾离子通道不断的开闭，由于脑电信号的产生十分复杂，调节过程也十分的随机，人的行为活动、思想感情又是时刻在变化的，所以由大脑皮层产生的脑电信号就十分的复杂。

（3）频域特性突出。如以上的节律介绍，分析脑电信号的频域特征对于处理EEG十分关键。

### 2.3.3 EEG信号预处理

EEG信号主要受测量系统和人体本身的干扰，测量系统的干扰主要包括：工频干扰、环境的电磁干扰，人体本身的干扰包括人体的其他组织、器官的动作干扰和其产生的生理电信号干扰，例如眨眼、眼动、呼吸、心跳等。

在以上干扰中，对脑电信号影响最大的是眼电信号。由于在脑电信号测量中一定存在有眼球的活动，眨眼、眼动产生了眼动电场，由于眼睛位置距离头皮极近，故形成的眼动信号对脑电的影响极大，在脑电的研究中一般都必须去除眼动伪迹。眼动伪迹的去除一般有主成分分析法（PCA）和独立成分分析法（ICA）[48]。

主成分分析：主成分分析是采用正交理论，将原始变量转换为一组相互独立的变量，放弃一些相对不重要的变量，选择其中相对比较重要的部分变量作为主要的自变量，然后运用最小二乘法得到主成分的估计值。

独立成分分析法（ICA）：独立成分分析法又叫做盲源分离，ICA是指不知道源信号和传输通道信号的情况下，由输入源信号的统计特征以及观测信号来恢复各个独立成分[49]。由于脑电信号中的扰动信号，如心跳，眼动伪迹，肌电信号等都可以视为由相互独立的信源产生的不同信号，故通过ICA理论便能从检测信号中去除伪迹信号，得到有用的脑电信号。

### 2.3.4 EEG信号情感特征

EEG信号和人类情感之间存在的关系可以由不同脑电特征表示，目前表征情感特征的EEG特征包括：

1. **线性特征**

时域分析：主要为Hjorth参数法[35]。Hjorth参数是一种时域分析的方法，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[50]。Hjorth法定义了3个时域参数：活动性、移动性及复杂度，这3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。

频域分析：频域分析是对脑电信号的频率谱估计，即把时域的脑电信号变换为脑电的频谱图，直观的展示脑电节律的变化[51]。功率谱估计法主要分为三类，即经典谱估计法，包括自相关法和基于快速傅里叶变换的周期图法；现代谱估计法主要是参数模型法；以及小波变换法。

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：

(2-4)

式中，表示数学期望 \*表示复共轭。功率谱密度函数的另一种表达方式是：

(2-5)

当随机信号的长度给定时，为得到周期图法的谱密度，式（2-5）求期望和取极限运算可以忽略不计：

(2-6)

当数据长度较短时，周期图法的分辨率很低，且周期图法存在有旁瓣泄露，故周期图对非平稳的随机信号即脑电信号的误差较大。

现代谱估计法：常用的模型包括AR、MA、自回归移动平移及谐波信号模型等。而对于脑电信号的现代谱估计，AR模型适用性较强，比较容易调节参数，在信噪比比较高的条件下也可以保持很好的分辨率，实时性较强，故多用于脑电信号的处理AR模型要求被测信号的有比较好的线性性和平稳性，对于以非平稳性著称的脑电信号，在使用AR模型估计功率谱时要对其进行分段处理[47]。

小波变换：小波变换是一种时频联合分析的方法，小波变换是傅里叶变换的进一步发展[1][52]。傅里叶变换把时域信号转变为频域信号，这样直观的呈现、定量分析频域信号，但是，傅里叶变换在描述信号频域信息的同时却无法描述信号时域信息，也就是说，当信号进行完傅里叶变换，得到频域信息的同时却完全丢失了时域信号的信息。在处理实际信号时，时域信号的信息却必不可少，例如，脑电信号的时域信号就包含着产生脑电信号的刺激的开始、过程和结束等。为了弥补傅里叶变换不能同时表达时域和频域的缺点，J.Morelt提出了小波变换。理论上，小波分析可以处理所有傅里叶变换可以处理的问题，而且小波变换同时包含时域和频域信息，可以很好的应用在多尺度。由电极采集到的脑电信号是由很多不同振荡频率混合而成的信号，这些频率包括有用的脑电信号同时也包括各种噪声。所以，小波变换可以应用于脑电信号模式分类，在脑电信号处理领域有着不可取代的作用。

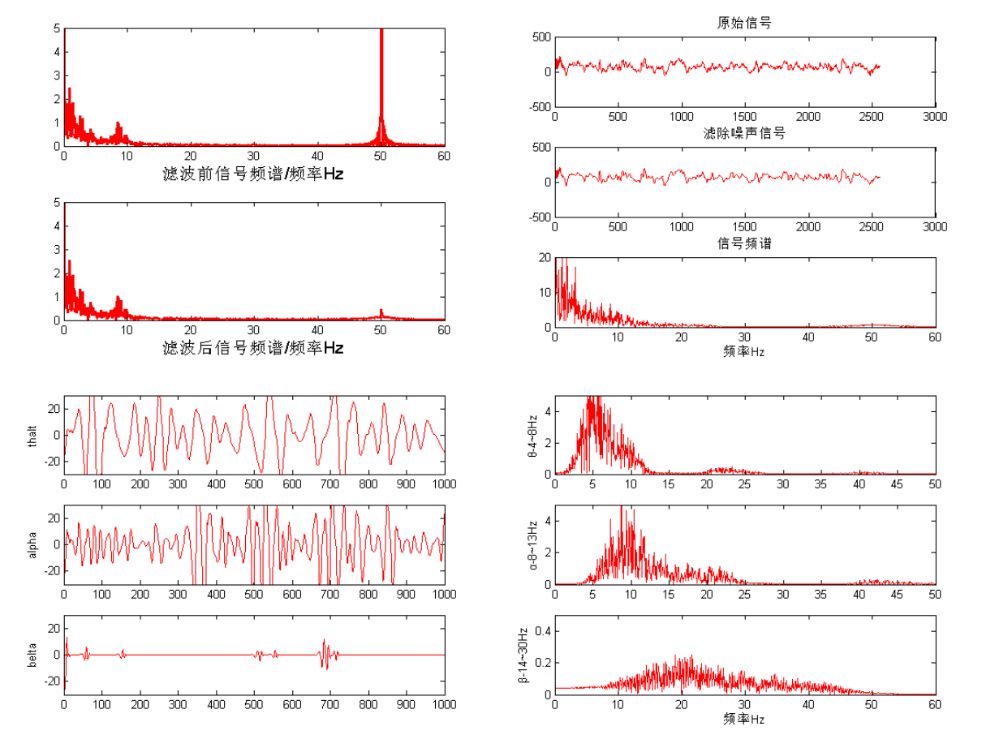
****

图2-8 小波变换提取脑电信号特征

1. **非线性特征**

随着对非线性系统的不断深入研究，非线性特征渐渐引起了脑电信号分析专家的注意。人脑是一个典型的非线性系统，于是一些非线性时间序列分析方法和信息学方法被应用于脑电分析。

近似熵（ApEn）：表示对时间序列的无需度测量，表征信号复杂性的大小[53][54]。近似熵算法相对简单，抗噪性能强，但分析数据量小。近似熵是一个正数，对于EEG信号，它的值越大表示复杂性越高或者无规律性越强。

样本熵：样本熵是Richman和Moorman[11]对近似熵进行的改进，降低了计算误差，提高了计算精度。样本熵通过度量信号产生新模式的概率来表征原始信号的复杂度的大小，样本熵的值越小，表示序列新模式产生的速率较低，原始信号的比较简单，反之则表示序列越复杂。相比于近似熵，样本熵具有更好的抗噪性和一致性[55]。

关联维和K熵:关联维是分维的一种，20世纪60年代Kolmogrov定义了Kolmogrov熵，可以用来表征混沌系统运动的混乱或无序的程度[56]。K熵又叫信息维数，Sinai对其进行了改进，故K熵又被叫做为Kolmogrov-Sinai熵，即为K-S熵。Crassbreger和Procaccia后来提出了相关维数的概念，他们利用相关熵来逼近Kolmogorov熵，得到k2熵，一般情况下可以同时得到关联维数和k熵的稳定估计，关联维数越大，脑电信号序列就越复杂。

LZ复杂度：20世纪60年代，由柯尔莫哥洛夫(Kolmogorov)等人提出，其基本思路是认为：一个序列的LZ复杂度是能够生产这一系列的最短计算程序比特数，即描述了序列接近随机序列的程度。因为刚开始提出的柯尔莫哥洛夫复杂度难以计算，A.Lempe1和J.Ziv提出了一种估计复杂性的算法，即Lempe1-ziv复杂度[57][58]。Lempel-Ziv复杂度算法表示一个给定的时间序列随其长度的增加引发新模式增加的速率。新模式随时间的增加出现的越多，意味着给定原数据发生的变化越快，这时LZ复杂度越大，表明这一时期的时间序列的变化是无规律的，新模式随时间的增加出现出现减小的趋势，意味着原数据序列发生的改变越慢，即新数据的出现时有规律可循的，这时LZ复杂度减小，出现很强的周期性。

最大李雅普诺夫指数：最大李雅普诺夫指数是否大于零是判断系统是否为混沌系统的标准：正的最大李雅普诺夫指数，表示在系统相空间中，不论初始系统轨道相距多么近，它们的轨道必定会随着时间的推移而发散，从而使系统轨道变得不能预测，这个系统就是混沌系统；负的最大李雅普诺夫指数，意味初始系统轨道必定会随着时间的推移而收敛，这个系统就不是混沌系统[59]。

C0复杂度：C0复杂度的算法思想主要是将要分析的时间序列进行分解，分为随机序列的部分和规则序列部分，对序列的复杂度测度就被定义为做随机运动部分的时间序列长度和全部原始时间序列长度分别与时间轴之间围成的面积之比[60]。

谱熵：即在频域内应用香农熵的概念，先对信号进行傅里叶变换，然后计算信号的功率分布，归一化单位功率[61]。

奇异谱分析：奇异谱分析就是将一维脑电时间序列进行延迟重构为多维相空间，经奇异值分解，按能量大小排序来确定被分解量的重要性[62]。奇异谱熵表征脑电信号能量在奇异谱分解下的不确定性，随脑电信号复杂性的增加，能量的分布也越来越发散。奇异谱熵可以衡量脑电信号非线性的程度。

2.4降维方法

### 2.4.1线性判别分析（LDA）

线性判别分析（LDA）是一种多变量统计分析方法，其利用被测数据的特征值判断属于哪一种分类。LDA是有监督的降维方法，它先对训练数据从高维映射到最佳识别的矢量空间，然后用一个线性判别函数进行判别。线性判别分析最重要的步骤是找到最佳的映射方法，将高维的特征投影到最佳特征矢量子空间，使得映射后的投影样本在子特征空间类与类之间距离最远，而类本身特征之间的距离最近。

### 2.4.2 多类簇特征选择(Multi Cluster Feature Selection，MCFS)

MCFS是由蔡登教授于2010年提出，MCFS是一种半监督的特征选择方法，目的是选择能最大保持数据类簇结构的特征，MCFS首先通过拉普拉斯映射计算数据的内在维度，然后利用稀疏表示的思想，根据内在维度计算每个特征在区分类簇时的贡献度[63]

### 2.4.3 平均影响值算法

平均影响值(Mean Impact Value,MIV)算法，用来表征神经网络的权重矩阵的变化，MIV算法通过量化输入特征对于网络训练结果的影响程度，从而达到筛选变量的目的[64]。MIV算法首先利用训练样本S训练得到一个比较稳定的神经网络，而后，将训练样本S中的每个特征在其原值基础上分别加和减10%构成两个新的训练样本S1和S2，然后将S1，S2作为输入特征重新进行训练神经网络，得到新的识别向量N1和N2，求N1和N2差值，即为变动此特征对识别带来的影响，最后将此影响值求均值就得到输入特征对识别正确率的平均影响值，依次可算出各个特征的MIV[65]，根据MIV的大小即可选择出对识别不同贡献程度的特征。

### 2.4.4遗传算法

遗传算法模拟自然界中生物物竞天择，适者生存的过程，通过搜索最优解进行特征筛选，GA算法首先编码问题的可能解，然后模拟自然界中生物的进化发生的选择、交叉和基因突变现象，不断进化产生接近最优解，在每一代选择时，通过适应度函数根据近似解选取比较优秀的个体，利用遗传算子对选择出的个体重新组合，组成下一次要迭代的可能解，对上述解群不断迭代，直到达到设定的某种收敛指标[66]。

### 2.4.5特征选择算法

CFS方法是一种被广泛应用的特征选择方法，它兼顾评价特征与分类之间的相关性以及特征与特征之间的冗余性[34,67~68]。CFS的主要思路是利用相关性分析的方法，从原始特征集中寻找出最优特征子集，最优特征子集的特征与特征之间相关性较低但特征与类标记相关度较高，从而去除原始特征的冗余子特征和无关子特征[51]。CFS计算方法如下所示：

(2-9)

式中的对特征子集的评价，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。

特征之间相关性可以用皮尔逊相关系数表示或者利用信息增益的方法求得。信息增益的计算方法如下：

假设数据属性为Y，y为Y的一个可能的取值，那么Y的熵的计算方法为：

(2-10)

若已知某一属性X，则X条件下计算Y熵的方法为：

(2-11)

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：

(2-12)

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：

(2-13)

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。

计算出不同特征子集的CFS值后，需要选择出最优特征子集，贪婪逐步搜索算法可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，, , ···,随后将特征子集(), (, ), (, ), (, , ), ···(, ···).进行分类器识别验证，根据结果的优劣即可得到特征子集的优劣。

2.5识别方法

### 2.5.1 BP神经网络

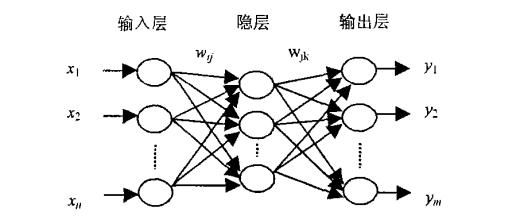
BP神经网络（Back Propagation network）是一种单向传播的多层神经网络学习算法，其结构简单，应用广泛。BP神经网络一般包括输入层，中间层（隐藏层），输出层三层，通过使用梯度下降法，通过反向传播来不断调整网络的权值和阈值，使网络的[误差平方和](http://baike.baidu.com/view/2135069.htm)最小。

图2-9 BP神经网络

图2-9中，,···是BP神经网络的输入值，, ···是BP神经网络的预测值，和是BP神经网络权值。

### 2.5.2 支持向量机

支持向量机（Support Vector Machine SVM）是一种二分类识别模型，是通过确定最佳的分界线，最小化经验损失，同时使得间隔最大的线性分类器。通过间隔最大化，就可以将许多分类问题转化为一个凸二次规划问题或者正则化的合页损失函数的最小化问题来求解。

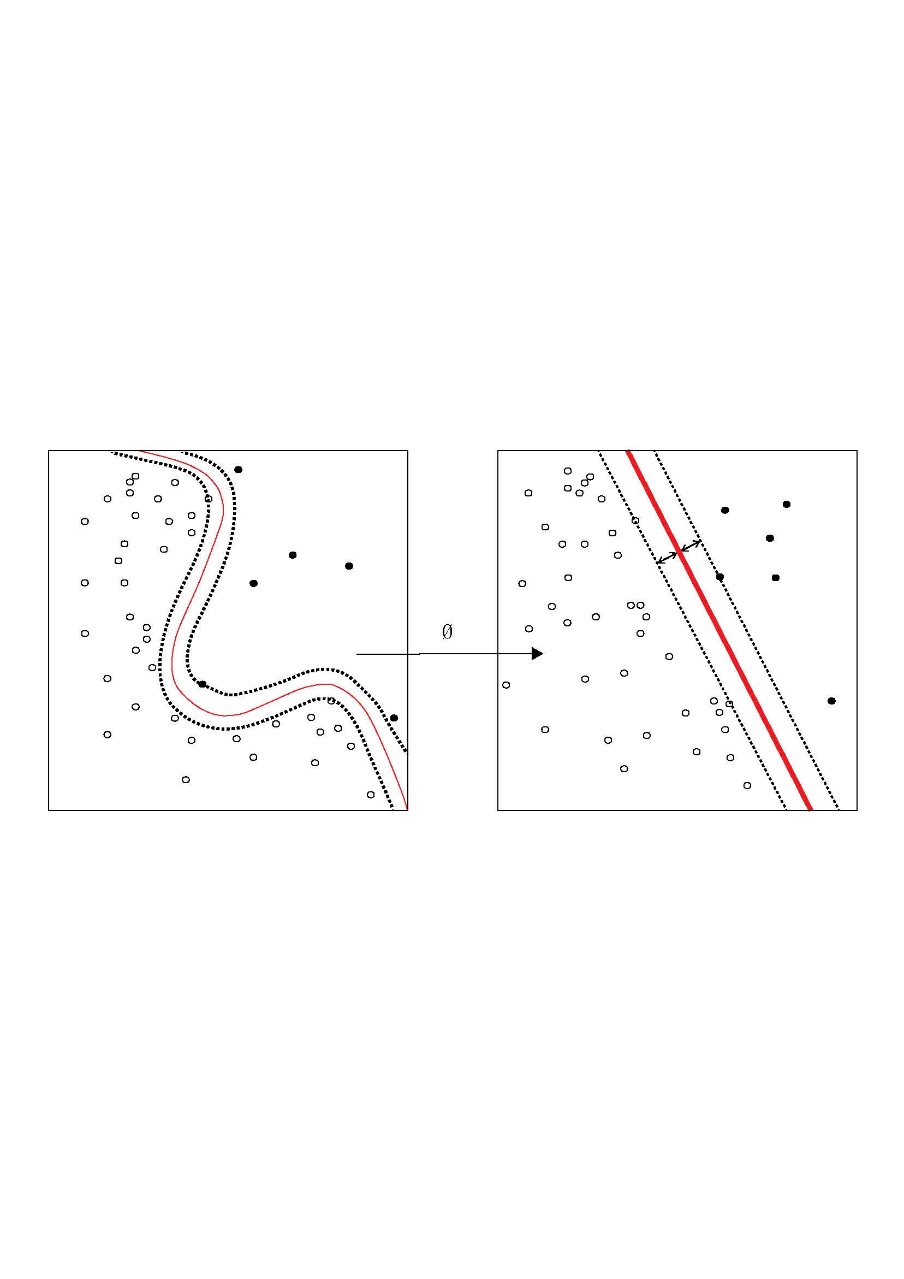


图2-10 支持向量机

### 2.5.3 决策树

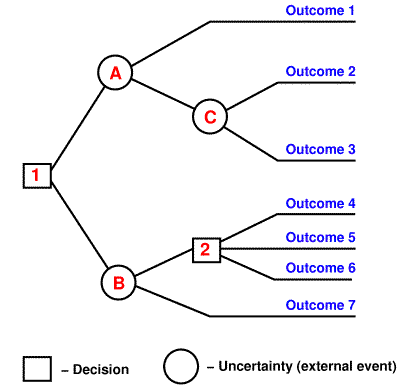
决策树（decision tree）是有监督学习的分类方法。决策树模型是一个预测模型，结构呈现树形，表示基于特征对实例进行分类的过程。决策树首先在给定数据集中寻找最优特征，然后在最优特征中寻找最优值，根据这个最优值将数据集分为两个子数据集，在子特征集中重复上述步骤，直到达到某种条件。决策树学习通常包括3个步骤:特征选择、决策树的生成和决策树的修剪。决策树分类模型中，C4.5算法产生的分类规则易于理解，准确率较高。

图2-11 决策树

# 第3章　语音信号情感识别

## 3.1 实验设计

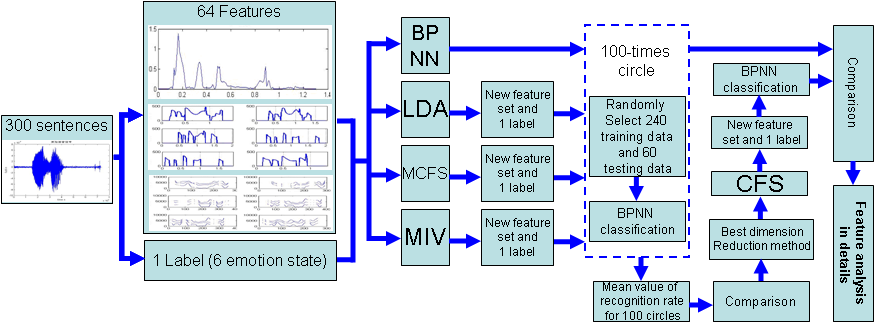
音频信号分为语音信号和音乐信号，音频信号情感的划分如上介绍，为了研究音频信号的情感识别，对于语音信号的情感识别利用离散化标签的数据库，即CASIA汉语情感语料库。CASIA汉语情感语料库由中国科学院自动化研究所录制，由4位录音人(2男2女)在纯净录音环境下(信噪比约为35db)分别在6类不同情感下(高兴、悲哀、生气、惊吓、难过、中性)对500句文本进行的演绎得到[13]。实验首先提取语音库中某女生包含喜怒哀乐平静惊讶六种感情，每种感情50句语音，共300句语音作为数据库。具体实验设计流程如图所示：

图3-1 语音情感识别流程图

(1) 首先提取每一句语音的64维特征统计向量。再将每种语音感情对应一类标号作为类别标识，随机打乱300句语音特征的排列顺序。最后从中随机抽取240组数据作为训练数据，60组作为预测数据，训练64维度输入，6维输出，隐藏层为65的BP神经网络并统计正确率。

(2) 将上述得到的64维特征向量按照MIV的方式找出对训练结果影响最大的特征向量，并组成新的特征向量组，训练BP神经网络并统计识别正确率，同上述识别正确率比较。在提取MIV时，采用每个感情特征的50组作为样本，其中随机提取45组数据作为训练样本，剩下5组数据作为预计输出，多次测量求出出现次数最多的特征向量，再从大到小排序,此顺序即为训练结果与原特征向量相关性大小排列顺序。

(3) 用数据降维的方法线性判别分析(Linear Discriminant Analysis,LDA)以及多类簇特征选择(Multi Cluster Feature Selection,MCFS)的方法降维原64维统计向量，训练BP神经网络并统计识别正确率，同上述识别正确率比较。

(4)由于降维后的数据仍然含有冗余成分，用互相关特征选择 (Correlation-Based Feature Selection，CFS)方法对得到的数据进行筛选分析，得到可以对识别贡献最大的特征向量，并对选择的特征向量进行分析，研究选择后的向量与识别结果的相关性。

(5)验证算法提取的正确性。即提取语音库中某男生的语音情感特征，比较其经上述提取后和未降维语音识别正确率。

## 3.2 选取特征

语音信号的特征与音乐信号的选取略有不同，因为音乐信号除了基本的语音信号特征，还包括诸如，韵律，节奏等特征，故语音信号的特征选取和音乐信号的选取采用不同的方法，音乐刺激同时为脑电信号的一种常用刺激信号，这将在下一章进行介绍。

语音信号特征选取：

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。

## 3.3 实验结果分析

### 3.3.1 MIV降维

通过MIV方法得到了共32维与原向量相关的向量。其中，第11,22,21,16,13，14,26,24维特征向量，即基频最大值，基频均值，基频二阶差分最大最小值，基频一阶差分最大值，基频中值，基频范围，基频方差，与识别结果相关性最大。除了基频特性外相关性比较大的特征向量还有41，第二共振峰方差；46，第三共振峰方差；36，第一共振峰的方差；32，第一共振峰最大值；27,发音帧数；28,不发音帧数；30，发音帧数和总帧数之比。经分析，发现，MIV降维后基频向量对识别结果起了主要影响，共振峰的特征值和语音帧数特征对识别结果起了重要的影响。

表3-1 MIV选取的32维向量及所占百分比

|  |  |  |  |
| --- | --- | --- | --- |
| 维度 | 百分比 | 维度 | 百分比 |
| 11 | 9.11% | 25 | 2.06% |
| 22 | 6.87% | 20 | 2.06% |
| 21 | 6.53% | 15 | 2.06% |
| 16 | 6.19% | 28 | 1.89% |
| 13 | 6.19% | 36 | 1.72% |
| 14 | 6.01% | 32 | 1.03% |
| 26 | 5.84% | 30 | 1.03% |
| 24 | 5.82% | 56 | 0.68% |
| 19 | 5.82% | 51 | 0.52% |
| 23 | 5.33% | 12 | 0.52% |
| 27 | 4.98% | 64 | 0.17% |
| 18 | 4.98% | 38 | 0.17% |
| 17 | 4.98% | 37 | 0.17% |
| 41 | 2.41% | 35 | 0.17% |
| 46 | 2.34% | 34 | 0.17% |
| 61 | 2.06% | 33 | 0.17% |

表3-2 未降维和MIV降维后识别结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 算法 | 识别结果 | | | | | |
| 喜 | 怒 | 惊 | 悲 | 恐 | 平静 |
| 未降维64维 | 80.49% | 80.36% | 77.34% | 93.48% | 74.49% | 86.58% |
| MIV降维32维 | 80.85% | 89.83% | 77.08% | 89.78% | 84.32% | 89.61% |

由表中测试结果知，用MIV降维取前32维后的特征向量训练BP神经网络所得到的平均识别率为85.21%，要比降维前识别率82.13%平均提高了3.08%。这说明，用MIV方法提取的特征向量基本上能保持原64维向量的特征。

### 3.3.2 LDA以及MCFS的方法降维

表3-3 MIV降维和LDA，MCFS降维至32维识别结果

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 算法 | 识别结果 | | | | | | | | | | | |
| 喜 | | 怒 | | 惊 | | 悲 | | 恐 | | 平静 | |
| LDA 降维至32维 | | 83.00% | | 88.12% | | 78.30% | | 67.94% | | 89.60% | | 89.63% |
| MCFS降维至32维 | | 91.89% | | 88.39% | | 91.22% | | 76.42% | | 46.05% | | 73.45% |
| MIV 降维至32维 | | 80.85% | | 89.83% | | 77.08% | | 89.78% | | 84.32% | | 89.61% |

由表可知用MIV算法提取前32维特征向量所得的语音情感平均识别率比 LDA和MCFS算法降维后得到的语音情感识别率分别提高了2.48%和7.35%。

### 3.3.4 用CFS方法降维

由表2可以看出，用MIV方法降维后的特征，每一维对与识别的贡献率不一。其中前23维所占比重皆大于1%，包括15维基频向量，5维共振峰向量，3维发音帧数向量。若单独用15维基频向量做特征向量进行实验，则识别率为 88.05%，用前23维，识别率为 88.90%。可见，这主要的23维向量中仍有冗余。为了提高识别率，减少冗余，包含大部分特征向量，本文用CFS的方法将排名前15的基频特征向量分别降维至3-8维的向量，将5维共振峰向量降维至2-3维，加之3维发音帧数向量(27,28,30)，选取组成8-14的维的新特征向量，验证具体特征向量与识别率的相关性。

表3-4 CFS法选择基频特征

|  |  |
| --- | --- |
| 基频共选取几维 | 所选取的维数特征 |
| 3 | 21,22,25 |
| 4 | 11,22,20,26 |
| 5 | 18,19,20,23,24 |
| 6 | 14,15,17,22,23,24 |
| 7 | 13,14,15,19,20,25,26 |
| 8 | 15,17,18,19,20,21,24,25 |

表3-5 CFS法选择共振峰特征

|  |  |
| --- | --- |
| 共振峰共选取几维 | 所选取的维数特征 |
| 2 | 41,36 |
| 3 | 41,36,32 |

表3-6 不同维度组合后识别结果

|  |  |  |  |
| --- | --- | --- | --- |
| 维度 | 识别率 | 维度 | 识别率 |
| 3+3+2 | 88.70% | 3+3+3 | 89.68% |
| 4+3+2 | 89.28% | 4+3+3 | 89.82% |
| 5+3+2 | 90.32% | 5+3+3 | 90.46% |
| 6+3+2 | 90.56% | 6+3+3 | 90.67% |
| 7+3+2 | 90.62% | 7+3+3 | 91.15% |
| 8+3+2 | 91.61% | 8+3+3 | 90.85% |

可见，向量降维后与原识别率差别并不是很大，为了使尽量减小特征向量选取维度，同时不影响正确率，则将基频向量降到5-8维，共振峰向量降到2维(36,41)比较合适。

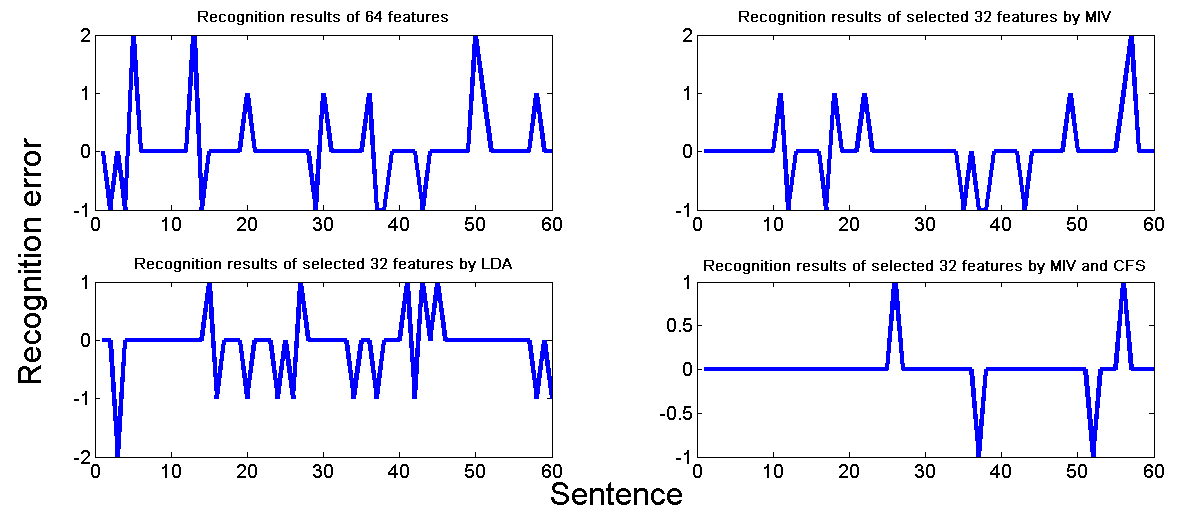
****

图3-2 一次实验降维前后识别分类误差

在图4中，对其中一次实验60个样本的分类错误进行统计，未降维前分类有13个识别错误；降维取前32维，分类错误有10个；用LDA降维至32维，分类错误有20个；降维改进后13维，分类错误仅有4个且误差较小。证明了新选取特征向量的有效性。

## 3.4 特征分析

由以上的研究可知，不同的特征对语音情感的区分度是不同的，为了分析不同特征同不同语音情感之间的关系，取表5出现频次最多的第15，19，20，22，24，27，28，30，36，41分析基频特征同语音情感之间的关系。

**E:\aa0011\语音识别\提取\eng-pict\new\new2\15wei.tif15wei E:\aa0011\语音识别\提取\eng-pict\new\new2\19wei.tif19wei**

(a) (b)

**E:\aa0011\语音识别\提取\eng-pict\new\new2\20wei.tif20wei E:\aa0011\语音识别\提取\eng-pict\new\new2\22wei.tif22wei**

(c) (d)

**E:\aa0011\语音识别\提取\eng-pict\new\new2\24wei.tif24wei E:\aa0011\语音识别\提取\eng-pict\new\new2\27wei.tif27wei**

(e) (f)

**E:\aa0011\语音识别\提取\eng-pict\new\new2\28wei.tif28wei E:\aa0011\语音识别\提取\eng-pict\new\new2\30wei.tif30wei**

(g) (h)

**E:\aa0011\语音识别\提取\eng-pict\new\new2\36-2wei.tif36-2weiE:\aa0011\语音识别\提取\eng-pict\new\new2\41wei.tif41wei**

(j) (j)

图3-3 选择特征的频率分布直方图（a）基频方差；（b）基频一阶差分中值；（c）基频一阶差分的方差（d）基频二阶差分最小值；（e）基频二阶差分中值；（f）发音帧帧数；（g）不发音帧帧数；（h）发音帧占比；（i）第一共振峰的方差；（j）第二个共振峰的方差

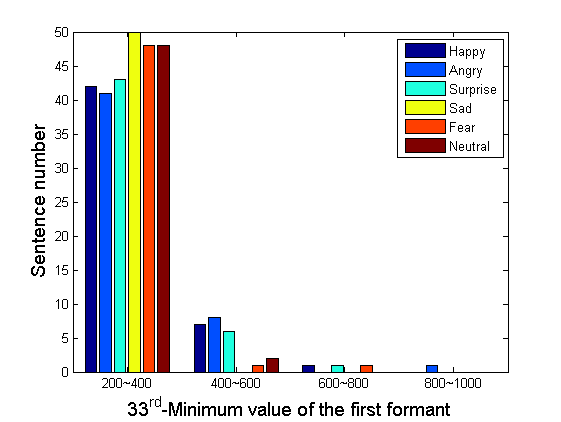
为直观研究所选的特征，可以对每个特征进行直方图分布处理，在图3-3中，直方图水平轴是特征值，竖轴是句子的个数。它的内涵是：对于一个特定的特征，有多少具有特定情感的句子同时分布在在一个数据区域。例如，在图5（a）中，表示选择的第十五个特征，基频方差的情感状态分布。这个特性的值在25到50的区域时，只有1句属于“快乐”， 1句属于“生气”，5句属于“平静”，没有“惊喜”，但有超过23句话属于“悲伤”以及24 句属于“恐惧”。因此，如果一个句子的基音频率在25到50之间时，这句话的情感状态更属于“悲伤”或“恐惧”是比较合理的。另一个例子，如图5（b）所示，如果基频的一阶差分中值小于350，这个句子的情感状态更可能是“快乐”或者“愤怒”。

宏观分析，由选择的基频特征即图3-3（a）（b）（c）（d）（e）上看，喜悦，愤怒，惊讶与悲伤，恐惧，平静可以很好的区分：悲伤，恐惧，平静的方差和一阶差分方差较小，一阶差分中值较大，而喜悦，愤怒，惊讶的方差较大且一阶差分方差较大，一阶差分中值较小，这说明前者基频平缓且变化平缓而后者基频尖锐且变化尖锐；而前者二阶差分大而后者二阶差分小，说明前者变化的速率大而后者变化速率小，其中，平静的基频和一二阶差分都在固定的值左右。喜悦，愤怒和惊讶有着相似的基频特性，悲伤和恐惧有着相似的基频特性。在实际的发音中，悲伤，恐惧，平静的发音频率比较稳定，而喜悦，愤怒，惊讶的发音频率变化较大，这符合上述结果。

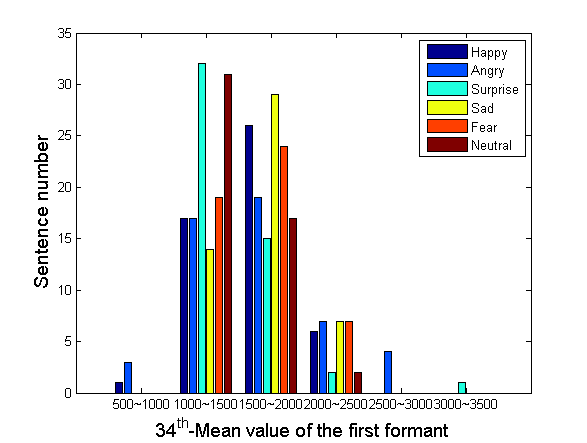
对于语音的发音帧数的特征，不同的情感也有着明显的差异。由上图3-3（f）（g）（h）知；恐惧和平静发音帧数较少，不发音帧数较多，发音帧数与总帧数之比较小；愤怒发音帧数较多不发音帧数较少，发音帧数与总帧数比大多在60%到80%之间。

共振峰特征上，喜悦的第一二共振峰方差变化较小，共振峰平缓温和；悲伤，恐惧，平静的第一二共振峰方差较大，变化剧烈。

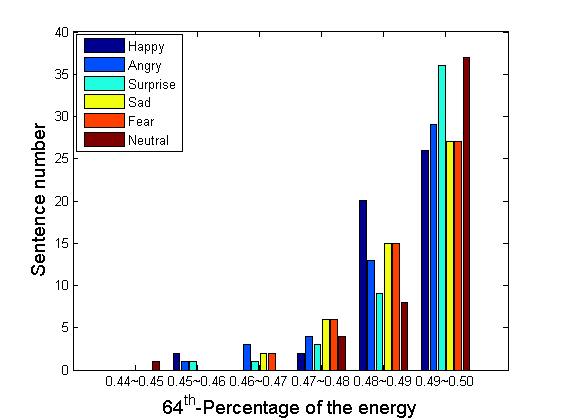
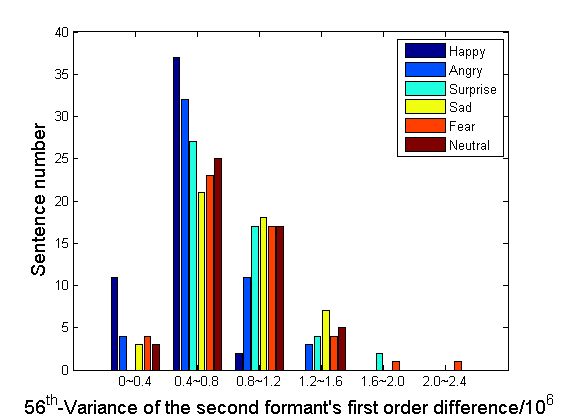
可见，已选取的特征与语音情感的相关性很大，为了证明未入选特征去相关性不大，下面列举了六个未选取特征的频率分布直方图。

**E:\aa0011\语音识别\提取\eng-pict\new\new2\12-2wei.tif12-2wei**

(a) (b)

**E:\aa0011\语音识别\提取\eng-pict\new\new2\51-2wei.tif51-2wei**

(c) (d)

****

(e) (f)

图3-4 未入选特征的频率分布直方图（a）基频最小值；（b）第一共振峰最小值；（c）第一共振峰平均值；（d）第一共振峰一阶差分；（e）第二共振峰一阶差分的方差；（f）4KHz 以上能量谱的百分比

由图3-4，六种感情分布相对均匀，没有哪种情感的特征分类比较明显，可见其对语音情感的识别贡献不大。

## 3.5 推广验证

为了验证经过MIV方法提取和CFS方法改进后的特征向量对于语音情感特征提取的有效性，设计了对照实验。由于女声和男声在发音的某些特征方面有所差异，例如女声的基频较高，如果上述提取的女声情感语音特征对于男声同样适用，这就证明了提取的特征是有一定普适性的。

表3-7应用于男声的识别结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 算法 | 识别结果 | | | | | |
| 喜 | 怒 | 惊 | 悲 | 恐 | 平静 |
| 未降维64维 | 79.63% | 64.99% | 79.98% | 62.68% | 48.36% | 79.06% |
| MIV 降维至32维 | 82.12% | 64.94% | 83.27% | 73.61% | 26.77% | 91.82% |
| MIV+CFS改进后13维 | 94.60% | 62.54% | 94.54% | 77.46% | 51.30% | 79.98% |

由上述识别结果可知，用MIV对原特征降维平均识别率提高了1.3%，证明了MIV方法提取的特征向量对减少冗余特征有一定的作用。用CFS方法对提取的特征向量改进，比原特征向量平均识别正确率提高了7.62%，证明了CFS方法减少了选取特征的冗余，所选取的向量对语音情感的识别比较有效。

## 3.6 本章小结

语音情感识别是人工智能的一个重要组成部分，已经受到广泛关注，越来越多的成果也相继出现。本章通过MIV和CFS算法优化了输入语音特征，再通过然后结合BP神经网络识别语音情感。实验结果证明，MIV和CFS算法可以有效的提取语音输入特征主要成分，语音信号的基频特征和共振峰特征对于语音情感识别具有重要贡献。利用MIV方法对64维原数据降维，取32维，语音情感识别率提高了3.08%，为了有效的提取情感特征向量，对MIV降维后32维数据用CFS的方法进行改进，取其中13维特征，语音情感识别率比原32维提高了9.48%。应用得到的特征向量进行说话人无关实验，结果证明了提取的10维特征向量具有一定的普适性。然而，由于数据样本不足，本文不得已采取多次测量求均值的方法，后续可以添加数据库数据，进行更好的研究。

# 4　音乐情感识别

## 4.1数据库

由于音乐信号的典型特征与语音信号略有不同，音乐信号情感分析数据库来源于1000 Songs Database，1000 Songs Database是一类由用于情感分析的歌曲数据库，歌曲的采样频率为44100赫兹，每首音乐的时长为45秒，且全部歌曲标注了效价维和激活维的均值和方差，故可以用二维情感模型对其进行分类[69]。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[13]。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。表1说明图中所示的四类感情可以视作无关的四种不同情感，从而进行分类验证。

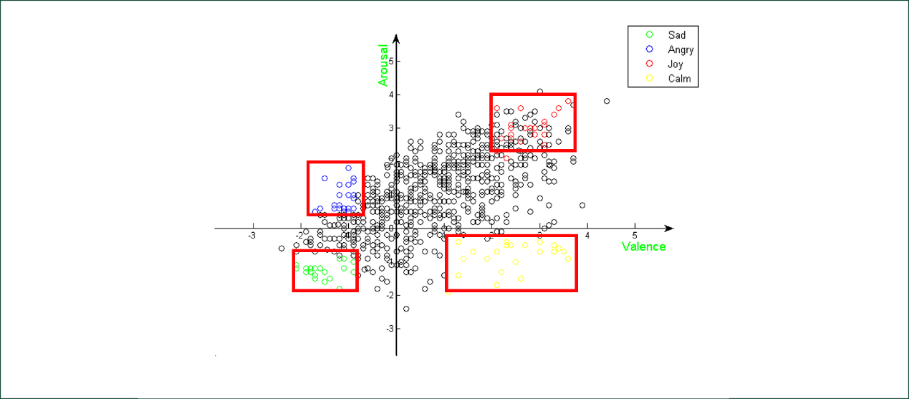


图4-1 音乐分类

表4-1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 |

## 4.2 音乐信号特征选取

MIRtoolbox工具包是基于Matlab平台的一种声音特征工具包。Mirtoolbox能够提取多种音频提取特征和统计特征[70，71]，包括时域特征、波形特征、节律、频谱、音质及声调等。实验提取了音频信号时域的均方差，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、峰值、质心等138个特征（表4-2）。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4-2 MIRtoolbox工具包提取的特征

|  |  |  |
| --- | --- | --- |
| 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 |

## 4.3 实验设计

系统提取的特征使用GA+CFS降维，降维的具体步骤如下：

a. 编码。将编码长度设定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，表示对应的特征被选中并可以进行建模；当染色体的值为0时，表示该特征未被选用，不能进行建模。

b. 设定初始群体和终止条件。本系统种群大小设定为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。

c. 设定适应度函数。适应度函数衡量种群中不同个体在遗传进化中能够达到或者相对近似解的距离，适应度函数评价得分高的个体则遗传到下一代的概率大。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。

d. 遗传算子的确定。通过轮盘赌方法、单点随机交叉方法和基因位变异方法，将交叉概率设定为0.8，变异概率取0.2来确定遗传算子。

e. 原始种群经过选择、交叉、变异计算之后选择比较优秀的特征集组成下一代种群。

f. 当循环结束时终止运算。最终得到最大适应度函数和最优特征子集。

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，选择以CFS方法作为适应度函数来评价选择建模特征。按照以上所述步骤和所选参数，做其中一次的实验适应度进化曲线，用来评价CFS改进的GA算法是否能很好的接近最优解。



图4-2 CFS作为评价准则适应度函数进化曲线

由图4-2知，在进化代数小于20时时，平均适应度值不能很好地接近最佳适应度值，当进化代数超过20代后，平均适应度值能很好的接近最佳适应度值，说明种群中每个个体都在最优解周围，即所选特征近似最优，故可以用CFS作为GA算法的适应度函数。

## 4.4 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。

利用GA+CFS算法，共筛选出8维特征，他们分别是：节奏特征的时间均值、时间斜率，谱特征的偏度均值、谱熵均值，音色特征过零率的周期幅度、最低能量的均值，音调模式的周期熵和谐波变化均值。

分类算法采用SVM和BP神经网络算法进行验证比较，其中SVM算法调用林智仁教授的libSVM包。实验采用100次十折交叉验证方法进行分类，验证结果见表4-3、4-4，对分类器的评价用召回率表征详见表4-5，分类效果如图4-3、4-4所示。

表4-3 SVM验证未降维和GA+CFS降维后识别结果

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | 60.00 | 71.43 | 60.00 | 100 | 81.90 | 5.146 |
| LCGA+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 | 98.10 | 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 | 99.60 | 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 |
| 音乐未降维 | 88.64 | 90.65 |  | 85.06 |  | 90.25 |  | 9.445 |
| 音乐降维 | 86.45 | 91.24 |  | 84.69 |  | 87.79 |  | 4.548 |

表4-4 BP验证未降维和GA+CFS降维后识别结果

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| 音乐未降维 | 91.23 | 89.15 |  | 85.64 |  | 90.60 |  | 81.455 |
| 音乐降维 | 93.54 | 92.56 |  | 91.38 |  | 98.76 |  | 10.450 |

表4-5 不同分类器的召回率

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 算法 | 召回率 | | | | | | |
| 喜/% | 怒/% | 惊/% | 悲/% | 恐/% | 平静/% | 平均/% |
| 语音SVM未降维 | 71.12 | 85.45 | 81.65 | 95.90 | 92.76 | 70.25 | 82.86 |
| 语音SVM降维 | 72.88 | 90.23 | 88.62 | 87.83 | 84.87 | 85.91 | 85.06 |
| 语音BP未降维 | 77.23 | 77.33 | 90.23 | 90.55 | 92.70 | 94.13 | 87.03 |
| 语音BP降维 | 76.75 | 88.36 | 100 | 92.33 | 100 | 100 | 92.91 |
| 音乐SVM未降维 | 92.14 | 90.85 |  | 95.25 |  | 94.23 | 93.12 |
| 音乐SVM降维 | 89.64 | 95.12 |  | 97.50 |  | 93.66 | 93.98 |
| 音乐BP未降维 | 88.63 | 84.56 |  | 92.43 |  | 90.14 | 88.94 |
| 音乐BP降维 | 85.95 | 86.13 |  | 93.36 |  | 92.22 | 89.42 |



（a） （b）

图4-3 SVM验证降维前后测试样本分类（a）SVM验证未降维测试集样本分类

（b）SVM验证GA+CFS降维测试集样本分类



（a） （b）

图4-4 BP验证降维前后测试样本分类（a）BP验证未降维测试集样本分类

（b）BP验证GA+CFS降维测试集样本分类

由表4-3可见，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约3s，建模速度提升了。由图4-3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。

由表4-4可见，对于BP来说，降维后的识别率比降维前均高出约20%，且降维后建模时间减少约70s。由图4-4某次训练降维前后样本分类直观可见，降维前的错误分类数为8个，降维后错误分类数为4个，降维后比降维前的错误分类率大幅减少。

由表4-5召回率来看，BP神经网络和SVM都有很好的分类性能。SVM相对于BP神经网络有较快的建模速度，而BP神经网络相对于SVM有较好的识别效果，在识别中若要求准确率和召回率可以使用BP分类器，而要求建模速度可以使用SVM分类器。

## 4.5 特征分析

由以上的识别验证可知，不同的特征组合对识别的结果影响不同，为了分析选择出的特征对不同情感的音乐分类的贡献，找出比较有效的特征，下面对选择的8个音乐特征进行了分析。



图4-5 选择的节奏特征频率分布直方图

节奏的时间均值特性，愤怒情绪集中在较小数值，喜悦和平静特性分布于相对分布于中部，而悲伤特性分布于较大数值部分。时间斜率特性愤怒情绪分布于较小数值部分。这与悲伤音乐的节奏较舒缓，而间隔时间较长，愤怒、喜悦音乐节奏感强，变化频率大有关。



图4-6 选择的谱特征频率分布直方图

对于谱偏度均值特征，愤怒、悲伤、喜悦、平静情感分布依次由小到大分布，具体说来，愤怒情感的谱偏度均值多分布于0-2区间，悲伤分布于2-4区间，喜悦分布于4-6区间，而平静分布于6-10区间。谱熵均值特征，喜悦、平静、悲伤、愤怒感情由小到大分布，喜悦和平静多分布于0.5-0.6区间，悲伤情感多落在0.6附近，如果歌曲谱熵均值大于0.7，则很可能在表达愤怒情绪。谱特征的分布比较明显，这与平静，喜悦等情感的音乐比较轻缓柔和，而悲伤，愤怒的音乐多慷慨激昂，能量较大相关。



图4-7 选择的音色特征频率分布直方图

对于音色特征，其最低能量均值愤怒和悲伤分布于较小数值部分，若最低能量在0.45到0.50区间，则可能是愤怒的情绪，而最低能量若在0.50到0.55区间，则可能表达悲伤的情感。



图4-8 选择的音调和谐波特征频率分布直方图

对于音调模式的周期熵，悲伤情感和愤怒情感能很好的区分，悲伤感情和平静感情不易区分，但其明显不同于其他两种情感，音调模式在0.45到0.50区间，则音乐大概率在表达悲伤的情感，若音乐模式的值大于0.50，则音乐大概率在表示愤怒的情感。音调是音乐基音的变化模式，音调能很好的区分这几种音乐，这可能与不同情感音乐发音的基音频率的不同有关。对于谐波变化均值，愤怒情感的表现比较明显，其大概率分布在0.10-0.12数值左右。

由以上选择出的情绪频数直方图来看，选择出的特征大部分能很好的区分出不同的音乐情感，特别是谱特征和音调特征，故可以认为选择出的特征与不同的音乐的情感有一定的对应关系。

# 第5章 音频信号引起的脑电信号分析

## 5.1方法设计

实验步骤如图1所示：

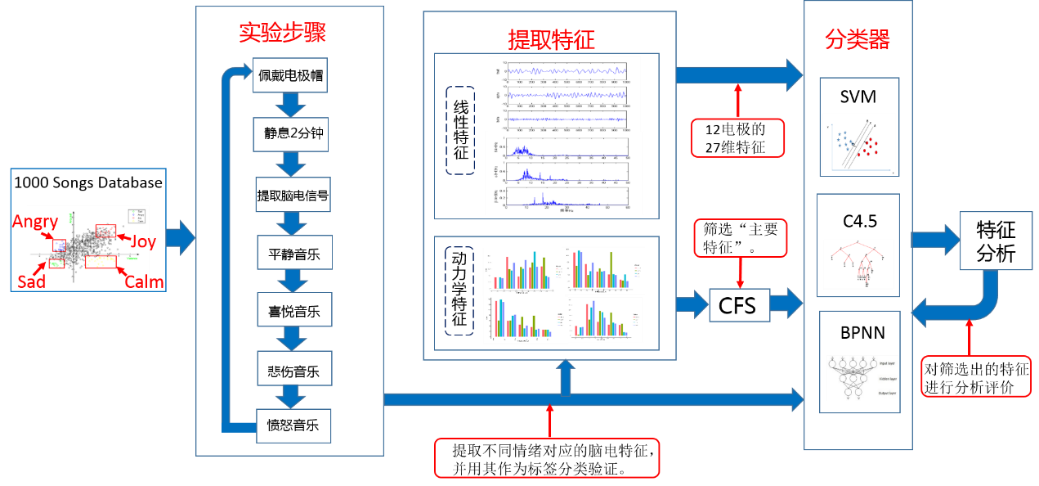


图5-1 音频信号刺激脑电实验流程图

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。

## 5.2 脑电采集步骤

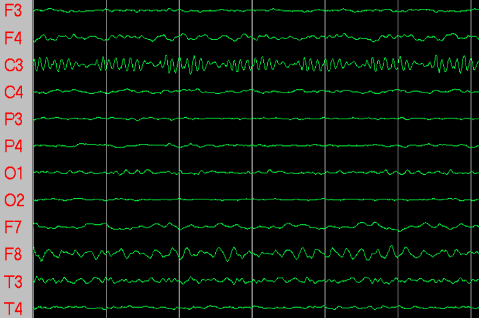


图5-2 脑电采集设备及采集的信号

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，每类感情播放由第四章选择出的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。

## 5.3数据分析

### 5.3.1 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[72~74]。后来非线性动力学得到了很大的发展，人们逐渐认识到大脑是一个高维、复杂的混沌系统，非线性动力学特征也被广泛应用与脑电研究[57,75]。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。

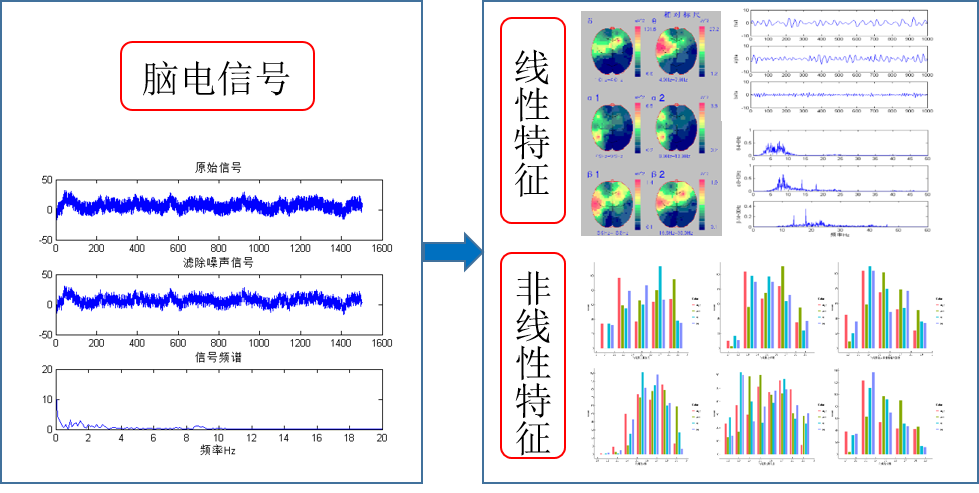


图5-3 脑电信号提取的特征

### 5.3.2 分类器验证

本文利用典型的机器学习方法即SVM，C4.5，BP神经网络三种方式验证降维后的识别率，并计算召回率评价不同分类器分类性能。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5-4 SVM，C4.5，BP验证降维前后的识别率

表5-1 SVM，C4.5，BP验证降维前后的召回率

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 召回率 | 喜/% | 怒/% | 悲/% | 平静/% | 平均/% |
| SVM未降维 | 72.84 | 75.44 | 85.70 | 78.28 | 78.07 |
| SVM降维 | 70.64 | 81.78 | 88.82 | 80.92 | 80.54 |
| C4.5未降维 | 80.25 | 83.39 | 91.52 | 93.17 | 87.08 |
| C4.5降维 | 80.77 | 86.54 | 93.41 | 89.15 | 87.47 |
| BP未降维 | 89.13 | 79.85 | 76.28 | 85.23 | 82.63 |
| BP降维 | 92.12 | 88.16 | 72.60 | 80.66 | 83.39 |

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，C4.5和BP的召回率都比较高，BP和SVM的识别率比较高。而对12个脑电电极进行评价，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。

由统计特征也可分析不同电极的特征之间的相关性，故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5-2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，故可以选择T4和Pz电极作为音乐引发的脑电情绪的分析研究。

表5-2 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 |

### 5.3.3 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5-5所示。



图5-5 T3，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5-6所示。



图5-6 T3，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。

## 5.4特征分析

### 5.4.1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。

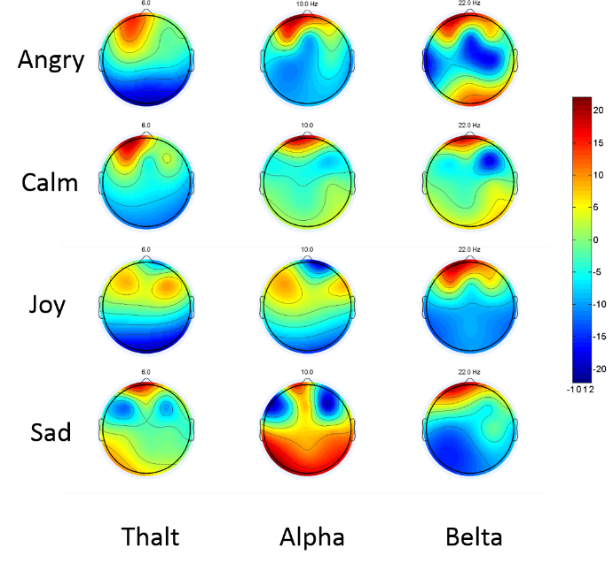


图5-7 不同波段四种感情音乐受众脑地形图

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5-7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。

### 5.4.2非线性特征

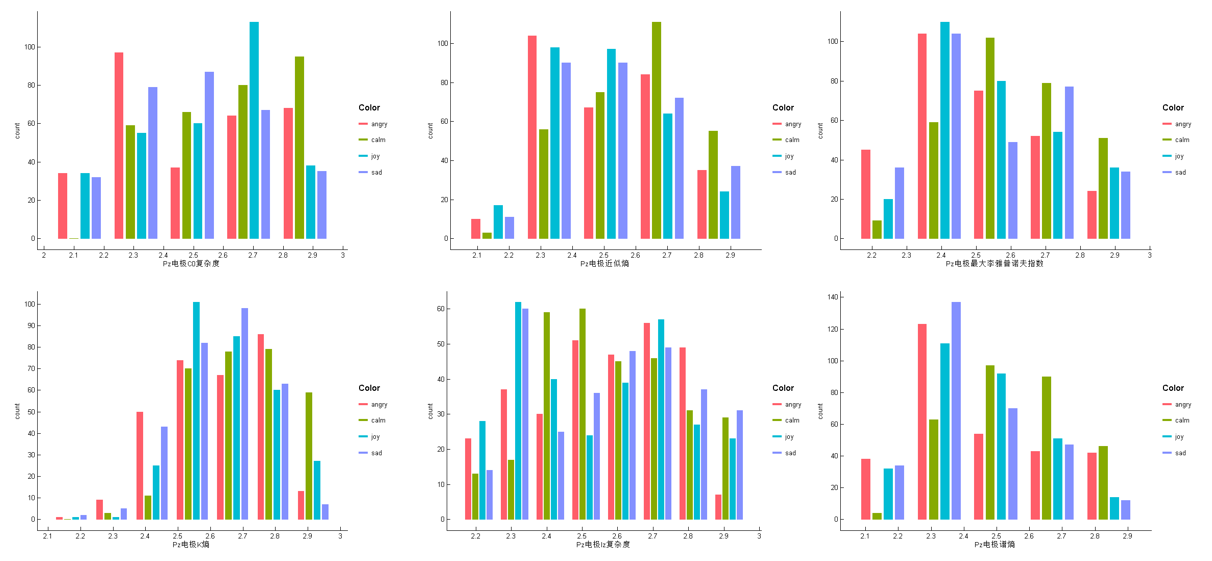
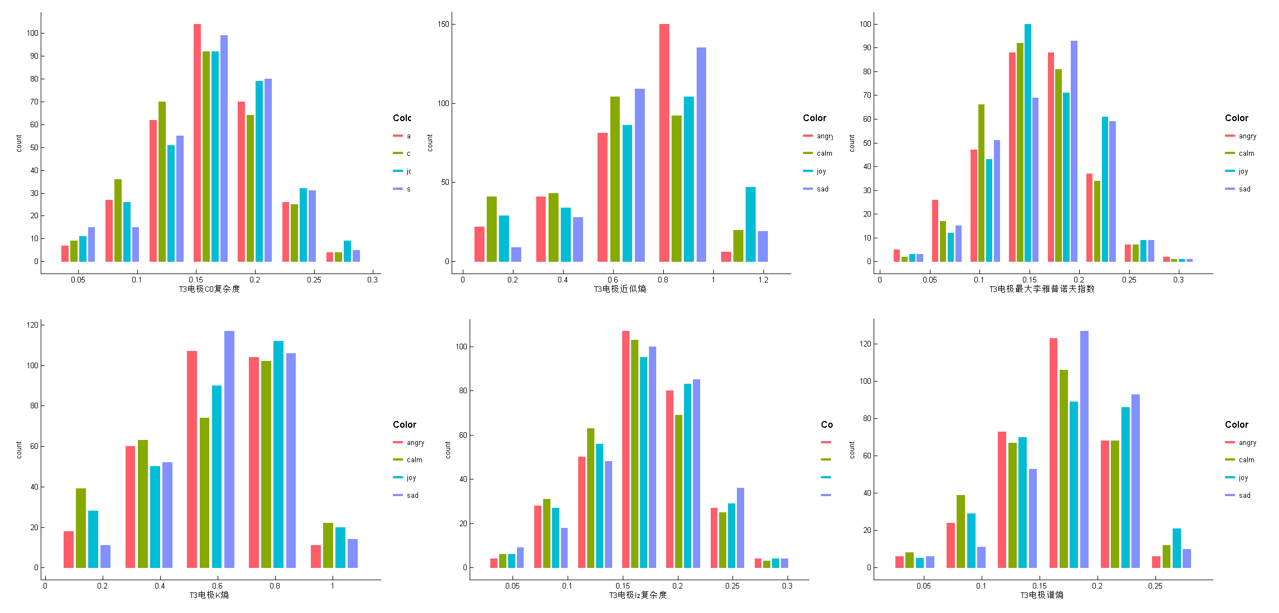
为了对选择的非线性特征进行分析，本文对选择出的特征做频率分布直方图，以直观的分析研究不同的情绪同非线性特征之间的关系。

图5-8 Pz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5-9 T3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。

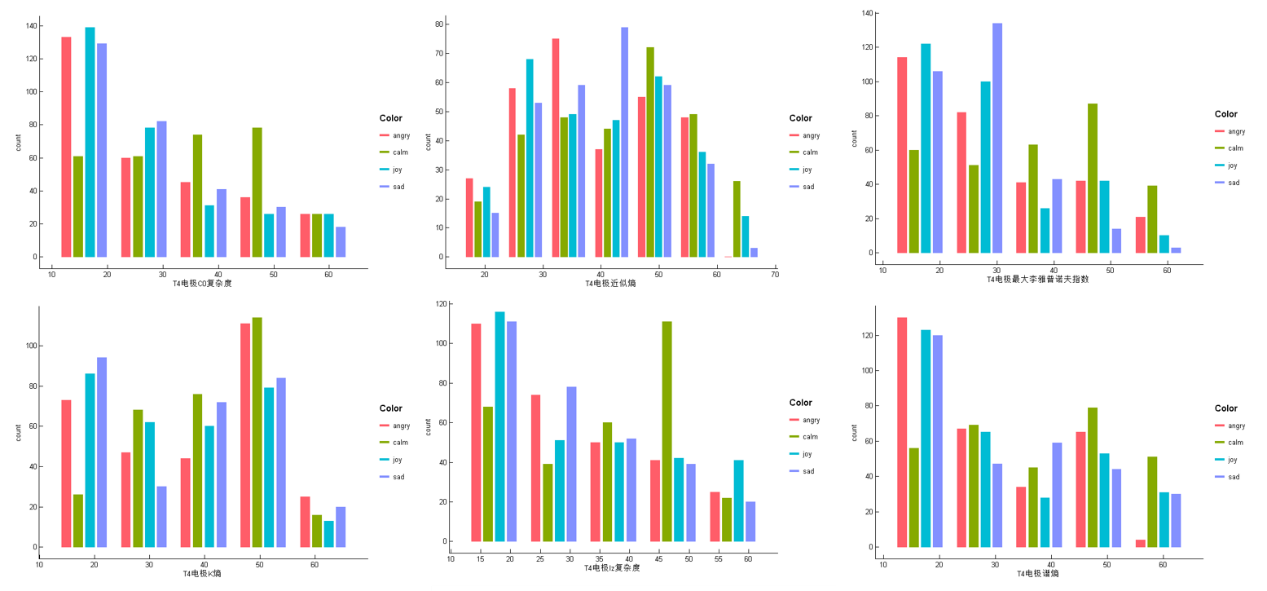


图5-10 T4电极非线性特征频数分布

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。

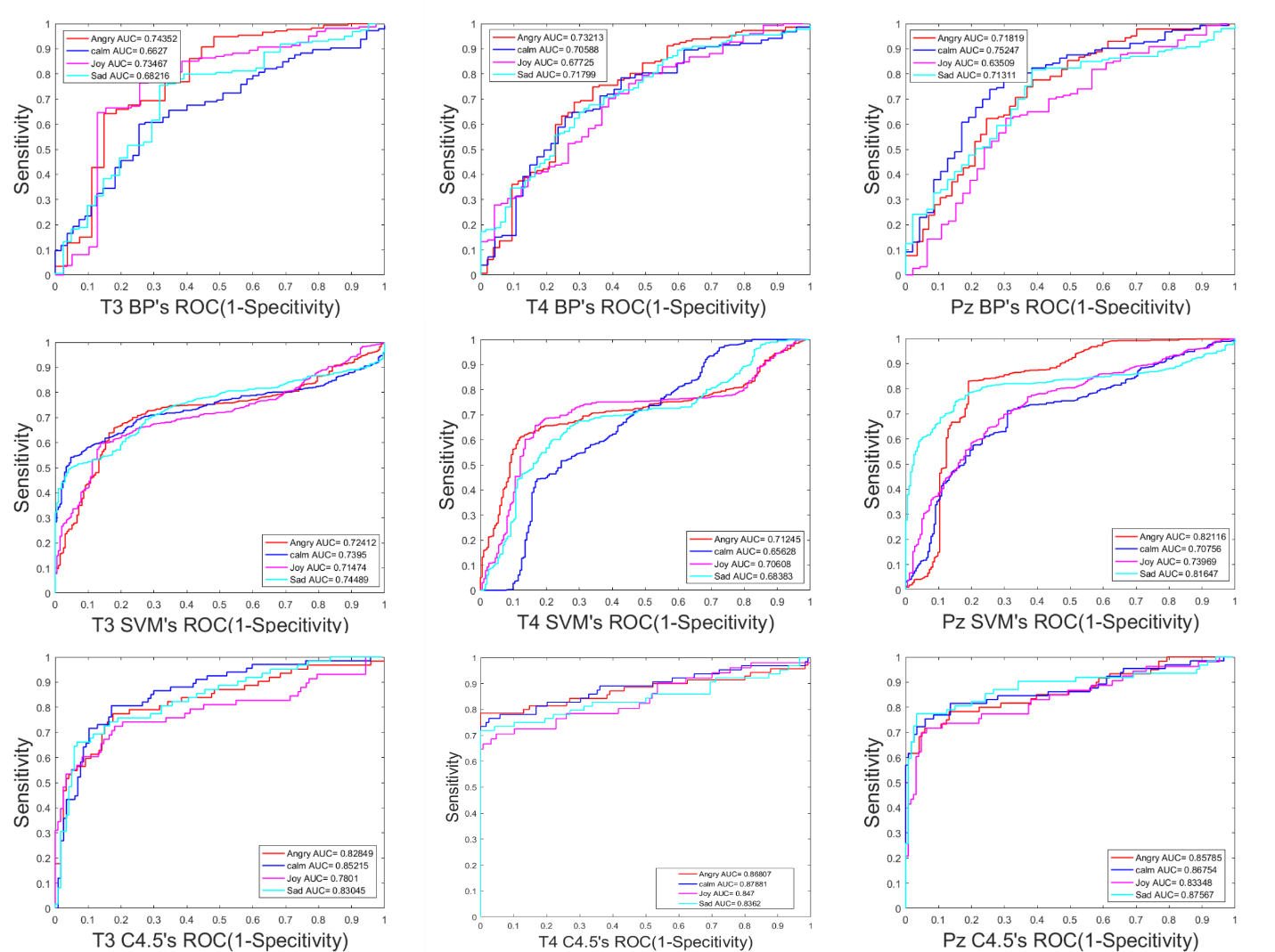
### 5.4.3 检验

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5-11 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。

ROC曲线和AUC是评价分类器分类好坏的标准，一般来说，ROC曲线越趋近于左上角，AUC的值越大，表明分类器的性能越好。为了证明所用分类器的分类效果，实验给出了三种不同分类器在不同电极上的ROC曲线和AUC的值。由图5-12，三种费雷起皆具有比较好的分类效果，其中，C4.5分类器的效果要略好于其他两种分类器。

图5-12 不同分类器的ROC曲线和AUC值

## 5.5本章小结

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，C4.5算法对于脑电情绪分类比较有效。

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6章　音频信号和脑电信号关联分析

## 6.1 方法设计

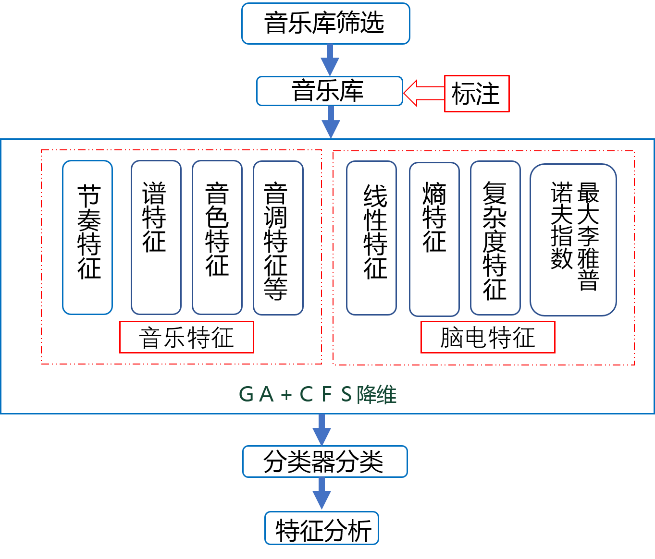


图6-1 音频脑电信号关联分析流程图

由第四章知，与音乐情感识别相关性比较大的特征有节奏特征的时间均值、时间斜率；谱特征的偏度均值、谱熵均值；音色特征过零率的周期幅度、最低能量的均值；音调模式的周期熵和谐波变化均值。故选择用这8维音乐特征同选择出来的脑电信号作关联分析。由第五章选择而出的12维脑电特征包括：6维线性特征和6维非线性特征。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）；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）。这20维特征可能对音乐情绪以及其引发的脑电特征的分类最有代表性。

## 6.2 特征选择

由于音乐特征和脑电特征是分别对应于音乐特征和脑电特征选择而出，且提取特征共20维，冗余特征相对较多，于是实验对上述选择出的20维特征进行降维。降维方法使用特征选择的方法，即GA+CFS的方法，利用上述降维方法分别对于每一样本进行特征选择，最后统计所有样本中选择出的特征，某个特征被选择最多，就证明这个特征比较能代表全局特征。不同特征频数此时如图6-2所示。



图6-2 CFS+GA选取特征的频数分布

由上图可以看到经CFS改进的GA算法筛选出的特征，在前8维音乐特征中，第3维谱特征的偏度均值，和第7维的音调周期熵最能代表总体特征。由第三章的特征分析可知，这两个维度同样能比较直观的区分出不同的音乐情感。在脑电特征，对于T3电极，比较有效的特征有第9维alpha波中心频率，第13维dalt波中心频率；对于T4电极，比较有效的特征有第17维最大李雅普诺夫指数，第18维C0复杂度，第20维lz复杂度；对于Pz电极，比较有效的特征有第17维最大李雅普诺夫指数，第18维C0复杂度，第19维谱熵。于是考虑用以上选择出的9维alpha波中心频率，第13维dalt波中心频率，17维最大李雅普诺夫指数，第18维C0复杂度，第19维谱熵，第20维lz复杂度共6个特征向量组成最终的特征向量数据集。

## 6.3 识别验证

为检验选择出的特征是否真正具有代表性，实验采用多种分类器对选择出的共20维特征以及进行降维之后的6维特征进行识别验证，并用召回率评价分类器性能。识别正确率如图所示：



图6-3 三种分类器降维前后识别正确率

由上图可见，选取特征识别正确率略低于原特征，但总体相差不多，这可能是由于特征的维数降低太多部分影响了识别的正确率。BP分类器和C4.5决策树对于情感的识别正确率相对较高，其中BP分类器识别率到80%以上，但BP分类器在实际应用中受网络训练次数限制，建模速度较慢。

为了评价分类器的性能，各个分类器的召回率如图所示：



图6-4 三种分类器降维前后识别召回率

由上图来看，BP分类器的召回率最高，C4.5原特征的召回率次之，SVM分类器最差。于是对于音频信号引起的脑电信号分析来看，BP的性能较好。

## 6.4 本章小结

结合前两章的实验结果，本章联合分析了音乐信号特征和由其引起的脑电信号特征。实验最终选出了20维的原联合特征以及由GA+CFS降维后的6维特征，并分别用BP，C4.5，SVM对其进行分类验证。实验结果证明，上述6维原特征可以很好的对脑电情绪进行分类。分类器性能上，使用BP分类器或者C4.5分类器。

# 第7章　基于LabVIEW的语音情感识别系统

## 7.1 设计流程

LabVIEW是一种可视化的程序开发环境，利用图标代替文本创建应用程序，用户可以方便的创建和展示自定义界面，LabVIEW提供了Matlab的Script节点，在LabView中可以很方便的调用Matlab进程进行混合编程。

针对音频信号情绪识别，本章利用LabVIEW和Matlab混合编程搭建了音频信号情感识别系统，其中特征的提取、训练和识别主要利用Matlab强大的计算功能完成，数据的可视化显示由LabVIEW完成。总体设计流程如图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。系统的主要识别模型为BP神经网络和SVM，对于音乐情感的识别效果，第三章音乐信号的情感识别已经介绍。

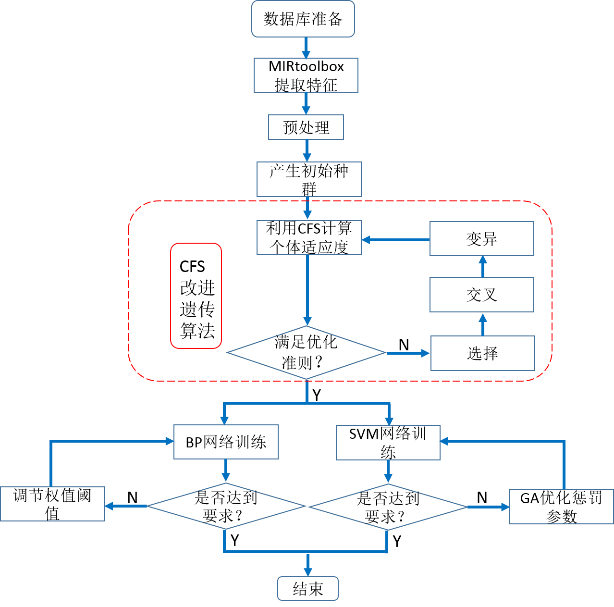


图7-1 系统设计流程图

## 7.2 程序设计

本实验采用Matlab2016a和LabVIEW2014混合编程，设计如图7-2所示的语音情感识别系统，包含录音与播放、特征提取、训练和识别4个模块。由于上述实验中，BP神经网络的识别结果优于SVM已被验证，故本语音识别系统训练和识别方法的内核采用BP神经网络。

### 7.2.1 录音与播放

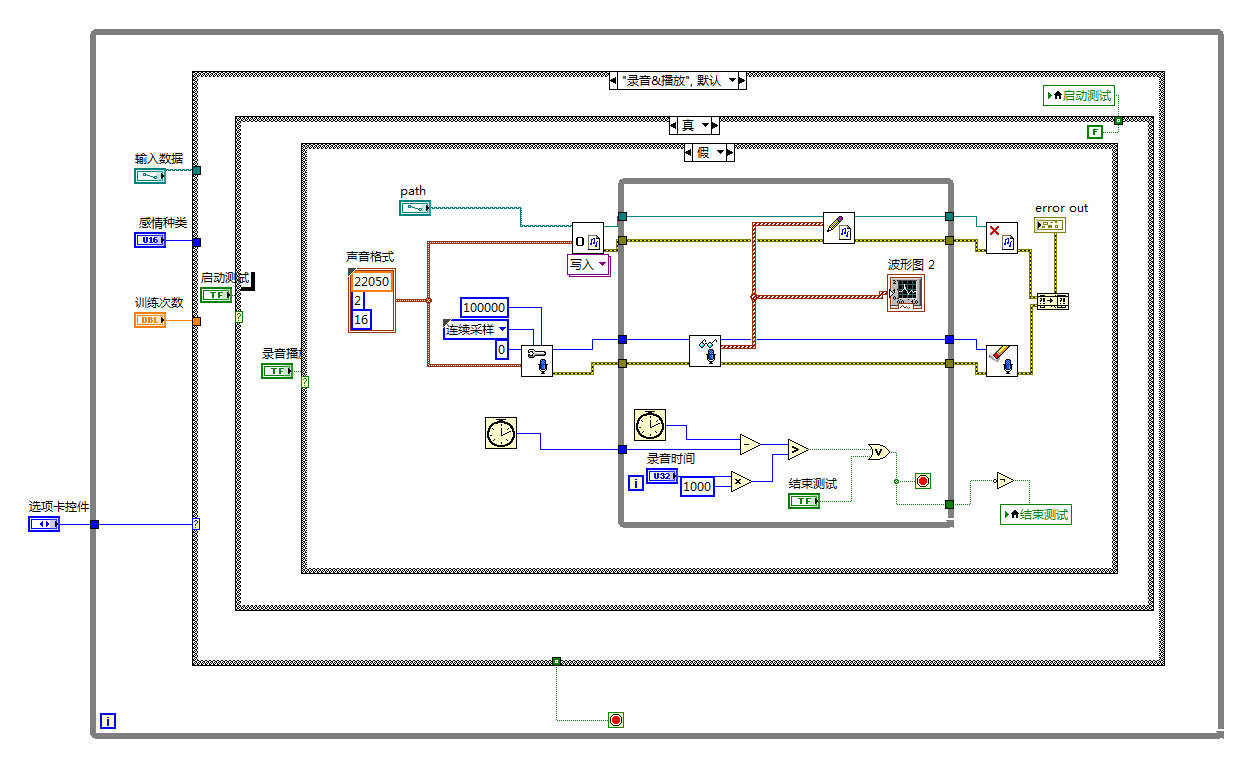
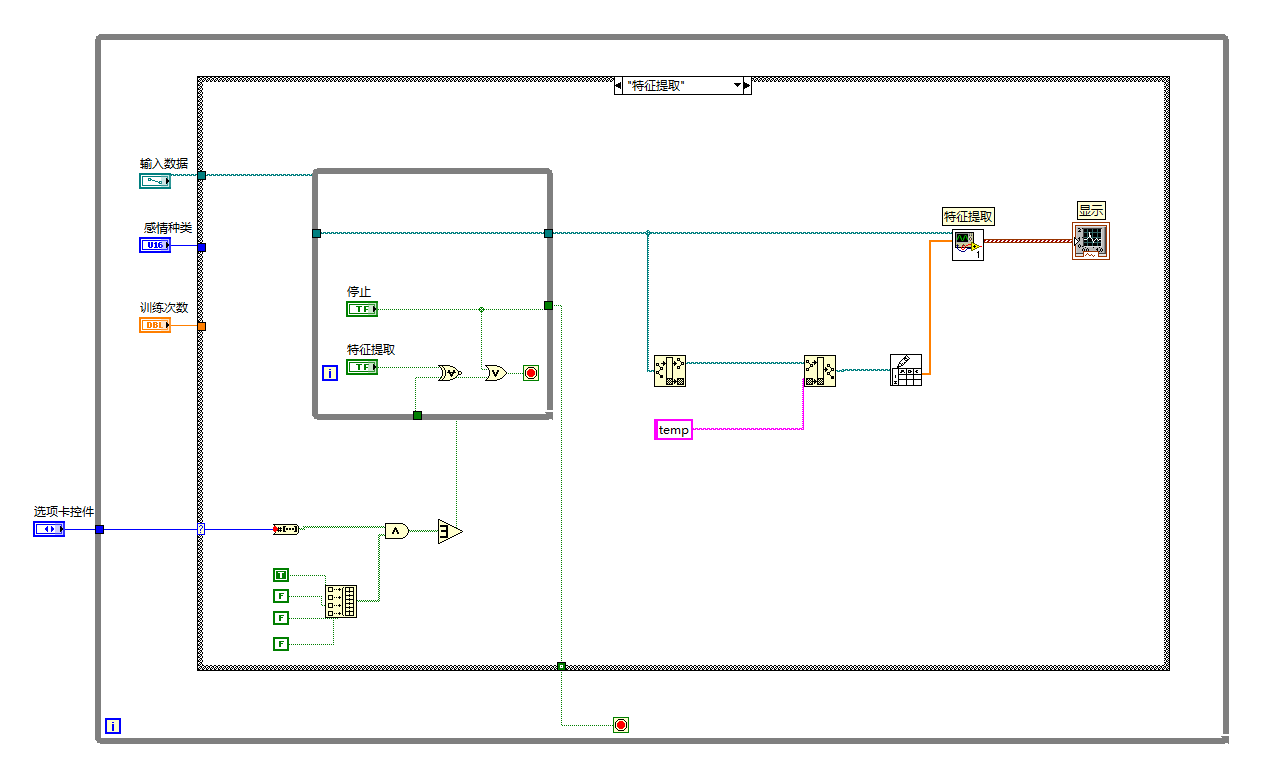


图7-2 录音与播放模块

录音与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，录制文件可以用来判别语音情感。利用windows自带声卡进行录制声音文件，保存格式为.wav格式，采样频率为22050HZ，录制文件存放到自定义文件夹中。播放功能可以播放选择的音频并显示音频波形。

### 7.2.2特征提取



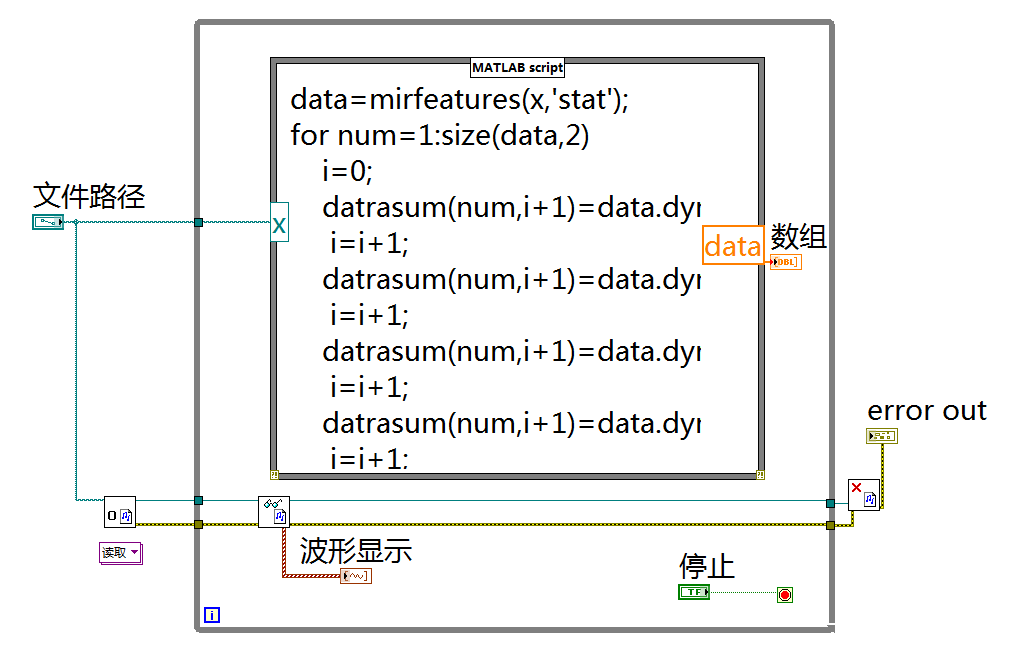
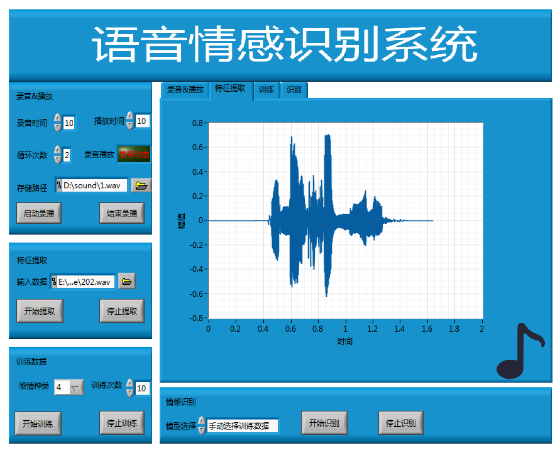
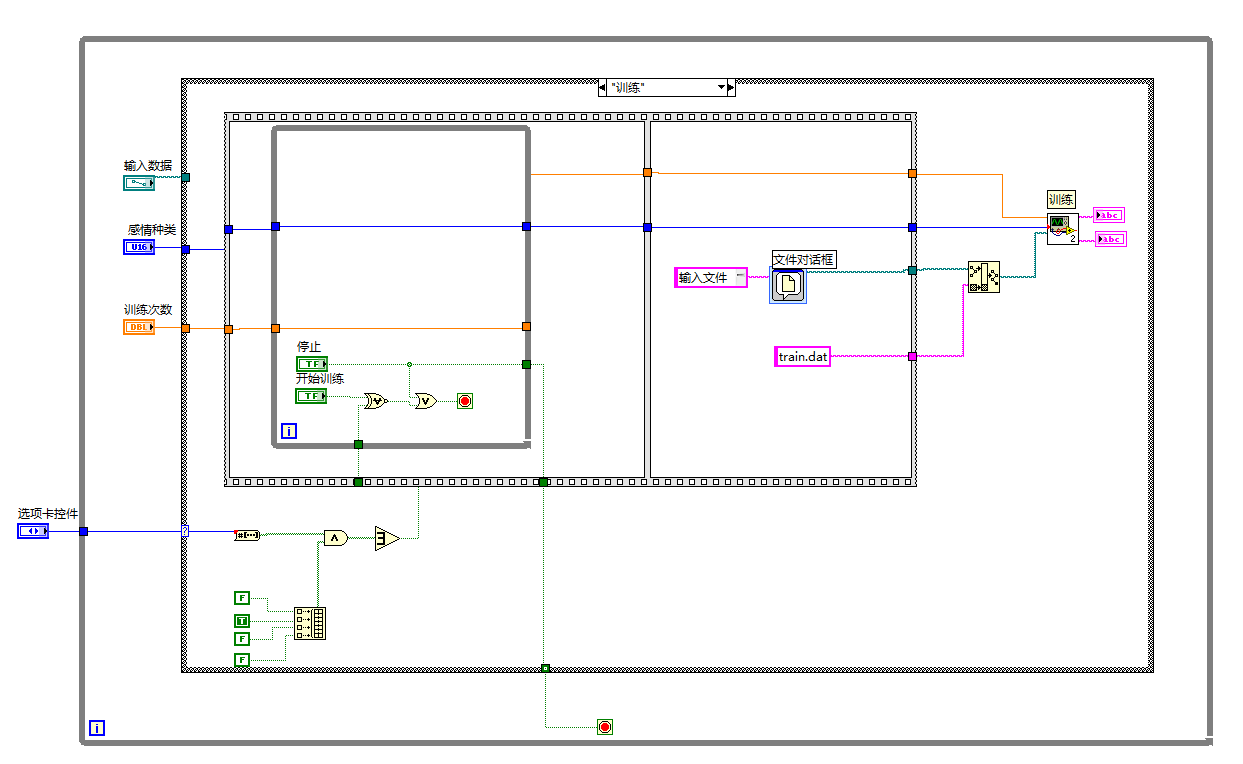
 

图7-3 特征提取模块及前面板

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形式保存在同一文件目录之下，以备下一步识别语音情感使用，提取的具体特征见第三章。

### 7.2.3模型训练



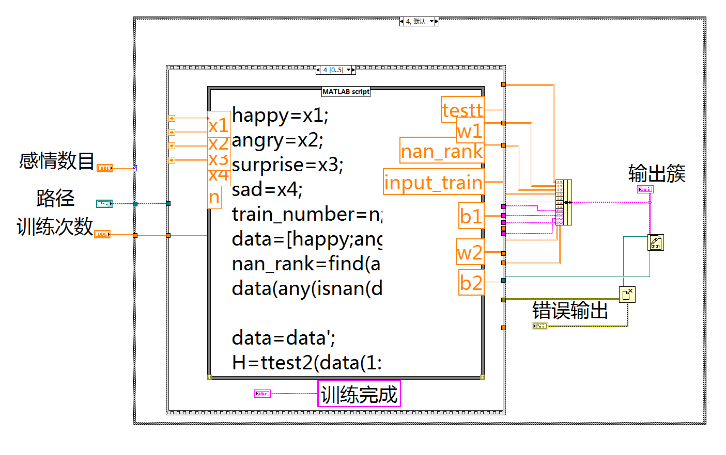
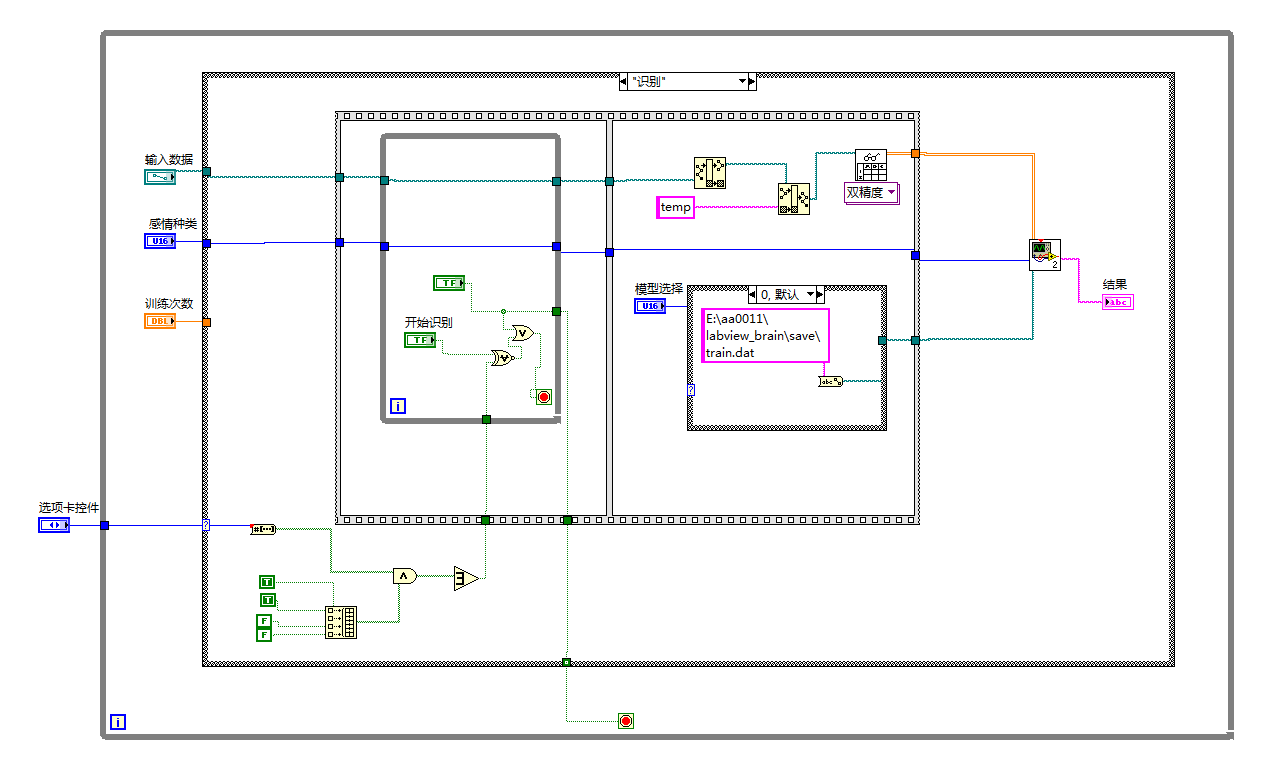


图7-4 模型训练模块及前面板

训练模块在使用时要将已标注好标签的同一感情的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形式保存在指定文件夹下，训练完成后的文件用于识别时选择是否使用此训练数据。

### 7.2.4 识别



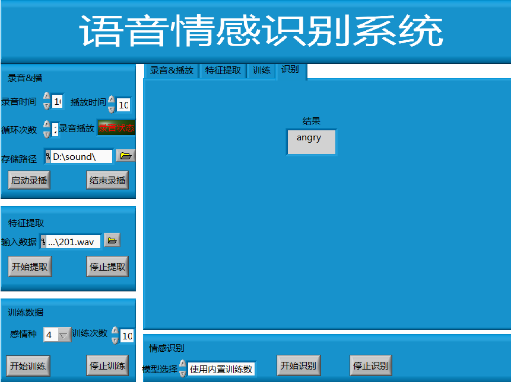
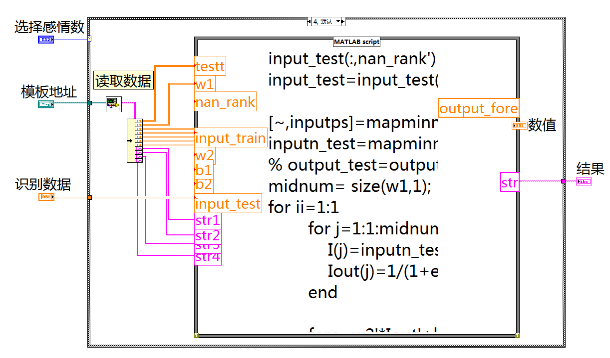


图7-5 识别模块及前面板

识别模块可选择使用内置训练数据或者手动选择使用哪一次的训练数据，内置训练数据是上述已经经过训练好的1000 songs database 数据以及以CASIA数据库基本情感为模板训练的数据，利用内置模板进行识别，可以很快的识别输入音乐的情感。手动选择数据可用来选择在模型训练用户自己训练的数据，达到针对不同人以及音乐的情感识别目的。

## 7.3本章小结

本章利用LabVIEW平台结合Matlab设计了一个具有语音信号采集、特征提取、模型训练、识别验证的语音情感识别系统，用于语音和音乐情感识别。采用MIRtoolbox工具箱提取音频信号的138维特征，并用经CFS改进的遗传算法对它进行降维处理，最终用SVM和BP神经网络对降维前后的特征进行识别验证。利用LabVIEW图形化界面以及Matlab强大的数学运算能力，将特征提取和识别的步骤直观的呈现在用户面前。

# 结　　论

不同情感类型的音乐引发的受众的脑电信号不同，例如轻快的音乐可能给受众正面积极的刺激，躁动不安的音乐令受众感到不适，悲伤低沉的音乐给受众抑郁消极的感觉等。如何获取受众情感，量化不同情感音乐对不同的受众产生什么样的影响是一个重要的课题。本文使用脑电仪提取不同受众的脑电信号，对不同情感的音乐引起的脑电信号进行了系统研究，取得了以下几方面的成果：

（1）提取语音信号的典型64维特征，利用MIV方法对原数据降维，取32维，语音情感识别率提高了3.08%，用CFS的方法进行改进，取其中13维特征，语音情感识别率比原32维提高了9.48%。应用得到的特征向量进行说话人无关实验，结果证明了提取的10维特征向量具有一定的普适性。

（2）首先利用T检验标注80首不同情绪（愤怒，喜悦，悲伤，平静）的音乐，接着利用Mirtoolbox工具箱提取音频信号时域的均方差，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、峰值、质心等138个统计特征。并用GA+CFS算法对上述特征进行降维，共筛选出8维特征，他们分别是：节奏特征的时间均值、时间斜率；谱特征的偏度均值、谱熵均值；音色特征过零率的周期幅度、最低能量的均值；音调模式的周期熵和谐波变化均值。随后利用SVM和BP对降维后的特征进行识别验证，对于SVM，总识别率相对不变，对于BP来说，降维后的识别率比降维前均高出约20%。

（3）采集了8位被试受众在听不同情绪音乐时的脑电信号，最后得到15组实验数据，提取12个电极脑电信号的线性特征和非线性特征，最后每个电极得到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： 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。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。

（4）结合上述实验结果，联合分析了音乐信号特征和由其引起的脑电信号特征。实验最终选出了20维的原联合特征以及由GA+CFS降维后的6维特征，并分别用BP，C4.5，SVM对其进行分类验证。实验结果证明，上述6维原特征可以很好的对脑电情绪进行分类。分类器性能上，BP分类器或C4.5分类器比较优秀。其中BP分类器识别率到80%以上，但BP分类器在实际应用中受网络训练次数限制，建模速度较慢。

（5）利用LabVIEW平台结合Matlab设计了一个语音情感识别系统，用于语音和音乐情感识别。

参考文献

1. 李昕，田彦秀，侯永捷，等. 小波变换结合经验模态分解在音乐干预脑电分析中的应用[J]. 生物医学工程学杂志，2016(4)：762-769.
2. KOELSCH S. Brain correlates of music-evoked emotions[J]. Nature Reviews Neuroscience, 2014, 15(3): 170-180.
3. OMIGIE D. Basic, specific, mechanistic Conceptualizing musical emotions in the brain[J]. Journal of Comparative Neurology, 2015, 524(8): 1676-1686.
4. BOGERT B, NUMMINEN-KONTTI T, GOLD B, ET AL. Hidden sources of joy, fear, and sadness: Explicit versus implicit neural processing of musical emotions[J]. Neuropsychologia, 2016, 89: 393-402
5. TRAINOR L J. The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation[J]. Phil. Trans. R. Soc. B, 2015, 370(1664): 20140089.
6. Schwenker F, Scherer S, Magdi Y M, et al. The GMM-SVM Supervector Approach for the Recognition of the Emotional Status from Speech[C]// Artificial Neural Networks - ICANN 2009, International Conference, Limassol, Cyprus, September 14-17, 2009, Proceedings. 2009:894-903.
7. 戴芮. 基于特征提取多模式结合的语音情感识别的研究[D]. 兰州理工大学, 2014.
8. 章琴. 基于HMM的中文情感语音合成的研究[D]. 合肥工业大学, 2014.
9. 陈洁, 张雪英, 孙颖. 基于HMM的可训练情感语音合成研究[J]. 电声技术, 2012, 36(3):43-46.
10. 陈立江, 毛峡, Mitsuru ISHIZUKA. 基于Fisher准则与SVM的分层语音情感识别[J]. 模式识别与人工智能, 2012, 25(4):604-609.
11. 曲利新. 基于深度信念网络的语音情感识别策略[D]. 大连理工大学, 2014.
12. Sainath T N, Kingsbury B, Saon G, et al. Deep Convolutional Neural Networks for Large-scale Speech Tasks[J]. Neural Networks, 2015, 64:39-48.
13. 韩文静, 李海峰, 阮华斌,等. 语音情感识别研究进展综述[J]. 软件学报, 2014, 25(1):37-50.
14. Origlia A, Galatà V, Ludusan B. Automatic classification of emotions via global and local prosodic features on a multilingual emotional database[J]. Atmospheric Environment, 2010, volume 36(30):4823-4837(15).
15. Malandrakis N, Potamianos A, Evangelopoulos G, et al. A supervised approach to movie emotion tracking[C]// IEEE International Conference on Acoustics. IEEE, 2011:2376-2379.
16. WIECZORKOWSKA A, SYNAK P, RA Z W. Multi-Label Classification of Emotions in Music// Intelligent Information Processing and Web Mining. 2006: 307-315
17. SHI Y Y, ZHU X, KIM H G, ET AL. A Tempo Feature via Modulation Spectrum Analysis and its Application to Music Emotion Classification.//IEEE International Conference on Multimedia and Expo, ICME 2006, July 9-12 2006, Toronto, Ontario, Canada. 2006: 1085-1088
18. LAURIER C, LARTILLOT O, EEROLA T, ET AL. Exploring relationships between audio features and emotion in music// Triennal Conference of European Society for Cognitive Sciences of Music. 2009:260-264
19. GURRIN C. EMIR: A novel emotion-based music retrieval system. 2012
20. PRESS C, Recognition M E. Music Emotion Recognition. Crc Press, 2011: 255-266
21. EM SCHMIDT, D TURNBULL, YE KIM. Feature selection for content-based, time-varying musical emotion regression.ACM proceedings of the international conference on multimedia information retrieval,2010: 267-273
22. ZHANG K, SUN S. Web music emotion recognition based on higher effective gene expression programming. Neurocomputing, 2013, 105(3): 100-106
23. MINGYU YOU, JIAMING LIU, GUOZHENG LI, ET AL. Embedded Feature Selection for Multi-label Classification of Music Emotions[J]. International Journal of Computational Intelligence Systems, 2012, 5(4): 668-678
24. Thammasan N, Moriyama K, Fukui K I, et al. Familiarity effects in EEG-based emotion recognition[J]. Brain Informatics, 2016, 4(1):1-12.
25. Kumagai Y, Arvaneh M, Tanaka T. Familiarity Affects Entrainment of EEG in Music Listening:[J]. Frontiers in Human Neuroscience, 2017, 11:384.
26. Bigliassi M, Karageorghis C I, Nowicky A V, et al. Effects of auditory distraction on voluntary movements: exploring the underlying mechanisms associated with parallel processing.[J]. Psychological Research, 2017:1-14.
27. Lopata J A, Nowicki E A, Joanisse M F. CREATIVITY AS A DISTINCT TRAINABLE MENTAL STATE: AN EEG STUDY OF MUSICAL IMPROVISATION[J]. Neuropsychologia, 2017.
28. Phneah S W, Nisar H. EEG-based alpha neurofeedback training for mood enhancement[J]. Australas Phys Eng Sci Med, 2017, 40(2):1-12.
29. Bridwell D A, Leslie E, Mccoy D Q, et al. Cortical Sensitivity to Guitar Note Patterns: EEG Entrainment to Repetition and Key[J]. Frontiers in Human Neuroscience, 2017, 11.
30. Martínez-Rodrigo A, García-Martínez B, Alcaraz R, et al. Study of Electroencephalographic Signal Regularity for Automatic Emotion Recognition[C]// International Conference on Ubiquitous Computing and Ambient Intelligence. Springer, Cham, 2017:766-777.
31. Zhou, L., et al. "Neural differences between the processing of musical meaning conveyed by direction of pitch change and natural music in congenital amusia. " Neuropsychologia 96(2017):29.
32. 谢康. 情绪音乐的脑电识别算法[D]. 成都：电子科技大学，2013.
33. 马勇，李娟，吕彬. 脑电信号驱动的个性化情绪音乐播放系统算法研究及初步实现[J]. 生物医学工程学杂志，2016(1)：38-41.
34. Li X, Hu B, Sun S, et al. EEG-based mild depressive detection using feature selection methods and classifiers[J]. Computer Methods & Programs in Biomedicine, 2016, 136(C):151-161.
35. 徐佳琳，左国坤. 基于互信息与主成分分析的运动想象脑电特征选择算法[J]. 生物医学工程学杂志，2016(2)：201-207.
36. THAYER R E. The biopsychology of mood and arousal.[J]. Cognitive & Behavioral Neurology, 1992(1): 65.
37. Cowie R, Douglascowie E, Tsapatsoulis N, et al. Emotion recognition in human-computer interaction[J]. IEEE Signal Processing Magazine, 2002, 18(1):32-80.
38. 孙亚新. 语音情感识别中的特征提取与识别算法研究[D]. 华南理工大学, 2015.
39. 查美丽. 基于情感的音乐分类系统的研究与实现[D]. 南京邮电大学, 2014.
40. Busso C, Lee S, Narayanan S. Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection[J]. IEEE Transactions on Audio Speech & Language Processing, 2010, 17(4):582-596.
41. Iliou T. Statistical Evaluation of Speech Features for Emotion Recognition[C]// International Conference on Digital Telecommunications. IEEE Computer Society, 2009:121-126.
42. 向磊. 语音情感特征提取与识别的研究[D]. 浙江理工大学, 2013.
43. Lugger M, Janoir M E, Yang B. Combining classifiers with diverse feature sets for robust speaker independent emotion recognition[C]// Signal Processing Conference, 2009, European. IEEE, 2009:1225-1229.
44. 胡冰洁. 基于特征向量的音乐情感分析的研究[D]. 西安电子科技大学, 2014.
45. 余华, 黄程韦, 金赟,等. 基于改进的蛙跳算法的神经网络在语音情感识别中的研究[J]. 信号处理, 2010, 26(9):1294-1299.
46. Wu S, Falk T H, Chan W Y. Automatic speech emotion recognition using modulation spectral features[J]. Speech Communication, 2011, 53(5):768-785.
47. 李颖洁, 邱意弘, 朱贻盛. 脑电信号分析方法及其应用:信息传输的基本原理[M]. 科学出版社, 2009.
48. 尹兵. 脑电波信号的去伪迹研究[D]. 南京邮电大学, 2014.
49. 李松, 伏云发, 杨秋红,等. 基于左右手运动想象单通道脑电信号的预处理研究[J]. 生物医学工程学杂志, 2016(5):862-866.
50. 李艳艳, 杨陈军, 野梅娜,等. 一种新的癫痫脑电融合特征提取方法[J]. 西北大学学报:自然科学版, 2016, 46(6):801-808.
51. 李小伟. 脑电、眼动信息与学习注意力及抑郁的中文相关性研究[D]. 兰州大学, 2015.
52. 李昕, 蔡二娟, 田彦秀,等. 一种改进脑电特征提取算法及其在情感识别中的应用[J]. 生物医学工程学杂志, 2017(4):510-517.
53. 杨默涵, 陈万忠, 李明阳. 基于总体经验模态分解的多类特征的运动想象脑电识别方法研究[J]. 自动化学报, 2017, 43(5):743-752.
54. 范金锋, 邵晨曦, 王剑,等. 醉酒者脑电和正常脑电非线性特性的比较评估[J]. 中国生物医学工程学报, 2008, 27(1):18-22.
55. 成娟, 陈勋, 彭虎. 基于样本熵的肌电信号起始点检测研究[J]. 电子学报, 2016, 44(2):479-484.
56. Gao L, Cheng W, Zhang J, et al. EEG classification for motor imagery and resting state in BCI applications using multi-class Adaboost extreme learning machine.[J]. Review of Scientific Instruments, 2016, 87(8):216-219.
57. 顿士君. 基于脑电多尺度非线性分析的睡眠分期研究[D]. 燕山大学, 2016.
58. 翟经纬. 脑电信号非线性处理方法在精神分裂症诊断中的应用[D]. 兰州大学, 2012.
59. Acharya U R, Sudarshan V K, Adeli H, et al. A Novel Depression Diagnosis Index Using Nonlinear Features in EEG Signals[J]. European Neurology, 2016, 74(1-2):79-83.
60. 蔡志杰, 孙洁. 改进的C0 复杂度及其应用[J]. 复旦学报(自然科学版), 2008, 47(6):791-796.
61. 王凯明. 基于改进功率谱熵的抑郁症脑电信号活跃性研究[J]. 物理学报, 2014, 63(17):178701-8.
62. 张小鹏, 范影乐, 杨勇. 基于奇异谱熵的脑电意识任务识别方法的研究[J]. 计算机工程与科学, 2009, 31(12):117-120.
63. Cai D, Zhang C, He X. Unsupervised feature selection for multi-cluster data[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010:333-342.
64. Huanjing Wang; Khoshgoftaar, T.M.; Napolitano, A..Stability of filter- and wrapper-based software metric selection techniques[C], 2014 IEEE 15th International Conference on Information Reuse and Integration. Redwod City, CA, 2014: 309-314.
65. 周莹. 基于MIV特征筛选和BP神经网络的滚动轴承故障诊断技术研究[D]. 北京交通大学, 2011.
66. 柳亚琴, 石洪波.基于GA-CFS属性选择的个人信用评估模型.计算机系统应用, 2011, 20(5):210-213
67. Hall M A. Correlation-based feature selection for machine learning[D]. The University of Waikato, 1999.
68. Bolón-Canedo V, Porto-Díaz I, Sánchez-Maroño N, et al. A framework for cost-based feature selection[J]. Pattern Recognition, 2014, 47(7):2481-2489.
69. Soleymani M, Caro M N, Schmidt E M, et al. 1000 songs for emotional analysis of music[C]// ACM International Workshop on Crowdsourcing for Multimedia. ACM, 2013:1-6.
70. Lartillot O, Toiviainen P. MIR in Matlab (II): A Toolbox for Musical Feature Extraction from Audio.//International Conference on Music Information Retrieval, Ismir 2007, Vienna, Austria, September. DBLP, 2007:127-130
71. Lartillot O, Toiviainen P, Eerola T. A Matlab Toolbox for Music Information Retrieval.//Data Analysis, Machine Learning and Applications - Proceedings of the, Conference of the Gesellschaft Für Klassifikation E.v. Albert-Ludwigs-Universität Freiburg, March. DBLP, 2008:261-268
72. 钟宏燕. 脑电信号的特征提取和分类研究[D].北京：北京工业大学，2014.
73. 蔡琼，陈鹏慧，魏武. 采用小波变换的脑电图信号分析及其应用[J].泉州：华侨大学学报(自然版)，2015，36(2)：166-170.
74. 卢英俊，戴丽丽，吴海珍，等. 不同类型音乐对悲伤情绪舒缓作用的EEG研究[J]. 心理学探新，2012，32(4)：369-375.
75. 秦积涛. 基于脑电分析方法的海洛因成瘾者的决策障碍研究[D].兰州：兰州大学，2016.

# 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**

# 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

致　　谢

三载悠悠，转瞬即过。回首硕士生涯，感慨颇多。三年里，我经历了从一个科研小白到一个能独立进行研究的合格研究生的转变，从刚开始的不知如何入手，到现在的明白学科研究的前沿所在并勇于在国际期刊上发出自己的声音，一路风雨，一路收获。不断的挫折打击，不断的独自思考，合作讨论，反复验证充斥在逐渐演变的过程中，这些经历不仅使我在学术上更加成熟，也是我生活上的宝贵财富。

首先，在此诚挚的感谢侯一民老师在本人研究生期间的教导和帮助。从确定研究方向到设计实验，得到成果，都离不开侯老师的指导。实验的刚刚起步阶段，根据实验结果我写了第一篇论文，由于没有经验，现在看来有诸多的错误和不合理之处，但侯老师并没有批评，而是指出不合理之处并鼓励我继续研究下去，在侯老师的循循善诱和耐心指导才使我有了长足的进步。

其次，应该感谢同一教研室的同学和同寝室的诸位好兄弟，正是有了各位同学和同寝的关心和互帮互助，才使得我的学习生活上更加成功。特别要感谢同门师兄弟姐妹的精诚合作，使得我在学术上能更加的专心，还有同寝室各位同学的互相谅解，创造了一个好的宿舍氛围，使相互更能投入到研究中去。

最后，感谢我的父母，他们不仅赐予我生命，更是坚定不移的支持我的学业，支持我所做的每一项决定。正是家庭的教育和父母的支持，让我有了茫茫学海中坚持下去的动力。