****

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

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，这是音乐厂商，唱片公司极其关注的问题，以期获得发行唱片的成功或者对音乐，歌曲的发行作提前的预期来进行改进并避免损失；同时，情感计算也是探究人类情感，实现人机交互的一个研究热点。脑电信号，作为人类的一个典型生物信号，包含有大量情感信息，更能客观的反应人类的情绪变化，近年来，随着科学技术的进步，脑电信号的提取更加方便快捷，对脑电信号情绪特征的提取也引起了广泛的研究。

在此背景下，本研究广泛提取了不同情感音频信号和对应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）在语音方面，利用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。利用各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位被试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。利用分类器对原始特征集和所选特征集的识别率进行测试。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。实验选择20维的原联合特征以及由GA+CFS降维后的6维特征，并分别用多种分类器对其进行分类验证。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。

（4）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。

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.

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

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

音频信号主要包括语音信号和音乐信号，用音频信号来研究情感信息，首先要对情感进行有效的分类。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，于是在识别市经常混合使用各个模型[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。

在音乐情感识别方面，Wieczorkowska A 等Alicja W 等将歌曲情感分成 11 个情感类别，提取音乐的频率、谐度等特征，使用改进的 K 最近邻（KNN）算法对乐曲进行多标签分类[3]。Shi Y Y,等基于 AdaBoost 分类器，提出了一个双层音乐情感分类系统[4]。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[5]。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(Support Vector Machine，SVM)，贝叶斯，决策树等）对音频情感进行了研究，取得了很好的效果[6~8]。浙江大学张克俊对生理信号标记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[9]。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[10]。

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波段的有统计性的差异。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。

本文的具体组织和安排如下：

第一章讲了课题研究的背景和意义，以及现在国内外研究的趋势和不足。

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。

第四章利用MIRtoolbox工具包提取了4种不同情感音乐的138维特征，并利用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，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。

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，提取采集的与情绪相关的12个电极每个电极的27维特征；利用互相关特征选择 (Correlation-Based Feature Selection，CFS)方法的方法选择出了与原特征集相关性较大冗余较小的最优特征集；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，利用互相关特征选择（Correlation-Based Feature Selection，CFS）方法对遗传算法（Genetic Algorithm，GA）进行改进，对音频信号特征降维，使原有的机器学习算法快速收敛，提高了识别正确率。在此基础上设计基于LabVIEW的语音情感识别系统，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，Wundt等提出了环形模型( 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。情感大部分都分布在这个环上，在这个环上，可以描述情感的相似程度。而Thayer二维情感模型更是广泛的维度情感模型，被大量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】。

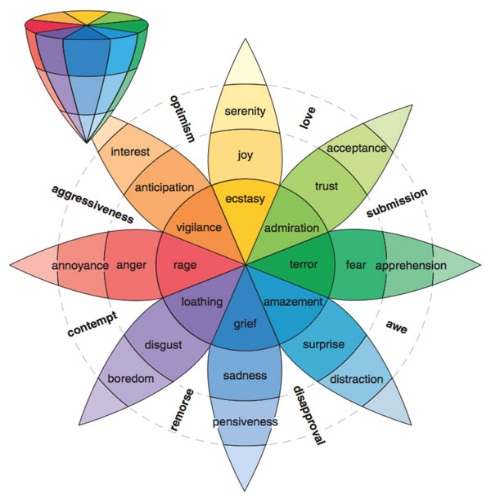
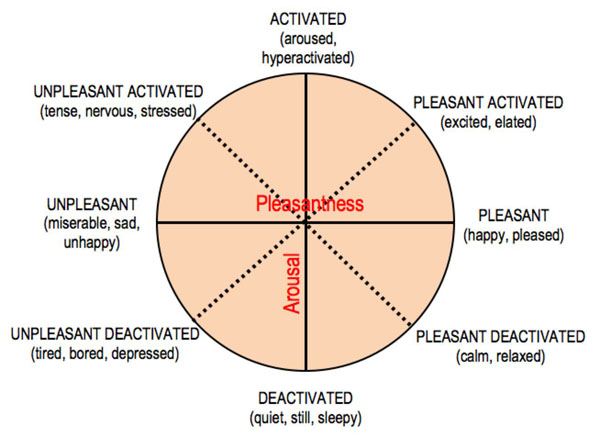


图 二维情感模型（左）三维情感模型（右）

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，包括语音的元音、辅音、节奏等；而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]。下面主要介绍全局统计特征。

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。

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

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]。

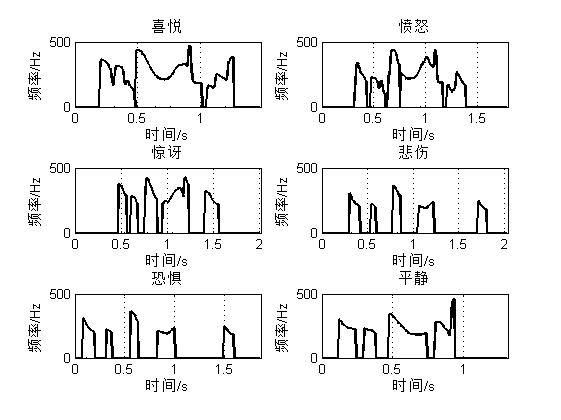


图2 六种情感“就是下雨也去”有音段基频比较

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]。

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

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]。

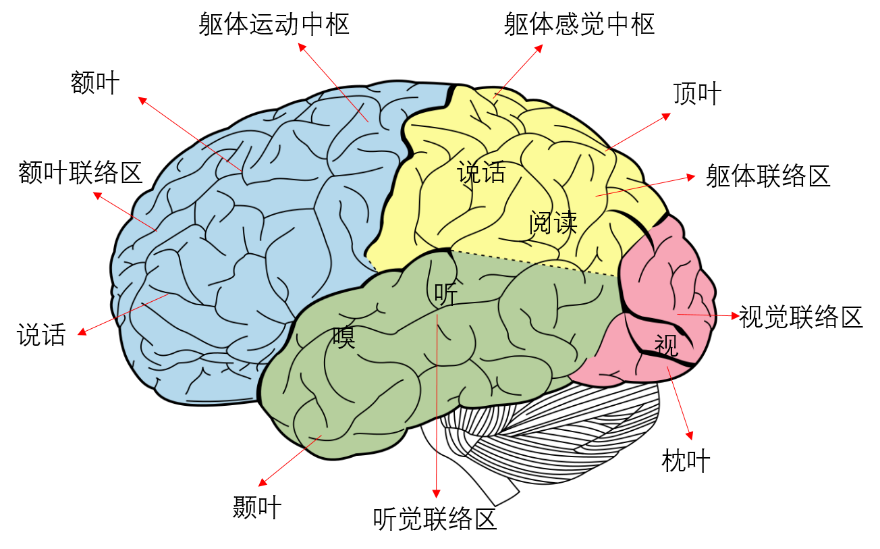
此外，谱特征，发音帧数特征也得了广泛研究[16][17][18]。

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。

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。

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：

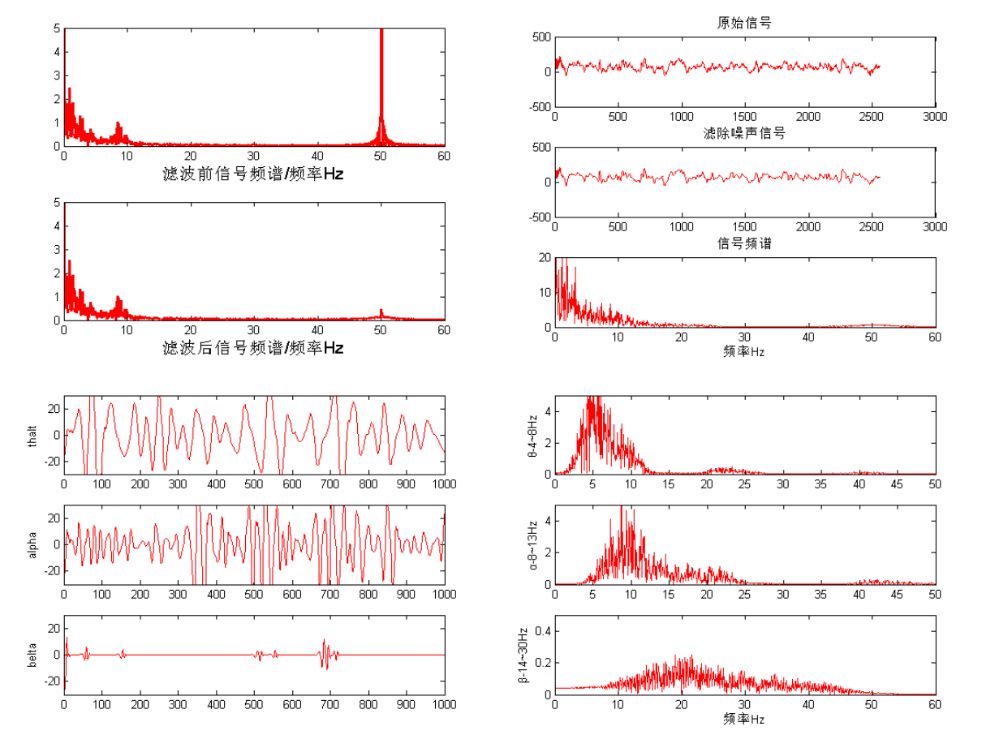
式中，表示数学期望 \*表示复共轭。功率谱密度函数的另一种定义是：

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。

小波变换：采集到的脑电往往综合着自发脑电和事件相关电位。脑电的频率与思维状态和神经功能都有着紧密联系,如何准确有效的提取到脑电特定频率的信息是脑电研究的一个关键问题。小波是在从信号处理领域的傅里叶变换中发展过来的。起初,在信号处理领域傅里叶变换扮演着重要的角色,它是一种将时域信号转换为包含频域信息的数学方法。然而,傅里叶变换的不足之处在于:无法同时包含时域和频域信息,也就是说,一旦将一组时域信号进行了傅里叶变换得到频域信号后,就完全失去了时域信息。但是,在现实生活中,大多数的时域信号都包含事件的开始、趋势以及结束都包含重要信息,此时傅里叶变换在分析这些信号时就显得无能为力。与傅里叶变换工作在单尺度上不同的是,小波变换可以很好的应用于多尺度。即共同含有时域和频域信息。而脑电信号本身是由许多振荡频率组成,这些频率中当然也包括噪声在内。脑电信号就是各种频率信号的重叠综合体所以,小波变换在脑电信号的处理领域扮演着重要**的角色。**

****

**图 脑电信号线性特征**

1. 非线性特征

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

随着对非线性系统的不断深入研究，非线性特征渐渐应用于脑电信号分析。非线性系统在自然界中广泛存在，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）：对时间序列的无需度测量，表示信号复杂性大小。近似熵算法简单，分析数据量小，抗干扰性强。近似熵是一个正数，对于EEG信号，它的值越大表示更高的复杂性或者无规律性。其定义为：

式中，Ⅳ为数据长度，m为子序列长度，r为有效阈值。简单地说，ApEn是对于m个相继观察值在，范围内模式接近、在子序列递增比较仍接近的对数相似程度的度量。(醉酒者脑电和正常脑电非线性特性的比较评估)

样本熵：样本熵是Richman和Moorman[11]对近似熵进行了改进，降低了计算误差，提高了计算精度。计算样本熵与数据长度无关且对于丢失数据不敏感。样本熵的物理意义表示非线性动力学系统产生新信息的速率，样本熵值越低，序列自我相似性越高，产生新模式的速率越低；样本熵值越大，序列越复杂。样本熵与近似熵的最大区别就是不进行自身模板匹配，亦即自身匹配不计入统计量。

关联维和K熵:关联维是分维的一种，1958年 Kolmogrov定义了测度熵,称为 Kolmogrov熵(又称信息维数),用来度量系统运动的混乱或无序的程度。随后 Sinai进行了改进,所以又称为 Kolmogrov-inai熵,简称为KS熵。考虑一个n维动力系统,将它的相空间分割成一个个边长为E的n维立方体盒子,对于状态空间的一个吸引子和一条落在吸引域的轨道x(t),取时间间隔为一很小量z,令P(,,…,)表示起始时刻轨道在第个格子中,t=z时在第动个格子中的联合概率,则 Kolmogorov熵定义为:

Crassbreger和Procaccia 于1983年提出了相关维数的概念，并定义了相关熵(Correlation Entropy)的概念来逼近Kolmogorov熵，简写为k2熵，典型情况下可以同时得到 关联熵和k熵的稳定估计。

LZ复杂度：20世纪60年代,柯尔莫哥洛夫(Kolmogorov)等人对复杂性作了定义:一个系统的复杂程度与该系统行为(空间结构或时间序列所表示的变化)的最小描述有关,一般称为柯尔莫哥洛夫复杂性。其基本思是认为:某事物的算法复杂度等于产生该事物的图形结构或符号序列的最短程序长度与该图形结构或符号序列本身大小之比的极限(当后者趋于极限时)122l由于柯尔莫哥洛夫复杂性是不可计算的,1976年A. Lempe1和J.Ziv提出一种度量符号序列复杂性的简单算法,称为 Lempe1-ziv复杂度。由于此法适用于符号序列,所以首先要对数据序列进行符号化。根据Lempel-Ziv复杂度算法可知，一个符号序列的复杂度越大，说明它的添加操作越多，新模式也越多，描述给定符号序列所需最少的、互不相同的“子串”个数越多，给定的符号序列周期性越弱，出现新模式的速率越快。反之，则复制操作越多，新模式越少，周期性越强，出现新模式的速率越慢。普遍认为复杂度的物理意义是它反映了一个时间序列随着序列长度的增加出现新模式的速率。复杂度越大，说明数据在窗口长度时期内随时间出现的新变化越多，发生新变化的速率越快，表明这一时期的数据变化是无序而复杂的。反之，复杂度越小，则说明发生新变化的速率越慢，数据变化是规则的，周期性越强。因此，计算一个时间序列的复杂度能够描述出系统的状态发生变化的情况，这是非常重要的性质。

最大李雅普诺夫指数：即指数是衡量系统动力学特性的一个重要定量指标,它表征了系统在

相空间中相邻轨道间收敛或发散的平均指数率。对于系统是否存在动力学混沌,可以从最大指数是否大于零非常直观的判断出来一个正的指数,意味着在系统相空间中,无论初始两条轨线的间距多么小,其差别都会随着的演化而呈指数率增加以致最后达到无法预测,这就是混沌现象。

C0复杂度：：计算序列功率谱的均方值，高于均方值的部分保持不变，而将低于均方值的部分置为零．计算新的信号频谱的Fourier逆变换，得到新的时间序列．新时间序列与原时间序列之差的平方和与原始信号的平方和之比定义为原始信号的c0复杂度。

最大李雅普诺夫指数：用以测量相空间内两点距离随时间（迭代次数）的指数分离程度，及估计系统的混乱程度。如果最大李雅普诺夫指数＜０，代表相邻初始值的两个轨线趋于一致；＞０，对应着与相邻初始值的两个轨线作指数偏离并且此系统对巧始值敏感，当然也表明该系统具有混浊特性。

谱熵：是把香农赌的概念应用到傅里叶变换，计算信号的功率分布，归一化单位功率。谱熵允许分离不同频率范围的成分。

奇异谱分析：通过对一维时间序列进行延迟重构和奇异值分解，按能量大小排序来确定被分解量的重要性。奇异谱熵通过对一维时间序列进行延迟秀构和奇异值分解来计算熵源，而并不是直接将原始信号作为有效熵源。它反映了肭电信号能量在奇异谱划分下的不确定性。信号越简单，能量越集中于少数几个模式；相反，信号越复杂，能量就越分散。奇异谱熵可以作为脑电信号非线性的一种度量。

2.4降维方法

2.4.1LDA

线性判别分（LDA）是根据研究对象的各种特征值判别其类型归属问题的一种多变量统计分析方法，LDA是有监督的方式，它先对训练数据进行降维，然后找出一个线性判别函数进行判别，其中最经典的则是“Fisher判别分析”。线性判别分析的核心思想是寻找到最佳的投影方法，将高维的样本投影到特征空间(feature space)，使得不同类别间的数据“距离”最大，而同一类别内的数据“距离”最小。MCFS是浙江大学教授蔡登在2010年提出，主要目标是选择能最大保持数据类簇结构的特征[11]。两种方法可用于模式识别和数据挖掘，降低数据的维度，优化算法性能和准确性。

2.4.2 平均影响值算法

平均影响值(Mean Impact Value,MIV)算法，是由Dombi等人提出，用来反映神经网络中权重矩阵的变化情况，MIV被认为是评价神经网络特征变量相关性最好的指标之一。

MIV算法旨在判断输入特征对于网络结果的影响程度，从而进行变量筛选[20]。首先用原数据训练一个稳定的神经网络，在训练结束后，将训练样本P中每一自变量特征在其原值基础上分别加和减10%构成两个新样本P1和P2，将P1，P2作为仿真样本重新进行仿真，得到两个仿真样本A1和A2，求A1和A2差值，即为变动此自变量对输出产生的影响值，最后将此影响值按照观测平均得到自变量对于因变量的平均影响值。依次即可算出各个变量的MIV[19]。

2.4.2遗传算法

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，逐代演化产生更好的近似解，在每一代选择时，通过个体适应度从解群中选取较优个体，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。

2.4.3特征选择算法

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，, , ···,随后将特征子集(), (, ), (, ), (, , ), (, ···).进行分类器识别验证。

2.5识别方法

2.5.1神经网络

BP神经网络是一种神经网络学习算法，是一种按误差逆传播算法训练的多层前馈网络，其应用广泛。一般拓扑结构包括输入层，中间层（隐藏层），输出层三层。它使用梯度下降法，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 BP神经网络拓扑结构图

图2中，x1,x2···xn是BP神经网络的输入值，Y1,Y2···Yn是BP神经网络的预测值，ωij和ωjk是BP神经网络权值。从图中可以看出，BP神经网络可以看做一个非线性函数，网络输入值和预测值分别为该函数的自变量和因变量。当输入节点为n，输出节点为m时，BP神经网络就表达了从n个自变量到m个因变量的函数映射关系。

2.5.2支持向量机

支持向量机csupportvectormachine SVM) 是一种二类分类模型、它的基本模型是定义在特征空间上的间隔最大的线性分类器，间隔最大使它有别于感知机; 支持向量机还包括核技巧，这使它成为实质上的非线性分类器，支持向量机的学习策略就是间隔最大化，可形式化为一个求解凸二次规划(convex quadraticprogramming) 的问题，也等价于正则化的合页损失函数的最小化问题。支持向量机的学习算法是求解凸二次规划的最优化算法。

2.5.3决策树

决策树(decision tree) 是一种基本的分类与回归方法。本章主要讨论用于分类的决策树。决策树模型呈树形结构，在分类问题中，表示基于特征对实例进行分类的过程。它可以认为是if-then规则的集合，也可以认为是定义在特征空间与类空间上的条件概率分布。其主要优点是模型具有可读性，分类速度快，学习时，利用训练数据，根据损失函数最小化的原则建立决策树模型。预测时，对新的数据，利用决策树模型进行分类.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。【统计学习方法】

# 第3章　语音信号情感识别

3.1实验设计

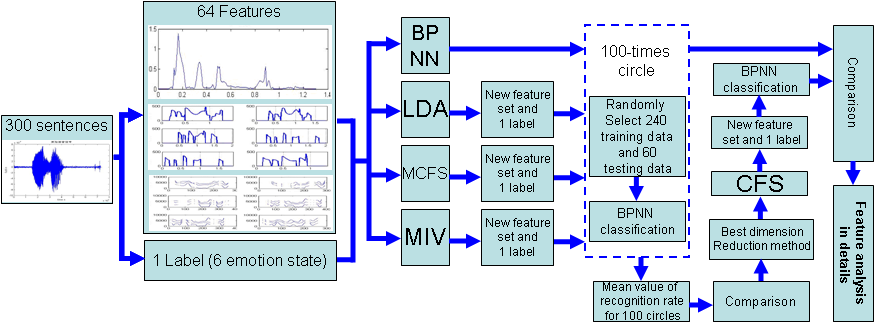
音频信号分为语音信号和音乐信号，音频信号情感的划分如上介绍，为了研究音频信号的情感识别，对于语音信号的情感识别利用离散化标签的数据库，即CASIA汉语情感语料库。CASIA汉语情感语料库由中国科学院自动化研究所录制，由4位录音人(2男2女)在纯净录音环境下(信噪比约为35db)分别在6类不同情感下(高兴、悲哀、生气、惊吓、难过、中性)对500句文本进行的演绎得到[13]。实验首先提取语音库中某女生包含喜怒哀乐平静惊讶六种感情，每种感情50句语音，共300句语音作为数据库。具体实验设计流程如图所示：

图 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。得到可以接受的特征向量，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

3.3.1MIV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11,22,21,16,1314,26,24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2 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 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。

3.3.2LDA以及MCFS的方法降维

表4 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5 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 |

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

表6 CFS法选择共振峰特征

表 不同维度组合后识别结果

|  |  |  |  |
| --- | --- | --- | --- |
| 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，而不影响正确率，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。

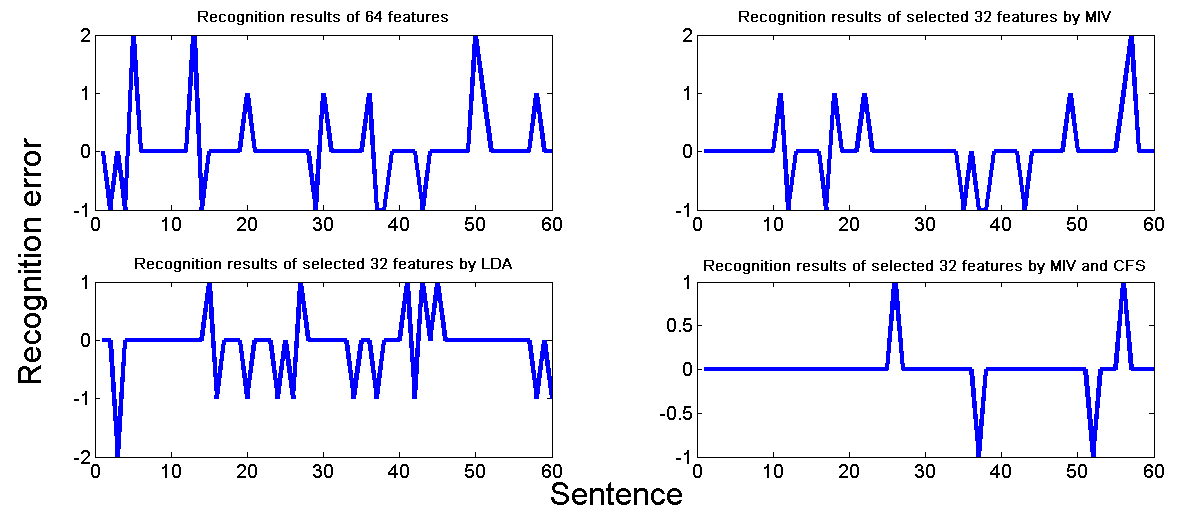
****

图4 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。

3.5特征分析

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分析基频特征同语音情感之间的关系。

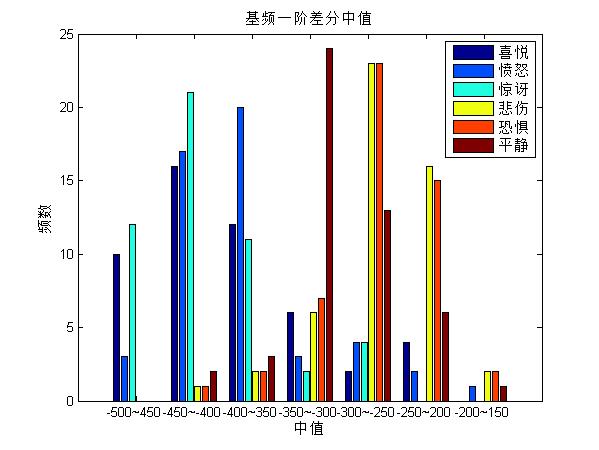
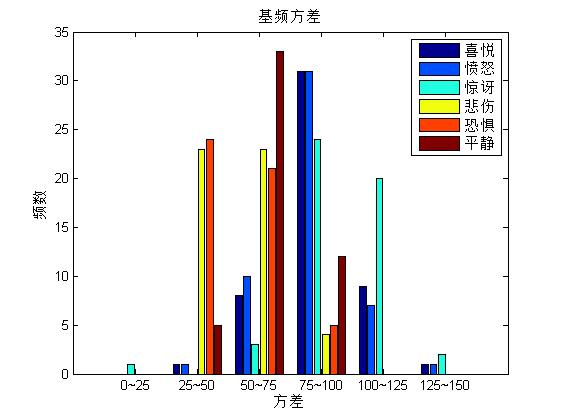


图5 第15维基频方差

图6 第19维 一阶差分中值

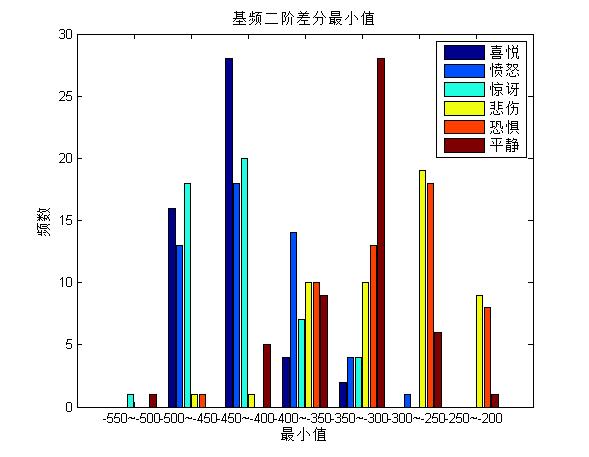
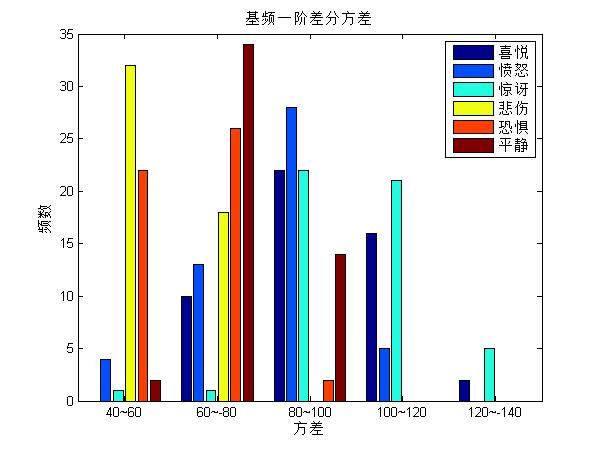


图7 第20维基频一阶差分方差 图8 第22维 基频二阶差分最小值

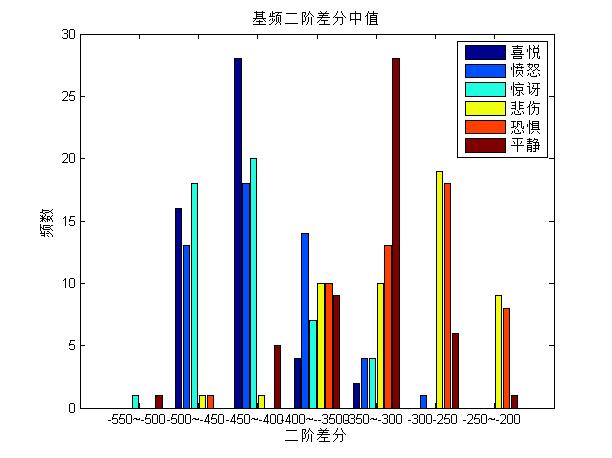


图9 第24维基频二阶差分中值

由基频特性看，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。

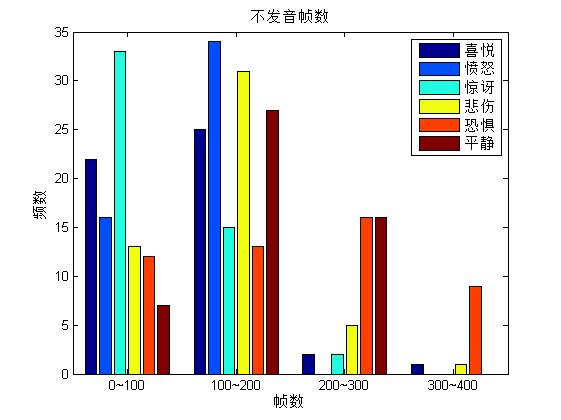
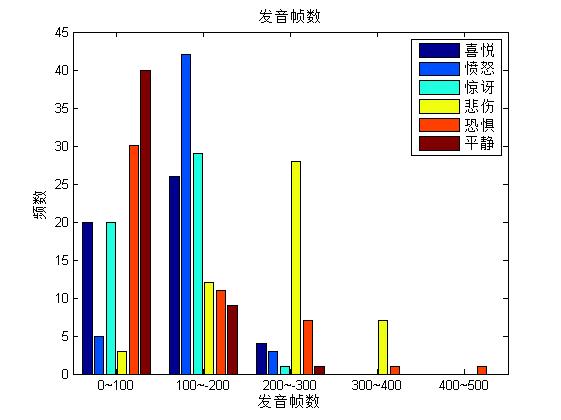


图10 第27维发音帧数 图11 第28维不发音帧数

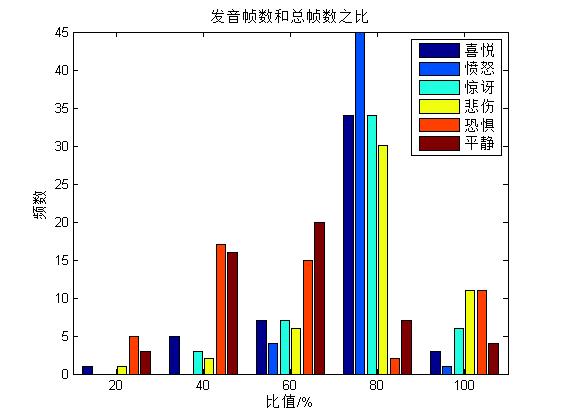


图12 第30维 发音帧数和总帧数比

语音的发音帧数也有着区分意义。由上图10-图12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80%左右。

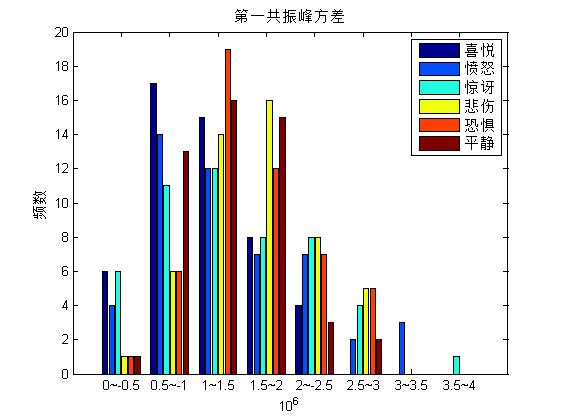


图13 第36维第一共振峰方差

图14 第41维第二共振峰方差

喜悦的第一二共振峰方差较小，共振峰平缓；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，需要对未选取特征进行分析。

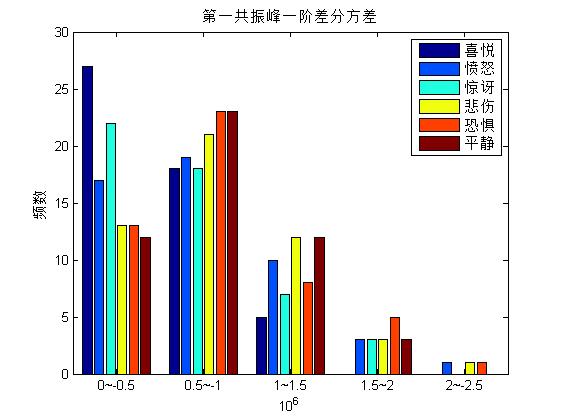
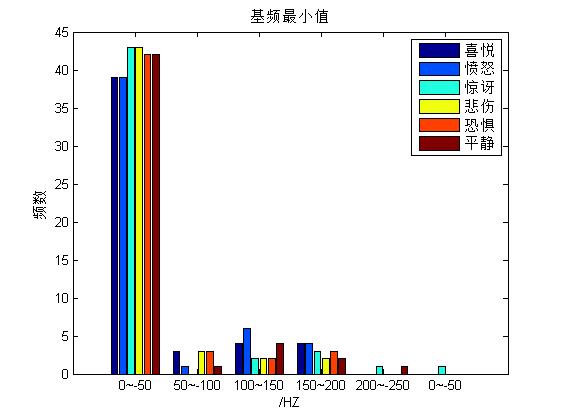


图15 第12维基频最小值

图16 第51维第一共振峰一阶差分方差

由上两图，六种感情分布均匀，可见其对语音情感的识别贡献不大。

3.6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的正确性。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。

3.7本章结论

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。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。

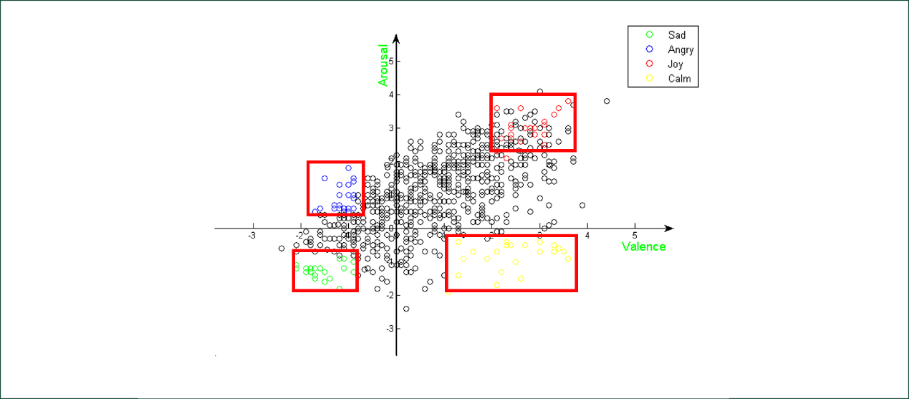


图2音乐分类

**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 |

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[11，12]，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。

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。

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，笔者将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. 种群经过选择、交叉、变异运算之后得到下一代种群。

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，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，其中一次的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

可以看出，进化代数较小时，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。

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，使它减少下次犯错的可能性。实验采用100次十折交叉验证方法进行分类，验证结果见表2、3，分类效果如图4、5所示。

表2 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3 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3 不同分类器的召回率

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 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. 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

图4 SVM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

图5 BP验证降维前后测试样本分类

由表2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3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5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，BP神经网络相对于SVM有较好的识别效果。

4.5特征分析



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。



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。谱熵均值特征，喜悦、平静、悲伤、愤怒感情由小到大分布，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。



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。



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，音调是音乐基音的变化模式【基于DBN】，音调能很好的区分这几种音乐，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，愤怒情感几种分布在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：

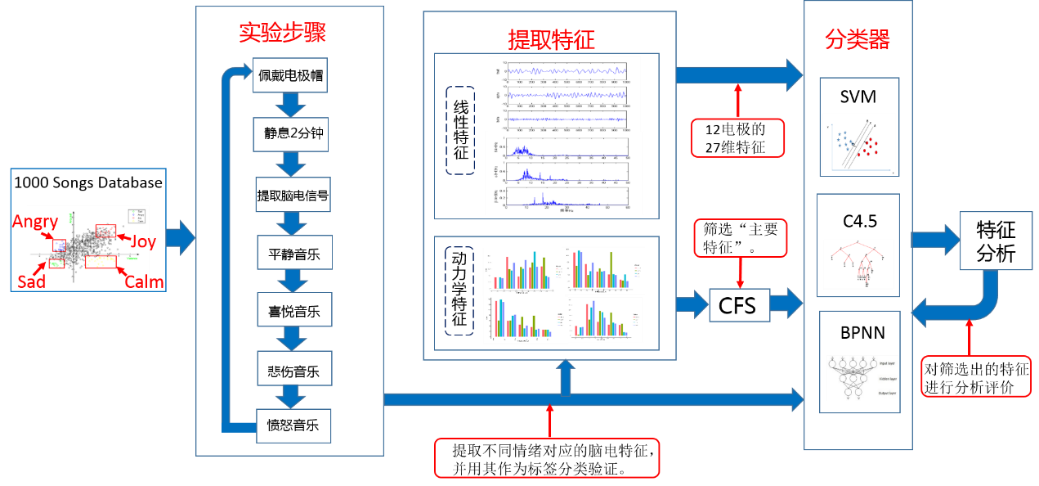


图1 实验流程图

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。

5.2脑电采集步骤

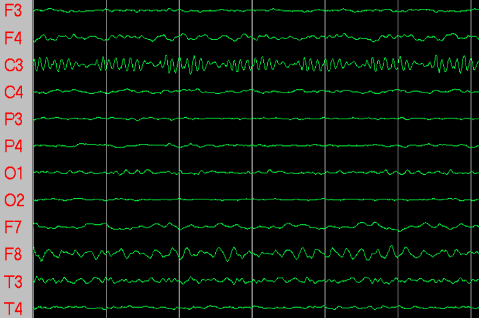


图3脑电采集设备及采集的信号

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。

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[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。

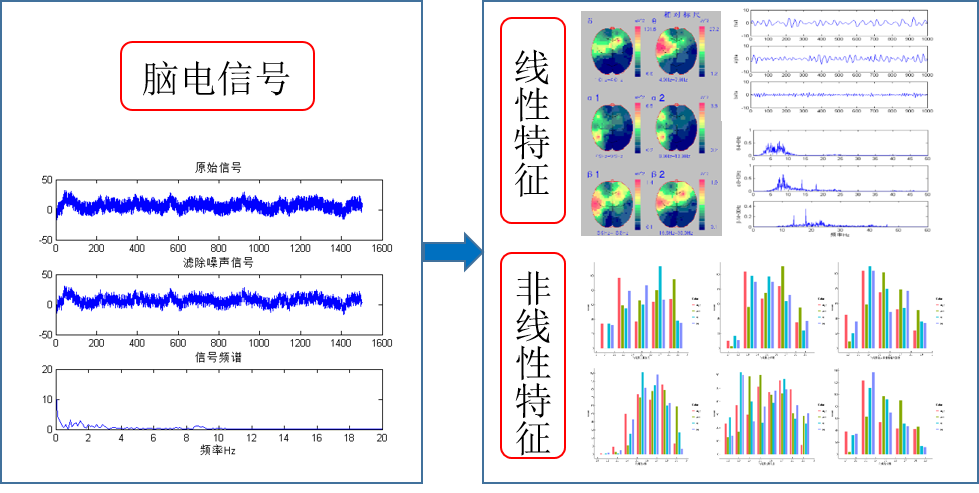


图4脑电信号提取的特征

5.3.2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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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；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（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。

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。

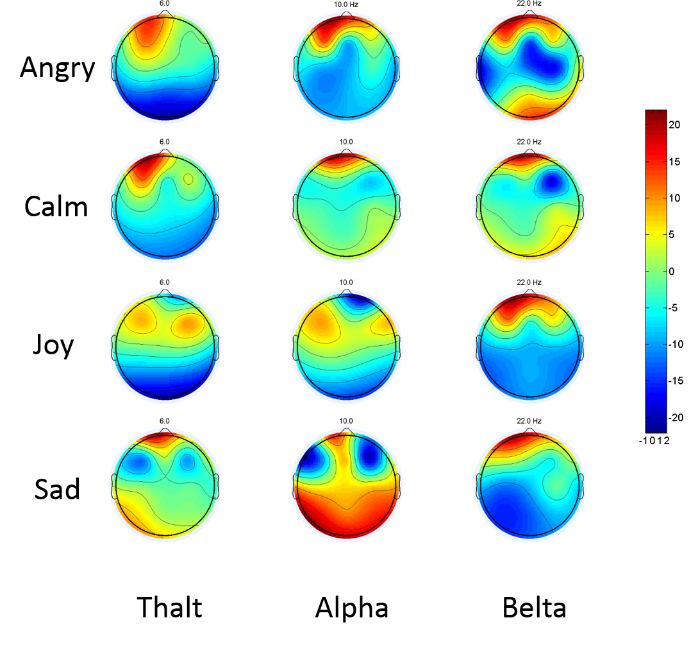


图8不同波段四种感情音乐受众脑地形图

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。

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。

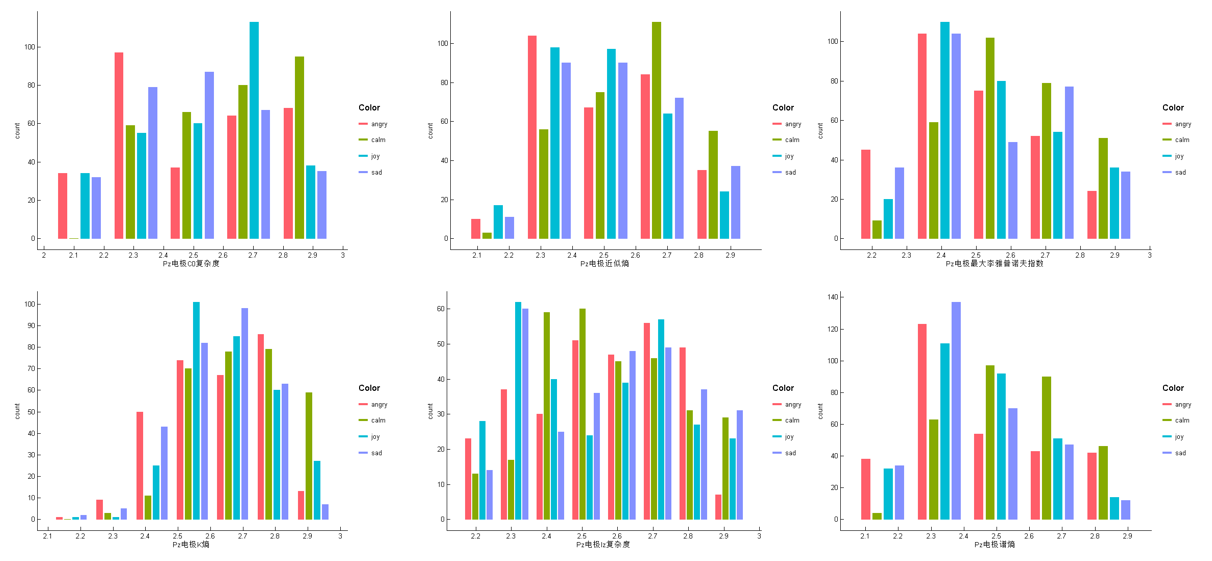


图9Pz电极非线性特征频数分布

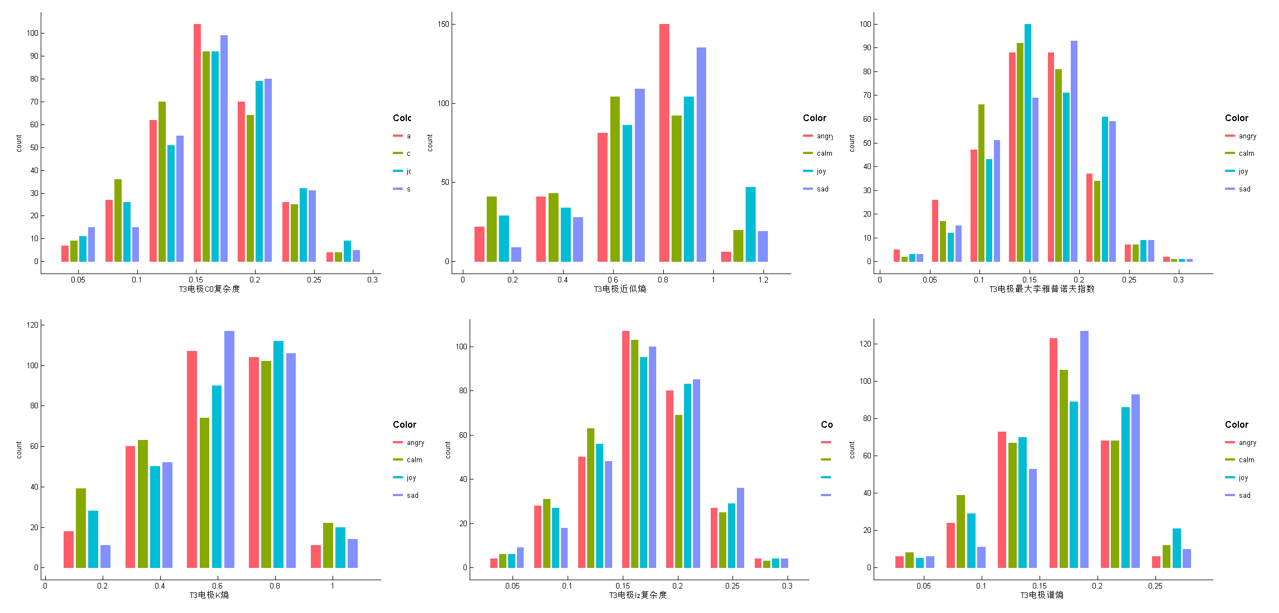
对于Pz电极，其C0复杂度对于二维情绪模型中效价维辨别度比较明显，愤怒和悲伤的效价维得分比较低，在C0复杂度多分布于较低数值段，喜悦和评价效价维得分比较高，在C0复杂度多分布于较高数值段； k熵和LZ复杂度在激活维得分较低情感（平静和悲伤）的数值多集中在较低数值，激活维得分较高情感（愤怒和喜悦）的数值多集中在较高数值；谱熵，近似熵和最大李雅普诺夫指数平静多分布在较高数值。

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。

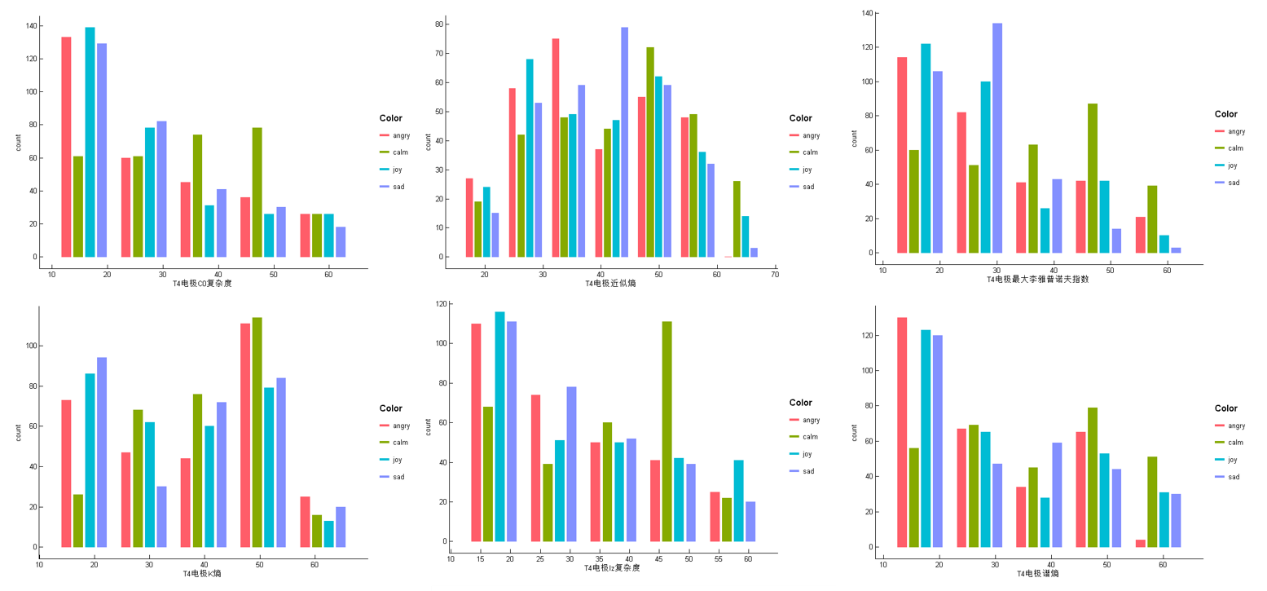


图11T4电极非线性特征频数分布

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。

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。

5.5本章结论

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：

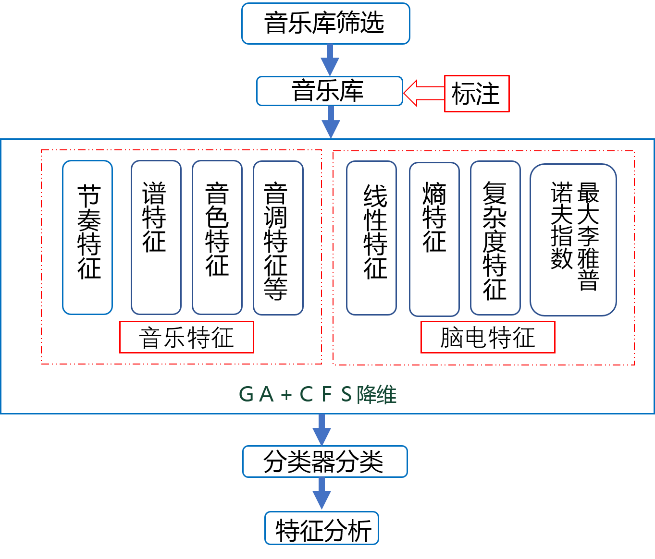
1）对于分类器而言，SVM和C4.5算法对于脑电情绪分类都比较有效。

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6章　音频信号和脑电信号关联分析

5.1 方法设计



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。

5.1 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 所示。



图 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。

5.3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 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 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 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。

5.4 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的语音和脑电情感识别系统

7.1 系统设计

LabVIEW是一种可视化的程序开发环境，其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 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（Back Propagation Neural Network，BPNN）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。

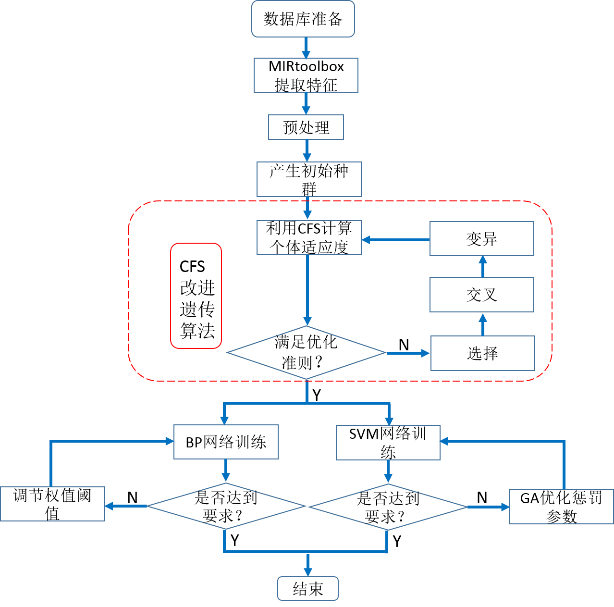
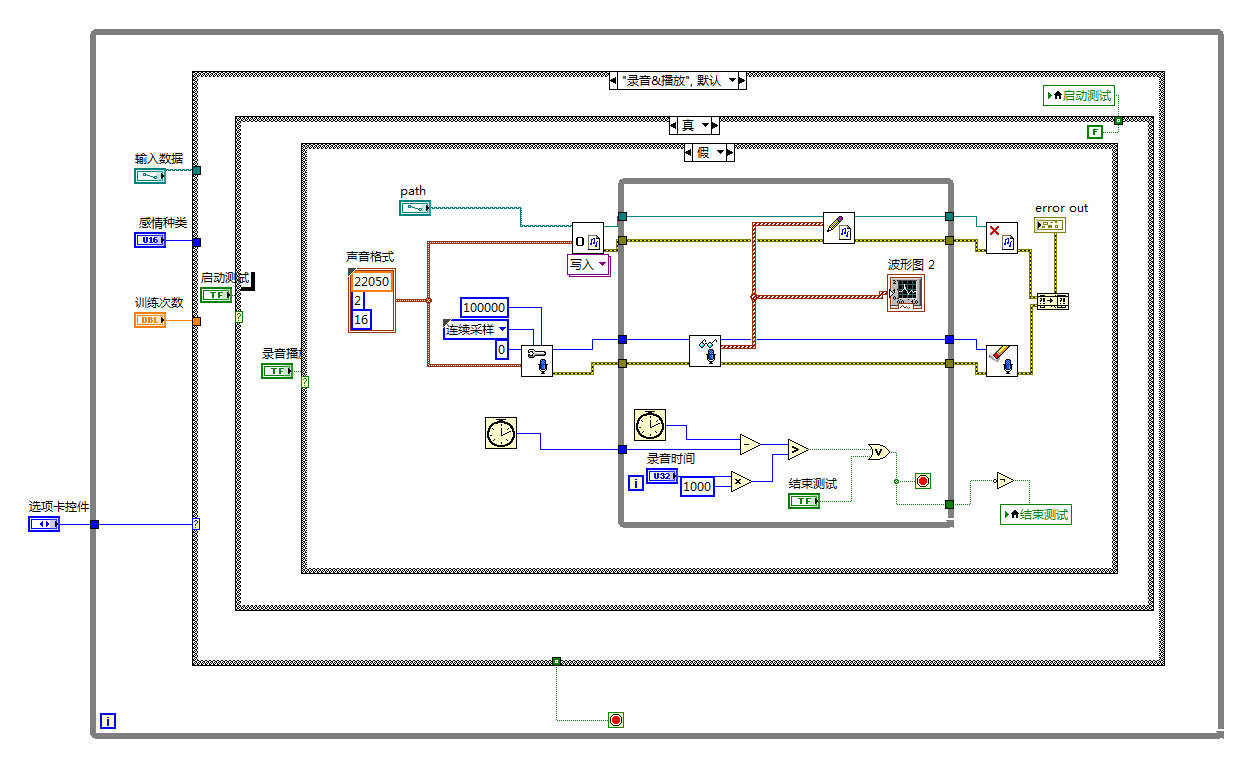


图1 总体设计流程

7.2 系统设计

LabVIEW可以直接调用Matlab程序。本实验采用Matlab2016a和LabVIEW2014混合编程，设计如图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BP神经网络的识别结果优于SVM的识别结果已被验证，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

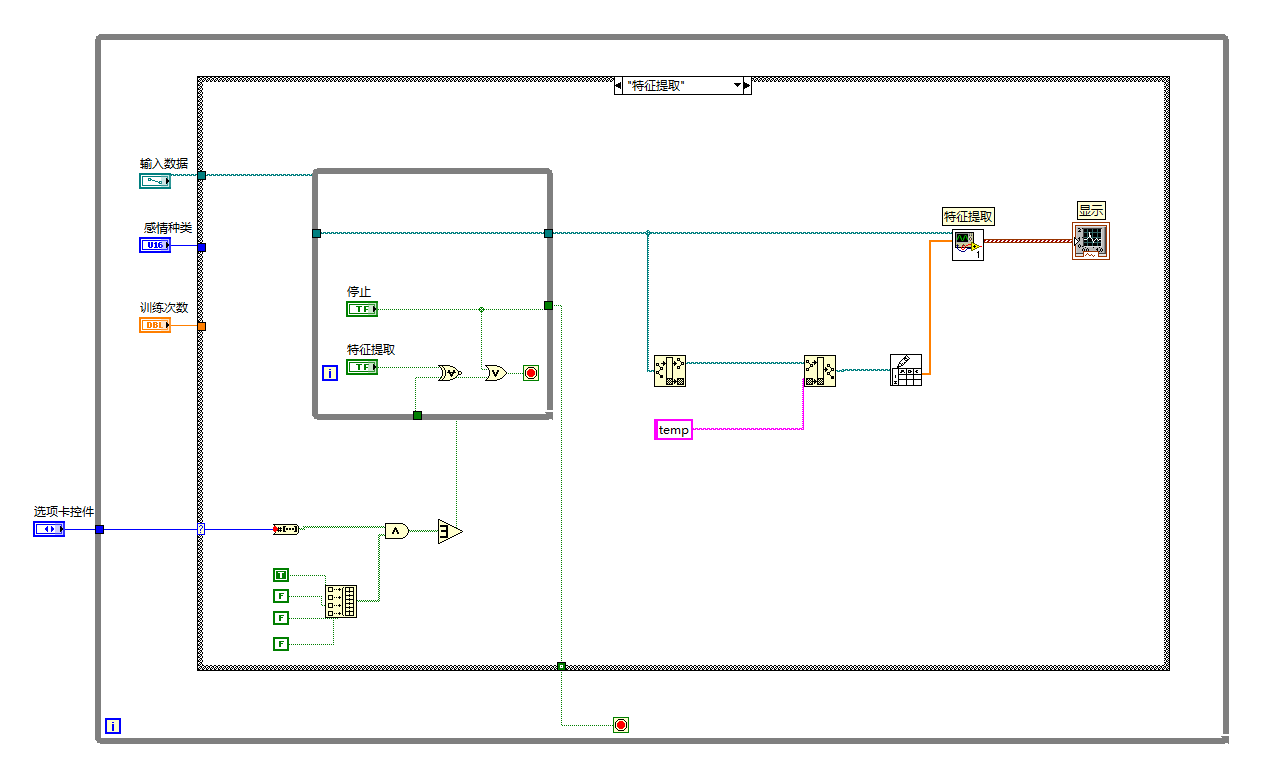


d. 录音与播放模块

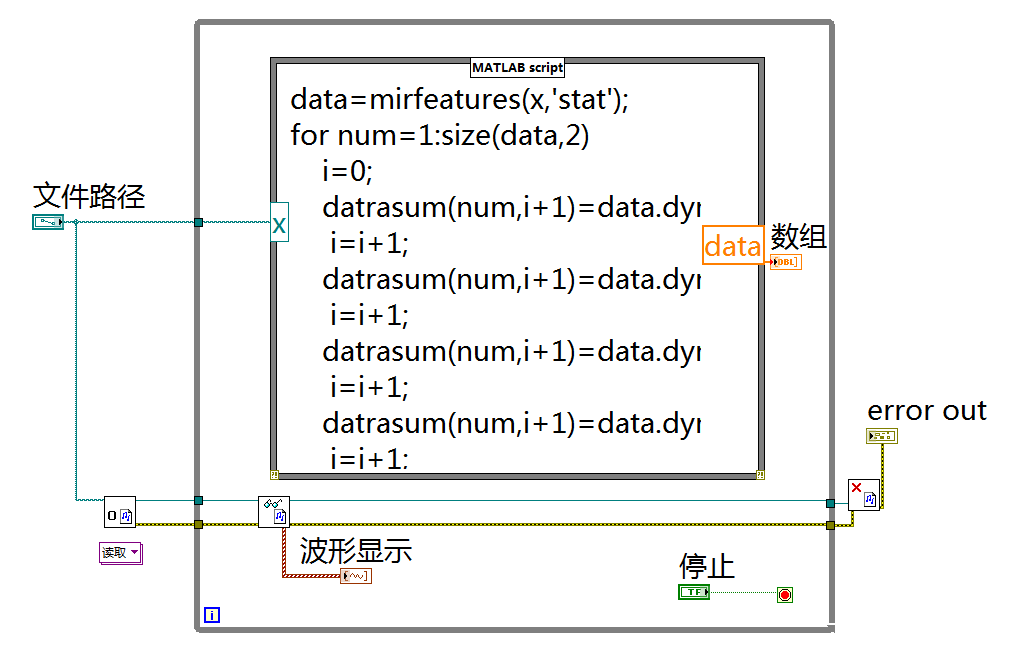
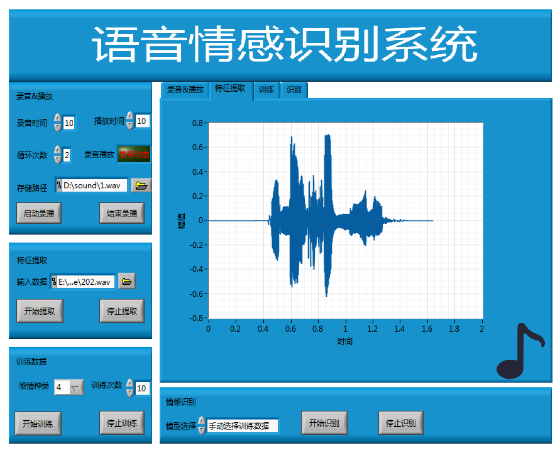
图2 语音情感识别系统

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



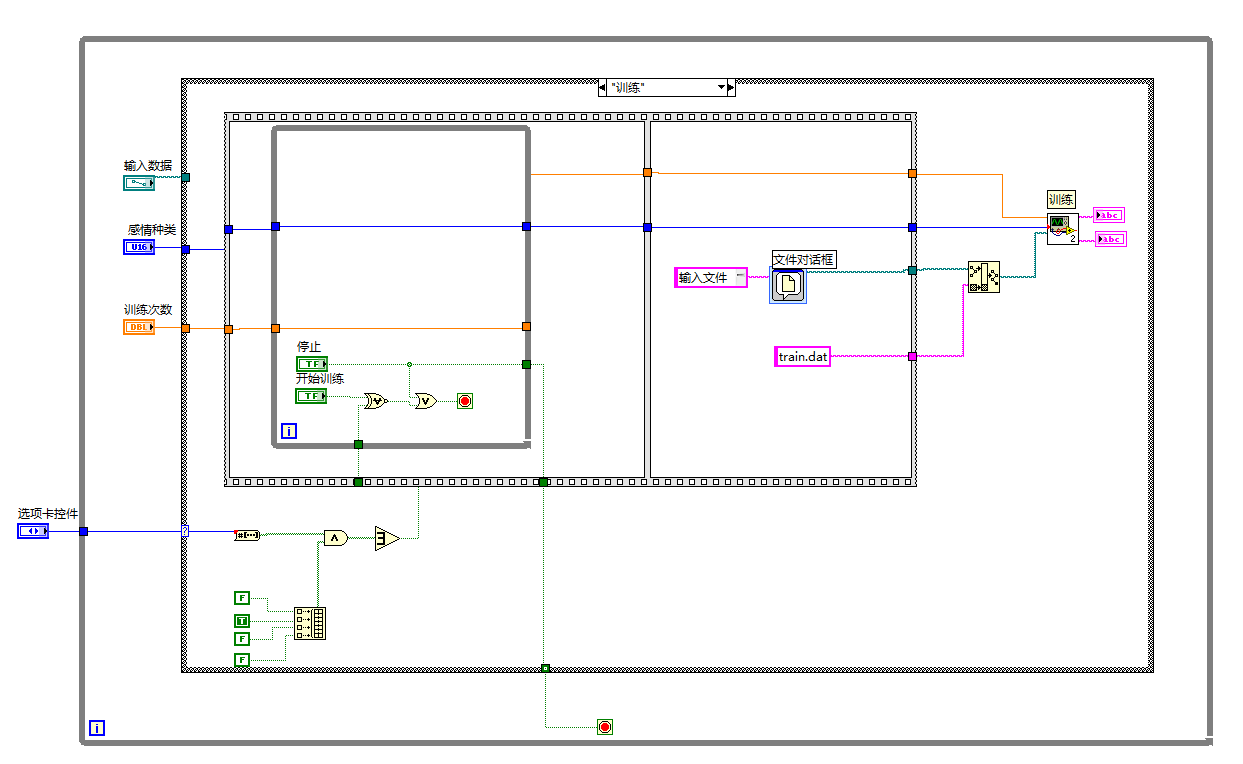
特征提取模块

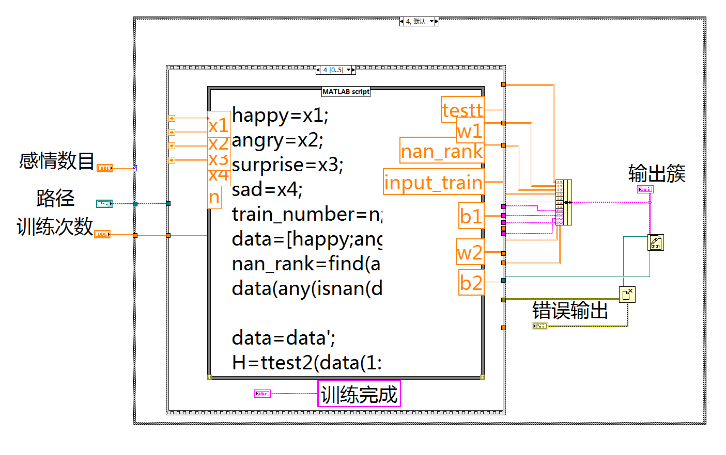
b. 特征提取子VI

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



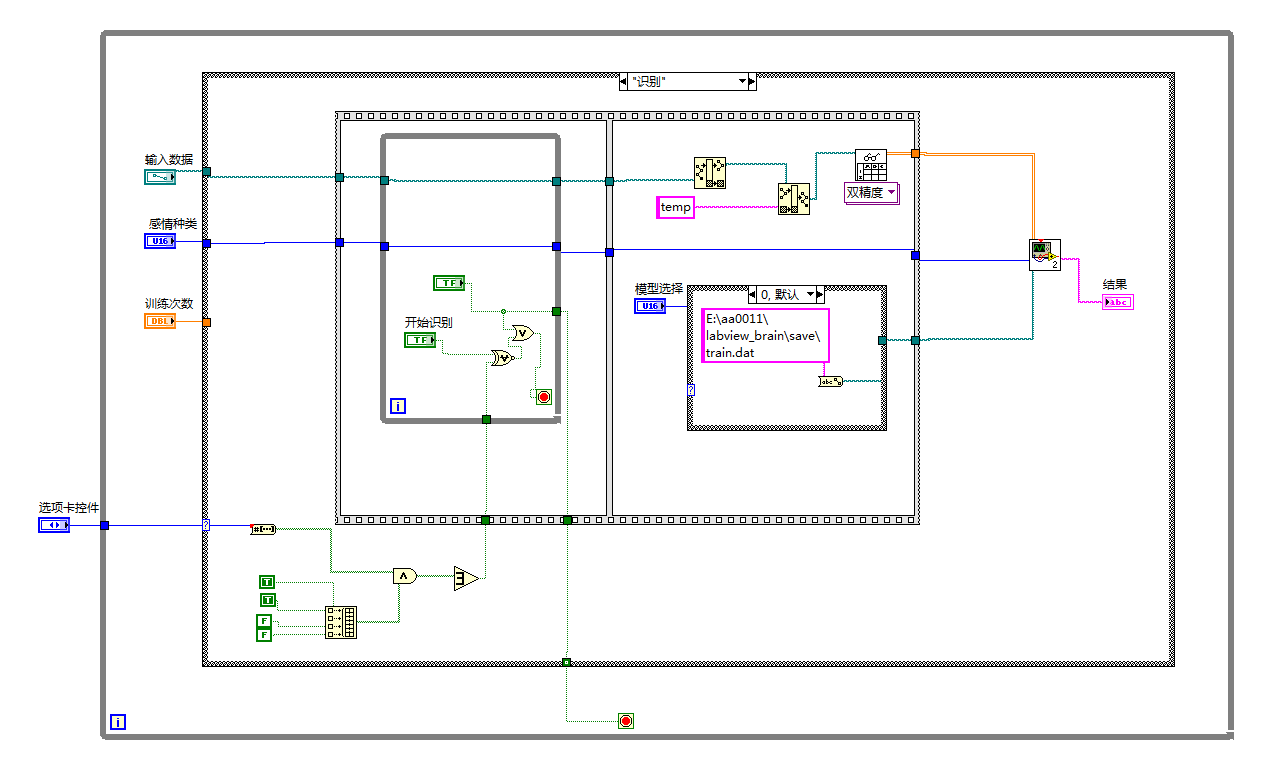
模型训练模块

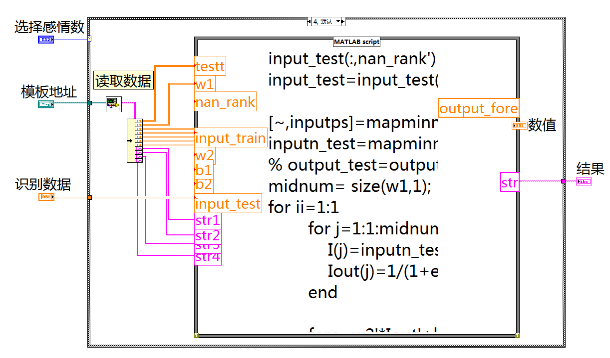
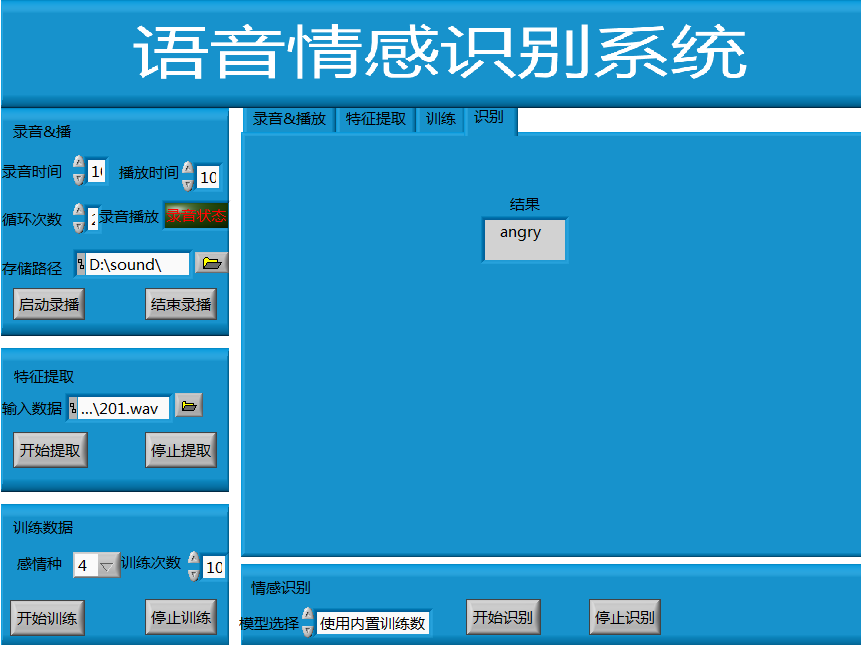
c. 训练子VI程序 训练前面板

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



识别模块

识别子程序 识别前面板

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

# 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

# 《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！