

## 语音信号的情感特征分析与识别研究综述\*

余伶俐, 蔡自兴, 陈明义

(中南大学 信息科学与工程学院, 湖南 长沙 410083)

**摘要:** 语音情感的分析与识别是近年来新兴研究课题之一, 本文介绍了近几年来国内外语音情感识别的状况, 阐述了各种人类情感分类的方法, 归纳了各种语音特征参数的提取方法以及各特征参数对情感识别的意义, 在此基础上综述了国内外在情感识别领域的研究进展与主要识别建模方法, 同时总结了各种识别建模方法的利弊。最后概括了语音情感识别领域的发展趋势, 并进行了展望。

**关键词:** 语音情感; 情感分类; 特征参数; 识别方法

**中图分类号:** TN912.34 **文献标识码:** A

## 1 引言

情感信息处理已成为提高人机交互能力的一个重要研究课题, 随着信息技术的高速发展, 人类对计算机的依赖性的不断增强。多媒体计算机系统情感信息的研究越来越受到研究者的重视, 分析情感特征、判断和模拟说话人的喜怒哀乐成为一个意义重大的研究课题之一。

斯坦福大学的 Reeves 和 Nass 通过研究发现, 在人机交互中需要解决的问题同人和人交流中是一致的, 最关键的都是“情感智能”的能力<sup>[1]</sup>。早在 1972 年, Williams 发现人的情感变化对语音的基音轮廓有很大的影响, 这是国外最早的语音情感方面的研究之一; 1990 年, 麻省理工大学多媒体实验室构造了一个“情感编辑器”对外界各种情感信号进行采样, 如人体的生理信号、脸部表情信号、语音信号来识别各种情感, 并让机器对各种情感做出适当的反应<sup>[2]</sup>; 1996 年, 日本东京 Seikei 大学提出情感空间的概念并建立了语音情感模型; 2000 年, Maribor 大学的 Vladimir Hozjan 研究了基于多种语言的语音情感识别<sup>[3]</sup>。近两年, 采用神经网络与隐马尔可夫模型为语音情感识别建模也取得一定的成功。而在国内, 语音情感分析与识别方面的研究起步比较晚, 2003 年, 北京科技大学信息工程学院提出基于 BDI Agent 技术的情感机器人语音识别技术研究方法<sup>[4]</sup>; 微软亚洲研究院, 台湾的一些大学和研究所以也进行该方面的研究; 同时, 国家自然科学基金委也立项资助哈工大进行心理紧张等情况下的 Robust 语音识别研究; 2004 年, 东南大学赵力申请了关于语音信号中情感特征分析与建模的国家自然科学基金; 近期, 江苏大学计算机学院也启动的语音情感方面的研究<sup>[5]</sup>。

语音信号的情感处理涉及到不同语种间的差异, 发展也不尽相同。英语、西班牙语、日语和德语的语音情感分析处理有较多的研究, 而汉语的情感分析与处理还处于初级阶段。本文主要从情感的分类、国内外语音特征参数提取分析方法以及国内外各种识别技术方法三方面进行综述, 并对语音情感识别发展趋势进行了总结和概括。

## 2 人的情感分类

要研究语音信号的情感, 首先需要  
根据某些特性标准, 对情感做一个有效合理的分类, 然后在不同类别的基础上研究特征参数的特性。

1954 年, Scholosberg 认为情感是由两个情感轴组成, 分别为高兴轴和赞成轴组成; 1962 年, Tomkin

表 1 情感的前期分类

研究组	主要的情感分类
Mozziacconacci S.	Joy, Anger, Sadness, Neutrality, Fear, Boredom, Indignation
Klasmeyer G.	Happiness, Anger, Sadness, Neutrality, Fear, Boredom, Disgust
Sheren KR.	Joy, Anger, Sadness, Fear, Disgust
McGilloway S.	Happiness, Anger, Sadness, Fear
Kang,Bong-Seok	Happiness, Anger, Sadness, Neutrality

\* 收稿日期: 2006-06-05 修订日期: 2006-07-23

基金项目: 国家基础研究资助项目 (A1420060159); 国家自然科学基金重点资助项目 (60234030)

建立了八种情感的分类模式<sup>[6]</sup>, 表1为各研究学者的情感分类类别。情感分类被应用到人机交互的场合, 1970年Russell<sup>[7]</sup>认为利用四个象限来定义情感空间, 如图1所示, 即四种主要的情感: 愤怒、高兴、悲伤和中性。

经过Plutchik<sup>[8]</sup>等人的多年研究, 在激活评价空间上对情感进行分析, 认为情感分布在一个圆形的结构上, 机构的中心是自然原点。通过向周围不同方向的扩展, 表现为不同的情感。情感点同自然原点之间的距离体现了情感强度。由于各种情感在自然原点的周围排成了一个圆形, 所以这种对情感进行分类的方法叫做“情感轮 (Emotion wheel)”。对于任何一个情感语句, 可以根据其情感强度和情感方向, 在情感轮所组成的二维平面中用一维情感矢量 $\vec{E}$ 来表示。其中情感强度表现为这个情感矢量的幅度值, 而情感方向则表现为该情感矢量的角度, 如图2所示。

不同于Plutchik分类手段, Fox提出的三级情感模型<sup>[9]</sup>, 则是按照情感中表现的主动和被动的程度不同将情感分成不同等级, 如表2所示。等级越低, 分类越粗糙, 等级越高, 分类越精细。

除了上面介绍的这几种分类方法外, 还有其他一些基于不同准则的分类方法。如基于不同情感表现行为的分类、基于生理特征的情感分类等。尽管这些方法在原理上有所区别, 但都有近似的表现形式。其实, 过细的情感分类并非对人际交互和情感识别有很大的意义。情感分的愈细, 情感特征愈加模糊, 识别率将会大大的降低。

所以现阶段的情感识别中, 多采用4~6种情感分类。

### 3 语音情感识别系统结构

语音情感识别是建立在对语音信号的产生机制深入研究与分析的基础上, 对语音中反映个人情感信息的一些特征参数进行提取, 并利用这些参数采用相应模式识别方法确定语音情感状态的技术, 语音情感识别系统如下图3所示。其主要处理目的是通过提取语音信号中的情感参数对隐藏在语句中的情感信息进行识别。

### 4 语音信号中情感特征参数分析及提取研究的概况

#### 4.1 根据韵律特征提取的参数

与平静语音相比, 带情感的语音在语音质量、音调和说话速度三组参数存在变化。在这三组特征参数中有一些不易测量, 语音质量和语速带有较多的主观色彩, 相比之下音调更具客观性<sup>[10]</sup>。在情感识别中, 文献[11]研究重点放在语音韵律方面, 根据韵律特征来提取情感被证明行之有效。通常与情

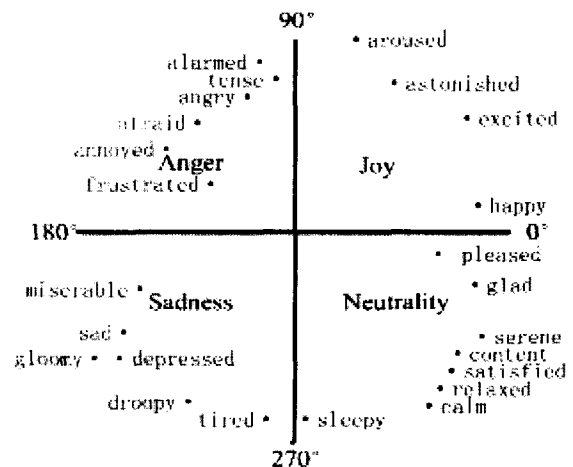


图1 情感轴模型

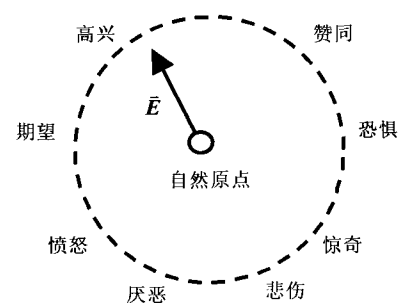


图2 情感轮模型

表2 FOX的情感三级分类模型

1st Level	Approach			Withdrawal		
2 <sup>nd</sup> Level	Joy	Interest	Anger	Distress	Disgust	Fear
3 <sup>rd</sup> Level	Pride	Concern	Hostility	Misery	Contempt	Horror
	Bliss	Responsibility	Jealousy	Agony	Resentment	Anxiety

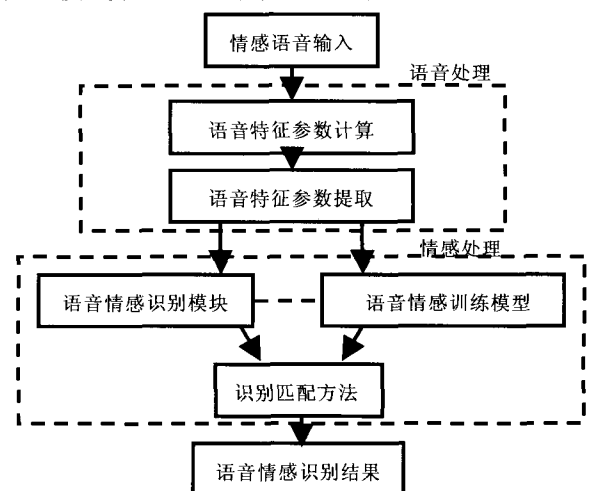


图3 语音情感识别系统框图

感相关的语音韵律特征包括：信号的持续时间、信号的振幅、基音周期和共振峰频率等。

#### (1) 持续时间<sup>[12]</sup>及语速<sup>[13]</sup>

情感语音的时间构造主要着眼于不同情感语音的发音时间构造的差别，情感语句的语速差异是基于不同情感说话速率的不同。在提取持续时间时应注意包括无声部分，因为无声部分本身对情感是有贡献的。对于汉语而言，一个汉字即为一个音节，所以用总音节数除以持续时间即得到语速。

#### (2) 能量（振幅）

信号的振幅特征与各种情感信息具有较强的相关性。求取语音信号每帧的短时振幅，分析随时间的变化情况。而且为了避免发音中无声部和噪音的影响，只考虑短时平均幅度超过某一阈值<sup>[14]</sup>时的绝对值的平均值。

#### (3) 基音周期<sup>[15]</sup>

基音是指发浊音时声带振动所引起的周期性，而基音周期是指声带振动频率的倒数，反映情感信息的重要特征参数。基音检测算法各种各样，如自相关函数法（ACF）、峰值提取算法（PPA）、平均幅度差函数法<sup>[16]</sup>（AMDF）、并行处理技术（PPROC）、倒谱法<sup>[17]</sup>（CEP）、简化逆滤波法<sup>[18]</sup>（SIFT）、小波变换法<sup>[19]</sup>等。无论采用哪种基音检测算法都可能产生检测错误，一般可对其进行后处理，如中值滤波、线性平滑处理以及组合平滑处理<sup>[20]</sup>为经典的后处理方法。

#### (4) 共振峰

声道可以看成是一根具有非均匀截面的声管，在发音时起共鸣器作用。当元音激励进入声道时会引起共振特性，产生一组共振频率，这就是共振峰，它一般包括共振峰频率的位置和频带宽度。语音信号的前4个共振峰经常用作情感识别的特征参数。共振峰亦为反映声道特性的一个重要参数，因为不同情感的发音可能使声道有不同的变化，所以能够预料到不同情感发音的共振峰的位置不同。提取共振峰参数主要有带通滤波器组<sup>[21]</sup>法、倒谱法<sup>[22]</sup>和线性预测<sup>[23]</sup>法（LPC）。其中，LPC法是先求出预测系数，然后用预测系数估计出声道的频响曲线，再用峰值检出法算出各共振峰的频率。此方法是频谱包络估计器的最新形式，提供了一个优良的声道模型。

### 4.2 结合韵律与语音学特征提取的参数

语音特征是声音的类型，包括元音、辅音和它们的发音。韵律特征指升调、降调、口音或重音。文献[24]中将韵律特征和语音特征结合：语音能量（ $P$ ）和高音（ $p$ ）；语音特征：12个LPC参数（ $C_1, C_2, C_3 \dots C_{12}$ ）Delta LPC参数（ $d$ ）。当一句话的起点和终点确定后，一句话被分为等间隔的20个部分，用向量  $f_1, f_2 \dots f_{20}$  表示，每个向量包含15个特征参数（ $P, p, C_1, C_2, C_3 \dots C_{12}, d$ ）。具体表示如下：

$$F_v = (f_1, f_2 \dots f_{20})$$

$$f_n = (P_n, p_n, C_1, C_2, C_3 \dots C_{12}, d_n)$$

### 4.3 全局与时序结构特征并用的参数提取

东南大学的赵力<sup>[25]</sup>在全局特征的基础上进行语音情感特征分析，并采用情感语句中各元音时序结构作为新的特征，针对不同语句中包含的不同元音个数的情况，进行零补齐、全局均值补齐、前均值补齐三种不同的规整方法。最终发现利用全局特征和时序特征相结合，对时序特征采用前均值补齐，并使用修正二次判别函数，获得94%的平均情感识别率。

### 4.4 非线性分析提取的特征参数

美国科罗拉多大学针对一种特殊语音：应激环境下语音，提出了基于TEO（Teager energy operator）的非线性分析方法。Teager<sup>[26]</sup>认为语音的生成是靠声门激励及声道中的涡流共同产生。Teager和Kaiser提出了非线性能量算子的连续形式：

$$\psi_c[x(t)] = \left[ \frac{d}{dt} x(t) \right]^2 - x(t) \left[ \frac{d^2}{dt^2} x(t) \right] \quad (1)$$

其中  $\psi_c$  是连续 Teager 的能量算子 (TEO),  $x(t)$  是一个连续语音信号。Kaiser 又提出离散能量算子,  $x(n)$  是离散语音信号:

$$\psi[x(n)] = x^2(n) - x(n+1)x(n-1) \quad (2)$$

以非线性涡流理论为基础, 提出了调幅-调频模型作为新的语音模型<sup>[27]</sup>, 语音信号通过中心频率在共振峰附近的 Gabor 带通滤波后, 就可以看作调幅-调频信号。该信号是以共振峰为载波频率, 进行幅度调制和频率调制, 且幅度与频率的变化是缓慢的, 远小于载波频率。语音信号就是由若干个单个共振峰幅度调制和频率调制累加的结果, 即语音信号可表示为:

$$s(t) = \sum_{m=1}^M r_m(t) \quad (3)$$

$$r_m(t) = a_m(t) \cos \left\{ 2\pi \left[ f_{cm}t + \int_0^t q_m(\tau) d\tau \right] + \theta \right\} \quad (4)$$

其中  $s(t)$  是语音信号,  $r_m(t)$  是以单个共振峰为载频的幅度调制和频率调制信号,  $M$  是语音的总的共振峰个数, 载波频率  $f_{cm}$  与单个共振峰频率相对应,  $a_m$  是时变幅度,  $q_m$  是频率调制信号。因此瞬时频率  $f(t) = f_{cm} + q(t)$  就反映了以共振峰为中心的频率调制过程。能量分离算法 (ESA energy separate algorithm) 根据能量算子的特性可实现在离散域调幅-调频信号的幅度包络  $|a(n)|$  和瞬时频率  $f(n)$  的估计。其中,  $x(n)$  是 AM-FM 信号,  $y(n) = x(n) - x(n-1)$  表示时域差分,  $f(n)$  表示采样信号的 FM 成分,  $a(n)$  表示采样信号的 AM 成分。

$$f(n) \approx \frac{1}{2\pi T} \arccos \left[ 1 - \frac{\psi[y(n)] - \psi[y(n+1)]}{4\psi[x(n)]} \right] \quad (5)$$

$$|a(n)| = \sqrt{\frac{\psi[x(n)]}{1 - \left( 1 - \frac{\psi[y(n)] - \psi[y(n+1)]}{4\psi[x(n)]} \right)^2}} \quad (6)$$

在研究稳健型语音识别系统时, 将多分辨率子带分析法 (Multirate Subband Analysis) 与 TEO 能量算子相结合<sup>[28]</sup>, 提出一种新的非线性特征参数 TEOCEP: 该特征识别性能优于使用短时能量计算得到的子带倒谱参数 SUBCEP。基于 TEO 的非线性特征在应激环境下的语音分析, 虽属起步但令人鼓舞, 预示 TEO 在语音分析中的前景乐观, 它体现一定的动态语音生成时的非线性变化, 同时也需要在该领域有更多的深入研究和完善。

## 5 语音信号中主要识别方法研究的概况

### 5.1 矢量分割型马氏距离判别法

利用情感语句的全局特征和各时序结构特征的组合<sup>[29]</sup>, 采用马氏距离进行语音信号中情感特征的识别。

$$d^2(X) = (X - \mu)^T \Sigma^{-1} (X - \mu) \quad (7)$$

这里  $X(x_1, x_2, \dots, x_p)$  是维数为  $P$  的输入特征向量,  $\mu$  是参考样本的均值向量,  $\Sigma$  是参考样本的协方差矩阵。式 (7) 计算量和内存所占容量会很大, 并且随着  $X$  的维数的增大, 协方差矩阵的推定误差将增大, 降低判别率。可改进为式 (8), (9) 形式, 偏置  $b$  是为了缓和特征值误差而引起的识别率下降, 其值由实验而定。利用一种新的矢量分割型马氏距离判别式。

$$d^2(X) = \sum_{k=1}^p \frac{1}{\lambda_k + b} (X - \mu, \phi_k)^2 \quad (8)$$

$$d^2(X) = (X - \mu)^T \Sigma^{-1} (X - \mu) = \sum_{i=1}^M (X_i - \mu_i)^T \Sigma_i^{-1} (X - \mu) = \sum_{i=1}^M \sum_{k=1}^K \frac{1}{\lambda_k + b} (X_i - \mu_i, \phi_{ik})^2 \quad (9)$$

### 5.2 主元分析法 (PCA)

文献[30]中对于要识别的数据, 假设有  $N$  个样本信号, 使用的特征参数个数为  $K$  个。对这个样本

信号求  $K$  个特征参数所对应的协方差阵, 然后对协方差阵进行特征值分解得到  $K$  个特征值和相应得特征矢量。对于不同的情感, 利用样本库中样本分别计算不同的主元  $k$  相对于不同的情感类型  $j$  所具有的均值  $\mu_{jk}$  和方差  $\sigma_{jk}$ 。并用下列各式进行最大可分性处理, 其框图如图 4 所示:

$$L_k = C_J^2 \sqrt{\sum_{i=1}^J \sum_{j=i+1}^J |\sigma_{ik} - \sigma_{jk}|^2} \quad (10)$$

$$M_k = \frac{1}{J} \left| \sum_{j=1}^J \sigma_{jk} \right| \quad (11)$$

$$H_k = \frac{L_k}{M_k} \quad (12)$$

$J$  是采用的情感的类型个数,  $L_k$  表示第  $k$  个主元在情感类别中的分离性,  $M_k$  表示第  $k$  个主元在情感类别中的集中性。用

$H_k$  来反映主元在情感类别中的辨别能力,  $H_k$  越大, 辨别能力越强。按顺序对  $K$  个主元进行排列, 选取  $p$  个  $H_k$  较大的主元作为识别用的主元。

$$Z_k = \langle A_k \bullet X_{std} \rangle \quad (13)$$

利用已知各参数在不同情感中的均值和方差对该特征参数进行标准化得到  $X_{std}$ , 然后将  $X_{std}$  对选取得各主元的基向量  $A_k$  分别进行投影求和, 获得待识别语音在各个有效主元的得分值  $Z_k$ 。再利用 (14) 求出各有效主元的综合概率:

$$P_j = \prod_{k=1}^p \left[ \frac{1}{\sqrt{2\pi}\sigma_{jk}} \exp \left\{ -\frac{|Z_k - \mu_{jk}|^2}{2\sigma_{jk}^2} \right\} \right] \quad (14)$$

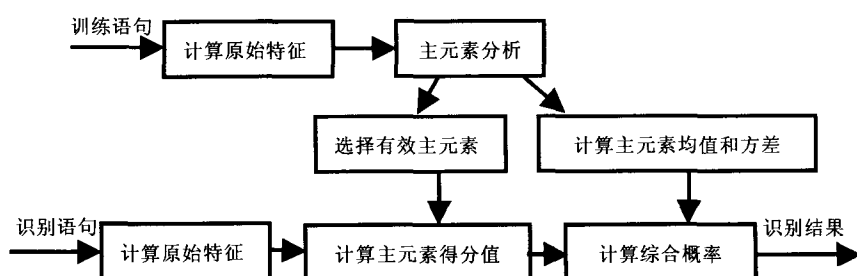


图 4 训练和识别方法框图

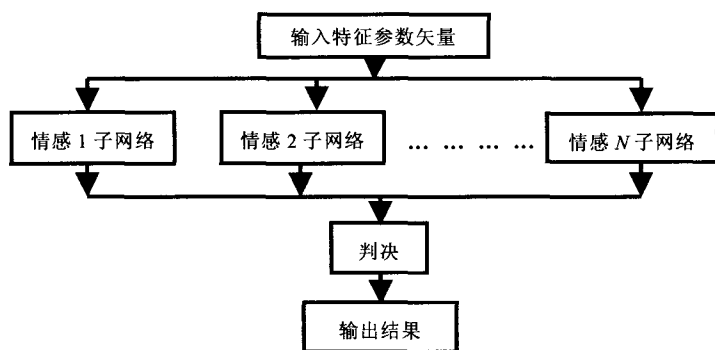


图 5 利用神经网络识别语音情感

### 5.3 神经网络方法 (ANN)

神经网络是一种具有大量连接的分布式处理器, 具有通过学习获得知识并解决问题的能力。对于情感识别的问题, 首先对每种情感构造一个网络。这种类型的网络叫 OCON (One-Class-In-One-Network)。可以采用图 5 所示的子网络并行结构, 将提取出的特征矢量输入到每一个子神经网络中, 每个子网采用三层神经元的 MLP 结构, 输出界于 0 到 1 之间的数值, 表示输入的参数矢量可能属于某种情感的似然程度。

对于每个子网络, 首先构造一个神经网络结构, 选定各层传递函数, 根据给定一个衡量理想输出和实际输出的能量函数。不同情感子网络具有相同的结构, 仅仅在权值上有所不同。利用这个函数对网络中各个节点的权值, 求偏导确定实际的权值, 使每个子网络逼近于实际的情感概率模型。在识别时选择似然程度最大的情感作识别结果。

正是因为神经网络具有通过自己组织和学习能够具有较强的分类能力, 也有一些研究者将它应用到了语音情感的识别研究方面。在 2000 年, J.Nicholson 等人<sup>[31]</sup>利用神经网络对 8 种情感进行了识别, 识别率达到 50%。在 2003 年, K.H.Kim 等人<sup>[32]</sup>采用了自适应神经网络对语音情感状态进行了研究。在他们的研究中, 除了采用传统的语音特征外, 还结合了发音时的一些生理特征参数 (如心跳、心电图等) 进行训练和识别, 也达到了 70% 左右的识别率。浙江大学的王青<sup>[33]</sup>对输入的特征矢量及神经网络训练数据作了主成分分析。降低输入的特征矢量的维数, 去除了冗余信息。采用三种类型的神经网络 (OCON, ACON 和 LVQ) 识别汉语语音信号中的四大情感。

神经网络之所有引起人们的兴趣, 主要在于其并行分布处理的能力, 这是与传统方法截然不同的, 同时也因为它具有以下几个方面的优点:

- (1) 高度的非线性和极强的分类能力;
- (2) 自组织和自学习的能力, 能够在学习的过程中发现并总结信号的特征;
- (3) 具有很强的鲁棒性和容错能力。

但同时, 神经网络的不足之处有:

- (1) 随着网络的增大, 需要的学习时间增加很快;
- (2) 网络的输入节点是固定的, 所有的输入必须是同样长度的, 输入的矢量在输入之前要进行预处理;
- (3) 网络的存储量问题, 即对于一个给定的任务, 网络设计为多大最好, 仍是一个在理论尚未解决的问题。

#### 5.4 隐马尔可夫模型 (HMM)

隐马尔可夫模型是一个离散时域有限自动机系统, 该模型首先在识别领域得到广泛的应用, 而语音情感识别作为语音识别的一个大类, 也将 HMM 引入到研究中。

利用马尔可夫链的信号模型技术, 以抽象的概率模型作为参考模板来反映信号的统计特性, 从而对随机过程建模。利用马尔可夫链的信号模型技术, 以抽象的概率模型作为参考模板来反映信号的统计特性, 从而对随机过程建模。假定有一个含有  $V$  种情感状态语音的待识别语音库, 每个情感状态都有  $K$  个语音样本。对于情感语音的识别, 要完成以下工作:

(1) 语音库中的每个情感状态  $V$  建立其对应的隐马尔可夫模型  $\lambda_v$ , 即得到和每种情感状态用于训练的所有观察序列最为匹配的参数模型  $(\pi, A, B)$ 。

(2) 对于任何一个待识别的语音样本, 执行图 6 中所示的步骤, 在图 6 中  $a_{ij}$  是状态转移矩阵  $A$  中的元素,  $b_j(0)$  是分布矩阵  $B$  的列向量, 自左向右的 HMM。即首先通过特征提取得到待识别单词的观测矢量序列  $O$ ; 随后, 依次对每个模型计算其匹配得分, 匹配得分可以通过 Viterbi 算法得到; 最后, 选择匹配得分最高的模型作为识别结果。即:

$$v^* = \arg \max_{1 \leq v \leq V} [P(O/\lambda_v)] \quad (15)$$

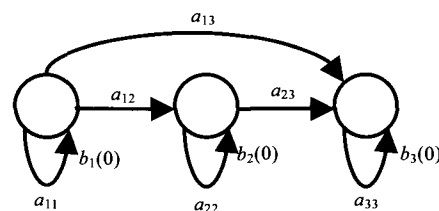


图 6 HMM 的状态转移图

采用 MEL 频率语音能量系数和 HMM 分类方法, 能够进行有效的识别语音所包含的情感。但不足反映情感的细节, 对情感进行精细的区分<sup>[34]</sup>。文献[35]中进行进一步的研究, 采用了两种方法: 一种是从语音信号的音高和能量轮廓中提取参数, 用单一状态 HMM 对全局进行计算; 另一种使用连续的隐马尔可夫模型, 使用瞬时特征取代全局参数。研究发现采用多状态的模型能取得更好的识别效果。

采用 HMM 模型进行语音情感识别对存在一些问题需要进一步研究:

- (1) 建立、训练都要较多的时间;
- (2) 计算的时间复杂度也较高;
- (3) 无法满足对语音情感最终达到实时识别的目标要求。

#### 5.5 混合高斯模型法(GMM)

高斯混合模型<sup>[36]</sup>是只有一个状态的模型, 在这个状态里具有多个高斯分布函数。

$$P_k = \sum_{i=1}^N \omega_i f_i(\vec{y}) \quad (16)$$

其中  $f_i$  是一个高斯分布函数, 不同高斯分布之间的加权系数  $\omega_i$  满足条件:

$$\sum_{i=1}^N \omega_i = 1 \quad (17)$$

在训练时首先利用矢量量化 (VQ) 抽样各类情感中有效主元矢量集的码本, 并对每个码字求出相应的方差, 这样每个码字和相应的方差就可以组成一个高斯分布函数。在识别时, 对于某种情感主元特征矢量  $\mathbf{Y}$  求相对于每个情感类别的概率值, 概率最大的为识别结果。

## 5.6 其它识别方法

目前, 关于情感的研究正处在不断深入之中, 各国都投入了大量的资金进行研究。而其中语音的情感识别是其中的热点之一, 正越来越受到人们的重视。各种模式识别方法, 如线性判别分类 (Linear Discriminant Classifier),  $k$  最近邻法 ( $k$ -Nearest Neighborhood)、支持向量机 (Support Vector Machine) 等, 都被应用于语音的情感识别, 也取得了不错的效果。

Chul Min Lee 等<sup>[37]</sup>把语音情感识别归结为模式识别问题, 并提出三种方法: 线性判别分类 (LDC)、 $k$  最近邻法 ( $K$ -NN) 和支持向量机 (SVM)。LDC 是带有高斯概率分布的参数方法, 估算完参数的均值和方差后, 利用 Bayes 准则计算出最大后验概率对情感进行分类。 $K$ -NN 通过计算  $k$  邻域的平均值估计每类情感的局部后验概率。在 SVM 中, 用非线性的映射方法将输入向量映射到更高维的特征空间进行计算。

## 6 发展趋势及结语

语音信号处理的情感信息处理是一综合听觉生理学、心理学、认知科学以及信号与信息处理等多学科的研究课题。通过对情感的分类、语音特征参数提取分析方法和情感识别方法的综述, 发现语音情感识别技术已经在很多方面取得了一系列的进展, 但仍有许多问题值得研究, 即其将来的发展趋势可概括如下:

(1) 情感计算及人工心理理论是新型交叉学科, 很多理论方法还很成熟等。人机交互与情感识别的对象一般是自然人, 与较小素材的人工材料实验样本有一定差异。所以将来的研究将努力扩展语音情感库的范畴, 由特定人群到普通人群, 由特定环境到普通环境。

(2) 对于情感语音的构成进行进一步深入的分析, 找出对于情感的表达有贡献的新的特征参数, 并将其加入到识别参数中, 以获得更高的识别率; 进一步研究更合理的语音情感分类识别方法, 使实时语音情感识别更方便、更有效。

(3) 组合多模式的情感识别: 情感表现是由面部表情、语音情感和身体姿势三个方面体现的, 因此情感交流是多模式信息组合。组合多模式的识别能大大提高识别效率。但如何对这些模型进行组合, 如何确定语音模式同其它模式之间的相互关系, 都是语音情感识别中需要继续研究的课题。

(4) 利用语义的语音情感识别: 由于说话人表达情感有其特定的环境, 这样在语音情感识别的研究中要考虑语义所具有的情感倾向性。目前在汉语词汇中词义和情感之间的联系还没有得到研究<sup>[38]</sup>, 如何利用词义, 将语音语义识别同情感识别相结合, 在更高层次上把握说话人的情感是一个重要研究课题。

(5) 在不同时间尺度上研究语音情感: 情感同时间尺度有密切的关系, 而目前仅局限于较短时间域研究<sup>[39]</sup>, 要通过语音参数对人的情感有更多了解, 需要进行多尺度分析。需要在较短的时间尺度上把握人的某个时刻状态, 也要更长时间尺度上把握某段时间内的心情。因此研究不同时间尺度上的语音情感, 对提高人机交互能力有着重要意义。

随着新型的人机交互技术的快速发展, 语音处理领域产生了许多热门的研究方向, 如个人机器人、语音识别、语音合成、语音的转换、语言翻译、个人隐私保护等。语音情感识别还可以应用在教学辅导、自动告警及娱乐等方面。随着 Internet 的普及以及计算机性能的大幅提高, 语音情感识别技术将被广泛应用在更多领域, 会有非常好的经济效益和社会价值。

## 参考文献:

- [1] Y M Cheng, D OShaughnessy, P Mermelstein. Statistical Recovery of Wideband Speech Form Narrowband Speech [J]. IEEE Trans. Speech

- and Audio Processing, 1994, 2(4): 544-548.
- [2] S Chennoukh, A Gerrits, G Miet, R Sluijter. Speech Enhancement via Frequency Extension using Spectral Frequency [A]. **Proc. ICASSP** [C]. Salt Lake City, 2001. 5.
  - [3] 陈建厦. 语音情感识别综述[A]. 第一届中国情感计算与智能交互学术会议[C]. 北京, 2003.
  - [4] 谷学静, 石志国, 王志良. 基于 BDI Agent 技术的情感机器人语音识别技术研究[J]. 计算机应用研究, 2003, (4): 24-26.
  - [5] 余华, 王治平, 赵力. 语音信号中情感特征的分析与识别[J]. 电声技术, 2004, (3): 30-34.
  - [6] Tomkins S. Affect Imagery [M]. Consciousness Springer Publishing Company, New York, 1962.
  - [7] Russell J A. A Circumflex Model Of Affect [J]. **Personality And Social Psychology**, 1980, 39: 1167-1178.
  - [8] R Plutchik. A General Psycho Evolutionary Theory Of Emotion [J]. **Emotion Theory Research and Experience**, 1980, 1.
  - [9] Kyung Hak Hyun, Eun Ho Kim, Yoon Keun Kwak. Improvement of Emotion Recognition by Bayesian Classifier Using Non-Zero-Pitch Concept [A]. **2005 IEEE International Workshop on Robots and Human Interactive Communication** [C].
  - [10] V Kaostov, S Fu+kuda. Emotion In User Interface ,Voice Interaction System Systems Man And Cybernetics [A]. **2000 IEEE International Conference On** [C]. 2000, 8(2): 798-803.
  - [11] D Galanis, V Darsinos, G Kokkinakis. Investigating Emotional Speech Parameter for Speech Sunrthesis [J]. **Icccs96**, 1996. 1227-1230.
  - [12] 陈雁翔, 戴蓓蓓, 周曦, 李辉. 基于对话语音的与文本无关的说话人确认系统的研究[J]. 中文信息学报, 2004, 18(2): 36-39.
  - [13] 赵杰. 论汉、日两语不同的音长特征[J]. 汉语学习, 2002, 8(4): 44-49.
  - [14] 钱向民. 包含在语音信号中情感特征的分析[J]. 电子技术应用, 2000, (5): 18-21.
  - [15] 王治平, 赵力, 邹采荣. 基于基音参数规整及统计分布模型距离的语音情感识别[J]. 2006, 31(1): 28-34.
  - [16] 张文耀, 许刚, 王裕国. 循环 AMDF 及其语音基音周期估计算法[J]. 电子学报, 2003, 6(6): 886-890.
  - [17] 殷柏涛, 张志涛. 用同态解卷估计揭稻虱鸣声的基音周期[J]. 声学与电子工程, 1994, (1): 1-8.
  - [18] 陈洪立. 一种基音周期估计方法[J]. 江西通信科技, 2004, 6(2): 32-34.
  - [19] 许开宇, 张曙. 基于小波分析的汉语语音识别[J]. 哈尔滨工程大学学报, 1996, 17(4): 20-26.
  - [20] 柏静, 韦岗. 一种基于线性预测与自相关函数法的语音基音周期检测新算法[J]. 电声技术, 2005, (8): 43-46.
  - [21] 章文义, 朱杰, 陈斐利. 一种新的共振峰参数提取算法及在语音识别中的应用[J]. 计算机工程, 2003, (13): 44-48.
  - [22] 王晓亚. 倒谱在语音的基音和共振峰提取中的应用[J]. 无线电工程, 2004, 34(1): 57-60.
  - [23] 郁伯康, 郁梅. LPC 方法提取语音信号共振峰的分析[J]. 电声技术, 2000, (3): 4-8.
  - [24] Nicholson J, Takahashi K, Nakatsu R. Emotion Recognition In Speech Using Neural Network [A]. **Neural Information Processing ICONIP'99.6<sup>th</sup> International Conference On** [C]. 1999, 20(2): 495-501.
  - [25] 赵力, 王治平, 卢韦, 等. 全局和时序结构特征并用的语音信号情感特征识别方法[J]. 自动化学报, 2004, 30(3): 423-429.
  - [26] 蒋丹宁, 蔡莲红. 基于语音声学特征的情感信息识别[J]. 清华大学学报(自然科学版), 2006, 46(1): 86-89.
  - [27] Maragos P, Kaiser JF, Fquatieri T. Energy Separation In Signal Modulation With Application To Speech Analysis [J]. **IEEE Trans On Signal Processing**, 1993, 41(10): 3024-3051.
  - [28] Jabloun F. Large Vocabulary Speech Recognition in Noisy Environments [D]. Ankara Turkey, Bilrent University, 1998.
  - [29] 赵力, 将春辉. 语音信号中情感特征分析和识别的研究[J]. 电子学报, 2004, 4(4): 606-609.
  - [30] 赵力, 钱向民, 邹采荣, 等. 语音信号中的特征分析与情感识别研究[J]. 通信学报, 2002, 12(10): 18-25.
  - [31] J Nicholson, K Takahashi, R Nakatsu. Emotion Recognition In Speech Using Neural Networks [J]. **Neural Computing and Applications**, 2000, (9): 290-296.
  - [32] K H Kim, S W Bang, S R Kim. Emotion Recognition System Using Short-term Monitoring of Physiological Signals [J]. **Medical& Biological Engineering&Computing**, 2004, 42: 419-427.
  - [33] 王青. 基于神经网络的汉语语音情感识别的研究[D]. 浙江: 浙江大学论文集, 2004-03.
  - [34] Björn Schuller, Gerhard Rigoll, Manfred Lang. Hidden Markov Model Based Speech Emotion Recognition [A]. **Acoustics, Speech, And Signal Processing, 2003(ICASSP'03). 2003 IEEE International Conference On** [C]. 2003-04, 2: 1-4.
  - [35] 赵力, 钱向民, 邹采荣, 等. 语音信号中情感识别研究[J]. 软件学报, 2001, 12(7): 1051-1054.
  - [36] Tin Lay New, Say Wei Foo, Liyanage C De Silva. Speech Emotion Recognition Using Hidden Markov Models [J]. **Speech Communication**, 2003, 41: 603-623.
  - [37] Chul Minlee, Narayanan S S, Pieraccini R. Classifying Emotions In Human-Machine Spoken Dialogs [A]. **Multimedia and Expo, 2002. Proceeding. 2002 IEEE International Conference On** [C]. 2002, 1: 737-740.
  - [38] 陈建厦. 语音情感识别综述[A]. 第一届中国情感计算与智能交互学术会议[C]. 北京, 2003.
  - [39] 高慧, 苏广川. 情绪化语音特征分析与识别的研究进展[J]. 航天医学与医学工程, 2004, 17(5): 77-80.

**作者简介:** 余伶俐 (1983-), 女, 博士生, 主要从事语音情感识别, 情感计算与情感建模的研究; 蔡自兴 (1938-), 男, 博导, 教授, 联合国专家, 纽约科学院院士, 从事人工智能和智能控制等的研究; 陈明义 (1964-), 男, 博士, 教授, 从事领域: 信号处理, 语音编码, 语音质量客观评估, 语音识别。



## Study on emotion feature analysis and recognition in speech signal: an overview

YU Ling-li, CAI Zi-xing, CHEN Ming-yi

(College of Information Science & Engineering, Central South University, Changsha 410083 China)

**Abstract:** Emotion recognition of speech is becoming one of new research projects. This paper introduces the status of emotion recognition in speech signal at home and abroad recently. And it also exposes various human emotion classifications. The parameters extraction methods and the meaning of different feature parameters are induced. Based on these, the research progress and the advantages and disadvantages of some capital constitutive modeling methods are summarized. At last, the future challenges and development trends of emotion recognition of speech are presented.

**Key words:** emotion speech; emotion classification; feature parameter; recognition methods

---

(续第 92 页) (from page 92)

## Advancement on IP watermarking techniques

LI Dong-xiao, ZHENG Wei, ZHANG Ming

(Dept. of Information Science & Electronic Engineering, Zhejiang Univ., Hangzhou 310027, China)

**Abstract:** Recently digital watermarking has emerged as a candidate solution for reusable IP core protection. The key idea of IP watermarking is try to use IP cores as cover media, and hide some digital information (watermarks) in their circuit designs, for intellectual property authentication and infringements tracing. In this paper, we present the basic research framework and survey the recent advancement on IP watermarking techniques. Several typical IP watermarking schemes are evaluated and compared. The technical difficulties are discussed, and future directions are outlined.

**Key words:** intellectual property; IP protection; digital watermarking

---

(续第 104 页) (from page 104)

## Design of adiabatic CMOS and traditional CMOS interface circuit

YU Jun-jun, WANG Peng-jun

(Institute of Circuits and Systems, Ningbo University, Ningbo 315211, China)

**Abstract:** In order to embed adiabatic CMOS circuits into traditional circuit systems to replace the parts that dissipate much energy, the study is conducted on the design of the interface circuits between adiabatic CMOS circuits and traditional CMOS circuits: Traditional CMOS to Adiabatic CMOS (TC/AC) interface circuit and Adiabatic CMOS to Traditional CMOS (AC/TC) interface circuit. In this way, traditional CMOS circuits can drive adiabatic CMOS circuits by using the TC/AC interface circuit, adiabatic CMOS circuits can drive traditional CMOS circuits by using the AC/TC interface circuit, and hence the energy consumption of a whole circuit system can be reduced by using adiabatic CMOS circuits which have the character of low energy consumption to strengthen the practicality of adiabatic CMOS circuits. Finally, the TC/AC interface circuit and the AC/TC interface circuit are verified by computer simulations.

**Key words:** TC/AC interface circuit; AC/TC interface circuit; low power; circuit design