语音的情感信息分析与编辑*

蔡莲红 崔丹丹 蒋丹宁 杨鸿武

清华大学计算机科学与技术系，北京 100084（clh-dcs@tsinghua.edu.cn）

摘 要：本文研究了语音声学参数的情感区分特征，并通过情感分类的方法确定声学特征对分类的贡献。设计实现了一个语音情感编辑器，它具有编辑、修改语音韵律参数的功能，以实现不同情感的表现。

关键词：情感；声学特征；情感编辑

1 前言

人们通过语音信号传递各种信息，包含“表事”，“表意”，“表情”等信息。话音反映说话人的意向和情感状态。近年来，情感语音逐渐成为的语音研究热点。

研究表明，语音的情感信息体现在多种声学参数的变化中，文献[1]将其归纳为基频、时长、能量和频谱四个方面。在基本情感类别中，愤怒和高兴均表现为基频均值、变化范围和方差的提高，能量的加强，以及频谱中高频成分的增加。相反，悲伤对应于基频均值和变化范围的降低，能量的减弱，语速的减慢，以及频谱中高频成分的减小。害怕的特征除了基频均值、变化范围和频谱中高频成分的增加外，还包括基频曲线上抖动的加强和语速的加快。惊讶则表现为很宽的基频变化范围，以及稍减慢的语速。此外，声学参数随时间的变化情况也负载了一定的情感信息。

本文重点研究了韵律特征对情感区分和情感表现的影响。首先建立了情感语料库，统计了语料库中语音的声学参数，并选用不同的分类器、不同的声学特征进行情感分类。分类结果表明韵律特征在情感分类中扮演重要的角色。为了研究和感知韵律特征与情感表现的关系，设计实现了一个语音情感编辑器，它具有编辑、修改语音韵律参数的功能，通过韵律修改表现不同的情感。

2 情感语音的区分特征

我们知道，与情感表现有关的声学参数包括基频、时长、能量和频谱参数。我们首先在句子范围内计算声学参数统计值，如平均值、标准差、最大值、变化范围等，以反映参数的全局特性。其次计算声学参数的时序特征，它是短时特征的序列，反映了参数随时间的变化情况。目前对声学特征的情感区分性的研究较少。实际上，不同的声学特征反映情感的不同侧面，从而在情感分类中具有互补性和区分性，因此研究情感特征的区分是非常必要的。

我们考虑了六种基本情感，为每类情感设计了 200 个语句。在每类情感的文本中，包含了不同的句子类型（陈述句和疑问句），句子长度，以及声调和重音分布等情况。语句的平均长度为 7 个到 13 个音节，最短的语句包含 2 个音节。语料的发音人为一名不带口音的女性发音人。

2.1 基本参数的统计特征

表 1 列出了情感语料中各种声学参数的统计平均值。可见，与中性语句的统计结果相比，愤怒、高兴、惊讶三类情感的基频明显升高，基频变化率提高，语速加快，能量增强，频谱中高频成分增加，频谱变化剧烈。害怕表现为基频升高，语速明显加快，以及语音信号中非周期成分明显增加。悲伤表现为基频和基频变化率下降，语速减慢，能量减弱，频谱中高频能量减少，以及频谱变化缓慢。这些统计结果与他人所总结的情感声学特征是基本一致的，说明论文所录制语料的情感表现是合理的。稍微有所区别的是，录制的害怕语音仅表现为基频的提高和语速的加快，而没有出现所述的能量增强和高频成分增加。这可能是由于论文所录制的害怕并不是极端的恐惧，在激发度上相对愤怒、高兴、惊讶三类情感较低。

* 国家自然科学重点基金项目资助（60433030，60418102）

表 1 汉语情感语料中声学参数的统计平均值

	愤怒	害怕	高兴	惊讶	悲伤	中性
基频（Hz）	385	330	396	435	256	288
基频变化率（Hz/ms）	0.82	0.54	0.78	0.89	0.24	0.53
时长（ms）	177	157	209	210	247	221
能量（dB）	68	57	65	69	50	57
频谱质心（Hz）	3024	2651	2777	2791	2479	2664
频谱变迁	0.87	0.59	0.77	0.76	0.29	0.58
频带周期性	0.59	0.57	0.68	0.73	0.62	0.62

2.2. 情感分类

本文通过分类实验研究了情感语音的区分特征。探讨了韵律参数、能量参数、频谱参数的统计特征和时序特征在情感分类中的作用。在提取出基本的声学参数之后，分别针对统计特征和时序特征进行了分类实验，并通过混淆矩阵度量声学特征在每两类情感之间的区分能力。实验表明，大部分声学参数的统计特征和频谱参数的时序特征能够较好地区分激发度不同的情感，而韵律参数的时序特征能够较好地区分激发度相近但评价性不同的情感。我们还研究了融合统计特征和时序特征的情感分类方法。该分类方法可提高情感分类的正确率，降低了情感间的混淆度。

分类数据是如上所述的汉语情感语料，包括六个情感类别：愤怒，害怕，高兴，悲伤，惊讶，中性。每类情感数据包含约 200 句语句，为了减小随机因素的影响，提高分类结果的稳定性，在分类实验中采用了交叉检验技术。所有语句被平均分为 5 份，而分类实验也相应地进行 5 次，每次分别将其中的 1 份数据作为测试集，其余的 4 份作为训练集。取 5 次实验的平均值作为最终的分类结果。

表.2 给出了分别采用 MLP（多层感知器）、PNN（概率神经网络）、SVM（支持向量机）作为分类模型时，韵律统计特征、能量统计特征、频谱统计特征的平均分类正确率。由表 2 可见，在单独采用一组参数的统计特征时，频谱统计特征和韵律统计特征的平均分类正确率较高，而能量统计特征的分类正确率较低。同时，MLP、PNN、SVM 三种分类模型的分类性能也有所差别。对于各组参数的统计特征，MLP 和 SVM 的性能优于 PNN，这可能是由于 PNN 通过欧式距离计算测试样本与训练集中各样本之间的距离，因此各维特征被等同对待，不能通过调整权值反映各维之间的相对重要程度；另外，PNN 网络中各个高斯核函数的宽度参数均设为相等的值，也会对分类性能产生影响。

表 2 统计特征的平均分类正确率(%)

	MLP	PNN	SVM
韵律统计特征	84.2	83.3	86.2
能量统计特征	74.3	68.9	76.7
频谱统计特征	88.5	85.1	89.6

3 语音的情感编辑

3.1. 情感编辑器

为了研究和感知韵律特征与情感表现的关系，设计实现了一个语音情感编辑器。它具有编辑、修改语音声学参数的

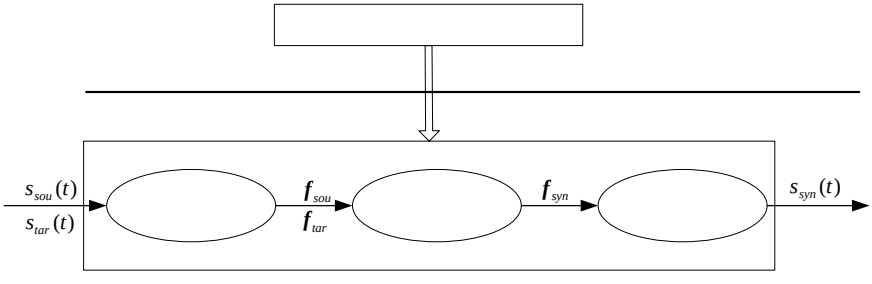

图 1 情感编辑器的系统框图

功能，通过修改韵律来表现不同的情感。图 1 显示了情感编辑器的系统框图，其核心部分包括声学特征分析、声学特征修改、以及语音重建三个模块。在输入原始语句 $s_{sou}(t)$ 以及与其文本相同、但情感表现不同的目标语句 $s_{tar}(t)$ 后，首先对它们进行声学分析，提取与情感相关的声学特征。随后根据目标语句的声学特征 $\boldsymbol{f}_{tar}$，以及用户编辑的情况，对原始语句的特征 $\boldsymbol{f}_{sou}$ 进行修改。最后，语音重建模块修改原始语句的声学信号，产生具有 $\boldsymbol{f}_{syn}$ 特征的语句。

3.2 声学特征分析及修改

声学特征分析模块提取的声学参数包括韵律参数和频谱参数。其中，基频和时长信息通过一个语音编辑和处理平台 VisualSpeech 标注，并保存为 Tag 文件传给情感编辑器。Tag 文件准确记录了每个音节的起、止点，以及每个基音周期中最大峰值（对应于声带闭合点）的位置。情感分析器根据 Tag 文件中的信息，恢复出完整的基频参数曲线，以及音节的时长参数。此外，Tag 文件中记录的基音周期最大峰值位置是语音重建算法所需要的信息。

提取的频谱参数包括共振峰参数和 H1-A3（基频分量与第三共振峰范围内最强的谐波分量之间的强度比）参数。其中，共振峰描述了声道作为一个共振腔的谐振特性，在所有表示声道函数的参数中具有最明显的物理意义。它的提取方法是对 12 阶 LPC 多项式求根，根据各元音的共振峰范围，从中选择出表示共振峰的根，并推导出相应的共振峰频率和带宽参数。H1-A3 参数是语音谱中基频所对应的频率分量与第三共振峰频率范围内最强的谐波分量之间的强度比。H1-A3 参数反映了频谱中高频成分的相对强弱，与音色的明亮程度相关。H1-A3 参数的数值越小，则说明语音频谱中的高频成分越强，音色越明亮。在提取出基本的声学参数之后，同时计算出它们的统计特征，包括平均值、最大值、最小值、以及在句中随时间变化的斜率。

声学特征的修改有两种方式。第一种方式是复制目标语音的特征 f_{tar}。由于原始语音和目标语音的长度不同，因此首先需要进行时间规整。在情感编辑器中，时间规整是以音节为单位进行的。对于基频和能量参数，在每个音节的浊音段范围内，简单地根据时间比例，找到原始参数曲线和目标参数曲线之间的对应关系。对于频谱参数，则采用了更为复杂的动态时间规整（DTW）算法。第二种修改方式是由用户通过拖动鼠标的方式直接编辑声学参数曲线。当采用这种方式修改声学特征时，也可以不向情感编辑器中输入目标语音。图 2 是情感编辑器的用户界面。用户可以在这个界面上手动修改时长、幅度，以及基频的平均值、音域和各时刻的频率数值。在修改声学特征时，除直接对参数曲线进行修改外，也可以只修改参数的某种统计特征。例如，可以在保持基频参数曲线变化形状不变的前提下，仅提高或降低基频的平均值。

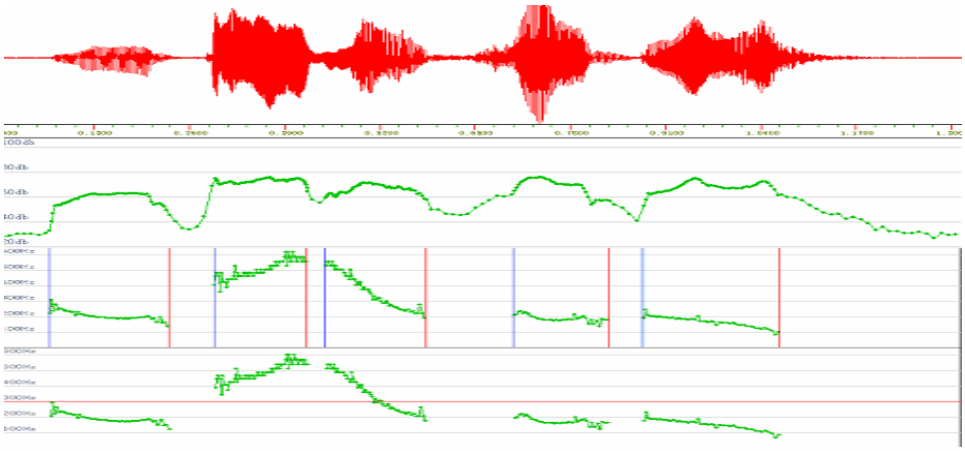

图 2 情感编辑器的功能示意

本文选用汉语 TTS 的输出和一部分朗读语句，通过编辑器修改成愤怒、害怕、高兴、悲伤、惊讶等情感。为了方便情感表达，各类情感语音的文本不必相同。但它们均包含了不同的句子类型（陈述句和疑问句）、句子长度，以及声调和重音分布等情况。所有的情感语句均由一名女性发音人在安静环境下录音得到。图 3 显示了一个通过直接编辑方式修改基频曲线的例子，将末音节“对”声调上升的斜率调高，以研究末音节声调曲线对情感表现的影响。

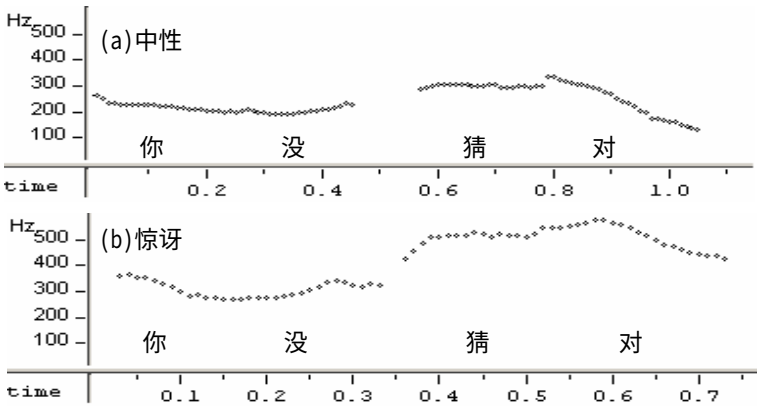

图 3 中性语句和惊讶语句的基频曲线

4 结束语

本文研究了韵律特征对情感区分的影响。建立了情感语料库，统计了语料库中语音的声学参数，最后选用不同的分类器、不同的声学特征进行情感分类。分类结果表明韵律特征在情感分类中扮演重要的角色。为了研究和感知韵律特征与情感表现的关系，设计实现了一个语音情感编辑器。通过修改韵律来表现不同的情感。语音的情感表现是语音参数的全面体现，除了韵律参数外，我们将进一步研究与情感信息相关的其他参数。

参 考 文 献

[1] Cowie R., Cowie E.D., Tsapatsoulis N., etc, “Emotion Recognition in Human-Computer Interaction”[J], *IEEE Signal Processing Magazine*, 2001, 18(1): 32-80.

[2] Paeschke A., Sendlmeier W.F., “Prosodic Characteristics of Emotional Speech: Measurements of Fundamental Frequency Movements”[A], *Proc. of ISCA Workshop on Speech and Emotion*[C], 2000, 75-80.

[3] 赵力，蒋春晖，邹采荣等，“语音信号中的情感特征分析和识别的研究”[J]，*电子学报*，2004，32(4)：606-609。

[4] Cheveign A.D., Kawahara H., “YIN, a Fundamental Frequency Estimator for Speech and Music” [J], *J. Acoust. Soc. Am*. 2002, 111(4): 1917-1930.

[5] 黄德智，张晓洲，蔡莲红，“一种数字语音处理研究平台的设计”，已被*数据采集与处理*录用。

[6] 蒋丹宁，蔡莲红，基于语音声学特征的情感信息识别，已被*清华大学学报*录用

[7] 蒋丹宁，博士论文，情感语音的声学特征分析及建模，2005，清华大学

[8] Kittler J., Hatef M., Duin R. P., etc, “On Combing Classifiers” [J] *IEEE Transactions on Pattern Analysis and Machine Learning*, 1998, 20(3): 226-239.