

文章编号: 1001-0920(2005)11-1311-03

多语种下情感语音基频参数变化的统计分析

田 岚^{1,2}, 姜晓庆¹, 侯正信²

(1. 山东大学 信息科学与工程学院, 济南 250100; 2. 天津大学 电子信息工程学院, 天津 300072)

摘要: 为研究韵律特征中最能反映语音情感信息的基频参数变化, 选取7种典型的情感状态、固定句式, 对同一说话人的汉语、英语、日语等多语种语音样本进行基频平均值、基频动态范围、基频抖动等参数的统计分析。统计结果表明, 情感语音的基频结构特征随情感状态改变有明显的变化, 且不同语种下这种结构的变化有较好的一致性。

关键词: 情感语音; 基频; 统计分析

中图分类号: TN 912.34

文献标识码: A

Statistical Study on the Diversity of Pitch Parameters in Multilingual Speech

TIAN Lan^{1,2}, JIANG Xiao-qing¹, HOU Zheng-xin²

(1. School of Information Science and Engineering, Shandong University, Jinan 250100, China; 2. School of Electronic and Information Engineering, Tianjin University, Tianjin 300072, China. Correspondent: TIAN Lan, E-mail: tianlan65@sdu.edu.cn)

Abstract: Pitch averages, pitch ranges and pitch jitters are calculated and statistical analysis of these parameters is carried out to evaluate the differences of the seven multi-typical emotion states. The test sentences are spoken in Chinese, English and Japanese by the same speaker and have the same pattern in syntax. The statistical analysis results show that the pitch structure has distinct diversity with the changes between different emotions and language factor does not effect pitch variation features of some given emotion in multilingual speech obviously.

Key words: Emotion speech; Pitch; Statistical analysis

1 引言

情感信息是人与人之间相互交流的重要信息。目前对于高质量的合成语音, 普遍采用基于大型数据库搜索拼接的方法, 若能加入情感信息, 并依据这一信息调整合成语音韵律和语句调式, 将大大提高合成语音的质量。情感语音处理技术应用广泛, 如情感语音通信模型、专为残疾人和无法打字的老年人设计的情感网络通信系统、虚拟视频通信系统、交互式家庭影院系统、自动对话系统、信息安全系统等^[1-4]。因此, 情感语音处理是将语音技术推向更广泛人机交互应用的一个突破点。

目前, 情感信息处理的研究还处于起步阶段。20

世纪90年代以来, 美国、日本、英国及其他国家先后有科研机构涉足语音情感的研究。近几年国内也开始从事相关的研究, 例如汉语语音情感特征的提取和识别^[5]。随着社会世界化、国际化的发展, 混合语种的情感语音处理已成为一种需要。研究指出: 语言因素不会影响情感语音的声学结构相关性^[6]。人类情感信息在语音上主要体现在韵律的变化上, 在韵律特征中, 基频结构是语音信号中能够提供说话者最明确情感信息的特征。本文通过实验, 比较同一说话人在不同情感状态下, 固定句式的汉、英、日语语音样本的基频曲线、基频平均值、基频范围、基频平均抖动值特征, 分析语种差异对不同情感下基频参

收稿日期: 2004-10-27; 修回日期: 2005-03-16

基金项目: 山东省自然科学基金项目(Y2002G13)。

作者简介: 田岚(1965—), 女, 济南人, 教授, 博士生, 从事模式识别、智能信息处理的研究; 姜晓庆(1981—), 女, 山东烟台人, 硕士生, 从事自然语言理解、语音信号处理的研究。

数内在特性的影响

2 人类情感的声学特征

与情感相关的人类生物特征信息是多种多样的。当人处于愤怒状态时,心跳加快,皮肤电压升高,血压升高,胸腔的回声和呼吸声在语音信号中的比重增加,语调提高,振幅变大,语速变快等。表现在语音上,情感影响的主要声学参数有:基频(范围、平均值、包络等)、调和性(呼吸频率、抖动等)、响度、时长(词长、元/辅音长、爆破脉冲密度)、谱信息(共振峰位置、带宽等)。

人类的情感多种多样,要准确区分或定位人类语音中的各种情感较为困难。Shaver等将人类的情感描述为一个模糊集,并对人类的135种情感通过聚类得到等级树上最顶端的6个节点:热爱、高兴、生气、惊讶、悲伤和恐惧。本文就热爱、高兴、生气、惊讶、悲伤、恐惧和平静这7种典型情感状态下固定句式语音样本的基频结构进行分析比较。

3 语音数据库

情感语音库的建立必须搜集足够数量且能比较准确地定位各种情感的语音样本。为避免因不同话者基频范围不同而引起的基频结构变化趋势相互抵消,本实验采用固定话者语音样本为统计对象。在样本采集过程中,说话人必须能对指定情感种类进行熟练、准确、到位的表达,实验中在适当控制环境下模拟各种情感进行录音。录音时,所选用的文本是日常生活中频繁使用,并在各种情感状态下不会产生歧义的短句,例如,“这是他的X”(汉语)、“This is his X”(英语)、“これはかれのXです”(日语),其中“X”为替换词。

对同一句式,选用3种语言分别在7种情感状态各采集50个样本。样本都转化为单声道、8 KHz采样、8比特的文件格式。对于同一文本,每种语言每个情感状态下的样本进行主观听力测验,选取每一类情感表达准确值最高的25个样本,然后对同一句式下3种语言的525个样本进行基频参数统计分析。

4 特征提取及分析^[7]

选取基频作为基本特征参数,并在此基础上计算基频曲线、基频范围、基频的抖动。基频参数提取采用改进的倒谱算法。

首先进行语音样本的预处理,根据语音的短时能量和过零率乘积判断成句语音样本的起终点,利用语音信号的倒频谱特征,检测出表征声门激励周期的基音信息。从理论上说,倒谱法可使基音信息和声道信息相对分离,再通过倒滤波方法检出激励源的信息。实际上,对于一帧浊音语音的倒谱,其倒谱域的基音信息和声道信息并不是完全分离的,尤其

在过渡音或背景噪声较大的情况下,周期激励信号能量相对变低,声道信息对基音的倒谱峰影响明显。为此作如下改进:首先对语音信号做LPC分析;然后对所得的不包含声道信息的预测残差信号进行倒谱分析,在频域取对数,去除600 Hz以上高频干扰成分,并在倒谱域以加权倒谱统计中值为清/浊检测阈值,计算出更准确的基频;最后对计算结果进行5点中值平滑处理,根据基频值可计算出基频范围和基频的抖动值。基频抖动由下式计算:

$$p_i^1 = p_i^0 - p_{i-1}^0, \quad i = 2, 3, \dots, N.$$

其中: p_i^1 为第*i*帧的基频抖动值, p_i^0 为第*i*帧的基音频率, N 为计算样本的帧数。

5 实验结果和分析

图1给出汉语样本“这是他的车”在惊讶、悲伤、愤怒、恐惧、高兴、热爱和平静状态下的基频曲线。可见,由于不同情感下的发音速度不同,语音样本长度不一,基频最大、最小值的分布差别较大,基频曲线呈现出一定的层次性。在不同情感下与平静状态下的基本调式相比,基频曲线在结构和变化趋势上也有很大差异。针对英语和日语样本做了同样分析,发现不同语种下同一种情感整句语调的变化存在相似性,且3语种基频曲线的最大、最小值分布相似。

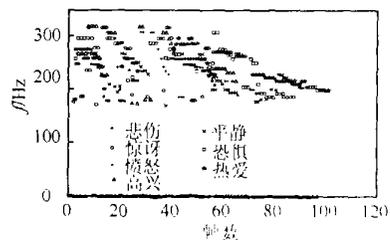


图1 不同情感状态下汉语样本的基频曲线的比较

进一步统计比较3种语言在7种情感状态下的基频平均值、基频平均范围、基频平均抖动值,并给出基频抖动图。为消除不同语种间语言学差异的影响,对计算结果做均值为0,方差为1的归一化处理。图2~图4分别给出了3种语言在不同情感状态下的基频平均值、基频平均范围和基频平均抖动值的比较。从统计结果可以看出,随着情感的变化,基频的结构特性也有显著的变化,但不同语种基频的变化趋势是相同的。3种语言的样本在惊讶、恐惧、高兴、愤怒的基频平均值都远远高于平静和悲伤状态下的平均值,而平静、悲伤和热爱状态的基频值则较为相似。同一情感状态下,3种语种的基频平均值相近。图3和图4中惊讶、高兴、愤怒、恐惧的基频取值动态范围较大,基频抖动跳跃性明显,对应于人在这些状态下多使用加速句和加强句,基频在重音处易发生突变,于是这些状态的基频波动远远大于其他状态。同

样, 悲伤、平静、热爱状态下基频动态取值范围较小, 基频抖动不大, 对应于人在这些情感状态下时语调降低、语句平坦的特点

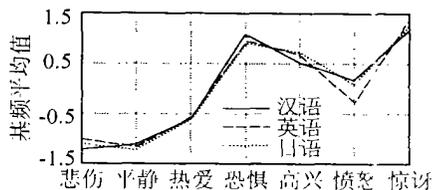


图2 3种语言在不同情感下基频平均值的比较

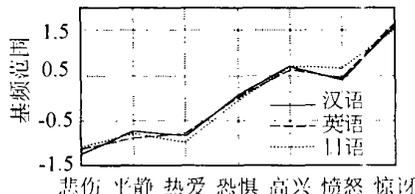


图3 3种语言不同情感状态下基频平均范围比较

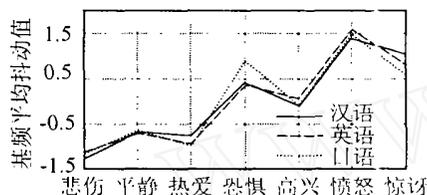


图4 不同语种归一化基频平均抖动值的比较

通过对基频参数的统计分析发现, 尽管基频参数在具体取值上存在着差异, 但3个语种在同一状态下的基频参数随情感转变的变化模式是相同的, 不受语种因素的影响

依据基频参数使用PCA方法对汉、英、日混合语音样本5种基本情感的识别结果如表1所示

表1 PCA方法对多语种语音情感识别结果

	悲伤	平静	高兴	愤怒	惊讶
悲伤	89.33	13.33	0	1.33	0
平静	10.67	68	2.67	2.67	1.33
高兴	0	8	61.33	8	10.67
愤怒	0	6.67	17.33	69.33	13.33
惊讶	0	4	8	18.67	73.33

6 结 论^[8,9]

本文对同一话者在固定内容时汉、英、日语样本进行的一系列分析和研究, 发现尽管不同语言对同一说话内容在语音样本细节上存在着语种本身的差异, 但从其韵律参数特性的相似性比较结果看, 要表达同一情感, 语种对该情感下的基频变化结构没有很大的影响。不考虑语种因素, 这几种典型情感在基频参数特性上都有自身的特征。如悲伤时, 基频平均值、基频范围都是最低的, 平静状态下虽然基频平均抖动值与悲伤时极为相近, 但是基频范围和基频平均值却比悲伤时高; 恐惧和高兴状态的基频平均

值和基频平均抖动值接近, 但是恐惧的基频抖动却比高兴时明显要大。通过这几个量的对照, 可以粗略地进行多语种语音情感的识别。同时, 这一结论可在多语种的情感语音合成系统、多语种的语音识别和说话人识别系统中加以应用, 对智能人机交互的实现也有重要的指导意义

参考文献 (References)

- [1] Tsuyosli Moriyama, Shinji Ozawa. Emotion Recognition and Synthesis System on Speech[A]. *Proc of the 6th Int Conf on Multimedia Computing and Systems*[C]. Florence: IEEE Press, 1999: 840-844
- [2] Akemi Iida, Nick Campbell, Michiaki Yasuura. Emotional Speech as an Effective Interface for People with Special Needs[A]. *Proc of the 3rd Asia Pacific Conf on Computer Human Interaction*[C]. Japan, 1998: 266-271.
- [3] Pierre Yves, Oudeyer. The Production and Recognition of Emotions in Speech: Features and Algorithms[J]. *Int J of Human Computer Studies*, 2003, 59(1-2): 157-183
- [4] Fabio Lavagetto. Emotional Representation and Animation of 3D Facial Models: The Interface Approach [A]. *Proc of 2001 Int Conf on Image Processing* [C]. Thessaloniki: IEEE Press, 2001: 594-597.
- [5] 赵力, 钱向民, 邹采荣, 等. 语音信号中的情感识别研究[J]. *软件学报*, 2001, 12(7): 1050-1055
(Zhao L, Qian X M, Zhou C R, et al. A Study on Emotion Recognition in Speech Signal [J]. *J of Software*, 2001, 12(7): 1050-1055)
- [6] Razak A A, Abidin M I Z, Komiya R. Emotion Pitch Variation Analysis in Malay and English Voice Samples [A]. *Proc of the 9th Asia-pacific Conf on Communications*[C]. IEEE, 2003: 108-112
- [7] 赵力. *语音信号处理*[M]. 北京: 机械工业出版社, 2003: 259-270
(Zhao L. *Speech Signals Processing* [M]. Beijing: China Machine Press, 2003: 259-270)
- [8] 田岚, 陆小珊. 一种数据驱动的汉语成句语音韵律特征产生模型的研究[J]. *控制与决策*, 2003, 18(6): 656-660
(Tian L, Lu X S. Study of Data-driven Hierarchical Prosody Generation Model for Chinese Sentence Utterance [J]. *Control and Decision*, 2003, 18(6): 656-660)
- [9] 张家喙. 论语音技术的发展[J]. *声学学报*, 2004, 29(3): 193-199
(Zhang J L. On the Development of Speech Technology [J]. *Acta Acustica*, 2004, 29(3): 193-199)