实用语音情感数据库的设计与研究

黄程韦[1]，金 赟[1,2]，赵 艳[1]，于寅骅[1]，赵 力[1]
(1. 东南大学水声信号处理教育部重点实验室，南京 210096；2. 徐州师范大学物理与电子工程学院，徐州 221116)

摘要：针对实用语音情感研究的需要，设计并且建立了诱发语音情感数据库。根据刺激评估表，设计了一种基于计算机游戏的情感诱发方式，首次针对汉语音的实用语音情感进行了情感数据的采集和情感语音库的建立。对烦躁、紧张和喜悦等三种具有实际应用意义的语音情感进行了有效的诱发，录制了情感语音数据，并进行了数据的检验与听辩测试。结果显示，采用计算机游戏诱发情感的方法能够有效地采集自然度较高的实用语音情感数据，为实用语音情感的特征分析与识别提供了必要的研究基础。

关键词：实用语音情感；刺激评估表；诱发语音；情感数据库
中图分类号：TP392 **文献标识码**：A **文章编号**：1000-3630(2010)-04-0396-04
DOI 编码：10.3969/j.issn1000-3630.2010.04.010

Design and establishment of practical speech emotion database

HUANG Cheng-wei[1], JIN Yun[1,2], ZHAO Yan[1], YU Yin-hua[1], ZHAO Li[1]
(1. *Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education*, *Southeast University, Nanjing* 210096, *China*;
2. *School of Physics and Electronics Engineering*, *Xuzhou Normal University*, *Xuzhou* 221116, *China*)

Abstract: An elicited speech emotion database is designed and established for practical speech emotion research. Based on Scherer's stimulus evaluation checks, the experiments of inducing emotional speech with a computer game are designed. A Chinese practical speech emotion database is established for the first time. Three practical speech emotions including impatience, nervousness and happiness are induced. And four categories of emotional data, including the neutral state, are collected, and emotional data are evaluated in a listening perception test. The results show that the method of emotional speech elicitation with a computer game is a reasonable and effective way of collecting high quality practical speech emotion data, which is a necessary support to feature analysis and recognition of practical speech emotion.

Key words: practical speech emotion; stimulus evaluation checks; elicited speech; emotion database

1 引 言

语音是人类交流信息最自然、最有效、最便捷的手段之一[1-3]。语音情感识别是情感计算的一个重要部分，近年来受到了计算机科学、心理学、认知科学与行为科学等领域的研究者们的越来越高的关注[4-8]。情感状态的识别与在此基础上的心理评估具有很高的实际应用价值[9,10]，特别是在载人航天等军事领域中，长时间的单调的高强度任务，会使得相关人员面临严酷的生理以及心理考验，引发某些负面的情绪。探讨这些情绪对工作能力的作用及其机制和影响因素，具有非常重要的应用价值，可以研究提高个体认知和工作效率的方法、避免影响认知和工作能力的因素。然而以往的语音情感识别，集中在对几种基本情感的研究上，实验手段上往往采取表演的方式来模仿实际环境中的真实情感。通过对基本的几类语音情感的分类研究，虽然能在理论上验证各种识别算法的优劣性能，能用于寻找对识别基本情感类别有效的声学特征，但仅停留在对基本情感类别的研究上，远远不能满足实际应用中的要求。本文针对在长期载人航天环境以及其它类似的高强度特殊作业环境中面临的实际问题，选择了具有实际应用价值的语音情感，采集了“烦躁”、“紧张”、“喜悦”和“平静”四种情感状态下的语音情感数据，建立了一个中文的实用语音情感数据库。

2 语音情感数据库建立过程

参考国内外著名语料库及其相关的规范[11-14]，设计了实用语音情感数据库建立的流程，如图 1 所示。

2.1 发音人的选择和语料的设计

选择 10 名大学生参加计算机游戏诱发的语音采集，5 名男性，5 名女性，年龄在 20 岁到 30 岁之

收稿日期：2009-08-04；**修回日期**：2009-11-16
基金项目：国家自然科学基金(60472058, 60975017)；江苏省自然科学基金(BK2008291)
作者简介：黄程韦(1984-)，男，江苏常熟人，博士，研究方向为语音情感识别。
通讯作者：黄程韦，E-mail: air1894@yahoo.com.cn

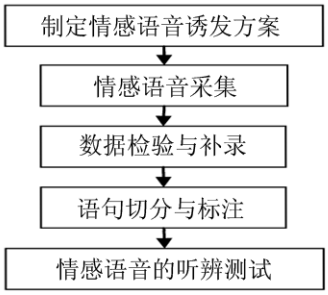

图 1　实用语音情感数据库的制作过程

Fig.1　Establishment process of practical speech emotion database

间，发音人具有良好的健康状况，近期无感冒，无喉部疾病，并且听力正常。发音人的选择主要考虑发音人的性别、年龄、生活背景、教育程度、职业、病理情况、听力状况、口音等。

文本材料的设计，考虑到实用语音情感识别的一个主要应用领域为长期的航空、航天和航海任务所引发的负面情绪的评估，无情感倾向性的工作用语单词和短句选自国际海事组织(IMO)发布的《标准航海通信用语》(SMCP)。工作用语中的单词和短句各举一例如下。

“螺旋桨”(单词)；“保持你船左后弦受浪”(短句)。

对于单词和短句的选择，主要考虑以下两方面：(1) 所选择的单词和语句必须不包含某一方面的情感倾向；(2) 必须具有较高的情感自由度，对同一个单词或语句能施加各种感情进行分析比较。

2.2　软硬件设备环境与数据检验

录音软件采用 Cool Edit pro 2.0，录音时采用单声道、11.025kHz 采样频率、16 位采样精度，录制语音保存为 PCM 编码的 wav 格式。录音软硬件包括：高性能计算机 1 台、M-audio mobilePre USB 声卡 1 台、M-audio Nova 大振膜电容话筒 1 支、监听耳机一副等。录音过程选在安静的实验室内进行。

每次录音后，应进行数据的检验与补录，及时对语音文件进行人工检验，以排除录音过程中可能出现的错误。

2.3　数据库的命名规则及其标注

为了便于语音库的管理，制定了语音文件的命名规则。一个语音文件的名称由三个部分组成。命名规则如表 1 所示。例如，“S2F1HAP1.WAV”表示 1 号女性发音人对短句 2 进行的喜悦情感的发音，采集次数为第 1 次。每个语音文件都对应一个标注文件，用于记录情感语音文件的信息，包括录音的时间、地点、发音人、目标情感、诱发方式等信息。

表 1　语音库文件命名规则

Table 1　The naming convention for the speech database

文本代号	文本编号	发音人性别	发音人编号	情感类别	采集次数
W(单词)	1~20	M(男性)	1~5	IMP(烦躁)	1~3
				NER(紧张)	1~3
S(短句)	1~20	F(女性)	1~5	HAP(喜悦)	1~3
				NEU(平静)	1~3

3　情感语音的诱发

3.1　情感语音的诱发过程

根据 Scherer 的观点[15-18]，人类声音中蕴含的情感信息，受到无意识的心理状态变化的影响，以及社会文化习惯导致的有意识的控制。对自然语音情感识别不适合采用表演数据，而需要通过诱发的方式采集自然度较高的数据。

本文采用计算机游戏进行情感诱发，通过游戏中画面和音乐的视觉、听觉刺激，提供一个互动的、具有较强感染力的人机交互环境，能够有效诱发出被试的正面与负面情感。特别是在游戏胜利时，被试由于在游戏虚拟场景中的成功与满足，被诱发出喜悦等正面情感；在游戏失败时，被试在虚拟场景中受到挫折，容易引发烦躁等负面情感；在游戏过程中，一些具有挑战性的游戏情节往往能引发被试的紧张情绪。

在游戏场景的设置方面，为了便于烦躁、紧张、喜悦情感的诱发，选用了既需要耐心又具有一定挑战的计算机小游戏。游戏中要求被试用鼠标移动一个小球通过复杂的管道，在通过管道的过程中如果小球碰到管壁，则游戏失败；在规定时间内顺利通过管道后，游戏胜利；游戏共有 5 个难度等级，以适合不同水平的被试。在游戏前，让被试者平静地读出指定的文本内容，录制平静状态的语音。在每次游戏胜利后，要求被试用喜悦的语气说出指定的文本内容，录制喜悦状态的语音。在每次游戏失败后，要求被试用烦躁的语气说出指定的文本内容，录制烦躁状态的语音。在游戏进行到一半时，暂停游戏，要求被试用紧张的语气说出指定的文本语句内容，录制紧张状态的语音。为了便于对数据进行检验，在每次录制情感语音后，让被试填写情绪的主观体验，包括五个选择：烦躁、紧张、喜悦、平静、其它，在实验结束后，根据被试的情绪主观体验表，剔除主观体验与诱发目标情感不一致的语音数据，必要时进行适当的补录。

3.2　刺激评估表：诱发实验的设计依据

Scherer 等情绪的刺激论者认为，人类根据不同的准则对感知觉获得的信息进行衡量，这些衡量的过程是我们的认知系统所固有的，但通常都是在潜

意识中进行。Scherer 将这些衡量机制称为“刺激评估表”(Stimulus Evaluation Checks，SECs)[16-19]。Scherer 分离出了五类不同的刺激评估项，及若干子项，如表 2 所示。

对喜悦、愤怒、惊讶和恐惧等人类的基本情感，它们与刺激评估表(SECs)之间的关系如下：喜悦是本征愉悦度取正值的结果；愤怒是相关度高、诱因低的结果，并且应对潜能项显示有足够的资源移除或忽略眼前的障碍；惊讶是新异度取正值的结果；恐惧是相关度高、诱因低的结果，并且应对潜能项显示没有足够的资源移除或忽略眼前的障碍。

当小球碰到管壁爆炸时，本征愉悦度低、控制度低、规则/自我兼容标准的内部标准低，连续失败时外部标准低，容易诱发烦躁情感状态的出现。当小球被顺利移出管道，到达终点时，本征愉悦度高、诱因高、控制度高、规则/自我兼容标准的内部标准高，连续胜利时外部标准高，容易诱发喜悦情感状态的出现。当游戏进行到一半时，小球处于管道中，相关度高、期待度较高、迫切度高、调节度低(游戏难度低时)，容易诱发紧张情感状态的出现。

4 情感语音数据的听辨与评选

为了保证所采集的情感语料的可靠性，对采集的语音情感数据进行了主观听辨与评选，每句样本由 10 名未参与录音的人员进行评测。标度 1、3、5、7、9 表示情感的五种强度，对应极弱，较弱，一般，较强，极强五个等级。

每个情感材料相对于每个听辨人都会产生一个评测的结果 $\boldsymbol{E}_i$：

$$\boldsymbol{E}_{ij}=\{e_1^{ij},e_2^{ij},e_3^{ij},e_4^{ij}\} \tag{1}$$

这里 j 用来标记情感材料，i 用于标记听辨人，矢量中的不同评价因子 e 代表该听辨人对于该情感语句的不同情感成分的评判值。

由于采取多人评测，为了得到第 j 条数据的评价结果，需要将所有听辨人的测评结果进行融合，得到该情感语句中情感成分的融合评判结果为：

$$\boldsymbol{E}_j=\sum_{i=1}^{M}a_i\boldsymbol{E}_{ij} \tag{2}$$

其中 a_i 是每个听辨人的评价结果的融合权重，M 为听辨人总数，它代表每个听辨人的评价结果的可靠程度，有：

$$\sum_{i=1}^{M}a_i=1 \tag{3}$$

选取合理的融合权重，需要对每个听辨人的评测质量进行评价。在多人的评测系统中，不同听辨人的评价结果必需是带有一定的联系或者相关性才能通过一定的方法进行判决融合。下面我们从听辨结果的一致度方面来计算融合权值。

表 2 游戏的刺激评估表

Table 2 The stimulus evaluation checks for a computer game

标度值	刺激评估子项	游戏情节
1	新异度(Novelty)	高：进入新的游戏级别 低：重玩旧的游戏级别
2	本征愉悦度(Intrinsic Pleasantness)	高：小球移出管道 低：小球碰撞管道
3	目标/需求重要度(Goal/Need Significance)	
	相关度(Relevance)	高：开始游戏，进入倒计时 低：未开始游戏，观察管道结构
	期待程度(Expectation)	高：小球离出口较远，且游戏刚开始 低：小球接近出口
	诱因(Conduciveness)	高：小球接近出口 低：小球离出口较远
	迫切度(Urgency)	高：剩余时间少 低：剩余时间多
4	应对潜能(Coping Potential)	
	原因(Causation)	事件发生的原因评价，分为内因和外因
	控制(Control)	高：鼠标移动平稳流畅 低：鼠标抖动
	力量(Power)	无
	调节(Adjustment)	高：当前游戏难度高 低：当前游戏难度低
5	规则/自我兼容标准(Norm/Self-compatibility)	
	外部标准(External standards)	高：游戏成绩进步 低：游戏成绩退步
	内部标准(Internal standards)	高：在游戏难度大的情况下胜利 低：在游戏难度小的情况下失败

对第 j 条数据，两个听辨人 p、q 之间的相似性度量定义为：

$$\rho_j^{pq}=\prod_{i=1}^{K}\frac{\min\{e_i^{pj},e_i^{qj}\}}{\max\{e_i^{pj},e_i^{qj}\}} \tag{4}$$

其中 K 为情感类别总数，这里 K=4。对每次测评，两个听辨人 p、q 之间的相似性度量为：

$$\rho^{pq}=\frac{1}{N}\sum_{j=1}^{N}\rho_j^{pq}=\frac{1}{N}\sum_{j=1}^{N}\prod_{i=1}^{K}\frac{\min\{e_i^{pj},e_i^{qj}\}}{\max\{e_i^{pj},e_i^{qj}\}} \tag{5}$$

其中 N 为测评数据的总数。根据两人之间的相

似性，可以得到一个一致度矩阵 $\boldsymbol{\rho}$，矩阵中的每个元素代表两个听辨人之间的相互支持程度，其中 M 为听辨人总数：

$$\boldsymbol{\rho}=\begin{bmatrix} 1 & \rho^{12} & \cdots & \rho^{1M} \\ \rho^{21} & 1 & \cdots & \rho^{2M} \\ \cdots & \cdots & \cdots & \cdots \\ \rho^{M1} & \rho^{M2} & \cdots & 1 \end{bmatrix} \tag{6}$$

计算平均一致度，即反映第 i 个听辨人与其他听辨人之间的一致程度：

$$\overline{\rho^{i}}=\frac{1}{M-1}\sum_{\substack{j=1\\ j\neq i}}^{M}\rho^{ij} \tag{7}$$

以归一化之后的一致度作为每个听辨人评测结果的融合权重 a_i，

$$a_i=\tilde{\rho}^{i}=\overline{\rho^{i}}\Big/\sum_{i=1}^{M}\rho^{i} \tag{8}$$

将其代入式(2)即得到的每条情感语句的评价结果 $\boldsymbol{E}_j$。

根据评价结果可对数据进行筛选，假设 $\boldsymbol{E}_j$ 中最大的元素是 e_p^j，则认为该情感语句为主情感为 p 的情感语料，作为最终评判的结果。如评判结果和该语句标注的情感类别(诱发的目标情感)相符合，则保留；否则，认为此条语句的情感诱发不成功，剔除。

5　结 论

烦躁、紧张以及喜悦等情感的出现，对人员的心理状态和工作能力，包括认知能力，具有非常重要的影响。因此对此类实用语音情感的研究具有十分重要的实际意义。针对目前实用情感语音数据库极为缺乏的现状，本文通过计算机游戏诱发的方式，采集诱发语音数据并且建立了一个实用语音情感数据库，包括烦躁、紧张、喜悦和平静四种情感状态的情感语音，共计 4800 条。合理地诱发出高自然度的情感语音，是采集情感数据的一个难点。Scherer 的刺激评估表(SECs)，能够给游戏情节如何诱发对应的目标情感，提供一个理论上的依据。建立实用语音情感数据库，是语音情感识别进入实际应用的一个重要基础和前提工作。在今后的工作中，需要研究更多的具有实际应用意义的语音情感，进一步开发情感识别在实际中的应用。在心理学诱发实验方面，根据 SECs 进行情景设置的专用的情感诱发游戏有待进一步开发，以便运用 SECs 进行更为准确和高效的情感数据诱发。

参　考　文　献

[1] 乔杰, 赵力, 邹采荣. 一种新的利用主观听觉特性的回波抵消算法[J]. 声学技术, 2007, **26**(4):704-708.
QIAO Jie, ZHAO Li, ZOU Cairong. Acoustic echo cancellation using human auditory[J]. Technical Acoustics, 2007, **26**(4): 704-708.

[2] 赵艳, 赵力, 邹采荣. 耳语音的语音处理研究综述[J]. 声学技术, 2008, **27**(4): 562-567.
ZHAO Yan, ZHSO Li, ZOU Cairong. A survey of whisper analysis and its processing[J]. Technical Acoustics, 2008, **27**(4): 562-567.

[3] 余华, 陈国明, 赵力, 邹采荣. 基于改进的 Kalman 滤波的语音增强算法[J]. 声学技术, 2009, **28**(6): 763-767.
YU Hua, CHEN Guoming, ZHAO Li, ZOU Cairong. Speech enhancement based on improved Kalman filtering method[J]. Technical Acoustics, 2009, **28**(6): 763-767.

[4] Vinciarelli A, Pantic M, Bourlard H, Pentland A. Social signal processing: Survey of An Emerging Domain[J]. Image Vis. Comput. J.(0262-8856), 2009, **27**(12): 1743-1759.

[5] Barbara A, Spellman, Daniel, Willingham T. Current directions in cognitive science[M]. Beijing: Beijing Normal University Press, 2007. 1-5.

[6] Zeng Z, Pantic M, Roisman G I, Huang T H. A survey of affect recognition methods: audio, visual and apontaneous expressions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence(0162-8828), 2009, **31**(1): 39-58.

[7] ZHAO Yan, ZHAO Li, ZOU Cairong, YU Yinhua. Speech emotion recognition using modified quadratic discriminatioin function[J]. Journal of Electronics (China), 2008, **25**(6): 840-844.

[8] Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor J G. Emotion recognition in human-computer interaction[J]. IEEE Signal Processing Magazine(1053-5888), 2001, **18**(1): 32-80.

[9] 董燕, 施承孙, 周晓梅, 侯桂芝. 飞行人员情绪表达特征对认知绩效的影响[J]. 第四军医大学学报, 2005, **26**(4): 374-377.
DONG Yan, SHI Chengsun, ZHOU Xiaomei, HOU Guizhi. Analysis of relationship between cattell-16PF emotional personality characters and cognition performance in pilots[J]. J. Fourth Mil. Med. Univ., 2005, **26**(4): 374-377.

[10] 王治平. 情感语音信号特征分析与识别[D]. 南京：东南大学, 2004.
WANG Zhiping. Feature analysis and emotino recognition in emotional speech[D]. Nanjing: Southeast University, 2004.

[11] LDC. TIMIT acoustic-phonetic continuous speech corpus [DB/OL]. 2004-11-10/2009-7-1. http://wave.ldc.upenn.edu/

[12] OGI. Numbers Corpus V1.2[DB/OL]. 2001-5-2/2009-7-1. http://cslu.cse.ogi.edu/corpora/numbers

[13] ELRA. ELRA Catalogue of Language Resources[DB/OL]. 2003-4-29/2009-7-1. http://catalog.elra.info/product_info.php

[14] 李爱军, 王天庆, 殷治刚. 863 语音识别语音语料库 RASC863----四大方言普通话语音库[C]. 第七届全国人机语音通讯学术会议, 2003, **7**: 41-44.
LI Aijun, WANG Tiangqing, YIN Zhigang. RASC-863:An annotated speech corpus with 4 regional accents[C]. National Conference on Man-Machine Speech Communication, 2003, **7**: 41-44.

[15] Klaus R. Scherer. Vocal communication of emotion: a review of research paradigms[J]. Speech Communication(0167-6393), 2003, **40**(1-2): 227-256.

[16] Scherer K. R. Vocal affect expression: a review and a model for future research[J]. Psychological Bulletin(0033-2909), 1986, **99**(2): 143-165.

[17] Johnstone T. Emotional speech elicited using computer games[C]. Fourth International Conference on Spoken Language, 1996, **3**: 1985-1988.

[18] Johnstone T, CM van Reekum, Hird K, Kirsner K, et al. Affective speech elicited with a computer game[J]. Emotion(1528-3542), 2005, **5**(4): 513-518.

[19] Presentations of Science base. Stimulus Evaluation Checks [DB/OL]. 2007-10-4/2009-7-1. http://www.posbase.uib.no/