



(12) 发明专利

(10) 授权公告号 CN 103198827 B

(45) 授权公告日 2015.06.17

(21) 申请号 201310100284.7

(22) 申请日 2013.03.26

(73) 专利权人 合肥工业大学

地址 230009 安徽省合肥市屯溪路 193 号

(72) 发明人 陈雁翔 龙润田

(74) 专利代理机构 安徽合肥华信知识产权代理

有限公司 34112

代理人 余成俊

(51) Int. Cl.

G10L 13/033(2013.01)

G10L 13/08(2013.01)

G10L 25/63(2013.01)

审查员 王玥

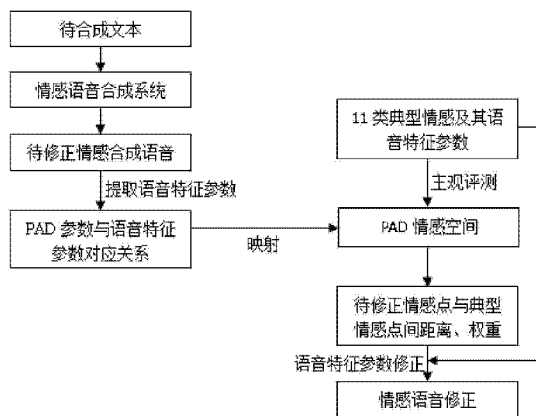
权利要求书2页 说明书5页 附图2页

(54) 发明名称

基于韵律特征参数和情感参数关联性的语音情感修正方法

(57) 摘要

本发明公开了一种基于韵律特征参数和情感参数关联性的语音情感修正方法,该方法包括:对 11 类典型情感进行情感建模;对待修正的情感语音提取特征参数,并将其映射至情感空间内;利用情感参数和语音特征参数的关联关系,以及两类语音参数基于不同的信息度量间的互补性;通过情感参数对语音特征参数的修正获得更客观准确的情感表达。本发明方法能够更细致的获得所需合成的情感状态,自然度高,可广泛应用于语音处理与先进智能领域。



1. 一种基于韵律特征参数和情感参数关联性的语音情感修正方法,其特征在于,具体包括以下步骤:

步骤 1:建立 PAD 情感模型并对 n 种典型情感进行 PAD 主观评分;

步骤 2:对经过情感语音合成系统合成出的待修正情感语音提取语音特征参数,利用语音特征参数与 PAD 情感参数的关联关系,将待修正情感语音映射至 PAD 情感空间内;待修正情感语音的情感空间值按如下步骤获取:

步骤 2.1:对于待修正情感合成语音提取梅尔倒谱参数(MFCC)和短时能量;

步骤 2.2:根据 PAD 情感空间中各项目与语音特征参数的关系,计算 PAD 空间各项目值;

步骤 2.3:待修正情感合成语音的 PAD 情感空间值由使用不同参数计算出的 PAD 空间值进行加权平均获得;加权比例为 70% 和 30%;

步骤 3:采用不均匀正态分布空间欧氏距离计算方法,计算待修正情感点到各个典型情感点的欧氏距离,PAD 情感空间为三维空间,两情感点间的最终距离由三维空间内两点间距离计算方法计算,最终获得待修正点到 n 类典型情感点的 n 个情感距离;PAD 情感空间中两情感点间距离按以下步骤获得:

步骤 3.1:采用不均匀正态分布空间欧氏距离计算方法,按式(1)计算待修正情感点到各典型情感点间的欧氏距离;

$$S(p_1 - p_2) = \frac{\|p_1^2 - p_2^2\|}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad (1);$$

其中 p_1, p_2 为两情感点在 PAD 空间内的 p 参数观察值, σ_a 和 σ_b 为两个情感类别的方差;当所建立的 PAD 情感空间为归一化 $(-1, +1)$ 的空间时,方差远远小于 1;

步骤 3.2:由 PAD 为三维情感空间,则 a, b 两情感间的最终距离 S 为:

$$S_{ab} = \sqrt{S_p^2 + S_A^2 + S_D^2} \quad (2);$$

其中 S_p, S_A, S_D 分别为 a, b 两情感点间 P、A、D 坐标距离;

步骤 4:根据待修正情感点到各个典型情感点的距离,计算修正情感点的基本情感组成,待修正点到各个典型情感点的距离与待修正点的典型情感组成权重成反比,采用排序组合方法获得待修正点的各典型情感组成权重;待修正情感点中各典型情感组成权重关系由式(3)获得:

$$M_{min} = \frac{S_{max}}{\sum_{i=1}^n S_i} \quad (3);$$

其中 S_{max} 为待修正点与距离其最远的典型情感点间的距离, M_{min} 为与待修正点距离 S_{max} 的典型情感所占组成权重;采用排序组合方法获得待修正点的各典型情感组成权重;

步骤 5:根据典型情感点对应的语音特征参数以及各典型情感在待修正情感中所占比重,采用线性修正算法,对合成情感特征参数进行修正;语音特征参数情感修正量,采用线性修正算法获得,其计算方法如式(4):

$$H' = \left(\sum_{i=1}^n M_i \bullet H_i \right) - H \quad (4);$$

其中， H 为原合成情感语音特征参数， H_i 为典型情感语音特征参数， H' 为参数校正量。

基于韵律特征参数和情感参数关联性的语音情感修正方法

技术领域

[0001] 本发明属于语音信号处理技术领域,涉及到一种语音合成中的关联性参数修正方法,更具体的涉及了一种基于韵律特征参数和情感参数关联性的语音情感修正方法。

背景技术

[0002] 随着计算机技术和信号处理技术的高速发展,传统的语音合成技术已经无法满足人们越来越高的要求。人们希望能够与机器进行自然和谐的交流,语音作为人类社会重要的交流方式,很大程度上决定了这种人机自然和谐交流的实现。因此语音合成在情感计算、信号处理领域具有很重要的研究意义。而细腻的情感表达则能够较大地提高合成语音的自然度。

[0003] 在语言表达过程中,说话人的言语中常常并不只包含有一种情感。很多语句有可能是几种情感融合后地表达。同时,人在情感表达的过程中语音和表情均对情感表现及其程度的产生影响。现有的情感表达研究主要集中在对单一语音韵律特征、情感特征或生理信息的情感合成,研究对象局限于实验室中的特定说话人或数据库中特定表演者的范围内,导致目前的情感合成方法无法合成出多种情感相互融合的综合化表情,使得合成语音情感表达不自然,工程应用效果不理想。

[0004] 在语音情感描述领域, Pereira 提出的“激励—评价—控制”(PAD)三维情感空间理论。用模糊化的情感空间标记方式,可以对语句进行维度空间的标注,这种维度空间中的标注可以较为准确的分析出每个语句的情感组成,从而能获得较为合理、精确的情感标记。

[0005] 在情感语音合成领域,目前语音情感合成方法主要分为两类:基于情感语料的情感语音合成和基于中性语料的情感语音合成。由于训练模型较小,因此合成语音的情感质量较差。采用基于韵律特征参数和情感参数关联性的语音情感修正方法可以较好的解决训练模型较小导致的情感数据不准确以及合成语音情感质量不高这一问题。因此提出一种有效的情感表达修正方法具有很强的现实意义和工程实践意义。

发明内容

[0006] 本发明的目的在于提出一种能够对合成语音所表达情感进行修正的方法,以解决合成语音情感表达不自然的问题。

[0007] 本发明采用的技术方案是:

[0008] 本发明提供了一种基于韵律特征参数和情感参数关联性的语音情感修正方法,包括以下步骤:

[0009] 步骤 1:建立 PAD 情感模型并对 n 种典型情感进行 PAD 主观评分。

[0010] 步骤 2:对经过情感语音合成系统合成出的待修正情感语音提取语音特征参数。利用语音特征参数与 PAD 情感参数的关联关系,将待修正情感语音映射至 PAD 情感空间内。

[0011] 步骤 3:采用不均匀正态分布空间欧氏距离计算方法,计算待修正情感点到各个典型情感点的欧氏距离。设 p_a, p_b 为 a、b 两情感点在 PAD 空间内的 p 参数观察值, σ_a 和 σ_b

为两个情感类别的方差。欧式距离计算公式如下：

$$[0012] \quad S(p_a - p_b) = \frac{\|p_a^2 - p_b^2\|}{\sqrt{\sigma_a^2 + \sigma_b^2}}$$

[0013] 由 PAD 为三维情感空间,则 a、b 两情感间的最终距离函数为：

$$[0014] \quad S_{ab} = \sqrt{S_P^2 + S_A^2 + S_D^2}$$

[0015] 最后,获得待修正点到 n 类典型情感点的 n 个情感距离。

[0016] 步骤 4 :如 :一个待修正点到各个典型情感点的距离分别为 : S_1, S_2, \dots, S_n 。则待修正点的典型情感组成权重为：

$$[0017] \quad M_{\min} = \frac{S_{\max}}{\sum_{i=1}^n S_i}$$

[0018] 其中 S_{\max} 为待修正点与距离其最远的典型情感点间的距离, M_{\min} 为与待修正点距离 S_{\max} 的典型情感所占组成权重。采用排序组合方法获得待修正点的各种典型情感组成权重。

[0019] 步骤 5 :根据典型情感点对应的语音特征参数以及各种典型情感在待修正情感中所占比重。采用线性修正算法,对合成情感特征参数进行修正。

[0020] 本发明提供的情感语音修正方法,其优点和积极效果在于：

[0021] 1、该方法是基于情感心理学和语音处理理论,从不同角度对情感状态进行了分析。

[0022] 2、该方法利用语音特征参数与情感参数的关联性关系,对合成语音情感状态进行了修正,提高了合成语音的自然度,这一成果还可以推广到其它多模态模型中。

附图说明

[0023] 图 1 是本发明方法的流程框图。

[0024] 图 2 是 PAD 主观评价方法图。

[0025] 图 3 是 PAD 情感空间模型示意图。

具体实施方式

[0026] 本发明的基本思想是通过语音所表达的情感进行建模,将情感参数和语音特征参数相关联,并利用情感参数与语音特征参数的关联关系对情感模型中的情感点进行语音特征参数的修正。最终获得修正情感合成语音。

[0027] 根据以上思想,本发明流程如图 1 所示,下面结合技术方案和附图对本发明的方法作进一步说明。

[0028] 首先分析待合成文本,利用 HMM 语音合成系统合成情感语音。同时对建立 PAD 情感模型并将 11 类典型情感映射至模型中。其次,对合成出的待修正情感语音提取语音特征参数。再次将获得情感参数利用 PAD 情感与特征参数的对应关系放入 PAD 三维情感空间中,判断合成参数的情感置信度,并对其语音特征参数进行校正。最终得到所需的情感合成语音。

[0029] 以男性情感语音修正为例,本方法的具体步骤如下：

[0030] 步骤 1 :对 11 类典型情感进行 PAD 主观评测,评测流程如图 2 所示。建立 PAD 情感模型如图 3 所示。将 11 类典型情感归一化映射到情感空间中。11 类典型情感 PAD 参考值如下所示 :

情感	PAD 参考值			情感	PAD 参考值		
	P	A	D		P	A	D
中性	0	-0.31	-0.06	厌恶	-0.44	0.22	0.36
放松	-0.01	-0.78	0.33	恐惧	-0.33	0.65	-0.72
温顺	-0.27	-0.07	-0.14	悲伤	-0.28	-0.36	-0.78
惊奇	0.23	0.64	0.01	焦虑	-0.46	0.58	-0.13
喜悦	0.66	0.74	0.32	愤怒	-0.86	0.66	0.91
轻蔑	-0.38	-0.7	0.6	-	-	-	-

[0032] 男性各个典型情感的语音特征参数如下所示 :

情感	基频均值	基频范围	基频方差	平均能量	平均时长	时长方差
中性	117.5	18.9	9.0	34.8	3428.3	16936.1
放松	127.9	23.1	88.6	32.6	3307.8	218954.0
温顺	125.7	22.8	310.0	32.4	2893.7	93080.2
惊奇	196.2	45.1	967.6	40.7	2278.0	83622.7
喜悦	187.5	44.3	1188.7	41.3	2447.6	107305.4
轻蔑	137.3	25.1	373.2	33.2	2895.6	122243.9
厌恶	177.4	43.6	701.7	37.2	2636.9	92265.5
恐惧	206.7	45.9	1532.0	37.6	2320.1	92265.6
悲伤	143.8	23.3	1226.6	32.8	3278.1	420285.8
焦虑	222.4	54.3	1150.3	40.2	1884.1	13279.1
愤怒	221.1	56.1	1006.3	40.9	1833.2	59716.6

[0034] 步骤 2 :对由情感语音合成系统所合成的情感语音提取相关语音特征参数,一般地提取梅尔倒谱参数和短时能量参数。同时利用如下关系将合成的待修正情感语音映射到具有典型情感点的情感空间中。

	特征	P	A	D	特征	P	A	D
[0035]	平均基频	-0.2712	0.6155	-0.0649	基频斜度	0.0042	-0.0512	0.0344
	最大基频	-0.2702	0.6196	-0.0608	首字基频	-0.2503	0.5808	-0.0658
	最小基频	-0.2543	0.5719	-0.0639	末尾基频	-0.2481	0.5690	-0.0566
	平均时长	0.0592	-0.2504	0.0250	短时能量	-0.4005	0.5749	0.0739
	停顿时长	0.0537	-0.1074	0.0006	语速	0.3056	-0.8137	0.0432

[0036] 对于男性而言, MFCC 特征与男性 PAD 情感空间参数的关系如下:

MFCC 系数	男性			MFCC 系数	男性			
	P	A	D		P	A	D	
[0037]	1	0.0141	-0.0641	-0.0015	8	0.0451	-0.1730	0.0703
	2	-0.0655	0.0984	-0.0134	9	0.0812	-0.2190	-0.0658
	3	0.0654	-0.2063	-0.0905	10	0.0178	0.1212	0.0205
	4	0.0514	-0.0627	-0.0856	11	0.0207	-0.0888	0.0407
	5	-0.0259	0.0954	-0.0107	12	0.0660	-0.1460	-0.0431
	6	0.1041	-0.3818	-0.1066	13	0.0071	-0.0265	-0.0457
	7	-0.0090	0.1713	-0.0372				

[0038] 待修正情感合语音的 PAD 情感空间值由使用不同参数计算出的 PAD 空间值进行加权平均获得, 加权比例为 70% 和 30%。

[0039] 步骤 3: 采用不均匀正态分布空间欧氏距离计算方法, 计算待修正情感点到各个典型情感点间的欧氏距离。

[0040] 3.1: 计算公式如式(1)所示:

$$[0041] \quad S(p_a - p_b) = \frac{\|p_a^2 - p_b^2\|}{\sqrt{\sigma_a^2 + \sigma_b^2}} \quad (1)$$

[0042] 其中 p_a, p_b 为 a、b 两情感点在 PAD 空间内的 p 参数观察值, σ_a 和 σ_b 为两个情感类别的方差。当所建立的 PAD 情感空间为归一化(-1, +1)的空间时, 方差远远小于 1。

[0043] 3.2: 由 PAD 为三维情感空间, 则 a、b 两情感间的最终距离为:

$$[0044] \quad S_{ab} = \sqrt{S_P^2 + S_A^2 + S_D^2} \quad (2)$$

[0045] 由空间距离测度聚类可知两种情感之间的距离越小, 则这两种情感的声学特征越接近, 采用此方法进行合成情感特征参数修正。

[0046] 步骤 4: 利用待修正情感点到各个典型情感点间的距离, 计算待修正情感点的基本情感组成。设一个待修正点到各个典型情感点的距离分别为: S_1, S_2, \dots, S_n 。则待修正点的典型情感组成权重为:

[0047]
$$M_{\min} = \frac{S_{\max}}{\sum_{i=1}^n S_i} \quad (3)$$

[0048] 其中 S_{\max} 为待修正点与距离其最远的典型情感点间的距离, M_{\min} 为与待修正点距离 S_{\max} 的典型情感所占组成权重。在本例中所用典型情感为 11 类, 故 n 取值为 11。采用排序组合方法获得待修正点的各个典型情感组成权重。

[0049] 步骤 5: 对于典型情感点对应的语音特征参数, 及待修正情感点的典型情感组成。采用线性修正算法, 获得各语音特征参数的修正量。对待修正情感点的各语音特征参数进行修正, 计算公式如式 (4) 所示:

[0050]
$$H' = \left(\sum_{i=1}^n M_i \bullet H_i \right) - H \quad (4)$$

[0051] 其中, H 为原合成情感语音特征参数, H_i 为典型情感语音特征参数, H' 为参数校正量。

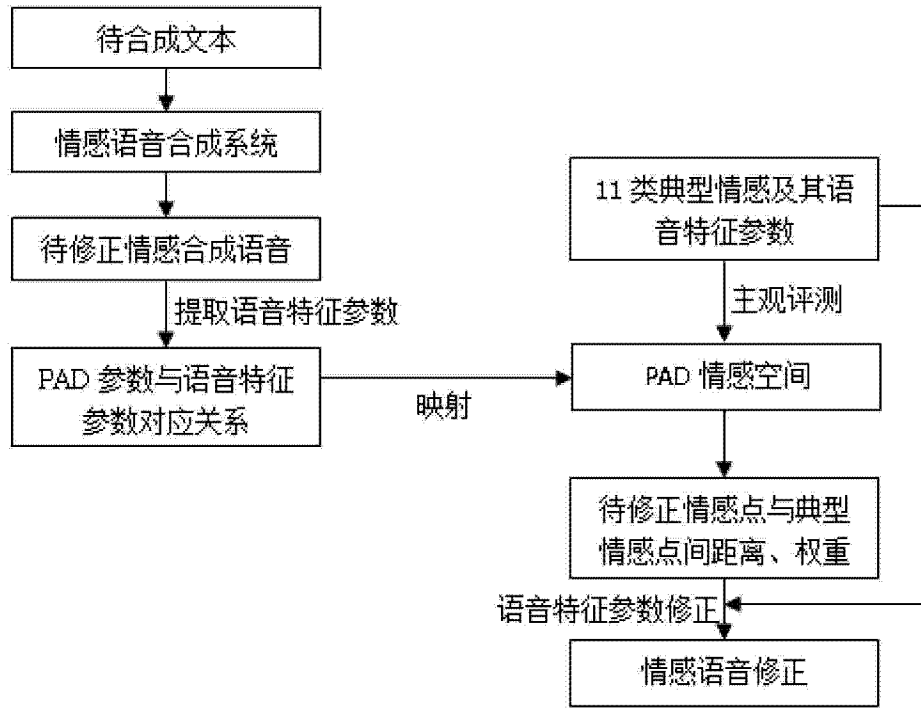


图 1

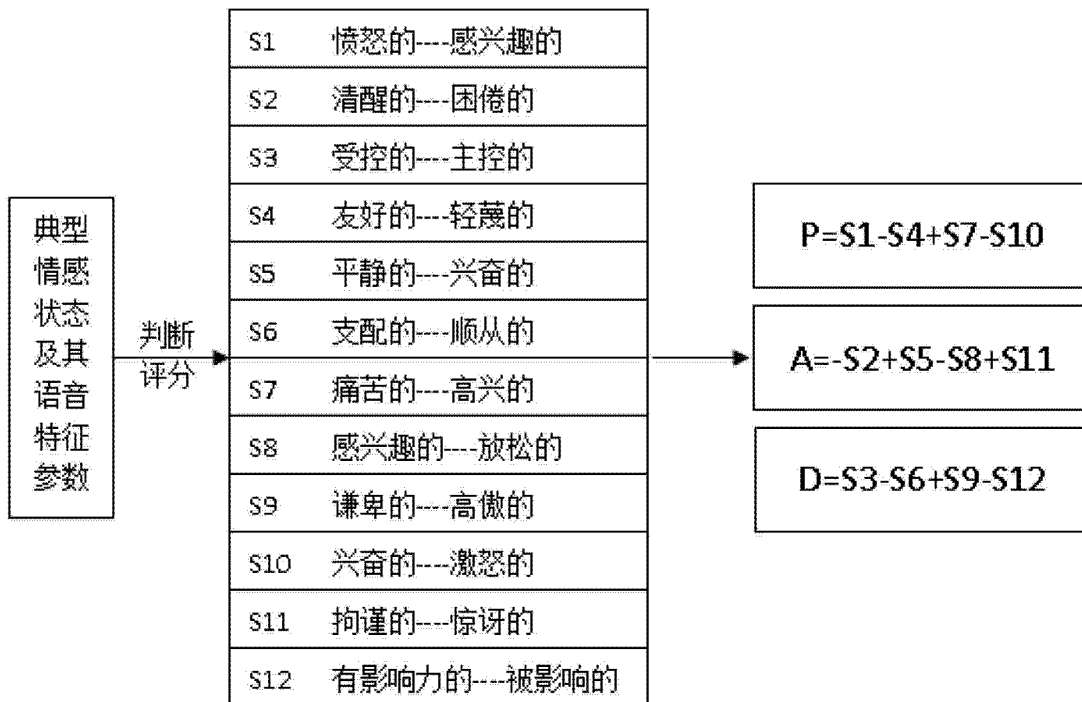


图 2

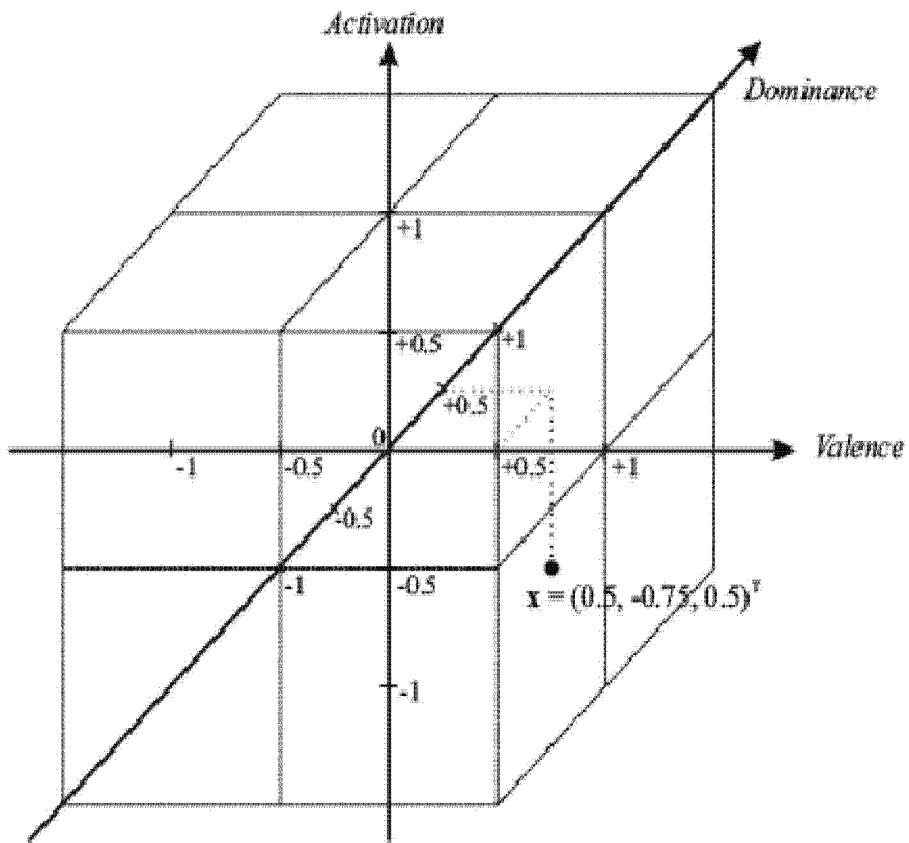


图 3