



(12)发明专利申请

(10)申请公布号 CN 107221344 A

(43)申请公布日 2017. 09. 29

(21)申请号 201710222674.X

G10L 19/02(2013.01)

(22)申请日 2017.04.07

G10L 25/27(2013.01)

(71)申请人 南京邮电大学

G10L 25/30(2013.01)

地址 210023 江苏省南京市栖霞区亚东新
城区文苑路9号

G06N 3/08(2006.01)

(72)发明人 李华康 杜阳阳 金旭 胡晓东
丘添元 张笑源 孙国梓 李涛

(74)专利代理机构 南京知识律师事务所 32207
代理人 李湘群

(51)Int.Cl.

G10L 25/63(2013.01)

G10L 13/02(2013.01)

G10L 15/02(2006.01)

G10L 15/06(2013.01)

G10L 15/16(2006.01)

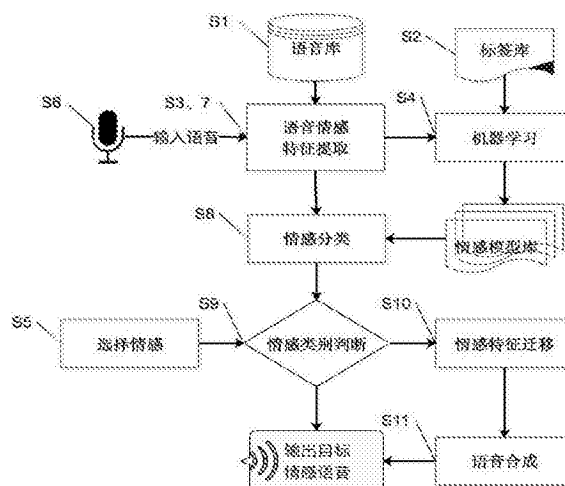
权利要求书1页 说明书4页 附图2页

(54)发明名称

一种语音情感迁移方法

(57)摘要

本发明公开了一种语音情感迁移方法,首先基于语音数据库生成语音情感数据集,完成标签标注,然后采用语音特征参数模型对音频文件进行音频特征抽取,得到语音特征集;接下来采用机器学习工具对语音特征集与语音情感标签进行机器学习,构建情感模型库。选择待迁移的目标,从多媒体终端输入语音信号,得到当前语音信号的特征集,通过情感分类得到当前情感类别,判断与输入的目标是否一致,如果一致则将原始输入语音信号直接作为目标情感语音输出,否则进行特征情感迁移;最后经过语音合成加工生成最终目标情感语音输出。本发明提出的基于情感分类和特征迁移的方法,能够在不失原始说话人发声特征的前提下实现语音情感的变化。



1. 一种语音情感迁移方法,其特征在于包含以下步骤:

步骤1、准备一个语音数据库,通过标准采样生成语音情感数据集 $S = \{s_1, s_2, \dots, s_n\}$;

步骤2、采用人工方式对步骤1的语音数据库打标签,标注每个语音文件的情感 $E = \{e_1, e_2, \dots, e_n\}$;

步骤3、采用语音特征参数模型对语音库中的每个音频文件 s_i 进行音频特征抽取,得到基本的语音特征集 $F_i = \{f_1^i, f_2^i, \dots, f_n^i\}$;

步骤4、采用机器学习工具对步骤3得到的每个语音特征集与步骤2得到的语音情感标签进行机器学习,得到每一类语音情感的特征模型,构建情感模型库 E_b ;

步骤5、通过一个多媒体终端,选择需要语音情感迁移的目标 $Target_e$;

步骤6、从多媒体终端输入语音信号 s_t ;

步骤7、将当前输入的 s_t 输入到语音情感特征提取模块,得到当前语音信号的特征集 $F_t = \{f_1^t, f_2^t, \dots, f_n^t\}$;

步骤8、采用与步骤4相同的机器学习算法,将步骤7得到的 s_t 的语音特征集 F_t 结合步骤4得到的情感模型库 E_b 进行情感分类,得到 s_t 的当前情感类别 s_e ;

步骤9、判断步骤8得到的 s_e 和步骤5输入的 $Target_e$ 是否一致,如果 $s_e = Target_e$,则将原始输入语音信号直接作为目标情感语音输出,如果 $s_e \neq Target_e$,则调用步骤10进行特征情感迁移;

步骤10、将当前语音情感主要特征向情感模型库中的语音情感主要特征进行迁移;

步骤11、采用语音合成算法对步骤10得到的特征迁移后的语音特征进行加工,合成最终目标情感语音输出。

2. 根据权利要求1所述的语音情感迁移方法,其特征在于步骤1中语音数据的采样频率为44.1KHz,录音时间在3~10s之间,并且保存为wav格式。

3. 根据权利要求1所述的语音情感迁移方法,其特征在于步骤1中为了获得较好的性能,采样数据的自然属性维度不能过于集中。

4. 根据权利要求1所述的语音情感迁移方法,其特征在于步骤6中所述输入可以是实时输入,也可以是录制完成后点击递交。

一种语音情感迁移方法

技术领域

[0001] 本发明属于语音识别技术领域,涉及语音情感的迁移方法,具体涉及一种基于不同语音提供者模型的语音情感的迁移方法。

背景技术

[0002] 随着智能芯片技术的发展,各种终端设备的智能化和集成化程度越来越高,设备的小型化、轻便化、网络化使得人们的生活越来越便捷。用户不断的通过网络终端进行语音视频的交流,积累了海量的多媒体数据。随着平台数据的积累,智能问答系统也逐渐应运而生。这些智能问答系统包括了语音识别、情感分析、信息检索、语义匹配、句子生成、语音合成等先端技术。

[0003] 语音识别技术是让机器通过识别技术和理解过程把语音信号转化为所对应的文本信息或者机器指令,让机器能够听懂人类的表达内容,主要包括语音单元选取、语音特征提取、模式匹配和模型训练等技术。语音单元包括单词(句)、音节和音速三种,具体按照场景和任务来选择。单词单元主要适合小词汇语音识别系统;音节单元更加适合于汉语语音识别;音素虽然能够很好地解释语音基本成分,但由于发音者的复杂多变导致无法得到稳定的数据集,目前仍在研究中。

[0004] 另一个研究方向是语音的情感识别,主要由语音信号采集、情感特征提取和情感识别组成。其中情感特征提取主要有韵律学特征、基于谱的相关特征和音质特征三种。这些特征一般以帧为最小粒度来实现提取,并以全局特征统计值的形式进行情感识别。在情感识别算法方面,主要包括离散语言情感分类器和维度语音情感预测器两大类。语音情感识别技术也被广泛应用于电话服务中心、驾驶员精神判别、远程网络课程等领域。

[0005] 智能体被誉为是下一代人工智能的综合产物,不仅能够识别周围环境因素,理解人的行为表达和语言描述,甚至在与人的交流过程中,更需要去理解人的情感,并且能够实现模仿人的情感表达,才能实现更为柔和的交互。目前智能体的情感研究主要集中在基于虚拟图像处理,涉及计算机图形学、心理学、认知学、神经生理学、人工智能等多个领域有研究者的成果。据研究,人虽然90%以上的环境感知信息来自视觉,但是绝大部分的情感感知是来自语音。如何从语音领域建立类人智能体的情感体系,至今尚未有公开的研究发布。

发明内容

[0006] 本发明的目的是以机器学习方法为主要手段,提出一种人的语音情感表述方法,并在此基础上使用深度学习和卷积网络算法,从系统上实现语音情感的迁移。不仅对语音识别、情感分析提供了一定的借鉴方法,更能在未来类人智能体上得到广泛应用。

[0007] 为实现上述目的,本发明提出的技术方案为一种语音情感迁移方法,具体包含以下步骤:

[0008] 步骤1、准备一个语音数据库,通过标准采样生成语音情感数据集 $S = \{s_1, s_2, \dots, s_n\}$;

[0009] 步骤2、采用人工方式对步骤1的语音数据库打标签,标注每个语音文件的情感 $E = \{e_1, e_2, \dots, e_n\}$;

[0010] 步骤3、采用语音特征参数模型对语音库中的每个音频文件 s_i 进行音频特征抽取,得到基本的语音特征集 $F_i = \{f_1^i, f_2^i, \dots, f_n^i\}$;

[0011] 步骤4、采用机器学习工具对步骤3得到的每个语音特征集与步骤2得到的语音情感标签进行机器学习,得到每一类语音情感的特征模型,构建情感模型库 E_b ;

[0012] 步骤5、通过一个多媒体终端,选择需要语音情感迁移的目标Target;

[0013] 步骤6、从多媒体终端输入语音信号 s_t ;

[0014] 步骤7、将当前输入的 s_t 输入到语音情感特征提取模块,得到当前语音信号的特征集 $F_t = \{f_1^t, f_2^t, \dots, f_n^t\}$;

[0015] 步骤8、采用与步骤4相同的机器学习算法,将步骤7得到的 s_t 的语音特征集 F_t 结合步骤4得到的情感模型库 E_b 进行情感分类,得到 s_t 的当前情感类别 s_e ;

[0016] 步骤9、判断步骤8得到的 s_e 和步骤5输入的Target是否一致,如果 $s_e = Target_e$,则将原始输入语音信号直接作为目标情感语音输出,如果 $s_e \neq Target_e$,则调用步骤10进行特征情感迁移;

[0017] 步骤10、将当前语音情感主要特征向情感模型库中的语音情感主要特征进行迁移;

[0018] 步骤11、采用语音合成算法对步骤10得到的特征迁移后的语音特征进行加工,合成最终目标情感语音输出。

[0019] 进一步,上述步骤1中,语音数据的采样频率为44.1KHz,录音时间在3~10s之间,并且保存为wav格式。

[0020] 步骤1中,为了获得较好的性能,采样数据的自然属性维度不能过于集中,采样数据尽量在不同年龄、性别、职业等人中采集。

[0021] 步骤6中,所述输入可以是实时输入,也可以是录制完成后点击递交。

[0022] 本发明具有以下有益效果:

[0023] 1、本发明首先提出语音情感迁移的概念,可以为未来虚拟现实提供情感构建方法。

[0024] 2、本发明提出的基于情感分类和特征迁移的方法,能够在不失原始说话人发声特征的前提下实现语音情感的变化。

附图说明

[0025] 图1是本发明提供的语音情感迁移方法示意图。

[0026] 图2是本发明原始输入语音样本的频谱特征图。

[0027] 图3是本发明原始语音样本经过情感转化的频谱特征图。

具体实施方式

[0028] 现结合附图对本发明作进一步详细的说明。

[0029] 本发明提供一种基于语音情感数据库的用户表达语音情感迁移方法,如图1所示,该方法涉及的模块或功能包括:

- [0030] 基础语音库,存有不同年龄、性别、场景下的语音原始数据。
- [0031] 标签库,对基础语音库进行情感标注,如平和、高兴、生气、愤怒、悲伤等。
- [0032] 语音输入装置,如麦克风,可以实现用户的实时语音输入。
- [0033] 语音情感特征提取,通过声音特征分析工具,得到一般的声音特征,并根据人的语音信号特点以及情感表现特点,选取所需的特征集作为语音情感特征。
- [0034] 机器学习,采用机器学习算法印证语音情感标签库,对语音情感特征集构建训练模型。
- [0035] 情感模型库,语音库数据通过机器学习得到的按照性别、年龄、情感等维度分类后的语音情感模型库。
- [0036] 选择情感,用户在输入语音信号前选择需要将当前语音实时转化为的情感模式。
- [0037] 情感类别判断,判断当前用户输入的情感是否与选择的情感一致。如果一致,则直接输出目标情感语音。如果不一致,调用情感迁移模块。
- [0038] 情感迁移,在用户输入语音和选择情感不一致的情况下,将输入语音情感特征集与选择情感特征集进行特征距离对比,调整输入语音情感特征空间表示,实现情感迁移。然后将调整好的情感语音作为目标情感语音输出。
- [0039] 现提供一个实施例,以说明语音情感的迁移过程,具体包含以下步骤:
- [0040] 步骤1、该方法需要准备一个语音数据库,作为优选,语音数据采用标准采样44.1KHz,录下某个测试人员一句话,时间在3~10s之间,并且保存为wav格式,得到语音情感数据集 $S = \{s_1, s_2, \dots, s_n\}$ 。为了获得较好的性能,采样数据尽力在不在年龄、性别、职业等人的自然属性维度过于集中。
- [0041] 步骤2、采用人工的方式,对步骤1准备的语音数据库打标签,标注每个语音文件的情感 $E = \{e_1, e_2, \dots, e_n\}$,如“担心”,“吃惊”,“生气”,“失望”,“悲伤”等
- [0042] 步骤3、采用语音特征参数模型对语音库中每个音频文件 s_i 进行音频特征抽取,得到基本的语音特征集 $F_i = \{f_1^i, f_2^i, \dots, f_n^i\}$ 等(图2所示为原始语音样本的频谱特征示意图),如“包络线(env)”,“语速(speed)”,“过零率(zcr)”,“能量(eng)”,“能量熵(eoe)”,“频谱质心(spec_cent)”,“频谱扩散(spec_spr)”,“梅尔频率(mfccs)”,“彩度向量(chrona)”等。
- [0043] 步骤4、采用机器学习工具(如Libsvm)对步骤3得到的每个语音文件的特征集与步骤2所得到的语音情感标签进行机器学习,得到每一类语音情感的特征模型,构建情感模型库 E_b 。
- [0044] 步骤5、通过一个多媒体终端,选择需要语音情感迁移目标 $Target_e$,如“悲伤”。
- [0045] 步骤6、从多媒体终端输入语音信号 s_t ,可以是实时输入,也可以是录制完成后点击递交。
- [0046] 步骤7、将当前输入的 s_t 输入到语音情感特征提取模块,得到当前语音信号的特征集 $F_t = \{f_1^t, f_2^t, \dots, f_n^t\}$ 。
- [0047] 步骤8、采用步骤4相同的机器学习算法,将步骤7得到的 s_t 的语音特征集 F_t 结合步骤4得到的情感模型库 E_b 进行情感分类,得到 s_t 的当前情感类别 s_e 。
- [0048] 步骤9、判断步骤8得到的 s_e 和步骤5输入的 $Target_e$ 是否一致,如果 $s_e = Target_e$,则将原始输入语音信号直接作为目标情感语音输出。如果 $s_e \neq Target_e$,则调用步骤10进行特

征情感迁移。

[0049] 步骤10、将当前语音情感主要特征向情感模型库中语音情感主要特征进行迁移(图3所示为迁移后的频谱特征),如包络线迁移 $\text{result}_{\text{env}} = (\text{s}_{\text{env}} + \text{Target}_{\text{env}}) / 2$,语速调整 $\text{result}_{\text{speed}} = (\text{s}_{\text{speed}} + \text{Target}_{\text{speed}}) / 2$ 。

[0050] 步骤11、采用一个语音合成算法(基音同步叠加技术,PSOLA)对步骤10得到的特征迁移过的语音特征进行加工合成最终目标情感语音输出。

[0051] 以上所述仅为本发明的优选实施案例而已,并不用于限制本发明,尽管参照前述实施例对本发明进行了详细的说明,对于本领域的技术人员来说,其依然可以对前述各实施例所记载的技术方案进行改进,或者对其中部分技术进行同等替换。凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

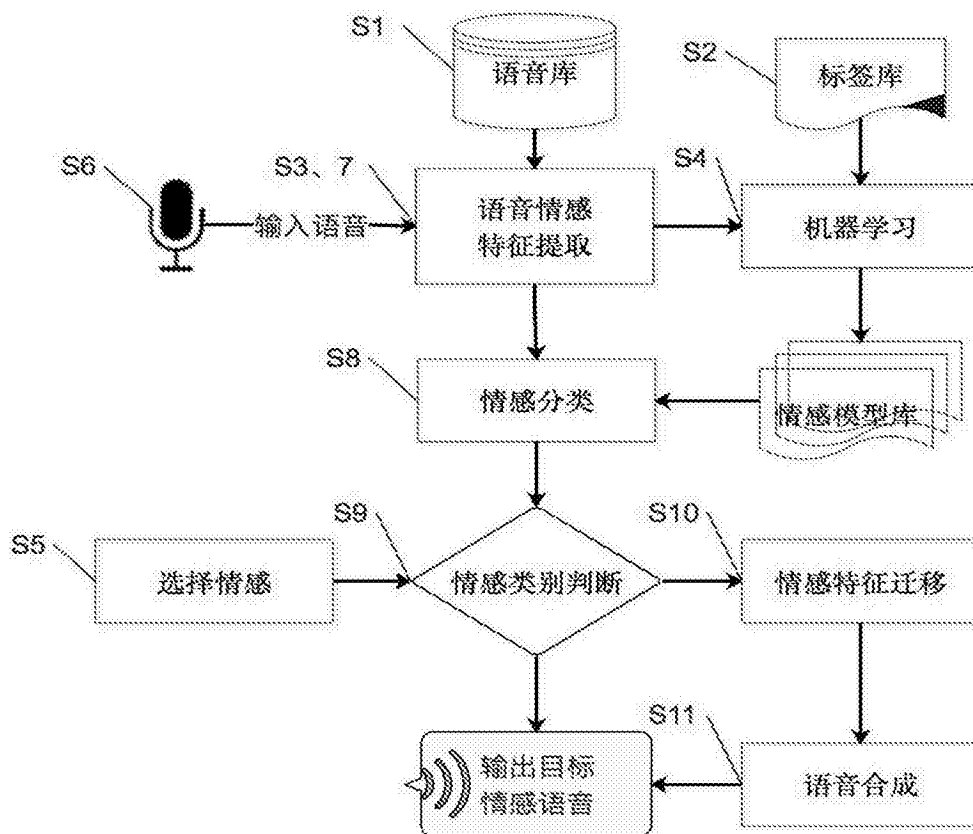


图1

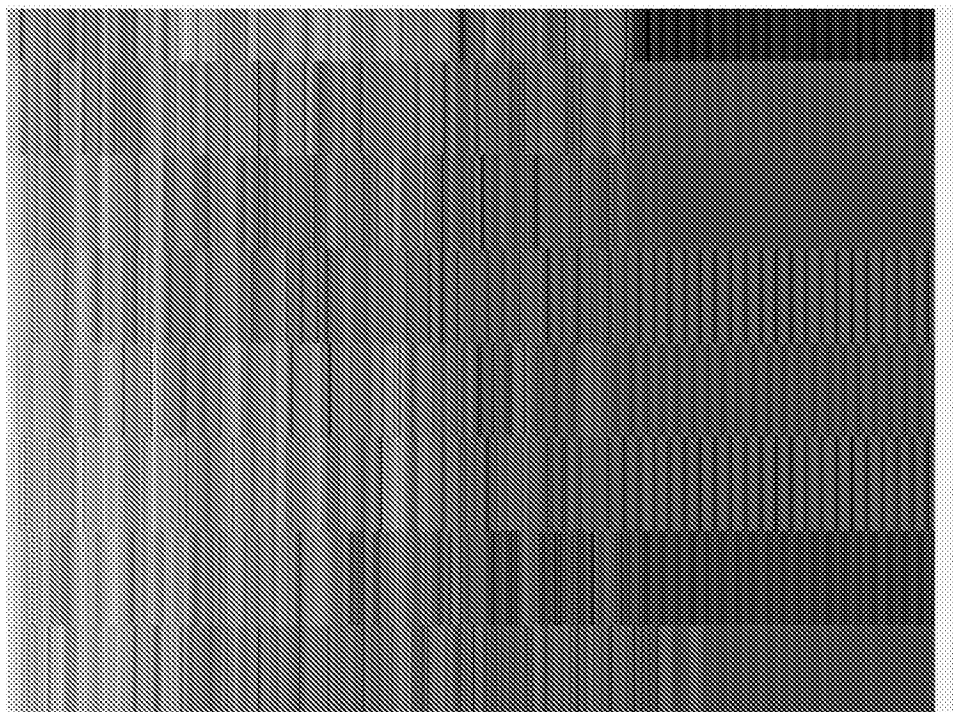


图2

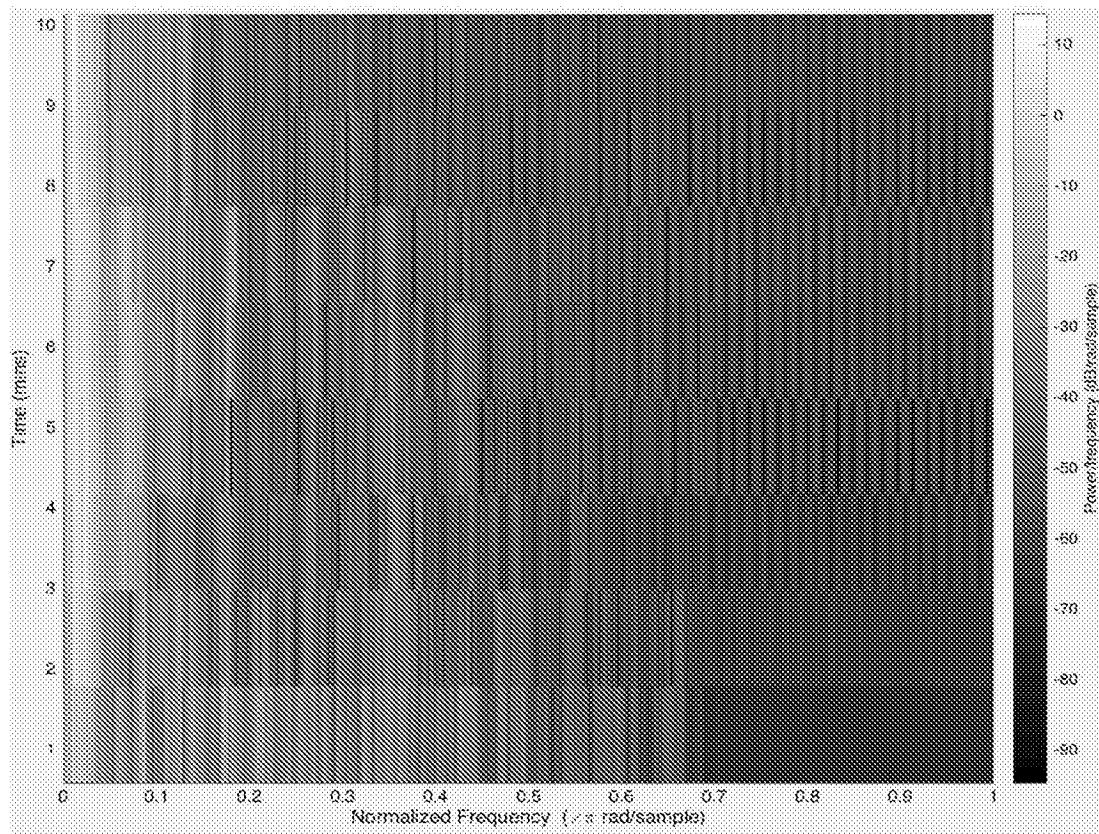


图3