(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107103900 A
(43)申请公布日 2017.08.29

(21)申请号 201710415814.5

(22)申请日 2017.06.06

(71)申请人 西北师范大学
地址 730000 甘肃省兰州市安宁东路967号

(72)发明人 杨鸿武 吴沛文

(74)专利代理机构 北京高沃律师事务所 11569
代理人 王戈

(51)Int.Cl.
G10L 13/02(2013.01)
G10L 13/10(2013.01)
G10L 15/06(2013.01)
G10L 15/18(2013.01)
G10L 15/187(2013.01)
G10L 15/19(2013.01)

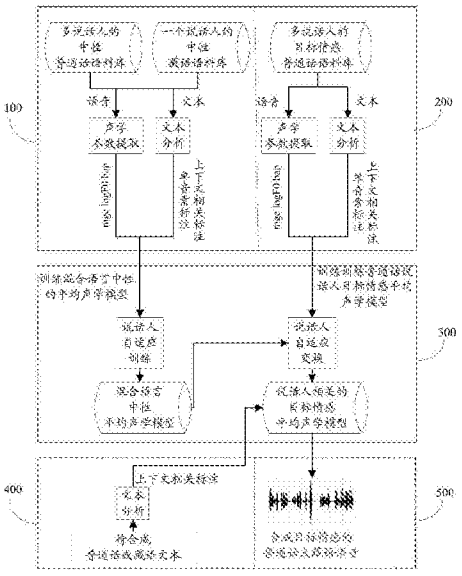
权利要求书4页 说明书12页 附图3页

(54)发明名称

一种跨语言情感语音合成方法及系统

(57)摘要

本发明公开一种跨语言情感语音合成方法及系统，首先，建立上下文相关标注格式和上下文相关聚类问题集；其次，确定第一语言标注文件、第二语言标注文件、目标情感普通话标注文件、待合成标注文件、第一语言声学参数、第二语言声学参数、目标情感声学参数；然后根据所述第一语言标注文件、所述第二语言标注文件、所述目标情感普通话标注文件、所述第一语言声学参数、所述第二语言声学参数和所述目标情感声学参数确定多说话人目标情感平均声学模型；最后，将所述待合成标注文件输入所述多说话人目标情感平均声学模型获得第一语言或/和第二语言目标情感语音合成文件，以实现合成同一说话人或不同说话人跨语言的情感语音。



CN 107103900 A

1.一种跨语言情感语音合成方法，其特征在于，包括以下步骤：

建立上下文相关标注格式和上下文相关聚类问题集，分别对多说话人的中性第一语言训练语料库、单说话人的中性第二语言训练语料库进行上下文相关文本标注，获得所述中性第一语言训练语料库对应的第一语言标注文件、所述中性第二语言训练语料库对应的第二语言标注文件；分别对所述中性第一语言训练语料库和所述中性第二语言训练语料库进行声学参数提取，获得所述中性第一语言训练语料库对应的第一语言声学参数、所述中性第二语言训练语料库对应的第二语言声学参数；

根据所述上下文相关标注格式和所述上下文相关聚类问题集对多说话人的目标情感普通话训练语料库进行上下文相关文本标注，获得目标情感普通话标注文件；对所述目标情感普通话训练语料库进行声学参数提取，获得目标情感声学参数；

根据所述第一语言标注文件、所述第二语言标注文件、所述目标情感普通话标注文件、所述第一语言声学参数、所述第二语言声学参数和所述目标情感声学参数确定多说话人目标情感平均声学模型；

对第一语言或/和第二语言的待合成文件进行上下文相关文本标注获得待合成标注文件；

将所述待合成标注文件输入所述多说话人目标情感平均声学模型获得第一语言或/和第二语言目标情感语音合成文件。

2.根据权利要求1所述的跨语言情感语音合成方法，其特征在于，所述建立上下文相关标注格式和上下文相关聚类问题集，分别对多说话人的中性第一语言训练语料库、单说话人的中性第二语言训练语料库进行上下文相关文本标注，获得所述中性第一语言训练语料库对应的第一语言标注文件、所述中性第二语言训练语料库对应的第二语言标注文件，具体步骤包括：

建立第一语言标注规则和第二语言标注规则；

根据第一语言标注规则和第二语言标注规则确定上下文相关标注格式，分别对多说话人的中性第一语言训练语料库、单说话人的中性第二语言训练语料库进行上下文相关文本标注，获得所述中性第一语言训练语料库对应的第一语言标注文件、所述中性第二语言训练语料库对应的第二语言标注文件；

根据第一语言和第二语言的相似性，建立上下文相关聚类问题集。

3.根据权利要求2所述的跨语言情感语音合成方法，其特征在于，所述建立第一语言标注规则和第二语言标注规则，具体步骤包括：

所述建立第一语言标注规则，具体步骤包括：

将SAMPA-SC普通话机读音标作为所述第一语言标注规则；

所述建立第二语言标注规则，具体步骤包括：

以国际音标为参考，基于SAMPA-SC普通话机读音标，获得输入第二语言拼音的国际音标；

判断所述第二语言拼音的国际音标与第一语言拼音的国际音标是否一致；若一致，则直接采用SAMPA-SC普通话机读音标来标记第二语言拼音；否，则按照简单化原则，利用自定义的未使用的键盘符号标记。

4.根据权利要求3所述的跨语言情感语音合成方法，其特征在于，所述根据第一语言标

注规则和第二语言标注规则确定上下文相关标注格式，具体步骤包括：

根据第一语言和第二语言的语法规则知识库和语法词典，对输入的第一语言和第二语言不规范的文本进行文本规范化、语法分析和韵律结构分析获得规范文本，韵律词、短语的长度信息，韵律边界信息，词语相关信息，声调信息；

将所述规范文本带入所述第一语言标注规则获得第一语言的单音素标注文件；或将所述规范文本带入所述第二语言标注规则获得第二语言的单音素标注文件；

根据韵律词、短语的长度信息，韵律边界信息，词语相关信息，声调信息和单音素标注文件确定上下文相关标注格式。

5.根据权利要求1所述的跨语言情感语音合成方法，其特征在于，所述根据第一语言标注文件、第二语言标注文件、目标情感普通话标注文件、第一语言声学参数、第二语言声学参数和目标情感声学参数确定多说话人目标情感平均声学模型，具体步骤包括：

将第一语言标注文件、第二语言标注文件、第一语言声学参数、第二语言声学参数作为训练集，基于自适应模型，通过说话人自适应训练，获得混合语言的中性平均声学模型；

根据混合语言的中性平均声学模型，将目标情感普通话标注文件、目标情感声学参数作为测试集，通过说话人自适应变换，获得多说话人目标情感平均声学模型。

6.根据权利要求5所述的跨语言情感语音合成方法，其特征在于，所述根据混合语言的中性平均声学模型，将目标情感普通话标注文件、目标情感声学参数作为测试集，通过说话人自适应变换，获得多说话人目标情感普通话说话人目标情感平均声学模型的具体步骤为：

采用约束最大似然线性回归算法，计算说话人的状态时长概率分布和状态输出概率分布的协方差矩阵和均值向量，用一组状态时长分布和状态输出分布的变换矩阵将中性平均声学模型的协方差矩阵和均值向量变换为目标说话人模型，具体公式为：

$$p_i(d)=N(d;\alpha m_i-\beta,\alpha\sigma_i^2\alpha)=|\alpha^{-1}|N(\alpha\psi;m_i,\sigma_i^2) \quad (7)$$

$$b_i(o)=N(o;Au_i-b,A\Sigma_iA^T)=|A^{-1}|N(W\xi;u_i,\Sigma_i) \quad (8)$$

其中，i为状态，d为状态时长，N为常数，$p_i(d)$为状态时长的变换方程，m_i为时长分布均值，σ_i^2为方差，$\psi=[d,1]^T$，o为特征征向量，$\xi=[o^T,1]$，u_i为状态输出分布均值，Σ_i为对角协方差矩阵，$X=[\alpha^{-1},\beta^{-1}]$为状态时长概率密度分布的变换矩阵，$W=[A^{-1},b^{-1}]$为目标说话人状态输出概率密度分布的线性变换矩阵；

通过基于MSD-HSMM的自适应变换算法，可对语音数据的基频、频谱和时长参数进行变换和归一化；对于长度为T的自适应数据O，可变换$\Lambda=(W,X)$进行最大似然估计：

$$\tilde{\Lambda}=(\tilde{W},\tilde{X})=\arg\max_{\Lambda}P(O|\lambda,\Lambda) \quad (9)$$

其中，λ为MSD-HSMM的参数集，O为长度为T的自适应数据，$\tilde{\Lambda}=\arg\max_{\Lambda}P(O|\lambda,\Lambda)$为最大似然估计；

对转化和归一化后的时长、频谱和基频参数进行最大似然估计，采用最大后验概率算法对说话人相关模型进行更新和修正，具体公式为：

$$k_t^d(i)=\frac{1}{P(O|\lambda)}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_{t-d}(j)p(d)\prod_{s=t-d+1}^{t}b_i(o_s)\beta_t(i) \quad (10)$$

MAP估计：

$$\hat{m}_i = \frac{\tau \overline{m}_i + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) d}{\tau + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) d} \tag{11}$$

$$\hat{u}_i = \frac{\omega \overline{u}_i + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) \sum_{s=t-d+1}^{t} o_s}{\omega + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) d} \tag{12}$$

其中，t为时间，λ为给定的MSD-HSMM参数集，T为长度，o为长度为T时自适应数据i为状态，d为状态时长，N为常数，s为训练语音数据模型，$k_t^d(i)$为状态i下连续观测序列$o_{t-d+1}...o_t$的概率，$\alpha_t(i)$为向前概率，$\beta_t(i)$为向后概率，$\overline{m}_i$和$\overline{u}_i$为线性回归变换后的均值向量，ω为状态输出的MAP估计参数，τ为时长分布MAP估计参数，$\hat{m}_i$和$\hat{u}_i$分别为自适应向量$\overline{m}_i$和$\overline{u}_i$的加权平均MAP估计值。

7.一种跨语言情感语音合成系统，其特征在于，包括：

语言语料库文本标注、参数提取模块，用于建立上下文相关标注格式和上下文相关聚类问题集，分别对多说话人的中性第一语言训练语料库、单说话人的中性第二语言训练语料库进行上下文相关文本标注，获得所述中性第一语言训练语料库对应的第一语言标注文件、所述中性第二语言训练语料库对应的第二语言标注文件；用于分别对所述中性第一语言训练语料库和所述中性第二语言训练语料库进行声学参数提取，获得所述中性第一语言训练语料库对应的第一语言声学参数、所述中性第二语言训练语料库对应的第二语言声学参数；

目标情感语料库文本标注、参数提取模块，用于根据所述上下文相关标注格式和所述上下文相关聚类问题集对多说话人的目标情感普通话训练语料库进行上下文相关文本标注，获得目标情感普通话标注文件；对所述目标情感普通话训练语料库进行声学参数提取，获得目标情感声学参数；

目标情感平均声学模型确定模块，用于根据所述第一语言标注文件、所述第二语言标注文件、所述目标情感普通话标注文件、所述第一语言声学参数、所述第二语言声学参数和所述目标情感声学参数确定多说话人目标情感平均声学模型；

待合成标注文件确定模块，用于对第一语言或/和第二语言的待合成文件进行上下文相关文本标注获得待合成标注文件；

语音合成文件确定模块，用于将所述待合成标注文件输入所述多说话人目标情感平均声学模型获得第一语言或/和第二语言目标情感语音合成文件。

8.根据权利要求7所述的跨语言情感语音合成系统，其特征在于，所述语言语料库文本标注模块，具体包括：

标注规则建立子模块，用于建立第一语言标注规则和第二语言标注规则；

语言语料库文本标注子模块，用于根据第一语言标注规则和第二语言标注规则确定上下文相关标注格式，分别对多说话人的中性第一语言训练语料库、单说话人的中性第二语言训练语料库进行上下文相关文本标注，获得所述中性第一语言训练语料库对应的第一语言标注文件、所述中性第二语言训练语料库对应的第二语言标注文件；

标音系统、问题集建立子模块，用于根据第一语言和第二语言的相似性，建立上下文相

关聚类问题集。

9.根据权利要求7所述的跨语言情感语音合成系统，其特征在于，所述目标情感平均声学模型确定模块，具体包括：

混合语言的中性平均声学模型确定子模块，用于将藏语标注文件、汉语标注文件、第一语言声学参数、第二语言声学参数作为训练集，基于自适应模型，通过说话人自适应训练，获得混合语言的中性平均声学模型；

目标情感平均声学模型确定子模块，用于根据混合语言的中性平均声学模型，将目标情感普通话标注文件、目标情感声学参数作为测试集，通过说话人自适应变换，获得多说话人目标情感平均声学模型。

一种跨语言情感语音合成方法及系统

技术领域

[0001] 本发明涉及多语种情感语音合成技术领域，特别是涉及一种跨语言情感语音合成方法及系统。

背景技术

[0002] 目前的语音合成技术，已经能够合成出较自然的中性语音，但当遇到机器人、虚拟助手等这些需要模仿人类行为的人机交互任务时，简单的中性语音合成则不能满足人们的需求。能够模拟表现出人类情感和说话风格的情感语音合成已经成为未来语音合成的发展趋势。

[0003] 对于使用人数众多的大语种汉语、英语等的情感语音合成来说，其研究投入较多，发展水平较高；但对于使用人数较少的小语种如藏语、俄语、西班牙语等情感语音合成来说，其发展却较缓慢，目前还没有一个公认的面向语音合成的高标准、高质量的小语种情感语料库，从而使得小语种情感语音的合成成为了语音合成领域的空白。

[0004] 目前，国内外对情感语音合成的研究技术包括波形拼接方法、韵律单元选择方法和统计参数方法。波形拼接方法需要给情感语音合成系统建立一个庞大的包含每一种情感的情感语料库库，之后对输入的文本进行文本和韵律分析，获得合成语音基本的单元信息，最后根据此单元信息在先前标注好的语料库库中选取合适的语音基元，并进行修改和调整拼接获得目标情感的合成语音，其合成的语音具有较好的情感相似度，但需要提前建立好一个大的、包含各种情感的语音基元语料库库，这在系统的实现中是非常困难的，而且也难以扩展到合成不同说话人、不同语言的情感语音上；韵律特征单元选择方法把韵律或语音体系的策略融入单位选择，用这种规则建立小的或混合的情感语料库库，用于修改目标f0和时长的轮廓，从而获得情感语音。韵律修改方法要对语音信号进行修改，合成语音的音质较差，也不能合成不同人、不同语言的情感语音。以上两种方法由于其局限性，不是现在的主流方法。统计参数语音合成方法虽然成为了主流的语音合成方法，但该方法只能合成出一种语言的情感语音，若需要合成不同语言的情感语音，就需要训练多个情感语音合成系统，每个情感语音合成系统都需要该种语言的情感语音训练语料库。

[0005] 针对上述情感语音合成方法的不足，如何克服上述问题，是目前多语种情感语音合成技术领域急需解决的技术问题。

发明内容

[0006] 本发明的目的是提供一种跨语言情感语音合成方法及系统，以实现用一种多说话人的目标情感普通话训练语料库训练一个普通话说话人目标情感平均声学模型，只需改变待合成文件就能合成同一说话人或不同说话人跨语言的情感语音。

[0007] 为实现上述目的，本发明提供了一种跨语言情感语音合成方法，包括以下步骤：

[0008] 建立上下文相关标注格式和上下文相关聚类问题集，分别对多说话人的中性第一语言训练语料库、单说话人的中性第二语言训练语料库进行上下文相关文本标注，获得所

述中性第一语言训练语料库对应的第一语言标注文件、所述中性第二语言训练语料库对应的第二语言标注文件；分别对所述中性第一语言训练语料库和所述中性第二语言训练语料库进行声学参数提取，获得所述中性第一语言训练语料库对应的第一语言声学参数、所述中性第二语言训练语料库对应的第二语言声学参数；

[0009] 根据所述上下文相关标注格式和所述上下文相关聚类问题集对多说话人的目标情感普通话训练语料库进行上下文相关文本标注，获得目标情感普通话标注文件；对所述目标情感普通话训练语料库进行声学参数提取，获得目标情感声学参数；

[0010] 根据所述第一语言标注文件、所述第二语言标注文件、所述目标情感普通话标注文件、所述第一语言声学参数、所述第二语言声学参数和所述目标情感声学参数确定多说话人目标情感平均声学模型；

[0011] 对第一语言或/和第二语言的待合成文件进行上下文相关文本标注获得待合成标注文件；

[0012] 将所述待合成标注文件输入所述多说话人目标情感平均声学模型获得第一语言或/和第二语言目标情感语音合成文件。

[0013] 可选的，所述建立上下文相关标注格式和上下文相关聚类问题集，分别对多说话人的中性第一语言训练语料库、单说话人的中性第二语言训练语料库进行上下文相关文本标注，获得所述中性第一语言训练语料库对应的第一语言标注文件、所述中性第二语言训练语料库对应的第二语言标注文件，具体步骤包括：

[0014] 建立第一语言标注规则和第二语言标注规则；

[0015] 根据第一语言标注规则和第二语言标注规则确定上下文相关标注格式，分别对多说话人的中性第一语言训练语料库、单说话人的中性第二语言训练语料库进行上下文相关文本标注，获得所述中性第一语言训练语料库对应的第一语言标注文件、所述中性第二语言训练语料库对应的第二语言标注文件；

[0016] 根据第一语言和第二语言的相似性，建立上下文相关聚类问题集。

[0017] 可选的，所述建立第一语言标注规则和第二语言标注规则，具体步骤包括：

[0018] 所述建立第一语言标注规则，具体步骤包括：

[0019] 将SAMPA-SC普通话机读音标作为所述第一语言标注规则；

[0020] 所述建立第二语言标注规则，具体步骤包括：

[0021] 以国际音标为参考，基于SAMPA-SC普通话机读音标，获得输入第二语言拼音的国际音标；

[0022] 判断所述第二语言拼音的国际音标与第一语言拼音的国际音标是否一致；若一致，则直接采用SAMPA-SC普通话机读音标来标记第二语言拼音；否，则按照简单化原则，利用自定义的未使用的键盘符号标记。

[0023] 可选的，所述根据第一语言标注规则和第二语言标注规则确定上下文相关标注格式，具体步骤包括：

[0024] 根据第一语言和第二语言的语法规则知识库和语法词典，对输入的第一语言和第二语言不规范的文本进行文本规范化、语法分析和韵律结构分析获得规范文本，韵律词、短语的长度信息，韵律边界信息，词语相关信息，声调信息；

[0025] 将所述规范文本带入所述第一语言标注规则获得第一语言的单音素标注文件；或

将所述规范文本带入所述第二语言标注规则获得第二语言的单音素标注文件；

[0026] 根据韵律词、短语的长度信息，韵律边界信息，词语相关信息，声调信息和单音素标注文件确定上下文相关标注格式。

[0027] 可选的，所述根据第一语言标注文件、第二语言标注文件、目标情感普通话标注文件、第一语言声学参数、第二语言声学参数和目标情感声学参数确定多说话人目标情感平均声学模型，具体步骤包括：

[0028] 将第一语言标注文件、第二语言标注文件、第一语言声学参数、第二语言声学参数作为训练集，基于自适应模型，通过说话人自适应训练，获得混合语言的中性平均声学模型；

[0029] 根据混合语言的中性平均声学模型，将目标情感普通话标注文件、目标情感声学参数作为测试集，通过说话人自适应变换，获得多说话人目标情感平均声学模型。

[0030] 可选的，所述根据混合语言的中性平均声学模型，将目标情感普通话标注文件、目标情感声学参数作为测试集，通过说话人自适应变换，获得多说话人目标情感普通话说话人目标情感平均声学模型的具体步骤为：

[0031] 采用约束最大似然线性回归算法，计算说话人的状态时长概率分布和状态输出概率分布的协方差矩阵和均值向量，用一组状态时长分布和状态输出分布的变换矩阵将中性平均声学模型的协方差矩阵和均值向量变换为目标说话人模型，具体公式为：

[0032] $p_i(d)=N(d;\alpha m_i-\beta,\alpha\sigma_i^2\alpha)=|\alpha^{-1}|N(\alpha\psi;m_i,\sigma_i^2)$ (7)；

[0033] $b_i(o)=N(o;Au_i-b,A\Sigma_iA^T)=|A^{-1}|N(W\xi;u_i,\Sigma_i)$ (8)；

[0034] 其中，i为状态，d为状态时长，N为常数，$p_i(d)$为状态时长的变换方程，m_i为时长分布均值，σ_i^2为方差，$\psi=[d,1]^T$，o为特征征向量，$\xi=[o^T,1]$，u_i为状态输出分布均值，Σ_i为对角协方差矩阵，$X=[\alpha^{-1},\beta^{-1}]$为状态时长概率密度分布的变换矩阵，$W=[A^{-1},b^{-1}]$为目标说话人状态输出概率密度分布的线性变换矩阵；

[0035] 通过基于MSD-HSMM的自适应变换算法，可对语音数据的基频、频谱和时长参数进行变换和归一化；对于长度为T的自适应数据O，可变换$\Lambda=(W,X)$进行最大似然估计：

[0036]

$$\tilde{\Lambda}=(\tilde{W},\tilde{X})=\arg\max_{\Lambda}P(O|\lambda,\Lambda) \quad (9);$$

[0037] 其中，λ为MSD-HSMM的参数集，O为长度为T的自适应数据，$\tilde{\Lambda}=\arg\max_{\Lambda}P(O|\lambda,\Lambda)$为最大似然估计；

[0038] 对转化和归一化后的时长、频谱和基频参数进行最大似然估计，采用最大后验概率算法对说话人相关模型进行更新和修正，具体公式为：

[0039]

$$k_t^d(i)=\frac{1}{P(O|\lambda)}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_{t-d}(j)p(d)\prod_{s=t-d+1}^{t}b_i(o_s)\beta_t(i) \quad (10);$$

[0040] MAP估计：

[0041]

$$\hat{m}_i = \frac{\tau \overline{m}_i + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) d}{\tau + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) d} \tag{11};$$

[0042]

$$\hat{u}_i = \frac{\omega \overline{u}_i + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) \sum_{s=t-d+1}^{t} o_s}{\omega + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) d} \tag{12};$$

[0043] 其中，t为时间，λ为给定的MSD-HSMM参数集，T为长度，o为长度为T时自适应数据i为状态，d为状态时长，N为常数，s为训练语音数据模型，$k_t^d(i)$为状态i下连续观测序列$o_{t-d+1}...o_t$的概率，$\alpha_t(i)$为向前概率，$\beta_t(i)$为向后概率，$\overline{m}_i$和$\overline{u}_i$为线性回归变换后的均值向量，ω为状态输出的MAP估计参数，τ为时长分布MAP估计参数，$\hat{m}_i$和$\hat{u}_i$分别为自适应向量$\overline{m}_i$和$\overline{u}_i$的加权平均MAP估计值。

[0044] 本发明还提供了一种跨语言情感语音合成系统，所述系统包括：

[0045] 语言语料库文本标注、参数提取模块，用于建立上下文相关标注格式和上下文相关聚类问题集，分别对多说话人的中性第一语言训练语料库、单说话人的中性第二语言训练语料库进行上下文相关文本标注，获得所述中性第一语言训练语料库对应的第一语言标注文件、所述中性第二语言训练语料库对应的第二语言标注文件；分别对中性第一语言训练语料库和中性第二语言训练语料库进行声学参数提取，获得所述中性第一语言训练语料库对应的第一语言声学参数、所述中性第二语言训练语料库对应的第二语言声学参数；

[0046] 目标情感语料库文本标注、参数提取模块，用于根据上下文相关标注格式和上下文相关聚类问题集对多说话人的目标情感普通话训练语料库进行上下文相关文本标注，获得目标情感普通话标注文件；对所述目标情感普通话训练语料库进行声学参数提取，获得目标情感声学参数；

[0047] 目标情感平均声学模型确定模块，用于根据所述第一语言标注文件、所述第二语言标注文件、所述目标情感普通话标注文件、所述第一语言声学参数、所述第二语言声学参数和所述目标情感声学参数确定多说话人目标情感平均声学模型；

[0048] 待合成标注文件确定模块，用于对第一语言或/和第二语言的待合成文件进行上下文相关文本标注获得待合成标注文件；

[0049] 语音合成文件确定模块，用于将所述待合成标注文件输入所述多说话人目标情感平均声学模型获得第一语言或/和第二语言目标情感语音合成文件。

[0050] 可选的，所述语言语料库文本标注模块，具体包括：

[0051] 标注规则建立子模块，用于建立第一语言标注规则和第二语言标注规则；

[0052] 语言语料库文本标注子模块，用于根据第一语言标注规则和第二语言标注规则确定上下文相关标注格式，分别对多说话人的中性第一语言训练语料库、单说话人的中性第二语言训练语料库进行上下文相关文本标注，获得所述中性第一语言训练语料库对应的第一语言标注文件、所述中性第二语言训练语料库对应的第二语言标注文件；

[0053] 标音系统、问题集建立子模块，用于根据第一语言和第二语言的相似性，建立上下文相关聚类问题集。

[0054] 可选的，所述目标情感平均声学模型确定模块，具体包括：

[0055] 混合语言的中性平均声学模型确定子模块，用于将藏语标注文件、汉语标注文件、第一语言声学参数、第二语言声学参数作为训练集，基于自适应模型，通过说话人自适应训练，获得混合语言的中性平均声学模型；

[0056] 目标情感平均声学模型确定子模块，用于根据混合语言的中性平均声学模型，将目标情感普通话标注文件、目标情感声学参数作为测试集，通过说话人自适应变换，获得多说话人目标情感平均声学模型。

[0057] 根据本发明提供的具体实施例，本发明公开了以下技术效果：

[0058] 1)、本发明利用一种多说话人的目标情感普通话训练语料库就能训练出一种多说话人目标情感平均声学模型，只需改变待合成文件就能合成出另一种语言或多种语言的情感语音合成，从而拓宽了语音合成范围。

[0059] 2)、本发明利用一种多说话人的目标情感普通话训练语料库就能训练出一种多说话人目标情感平均声学模型，既能合成出同一个说话人不同语言的情感语音，还能合成出不同说话人说不同语言的情感语音。

附图说明

[0060] 为了更清楚地说明本发明实施例或现有技术中的技术规则，下面将对实施例中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0061] 图1为本发明实施例跨语言情感语音合成方法流程图；

[0062] 图2为本发明实施例藏语标注规则的具体流程图；

[0063] 图3为本发明实施例建立上下文相关标注格式的具体流程图；

[0064] 图4为本发明实施例声学参数提取的具体流程图；

[0065] 图5为本发明实施例跨语言情感语音合成系统结构框图。

具体实施方式

[0066] 下面将结合本发明实施例中的附图，对本发明实施例中的技术规则进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0067] 本发明的目的是提供一种跨语言情感语音合成方法及系统，以实现用一种多说话人的目标情感普通话训练语料库训练一个普通话说话人目标情感平均声学模型，只需改变待合成文件就能合成同一说话人或不同说话人跨语言的情感语音。

[0068] 为使本发明的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0069] 本发明公开了第一语言和第二语言，所述第一语言为汉语、英语、德语、法语中任意一种；所述第二语言为藏语、西班牙语、日语、阿拉伯语、韩语、葡萄牙语中任意一种。本发明具体实施例将汉语作为第一语言，将藏语作为第二语言为例进行论述，图1为本发明实施例跨语言情感语音合成方法流程图，具体详见图1。

[0070] 本发明具体提供了一种跨语言情感语音合成方法，具体步骤包括：

[0071] 步骤100：建立汉语和藏语通用的上下文相关标注格式和上下文相关聚类问题集，分别对多说话人的中性汉语训练语料库、单说话人的中性藏语训练语料库进行上下文相关文本标注，获得所述中性汉语训练语料库对应的汉语标注文件、所述中性藏语训练语料库对应的藏语标注文件；分别对所述中性汉语训练语料库和所述中性藏语训练语料库进行声学参数提取，获得所述中性汉语训练语料库对应的汉语声学参数、所述中性藏语训练语料库对应的藏语声学参数。

[0072] 步骤200：根据所述上下文相关标注格式和所述上下文相关聚类问题集对多说话人的目标情感普通话训练语料库进行上下文相关文本标注，获得目标情感普通话标注文件；对所述目标情感普通话训练语料库进行声学参数提取，获得目标情感声学参数。

[0073] 步骤300：根据所述汉语标注文件、所述藏语标注文件、所述目标情感普通话标注文件、所述汉语声学参数、所述藏语声学参数和所述目标情感声学参数确定多说话人目标情感平均声学模型。

[0074] 步骤400：对汉语或/和藏语的待合成文件进行上下文相关文本标注获得待合成标注文件。

[0075] 步骤500：将所述待合成标注文件输入所述多说话人目标情感平均声学模型获得汉语或/和藏语目标情感语音合成文件。

[0076] 下面对各个步骤进行详细的介绍：

[0077] 步骤100：建立汉语和藏语通用的上下文相关标注格式和上下文相关聚类问题集，分别对多说话人的中性汉语训练语料库、单说话人的中性藏语训练语料库进行上下文相关文本标注，获得所述中性汉语训练语料库对应的汉语标注文件、所述中性藏语训练语料库对应的藏语标注文件；分别对所述中性汉语训练语料库和所述中性藏语训练语料库进行声学参数提取，获得所述中性汉语训练语料库对应的汉语声学参数、所述中性藏语训练语料库对应的藏语声学参数。

[0078] 步骤101：建立汉语标注规则和藏语标注规则。

[0079] 步骤1011：将SAMPA-SC普通话机读音标作为所述汉语标注规则。

[0080] 步骤1012：所述建立藏语标注规则，具体步骤包括：

[0081] 目前汉语普通话机读音标SAMPA-SC已趋于成熟并广泛应用，而藏语和汉语在发音上有很多相似之处，例如，汉藏语系中，汉语与藏语在发音上既有共性又有差异，藏语拉萨方言和汉语普通话都是由音节组成，每个音节都包含1个韵母和1个声母，藏语拉萨方言有45个韵母和36个声母，普通话有39个韵母和22个声母，它们共享13个韵母和20个声母，且都有4个声调只是调值不同。因此本发明以SAMPA-SC为基础，根据藏语的发音特点，设计出一套藏语计算机可读音标SAMPA-T，即藏语标注规则。具体详见图2。

[0082] 以国际音标为参考，基于SAMPA-SC普通话机读音标，获得输入藏语拼音的国际音标。

[0083] 判断所述藏语拼音的国际音标与汉语拼音的国际音标是否一致，若一致，则直接采用SAMPA-SC普通话机读音标来标记藏语拼音，否，则按照简单化原则，利用自定义的未使用的键盘符号标记。

[0084] 步骤102：根据汉语标注规则和藏语标注规则确定汉语和藏语通用的上下文相关

标注格式，根据上下文相关标注格式分别对多说话人的中性汉语训练语料库、单说话人的中性藏语训练语料库进行上下文相关文本标注，分别获得所述中性汉语训练语料库对应的汉语标注文件、所述中性藏语训练语料库对应的藏语标注文件，具体详见图3。

[0085] 步骤1021：根据汉语和藏语的语法规则知识库和语法词典，对输入的汉语和藏语不规范的文本进行文本规范化、语法分析和韵律结构分析获得规范文本，韵律词、短语的长度信息，韵律边界信息，词语相关信息，声调信息。

[0086] 步骤1022：将所述规范文本带入所述汉语标注规则获得汉语的单音素标注文件；或将所述规范文本带入所述藏语标注规则获得藏语的单音素标注文件。

[0087] 步骤1023：根据韵律词、短语的长度信息，韵律边界信息，词语相关信息，声调信息和单音素标注文件确定汉语和藏语通用的上下文相关标注格式。

[0088] 上下文相关标注格式用来标注发音基元(声韵母)的上下文信息。上下文相关标注格式包括声韵母音、音节、词、韵律词、韵律短语和语句6层，用来表示发音基元(声韵母)及其在不同语境下的上下文相关信息。

[0089] 步骤1024：根据上下文相关标注格式分别对多说话人的中性汉语训练语料库、单说话人的中性藏语训练语料库进行上下文相关文本标注，分别获得所述中性汉语训练语料库对应的汉语标注文件、所述中性藏语训练语料库对应的藏语标注文件。

[0090] 步骤103：根据汉语和藏语的相似性，建立汉语和藏语通用的上下文相关聚类问题集。

[0091] 步骤104：分别对中性汉语训练语料库和中性藏语训练语料库进行声学参数提取，获得所述中性汉语训练语料库对应的汉语声学参数、所述中性藏语训练语料库对应的藏语声学参数，具体详见图4。

[0092] 声学参数提取时，通过对语音信号进行分析，提取语音信号的基频和谱特征等声学特征。本发明中用广义梅尔倒谱系数(Mel-generalized cepstral，mgc)作为谱特征，用来表示频谱包络，即：源滤波器模型中的滤波器部；用对数基频logF0作为基频特征。因为语音信号不是纯粹的、稳定的周期信号，基频的错误直接影响对频谱包络的提取，因此，提取频谱包络(广义梅尔倒谱系数mgc)同时也要提取基频特征(对数基频logF0)。

[0093] 所述声学参数提取包括：广义梅尔倒谱系数mgc提取，对数基频logF0提取，非周期分量bap提取。

[0094] 广义梅尔倒谱系数mgc提取公式具体为：

[0095]

$$H(z)=\begin{cases}\left(1+\gamma\sum_{m=0}^{M}c_{\alpha,\gamma}(m)z_{\alpha}^{-m}\right)^{1/\gamma}, & -1\le\gamma<0\\ \exp\sum_{m=0}^{M}c_{\alpha,\gamma}(m)z_{\alpha}^{-m}, & \gamma=0\end{cases} \quad (1);$$

[0096] 其中，$z_{\alpha}^{-m}=\frac{z^{-m}-\alpha}{1-\alpha z^{-m}}$ ($|\alpha|<1$)为m阶全通函数，γ 为系统函数的属性，$c_{\alpha,\gamma}(m)$ 为系数，M为滤波器系数总个数，z为离散信号的z变换，m为滤波器系数阶数。

[0097] 如果 $\gamma=0$，$c_{\alpha,\gamma}(m)$ 为mgc模型；γ 等于-1，则该模型为自回归模型；如果 γ 等于0，则为指数模型。

[0098] 对数基频logF0提取：

[0099] 采用归一化自相关函数法提取基频特征，其具体步骤为：

[0100] 对于语音信号s(n)，n≤N,n∈N+，其自相关函数为：

[0101]

$$acf(k)=\sum_{n=0}^{N-K}s(n)s(n+k),0\le k\le K-1, \tag{2};$$

[0102] 其中，k为延时时间，应设置为基音周期的整数倍，s(n+k)为s(n)相邻的语音信号，N整数，K为延时时间的最大数。

[0103] 对自相关函数acf(k)进行归一化处理，便得到归一化自相关函数：

[0104]

$$nccf(k)=\frac{\sum_{n=0}^{N-K}s(n)s(n+k)}{\sqrt{e_0e_k}} \tag{3};$$

[0105] 其中，$e_k=\sum_{n=k}^{n-k+N-K}s^2(n)$，e0为0时刻的ek。

[0106] 当自相关函数的最大值时，函数的延迟值k即为基音周期。基音周期取倒数就是基频，基频对数就是需要提取的对数基频logF0。

[0107] 非周期分量bap提取：

[0108] 语音信号的非周期成分在频域被定义为非周期成分的相对能量水平，并通过非谐波成分的能量与固定基频值结构规整后的谱的总能量的比值计算线性域的非周期成分值ap，也就是说用上下谱包络相减就能确定线性域的非周期成分值ap，具体公式为：

[0109]

$$P_{AP}(\omega')=\frac{\int w_{ERB}(\lambda';\omega')\left|S(\lambda')\right|^2\Gamma(\frac{\left|S_L(\lambda')\right|^2}{\left|S_U(\lambda')\right|^2})\mathrm{d}\lambda'}{\int w_{ERB}(\lambda';\omega')\left|S(\lambda')\right|^2\mathrm{d}\lambda'} \tag{4};$$

[0110] PAP(ω′)为lg域非周期成分值；S(λ′)代表谱能量，SL(λ′)表示谱下包络的谱能量，SU(λ′)为谱上包络的谱能量；wERB(λ′；ω′)为平滑声学滤波器，λ′为基频，ω′为频率。

[0111] 在每帧的每个频带内对ap求取平均值就能确定非周期分量bap，具体公式为：

[0112] $bap(\omega')=10^{\frac{P_{AP}(\omega')}{20}}$

[0113] 其中，bap(ω′)为非周期分量bap。

[0114] 步骤200：根据上下文相关标注格式和上下文相关聚类问题集对多说话人的目标情感普通话训练语料库进行上下文相关文本标注，获得目标情感普通话标注文件；对所述目标情感普通话训练语料库进行声学参数提取，获得目标情感声学参数。

[0115] 对所述目标情感普通话训练语料库进行声学参数提取与对中性汉语训练语料库和中性藏语训练语料库进行声学参数提取的声学参数提取方式相同。具体详见公式(1)-(4)。

[0116] 步骤300：根据所述汉语标注文件、所述藏语标注文件、所述目标情感普通话标注文件、所述汉语声学参数、所述藏语声学参数和所述目标情感声学参数确定多说话人目标

情感平均声学模型。

[0117] 步骤301:将汉语标注文件、藏语标注文件、汉语声学参数、藏语声学参数作为训练集,基于自适应模型,通过说话人自适应训练,获得混合语言的中性平均声学模型。所述自适应模型为深度学习模型、长短时记忆模型、隐马尔科夫模型中的任意一种。本发明采用半隐马尔科夫模型进行分析。

[0118] 本发明采用约束最大似然线性回归算法,将平均声学模型和训练中说话人的语音数据之间的差异用线性回归函数表示,用一组状态时长分布和状态输出分布的线性回归公式归一化训练说话人之间的差异,训练得到上下文相关的半隐马尔科夫模型(Multi-Space Hidden semi-Markov models,MSD-HSMM)。采用基于半隐马尔科夫模型MSD-HSMM的说话人自适应训练算法来提高合成语音的音质,减少各说话人之间的差异对合成语音质量的影响。状态时常分布和状态输出分布的线性回归公式具体为:

[0119]

$$\hat{d}_i^{(s)} = \alpha^{(s)} d_i + \beta^{(s)} = X^{(s)} \xi_{(i)} \tag{5};$$

[0120]

$$\hat{o}_i^{(s)} = A^{(s)} o_i + b^{(s)} = W^{(s)} \xi_{(i)} \tag{6};$$

[0121] 其中,公式(5)所示为状态时长分布变换方程,i为状态,右下角的i表示在状态i下,s为训练语音数据模型,s标记在右上角表示属于语音数据模型s的,$\hat{d}_i$表示训练语音数据模型s的状态时长的均值向量。X=[α,β]为训练语音数据模型s的状态时长分布与平均音模型之间差异的变换矩阵,d_i为其平均时长,其中,ξ=[o^T,1]。公式(6)所示为状态输出分布变换方程,$\hat{o}_i$表示训练语音数据模型s的状态输出的均值向量,W=[A,b]为训练语音数据模型s的状态输出分布与平均音模型之间差异的变换矩阵,o_i为其平均观测向量。

[0122] 步骤302:根据混合语言的中性平均声学模型,将目标情感普通话标注文件、目标情感声学参数作为测试集,通过说话人自适应变换,获得多说话人目标情感平均声学模型;其具体步骤为:

[0123] 步骤3021:采用约束最大似然线性回归算法,计算说话人的状态时长概率分布和状态输出概率分布的协方差矩阵和均值向量,用一组状态时长分布和状态输出分布的变换矩阵将中性平均声学模型的协方差矩阵和均值向量变换为目标说话人模型,具体公式为:

[0124] $p_i(d) = N(d;\alpha m_i - \beta, \alpha\sigma_i^2\alpha) = |\alpha^{-1}| N(\alpha\psi; m_i, \sigma_i^2)$ (7)

[0125] $b_i(o) = N(o;Au_i - b, A\Sigma_i A^T) = |A^{-1}| N(W\xi; u_i, \Sigma_i)$ (8)

[0126] 其中,i为状态,d为状态时长,N为常数,$p_i(d)$为状态时长的变换方程,m_i为时长分布均值,σ_i^2为方差,ψ=[d,1]^T,o为特征征向量,ξ=[o^T,1],u_i为状态输出分布均值,Σ_i为对角协方差矩阵,X=[α^{-1},β^{-1}]为状态时长概率密度分布的变换矩阵,W=[A^{-1},b^{-1}]为目标说话人状态输出概率密度分布的线性变换矩阵;

[0127] 步骤3022:通过基于MSD-HSMM的自适应变换算法,可对语音数据的基频、频谱和时长参数进行变换和归一化;对于长度为T的自适应数据O,可变换Λ=(W,X)进行最大似然估计:

[0128]

$$\tilde{\Lambda} = (\tilde{W}, \tilde{X}) = \arg\max_{\Lambda} P(O|\lambda, \Lambda) \tag{9}$$

[0129] 其中，λ为MSD-HSMM的参数集，O为长度为T的自适应数据，$\tilde{\Lambda} = \arg\max_{\Lambda} P(O|\lambda, \Lambda)$为最大似然估计。

[0130] 步骤3023：对转化和归一化后的时长、频谱和基频参数进行最大似然估计，采用最大后验概率算法对说话人相关模型进行更新和修正，具体公式为：

[0131]

$$k_t^d(i) = \frac{1}{P(O|\lambda)} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_{t-d}(j) p(d) \prod_{s=t-d+1}^{t} b_i(o_s) \beta_i(i) \tag{10}$$

[0132] MAP估计：

[0133]

$$\hat{m}_i = \frac{\tau \overline{m}_i + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) d}{\tau + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) d} \tag{11}$$

[0134]

$$\hat{u}_i = \frac{\omega \overline{u}_i + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) \sum_{s=t-d+1}^{t} o_s}{\omega + \sum_{t=1}^{T} \sum_{d=1}^{t} k_t^d(i) d} \tag{12}$$

[0135] 其中，t为时间，λ为给定的MSD-HSMM参数集，T为长度，o为长度为T时自适应数据i为状态，d为状态时长，N为常数，s为训练语音数据模型，$k_t^d(i)$为状态i下连续观测序列$o_{t-d+1} \ldots o_t$的概率，$\alpha_t(i)$为向前概率，$\beta_t(i)$为向后概率，$\overline{m}_i$和$\overline{u}_i$为线性回归变换后的均值向量，ω为状态输出的最大后验概率(Maximum aposteriori，MAP)估计参数，τ为时长分布MAP估计参数，$\hat{m}_i$和$\hat{u}_i$分别为自适应向量$\overline{m}_i$和$\overline{u}_i$的加权平均MAP估计值。

[0136] 步骤400：对汉语或/和藏语的待合成文件进行上下文相关文本标注获得待合成标注文件。

[0137] 所述待合成文件包括汉语和/藏语待合成文件，待合成文件为字、词、短语、句子任意一种，将所述汉语和/藏语待合成文件根据所述上下文相关文本标注格式进行上下文相关文本标注获得待合成标注文件。

[0138] 也就是说，当待合成文本为藏语待合成文本时，根据所述上下文相关文本标注格式进行上下文相关文本标注获得藏语待合成标注文件；当待合成文本为汉语待合成文本时，根据所述上下文相关文本标注格式进行上下文相关文本标注获得汉语待合成标注文件；当待合成文本为藏语和汉语待合成文本时，根据所述上下文相关文本标注格式进行上下文相关文本标注获得藏语和汉语待合成标注文件。

[0139] 步骤500：将所述待合成标注文件输入所述多说话人目标情感平均声学模型获得目标情感语音合成文件。

[0140] 对于待合成文本的待合成标注文件，利用问题集，根据每个发音基元(声韵母)的上下文相关信息获得每个发音基元的说话人相关的目标情感平均声学模型，再通过聚类确定整个待合成句子的说话人相关的目标情感平均声学模型，然后根据此说话人相关的目标

情感平均声学模型获得普通话和/或藏语的目标情感的声学参数文件，最后利用声学参数文件通过语音波形生成器来合成出藏语和/或汉语目标情感语音合成文件。

[0141] 也就是说，将所述藏语待合成标注文件输入所述多说话人目标情感平均声学模型获得藏语目标情感语音合成文件；将所述汉语待合成标注文件输入所述多说话人目标情感平均声学模型获得汉语目标情感语音合成文件；将所述汉语和藏语待合成标注文件输入所述多说话人目标情感平均声学模型获得汉语和藏语混合目标情感语音合成文件。

[0142] 为实现上述目的，本发明还提供了一种跨语言情感语音合成系统。

[0143] 图5为本发明实施例跨语言情感语音合成系统结构框图，如图5所示，所述系统包括：语言语料库文本标注、参数提取模块1，目标情感语料库文本标注、参数提取模块2，目标情感平均声学模型确定模块3，待合成标注文件确定模块4，语音合成文件确定模块5。

[0144] 语言语料库文本标注、参数提取模块1，用于建立上下文相关标注格式和上下文相关聚类问题集，分别对多说话人的中性汉语训练语料库、单说话人的中性藏语训练语料库进行上下文相关文本标注，获得所述中性汉语训练语料库对应的汉语标注文件、所述中性藏语训练语料库对应的藏语标注文件；用于分别对中性汉语训练语料库和中性藏语训练语料库进行声学参数提取，获得所述中性汉语训练语料库对应的汉语声学参数、所述中性藏语训练语料库对应的藏语声学参数。

[0145] 所述语言语料库文本标注、参数提取模块1具体包括：所述语言语料库文本标注模块和所述语言语料库参数提取模块。

[0146] 所述语言语料库文本标注模块，具体包括：标注规则建立子模块，语言语料库文本标注子模块，标音系统、问题集建立子模块。

[0147] 所述标注规则建立子模块，用于建立汉语标注规则和藏语标注规则；

[0148] 所述语言语料库文本标注子模块，用于根据汉语标注规则和藏语标注规则确定上下文相关标注格式，分别对多说话人的中性汉语训练语料库、单说话人的中性藏语训练语料库进行上下文相关文本标注，获得所述中性汉语训练语料库对应的汉语标注文件、所述中性藏语训练语料库对应的藏语标注文件；

[0149] 所述标音系统、问题集建立子模块，用于根据汉语和藏语的相似性，建立汉语和藏语通用的上下文相关聚类问题集。

[0150] 所述语言语料库参数提取模块，用于分别对中性汉语训练语料库和中性藏语训练语料库进行声学参数提取，获得所述中性汉语训练语料库对应的汉语声学参数、所述中性藏语训练语料库对应的藏语声学参数。

[0151] 目标情感语料库文本标注、参数提取模块2，用于对多说话人的目标情感普通话训练语料库进行上下文相关文本标注，获得目标情感普通话标注文件；对所述目标情感普通话训练语料库进行声学参数提取，获得目标情感声学参数；

[0152] 目标情感平均声学模型确定模块3，用于根据所述汉语标注文件、所述藏语标注文件、所述目标情感普通话标注文件、所述汉语声学参数、所述藏语声学参数和所述目标情感声学参数确定多说话人目标情感平均声学模型；

[0153] 所述目标情感平均声学模型确定模块3，具体包括：混合语言的中性平均声学模型确定子模块、目标情感平均声学模型确定子模块。

[0154] 所述混合语言的中性平均声学模型确定子模块，用于将藏语标注文件、汉语标注

文件、汉语声学参数、藏语声学参数作为训练集，基于自适应模型，通过说话人自适应训练，获得混合语言的中性平均声学模型；

[0155] 所述目标情感平均声学模型确定子模块，用于根据混合语言的中性平均声学模型，将目标情感普通话标注文件、目标情感声学参数作为测试集，通过说话人自适应变换，获得多说话人目标情感平均声学模型。

[0156] 待合成标注文件确定模块4，用于对汉语或/和藏语的待合成文件进行上下文相关文本标注获得待合成标注文件。

[0157] 语音合成文件确定模块5，用于将所述待合成标注文件输入所述多说话人目标情感平均声学模型获得汉语或/和藏语目标情感语音合成文件。

[0158] 具体举例：

[0159] 本发明录制一个女性藏语说话人的800句作为单说话人的中性藏语训练语料库，将汉英双语语音数据库作为多说话人的中性汉语训练语料库，录制了一个9个女性说话人11种情感共9900句作为多说话人的目标情感普通话训练语料库，即11中情感包括悲伤、放松、愤怒、焦虑、惊奇、恐惧、轻蔑、温顺、喜悦、厌恶、中性。实验证明，随着普通话目标情感训练语料的增加，合成的目标情感的藏语或汉语语音的情感相似度评测得分(EmotionalMean OpinionScore,EMOS)逐渐提高。

[0160] 本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。对于实施例公开的系统而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。

[0161] 本文中应用了具体个例对本发明的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本发明的方法及其核心思想；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处。综上所述，本说明书内容不应理解为对本发明的限制。

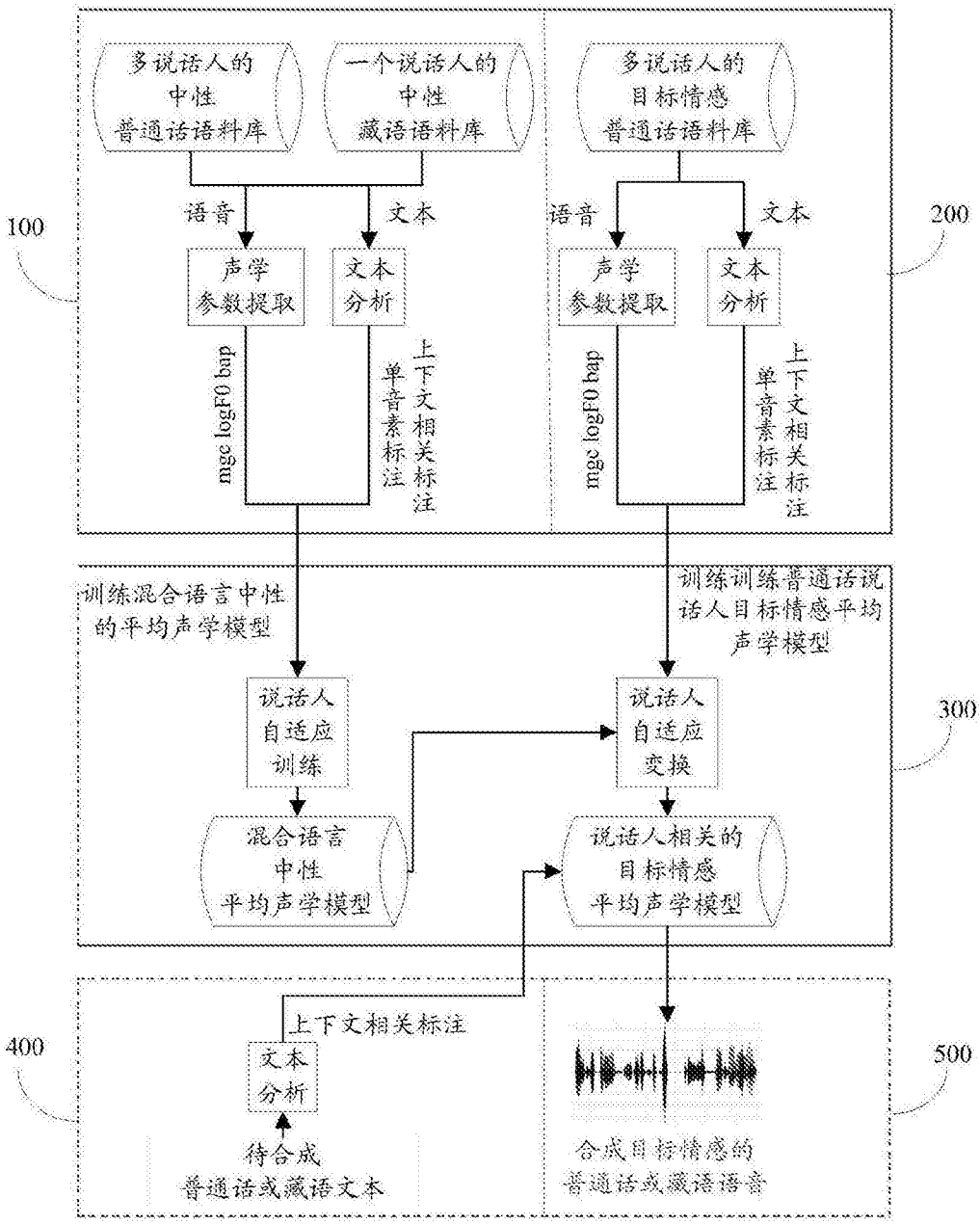

多说话人的
中性
普通话语料库
一个说话人的
中性
藏语语料库
多说话人的
目标情感
普通话语料库
语音
文本
语音
文本
100
200
声学
参数提取
文本
分析
声学
参数提取
文本
分析
mgc logF0 bap
单音素标注
上下文相关标注
mgc logF0 bap
单音素标注
上下文相关标注
训练混合语言中性
的平均声学模型
训练训练普通话说
话人目标情感平均
声学模型
300
说话人
自适应
训练
说话人
自适应
变换
混合语言
中性
平均声学模型
说话人相关的
目标情感
平均声学模型
上下文相关标注
400
500
文本
分析
待合成
普通话或藏语文本
合成目标情感的
普通话或藏语语音

图1

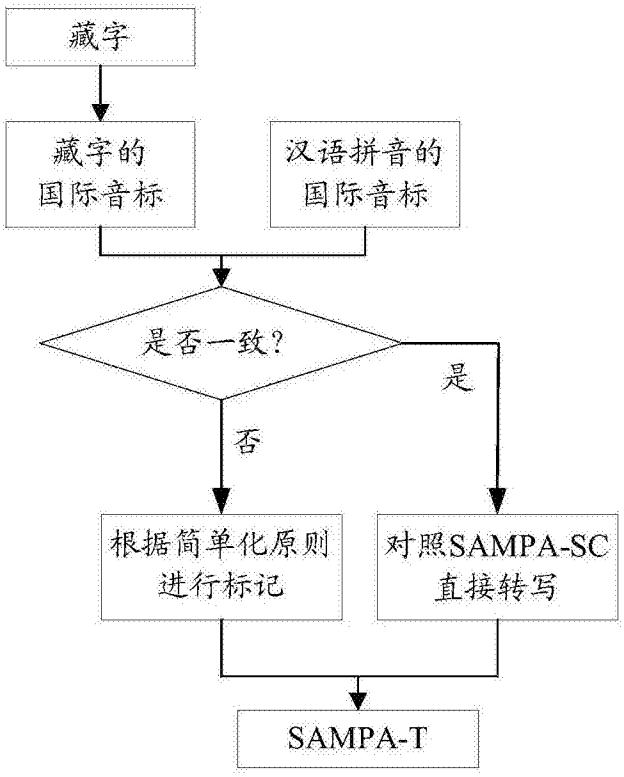

藏字
藏字的
国际音标
汉语拼音的
国际音标
是否一致?
是
否
根据简单化原则
进行标记
对照SAMPA-SC
直接转写
SAMPA-T

图2

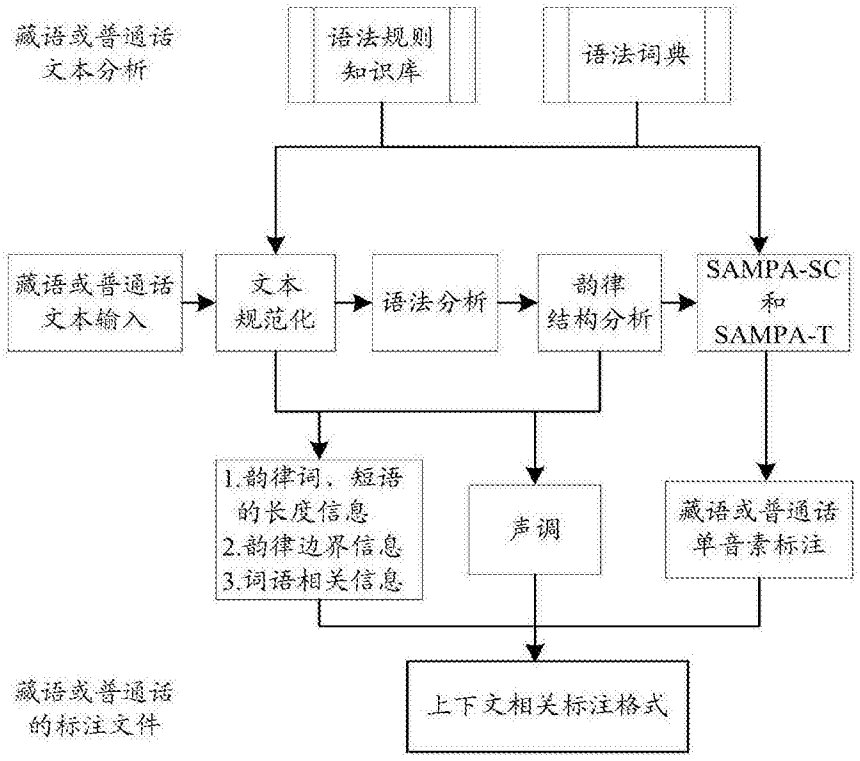

藏语或普通话
文本分析
语法规则
知识库
语法词典
藏语或普通话
文本输入
文本
规范化
语法分析
韵律
结构分析
SAMPA-SC
和
SAMPA-T
1.韵律词，短语
的长度信息
2.韵律边界信息
3.词语相关信息
声调
藏语或普通话
单音素标注
藏语或普通话
的标注文件
上下文相关标注格式

图3

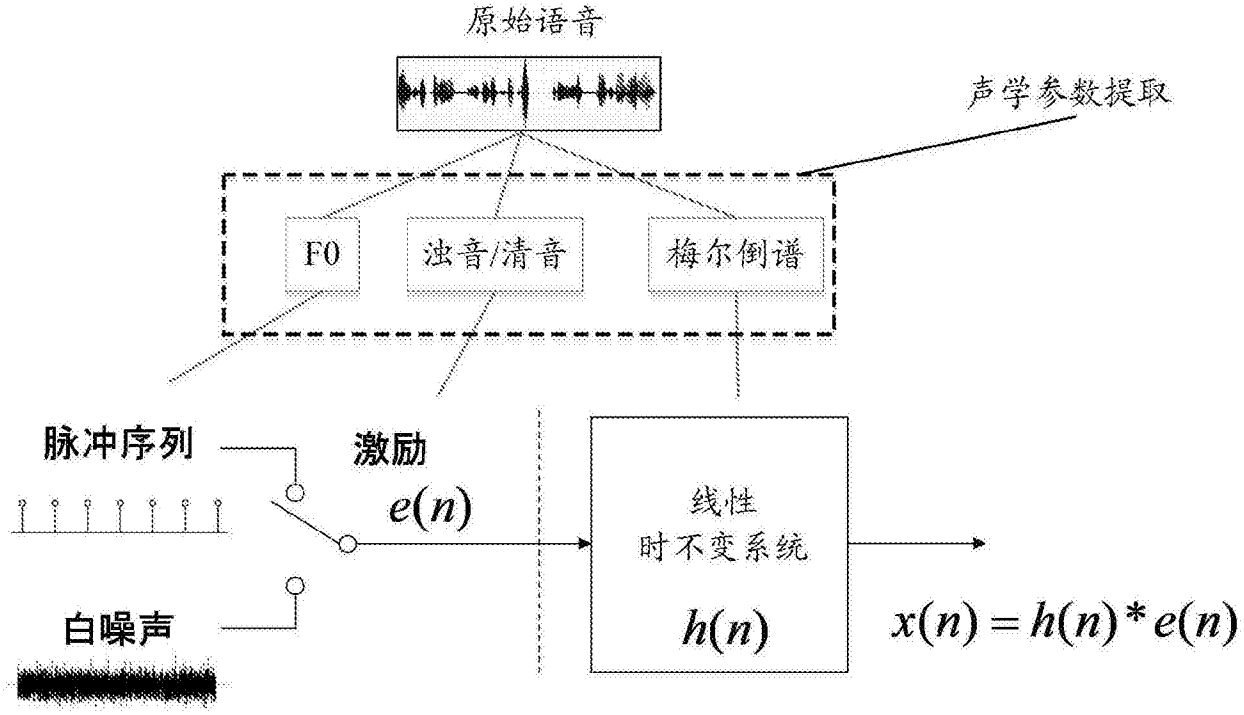

原始语音
声学参数提取
F0
浊音/清音
梅尔倒谱
脉冲序列
激励
e(n)
白噪声
线性
时不变系统
h(n)
x(n) = h(n) * e(n)

图4

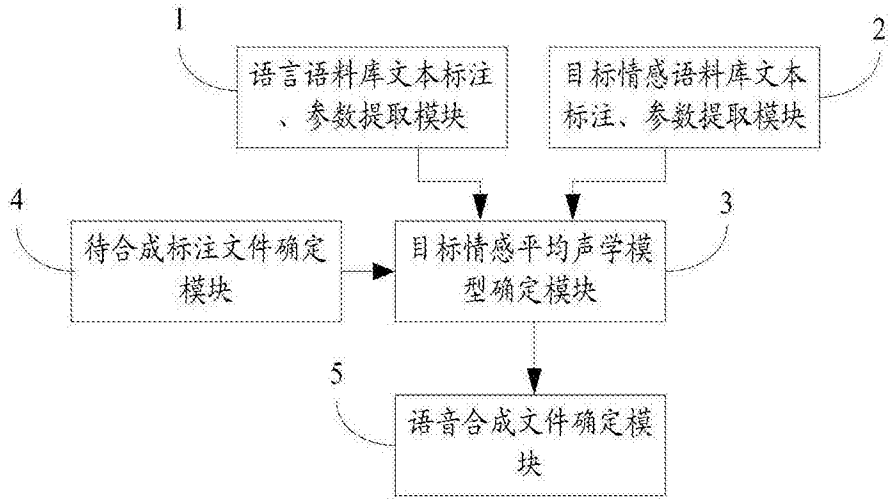

1
语言语料库文本标注、参数提取模块
2
目标情感语料库文本标注、参数提取模块
4
待合成标注文件确定模块
3
目标情感平均声学模型确定模块
5
语音合成文件确定模块

图5