(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101894548 B
(45)授权公告日 2012.07.04

(21)申请号 201010207237.9

(22)申请日 2010.06.23

(73)专利权人 清华大学
地址 100084 北京市 100084-82 信箱

(72)发明人 何亮 张卫强 刘加

(74)专利代理机构 北京市立方律师事务所 11330
代理人 马佑平

(51)Int.Cl.
G10L 15/02(2006.01)
G10L 15/06(2006.01)
G10L 21/02(2006.01)

(56)对比文件
CN 101702314 A,2010.05.05,
CN 1588535 A,2005.03.02,
US 2008/0147380 A1,2008.06.19,
CN 101178705 A,2008.05.14,
CN 101727903 A,2010.06.09,
钟山
刘加
.MLLR 特征的 SVM 语种识别算法 .《清华大学学报(自然科学版)》.2009, 第 1284 页第 1.1 节至第 1286 页第 3.3 节 .

审查员 王玥

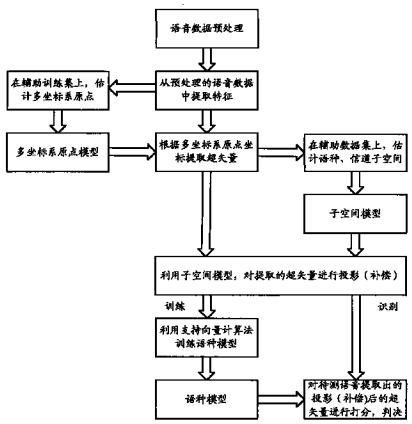
权利要求书 3 页 说明书 9 页 附图 2 页

(54)发明名称

一种用于语种识别的建模方法及装置

(57)摘要

本发明的实施例提出了一种用于语种识别的建模方法,包括输入语音数据,对语音数据预处理得到特征序列,将特征矢量映射为超矢量,对超矢量进行投影补偿,通过支持向量机算法建立训练语种模型;对待测语音采用上述步骤得到待测超矢量,对待测超矢量进行投影补偿,利用语种模型对所述待测超矢量打分,识别待测语音的语言种类。本发明实施例还提出了一种用于语种识别的建模装置包括语音预处理模块、特征提取模块、多坐标系原点选择模块、特征矢量映射模块、子空间提取模块、子空间投影补偿模块、训练模块和识别模块。根据本发明实施例提供的方法及装置,去除高维统计量中对识别无效的信息,提高语种识别的正确率,降低在集成电路上的运算复杂度。



CN 101894548 B

1. 一种用于语种识别的方法，其特征在于，包括如下步骤：

输入语音数据，对所述语音数据预处理得到特征序列，所述特征序列由特征矢量组成，并根据坐标系选择算法和特征矢量映射算法，将所述特征序列的特征矢量映射，并在时间上取平均，得超矢量。对所述超矢量进行投影和补偿，通过支持向量机算法建立并训练语种模型；

上述步骤包括：

1) 对所述语音数据进行预处理，从所述预处理后的语音数据中提取特征序列，特征序列由特征矢量组成；

2) 从所述特征矢量所在的空间中选择各个坐标系原点，确定所述特征矢量与坐标系原点之间的度量关系，根据坐标系选择算法和特征矢量映射算法，将所述特征序列的特征矢量映射，并在时间上取平均，得到超矢量；

3) 根据所述超矢量，训练信道子空间和语种子空间，利用所述信道子空间和语种子空间对超矢量进行投影和补偿，提取超矢量仅存在于语种子空间的部分；

4) 通过支持向量机算法，建立并训练语种模型；

输入待测语音，对所述待测语音预处理得到特征序列，所述特征序列由特征矢量组成，并根据坐标系选择算法和特征矢量映射算法，将所述特征序列的特征矢量映射，并在时间上取平均，得待测超矢量，对所述待测超矢量进行投影和补偿，利用所述语种模型对所述待测超矢量打分，识别所述待测语音的语言种类。

2. 如权利要求 1 所述的方法，其特征在于，所述从特征矢量所在的空间中选择各个坐标系原点包括以下两种方式之一：

采用 EM 算法训练高斯混合模型，并将高斯混合模型均值作 为各个坐标系原点；

采用 VQ 算法，选用 VQ 码本作为各个坐标系的原点。

3. 如权利要求 1 所述的方法，其特征在于，所述利用语种模型对所述待测超矢量打分，识别所述待测语音的语言种类进一步包括：

1) 对所述待测语音进行预处理，从所述预处理后的待测语音中提取特征序列，特征序列由特征矢量组成；

2) 利用根据坐标系选择算法和特征矢量映射算法，将所述特征序列的特征矢量映射，并在时间上取平均，得待测超矢量；

3) 根据所述待测超矢量，利用信道子空间和语种子空间对所述待测超矢量进行投影和补偿，提取所述待测超矢量仅存在于语种子空间的部分；

4) 利用所述语种模型对所述待测超矢量进行打分，与判决门限比较，识别所述待测语音的语言种类。

4. 如权利要求 1 所述的方法，其特征在于，所述训练信道子空间和语种子空间通过以下算法之一：

主成分分析算法、概率主成分分析算法或者基于核方法的主成分分析算法。

5. 如权利要求 1 或 3 所述的方法，其特征在于，利用所述信道子空间和语种子空间对所述待测超矢量进行投影和补偿进一步包括：

对所述语音数据，选取所述超矢量仅存在于语种子空间的部分；

对所述待测语音，选取所述待测超矢量仅存于语种子空间的部分。

6. 一种用于语种识别的装置，其特征在于，包括语音预处理模块、特征提取模块、多坐标系原点选择模块、特征矢量映射模块、子空间提取模块、子空间投影补偿模块、训练模块和识别模块，

其中，语音预处理模块，用于降噪，并去除与语种识别无关的内容，输出去除后的纯语音；

特征提取模块，用于读入所述预处理模块的语音，并提取特征，输出特征序列，特征序列由特征矢量组成；

多坐标系原点选择模块，用于选取辅助训练集，在特征矢量空间选择各个坐标系原点；

特征矢量映射模块，用于根据选定的各个坐标系原点，将所述特征提取模块输出的特征序列映射成为超矢量；

子空间提取模块，用于利用辅助训练集上的超矢量，训练语种子空间和信道子空间；

子空间投影补偿模块，用于利用所述语种子空间和信道子空间，对所述语音数据的超矢量和待测语音的超矢量进行投影补偿；

训练模块，用于对经过子空间投影补偿的超矢量，采用支持向量机算法建立并训练语种模型；

识别模块，利用所述语种模型对待测超矢量打分，识别所述待测语音的语言种类。

7. 如权利要求 6 所述的装置，其特征在于，所述装置通过支持向量机算法建立并训练语种模型进一步包括：

1) 语音预处理模块对所述语音数据进行预处理，特征提取模块从所述预处理后的语音数据中提取特征序列，特征序列由特征矢量组成；

2) 所述多坐标系原点选择模块从所述特征矢量所在的空间中选择各个坐标系原点，确定所述特征矢量与坐标系原点之间的度量关系，所述特征矢量映射模块根据坐标系选择算法和特征矢量映射算法，将所述特征矢量映射，并在时间上取平均，得超矢量；

3) 所述子空间提取模块根据所述超矢量，训练信道子空间和语种子空间，所述子空间投影补偿模块利用所述信道子空间和语种子空间对超矢量进行投影和补偿，提取超矢量仅存在于语种子空间的部分；

4) 所述训练模块通过支持向量机算法，建立并训练语种模型。

8. 如权利要求 7 所述的装置，其特征在于，所述多坐标系原点选择模块从特征矢量所在的空间中选择各个坐标系原点包括以下两种方式之一：

采用 EM 算法训练高斯混合模型，并将高斯混合模型均值作为各个坐标系原点；

采用 VQ 算法，选用 VQ 码本作为各个坐标系的原点。

9. 如权利要求 6 所述的装置，其特征在于，所述装置还包括多坐标系原点存储模块、子空间模型存储模块和语种模型存储模块，

所述多坐标系原点存储模块，用于存储来自所述多坐标系原点选择模块的坐标系原点；

所述子空间模型存储模块，用于存储来自子空间提取模块的语种子空间和信道子空间；

所述语种模型存储模块，用于存储来自所述训练模块的语种模型。

10. 如权利要求 9 所述的装置，其特征在于，所述装置利用语种模型对所述待测超矢量打分，识别所述待测语音的语言种类进一步包括：

1) 所述语音预处理模块对所述待测语音进行预处理，所述特征提取模块从所述预处理后的待测语音中提取特征序列，特征序列由特征矢量组成；

2) 根据所述多坐标系原点存储模块存储的坐标系原点，特征矢量映射模块，通过坐标系选择算法和特征序列映射算法，将所述特征矢量映射，并在时间上取平均，得待测超矢量；

3) 所述子空间投影补偿模块根据所述待测超矢量，利用所述子空间模型存储模块存储的语种子空间和信道子空间，对所述待测超矢量进行投影和补偿，提取所述待测超矢量仅存在于语种子空间的部分；

4) 所述识别模块利用所述语种模型存储模块存储的语种模型对所述待测超矢量进行打分，与判决门限比较，识别所述待测语音的语言种类。

11. 如权利要求 7 所述的装置，其特征在于，所述子空间提取模块训练信道子空间和语种子空间通过以下算法之一：

主成分分析算法、概率主成分分析算法或者基于核方法的主成分分析算法。

12. 如权利要求 7 或 10 所述的装置，其特征在于，所述子空间投影补偿模块利用所述信道子空间和语种子空间对所述待测超矢量进行投影和补偿进一步包括：

对所述语音数据，选取所述超矢量仅存在于语种子空间的部分；

对所述待测语音，选取所述待测超矢量仅存于语种子空间的部分。

一种用于语种识别的建模方法及装置

技术领域

[0001] 本发明涉及语音识别、模式识别和信号处理，具体而言，本发明涉及一种用于语种识别的建模方法及装置。

背景技术

[0002] 语种识别是指利用机器判别给定语音语言种类的技术。语种识别技术是多语言处理系统的前端，可用于语音人性化服务、语音安全监控等领域。

[0003] 目前，语种识别领域最流行的系统建模方法是：对预处理后的语音提取频谱层特征，随后采用 GMM(Gaussian Mixture Models，高斯混合模型）或 SVM(Support Vector Machine，支持向量机）进行系统建模。

[0004] 常用的频谱层特征有 Mel 频率倒谱系数 (MFCC)、线性预测倒谱系数 (LPCC) 和感知线性预测 (PLP) 及它们的衍生特征。经过特征提取过程，预处理的语音信号转化为更容易进行语种识别的时间序列。GMM 和 SVM 这两种建模方法试图从两种角度对时间序列进行识别。前者利用模型参数对时间序列的分布进行拟合；后者在高维空间寻找最优分类面。两类建模方法各有所长：GMM 建模方法参数物理意义明确，在训练、识别数据充分的情况下有较好的性能；SVM 建模方法基于结构风险最小化原则，在训练数据稀少的情况下有较好的识别能力。最近提出的 GMM-SVM 建模方法将 GMM 模型本身作为 SVM 分类器的输入。

[0005] 与 GMM 或 SVM 建模方法相比，GMM-SVM 建模方法具有两个明显优点：1) 利用支持向量机算法对 GMM 的权重、权重或方差进行鉴别式建模，提高语种识别率；

[0006] 2) 融合子空间投影（补偿）技术，可以解决训练数据与待识别语音数据信道不匹配的问题，并进一步解决待识别语音数据稀少的问题。

[0007] GMM-SVM 建模方法的不足之处在于：

[0008] 1)GMM 的协方差矩阵通常被简化为对角阵，协方差矩阵的非对角阵元素所含有的鉴别式信息并没有被利用；

[0009] 2)GMM 模型不包含高阶统计量 (3 阶以及 3 阶以上)，而合理使用高阶统计量可以有效提高语种识别率；

[0010] 3)GMM-SVM 的子空间投影和子空间补偿技术都基于线性空间，而时间序列所隐含的非线性信息没有被有效利用。

发明内容

[0011] 本发明的目的旨在至少解决上述技术缺陷之一，特别针对有效利用时间序列的高阶统计量，更可以采用线性子空间、非线性子空间技术对提出的统计量进行投影补偿，进一步提升语种识别系统性能，提出了一种用于语种识别的建模的方法及装置。

[0012] 为实现上述目的，本发明实施例一方面提出了一种用于语种识别的建模方法，包括如下步骤：

[0013] 输入语音数据，对所述语音数据预处理得到特征序列，所述特征序列由特征向量

组成，并根据坐标系选择算法和特征矢量映射算法，将所述特征矢量映射为超矢量，对所述超矢量进行投影和补偿，通过支持向量机算法建立并训练语种模型；

[0014] 输入待测语音，对所述待测语音预处理得到特征序列，所述特征序列由特征向量组成，并根据坐标系选择算法和特征矢量映射算法，将所述特征矢量映射为待测超矢量，对所述待测超矢量进行投影和补偿，利用所述语种模型对所述待测超矢量打分，识别所述待测语音的语言种类。

[0015] 本发明实施例另一方面提出了一种用于语种识别的建模装置，包括语音预处理模块、特征提取模块、多坐标系原点选择模块、特征矢量映射模块、子空间提取模块、子空间投影补偿模块、训练模块和识别模块。

[0016] 其中，语音预处理模块，用于降噪，并去除与语种识别无关的内容，输出去除后的纯语音；

[0017] 特征提取模块，用于读入所述预处理模块的语音，并提取特征，输出特征序列，所述特征序列由特征向量组成；

[0018] 多坐标系原点选择模块，用于选取辅助训练集，在特征序列空间选择各个坐标系原点；

[0019] 特征矢量映射模块，用于根据选定的各个坐标系原点，将所述特征提取模块输出的特征矢量映射成为超矢量；

[0020] 子空间提取模块，用于利用辅助训练集上的超矢量训练语种子空间和信道子空间；

[0021] 子空间投影补偿模块，用于利用所述语种子空间和信道子空间，对所述语音数据的超矢量和待测语音的超矢量进行投影补偿；

[0022] 训练模块，用于对经过子空间投影补偿的超矢量，采用支持向量机算法建立并训练语种模型；

[0023] 识别模块，利用所述语种模型对所述待测超矢量打分，识别所述待测语音的语言种类。

[0024] 根据本发明实施例提供的用于语种识别的建模方法及装置，通过对语音信号特征序列的高维统计量有效建模，并采用子空间技术，去除了高维统计量中对识别无效的信息，提高了语种识别的正确率，又降低了在集成电路上的运算复杂度。

[0025] 本发明提出的上述方案，对现有系统的改动很小，不会影响系统的兼容性，而且实现简单、高效。

[0026] 本发明附加的方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本发明的实践了解到。

附图说明

[0027] 本发明上述的和 / 或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解，其中：

[0028] 图 1 为根据本发明实施例的用于语种识别的建模方法结构框图；

[0029] 图 2 为图 1 中用于语种识别的建模方法的实施流程图；

[0030] 图 3 为根据本发明实施例的用于语种识别的建模装置的结构框图。

具体实施方式

[0031] 下面详细描述本发明的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，仅用于解释本发明，而不能解释为对本发明的限制。

[0032] 为实现本发明之目的，本发明实施例公开了一种用于语种识别的建模方法。图 1 示出了该建模方法的流程框图。如图 1 所示，该方法包括如下步骤：

[0033] S101：输入语音数据，对语音数据预处理得到特征序列，并根据坐标系选择算法和特征矢量映射算法，将特征矢量映射为超矢量，对超矢量进行投影和补偿，通过支持向量机算法建立并训练语种模型；

[0034] 具体的说，结合图 2 所示，首先输入语音数据，然后执行如下步骤：

[0035] A1：语音数据预处理。

[0036] A11：对语音数据即语音信号进行零均值化和预加重，其中零均值化为整段语音减去其均值。预加重为语音进行高通滤波。

[0037] 其中，高通滤波器传输函数为 $H(z) = 1-\alpha z^{-1}$，其中 $0.95 \leqslant \alpha \leqslant 1$。

[0038] A12：对语音信号分帧。其中，帧长为 20ms，帧移为 10ms。

[0039] A2：从预处理的语音数据中提取特征序列。

[0040] 特征序列是由一系列的特征向量组成。

[0041] A21：对语音信号加汉明窗，其中窗函数为：

[0042] $$\omega_H(n) = \begin{cases} 0.54 - 0.46\cos(\dfrac{2\pi n}{N-1}) & 0 \leq n \leq N-1 \\ 1 & others \end{cases}$$

[0043] A22：对加汉明窗的数据做离散傅立叶变换 (DFT)

[0044] $$X(\omega_k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{M}nk}$$

[0045] 其中 ω_k 代表频率，k 代表频率标号，N 是 DFT 变换点数。

[0046] A23：选择有 M(m = 1,2,…,M) 个滤波器的滤波器组，其中第 m 个三角型滤波器如下定义

[0047] $$H_m[k] = \begin{cases} 0 & k < f[m-1] \\ \dfrac{(k - f[m-1])}{(f[m] - f[m-1])} & f[m-1] \leq k \leq f[m] \\ \dfrac{(f[m+1] - k)}{(f[m+1] - f[m])} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m] \end{cases},$$

[0048] 其中，$\sum_{m=1}^{M} H_m[k] = 1$，f[m] 为三角窗的边界点，由下式确定：

[0049] $$f[m] = \frac{N}{F_s} B^{-1}\left(B(f_l) + m\frac{B(f_h) - B(f_l)}{M+1}\right)$$

[0050] f_l 和 f_h 为给定滤波器组的最低频率和最高频率，B 为频率向梅尔频标的映射函数

[0051] $B(f) = 1125\ln(1+(f/700))$，

[0052] B^{-1} 为梅尔频标向频率的映射函数：

[0053] $B^{-1}(b) = 700\exp((b/1125)-1)$。

[0054] A24：计算每个滤波器输出的对数能量

[0055] $$S[m] = \ln\left[\sum_{k=0}^{N-1}|X_\omega[k]|^2 H_m[k]\right], 0 < m \le M$$。

[0056] A25：离散余弦变换，并计算 MFCC 系数

[0057] $$c[n] = \sum_{m=0}^{M-1} S[m]\cos(\pi n(m-1/2)/M), 0 < m \le M$$，

[0058] 取前 7 维系数，拼接成为 MFCC 基本特征 $c = [c_0, c_1, \cdots, c_6]$。

[0059] A26：计算第 n 时刻的偏移差分倒谱特征 (SDC)，

[0060] $s_{(iN+j)}(n) = c_j(n+iS+b)-c_j(n+iS-b)$，$j = 1,2,\cdots,N-1$；$i = 0,1,\cdots,K-1$

[0061] 其中，b 为计算一阶差分特征时的帧数差，一般取值为 1；K 为块数，一般取值为 7；S 为各块之间的偏移帧数，一般取值为 3。

[0062] 在本实施例中，$b = 1, K = 7, S = 3$

[0063] A27：将基本特征与差分特征拼接，形成新的特征矢量。

[0064] $y(n) = \{c_j(n), j = 0,1,\cdots,N-1; s_{iN+j}(n), j = 0,1,\cdots,N-1, i = 0,1,\cdots,K-1\}$。

[0065] A3：选取各个坐标系原点，提取高维统计量。

[0066] A31：在辅助训练数据集上，通过 EM 算法选取多坐标系原点坐标 $o = \{o_1, o_2, \cdots, o_C\}$，C 为坐标系数目。

[0067] A32：选择特征矢量 y(n) 与原点坐标 o_j 的度量 $f[y(n), o_c]$，$1 \le c \le C$，并计算特征矢量 y(n) 在每个子坐标系的占有率

[0068] $$\gamma[y(n)|o_j] = \frac{f[y(n), o_j]}{\sum_{c=1}^{C} f[y(n), o_c]}$$。

[0069] A33：选择特征矢量 y(n) 在坐标系内的扩展函数 $g[y(n), c_c]$，结合步骤 A32 计算所得的占有率，将特征矢量 y(n) 映射为超矢量

[0070] $v(n) = [r[y(n)|o_1]g[y(n), o_1], r[y(n)|o_2]g[y(n), o_2], \cdots, r[y(n)|o_C)g(y(n), o_C]]$

[0071] A34：超矢量序列 v(n) 对时间取平均，得到该段语音的超矢量 $\mathbf{v} = \frac{1}{T}\sum_{n=1}^{T}\mathbf{v}(n)$。

[0072] A4：采用子空间技术，寻找信道子空间和语种子空间。

[0073] 其中，信道子空间为超矢量所属空间所包含的，不利于识别的子空间。语种子空间为超矢量所属空间所包含的，有利于识别的子空间。

[0074] 通过子空间投影、补偿技术，提取超矢量 v 中仅有利于语种识别的部分。

[0075] A41：在辅助训练数据集上，提取语音超矢量 $\{v_0, v_1, v_2, \cdots, v_L\}$。对辅助训练集的要求是，尽量包含训练、识别语音的语言种类，并且每个语种要对应多段语音。

[0076] A42：对提取的语音超矢量 $\{v_0, v_1, v_2, \cdots, v_L\}$ 采用主成分分析算法 (PCA，包括直接求解矩阵方法和迭代求解法)，提取语种子空间 L。

[0077] A43：对提取的语音超矢量 $\{v_0, v_1, v_2, \cdots, v_L\}$ 进行修正，每个超矢量减去对应语种超矢量的期望，得到新的语音超矢量 $\{v'_0, v'_1, v'_2, \cdots, v'_L\}$。对新的语音超矢量采用主成分分析算法 (PCA，包括直接求解矩阵方法和迭代求解法)，提取信道子空间 U。

[0078] A5：利用语种子空间 L、信道子空间 U，对超矢量 v 进行投影和补偿，提取超矢量 v 仅存在于语种子空间 L 的部分 v^L。

[0079] A6：通过支持向量机算法，建立语种模型。

[0080] A61：支持向量机训练算法；

[0081] 令输入样本集为$(\mathbf{v}_p^{\mathbf{L}}, \theta_p)$，p = [1，2，…，P]，$\theta_p \in \{+1, -1\}$，通常，$\theta_p = +1$ 的样本称为正样本，$\theta_p = -1$ 的样本称为负样本。SVM 算法寻找最优分类面 w，使得正负样本集之间的距离最大。最优分类面 w 是通过求解下述优化函数而得

[0082] $$\min L = \frac{1}{2}\|w\|^2 + C\left(\sum_{p=1}^{P}\xi_p\right)$$

[0083] 其中，$\|w\|^2$ 与正负样本之间距离成反比，ξ_p 是在样本线性不可分的情况下引入的松弛变量，C 是控制错分样本的惩罚程度。上式在对偶空间求解，优化函数变为

[0084] $$\max \sum_{p=1}^{P}\alpha_p - \frac{1}{2}\sum_{p,q=1}^{P}\alpha_p\alpha_q\theta_p\theta_q K(v_p^{\mathbf{L}}, v_q^{\mathbf{L}})$$

[0085] 其中，$\sum_{p=1}^{P}\theta_p\alpha_p = 0$，$\alpha_p \geqslant 0$，p = 1，2，…，P，$K(v_p, v_q)$ 为 $v_p^{\mathbf{L}}$ 和 $v_q^{\mathbf{L}}$ 的核函数。

[0086] 设最优解 α^*，则最优分类面是训练样本的组合$\{\alpha_p^*\theta_p v_p^{\mathbf{L}}\}$，p = [1，2，…，P]。

[0087] A62：对步骤 A5 中获得的超矢量，采用步骤 A61 中的支持向量机算法建立并训练语种模型。

[0088] S102：输入待测语音，对待测语音预处理得到特征序列，并根据坐标系选择算法和特征矢量映射算法，将特征矢量映射为待测超矢量，对待测超矢量进行投影和补偿，利用语种模型对待测超矢量打分，识别待测语音的语言种类。

[0089] 具体的说，首先输入待测语音，然后采用上述步骤 A1、A2、A3 和 A5 中的方法提取超矢量。

[0090] B1：根据输入的待测语音，采用上述步骤 A1、A2、A3 和 A5 中的方法提取超矢量。具体的说，

[0091] B11：对待测语音进行预处理，从预处理后的待测语音中提取特征序列，特征序列是由一系列的特征向量组成；

[0092] B12：根据步骤 A3 中得到的各个坐标系原点，利用根据坐标系选择算法和特征序列映射算法，将特征矢量映射成待测超矢量；

[0093] B13：根据待测超矢量，通过步骤 A4 中得到的信道子空间和语种子空间，利用信道子空间和语种子空间对待测超矢量进行投影和补偿，提取待测超矢量仅存在于语种子空间的部分；

[0094] B2：利用步骤 A62 中训练的语种模型，对步骤 B1 中输出的超矢量进行打分，得到输出分数。其中打分函数为：

[0095] $$f(v) = \sum_{p=1}^{P}\alpha_p^*\theta_p K(v_p^{\mathbf{L}}, v^{\mathbf{L}}) + b^*$$

。

[0096] B3：对步骤 B2 的输出分数进行后处理，与判决门限比较，判别该段语音的语言种

类。

[0097] 根据本发明实施例提供的用于语种识别的建模方法，通过对语音信号特征序列的高维统计量有效建模，并采用子空间技术，去除了高维统计量中对识别无效的信息，提高了语种识别的正确率，又降低了在集成电路上的运算复杂度。

[0098] 本发明实施例还提出了一种用于语种识别的建模装置。图 3 示出了该建模装置的结构框图。如图 3 中所示，该装置包括特征提取模块、多坐标系原点选择模块、特征矢量映射模块、子空间提取模块、子空间投影补偿模块、训练模块和识别模块。

[0099] 其中，语音预处理模块，用于降噪，并去除彩铃、音乐等与语种识别无关的部分，输出纯净语音供特征提取模块；

[0100] A11：语音预处理模块对语音数据即语音信号进行零均值化和预加重，其中零均值化为整段语音减去其均值。预加重为语音进行高通滤波。

[0101] 其中，高通滤波器传输函数为 $H(z) = 1-\alpha z^{-1}$，其中 $0.95 \leqslant \alpha \leqslant 1$。

[0102] A12：语音预处理模块对语音信号分帧。其中，帧长为 20ms，帧移为 10ms。

[0103] 特征提取模块，用于读入预处理模块的语音，并提取特征，输出特征序列。其中，特征序列由特征向量组成。

[0104] 特征序列是由一系列的特征向量组成。

[0105] A21：特征提取模块对语音信号加汉明窗，其中窗函数为：

[0106] $$\omega_H(n) = \begin{cases} 0.54 - 0.46\cos(\dfrac{2\pi n}{N-1}) & 0 \leq n \leq N-1 \\ 1 & others \end{cases}$$

[0107] A22：特征提取模块对加汉明窗的数据做离散傅立叶变换 (DFT)

[0108] $$X(\omega_k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{M}nk}$$

[0109] 其中 ω_k 代表频率，k 代表频率标号，N 是 DFT 变换点数。

[0110] A23：特征提取模块选择有 M(m = 1,2,…,M) 个滤波器的滤波器组，其中第 m 个三角型滤波器如下定义

[0111] $$H_m[k] = \begin{cases} 0 & k < f[m-1] \\ \dfrac{(k - f[m-1])}{(f[m] - f[m-1])} & f[m-1] \leq k \leq f[m] \\ \dfrac{(f[m+1] - k)}{(f[m+1] - f[m])} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m] \end{cases},$$

[0112] 其中，$\sum_{m=1}^{M} H_m[k] = 1$，f[m] 为三角窗的边界点，由下式确定：

[0113] $$f[m] = \frac{N}{F_s} B^{-1}\left(B(f_l) + m\frac{B(f_h) - B(f_l)}{M+1}\right)$$

[0114] f_l 和 f_h 为给定滤波器组的最低频率和最高频率，B 为频率向梅尔频标的映射函数

[0115] B(f) = 1125ln(1+(f/700))，

[0116] B^{-1} 为梅尔频标向频率的映射函数：

[0117] $B^{-1}(b) = 700\exp((b/1125)-1)$。

[0118] A24：特征提取模块计算每个滤波器输出的对数能量

[0119] $$S[m]=\ln\left[\sum_{k=0}^{N-1}|X_{\omega}[k]|^2H_m[k]\right],0<m\leq M$$

[0120] A25：离散余弦变换，并计算 MFCC 系数

[0121] $$c[n]=\sum_{m=0}^{M-1}S[m]\cos\left(\pi n(m-1/2)/M\right),0<m\leq M$$

[0122] 取前 7 维系数，拼接成为 MFCC 基本特征 c = [c_0, c_1, …, c_6]。

[0123] A26：特征提取模块计算第 n 时刻的偏移差分倒谱特征 (SDC)，

[0124] $s_{(iN+j)}(n) = c_j(n+iS+b)-c_j(n+iS-b)$，j = 1,2, …,N-1；i = 0,1, …,K-1

[0125] 其中，b 为计算一阶差分特征时的帧数差，一般取值为 1；K 为块数，一般取值为 7；S 为各块之间的偏移帧数，一般取值为 3。

[0126] 在本实施例中，b = 1，K = 7，S = 3

[0127] A27：特征提取模块将基本特征与差分特征拼接，形成新的特征矢量。

[0128] $y(n) = \{c_j(n), j = 0,1, \cdots,N-1; s_{iN+j}(n), j = 0,1, \cdots,N-1, i = 0,1, \cdots,K-1\}$

[0129] 多坐标系原点选择模块，用于选取辅助训练集，在特征序列空间选择各个坐标系原点。

[0130] A31：多坐标系原点选择模块在辅助训练数据集上，通过 EM 算法选取多坐标系原点坐标 $o = \{o_1, o_2 \cdots, o_C\}$，C 为坐标系数目。

[0131] A32：多坐标系原点选择模块选择特征矢量 y(n) 与原点坐标 o_j 的度量 $f[y(n), o_c]$，$1 \leqslant c \leqslant C$，并计算特征矢量 y(n) 在每个子坐标系的占有率

[0132] $$\gamma[y(n)|o_j]=\frac{f[y(n),o_j]}{\sum_{c=1}^{C}f[y(n),o_c]}。$$

[0133] 特征矢量映射模块，用于根据选定的各个坐标系原点，将特征提取模块输出的特征矢量映射成为超矢量。

[0134] A33：多坐标系原点选择模块选择特征矢量 y(n) 在坐标系内的扩展函数 $g[y(n), c_c]$，根据计算所得的占有率，特征矢量映射模块将特征矢量 y(n) 映射为超矢量

[0135] $v(n) = [r[y(n)|o_1]g[y(n),o_1], r[y(n)|o_2]g[y(n),o_2], \cdots, r[y(n)|o_C)g(y(n), o_C]]$

[0136] A34：超矢量序列 v(n) 对时间取平均，得到该段语音的超矢量$\mathbf{v}=\frac{1}{T}\sum_{n=1}^{T}\mathbf{v}(n)$。

[0137] 子空间提取模块，用于利用辅助训练集上的超矢量训练语种子空间和信道子空间。

[0138] 其中，信道子空间为超矢量所属空间所包含的，不利于识别的子空间。语种子空间为超矢量所属空间所包含的，有利于识别的子空间。

[0139] 通过子空间投影、补偿技术，提取超矢量 v 中仅有利于语种识别的部分。

[0140] A41：在辅助训练数据集上，子空间提取模块提取语音超矢量 {v_0, v_1, v_2, …, v_L}。对辅助训练集的要求是，尽量包含训练、识别语音的语言种类，并且每个语种要对应多段语音。

[0141] A42：子空间提取模块对提取的语音超矢量 {v_0, v_1, v_2, …, v_L} 采用主成分分析算法 (PCA，包括直接求解矩阵方法和迭代求解法)，提取语种子空间 L。

[0142] A43：子空间提取模块对提取的语音超矢量 $\{v_0, v_1, v_2, \cdots, v_L\}$ 进行修正，每个超矢量减去对应语种超矢量的期望，得到新的语音超矢量 $\{v'_0, v'_1, v'_2, \cdots, v'_L\}$。对新的语音超矢量采用主成分分析算法（PCA，包括直接求解矩阵方法和迭代求解法），提取信道子空间 U。

[0143] 子空间投影补偿模块，用于利用语种子空间和信道子空间，对语音数据的超矢量和待测语音的超矢量进行投影补偿。

[0144] 利用语种子空间 L、信道子空间 U，对超矢量 v 进行投影和补偿，提取超矢量 v 仅存在于语种子空间 L 的部分 v^L。

[0145] 训练模块，用于对经过子空间投影补偿的超矢量，采用支持向量机算法建立并训练语种模型。

[0146] A61：支持向量机训练算法；

[0147] 训练模块令输入样本集为 $(\mathbf{v}_p^{\mathbf{L}}, \theta_p)$，$p = [1,2, \cdots, P]$，$\theta_p \in \{+1, -1\}$，通常，$\theta_p = +1$ 的样本称为正样本，$\theta_p = -1$ 的样本称为负样本。SVM 算法寻找最优分类面 w，使得正负样本集之间的距离最大。最优分类面 w 是通过求解下述优化函数而得

[0148] $$\min L = \frac{1}{2}\|w\|^2 + C\left(\sum_{p=1}^{P}\xi_p\right)$$

[0149] 其中，$\|w\|^2$ 与正负样本之间距离成反比，ξ_p 是在样本线性不可分的情况下引入的松弛变量，C 是控制错分样本的惩罚程度。上式在对偶空间求解，优化函数变为

[0150] $$\max \sum_{p=1}^{P}\alpha_p - \frac{1}{2}\sum_{p,q=1}^{P}\alpha_p\alpha_q\theta_p\theta_q K(v_p^{\mathbf{L}}, v_q^{\mathbf{L}})$$

[0151] 其中，$\sum_{p=1}^{P}\theta_p\alpha_p = 0$，$\alpha_p \geq 0$，$p = 1,2, \cdots, P$，$K(v_p, v_q)$ 为 $v_p^{\mathbf{L}}$ 和 $v_q^{\mathbf{L}}$ 的核函数。

[0152] 设最优解 α^*，则最优分类面是训练样本的组合 $\{\alpha_p^*\theta_p v_p^{\mathbf{L}}\}$，$p = [1,2, \cdots, P]$。

[0153] A62：训练模块对已获得的超矢量，采用上述步骤 A61 中的支持向量机算法建立并训练语种模型。

[0154] 本发明实施例提供的用于语种识别的建模装置还包括多坐标系原点存储模块、子空间模型存储模块和语种模型存储模块，

[0155] 其中，多坐标系原点存储模块，用于存储来自多坐标系原点选择模块的坐标系原点；子空间模型存储模块，用于存储来自子空间选择模块的语种子空间和信道子空间；语种模型存储模块，用于存储来自训练模块的语种模型。

[0156] 本发明实施例提供的用于语种识别的建模装置进一步包括识别模块，利用语种模型对待测超矢量打分，识别待测语音的语言种类。

[0157] 具体的说，识别模块输入待测语音，对待测语音预处理得到特征序列，并根据坐标系选择算法和特征矢量映射算法，将特征矢量映射为待测超矢量，对待测超矢量进行投影和补偿，利用语种模型对待测超矢量打分，识别待测语音的语言种类。

[0158] 首先输入待测语音，然后采用上述步骤 A1、A2、A3 和 A5 中的算法提取超矢量。

[0159] B1：语音预处理模块根据输入的待测语音，采用上述步骤 A1、A2、A3 和 A5 中的算法提取超矢量：包括：

[0160] B11：特征提取模块对待测语音进行预处理，从预处理后的待测语音中提取特征序列，特征序列是由一系列的特征向量组成；

[0161] B12：根据多坐标系原点存储模块存储的坐标系原点，特征矢量映射模块通过坐标系选择算法和特征序列映射算法，将所述特征矢量映射成待测超矢量；

[0162] B13：根据待测超矢量以及子空间模型存储模块存储的语种子空间和信道子空间，子空间投影补偿模块利用信道子空间和语种子空间对待测超矢量进行投影和补偿，提取待测超矢量仅存在于语种子空间的部分；

[0163] B2：识别模块利用语种模型存储模块存储的语种模型，根据子空间投影补偿模块输出的超矢量进行打分，得到输出分数。其中打分函数为：

[0164] $$f(v)=\sum_{p=1}^{P}\alpha_p^*\theta_p K(v_p^{\mathbf{L}},v^{\mathbf{L}})+b^*$$ 。

[0165] B3：识别模块对输出分数进行后处理，与判决门限比较，判别该段语音的语言种类。

[0166] 根据本发明实施例提供的用于语种识别的建模装置，通过对语音信号特征序列的高维统计量有效建模，并采用子空间技术，去除了高维统计量中对识别无效的信息，提高了语种识别的正确率，又降低了在集成电路上的运算复杂度。

[0167] 本领域普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，该程序在执行时，包括方法实施例的步骤之一或其组合。

[0168] 另外，在本发明各个实施例中的各功能单元可以集成在一个处理模块中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用时，也可以存储在一个计算机可读取存储介质中。

[0169] 上述提到的存储介质可以是只读存储器，磁盘或光盘等。

[0170] 以上所述仅是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

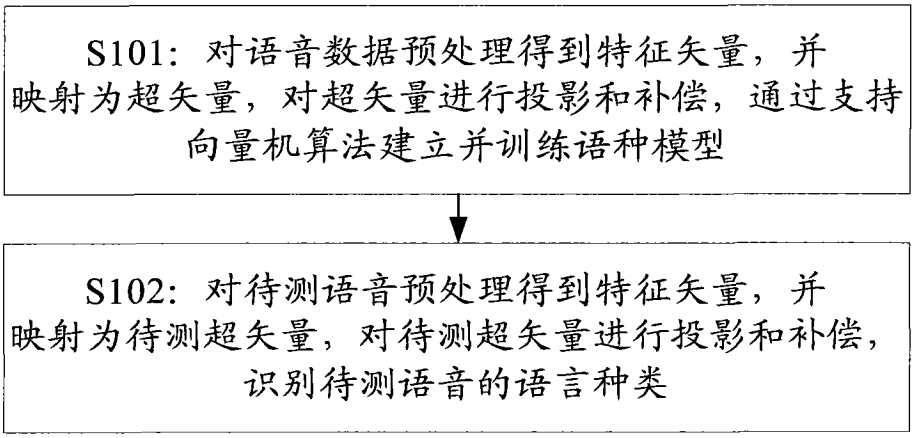

S101：对语音数据预处理得到特征矢量，并映射为超矢量，对超矢量进行投影和补偿，通过支持向量机算法建立并训练语种模型
S102：对待测语音预处理得到特征矢量，并映射为待测超矢量，对待测超矢量进行投影和补偿，识别待测语音的语言种类

图 1

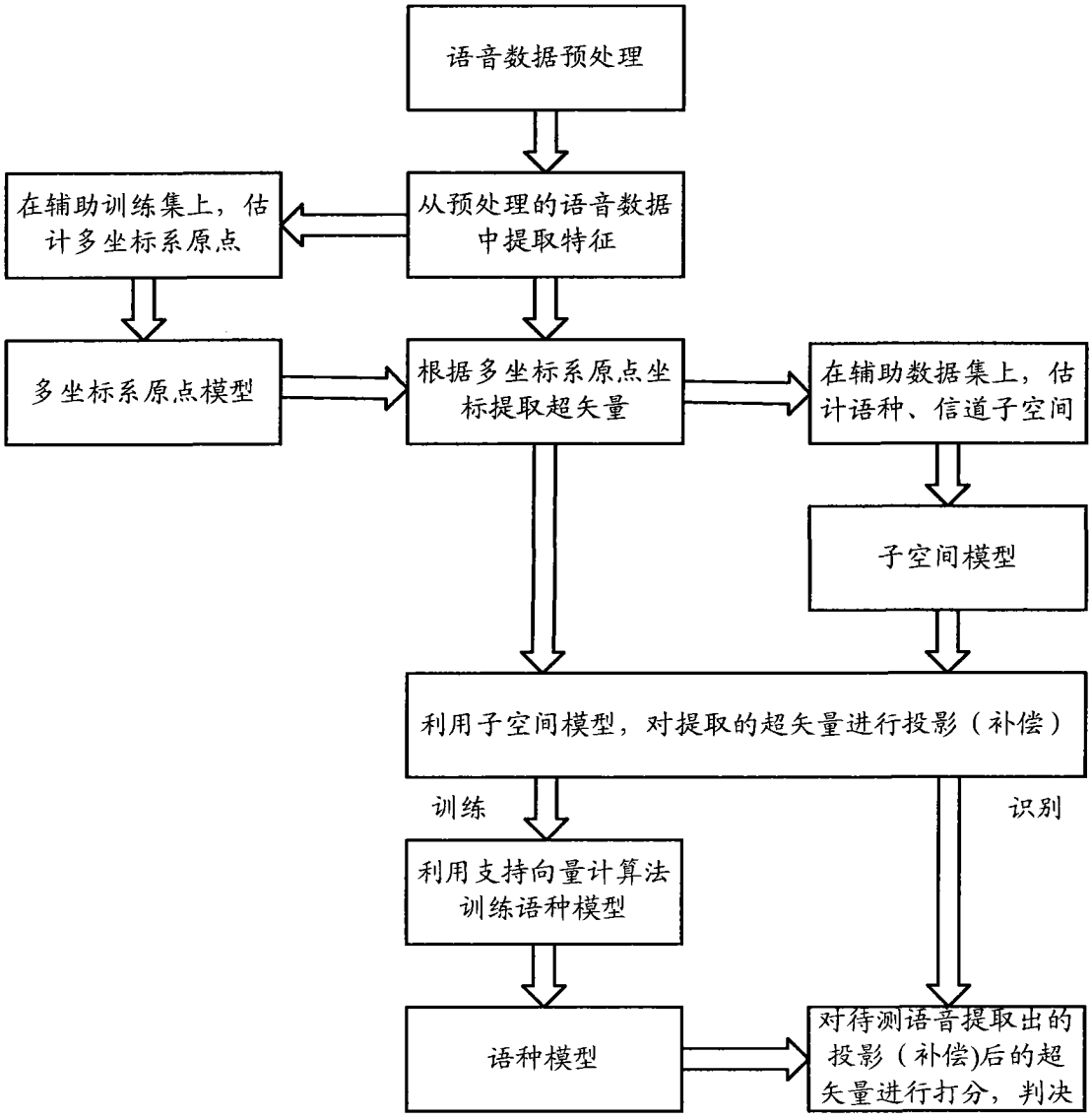

语音数据预处理
从预处理的语音数据中提取特征
在辅助训练集上，估计多坐标系原点
多坐标系原点模型
根据多坐标系原点坐标提取超矢量
在辅助数据集上，估计语种、信道子空间
子空间模型
利用子空间模型，对提取的超矢量进行投影（补偿）
训练
识别
利用支持向量计算法训练语种模型
语种模型
对待测语音提取出的投影（补偿)后的超矢量进行打分，判决

图 2

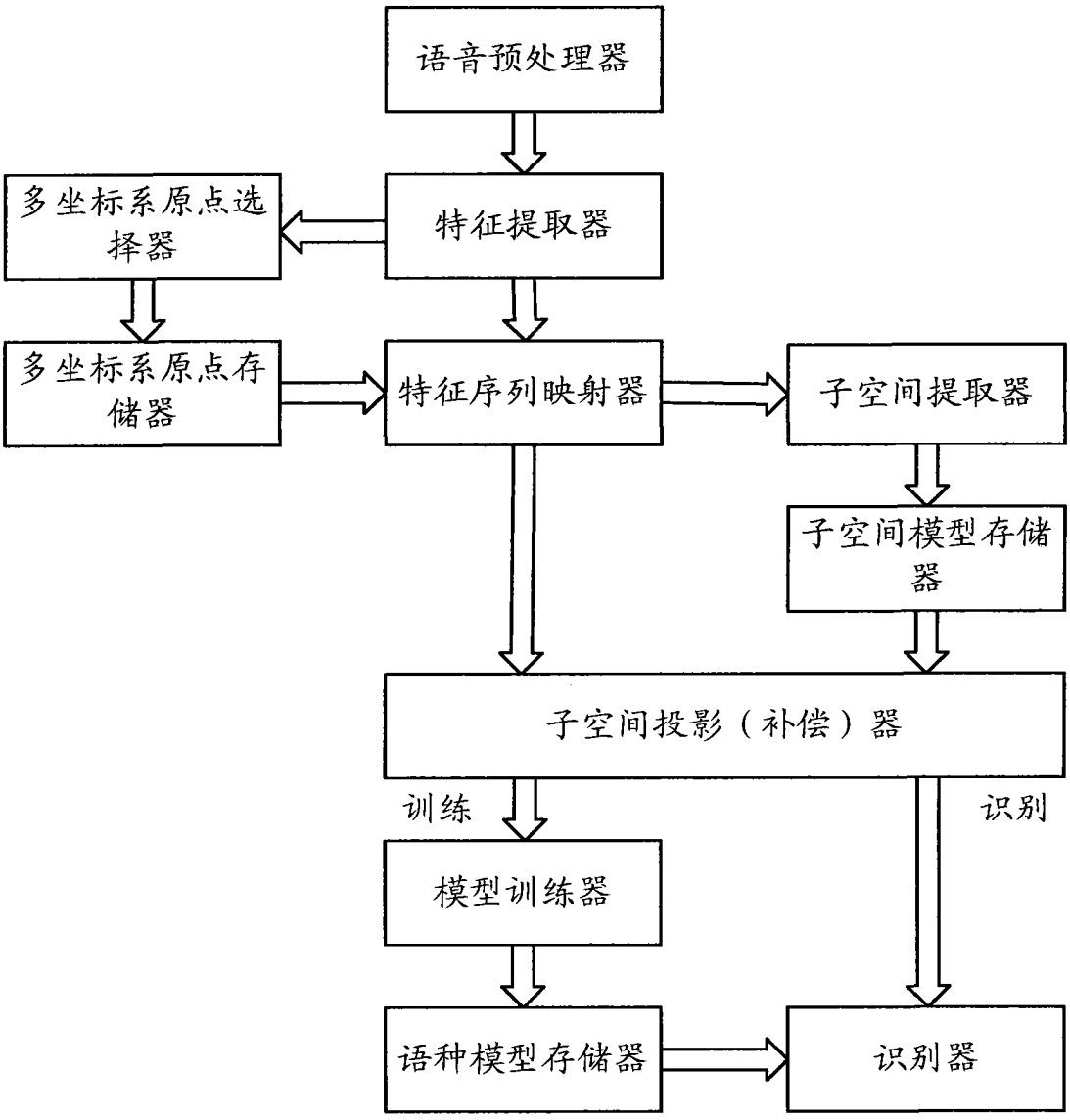

语音预处理器
特征提取器
多坐标系原点选择器
多坐标系原点存储器
特征序列映射器
子空间提取器
子空间模型存储器
子空间投影（补偿）器
训练
识别
模型训练器
语种模型存储器
识别器

图 3