



(12) 发明专利

(10) 授权公告号 CN 101809652 B

(45) 授权公告日 2013. 07. 10

(21) 申请号 200880108790. 8

(22) 申请日 2008. 09. 22

(30) 优先权数据

2007-247572 2007. 09. 25 JP

(85) PCT申请进入国家阶段日

2010. 03. 25

(86) PCT申请的申请数据

PCT/JP2008/067125 2008. 09. 22

(87) PCT申请的公布数据

W02009/041402 JA 2009. 04. 02

(73) 专利权人 日本电气株式会社

地址 日本东京都

(72) 发明人 江森正

(74) 专利代理机构 北京东方亿思知识产权代理

有限责任公司 11258

代理人 宋鹤 南霆

(51) Int. Cl.

G10L 15/02 (2006. 01)

G10L 13/033 (2013. 01)

(56) 对比文件

CN 1953050 A, 2007. 04. 25,

JP 特开平 6-214596 A, 1994. 08. 05,

JP 特开 2001-125588 A, 2001. 05. 11, 全文.

CN 1494053 A, 2004. 05. 05,

JP 特开平 11-327592 A, 1999. 11. 26,

审查员 王玥

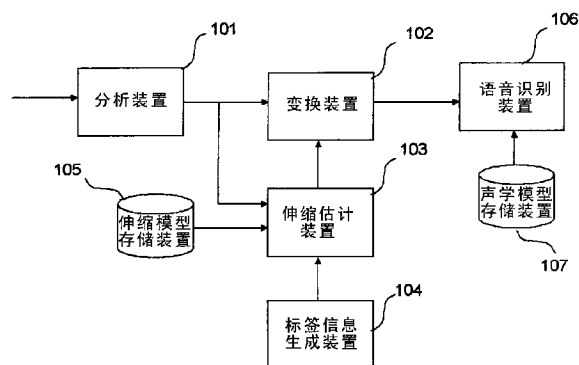
权利要求书2页 说明书7页 附图4页

(54) 发明名称

频率轴伸缩系数估计设备、系统方法

(57) 摘要

本发明提供抑制用于估计伸缩系数的成本并且估计高精度的值的伸缩系数估计系统。本发明包括：输出语音/非语音的标签信息的标签信息生成装置 104、存储表示语音/非语音的出现概率的概率模型的伸缩模型存储装置 105、利用表示语音/非语音的出现概率的概率模型、语音/非语音的标签、以及倒谱来计算频率轴方向的伸缩系数的伸缩估计装置 103。



1. 一种语音识别系统,其特征在于,包括:

伸缩系数估计装置,其利用语音 / 非语音的两个级别的概率模型,来估计表示说话人的声道长度的差异的伸缩系数;

变换装置,其使用由所述伸缩系数估计设备求出的伸缩系数来进行倒谱变换;

声学模型存储装置,其存储表示音韵信息的出现概率的概率模型;以及

语音识别装置,其利用变换后的倒谱、以及存储在所述声学模型存储装置中的所述表示音韵信息的出现概率的概率模型来进行语音识别;

所述伸缩系数估计装置包括:

标签信息生成装置,其输出语音 / 非语音的标签信息;

伸缩模型存储装置,其存储表示语音 / 非语音的出现概率的概率模型;以及

伸缩估计装置,其利用所述表示语音 / 非语音的出现概率的概率模型、所述语音 / 非语音的标签、以及倒谱,来计算频率轴方向上的伸缩系数。

2. 如权利要求 1 所述的语音识别系统,其特征在于,

所述音韵信息是音素单位的音韵信息。

3. 如权利要求 1 所述的语音识别系统,其特征在于,

还包括分析装置,该分析装置将输入语音信号截出预定区间,并对截出的预定区间实施预定变换来导出倒谱,

所述变换装置利用来自所述伸缩估计装置的伸缩系数,对由所述分析装置求出的倒谱进行变换。

4. 一种语音合成系统,其特征在于,包括:

标签信息生成装置,其输出语音 / 非语音的标签信息;

伸缩模型存储装置,其存储表示语音 / 非语音的出现概率的概率模型;

伸缩系数估计装置,其利用语音 / 非语音的两个级别的概率模型来估计表示说话人的声道长度的差异的伸缩系数,并且,其利用所述表示语音 / 非语音的出现概率的概率模型、所述语音 / 非语音的标签、以及倒谱,来计算频率轴方向上的所述伸缩系数;

变换装置,其使用由所述伸缩系数估计设备求出的伸缩系数来进行倒谱变换;以及

波形合成装置,其利用由所述变换装置进行了变换的倒谱来合成波形。

5. 如权利要求 4 所述的语音合成系统,其特征在于,

还包括分析装置,该分析装置将输入语音信号截出预定区间,并对截出的预定区间实施预定变换来导出倒谱,

所述变换装置利用来自所述伸缩估计装置的伸缩系数,对由所述分析装置求出的倒谱进行变换。

6. 如权利要求 1 所述的语音识别系统,其特征在于,

所述伸缩模型存储装置具有多个表示语音 / 非语音的出现概率的概率模型,

所述伸缩系数估计装置计算以下次数的伸缩系数,所述次数等于存储在所述伸缩模型存储装置中的模型个数,

所述语音识别系统还包括:

伸缩系数决定装置,其从所述伸缩系数估计装置所求出的伸缩系数中决定伸缩系数,并输出该伸缩系数和与该伸缩系数对应的模型信息;以及

声学模型切换装置,其基于所述模型信息来选择存储在所述声学模型存储装置中的声学模型并将所选声学模型提供给所述语音识别装置,

所述变换装置利用由所述伸缩系数决定装置决定的伸缩系数来进行倒谱变换。

7. 一种语音识别方法,其特征在于,

在存储装置中准备表示语音 / 非语音的出现概率的概率模型,

并且,所述语音识别方法包括:

输出语音 / 非语音的标签信息的步骤;

利用语音 / 非语音的两个级别的概率模型来估计表示说话人的声道长度的差异的伸缩系数,且利用所述表示语音 / 非语音的出现概率的概率模型、所述语音 / 非语音的标签、以及倒谱来计算频率轴方向上的所述伸缩系数的步骤;

利用所述伸缩系数来进行倒谱变换的步骤;以及

利用通过在所述进行倒谱变换的步骤中进行了变换的所述倒谱、以及表示音韵信息的出现概率的概率模型来进行语音识别的步骤。

8. 如权利要求 7 所述的语音识别方法,其特征在于,

还包括分析步骤,该分析步骤将输入语音信号截出预定区间,并对截出的预定区间实施预定变换来导出倒谱,

所述进行倒谱变换的步骤利用所述伸缩系数来对在所述分析步骤中求出的倒谱进行变换。

9. 一种语音合成方法,其特征在于,

在存储装置中准备表示语音 / 非语音的出现概率的概率模型,

并且,所述语音合成方法包括:

输出语音 / 非语音的标签信息的步骤;

利用语音 / 非语音的两个级别的概率模型来估计表示说话人的声道长度的差异的伸缩系数,且利用所述表示语音 / 非语音的出现概率的概率模型、所述语音 / 非语音的标签、以及倒谱来计算频率轴方向上的所述伸缩系数的步骤;

利用所述伸缩系数来进行倒谱变换的步骤;以及

利用通过所述进行倒谱变换的步骤进行了变换的倒谱来合成波形的步骤。

10. 如权利要求 7 所述的语音识别方法,其特征在于,

在存储装置中准备多个表示语音 / 非语音的出现概率的概率模型,

并且,所述语音识别方法包括以下步骤:

计算与所述存储装置中准备的所述概率模型的个数相等次数的伸缩系数;

从多个伸缩系数中决定伸缩系数,并选择与该伸缩系数对应的模型信息;

利用所述决定的伸缩系数来进行倒谱变换;

基于所选的模型信息来选择声学模型,并将所选声学模型提供给语音识别。

频率轴伸缩系数估计设备、系统方法

技术领域

[0001] 本身请主张在先日本专利申请第 2007-247572 号(2007 年 9 月 25 日申请)的优选权,所述在先申请的全部记载内容以引用的方式合并在本申请文件中。

[0002] 本发明涉及语音处理技术,尤其涉及用于对以语音在频率轴上的伸缩表示的说话人特性进行归一化的伸缩系数估计设备、倒谱变换系统、语音识别系统、波形合成系统、方法以及程序。

背景技术

[0003] 在语音识别领域中,为了提高识别性能的精度,提出了修正由说话人的声道长度的差异引起的声谱的共振频率的波动的方法。这种技术被称作声道长度归一化。作为声道长度归一化的方法,例如,在专利文件 1(日本专利公告公报 3632529 号)中,利用倒谱空间上的线性变换将共振频率的偏差模型化,由此有效地估计了偏差的大小。

[0004] 专利文件 1 中提出的声道长度归一化的构成包括:由分析语音并输出倒谱的分析部、基于倒谱估计表示伸缩程度的伸缩系数值的伸缩估计部、以及利用伸缩系数的值对倒谱进行线性变换的变换部。

[0005] 变换部中使用的线性变换利用了通过利用全通滤波器的逆变换来将频率轴上的变换表示为倒谱的线性变换的方式。在该变换中使用的参数的个数为 1。

[0006] 专利文件 1 在伸缩系数的估计中使用了在语言识别中使用的 HMM(Hidden Markov Model,隐式马尔可夫模型)。如在专利文件 1 中也记载的那样,在语言识别中使用的 HMM 是将以单词或音素等为单位的音韵信息的输出概率模型化的模型。

[0007] 例如,将日语“はい”分解成音素,作为“h a i”,并将这些每个 h,a,i 的出现概率模型化。如果例举作为经常使用的出现概率形式的正态分布,则此时,对每个音素事先计算倒谱等特征量的值的平均和方差,并在进行识别等时使用。在专利文件 1 中,为了估计伸缩系数,使用下式(1)。

$$[0008] \quad \alpha = \frac{\sum_{j=1}^J \sum_{t=1}^T \gamma_t(j) \left[\sum_{m=1}^M \frac{(c_{mt} - \mu_{mj}) \{ (m-1)c_{(m-1)t} - (m+1)c_{(m+1)t} \}}{\sigma_{mj}^2} \right]}{\sum_{j=1}^J \sum_{t=1}^T \gamma_t(j) \left[\sum_{m=1}^M \frac{\{ (m-1)c_{(m-1)t} - (m+1)c_{(m+1)t} \}^2}{\sigma_{mj}^2} \right]} \quad (1)$$

[0009] 其中,J、j 表示音素或状态的个数以及用于标别音素或状态的 ID,t 表示时刻,M、m 表示倒谱的维数和维, c_{mt} 表示 t 时刻的第 m 维的倒谱系数, μ_{mj} 、 σ_{mj} 表示音素 j、第 m 维的 HMM 中的平均向量和标准偏差的值。

[0010] 挡利用上式来估计伸缩系数的值时,需要用于标识平均向量或方差的信息。

[0011] 用于标识的 ID 信息可通过提供例如描述语音内容的单词信息等来计算。即,在上述例子的情况下,当提供日语“はい”时,可以确定“h ai”这样的音素序列信息,并能够展开成 h,a,i 各自的状态序列,还能够确定属于状态的概率分布。

[0012] 专利文件 1 :日本专利公告公报第 3632529 号 ;

[0013] 非专利文件 1 :HTKBook Ver. 3. 3、第 35 ~ 40 页、第 54 ~ 64 页、第 127 ~ 130 页。

发明内容

[0014] 将上述的专利文件 1、非专利文件 1 的全部公开内容以引用的方式合并在本申请文件中。以下,分析本发明的相关技术。

[0015] 作为获取音素序列等的方法,在预先知道发声内容的情况下可人工提供音素信息。

[0016] 另一方面,在谈话等事先无法知道说话的内容的情况下,通过人工提供音素信息需要很大成本,不现实。此时,多数采用暂时通过语音识别来自动获取的方法。

[0017] 然而,语音识别的处理需要较大的运算量,因此在有限的运算量中需要避免仅仅为了生成用于估计的音素序列信息而进行上述计算的情形。

[0018] 另外,从语音识别所获取的音素序列信息包含有错误,因此无法避免估计精度的劣化。

[0019] 因此,本发明的目的在于,提供能够以少的运算量执行当进行声道长度归一化处理时的伸缩系数的估计的伸缩系数估计设备、具有该设备的倒谱变换系统、语音识别系统、波形合成系统、方法以及程序。

[0020] 本发明的其他目的在于,提供能够提高表示声道长度的差异的伸缩系数的估计精度的伸缩系数估计设备、具有该设备的倒谱变换系统、语音识别系统、波形合成系统、方法以及程序。

[0021] 本申请公开的发明为了解决上述问题,大致具有以下构成。

[0022] 根据本发明的第一方面,提供一种伸缩系数估计设备,包括:标签信息生成装置,其输出语音/非语音的标签信息;伸缩模型存储装置,其存储表示语音/非语音的出现概率的概率模型;伸缩估计装置,其利用表示语音/非语音的出现概率的概率模型、语音/非语音的标签、以及倒谱,来计算频率轴方向上的伸缩系数。

[0023] 根据本发明的另一方面,提供一种倒谱变换系统,包括:所述伸缩系数估计设备;以及使用伸缩系数来进行倒谱变换的变换装置。

[0024] 根据本发明的又一方面,提供一种语音识别系统,包括:所述倒谱变换系统;声学模型存储装置,其存储表示音素单位等音韵信息的出现概率的概率模型;以及语音识别装置,其利用变换后的倒谱、以及表示音素单位等音韵信息的出现概率的概率模型来进行语音识别。

[0025] 根据本发明的再一方面,提供一种波形合成系统,其包括利用通过所述倒谱变换系统变换后的倒谱来合成波形的波形合成装置。

[0026] 在本发明的再一方面中,也可以如下构成:所述伸缩模型存储装置具有多个表示语音/非语音的出现概率的概率模型,所述伸缩估计装置计算以下次数的伸缩系数,所述次数等于存储在所述伸缩模型存储装置中的模型个数,并且,

[0027] 所述语音识别系统还包括:伸缩系数决定装置,其从所述伸缩估计装置所求出的伸缩系数中决定伸缩系数,并输出该伸缩系数和与该伸缩系数对应的模型信息;以及声学模型切换装置,其基于所述模型信息来选择存储在所述声学模型存储装置中的声学模型,

并将所选声学模型提供给声学识别装置,并且,

[0028] 所述变换装置利用由所述伸缩系数决定装置决定的伸缩系数来进行倒谱变换。

[0029] 根据本发明,提供一种伸缩系数估计方法,其中在存储装置中准备表示语音/非语音的出现概率的概率模型,并且,

[0030] 所述伸缩系数估计方法包括以下步骤:

[0031] 输出语音/非语音的标签信息;以及

[0032] 利用所述表示语音/非语音的出现概率的概率模型、所述语音/非语音的标签、以及倒谱来计算频率轴方向的伸缩系数。

[0033] 根据本发明,提供一种倒谱变换方法,其包括:利用通过所述伸缩系数估计方法求出的所述伸缩系数来进行倒谱变换的步骤。

[0034] 根据本发明,提供一种语音识别方法,其包括:利用通过所述倒谱变换方法变换后的所述倒谱、以及表示音韵信息的出现概率的概率模型来进行语音识别的步骤。

[0035] 根据本发明,提供一种波形合成方法,其包括利用通过倒谱变换方法变换后的所述倒谱来合成波形的步骤。

[0036] 根据本发明,提供一种伸缩估计程序,其使计算机执行以下处理:在存储装置中准备表示语音/非语音的出现概率的概率模型;输出语音/非语音的标签信息;以及利用所述表示语音/非语音的出现概率的概率模型、所述语音/非语音的标签、以及倒谱来计算频率轴方向上的伸缩系数。

[0037] 根据本发明,提供一种倒谱变换程序,其使计算机执行以下处理:利用通过所述伸缩系数估计程序求出的所述伸缩系数来进行倒谱变换。

[0038] 根据本发明,提供一种语音识别程序,其使计算机执行以下处理:利用通过所述倒谱变换程序变换后的所述倒谱、以及表示音韵信息的出现概率的概率模型来进行语音识别。

[0039] 根据本发明,提供一种波形合成程序,其使计算机执行以下处理:利用通过倒谱变换程序变换后的所述倒谱来合成波形。

[0040] 发明效果

[0041] 根据本发明,通过使用以使其表示说话人的全局信息的方式被设计的HMM、即归纳到语音和非语音两个级别的概率模型,能够以少的运算量执行当进行声道长度归一化处理时的伸缩系数的估计。

[0042] 根据本发明,能够提高表示声道长度的差异的伸缩系数的估计精度。

附图说明

[0043] 图1是示出本发明第一实施例的构成的框图;

[0044] 图2是示出本发明第一实施例的动作的流程图;

[0045] 图3是示出本发明第二实施例的构成的框图;

[0046] 图4是示出本发明第二实施例的动作的流程图。

[0047] 标号说明

[0048] 101 分析装置

[0049] 102 变换装置

- [0050] 103 伸缩估计装置
- [0051] 104 标签信息生成装置
- [0052] 105 伸缩模型存储装置
- [0053] 106 语音识别装置
- [0054] 107 声学模型存储装置
- [0055] 201 伸缩系数决定装置
- [0056] 202 声学模型切换装置

具体实施方式

[0057] 本发明包括：输出语音 / 非语音的标签信息的标签信息生成装置 (104)；存储表示语音 / 非语音的出现概率的概率模型的伸缩模型存储装置 (107)；利用表示语音 / 非语音的出现概率的概率模型、语音 / 非语音的标签、以及倒谱来计算频率轴方向上的伸缩系数的伸缩估计装置 (103)。

[0058] 本发明在用于抑制由于估计说话人的声道长度的差异而产生的共振峰频率的波动的、频率轴上的伸缩函数的伸缩系数的估计中，通过使用以使其表示说话人的全局信息的方式被设计的 HMM、即归纳到语音和非语音两个级别的概率模型，可以以少的运算量估计伸缩系数。

[0059] 在本发明中，通过归纳到语音 / 非语音两个级别，不需要在估计中使用不确定的音素信息。根据本发明，语音 / 非语音的音素信息不需要事先进行语音识别等处理，因此不需要运算量。

[0060] 在估计作为说话人的全局信息的声道长度的差异时，以语音和非语音这样的大的单位进行估计的做法好于按音素单位精细地比较其差异的做法。另外，所提供的音素信息等也一样越精细就越容易包含错误，与此相对，归类为语音和非语音时不会出错，因此可高精度地进行估计而不受所提供的音素信息的影响。

[0061] 因此，与使用现有方法的情况相比能够高精度地估计表示声道长度的差异的伸缩系数。以下，利用实施例进行详细说明。

[0062] 实施例

[0063] 图 1 是示出本发明一个实施例的构成的图。下面，对用于语音识别中的特征量变换装置进行说明。参考图 1 可知，本实施例包括：分析装置 101、变换装置 102、伸缩估计装置 103、标签信息生成装置 104、伸缩模型存储装置 105、语音识别装置 106、声学模型存储装置 107。

[0064] 虽然不进行特别限定，但分析装置 101 输出如下求出的 mel 倒谱，即：例如从采样频率为 44.1kHz、并以一个样本对应 16 位的方式进行了 A/D 转换的输入语音信号中截出 10 毫秒 (msec) 左右的固定区间，对该固定区间进行预加重 (preemphasis) 处理、FFT (快速傅立叶变换) 处理、滤波器组 (filter bank) 处理，然后通过进行余弦变换来计算出该 mel 倒谱。

[0065] 声学模型存储装置 107 中存储有通过将考虑了前后音素上下文的三音子 (Triphone) 作为音素的 HMM 等来表述的声学模型，这种声学模型例如被记载在剑桥大学发表的非专利文件 1 (HTKBook Ver. 3.3 的第 35 ~ 40 页) 中。

[0066] 伸缩模型存储装置 105 使用只表示语音 / 非语音两个级别的 HMM。将该 HMM 称作 GMM(Gaussian Mixture Model, 高斯混合模型), 以与存储在声学模型存储装置 107 中的 HMM 进行区别。

[0067] 相对于 HMM 使用考虑了音素上下文的三音子等作为音素, GMM 根据针对每个说话人来利用频率轴上的伸缩进行修正的声道长度归一化的特征, 只保持两个级别, 而并不使用以精细的音素级别构建模型的 HMM。

[0068] 对 GMM 的学习进行说明。首先, 使用上述的倒谱、如非专利文件 1(第 54 ~ 64 页)中所记载的功率差或前后时刻差, 作为语音的声学特征。

[0069] 使用标签数据来计算前向 / 后向概率, 由此将特征和标签数据对应起来, 其中标签数据利用声学特征和与其对应的写入文本而得。

[0070] 在本实施例中, 使用“sil voice sil”这样的两个级别的标签信息。其中, “sil”表示非语音, “voice”表示语音。

[0071] 利用相对应的特征量和标签数据来计算表示每个对应的标签的特征量分布的参数。

[0072] 表示分布的函数多采用正态分布。此时, 参数是平均值和方差。

[0073] 考虑到此后的使用了 HMM 的语音识别, 用于 GMM 学习的语音最好使用用于 HMM 学习的语音数据。

[0074] 标签信息生成装置 104 与学习时同样地生成并输出“sil voice sil”的标签数据。

[0075] 此时, 通过根据语音长度来改变 voice 标签数, 能够更加正确地进行对应。当语音数据长时, 可推测出具有更多的语音区间, 因此根据语音数据的长度来增加 voice 标签数, 由此可获得减少发生不必要的迁移的概率的效果。

[0076] 伸缩估计装置 103 利用上式 (1) 来计算表示频率轴上的伸缩的参数值。

[0077] 上式 (1) 中的 J 表示级别数, 在本实施例中, 该 J 为 1。

[0078] 在式 (1) 中, T 表示特征量的个数。在此情况下, 该 T 是将语音长度以进行截取的时间单位进行分割的情况下的个数。

[0079] $y_t(j)$ 表示 t 时刻的语音级别的优势度 (占有度数)。该值可通过上述的前向 / 后向算法来计算。

[0080] 作为倒谱系数的维数的 c_{mt} 是 t 时刻的第 m 维倒谱系数的值。

[0081] μ_{mj} 表示语音级别的第 m 维平均向量、 σ_{mj} 表示语音级别的第 m 维标准偏差。

[0082] 伸缩系数 α 与专利文件 1 中所述的一样是表示频率轴的伸缩程度的变形参数, 其值由伸缩估计装置 103 计算。

[0083] 变换装置 102 使用由伸缩估计装置 103 算出的伸缩系数 α 来变换倒谱, 并输出变换后的倒谱。倒谱的变换与专利文件 1 同样地通过下式 (2) 进行。

$$[0084] \quad \hat{c}_m^{(i)} = \begin{cases} c_{-i} + \alpha \hat{c}_0^{(i-1)}, & m = 0 \\ (1 - \alpha^2) \hat{c}_{m-1}^{(i-1)} + \alpha \hat{c}_1^{(i-1)}, & m = 1 \\ \hat{c}_{m-1}^{(i-1)} + \alpha (\hat{c}_m^{(i-1)} - \hat{c}_{m-1}^{(i)}), & m \geq 2 \end{cases} \quad (2)$$

[0085] $i = -\infty, \dots, -1, 0.$

[0086] 其中,标上 $\hat{\cdot}$ 的 c 表示变换后的倒谱。

[0087] 语音识别装置 106 将通过变换装置 102 变换的倒谱作为特征量并利用 HMM 进行识别。

[0088] 作为本实施例的变形,也可以构成为具有波形合成装置,该波形合成装置利用通过变换装置 102 变换的倒谱来进行波形的合成。即,该伸缩估计装置 103 也能够应用于语音合成装置。

[0089] 当然,分析装置 101、变换装置 102、伸缩估计装置 103、语音识别装置 106 也可以通过由计算机(处理器、或者数字信号处理器)执行的程序来实现其功能。

[0090] 图 2 是用于说明本发明第一实施例的处理的流程图。参考图 1 和图 2,详细说明本发明第一实施例的整体动作。首先,输入语音数据信号(图 2 的步骤 A101),通过分析装置 101 计算 mel 倒谱(步骤 A102)。

[0091] 伸缩估计装置 103 利用从分析装置 101 输出的 mel 倒谱、伸缩模型存储装置 105 的 GMM(A104)、以及来自标签信息生成装置 104 的语音/非语音的标签信息(A105)来计算伸缩参数(步骤 A103)。

[0092] 变换装置 102 利用上式(2)中的任一变换函数并使用伸缩系数 α 来变换倒谱并输出变换后的倒谱(步骤 A106)。

[0093] 语音识别装置 106 将通过变换装置 102 变换的倒谱作为特征量并利用声学模型存储装置 107 的 HMM 进行识别(步骤 A107)。

[0094] 图 3 是示出本发明第二实施例的构成的图。本实施例在上述的第一实施例的基础上还具备伸缩系数决定装置 201 和声学模型切换装置 202。

[0095] 伸缩模型存储装置 105 存储多个与在上述第一实施例中进行说明的 GMM 相同的模型。其中,多个是指所学习的语音数据不同的多个 GMM。例如,可以想到男性和女性这两个 GMM 的情况、或者按照每个说话人进行学习的 GMM 等。

[0096] 伸缩估计装置 103 与所述第一实施例同样地计算伸缩系数(变形参数) α 。

[0097] 在本实施例的情况下,伸缩估计装置 103 的伸缩估计动作被重复多次,变形参数被计算以下次数,所述次数等于存储在伸缩模型存储装置 105 中的 GMM 的个数。另外,为了计算变形参数也多次计算前向/后向概率。

[0098] 伸缩系数决定装置 201 选择由伸缩估计装置 103 算出的每个 GMM 的前向/后向概率中最高的 GMM,输出 GMM 信息和变形参数 α 的值。

[0099] GMM 信息是指被选中的 GMM 的信息,例如在多个 GMM 为男性和女性的情况下,输出男性、女性中的任一个。

[0100] 另外,在 GMM 针对每个说话人的情况下,输出说话人姓名。声学模型切换装置 202 基于伸缩系数决定装置 201 所输出的 GMM 信息来选择存储在声学模型存储装置 107 中的 HMM,并输出所选 HMM。

[0101] 图 4 是用于说明第二实施例的处理的流程图。参考图 3 和图 4,详细说明本发明第二实施例的整体动作。首先,输入语音数据信号(图 4 的步骤 A201),由分析装置 101 计算 mel 倒谱(步骤 A202)。

[0102] 伸缩估计装置 103 利用从分析装置 101 输出的模式(mel 倒谱)、以及来自标签信息生成装置 104 的语音/非语音标签信息(A205),计算以下次数的伸缩参数,所述次数等于

存储在伸缩模型存储装置 105 中的 GMM 的个数 (A204) (步骤 A203)。

[0103] 伸缩系数决定装置 201 输出所选 GMM 的信息和伸缩系数 (变形参数) α 的值 (步骤 A206)。

[0104] 变换装置 102 利用上式 (2) 的任一变换函数并使用伸缩系数 α 来变换倒谱并输出变换后的倒谱 (步骤 A207)。

[0105] 声学模型切换装置 202 基于伸缩系数决定装置 201 所输出的 GMM 信息, 选择存储在声学模型存储装置 107 中的 HMM 并输出所选 HMM (步骤 A208)。

[0106] 语音识别装置 106 将通过变换装置 102 变换的倒谱作为特征量并利用声学模型存储装置 107 的 HMM 来进行识别 (步骤 A209)。

[0107] 根据本实施例, 通过如上构成, 能够同时进行面向特定说话人的模型的选择和变形参数的估计。

[0108] 作为本实施例的变形, 也可以构成为具有波形合成装置, 该波形合成装置利用通过变换装置 102 变换的倒谱来进行波形的合成。即, 伸缩估计装置 103 也可以应用于语音合成装置。

[0109] 当然, 在本实施例中, 分析装置 101、变换装置 102、伸缩估计装置 103、语音识别装置 106、伸缩系数决定装置 201、声学模型切换装置 202 也可以通过由计算机 (处理器、或者数字信号处理器) 执行的程序来实现其功能。

[0110] 根据本发明, 也可以应用于进行语音识别的语音识别装置、用于在计算机中实现语音识别装置的程序。

[0111] 在本发明的全部公开内容 (包括权利要求书) 的范围内, 并基于其基本技术构思, 可对实施方式乃至实施例进行变更 / 调整。另外, 也可以在本发明的权利要求书的范围内对各种公开构成要素进行多种组合以及选择。即, 本发明显然包括本领域的普通技术人员根据包含权利要求书在内的全部公开内容以及技术构思能够得到的各种变形和修正。

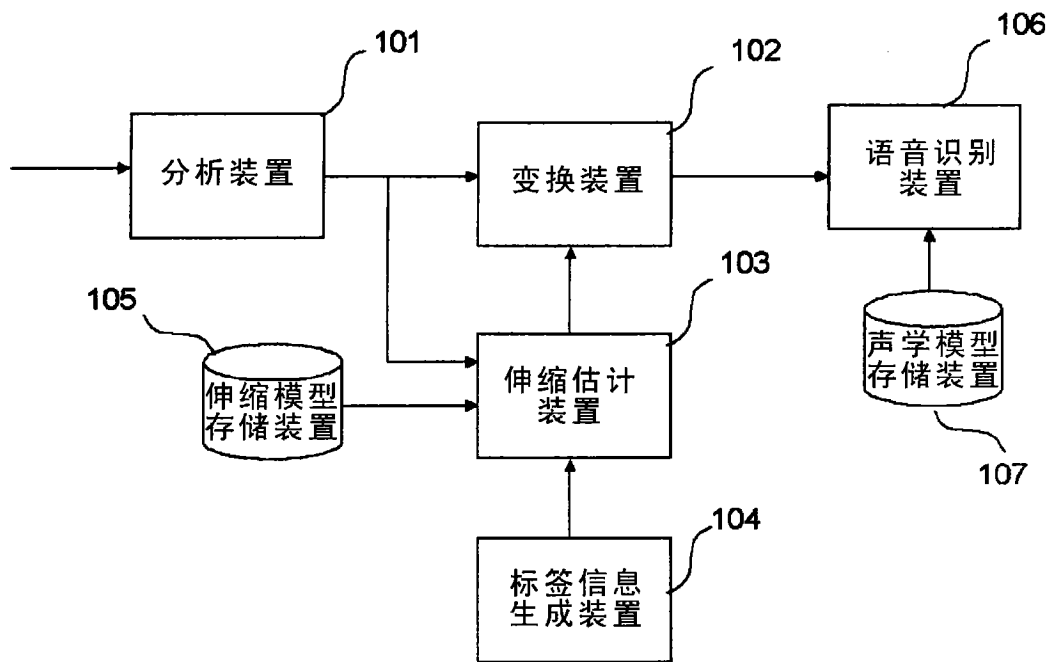


图 1

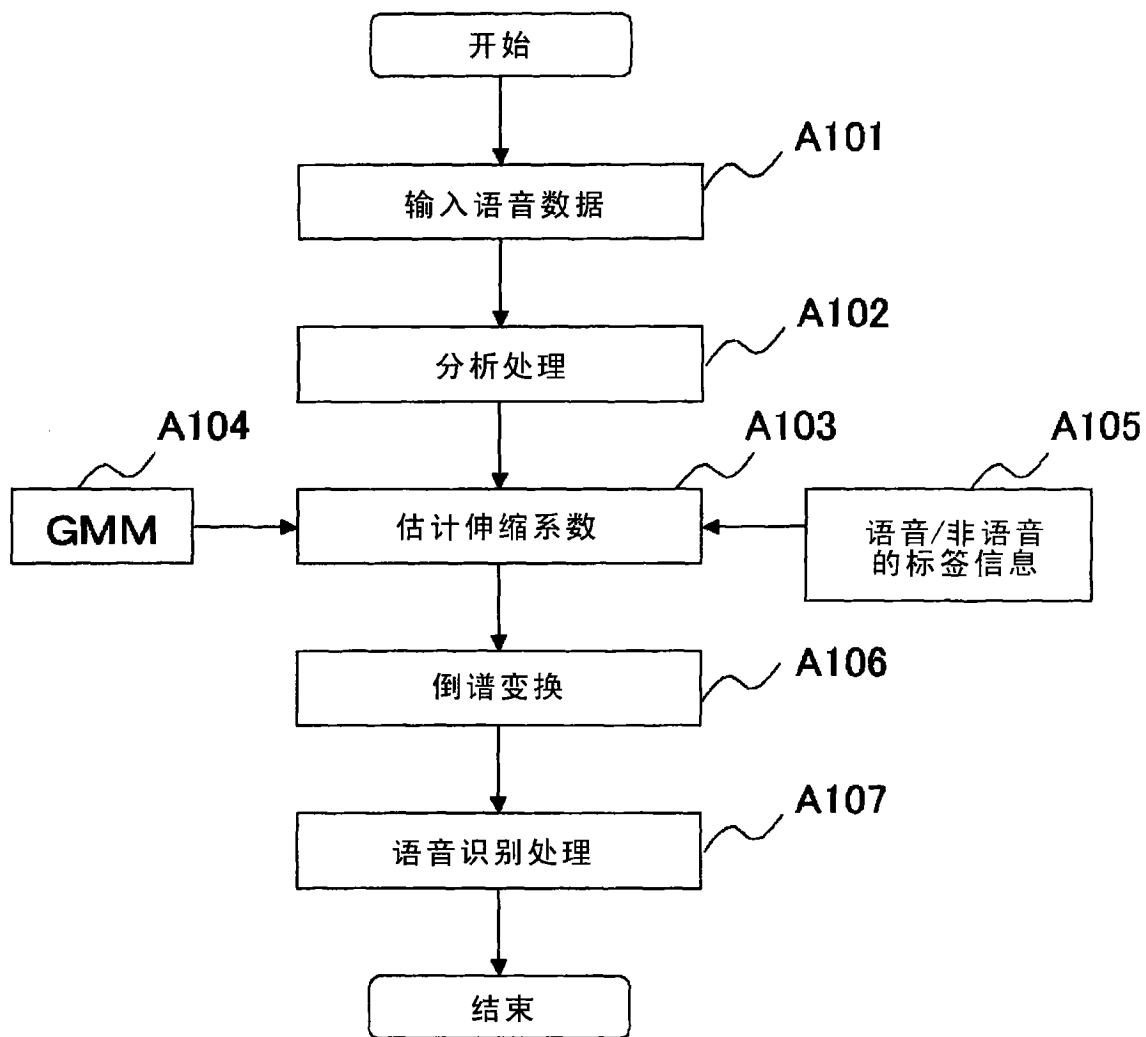


图 2

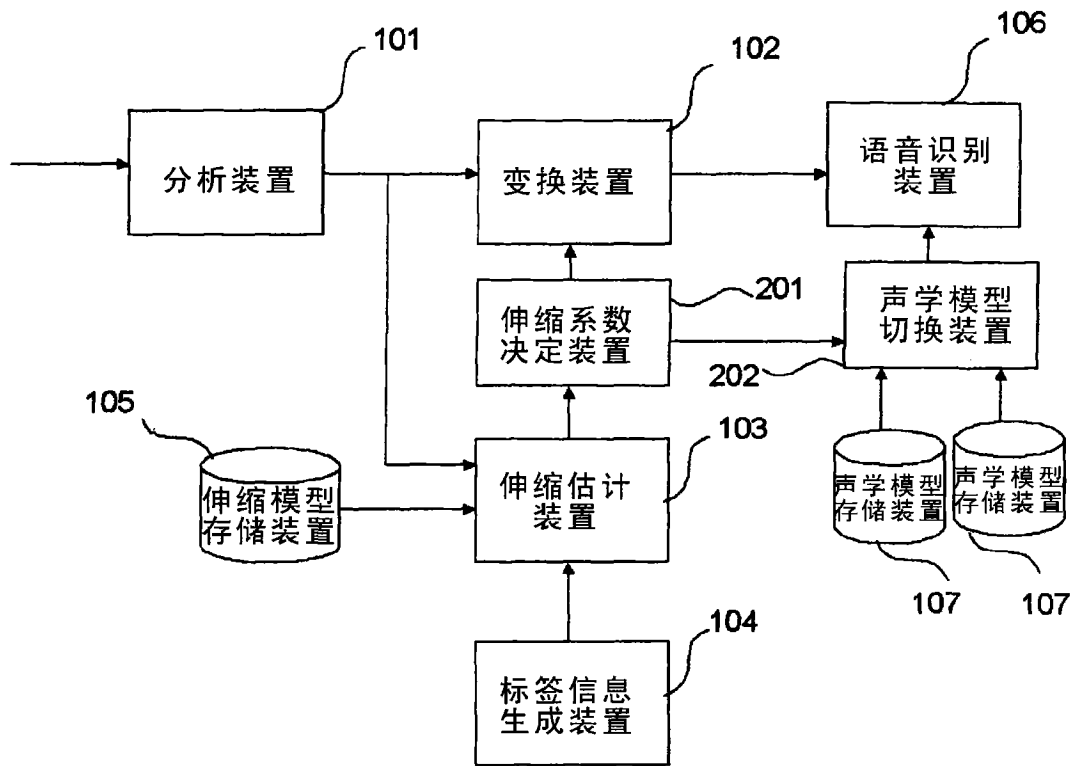


图 3

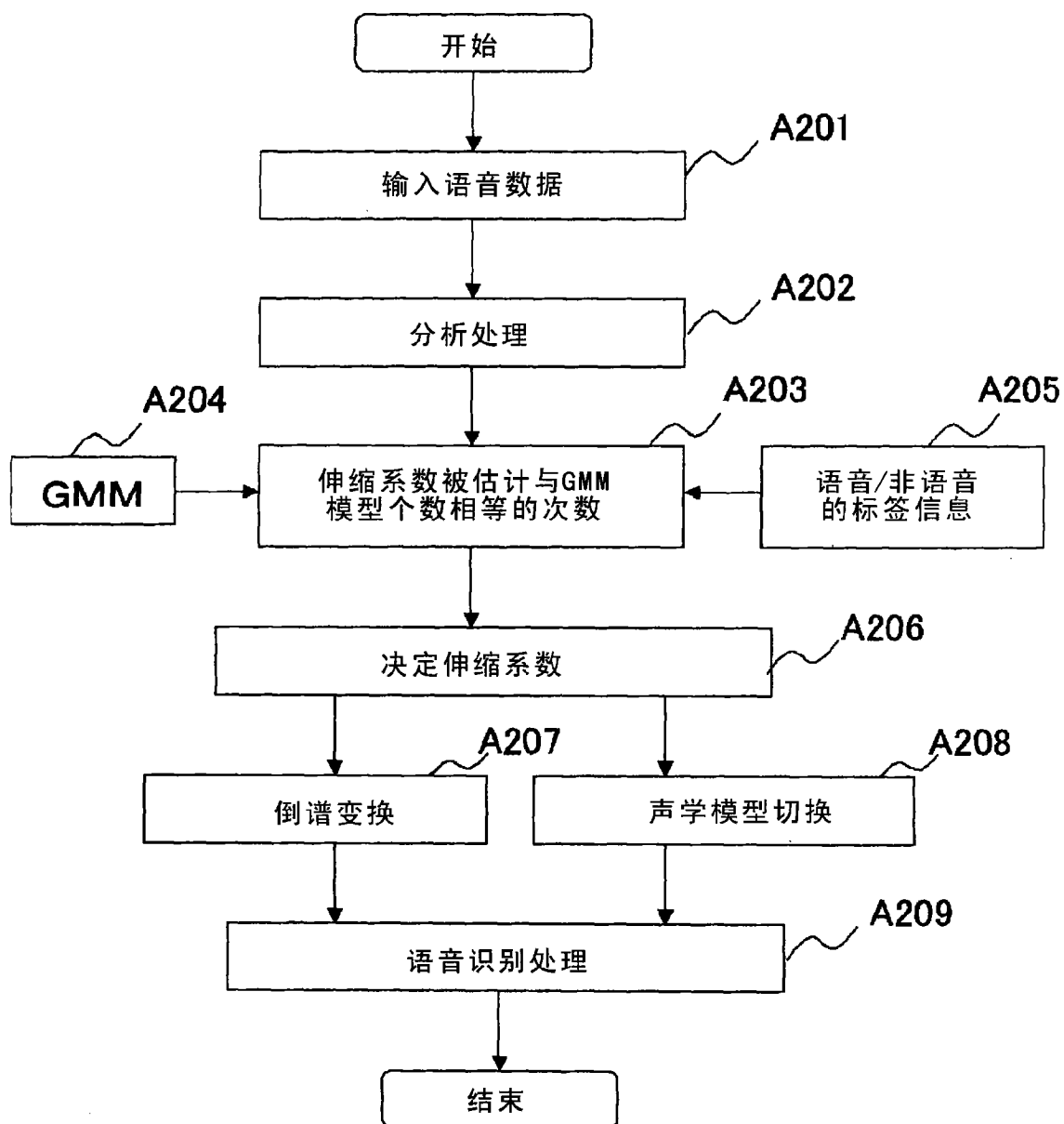


图 4