



(12) 发明专利

(10) 授权公告号 CN 101154380 B

(45) 授权公告日 2011.01.26

(21) 申请号 200610141241.3

CN 1247363 A, 2000.03.15, 说明书第2页第2段至说明书第9页第1段和图3、4.

(22) 申请日 2006.09.29

WO 99/54846 A1, 1999.10.28, 全文.

(73) 专利权人 株式会社东芝

JP 特开 2005-345683 A, 2005.12.15, 全文.

地址 日本东京都

审查员 万济萍

(72) 发明人 栾剑 郝杰

(74) 专利代理机构 北京市中咨律师事务所

11247

代理人 李峥 杨晓光

(51) Int. Cl.

G10L 17/00(2006.01)

G10L 15/00(2006.01)

G10L 15/08(2006.01)

(56) 对比文件

US 2002/0116189 A1, 2002.08.22, 全文.

US 2006/0178885 A1, 2006.08.10, 全文.

US 6107935 A, 2000.08.22, 全文.

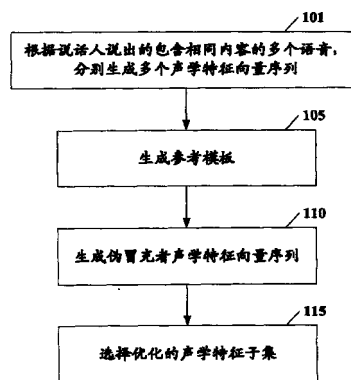
权利要求书 3 页 说明书 7 页 附图 5 页

(54) 发明名称

说话人认证的注册及验证的方法和装置

(57) 摘要

本发明提供了说话人认证的注册方法和装置、说话人认证的验证方法和装置、以及说话人认证系统。上述说话人认证的注册方法,包括:根据说话人说出的包含相同内容的多个语音,分别生成多个声学特征向量序列;根据上述多个声学特征向量序列生成一个参考模板;根据一个码本,为上述多个声学特征向量序列的每一个分别生成相应的伪冒充者声学特征向量序列,其中上述码本包含多个码字以及每个码字对应的特征向量;以及根据上述多个声学特征向量序列、上述参考模板和上述多个伪冒充者声学特征向量序列,选择优化的声学特征子集。



1. 一种说话人认证的注册方法,包括:

根据说话人说出的包含相同内容的多个语音,分别生成多个声学特征向量序列;

根据上述多个声学特征向量序列生成一个参考模板;

根据一个码本,为上述多个声学特征向量序列的每一个分别生成相应的伪冒充者声学特征向量序列,其中上述码本包含多个码字以及每个码字对应的特征向量;以及

根据上述多个声学特征向量序列、上述参考模板和上述多个伪冒充者声学特征向量序列,选择优化的声学特征子集。

2. 根据权利要求1所述的说话人认证的注册方法,其中,为上述多个声学特征向量序列的每一个分别生成相应的伪冒充者声学特征向量序列的步骤包括:

为该声学特征向量序列中的每个特征向量指定一个码字,构成一个与该声学特征向量序列对应的码字序列;以及

根据上述码本中的码字和对应的特征向量,将上述码字序列转换为对应的特征向量序列,作为上述伪冒充者声学特征向量序列。

3. 根据权利要求2所述的说话人认证的注册方法,其中,为该声学特征向量序列中的每个特征向量指定一个码字的步骤包括:

为该声学特征向量序列中的每个上述特征向量查找码本中与之最接近的特征向量;以及

将上述最接近的特征向量对应的码字指定给该声学特征向量序列中的该特征向量。

4. 根据权利要求1所述的说话人认证的注册方法,其中,选择优化的声学特征子集的步骤包括:

遍历每一种可能的声学特征子集;以及

选择使得相对于上述多个伪冒充者声学特征向量序列上述参考模板的识别率最高的声学特征子集作为上述优化的声学特征子集。

5. 根据权利要求1所述的说话人认证的注册方法,其中,选择优化的声学特征子集的步骤包括:

遍历每一种可能的声学特征子集;

按照该声学特征子集,分别计算上述多个声学特征向量序列和上述参考模板的DTW距离 $d_e(i)$ ,以及上述多个伪冒充者声学特征向量序列和上述参考模板的DTW距离 $d_p(i)$ ;以及

选择上述多个声学特征向量序列和上述参考模板的DTW距离 $d_e(i)$ 与上述多个伪冒充者声学特征向量序列和上述参考模板的DTW距离 $d_p(i)$ 的比值最小的声学特征子集,作为上述优化的声学特征子集。

6. 根据权利要求1所述的说话人认证的注册方法,其中,选择优化的声学特征子集的步骤包括:

遍历每一种可能的声学特征子集;

按照该声学特征子集,分别计算上述多个声学特征向量序列和上述参考模板的DTW距离 $d_e(i)$ ,以及上述多个伪冒充者声学特征向量序列和上述参考模板的DTW距离 $d_p(i)$ ;以及

选择上述多个声学特征向量序列和上述参考模板的DTW距离 $d_e(i)$ 与上述多个伪冒充

者声学特征向量序列和上述参考模板的 DTW 距离  $dp(i)$  的差、除以上述多个声学特征向量序列和上述参考模板的 DTW 距离  $de(i)$  与上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $dp(i)$  的和的比值最小的声学特征子集,作为上述优化的声学特征子集。

7. 根据权利要求 4-6 的任意一项所述的说话人认证的注册方法,其中,上述遍历每一种可能的声学特征子集的步骤,是在指定的范围内进行的。

8. 根据权利要求 7 所述的说话人认证的注册方法,其中,上述指定的范围是声学特征的个数大于一个指定的数的可能的声学特征子集。

9. 根据权利要求 1 所述的说话人认证的注册方法,进一步包括:

根据上述优化的声学特征子集,对上述参考模板中的声学特征向量的维数进行压缩。

10. 根据权利要求 1 所述的说话人认证的注册方法,进一步包括:

根据上述码本,对上述参考模板中的声学特征向量的个数进行压缩。

11. 一种说话人认证的注册装置,包括:

声学特征提取单元,其根据说话人说出的语音,生成声学特征向量序列;

模板生成单元,其根据说话人说出的包含相同内容的多个语音的多个声学特征向量序列,生成参考模板;

伪冒充者数据生成单元,其根据一个码本,为上述多个声学特征向量序列的每一个分别生成相应的伪冒充者声学特征向量序列,其中上述码本包含多个码字以及每个码字对应的特征向量;以及

优化单元,其根据上述多个声学特征向量序列、上述参考模板和上述多个伪冒充者声学特征向量序列,选择优化的声学特征子集。

12. 根据权利要求 11 所述的说话人认证的注册装置,其中,伪冒充者数据生成单元包括:

向量-码字转换单元,其为该声学特征向量序列中的每个特征向量指定一个码字,构成一个与该声学特征向量序列对应的码字序列;

码字-向量转换单元,其根据上述码本中的码字和对应的特征向量,将上述码字序列转换为对应的特征向量序列,作为上述伪冒充者声学特征向量序列。

13. 根据权利要求 12 所述的说话人认证的注册装置,其中,向量-码字转换单元,为该声学特征向量序列中的每个上述特征向量查找码本中与之最接近的特征向量;并且将上述最接近的特征向量对应的码字指定给该声学特征向量序列中的该特征向量。

14. 根据权利要求 11 所述的说话人认证的注册装置,其中,优化单元,遍历每一种可能的声学特征子集;以及选择使得相对于上述多个伪冒充者声学特征向量序列上述参考模板的识别率最高的声学特征子集作为上述优化的声学特征子集。

15. 根据权利要求 11 所述的说话人认证的注册装置,其中,优化单元,遍历每一种可能的声学特征子集;按照该声学特征子集,分别计算上述多个声学特征向量序列和上述参考模板的 DTW 距离  $de(i)$ ,以及上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $dp(i)$ ;以及选择上述多个声学特征向量序列和上述参考模板的 DTW 距离  $de(i)$  与上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $dp(i)$  的比值最小的声学特征子集,作为上述优化的声学特征子集。

16. 根据权利要求 11 所述的说话人认证的注册装置,其中,优化单元,遍历每一种可能

的声学特征子集;按照该声学特征子集,分别计算上述多个声学特征向量序列和上述参考模板的DTW距离 $d_e(i)$ ,以及上述多个伪冒充者声学特征向量序列和上述参考模板的DTW距离 $d_p(i)$ ;以及选择上述多个声学特征向量序列和上述参考模板的DTW距离 $d_e(i)$ 与上述多个伪冒充者声学特征向量序列和上述参考模板的DTW距离 $d_p(i)$ 的差、除以上述多个声学特征向量序列和上述参考模板的DTW距离 $d_e(i)$ 与上述多个伪冒充者声学特征向量序列和上述参考模板的DTW距离 $d_p(i)$ 的和的比值最小的声学特征子集,作为上述优化的声学特征子集。

17. 根据权利要求14-16的任意一项所述的说话人认证的注册装置,其中,优化单元,在指定的范围内遍历每一种可能的声学特征子集。

18. 根据权利要求17所述的说话人认证的注册装置,其中,上述指定的范围是声学特征的个数大于一个指定的数的可能的声学特征子集。

19. 根据权利要求11所述的说话人认证的注册装置,进一步包括:

压缩单元,其根据上述优化的声学特征子集,对上述参考模板中的声学特征向量的维数进行压缩。

20. 根据权利要求11所述的说话人认证的注册装置,进一步包括:

压缩单元,其根据上述码本,对上述参考模板中的声学特征向量的个数进行压缩。

## 说话人认证的注册及验证的方法和装置

### 技术领域

[0001] 本发明涉及信息处理技术,具体地涉及说话人认证 (speaker authentication) 的技术。

### 背景技术

[0002] 利用每个人说话时的发音特点可以识别出不同的说话人,从而可以进行说话人的认证。在 K. Yu, J. Mason, J. Oglesby 发表的文章“Speaker recognition using hidden Markov models, dynamic time warping and vector quantisation”(Vision, Image and Signal Processing, IEE Proceedings, Vol. 142, Oct. 1995, pp. 313-18) 中介绍了常见的三种说话人识别引擎技术:HMM, DTW 和 VQ。

[0003] 通常,说话人认证的过程包括注册 (enrollment) 和验证 (verification) 两个阶段。在注册阶段,根据说话人(用户)本人朗读的包含密码的语音,生成该说话人的说话人模板;在验证阶段,根据说话人模板判断测试语音是否为该说话人本人说出的相同密码的语音。因此,说话人模板的质量对于整个认证过程非常重要。

[0004] 对于一个基于动态时间弯折技术的说话人验证系统,需要从每帧语音提取出一定数量的特征以获得可靠的性能。通常对于所有说话人都是用相同的方法提取特征而忽视了各个说话人的不同特性。因此,出现了为每个说话人定制一套最优特征的方案。这种最优特征集是通过在特征里选取合适的特征子集实现的。通过这种方法,系统的性能可以得到提高,而同时存储说话人模板需要的空间也降低了。然而制定特征选择的有效标准常常是个难题,尤其在可使用的信息非常少的时候。

[0005] 已知一种最优化方法可以根据其两个组成部分来唯一确定:性能标准和搜索方式。对于性能标准,通常需要一个冒充者数据库,例如 B. Sabac 的文章“Speaker recognition using discriminative features selection”(发表于 ICSLP-2002, pp. 2321-2324) 中使用的错误接受率。即,需要用大量的用户数据和冒充者数据来测试不同特征组合的识别性能来做判断并最终找到最优的组合。然而,冒充者的数据并不总是存在的。在一个密码可以自由选择的话人验证系统中,就很难为所有可能的密码准备冒充者的数据。

### 发明内容

[0006] 为了解决上述现有技术中存在的问题,本发明提供了说话人认证的注册方法和装置、说话人认证的验证方法和装置、以及说话人认证系统。

[0007] 根据本发明的一个方面,提供了一种说话人认证的注册方法,包括:根据说话人说出的包含相同内容的多个语音,分别生成多个声学特征向量序列;根据上述多个声学特征序列生成一个参考模板;根据一个码本,为上述多个声学特征向量序列的每一个分别生成相应的伪冒充者声学特征向量序列,其中上述码本包含多个码字以及每个码字对应的特征向量;以及根据上述多个声学特征向量序列、上述参考模板和上述多个伪冒充者声学特征

向量序列,选择优化的声学特征子集。

[0008] 根据本发明的另一个方面,提供了一种说话人认证的验证方法,包括:根据测试语音,生成测试声学特征向量序列;根据优化的声学特征子集,对上述测试声学特征向量序列进行优化得到优化的测试声学特征向量序列,其中,上述优化的声学特征子集是在注册过程中生成的;根据参考模板和上述优化的测试声学特征向量序列,判断该输入的测试语音是否为说话人本人说出的注册语音。

[0009] 根据本发明的另一个方面,提供了一种说话人认证的注册装置,包括:声学特征提取单元,其根据说话人说出的语音,生成声学特征向量序列;模板生成单元,其根据说话人说出的包含相同内容的多个语音的多个声学特征序列,生成参考模板;伪冒充者数据生成单元,其根据一个码本,为上述多个声学特征向量序列的每一个分别生成相应的伪冒充者声学特征向量序列,其中上述码本包含多个码字以及每个码字对应的特征向量;以及优化单元,其根据上述多个声学特征向量序列、上述参考模板和上述多个伪冒充者声学特征向量序列,选择优化的声学特征子集。

[0010] 根据本发明的另一个方面,提供了一种说话人认证的验证装置,包括:测试声学特征提取单元,其根据测试语音,生成测试声学特征向量序列;测试优化单元,根据优化的声学特征子集,对上述测试声学特征向量序列进行优化得到优化的测试声学特征向量序列,其中,上述优化的声学特征子集是在注册过程中生成的;判断单元,其根据参考模板和上述优化的测试声学特征向量序列,判断该输入的测试语音是否为说话人本人说出的注册语音。

[0011] 根据本发明的另一个方面,提供了一种说话人认证系统,包括:前面所述的说话人认证的注册装置;以及前面所述的说话人认证的验证装置。

## 附图说明

[0012] 相信通过以下结合附图对本发明具体实施方式的说明,能够使人们更好地了解本发明上述的特点、优点和目的。

[0013] 图 1 是根据本发明一个实施例的说话人认证的注册方法的流程图;

[0014] 图 2 是根据本发明另一个实施例的说话人认证的注册方法的流程图;

[0015] 图 3 是根据本发明一个实施例的说话人认证的验证方法的流程图;

[0016] 图 4 是根据本发明的一个实施例的说话人认证的注册装置的框图;

[0017] 图 5 是根据本发明的一个实施例的说话人认证的验证装置的框图;以及

[0018] 图 6 是根据本发明一个实施例的说话人认证系统的框图。

## 具体实施方式

[0019] 下面就结合附图对本发明的各个优选实施例进行详细的说明。

[0020] 图 1 是根据本发明一个实施例的说话人认证的注册方法的流程图。如图 1 所示,首先在步骤 101,根据说话人说出的包含相同内容的多个语音,分别生成多个声学特征向量序列。每个声学特征向量可以包含,例如,多个 MFCC(Mel Frequency Cepstrum Coefficient, Mel 频率倒谱系数)的方式来表示语音的声学特征。但是,本发明对此并没有特别的限制,也可以采用已知的和未来的其它方式来表示语音的声学特征,例如,LPCC(Linear

Predictive Cepstrum Coefficient, 线性预测倒谱系数) 或者其它基于能量、基音频率或小波分析等得到的各种系数等, 以及音高 (pitch) 和时长 (duration) 等信息, 以及它们在时间轴上的一阶差分 (first differential) 和二阶差分 (second differential) 等等。所有被认为适用于说话人识别的特征都可以合并在一起作为特征全集, 然后, 利用后面描述的特征选择方法, 在注册过程中自动地针对各说话人的特色, 从特征全集中选择特征, 为说话人定制一套优化的特征子集。

[0021] 接着, 在步骤 105, 根据上述多个声学特征向量序列生成一个参考模板。例如, 首先选定一个声学特征向量序列作为初始模板, 然后用 DTW 的方法将第二个声学特征向量序列与之时间对齐, 并用两个声学特征向量序列中相对应的特征向量的平均来生成一个新的模板, 然后再将第三个声学特征向量序列与新模板时间对齐, 如此循环直到所有的声学特征向量序列都结合到一个独立的模板中, 即所谓的模板合并。详细内容可以参考 W. H. Abdulla、D. Chow 和 G. Sin 发表的文章“Cross-words referencetemplate for DTW-based speech recognition systems”(IEEE TENCON2003, pp. 1576-1579)。应当指出, 本发明对于生成参考模板的方式没有特别的限制。

[0022] 接着, 在步骤 110, 根据一个码本, 为上述多个声学特征向量序列的每一个分别生成相应的伪冒充者声学特征向量序列。在本实施例中使用的码本是在整个应用的声学空间中训练出的码本, 例如, 对于中文语言应用环境来说, 该码本需要能够涵盖中文语音的声学空间; 对于英文语言应用环境来说, 该码本则需要能够涵盖英文语音的声学空间。当然, 对于一些特殊用途的应用环境, 也可以相应的改变码本所涵盖的声学空间。

[0023] 本实施例的码本包含多个码字以及每个码字对应的特征向量。码字的数量取决于声学空间的大小、希望的压缩比例和希望的压缩质量。声学空间越大需要的码字的数量越大。在同样的声学空间的条件下, 码字的数量越小, 压缩比例越高; 码字的数量越大, 压缩的模板质量越高。根据本发明的一个优选实施例, 在普通中文语音的声学空间下, 码字的数量优选为 256 至 512。当然, 根据不同需要, 可以适当调节码本的码字数量和涵盖的声学空间。

[0024] 具体地, 本步骤中, 首先为该声学特征向量序列中的每个特征向量指定一个码字, 从而将该声学特征向量序列转换为一个对应的码字序列。例如, 可以通过计算该声学特征向量序列中的该特征向量和码本中每个特征向量的距离 (例如, 欧氏距离), 来找出最接近的特征向量。将码本中与该最接近的特征向量对应的码字指定给该声学特征向量序列中的该特征向量。

[0025] 然后, 根据上述码本中的码字和对应的特征向量, 将上述码字序列转换为对应的特征向量序列, 作为上述伪冒充者声学特征向量序列。

[0026] 接着, 在步骤 115, 根据上述多个声学特征向量序列、上述参考模板和上述多个伪冒充者声学特征向量序列, 选择优化的声学特征子集。具体地, 遍历每一种可能的声学特征子集, 并且选择使得相对于上述多个伪冒充者声学特征向量序列上述参考模板的识别率最高的声学特征子集作为上述优化的声学特征子集。

[0027] 根据本发明的一个实施例, 按照每个可能的声学特征子集, 分别计算上述多个声学特征向量序列和上述参考模板的 DTW 距离  $d_e(i)$  (在此被称为说话人内距离), 以及上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $d_p(i)$  (在此被称为说话人间距离)。选择上述多个声学特征向量序列和上述参考模板的 DTW 距离  $d_e(i)$  与上述多个伪

冒充者声学特征向量序列和上述参考模板的 DTW 距离  $dp(i)$  的比值最小的声学特征子集, 作为上述优化的声学特征子集。

[0028] 根据本发明的另一个实施例, 按照每个可能的声学特征子集, 分别计算上述多个声学特征向量序列和上述参考模板的 DTW 距离  $de(i)$ , 以及上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $dp(i)$ 。选择上述多个声学特征向量序列和上述参考模板的 DTW 距离  $de(i)$  与上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $dp(i)$  的差、除以上述多个声学特征向量序列和上述参考模板的 DTW 距离  $de(i)$  与上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $dp(i)$  的和的比值最小的声学特征子集, 作为上述优化的声学特征子集。

[0029] 另外, 根据本发明的一个实施例, 上述遍历每一种可能的声学特征子集的步骤是在指定的范围内进行的。上述指定的范围, 例如, 是声学特征的个数大于一个指定的数的可能的声学特征子集。

[0030] 通过以上说明可以看出, 本实施例的说话人认证的注册方法可以在没有冒充者数据库的情况下选择优化的声学特征子集, 根据每个说话人的特点选择的最佳特征组合将比传统方法更具有分辨力。将它应用在与文本相关的说话人验证系统中, 系统的性能得到显著提高。

[0031] 图 2 是根据本发明另一个实施例的说话人认证的注册方法的流程图。下面就结合该图, 对本实施例进行描述。对于那些与前面实施例相同的部分, 适当省略其说明。

[0032] 如图 2 所示, 本实施例的说话人认证的注册方法的步骤 101 至 115 与图 1 所示的实施例相同, 在此不再重复。

[0033] 与前面描述的实施例相比, 本实施例还包括步骤 220, 压缩参考模板。具体地, 可以包括: 根据上述优化的声学特征子集, 对上述参考模板中的声学特征向量的维数进行压缩; 或者, 根据上述码本, 对上述参考模板中的声学特征向量的个数进行压缩。

[0034] 其中, 根据上述码本, 对上述参考模板中的声学特征向量的个数进行压缩方式可以参考本申请人在 2005 年 11 月 11 日提交的中国专利申请 200510115300.5 (发明名称为“说话人模板的压缩、合并装置和方法, 以及说话人认证”)。将参考模板中相邻且被指定的码字相同的多个特征向量用一个特征向量代替。例如, 首先, 计算上述相邻且码字相同的一组特征向量的平均向量, 然后, 用计算的平均向量代替上述相邻且码字相同的一组特征向量。

[0035] 如果在参考模板中存在有多组这样的相邻且码字相同的多个特征向量, 则可以按照上述方式逐一地进行替换。这样, 逐一地将多个特征向量替换为一个特征向量, 参考模板中的特征向量的数量就减少了, 模板也就被压缩了。

[0036] 通过以上说明可以看出, 本实施例的说话人认证的注册方法不仅可以在没有冒充者数据库的情况下选择优化的声学特征子集, 而且参考模板也被相应地压缩, 使得模板所需的存储量下降, 同时, 计算量也得到了降低。将它应用在与文本相关的说话人验证系统中, 系统的性能得到显著提高。

[0037] 在同一发明构思下, 图 3 是根据本发明一个实施例的说话人认证的验证方法的流程图。下面就结合图 3, 对该实施例进行描述。对于那些与前面实施例相同的部分, 适当省略其说明。



[0038] 如图 3 所示,首先在步骤 301,根据测试语音,生成测试声学特征向量序列。与前面描述的图 1 的步骤 101 类似,每个声学特征向量可以包含,例如,多个 MFCC(Mel Frequency Cepstrum Coefficient, Mel 频率倒谱系数)的方式来表示语音的声学特征。但是,本发明对此并没有特别的限制,也可以采用已知的和未来的其它方式来表示语音的声学特征,例如,LPCC(Linear Predictive Cepstrum Coefficient,线性预测倒谱系数)或者其它基于能量、基音频率或小波分析等得到的各种系数等,以及音高(pitch)和时长(duration)等信息,以及它们在时间轴上的一阶差分(first differential)和二阶差分(second differential)等等。所有被认为适用于说话人识别的特征都可以合并在一起作为特征全集。

[0039] 接着,在步骤 305,根据优化的声学特征子集,对上述测试声学特征向量序列进行优化得到优化的测试声学特征向量序列,其中,上述优化的声学特征子集是在注册过程中生成的。前面实施例中已经对如何选择优化的声学特征子集进行了说明,在此不再重复。

[0040] 接着,在步骤 310,根据参考模板和上述优化的测试声学特征向量序列,判断该输入的测试语音是否为说话人本人说出的注册语音。具体地,例如,首先计算上述参考模板和上述优化的测试声学特征向量序列的 DTW 匹配得分;然后比较上述计算出的 DTW 匹配得分和一个分辨阈值,判断输入的语音是否为说话人本人说出的注册语音。

[0041] 需要指出,已知的和未来的各种计算参考模板和测试声学特征向量序列的 DTW 匹配得分的方式都可以适用于本发明。而且,已知的和未来的各种分辨阈值的设定方式也都可以适用于本发明。

[0042] 通过以上说明可以看出,本实施例的说话人认证的验证方法可以利用在注册阶段选择的优化的声学特征子集,根据每个说话人的特点选择的最佳特征组合,比传统方法更具有分辨力。系统的性能可以得到显著提高。

[0043] 在同一发明构思下,图 4 是根据本发明的一个实施例的说话人认证的注册装置的框图。下面就结合图 4,对该实施例进行描述。对于那些与前面实施例相同的部分,适当省略其说明。

[0044] 如图 4 所示,本实施例的说话人认证的注册装置 400,包括:声学特征提取单元 401,其根据说话人说出的语音,生成声学特征向量序列;模板生成单元 402,其根据说话人说出的包含相同内容的多个语音的多个声学特征向量序列,生成参考模板;伪冒充者数据生成单元 403,其根据一个码本 704,为上述多个声学特征向量序列的每一个分别生成相应的伪冒充者声学特征向量序列,其中上述码本 704 包含多个码字以及每个码字对应的特征向量;以及优化单元 405,其根据上述多个声学特征向量序列、上述参考模板和上述多个伪冒充者声学特征向量序列,选择优化的声学特征子集。

[0045] 其中,伪冒充者数据生成单元 403 包括:向量-码字转换单元 4031,其为该声学特征向量序列中的每个特征向量指定一个码字,构成一个与该声学特征向量序列对应的码字序列;码字-向量转换单元 4032,其根据上述码本中的码字和对应的特征向量,将上述码字序列转换为对应的特征向量序列,作为上述伪冒充者声学特征向量序列。

[0046] 根据本发明的一个实施例,向量-码字转换单元 4031,为该声学特征向量序列中的每个上述特征向量查找码本中与之最接近的特征向量;并且将上述最接近的特征向量对应的码字指定给该声学特征向量序列中的该特征向量。

[0047] 根据本发明的一个实施例,优化单元 405,遍历每一种可能的声学特征子集;以及选择使得相对于上述多个伪冒充者声学特征向量序列上述参考模板的识别率最高的声学特征子集作为上述优化的声学特征子集。

[0048] 根据本发明的另一个实施例,优化单元 405,遍历每一种可能的声学特征子集;按照该声学特征子集,分别计算上述多个声学特征向量序列和上述参考模板的 DTW 距离  $d_e(i)$ ,以及上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $d_p(i)$ ;以及选择上述多个声学特征向量序列和上述参考模板的 DTW 距离  $d_e(i)$  与上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $d_p(i)$  的比值最小的声学特征子集,作为上述优化的声学特征子集。

[0049] 根据本发明的另一个实施例,优化单元 405,遍历每一种可能的声学特征子集;按照该声学特征子集,分别计算上述多个声学特征向量序列和上述参考模板的 DTW 距离  $d_e(i)$ ,以及上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $d_p(i)$ ;以及选择上述多个声学特征向量序列和上述参考模板的 DTW 距离  $d_e(i)$  与上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $d_p(i)$  的差、除以上述多个声学特征向量序列和上述参考模板的 DTW 距离  $d_e(i)$  与上述多个伪冒充者声学特征向量序列和上述参考模板的 DTW 距离  $d_p(i)$  的和的比值最小的声学特征子集,作为上述优化的声学特征子集。

[0050] 根据本发明的另一个实施例,优化单元 405 在指定的范围内遍历每一种可能的声学特征子集。上述指定的范围,例如,是声学特征的个数大于一个指定的数的可能的声学特征子集。

[0051] 如图 4 所示,本实施例的说话人认证的注册装置 400,进一步包括:压缩单元 406,其根据上述优化的声学特征子集,对上述参考模板中的声学特征向量的维数进行压缩。

[0052] 根据本发明的另一个实施例,压缩单元 406 进一步根据上述码本,对上述参考模板中的声学特征向量的个数进行压缩。

[0053] 通过以上说明可以看出,本实施例的说话人认证的注册装置操作上可以实现前面实施例描述的说话人认证的注册方法,可以在没有冒充者数据库的情况下选择优化的声学特征子集,根据每个说话人的特点选择的最佳特征组合将比传统方法更具有分辨力。将它应用在与文本相关的说话人验证系统中,系统的性能得到显著提高。

[0054] 另外,参考模板也可以被相应地压缩,使得模板所需的存储量下降,同时,计算量也得到了降低。

[0055] 在同一发明构思下,图 5 是根据本发明的一个实施例的说话人认证的验证装置的框图。下面就结合图 5,对该实施例进行描述。对于那些与前面实施例相同的部分,适当省略其说明。

[0056] 如图 4 所示,本实施例的说话人认证的验证装置 500,包括:测试声学特征提取单元 501,其根据测试语音,生成测试声学特征向量序列;测试优化单元 502,根据优化的声学特征子集,对上述测试声学特征向量序列进行优化得到优化的测试声学特征向量序列,其中,上述优化的声学特征子集是在注册过程中生成的;判断单元 503,其根据参考模板和上述优化的测试声学特征向量序列,判断该输入的测试语音是否为说话人本人说出的注册语音。

[0057] 其中,判断单元 503 包括:DTW 计算单元 5031,其计算上述参考模板和上述优化的

测试声学特征向量序列的 DTW 匹配得分。判断单元 503, 比较上述计算出的 DTW 匹配得分和一个阈值, 判断输入的语音是否为说话人本人说出的注册语音。

[0058] 通过以上说明可以看出, 通过以上说明可以看出, 本实施例的说话人认证的验证装置操作上可以实现前面实施例描述的说话人认证的验证方法, 可以利用在注册阶段选择的优化的声学特征子集, 根据每个说话人的特点选择的最佳特征组合, 比传统方法更具有分辨力。系统的性能可以得到显著提高。

[0059] 在同一发明构思下, 图 6 是根据本发明一个实施例的说话人认证系统的框图。下面就结合图 6, 对该实施例进行描述。对于那些与前面实施例相同的部分, 适当省略其说明。

[0060] 如图 6 所示, 本实施例的说话人认证系统包括: 注册装置 400, 其可以为前面实施例描述的说话人认证的注册装置; 以及验证装置 500, 其可以为前面实施例描述的说话人认证的验证装置。由注册装置 400 生成的参考模板和优化的特征子集, 通过任意的通信方式, 例如, 网络、内部信道、磁盘等记录媒体等, 传递给验证装置 500。

[0061] 通过以上说明可以看出, 本实施例的说话人认证系统, 在注册阶段, 可以在没有冒充者数据库的情况下选择优化的声学特征子集, 根据每个说话人的特点选择的最佳特征组合; 在验证阶段, 可以利用在注册阶段选择的优化的声学特征子集, 根据每个说话人的特点选择的最佳特征组合。因此, 比传统方法更具有分辨力, 系统的性能可以得到显著提高。另外, 参考模板也可以被相应地压缩, 使得模板所需的存储量下降, 同时, 计算量也得到了降低。

[0062] 需要指出, 本发明上述各个实施例的说话人认证的注册装置 400、说话人认证的验证装置 500 及其各个组成部分, 可以由专用的电路或芯片构成, 也可以通过计算机 ( 处理器 ) 执行相应的程序来实现。

[0063] 以上虽然通过一些示例性的实施例对本发明的说话人认证的注册方法和装置、说话人认证的验证方法和装置、以及说话人认证系统。进行了详细的描述, 但是以上这些实施例并不是穷举的, 本领域技术人员可以在本发明的精神和范围内实现各种变化和修改。因此, 本发明并不限于这些实施例, 本发明的范围仅由所附权利要求为准。

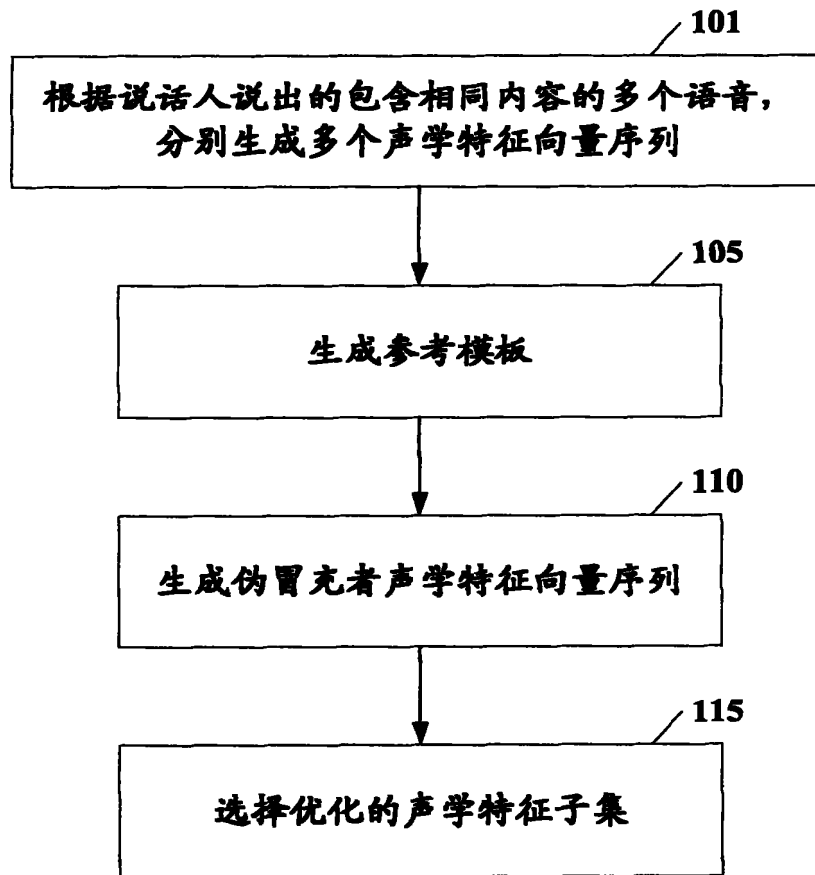


图 1

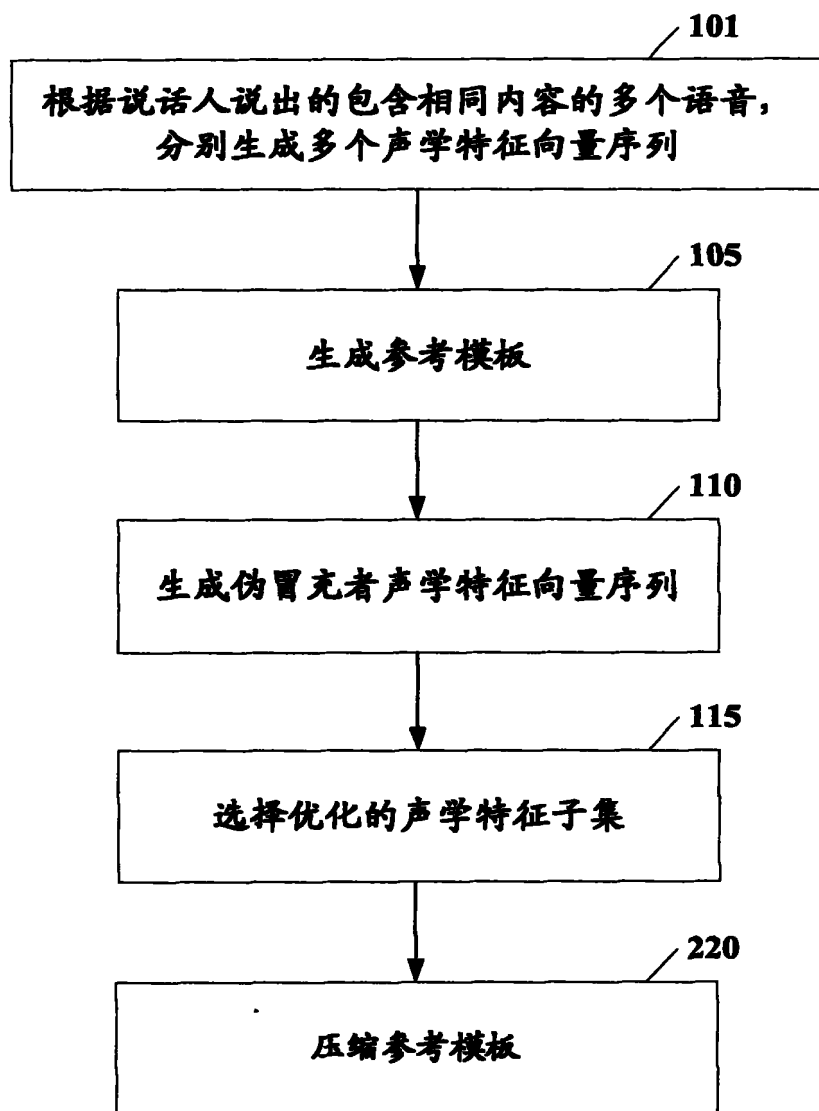


图 2

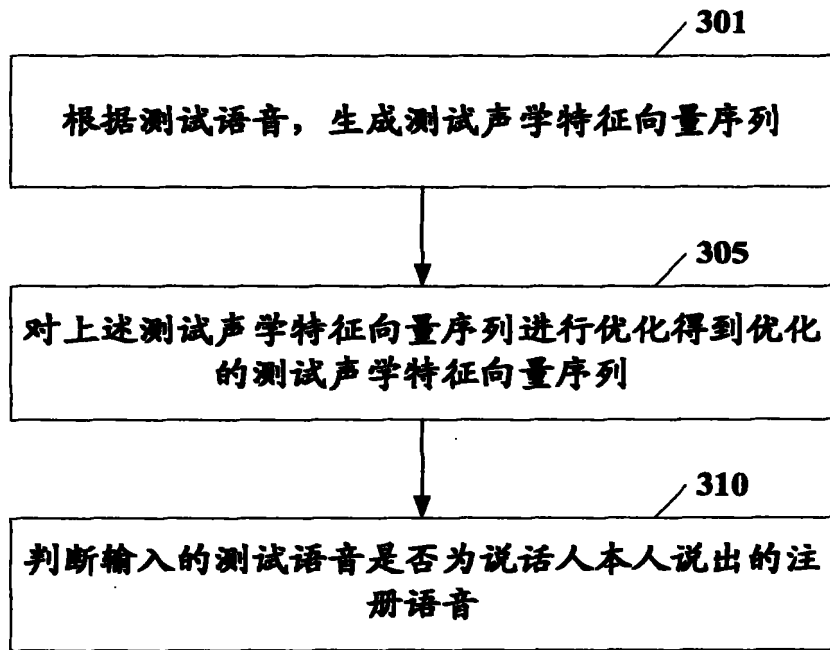


图 3

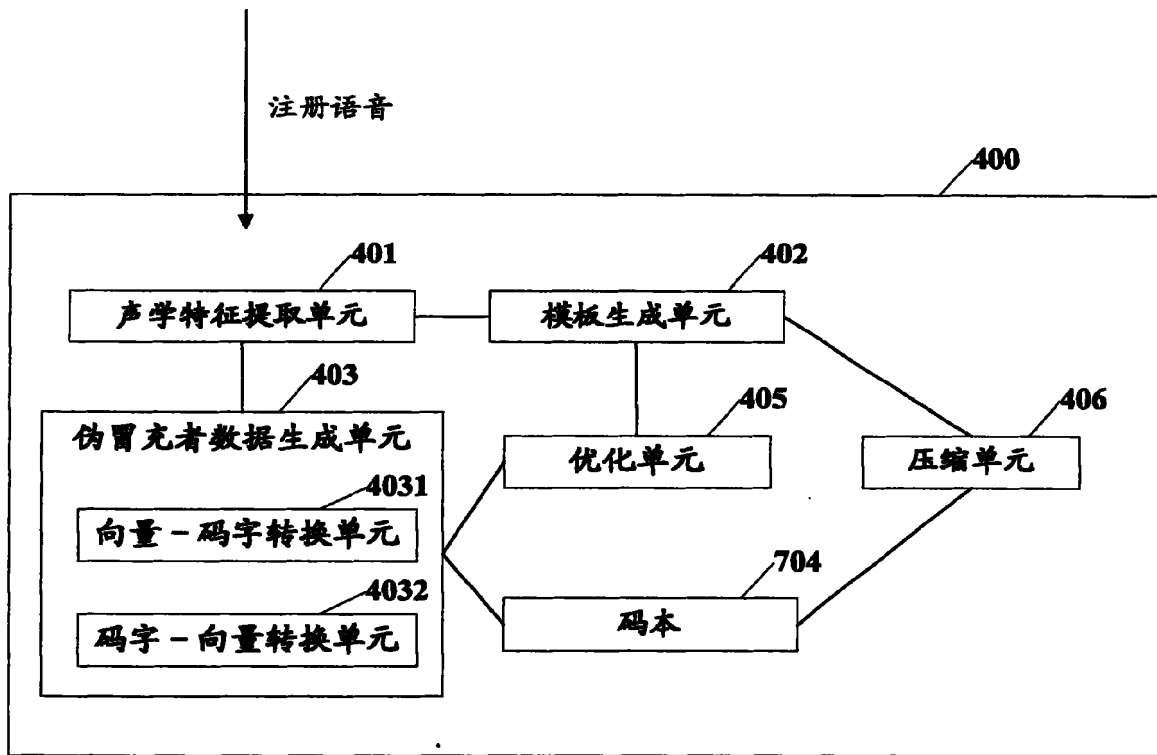


图 4

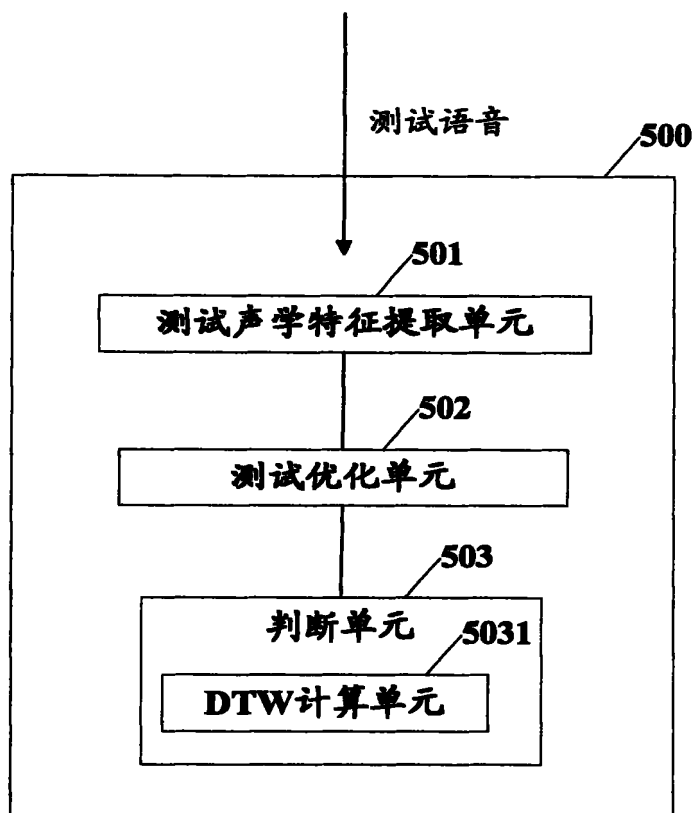


图 5

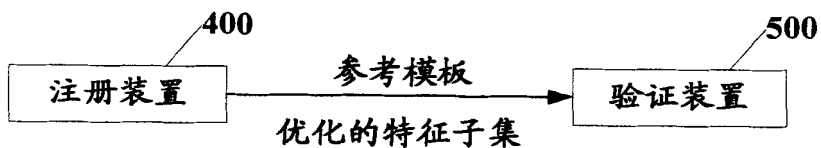


图 6