(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 1879146 B
(45)授权公告日 2011.06.08

(21)申请号 200480032825.6

(22)申请日 2004.10.27

(30)优先权数据
03104078.5 2003.11.05 EP

(85)PCT申请进入国家阶段日
2006.05.08

(86)PCT申请的申请数据
PCT/IB2004/052218 2004.10.27

(87)PCT申请的公布数据
WO2005/045803 EN 2005.05.19

(73)专利权人 皇家飞利浦电子股份有限公司
地址 荷兰艾恩德霍芬

(72)发明人 H·施拉姆

(74)专利代理机构 中国专利代理(香港)有限公司 72001
代理人 李亚非 刘杰

(51)Int.Cl.
G10L 13/04(2006.01)

(56)对比文件
EP 0962914 A2,1999.12.08, 全文.
US 5987405 A,1999.11.16, 第3栏7-29行, 第4栏64行-第5栏12行、图2.
CN 1255011 A,2000.05.31, 全文.

审查员 张亚峰

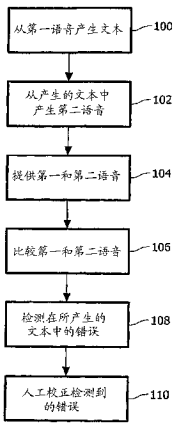
权利要求书 2 页 说明书 6 页 附图 4 页

(54)发明名称

用于语音到文本的转录系统的错误检测

(57)摘要

本发明涉及一种用于在由语音到文本的转录系统产生的文本内错误检测的方法、系统和计算机程序产品。借助于文本到语音的转录系统将转录文本再变换成人工语音信号。原始自然语音信号和人工产生的语音被提供给校对者以比较两个声学信号。在原始语音信号与从转录文本中变换的语音之间的偏差表示在语音到文本的变换过程中可能出现了错误,这不得不人工进行校正。优选地通过使用从所述两个语音信号推导出的比较信号,要比较的语音信号能够在视觉上和/或在听觉上被提供给校对者。在校对过程中能够跳过大部分正确转录的文本,从而节省了时间,并且增强了整个校对过程的有效性。



CN 1879146 B

1. 一种用于在利用自动的语音到文本的转录系统从第一语音信号转录的文本内进行错误检测的方法，包括：

从转录的文本中合成第二语音信号；

基于第一语音信号和第二语音信号，产生比较信号，所述比较信号指示这两个语音信号之间的偏差；和

在听觉上和/或在视觉上将所述比较信号提供给校对者，

其中该方法进一步包括比较信号的模式识别，以识别表示文本中的错误类型的比较信号的预先训练的模式。

2. 根据权利要求 1 所述的方法，其中第二语音信号的速度和/或音量与第一语音信号的速度和/或音量相匹配。

3. 根据权利要求 1 或 2 所述的方法，其中一组滤波器函数被应用于第一语音信号，以使得第一语音信号的频谱接近第二语音信号的频谱。

4. 根据权利要求 1 或 2 所述的方法，其中第二语音信号是通过下述产生的：应用逆语音转录过程；从文本中产生特征向量序列；使用 (a) 语音到文本的转录系统的统计模型和 (b) 在从第一语音信号转录文本的过程中获得的状态序列。

5. 根据权利要求 1 或 2 所述的方法，其中通过对第一和第二语音信号进行相减或叠加来产生比较信号。

6. 根据权利要求 5 所述的方法，其中当所述比较信号的幅度超过预定范围时，输出出错指示。

7. 根据权利要求 6 所述的方法，其中在图形用户接口上在转录的文本内在视觉上输出出错指示。

8. 根据权利要求 1 所述的方法，其中给校正建议提供在所产生的文本中检测到的错误类型。

9. 根据权利要求 1 所述的方法，进一步包括：

提供第一和第二语音信号输出给校对者，以便在第一和第二语音信号之间进行比较，从而识别在转录的文本中可能的错误。

10. 一种用于语音到文本的转录系统的错误检测系统，该语音到文本的转录系统从第一语音信号 (400) 中提供转录的文本 (412)，所述错误检测系统包括：

- 用于从转录的文本 (412) 中合成第二语音信号 (416) 的装置；

- 用于基于第一和第二语音信号来产生比较信号的装置，所述比较信号指示这两个语音信号之间的偏差；和

- 用于在听觉上和/或在视觉上将所述比较信号提供给校对者的装置，

其中所述比较信号中的不同模式被指定给转录的文本 (412) 中的某一错误类型。

11. 根据权利要求 10 所述的错误检测系统，其中通过对第一和第二语音信号进行相减或叠加来产生比较信号。

12. 根据权利要求 10 或 11 所述的错误检测系统，其中为了错误检测的目的而在听觉上或在视觉上提供第一和第二语音信号。

13. 根据权利要求 11 所述的错误检测系统，其中当所述比较信号超过预定范围时，输出出错指示。

14. 根据权利要求 11 所述的错误检测系统，其中给校正建议提供在转录的文本中检测到的错误类型。

15. 根据权利要求 10 所述的错误检测系统，还包括：

用于提供第一和第二语音信号的装置，以便在第一和第二语音信号之间进行比较，从而识别在转录的文本 (412) 中可能的错误。

用于语音到文本的转录系统的错误检测

技术领域

[0001] 本发明涉及语音到文本的转录（transcription）系统和方法的领域，并且更具体地涉及在语音到文本的转录系统中错误的检测。

背景技术

[0002] 语音转录和语音识别系统识别例如口头口授的语音，并且将所识别的语音转录成文本。现今语音转录系统已经被广泛地用在例如医学部门或法律实践中。存在多种语音转录系统，例如Philips ElectronicsNV的Speed Magic™和IBM Corporation的Via Voice™，它们在商业上是可用的。与人类转录人员相比，一方面，语音转录系统节省了时间和成本，但另一方面，它与人类转录人员相比不能够提供高精度的语音理解和命令解释。

[0003] 由语音到文本的转录系统产生的文本不可避免地包括错误的文本部分。这种错误的文本部分的出现是由于许多原因，例如在其中记录语音的不同的环境条件，比如噪声，或者系统未被正确适配的不同扬声器。口授中涉及标点、文本格式或字体的口头命令必须由语音到文本的转录系统进行适当地解释，而不是按照字面意义转录为词语。

[0004] 由于语音到文本的转录系统的特征在于有限的语音识别能力以及有限的命令解释能力，所以它们在转录的文本中不可避免地产生错误。为了确保将口授正确地转录为文本，在校对步骤中必须对语音到文本的转录系统所产生的文本检查错误以及错误的文本部分。校对通常必须由人类校对者执行。校对者比较口授的原始语音信号与由语音到文本的转录系统产生的转录文本。

[0005] 通常通过听取原始语音信号并同时读取转录文本来执行以比较的形式的校对。特别是，这类比较使得校对者非常疲劳，因为以视觉信息的形式的文本必须与以声学信息的形式提供的语音信号进行比较。因此，该比较需要校对者在与口授的持续时间相对应的时间高度集中注意力。

[0006] 考虑到语音到文本的转录系统的出错率可以在20％以下并且在不久的将来甚至可以减少，显然对于转录文本的大部分而言校对不是必须的。然而，文本的原始来源仅作为语音信号而可得到，该语音信号只可以通过听取其而以顺序的方式进行访问。比较书面文本和声学信号只可以通过听取全部的声学信号来执行。因此，与转录过程本身相比，校对可能甚至更加消耗时间。

发明内容

[0007] 本发明旨在提供一种用于在由自动的语音到文本的转录系统所产生的文本内有效的错误检测的方法、系统和计算机程序产品。

[0008] 特别地，本发明提供一种用于在利用自动的语音到文本的转录系统从第一语音信号转录的文本内进行错误检测的方法，包括：从转录的文本中合成第二语音信号；基于第一语音信号和第二语音信号，产生比较信号，所述比较信号指示这两个语音信号之间的偏差；和在听觉上和／或在视觉上将所述比较信号提供给校对者。

[0009] 本发明还提供一种用于语音到文本的转录系统的错误检测系统，该语音到文本的转录系统从第一语音信号中提供转录的文本，所述错误检测系统包括：用于从转录的文本中合成第二语音信号的装置；用于基于第一和第二语音信号来产生比较信号的装置，所述比较信号指示这两个语音信号之间的偏差；和用于在听觉上和/或在视觉上将所述比较信号提供给校对者的装置。

[0010] 本发明提供一种用于语音到文本的转录系统的错误检测的方法。语音到文本的转录系统接收第一语音信号，并且将该第一语音信号转录成文本。为了便于必须由人类校对者执行的校对或校正过程，将转录的文本再变换成第二合成语音信号。这样，校对者仅必须比较两个声学信号，即第一和第二语音信号，而不是比较第一语音信号和转录的文本。例如经由立体声耳机将第一和第二语音信号提供给校对者。这样，校对者同时听取第一和第二语音信号，并且能够容易地检测到两个语音信号之间可能的偏差，该偏差表示在语音到文本的转录过程中已经出现错误。

[0011] 通过所谓的文本到语音的合成系统来执行将转录文本再变换成第二语音信号。例如在 EP 0363233 和 EP 0706170 中公开了文本到语音的合成系统的实例。典型的文本到语音的合成系统基于包含在其中存储有记录的声音部分的数据库的双音合成技术或单元选取合成技术。

[0012] 根据本发明的一个优选实施例，从转录文本产生与第一语音信号同步的合成第二语音信号的方法是颠倒语音识别过程。代替从输入特征向量（表示例如第一语音信号的 10ms 的部分）产生输出文本，也应用语音识别系统来从输入文本产生输出特征向量。这能够通过首先将文本变换成（上下文相关的）音素序列并且接着将音素序列变换成隐含马尔可夫模型序列 (HMMs) 来实现。连接的 HMMs 根据不同的 HMM 状态序列又产生输出特征向量序列。为了支持第一和第二语音信号之间的同步，用于产生第二语音信号的 HMM 状态序列是在先前语音识别步骤中获得的最佳（维特比）状态序列，在先前的语音识别步骤中已经将第一语音信号变换成文本。该状态序列将每个特征向量对准不同的隐含马尔可夫模型状态，并且因此对准转录文本的不同部分。

[0013] 根据本发明的另一个优选实施例，从第一语音信号的转录文本中提取的第二语音信号的速度和/或音量与第一语音信号的速度和/或音量相匹配。因此，相对于第一自然语音信号的速度和/或音量来执行从转录文本到第二语音信号的合成。这是有利的，因为两个同步的声学信号之间的比较比两个不同步的声学信号之间的比较容易得多。因此，转录文本的同步取决于转录文本体本身以及取决于第一、因而的自然语音信号的速度和动态范围。

[0014] 根据本发明的另一个优选实施例，第一语音信号也是变换的对象。优选地，对第一语音信号应用一组滤波器函数以便变换第一语音信号的频谱。这样，使得第一语音信号的频谱与合成的第二语音信号的频谱相同。结果，自然的第一语音信号的声音和合成的第二语音信号的声音接近，这再一次便于由人类校对者执行的两个语音信号的比较。最后，比较两个人工产生的或人工发声的声学信号，而不是比较一个人工声学信号与一个自然声学信号。

[0015] 根据本发明的另一个优选实施例，通过对第一和第二语音信号进行相减或叠加而产生一个附加信号。当通过对第一和第二语音信号进行相减而产生该类型的比较信号时，

该比较信号的幅度指示在第一和第二语音信号之间的偏差。特别是，第一和第二语音信号之间大的偏差是语音到文本的转录系统已经产生错误的指示。因此，比较信号给出了在语音到文本的转录过程中是否出现错误的直接指示。比较信号没有必要非得通过两个语音信号相减来产生。通常，可以想到根据第一和第二语音信号来产生比较信号的非常多的方法，例如借助于语音信号的叠加或卷积。

[0016] 根据本发明的另一个优选实施例，在听觉上和 / 或在视觉上将比较信号提供给校对者。这样，所产生的比较信号被提供给校对者。通过使用该比较信号，校对者能够更容易地识别转录文本的错误部分。特别是，当在转录文本中在视觉上提供比较信号时，校对者的注意力被吸引到与值得重视的比较信号相对应的那些文本部分。在校对过程中能够跳过与具有低幅度的比较信号相关联的大部分正确转录的文本。因而，显著地增强了校对者和校对过程的效率。

[0017] 根据本发明的另一个优选实施例，当比较信号的幅度超过预定范围时，用于错误检测的方法产生出错指示。当例如通过对第一和第二语音信号相减来产生比较信号的时候，在比较信号的幅度超过预定阈值时输出出错指示给校对者。出错指示的输出能够在听觉上以及在视觉上出现。借助于该出错指示，校对者不再非得观察或听取笨拙发声的比较信号。例如可以通过不同的振铃音来实现出错指示。

[0018] 根据本发明的另一个优选实施例，借助于图形用户接口在转录文本中在视觉上输出出错指示。这样，校对者不必非得在听觉上听取和比较两个语音信号。而且，第一和第二语音信号之间的比较完全由比较信号来表示。只有在比较信号超过预定阈值的这种情况中，才在转录文本中输出出错指示。于是校对者的任务减少为人工控制分配有出错指示的那些文本部分。校对者可以系统地选择可能出错的这些文本部分。为了检查语音到文本的转录系统是否产生错误，校对者仅听取与分配有出错指示的文本部分相对应的第一和第二语音信号的那些剪辑。

[0019] 因此，所述方法提供一种仅对可能是错误的转录文本的那些文本部分进行滤波的有效途径。因而不再需要为了校对的目的而听取完整的第一语音信号和读取整个转录文本。必须由人类校对者执行的校对有效地减少为由错误检测系统将其识别为可能出错的那些文本部分。与校对过程的时间曝光 (exposure) 减少相同，校对的总效率被提高。

[0020] 根据本发明的另一个优选实施例，对比较信号执行模式识别，以便对表示文本中不同错误类型的比较信号的预定义模式进行识别。由语音到文本的转录系统产生的错误通常是由于部分第一自然语音信号的错误判断。特别是对于自然语音信号的含糊部分出现这种错误，例如具有不同意义并因此具有不同拼写的相似发声的词语。例如，当例如将不同的口语词错误识别为相似发声的词语时，语音到文本的转录系统就可能产生无意义的词语。在转录过程期间，这种混淆可能出现若干次。当现在又将转录文本再变换成第二语音信号时以及当借助于上述的比较信号来比较第一和第二语音信号时，两个词之间的这种混淆可能导致比较信号中的不同模式。

[0021] 借助于应用于比较信号的模式识别，可以直接识别由转录系统产生的某一错误类型。与由语音到文本的转录系统产生的某些错误类型相对应的不同模式典型地通过某类的存储装置来存储，并且被提供给错误检测方法以便识别不同的错误类型。而且，在比较信号中与指示某个错误类型的任一已知模式不匹配的模式可以被指定为一个错误和由校对者

人工执行的校正过程。这样，用于错误检测的方法可以收集在比较信号中将被指定为不同错误类型的各种模式。可以将这种功能解释为一种自主学习。

[0022] 根据本发明的另一个优选实施例，校正建议具备由语音到文本的转录系统产生的检测到的错误类型。由于借助于比较信号的相应模式来识别转录文本中不同的错误类型，所以能够解析错误的来源、语音信号的错误识别部分。优选的是借助于图形用户接口在视觉上提供校正建议。必须由人类校对者执行的校对理想地减少为接受或拒绝由错误检测系统提供的校正建议的步骤。当校对者接受错误校正时，错误检测系统自动地用产生的校正建议代替转录文本的错误文本部分。假设校对者拒绝了由错误检测系统提供的校正建议的别的情况，那么校对者不得不人工校正转录文本的错误文本部分。

[0023] 用于在由语音到文本的转录系统所产生的文本内的错误检测的所述方法和系统为转录文本的校对提供了有效的和较少时间消耗的途径。不可缺少的人类校对者的主要任务减少为转录文本内最小数量的可能错误识别的文本部分。与校对的传统方法相比，校对者不再非得听取由语音到文本的转录系统所转录的整个自然语音信号。

附图说明

[0024] 在下面将通过参考附图来更加详细地描述本发明的优选实施例，其中：

[0025] 图 1 是说明错误检测方法的流程图，

[0026] 图 2 是说明错误检测方法的流程图，

[0027] 图 3 是说明包括比较信号的模式识别的错误检测方法的流程图，

[0028] 图 4 示出具有错误检测装置的语音到文本的转录系统的框图。

具体实施方式

[0029] 图 1 示出本发明的错误检测方法的流程图。在第一步骤 100 中，借助于常规的语音到文本的转录系统从第一自然语音信号中产生文本。在下一步骤 102 中，借助于常规的文本到语音的合成系统将步骤 100 的转录文本再变换成第二语音信号。在随后的步骤 104 中，将第一自然语音信号和第二人工产生的语音信号提供给人类校对者。在步骤 106，所述校对者同时听取第一和第二语音信号。典型地，使得第一和第二语音信号同步以便有助于由所述校对者执行的声学比较。在步骤 108，校对者检测到在第一和第二语音信号之间的偏差。所述偏差指示在将第一自然语音信号转录成文本的步骤 100 中出现错误。当在步骤 108 中校对者检测到错误时，人工执行该文本中检测到的错误的校正。

[0030] 这样，校对即初始自然语音信号与转录文本之间的比较不再基于有关声学与视觉信号的比较。代之以，校对者仅须听取两个不同的声学信号。仅在检测到错误的情况中，校对者不得不在转录文本中找到相应的文本部分并且执行校正。

[0031] 图 2 是说明根据本发明优选实施例的错误检测方法的流程图。与图 1 所说明的类似，在第一步骤 200 中，通过常规的文本到语音的转录系统从第一语音信号中转录文本。基于转录的文本，在下一步骤 202 中，借助于文本到语音的合成系统来合成人工语音信号。为了便于两个语音信号之间的比较，在步骤 204 中将第一自然语音信号应用到一组滤波器函数，以使得所述自然语音信号的频谱接近第二人工产生的语音信号的频谱。

[0032] 之后，该方法进行步骤 206 或进行步骤 208。在步骤 206 中，将滤波的第一自然语

音信号以及第二人工产生的语音信号在听觉上提供给校对者。相反在步骤 208 中，将滤波的第一自然语音信号以及第二人工产生的语音信号在视觉上提供给校对者。在将第一和第二语音信号提供给校对者之后，该方法继续进行步骤 210，在该步骤中，校对者在听觉上和 / 或在视觉上比较第一和第二语音信号。在下一步骤 212 中，校对者通过听取两个不同的语音信号和 / 或通过两个语音信号的图形表示来检测所产生的文本中的错误。在最后的步骤 214 中，由校对者来人工校正检测到的错误。

[0033] 在图 3 中，示出用于说明根据本发明的错误检测方法的另一流程图。再次在第一步骤 300 中，借助于常规的语音到文本的转录系统从第一自然语音信号中转录文本。在下一步骤 302 中，借助于文本到语音的合成系统将转录文本再变换成第二语音信号。与图 2 所描述的类似，在步骤 304 中，将第一自然语音信号应用到一组滤波器函数，以便使得第一语音信号的声音和频谱与人工产生的第二语音信号的声音和频谱相同。

[0034] 在随后的步骤 306 中，借助于例如第一和第二语音信号相减或叠加来产生第一和第二语音信号之间的比较信号。代替直接提供语音信号，该方法现在限定为提供所产生的比较信号。在步骤 308 在听觉上提供比较信号或在步骤 310 在视觉上提供比较信号。在步骤 312 中，通过所述比较信号能够容易地检测到文本中可能的错误。

[0035] 当例如通过两个语音信号相减而产生所述比较信号时，在该比较信号的幅度高于预定阈值时，能够容易地检测到文本中可能的错误。在步骤 312 检测到可能的错误文本部分之后，在步骤 318 能够人工地对检测到的错误执行校正，或者可以使用可选步骤 314 和 316。在步骤 314，对所述比较信号应用模式识别。当比较信号的不同部分与系统中所存储的两个特征模式相匹配时，转录文本的相应文本部分被识别为可能的错误。在随后的步骤 316 中，将这些可能的错误文本部分指定为一种不同的错误类型。以这种方式收集的出错信息可以被进一步使用，以便产生建议校正来消除转录文本中的这些错误。

[0036] 图 4 示出用于语音到文本的转录系统的错误检测系统的框图。第一语音信号 400 被输入到错误检测模块 402。该错误检测模块 402 包括用于语音到文本的转录的装置并且产生文本 412，该文本从错误检测模块 402 输出。此外，错误检测模块 402 被连接到图形用户接口 406 以及被连接到声学用户接口 404。错误检测模块 402 进一步包括语音合成模块 408、语音到文本的转录模块 410、文本到语音的变换模块 414 以及文本 412、第一语音信号 418 和第二语音信号 416。

[0037] 表示口授的自然语音信号 400 被输入到错误检测模块 402 的语音合成模块 408 和语音到文本的转录模块 410。语音到文本的转录模块 410 将语音信号 400 转录成文本 412。所产生的文本 412 作为转录文本被输出，并且在错误检测模块 402 中进行进一步的处理。因此，将文本 412 提供给文本到语音的变换模块 414，其将转录文本 412 再变换成第二人工产生的语音信号 416。

[0038] 文本到语音的变换模块 414 基于从文本到语音的合成系统获知的常规技术。现在借助于声学用户接口 404 能够将人工产生的语音信号 416 与进入错误检测模块 402 的初始自然语音信号 400 进行比较。声学用户接口 404 例如能够通过立体声耳机来实施。在立体声耳机的左声道上可以提供自然语音信号 400，而在该耳机的右声道上可以提供人工产生的语音信号 416。

[0039] 同时听取两个语音信号的人类校对者因此能够容易地检测到由语音到文本的转

录模块 410 执行的错误判断和错误所引起的两个语音信号 400 和 416 之间的偏差。

[0040] 由于自然语音信号 400 和机器产生的语音信号 416 之间的比较可能对于校对者而言是混淆的或是笨拙发声的，所以通过语音合成模块 408 对自然语音信号 400 应用一组滤波器函数而能够对自然语音信号 400 进行滤波，以便使得自然语音信号 400 的频谱和声音与合成的语音信号 416 相同。因此，语音合成模块 408 将自然语音信号 400 变换成滤波的语音信号 418。与上述两个语音信号的描述类似，借助于声学用户接口 404 能够在听觉上将所述滤波的语音信号 418 以及所述合成的语音信号 416 提供给校对者。

[0041] 另外地或可选地，借助于图形用户接口 406 能够以图形表示来提供所述两个产生的语音信号。在语音信号 416 和 418 的图形表示的帮助下，校对者可以跳过已经正确转录的转录文本的大部分。特别是，当错误检测模块 402 通过产生指示两个语音信号的巨大偏差的比较信号来提供两个语音信号 416 和 418 的进一步处理时，由语音到文本的变换模块 410 所产生的错误的校对过程和检测以及校正变得更加有效并且具有较少的时间消耗。进一步的优点是通过其中不同模式能够被指定为特定的错误类型的模式识别对产生的比较信号进行进一步的处理，以便有助于由人类校对者执行检测和校正任务。

[0042] 参考数字的列表

[0043] 400 :第一语音信号

[0044] 402 :错误检测模块

[0045] 404 :声学用户接口

[0046] 406 :图形用户接口

[0047] 408 :语音合成模块

[0048] 410 :语音到文本的转录模块

[0049] 412 :文本

[0050] 414 :文本到语音的变换模块

[0051] 416 :第二语音信号

[0052] 418 :滤波的语音信号

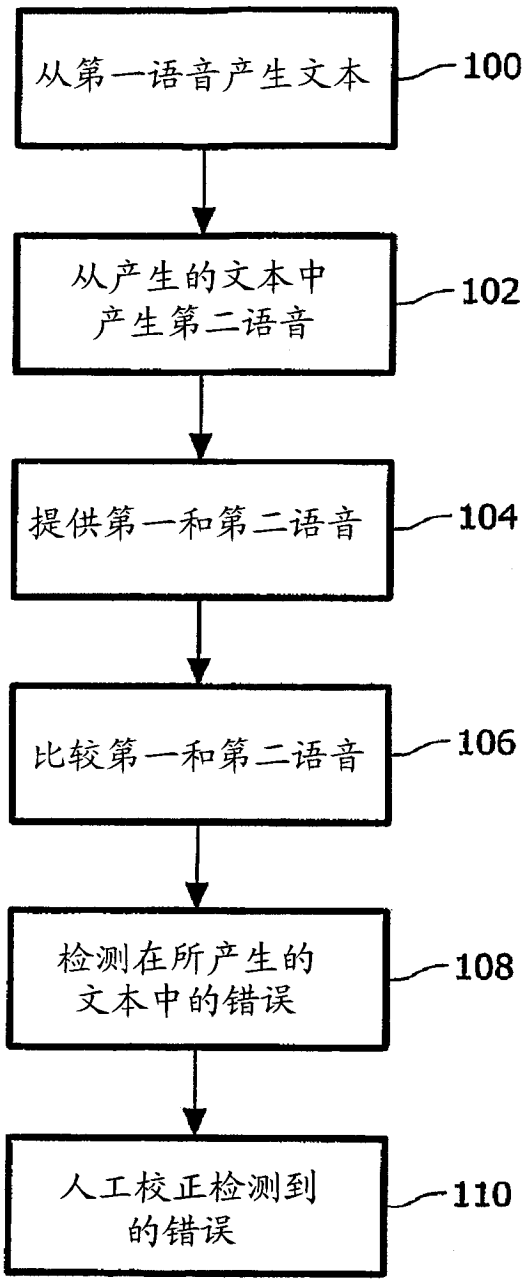

从第一语音产生文本
100
从产生的文本中产生第二语音
102
提供第一和第二语音
104
比较第一和第二语音
106
检测在所产生的文本中的错误
108
人工校正检测到的错误
110

图 1

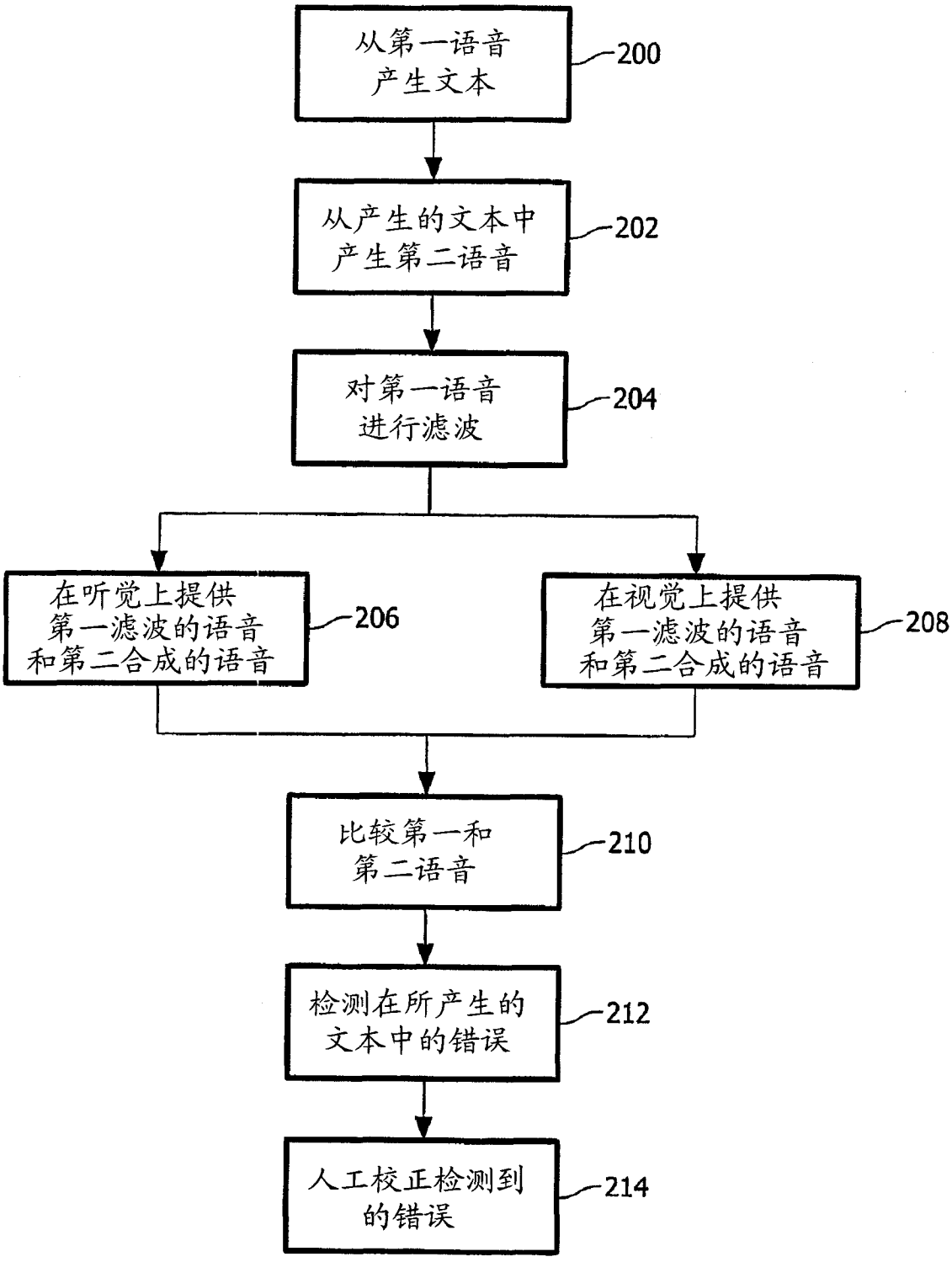

从第一语音
产生文本
200
从产生的文本中
产生第二语音
202
对第一语音
进行滤波
204
在听觉上提供
第一滤波的语音
和第二合成的语音
206
在视觉上提供
第一滤波的语音
和第二合成的语音
208
比较第一和
第二语音
210
检测在所产生的
文本中的错误
212
人工校正检测到
的错误
214

图 2

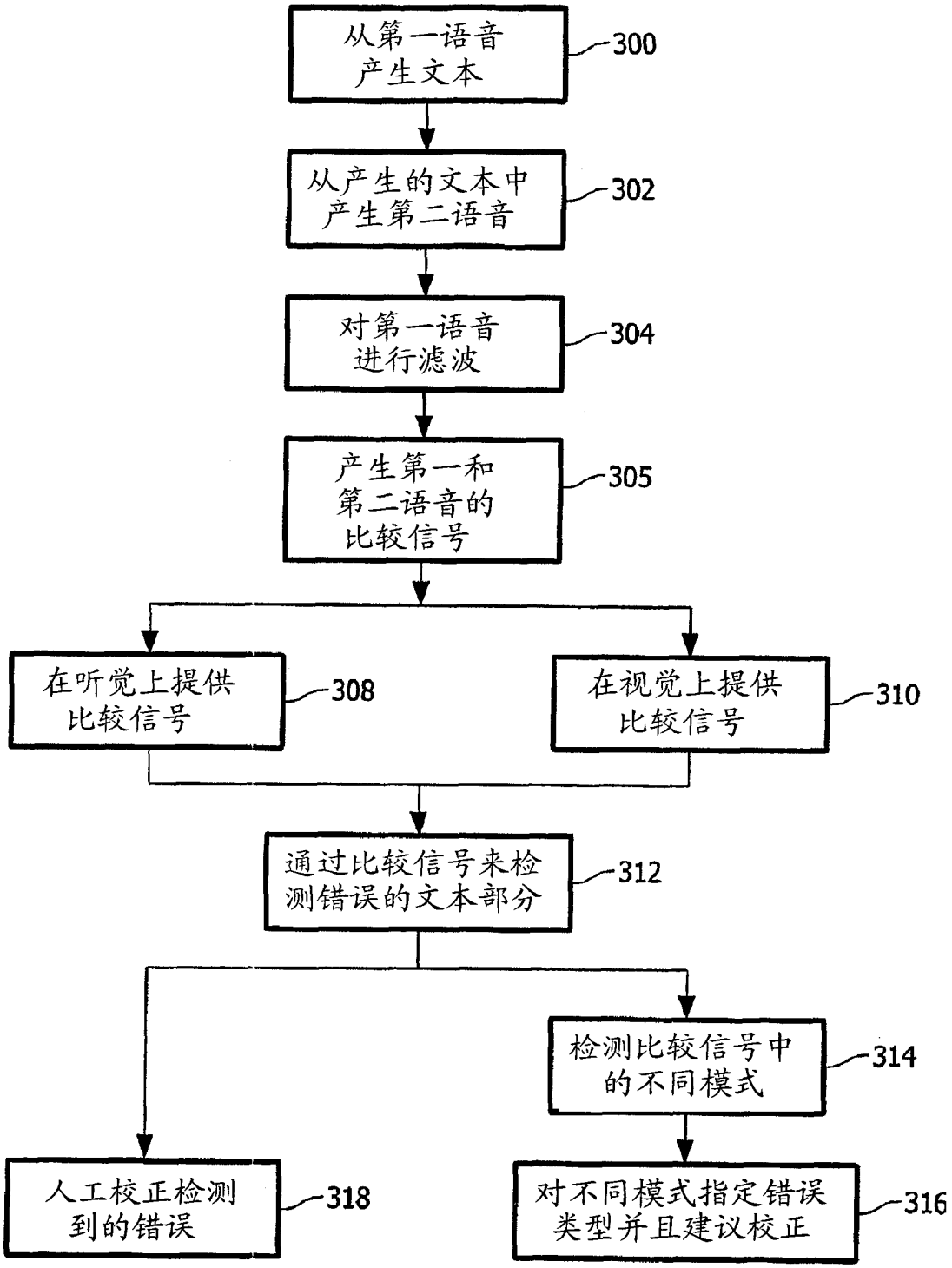

从第一语音产生文本
300
从产生的文本中产生第二语音
302
对第一语音进行滤波
304
产生第一和第二语音的比较信号
305
在听觉上提供比较信号
308
在视觉上提供比较信号
310
通过比较信号来检测错误的文本部分
312
检测比较信号中的不同模式
314
人工校正检测到的错误
318
对不同模式指定错误类型并且建议校正
316

图 3

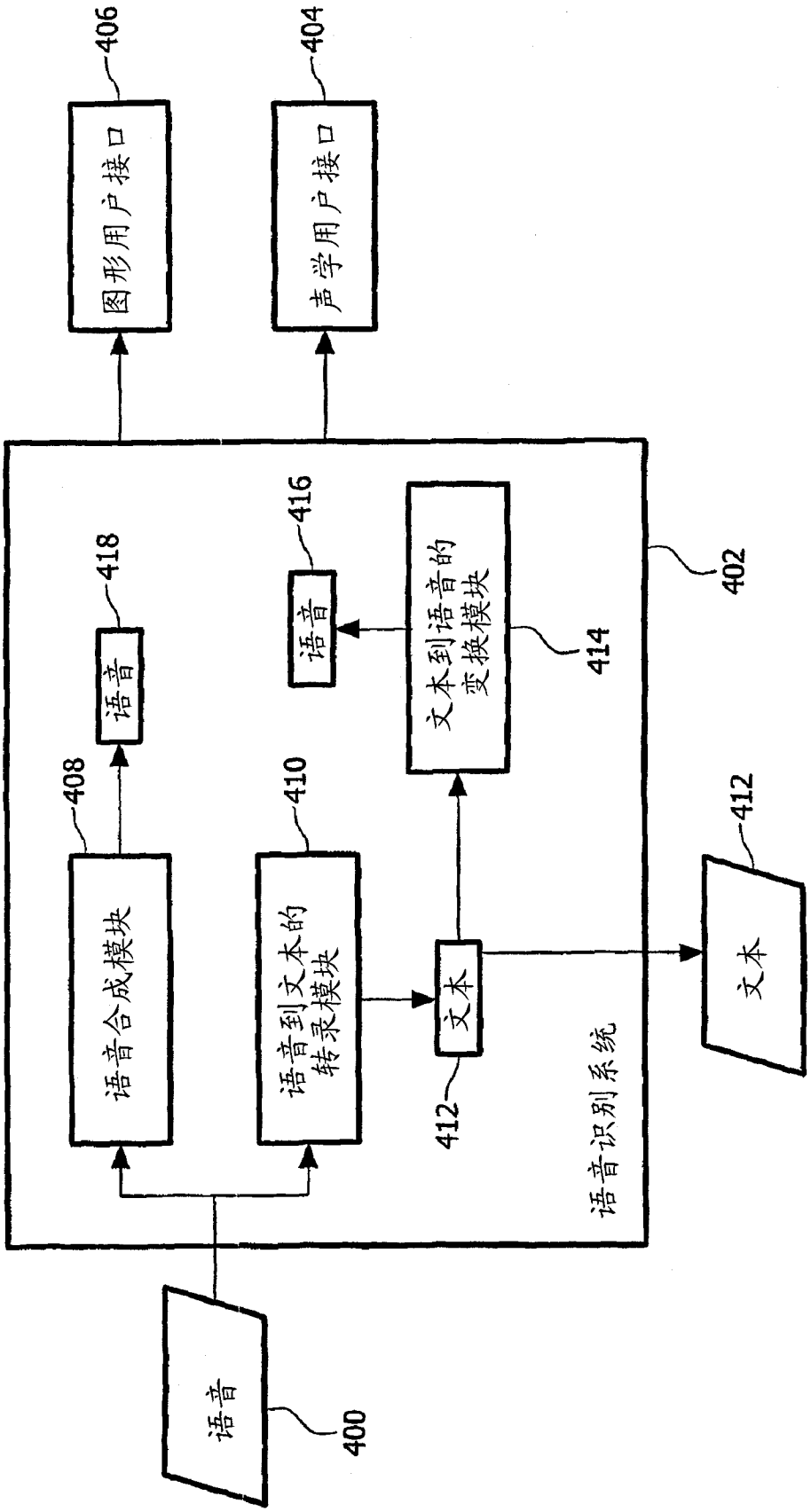

406
图形用户接口
404
声学用户接口
418
语音
416
语音
408
语音合成模块
410
语音到文本的转录模块
文本到语音的变换模块
414
402
412
文本
412
文本
语音识别系统
语音
400

图 4