



(12) 发明专利

(10) 授权公告号 CN 1540625 B

(45) 授权公告日 2010.06.09

(21) 申请号 200410032631.8

US 6141642 A, 2000.10.31, 说明书第4栏第

(22) 申请日 2004.03.24

22-23、65-67行、第5栏第1-15、30-65行。

(30) 优先权数据

审查员 时鹏

10/396,944 2003.03.24 US

(73) 专利权人 微软公司

地址 美国华盛顿州

(72) 发明人 初敏 彭湖 赵勇

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 骆希聪

(51) Int. Cl.

G10L 13/08 (2006.01)

G10L 13/00 (2006.01)

G10L 15/26 (2006.01)

(56) 对比文件

US 6151576 A, 2000.11.21, 说明书第4栏第
22-23、65-67行、第5栏第1-15、30-65行。

EP 1213705 A2, 2001.12.03, 全文。

EP 0984426 A2, 1999.08.31, 全文。

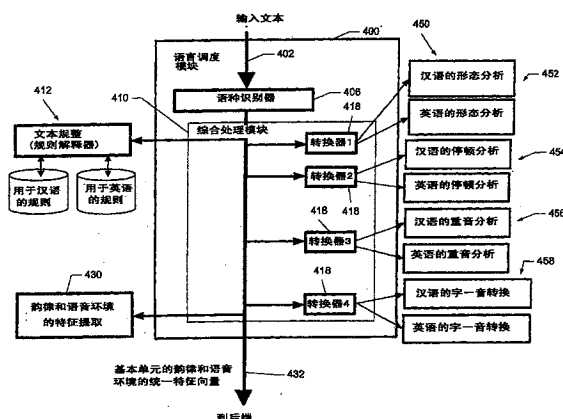
权利要求书 2 页 说明书 7 页 附图 7 页

(54) 发明名称

多语种文本 - 语音系统的前端结构

(57) 摘要

一种用于处理语音合成器的多语种文本的文本处理系统,包括的第一语种相关模块,用于对包括第一种语言的输入文本的一部分进行文本和韵律分析中的至少一个。第二语种相关模块对包括第二种语言的输入文本的第二部分进行文本和韵律分析中的至少一个。第三模块适合于接收第一和第二语种相关模块的输出,并且基于多语种文本对输出进行韵律和语音环境的提取。



1. 一种用于处理语音合成器的多语种文本的文本处理系统,该多语种文本包含由至少两种语言混合组成的句子,该文本处理系统包括:

第一语种相关模块,用于对包括第一种语言的输入文本的第一部分进行文本和韵律分析中的至少一个;

第二语种相关模块,用于对包括第二种语言的输入文本的第二部分进行文本和韵律分析中的至少一个;以及

第三模块,用于接收第一和第二语种相关模块的输出,并且基于多语种文本对输出的每一完整句子进行韵律和语音环境的分析。

2. 如权利要求1所述的文本处理系统,还包括一个文本规整模块,用于对文本进行规整以便于第一语种相关模块和第二语种相关模块进行处理。

3. 如权利要求1或2所述的文本处理系统,还包括一个语种识别器模块,该语种识别器模块适合于接收多语种文本,并且把包括第一种语言的部分的标识符和包括第二种语言的部分的标识符相关联。

4. 如权利要求3所述的文本处理系统,还包括一个综合处理器模块,该综合处理器模块适合于接收每个模块的输出并适当地将所述处理的输出转发到另一个模块。

5. 如权利要求4所述的文本处理系统,其中根据关联的标识符的功能,所述综合处理器将所述输出转发到第一语种相关模块和第二语种相关模块。

6. 如权利要求5所述的文本处理系统,其中所述的第一语种相关模块和第二语种相关模块适合于进行语言形态分析。

7. 如权利要求5所述的文本处理系统,其中所述的第一语种相关模块和第二语种相关模块适合于进行停顿分析。

8. 如权利要求5所述的文本处理系统,其中所述的第一语种相关模块和第二语种相关模块适合于进行重音分析。

9. 如权利要求5所述的文本处理系统,其中所述的第一语种相关模块和第二语种相关模块适合于进行字形-至-语音转换。

10. 一种用于语音合成器中多语种文本的文本处理的方法,该多语种文本包含由至少两种语言混合组成的句子,该方法包括:

接收输入文本,并标识包括第一种语言的部分和包括第二种语言的部分;

利用第一语种相关模块对包括第一种语言的部分进行文本和韵律分析中的至少一个,并且利用第二语种相关模块对包括第二种语言的部分进行文本和韵律分析中的至少一个;以及

接收第一和第二语种相关模块的输出,并且基于多语种文本对输出的每一完整句子进行韵律和语音环境的分析。

11. 如权利要求10所述的方法,还包括对输入文本进行规整。

12. 如权利要求10或11所述的方法,其中所述标识部分包括把标识符与各个部分相关联。

13. 如权利要求12所述的方法,还包括根据与部分相关联的标识符,将所述部分转发给第一语种相关模块和第二语种相关模块。

14. 如权利要求10、11、12或13所述的方法,还包括根据文本中的顺序来标识文本的部

分。

15. 如权利要求 12 所述的方法,还包括根据文本中的顺序来标识文本的部分。

16. 如权利要求 10、11 或 13 所述的方法,其中执行韵律和语音环境的提取包括输出所述多语种文本的韵律的符号描述。

17. 如权利要求 12 所述的方法,其中执行韵律和语音环境的提取包括输出所述多语种文本的韵律的符号描述。

18. 如权利要求 14 所述的方法,其中执行韵律和语音环境的提取包括输出所述多语种文本的韵律的符号描述。

19. 如权利要求 10、11 或 13 所述的方法,其中执行韵律和语音环境的提取包括输出所述多语种文本的韵律的数字描述。

20. 如权利要求 12 所述的方法,其中执行韵律和语音环境的提取包括输出所述多语种文本的韵律的数字描述。

21. 如权利要求 14 所述的方法,其中执行韵律和语音环境的提取包括输出所述多语种文本的韵律的数字描述。

22. 一种适合于实施权利要求 10-21 的方法中的任何一种方法的系统。

多语种文本 – 语音系统的前端结构

[0001] 发明背景

[0002] 本发明涉及语音合成,尤其是涉及一种多语种语音合成系统。

背景技术

[0003] 已经开发的文本 – 至 – 语音系统能够使计算机化的系统通过合成语音与用户交流,这些应用系统包括口语对话系统、呼叫中心服务系统、语音激活的网络和电子邮件服务等。虽然在过去几年里文本 – 至 – 语音系统有了很大发展,但是还存在一些不足之处。例如,许多文本 – 至 – 语音系统都仅仅是基于一种语言而设计的,而许多应用都需要有一种系统能够提供多种语言词汇的语音合成,尤其是在同一句话中包含两种以上语言的词汇的语音合成。

[0004] 目前已经研制出来的提供包含多种语言词汇的语句的语音合成的系统需要使用不同的文本 – 至 – 语音引擎以合成句子中的每一种相应的语言的词汇,其中每个引擎分别生成其所合成词汇的波形。然后将波形连接或是用其他方式连续地输出所述的波形,以便合成出完整的句子来。这种设备的主要缺点在于从两个引擎产生的声音听起来通常会有差别。这种合成出来的话语通常会令用户感到厌烦,因为听起来像两个不同的讲话者在讲话。并且,整个句子的语调也被破坏了,导致句子的可懂度降低。

[0005] 因此,一种能克服上述至少一部分缺点的多语种语音合成系统将是有用的,并且能够改善多语种语音合成。

发明内容

[0006] 一种用于语音合成系统的文本处理系统接收包含至少两种语言的混合输入文本,提供适合于语音合成器的后端部分使用的输出。一般地,该文本处理系统包括执行文本处理的语种无关模块和语种相关模块。这种体系结构有利于在不同语种间平滑切换,并且能够使混合语言的语句的语调保持流畅。

附图说明

[0007] 图 1 是能够实施本发明的一般计算机环境的方框图。

[0008] 图 2 是能够实施本发明的一个移动装置的方框图。

[0009] 图 3A 是一个现有技术的语音合成系统的第一实施例方框图。

[0010] 图 3B 是一个现有技术的语音合成系统的第二实施例方框图。

[0011] 图 3C 是一个现有技术的语音合成系统的前端部分的方框图。

[0012] 图 4 是包括语音合成器的文本处理系统的本发明第一实施例的方框图。

[0013] 图 5 是包括语音合成器的文本处理系统的本发明第二实施例的方框图。

具体实施方式

[0014] 在对本发明进行说明之前,有必要先介绍一下可以应用本发明的典型的计算机环

境。图 1 给出了一个可以实施本发明的适合的计算机系统环境 100。该计算机系统环境 100 只是可以实施本发明的计算机系统环境的一个实施例,而并非是对本发明的应用范围或功能进行任何限制。计算环境 100 也不应被认为是对在实例性操作环境 100 中所示的任何一个部件或者部件的组合有任何依赖性或要求。

[0015] 本发明可以用于众多的特定或非特定目的的计算系统环境或配置。常见的适于应用本发明的计算机系统、环境或配置包括(但不限于)如下系统或装置:个人计算机,服务器计算机,手持或膝上型装置,多处理器系统,基于微处理器的系统,机顶盒,可编程电子消费装置,网络 PC,小型计算机,大型计算机,包括任何上述的系统或装置的分布式计算环境,等等。

[0016] 本发明可以使用计算机可执行指令的一般模式来描述,例如计算机执行的程序模块。一般地,程序模块包括子程序、程序、对象、组件、数据结构等,它们用来执行特定的任务或者实现特定的抽象数据类型。本发明也可以被应用于分布式计算环境,其中通过利用通信网络链接的远程处理装置来执行任务。在分布式计算环境中,程序模块可以同时存放在包括存储器存储装置的本地和远程计算机存储介质中。下面就对照附图来说明程序和模块所要执行的任务。本领域普通技术人员能够将下述说明和附图实施成写入任何形式的计算机可读介质的处理器可执行指令。

[0017] 图 1 给出了一个实施本发明的典型系统包括一个用于非特定目的的计算装置,该计算装置是一台计算机 110。计算机 110 的部件包括(但不限于)一个处理单元 120、一个系统存储器 130 和用于将包括系统存储器在内的各个系统部件连接到处理单元 120 的系统总线 121。系统总线 121 可以是包括如下几种类型的总线结构中的任何总线结构:一个存储总线或存储控制器、一个外围总线和使用各种总线结构中的任何总线的局部总线。这些总线结构例如但不限于工业标准结构 (ISA) 总线、微通道体系结构 (MCA) 总线、增强的 ISA 总线、视频电子标准协会 (VESA) 局部总线以及外部设备互连 (PCI) 总线(也叫夹层总线 (Mezzanine bus))。

[0018] 典型地,计算机 110 包括一些计算机可读介质。计算机可读介质是能被计算机 110 访问的任一可用介质,包括非易失性介质和易失性介质,可移动介质和不可移动介质。例如但不限于,计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括使用任何方法或技术实现信息存储的非易失性的和易失性的、可移动的和不可移动的介质,这些信息可以是计算机可读指令、数据结构、程序模块或者其他数据。计算机存储介质包括但不限于随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除只读存储器 (EEPROM)、闪存 (flash memory) 或其他存储器技术、只读光盘存储器 (CD-ROM)、数字多功能光盘 (DVD) 或其他光盘存储器、盒式磁带、磁带、磁盘存储器或其他磁存储装置,或者其他任何可以用于存储所述信息并且可被计算机 100 访问的介质。

[0019] 通信介质包括任何信息传输介质,通常包含计算机可读指令、数据结构、程序模块或其他存在于如载波或其他传输技术使用的调制的数据信号中的数据。术语“调制的数据信号”是指以一种方式设置或改变它的一个或多个特征以在其中对信息进行编码的信号。通信介质包括例如但不限于有线介质(例如有线网络或直接有线连接)以及无线介质(例如声波、FR、红外线和其它的无线介质。以上所述介质的任何组合也包括在所述计算机可读介质的范围之内。

[0020] 系统存储器 130 包括非易失性和易失性的计算机存储介质,例如只读存储器 (ROM) 131 和随机存取存储器 (RAM) 132。典型地,基本输入 / 输出系统 133 (BIOS) 存储在只读存储器 131 中,其中包含用于协助在计算机 110 内的元件之间例如在开机过程中传递信息的基本程序。典型地,随机存取存储器 132 中存储操作单元 120 所要立即访问和 / 或即将操作的数据和 / 或程序模块。图 1 中举例性地而不是限制性地示出了操作系统 134、应用程序 135、其他程序模块 136 以及程序数据 137。

[0021] 计算机 110 也可以包括其他的可移动 / 不可移动的、易失性 / 非易失性的计算机存储介质。仅仅是作为一个例子,图 1 中给出了用于读取或写入不可移动的、非易失性的磁介质的一个硬盘驱动器 141、用于读取或写入可移动的、非易失性的磁盘 152 的一个磁盘驱动器 151、用于读取或写入可移动的、非易失性的光盘 156 如 CD ROM 或其他光学介质的一个光盘驱动器 155。其他可以被用在所述的典型操作环境中的可移动 / 不可移动、易失性 / 非易失性的计算机存储介质包括但不限于盒式磁带、闪存卡、数字多功能光盘、数字录像带、固态 RAM、固态 ROM 等等。硬盘驱动器 141 典型地通过不可移动存储接口例如接口 140 与系统总线 121 相连接,磁盘驱动器 151 和光盘驱动器 155 典型地通过移动存储接口例如接口 150 与系统总线 121 相连接。

[0022] 以上所述并显示在图 1 中的驱动器以及与其相应的计算机存储介质用于为计算机 110 存储计算机可读指令、数据结构、程序模块及其它数据。例如,图 1 中的硬盘驱动器 141 作为存储操作系统 144、应用程序 145、其它程序模块 146 以及程序数据 147 示出。要注意的是这些组件可以相同或不同于操作系统 134、应用程序 135、其它程序模块 136 以及程序数据 137。这里,操作系统 144、应用程序 145、其它程序模块 146 以及程序数据 147 用不同于上述组件的附图标记来表示以至少说明它们是不同的拷贝。

[0023] 用户可以通过输入装置向计算机 110 输入命令和信息,这些输入装置可以是键盘 162、麦克风 163 以及指点装置如鼠标、轨迹球或触摸板。还可以是其他的输入装置 (图上未画出),例如操纵杆、游戏垫、圆盘式卫星电视天线 (satellitedish)、扫描仪等。上述输入装置通常是通过一个耦合到系统总线的用户输入接口 160 连接到处理单元 120 的,但也可以是通过别的接口和总线结构连接,例如一个并行端口、游戏端口或者通用串行总线 (USB)。监视器 191 或者其他类型的显示装置通过一个接口例如视频接口 190 也连接到系统总线 121。除了该监视器外,计算机也可以包括其他的外部输出设备例如扬声器 197 以及打印机 196,它们通过一个外部输出接口 190 连接。

[0024] 计算机 110 可以通过逻辑连接的方式连接到一台或更多远程计算机 (例如远程计算机 180) 从而在网络环境中进行操作。远程计算机 180 可以是个人计算机、手持装置、服务器、路由器、网络 PC、对等装置 (peer device) 或者其他常用的网络节点,通常包括多个或所有上述与计算机 110 相关的部件。图 1 中所示的逻辑连接包括一个局域网 (LAN) 171 和一个广域网 (WAN) 173,但也可以包括其它网络。这种网络环境常见于办公室、企业范围内的计算机网络、内部网和因特网。

[0025] 当被用于局域网环境时,计算机 110 通过一个网络接口或适配器 170 连接到局域网 171。当被用于广域网的网络环境时,计算机 110 通常包括一个调制解调器 172 或在广域网 173 (比如因特网) 上建立通信的其它装置。调制解调器 172 可以是内置的或者外置的,并通过用户输入接口 160 或其他适当的机构连接到系统总线 121。在网络环境中,所述

与计算机 110 相关的程序模块或其相关部分可以存储在远程存储器存储装置中。图 1 中实例性而非限制性地示了驻留在远程计算机 180 上的远程应用程序 185。在此要指出的是,所示的网络连接仅仅是示例性的给出,并也可以使用在计算机之间建立通信的其它的装置。

[0026] 图 2 给出了又一个示例性的计算环境——移动装置 200 的方框图。移动装置 200 包括微处理器 202、存储器 204、输入 / 输出 (I/O) 组件 206 和用于与远程计算机或其他移动装置通信的通信接口 208。在一个实施例中,上述组件通过适当的总线 210 连接以实现相互之间的通信。

[0027] 存储器 204 以非易失性电子存储器如随机存取存储器 (RAM) 实施,并且具有一个备用电池模块 (图中未画出) 从而能够保证当移动装置 200 断电时存储器 204 中的信息不会丢失。存储器 204 的一部分优选被分配用作程序运行的可寻址存储器,而存储器 204 的另一部分优选用于实现模拟磁盘驱动器的存储的存储功能。

[0028] 存储器 204 包括一个操作系统 212、应用程序 214 以及一个对象仓库 216。在操作中,最好是由处理器 202 从存储器 204 来运行操作系统 212。在一个实施例中,操作系统 212 采用了从微软公司可购买的 **WINDOWS®** CE 商标的操作系统。操作系统 212 最好是专为移动装置设计的,并且具有数据库特征,从而应用程序 214 可以通过一些已公开的应用编程接口和方法来利用该数据库特征。至少部分地响应对所述已公开的应用编程接口和方法的访问,对象仓库 216 中的对象受到应用程序 214 和操作系统 212 的支持。

[0029] 通信接口 208 代表了允许移动装置 200 发送和接收信息的多种装置和技术。该装置包括有线和无线调制解调器、卫星接收器以及广播调谐器等等。移动装置 200 还可以直接连接到计算机上与之交换数据,在这种情况下,通信接口 208 可以是红外无线电收发器或串行或并行通信连接,它们都能够传输流信息。

[0030] 输入 / 输出组件 206 包括多种输入装置 (例如触摸感应屏、按钮、滚轮和麦克风) 以及多种输出装置 (包括音频产生器、振荡装置和显示器)。上述装置仅是作为示例,并不需要都出现在移动装置 200 中,此外,其他的输入 / 输出装置也可以被加入到或者出现在移动装置 200 中,这种情况也包括在本发明的保护范围之内。

[0031] 为了更好地理解本发明,下面对在图 3A 和 3B 分别示出的现有技术的语音合成器 (或语音引擎) 300 和 302 作一简单介绍是有益的。首先参照图 3A,语音合成器 300 包括一个前端部分或者文本处理系统 304,用于处理从 306 接收的输入文本并通过模块 303 完成文本分析和韵律分析。模块 303 的输出 308 中包含对输入文本 306 的韵律的符号描述。输出 308 提供给在引擎 300 的后端部分或者合成模块 312 中的单元选择和拼接模块 310。单元选择和拼接模块 310 利用采样语音单元的语料库 316 产生合成语音波形 314。合成语音波形 314 通过直接语音拼接单元产生,通常不进行任何基频或时长的修改,只要语音语料库 316 已经涵盖了所有合成单元的足够的韵律和频谱变形并且总能找到合适的段。

[0032] 语音合成器 302 也包括文本和韵律分析模块 303,用于接收输入文本 306 并输出其韵律的符号描述 308。此外,由图可知,前端部分 304 还包括一个韵律预测模块 320,该韵律预测模块 320 用于接收韵律符号描述 308 并在输出 322 上提供韵律的数字描述。根据众所周知的做法,韵律预测模块 320 将某些上层韵律约束 (例如词性、语法、重音和强调等) 作为输入并对基频、时长、能量等进行预测,由此生成确定的数值从而构成输出 322。输出 322 被提供给后端部分 312,在这种形式中该后端部分 312 包括一个语音产生模块 326,用于生

成具有与 322 中的韵律数字描述相匹配的韵律特征的合成语音波形 314。这个过程可以通过方式实现：在基于共振峰或者 LPC 的合成器后端设定相应的参数，或者在拼接的后端运用韵律标度算法例如 PSOLA 或 HNM。

[0033] 附图 3C 示出了可以形成在语音合成器 300 和 302 的前端部分 304 中用于提供韵律的符号描述 308 的文本和韵律分析模块 303 的各种模块。典型的处理模块包括一个文本规整模块 340，该文本规整模块 340 用于接收输入文本 306 并将输入文本 306 的货币、日期或其他的部分转换成可读的字词。

[0034] 在文本规整过程中，可以使用形态分析模块 342 对输入文本进行形态分析以确定复数、过去时态等信息。必要的话，还可以使用模块 344 进行句法 / 语义分析以辨别单词的词性 (part of speech (POS)) 或者预测句子的句法 / 语义结构。如果需要的话，进一步的处理可以在模块 346 中完成，模块 346 根据模块 344 的输出（即 POS 标记或句法 / 语义结构）或者简单地依据逗号、句号等来将单字分组成短语。包括重读、重音和 / 或焦点 (focus) 的语义特征通过模块 348 来预测。字形 - 至 - 语音转换模块 350 依据正确的发音将文字转换成语音符号。303 的输出是具有韵律符号描述 308 的语音单元序列。

[0035] 要强调的是文本韵律分析模块 303 的上述组成模块只是示例性的，只是当需要时才被包括在 303 中以用于从前端部分 304 生成后端部分 312 所使用的输出，如图 3A 或 3B 中所示。

[0036] 对于多语种文本，其中每一种语言都要有一个语音引擎 300 或 302 来合成。文本中属于每一种语言的部分都要提供给相应的单一语种语音合成器并分别进行处理，然后再利用合适的硬件将产生的多个输出 314 相互连接或者以其他方式连续的输出。正如在背景技术部分指出的，其缺点在于导致了整句语调的丢失以及一单句话的部分听起来好像是两个或更多不同的讲话者所说的。

[0037] 图 4 给出了语音合成系统的文本和韵律分析系统 400 的第一个实施例，文本和韵律分析系统 400 用于接收含有由一种或至少两种语言混合组成的句子的输入文本 402，并产生适合于通常如图 3A 或 3B 所示的语音合成器的后端部分处理的输出 432。一般地，前端部分 400 包括语种无关模块和语种相关模块以实现如图 3C 中所示的所需功能。这种结构的优点在于能够实现不同语种间的平滑切换以及保持混合语种的语句的语调流畅。图 4 中自上而下的显示出了这种方法的处理流程。

[0038] 在实例性实施例中，文本和韵律分析部分 400 中有一个语言调度模块，该语言调度模块包括一个语种辨别模块 406 和一个综合处理模块。语种辨别模块 406 接收输入文本 402 并包括或联合为其中的句子和 / 或单字加上标志其所属语言的适当的语种标识符 (Id) 或标记。由所示的实施例中，由于中文汉字和英文单词使用明显不同的编码形成输入文本 402，因而能够相对容易地辨别出输入文本 402 中的中文或英文部分。而对于像法语、德语或者西班牙语这些具有某些相同字符的语种，则还需要进一步的处理。

[0039] 含有适当的语种标识符的输入文本被提供到综合处理模块 410，一般地，综合处理模块 410 用于管理语种无关模块和语种相关模块之间的数据流动，以及维持数据的统一流动以保证正确处理所接收到的每个模块的输出。典型地，综合处理模块 410 首先把带有语种标识符的输入文本传送给文本规整模块 412。在所示实施例中，文本规整模块 412 是一个与语种无关的规则解释器。模块 412 包括两个组件，一个是模式识别器，另一个是模式解释

器,该模式解释器用于依据规则将一个匹配模式转换为一个可读的文本串。每个规则由两部分组成,开始部分是一个模式的定义,其余部分是该模式的转换规则。定义部分可以是两种语言所共用的也可以是指定给它们中的一种语言。转换规则通常是特定的语言。如果有新的语种加入,只需要添加针对新语种的新规则即可,而无需改动规则解释模块。正如本领域普通技术人员所理解的是,文本规整模块 412 也可以设在语种辨别模块 410 之前,只要在文本规整模块 412 中提供适当的操作以辨别输入文本中每种语言的词汇即可。

[0040] 当接收到文本规整模块 412 的输出后,综合处理模块 410 将适当的单词和 / 或短语传送到适当的语种相关模块进行文本和韵律分析。在所示实施例中提供了一个汉语普通话模块 420 和一个英语模块 422。汉语模块 420 和英语模块 422 用于处理所有的语种细节处理,例如两种语言的语法和字 - 音转换、汉语的词的分切以及英语的缩写还原等等。在图 4 中,转换器 418 示意性的说明了综合处理器 (integrator) 410 按照语种标识符的指示将部分输入文本传送到适当的语种相关模块的功能。

[0041] 除了语种标识符以外,输入文本 402 的分段中还包括指示其在输入文本 402 中的位置的标识符或与其关联,以使一旦在接收到各个语种无关模块和语种相关模块的输出时综合处理器 410 可以重构该分段的正确顺序,原因是并非所有的分段都是被同一个模块处理。这样就可以进行并行处理,从而加快了对输入文本 402 处理的速度。当然,也可以按在输入文本 402 中出现的顺序逐段地处理输入文本 402。

[0042] 然后,各个语种相关模块的输出还要经过一个用于韵律和语音环境的统一的特征提取模块 430 进行处理。在这种方式下,由于在通过模块 420 和 422 对汉语和英语分段适当地进行文本和韵律分析之后对整个句子的韵律和语音环境进行了分析,因此整个句子的句调也没有丢失。在所示实施例中,文本和韵律分析部分 400 的输出 432 是一个顺序的单元序列 (包括英语单元和汉语普通话单元),并且具有包括韵律和语音环境在内的统一的特征向量。这样,就可以通过如图 3A 中所示的后端部分中提供单元拼接,下面将进一步描述单元拼接的实例性实施例。。可选的,如果需要,文本和韵律分析部分 400 可以与一个适当的语种无关模块连接以用于进行韵律预测 (类似于模块 320) 并提供韵律的数字描述作为输出。再将韵律数字描述提供给后端部分 312,如图 3B 所示。

[0043] 图 5 给出了作为本发明的另一个实施例的一个双语种文本和韵律分析系统 450,其中文本和韵律分析被划分成 4 个孤立的示例性功能模块,包括语形分析 452、停顿分析 454、强调 / 重音分析 456 和字形 - 至 - 语音转换 458。其中每个功能都包括分别支持英语和汉语普通话的两个模块。与图 4 类似,图 5 中也自上而下的显示了对输入文本的处理顺序流程。虽然图中只出现了英语和汉语普通话两种语言,但是显然可以很容易地对所述文本和韵律分析部分 400、450 加以改变以适用于所需的更多语言。此外,还需要注意的是,其他语种相关模块和 / 或语种无关模块也可以根据需要而很容易地集成在所述的文本处理系统结构中。

[0044] 在一个实施例中,后端部分 312 可以使用如提供单元拼接的图 3A 中所示的形式。对于一个包括汉语普通话和英语的多语种系统来说,汉语普通话的最小单元是音节,而英语的最小单元是音素。单元选择算法应该从具有合适韵律的候选单元库中选取音段,以便尽可能地实现自然、完美的拼接。这里需要考虑 7 个韵律约束条件。它们包括:在短语中的位置、在单词中的位置、在音节中的位置、左音调、右音调、在单词中的重音等级以及在短语

中的重音等级。其中,在音节中的位置和在单词中的重音等级仅适用于英语,而左 / 右音调则仅适用于汉语普通话。

[0045] 使用 CART(分类回归树)通过查询韵律约束条件对基本单元的所有样本进行聚类。CART 的分裂依据是使 f_0 均值、 f_0 的动态幅度和时长三个特征的特征的 MSE(均方差)的加权总和的减小最大,其中每个特征的特征的 MSE 被定义为所有样本的特征值与其主叶的均值之间的距离的平方的均值。当树生成以后,位于同一个叶子节点的所有样本具有相似的韵律特征。单元之间拼接的连续性通过使用两个语音约束条件、左右语音环境和平滑代价来保证。拼接代价被定义为在 7 个韵律约束条件、2 个语音约束条件和平滑代价的源点 - 目标距离的加权和。每个韵律 / 语音约束的距离表以及所有部分的权值首先手动设定,而后以如下的方法自动调整,该方法参见“Perpetually optimizing the cost function for unit selection in a TTS system for one single run of MOS evaluation”, Proc. of ICSLP' 2002, Denver, by H. Peng, Y. Zhao and M. Chu。当合成一个语句时,首先利用韵律约束条件为每个单元找到一个样本聚类集(CART 树的一个叶子节点),然后应用 Viterbi 搜索算法来为每个单元查找产生最小总体拼接代价的最佳样本。所选的分段随后被逐个拼接以得到所要合成的语句。较佳地,单元语料库是由同一个双语的发音人的发音来建立的。虽然两种语言采用的单元大小不同,但它们具有相同的单元选择算法和特征集。因此,语音合成器的后端部分可以处理单语种或者双语种混合的单元序列。以上所述的单元样本的选择方法在 2002 年 7 月 25 日公开的专利申请号 20020099547A1、名称为“Method and Apparatus for Speech Synthesis Without Prosody Modification”的美国专利中有更为详尽的描述,在此以引用参考的方式将其全部内容结合在本申请中。

[0046] 虽然本发明是参照特定的实施例来描述的,本领域普通技术人员应能认识到在其形式和细节上所作的任何改变都不脱离本发明的精神和范围。

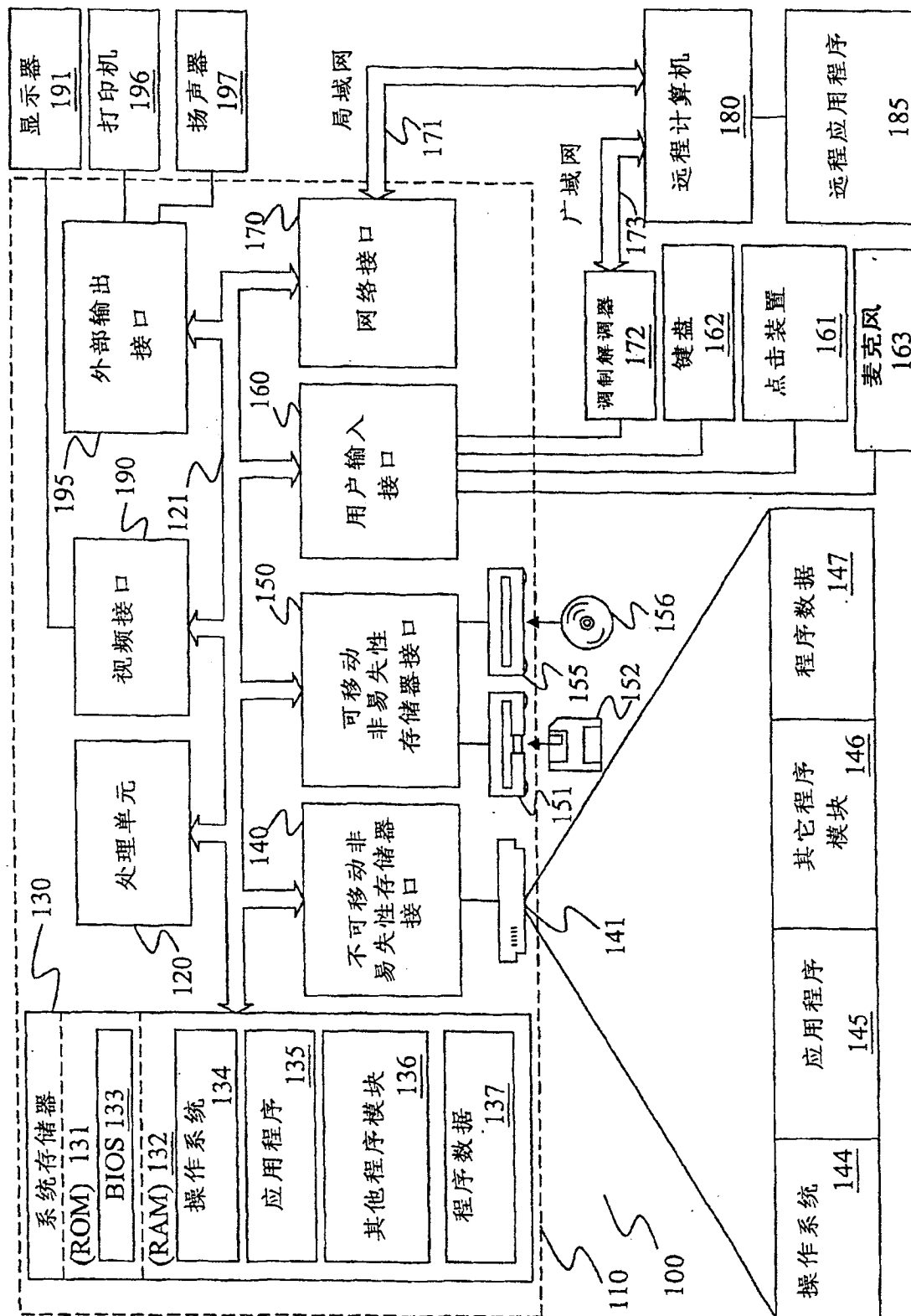


图 1

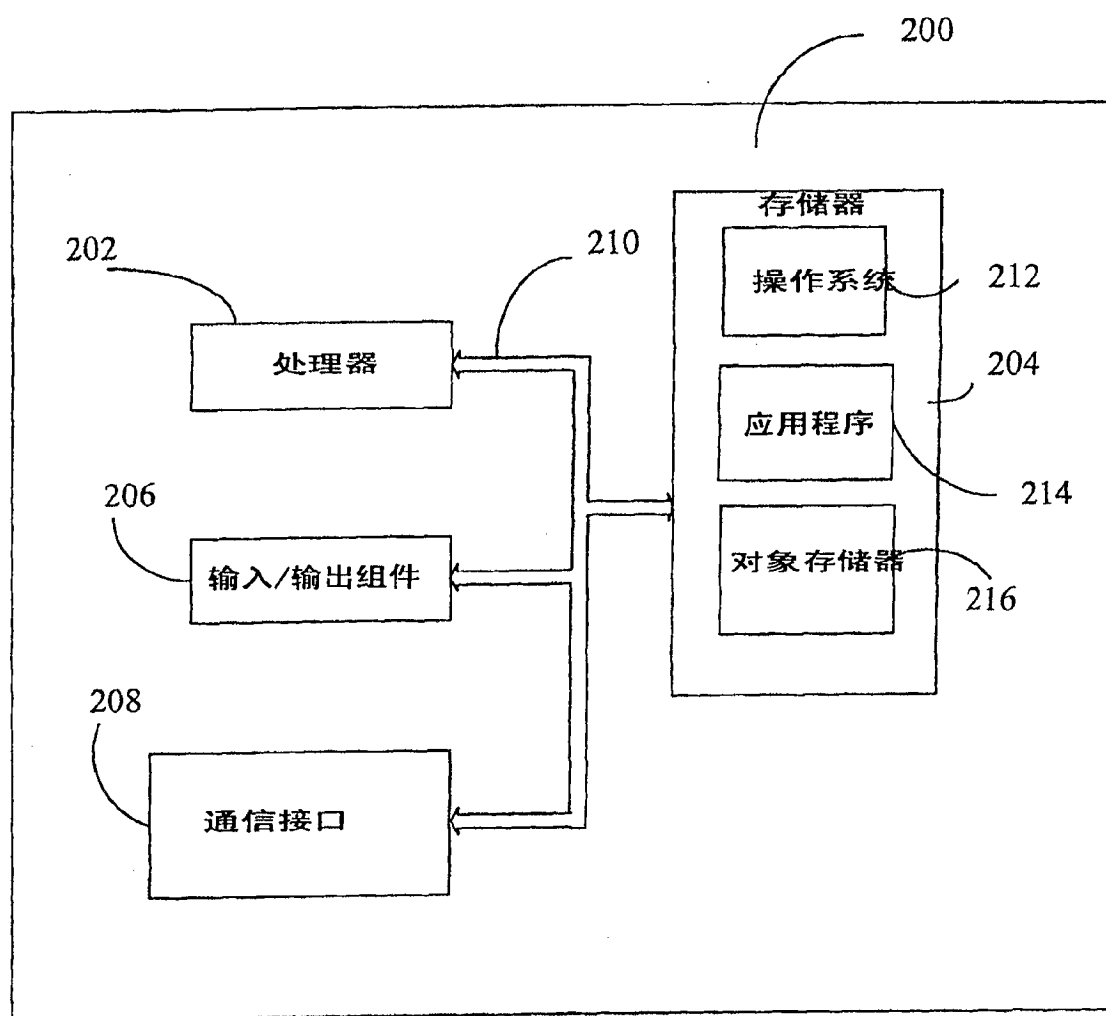
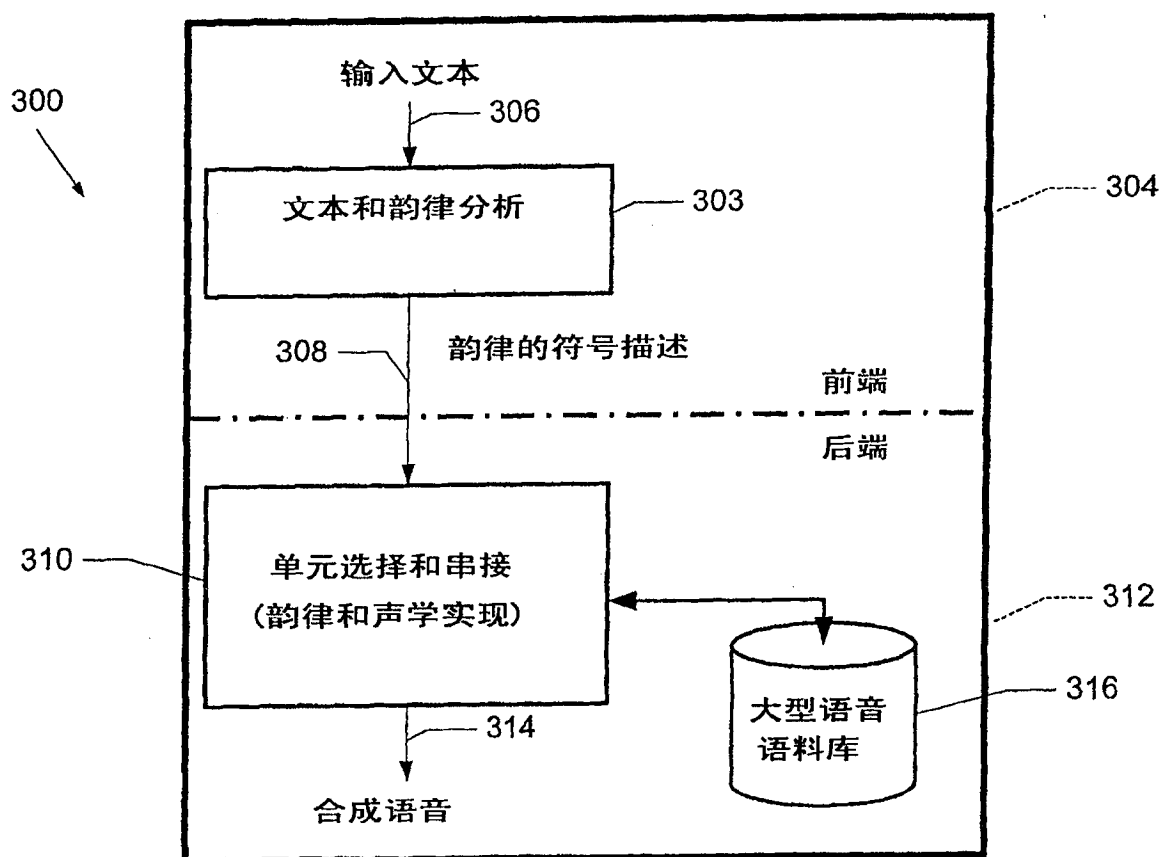
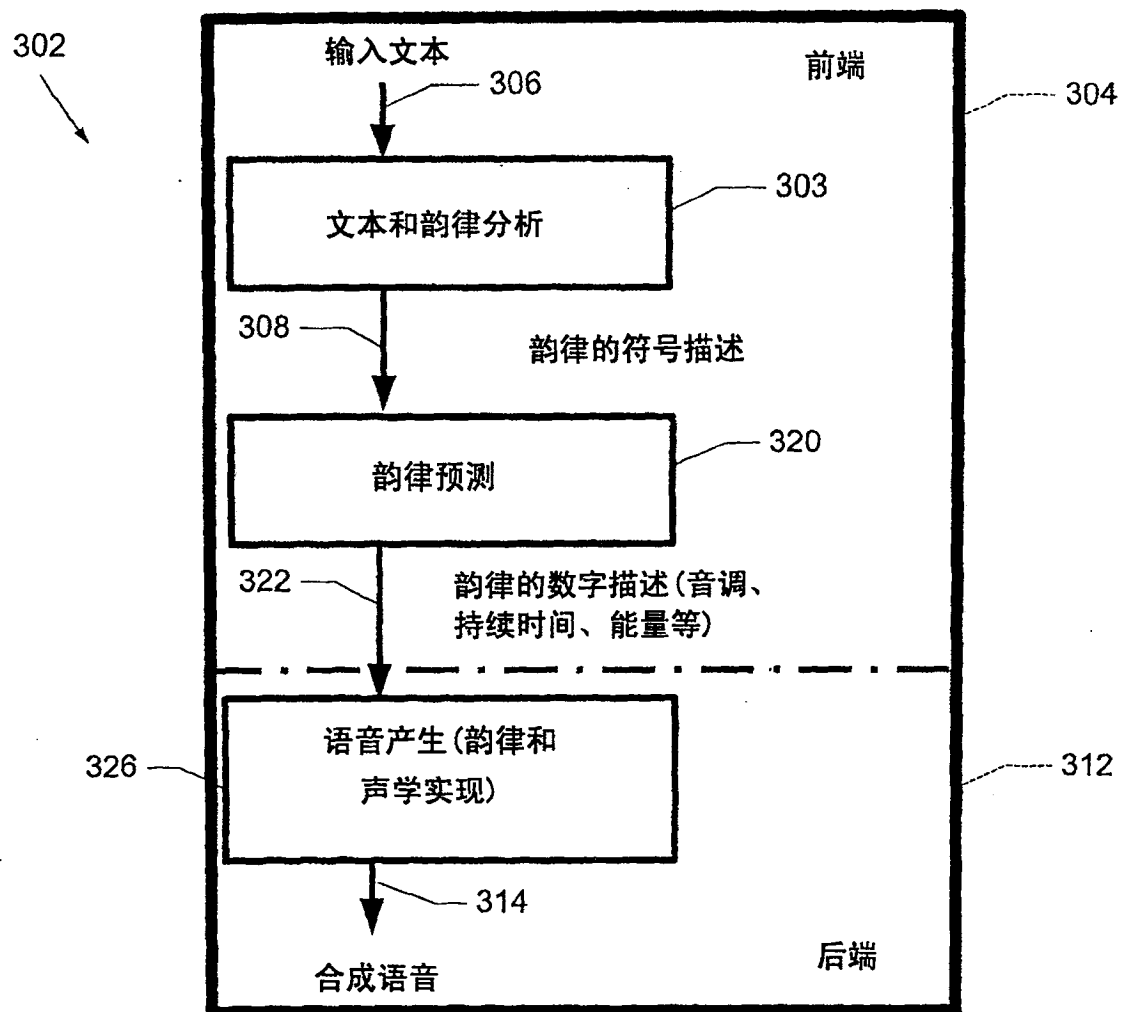


图 2



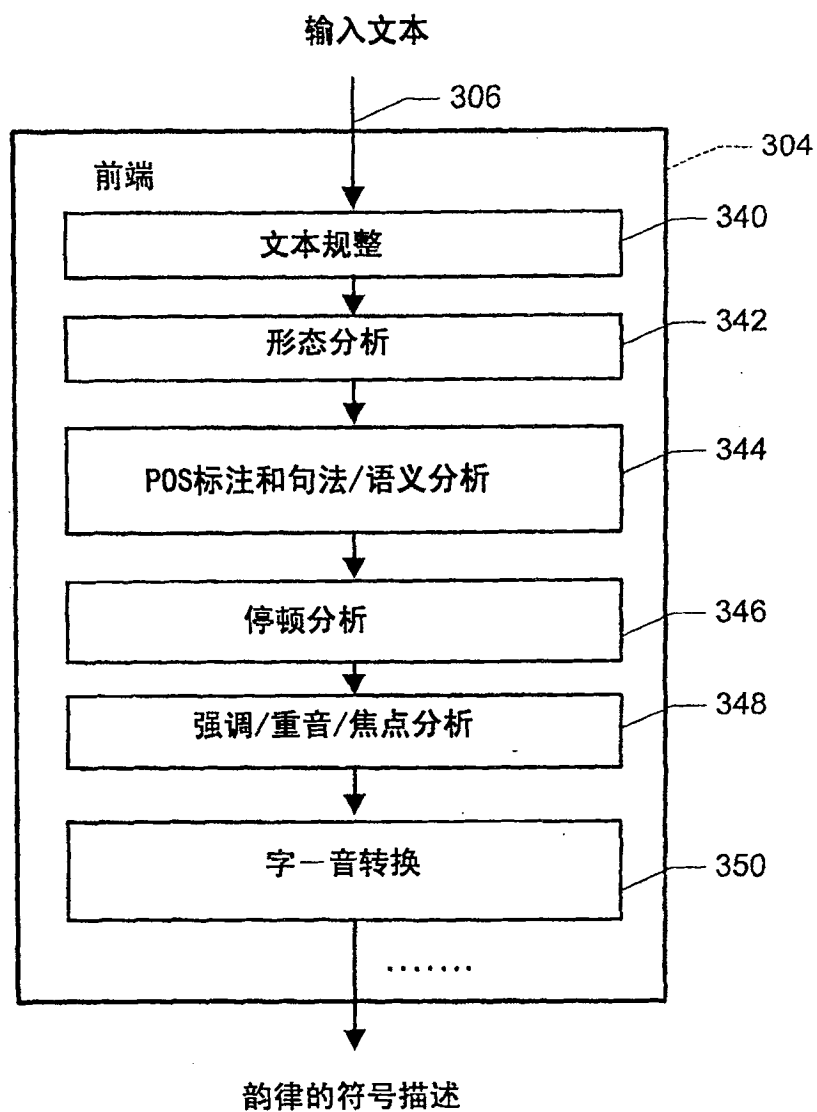
现有技术

图 3A



现有技术

图 3B



现有技术

图 3C

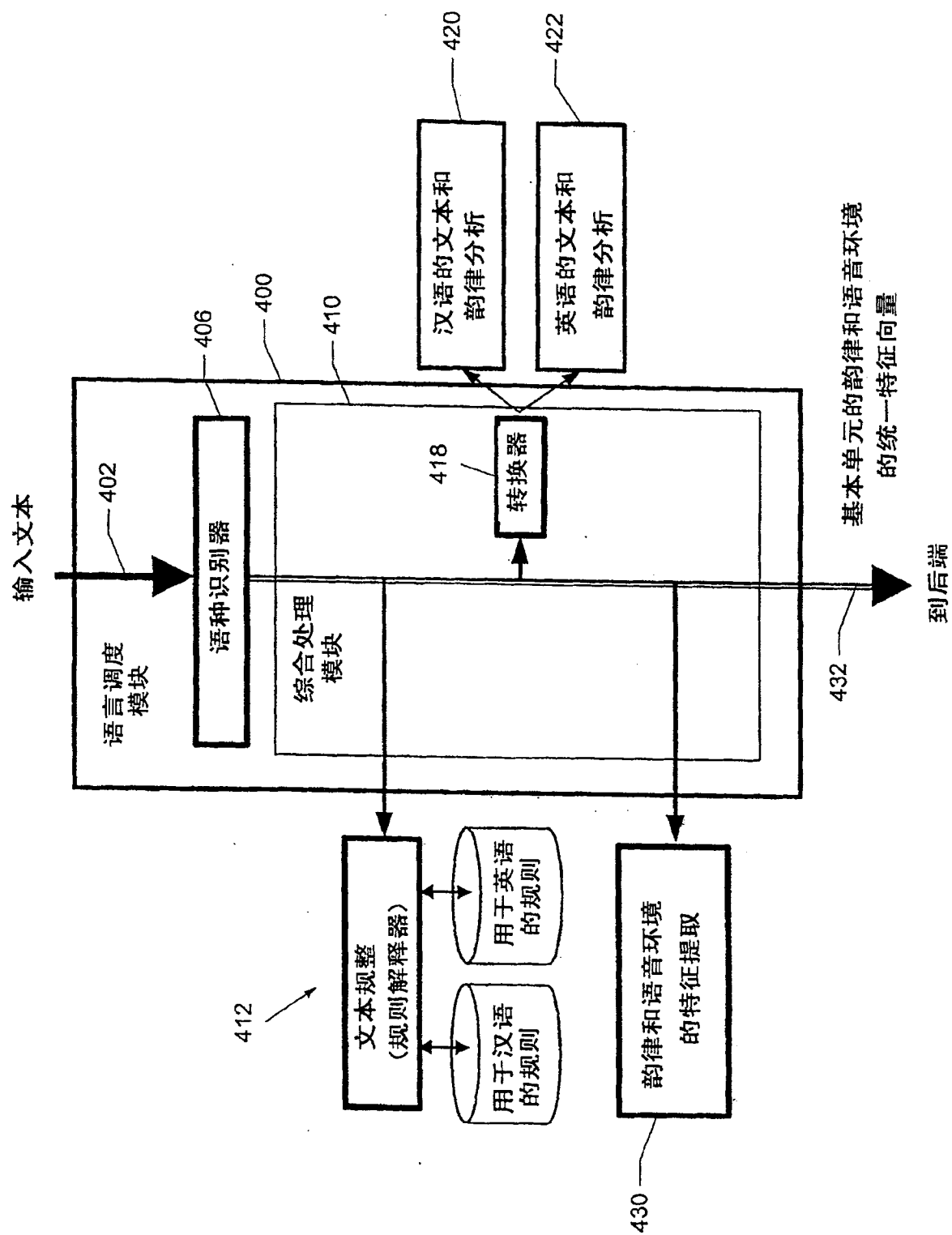


图 4

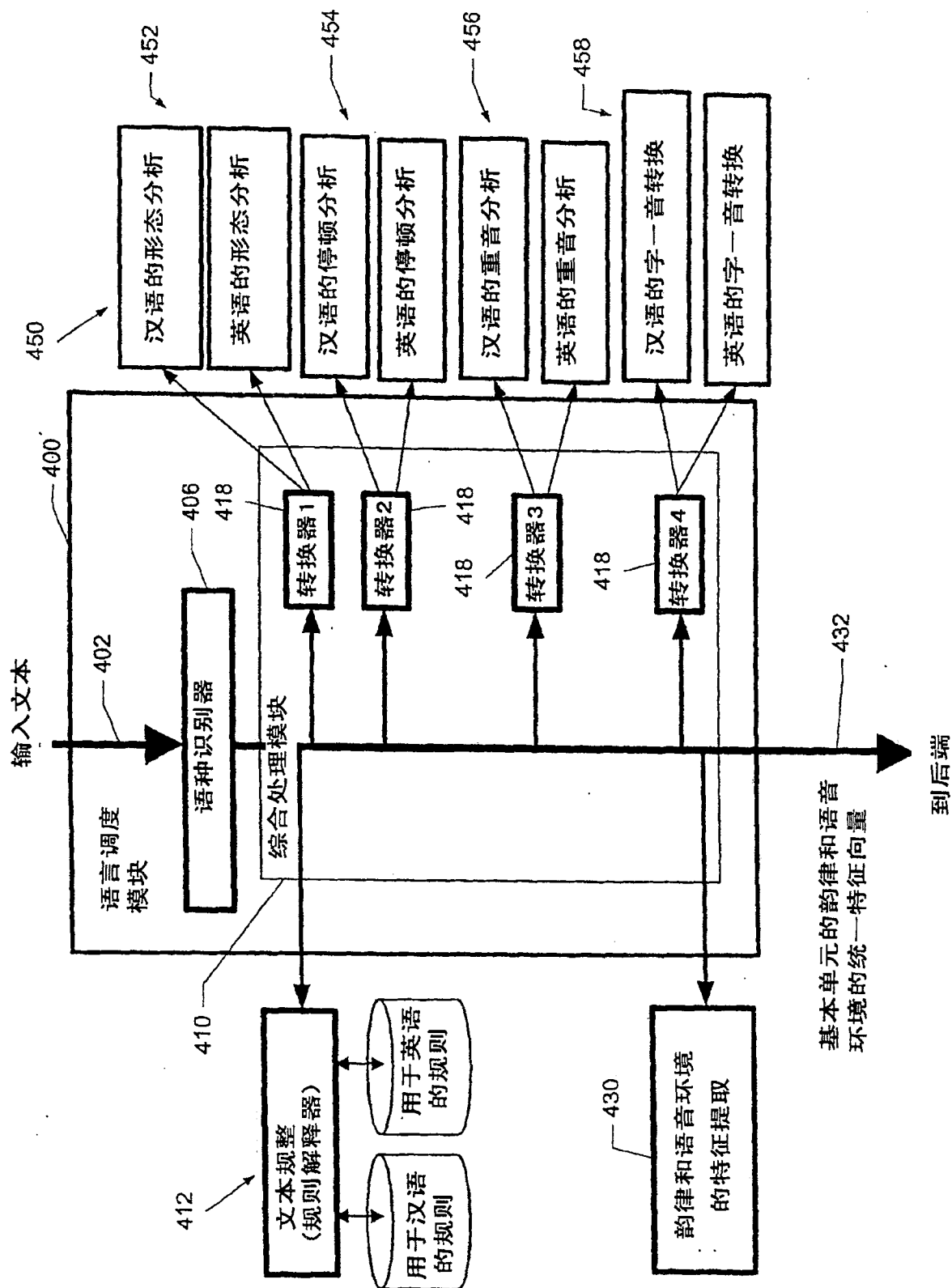


图 5