



## [12] 发明专利申请公布说明书

[21] 申请号 200580016369.0

[43] 公开日 2007 年 12 月 26 日

[11] 公开号 CN 101095287A

[22] 申请日 2005.4.20

[21] 申请号 200580016369.0

[30] 优先权

[32] 2004.4.20 [33] US [31] 60/563,754

[86] 国际申请 PCT/US2005/013478 2005.4.20

[87] 国际公布 WO2005/104092 英 2005.11.3

[85] 进入国家阶段日期 2006.11.21

[71] 申请人 语音信号科技公司

地址 美国马萨诸塞

[72] 发明人 丹尼尔·L·罗森

[74] 专利代理机构 中国国际贸易促进委员会专利商标事务所  
代理人 李春晖

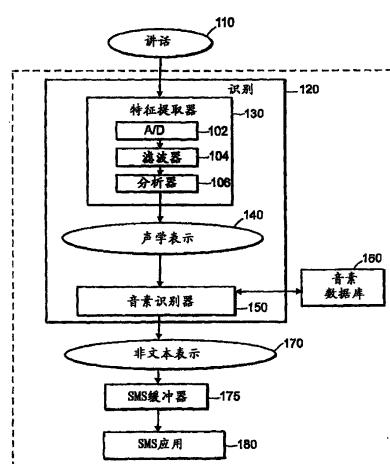
权利要求书 3 页 说明书 11 页 附图 3 页

## [54] 发明名称

基于短消息的话音服务

## [57] 摘要

一种通过移动通信设备发送话音消息的方法，该方法包括：接收来自移动通信设备用户的讲话；生成所接收讲话的非文本表示；将非文本表示插入到文本消息的主体中；并且在无线消息发送信道上将文本消息从移动通信设备发送到接收者的设备。



1. 一种经由移动通信设备发送话音消息的方法，所述方法包括：  
接收来自移动通信设备的用户的讲话；  
生成所接收的讲话的非文本表示；  
将该非文本表示的至少一部分插入到文本消息的主体中；以及  
通过无线消息发送信道，将所述文本消息从所述移动通信设备发  
送到接收者的设备。
2. 如权利要求 1 所述的方法，其中所述移动通信设备是蜂窝电话。
3. 如权利要求 1 所述的方法，其中，插入所述非文本表示的至少  
一部分的步骤涉及将所有非文本表示插入到文本消息的主体中。
4. 如权利要求 1 所述的方法，其中，生成所接收讲话的非文本表  
示的步骤涉及对从所接收的讲话中获得的信号进行识别以生成符  
号串，其中符号串为所述非文本表示。
5. 如权利要求 4 所述的方法，其中，所述符号串中的符号选自音  
素、双音素、三音素、音节和半音节构成的组。
6. 如权利要求 4 所述的方法，其中，所述识别为音素识别，并且  
在所生成的符号串中的符号为音素。
7. 如权利要求 1 所述的方法，其中，生成所接收讲话的非文本表  
示的步骤涉及对从所接收的讲话中获得的信号进行音素识别，以生成  
音素串，其中音素串为所述非文本表示。
8. 如权利要求 1 所述的方法，其中，所述无线消息发送信道为 SMS  
信道，并且所述文本消息为 SMS 消息。
9. 如权利要求 1 所述的方法，还包括：包括文本消息所带的指示  
符，用于标识该文本消息包含讲话的非文本表示。
10. 如权利要求 1 所述的方法，其中，所述非文本表示是所接收  
的讲话的压缩形式。
11. 如权利要求 1 所述的方法，还包括：将非文本表示分成多个  
部分，并且将每个部分插入到多个文本消息的不同相应文本消息的主

体中，其中首先提到的文本消息为所述多个文本消息中的一个，并且还包括：通过无线消息发送信道将多个文本消息从移动通信设备发送到接收者的设备。

**12. 一种操作移动通信设备的方法，所述方法包括：**

在无线消息发送信道上接收文本消息，其中该文本消息包含讲话的非文本表示；

从所述文本消息中提取所述非文本表示；

由所述非文本表示合成口头讲话的音频表示；以及

通过所述移动通信设备上的音频输出装置播放合成的音频表示。

**13. 如权利要求 12 所述的方法，其中所述移动通信设备为蜂窝电话。**

**14. 如权利要求 12 所述的方法，其中，所述讲话的非文本表示是代表该讲话的声音的符号串。**

**15. 如权利要求 14 所述的方法，其中，符号串中的符号选自音素、双音素、三音素、音节和半音节构成的组。**

**16. 如权利要求 14 所述的方法，其中，符号串中的符号为音素。**

**17. 如权利要求 12 所述的方法，其中，所述无线消息发送信道为 SMS 信道，所述文本消息为 SMS 消息。**

**18. 如权利要求 12 所述的方法，其中，所接收的文本消息包括指示符，用于标识该文本消息包含讲话的非文本表示。**

**19. 如权利要求 12 所述的方法，其中，所述非文本表示为讲话的压缩形式。**

**20. 如权利要求 12 所述的方法，还包括：**

通过无线消息发送信道接收除了首先提到的文本消息之外的多个文本消息，所述首先提到的文本消息和所述多个文本消息形成一组文本消息，其中该组文本消息中的每个文本消息都包含讲话的不同部分的非文本表示；

从多个文本消息中提取非文本表示；以及

由提取自该组文本消息的非文本表示合成口头讲话的音频表示。

21. 一种用于发送话音消息的移动通信设备，所述移动通信设备包括：

处理器系统；

用于接收来自移动通信设备用户的讲话的麦克风；

收发器；以及

存储有代码的存储器，该代码当在处理器系统上执行时使得该移动通信设备：

生成所接收讲话的非文本表示；

将该非文本表示插入到文本消息的主体中；以及

经由所述收发器，通过无线消息发送信道将所述文本消息从该移动通信设备发送到接收者的设备。

22. 一种用于接收话音消息的移动通信设备，所述移动通信设备包括：

处理器系统；

收发器，用于接收包含讲话的非文本表示的文本消息；

音频输出装置；和

存储有代码的存储器，该代码当在处理器系统上执行时使得该移动通信设备：

从接收到的文本消息中提取非文本表示；

由该非文本表示合成口头讲话的音频表示；以及

通过音频输出装置播放合成的音频表示。

## 基于短消息的话音服务

本申请要求2004年4月20日提交的美国临时专利申请序列号No.60/563,754的权益，该文献的全部内容在这里被引用作为参考。

### 技术领域

本发明总体涉及在移动通信设备例如蜂窝电话上可用的信道上输送话音消息。

### 背景技术

为了最小化在无线通信网络上传送的话音信息量，因而最大化在任意时刻可以在网络上承受的电话呼叫数量，蜂窝电话采用话音编码器/解码器或编解码器。编解码器从语音信号中去除许多冗余或不必要的信息。然后，将该语音信号的基本元素在网络上传送给正在接收的蜂窝电话，在那里这些元素被解码或者与类似前面已去除信息的数据重新组合。这导致能够由终端用户识别的重构语音。编解码器必须在最小数据传输的需求与在接收端上解码时为保持足够的原始语音信息以听起来自然的需求之间寻求平衡。在通常的话音编解码器中，目前能够将语音信号压缩至每秒4.5k-8k字节之间，并且每秒2.4k字节大致为保持自然发声语音所需的最小速率。尽管能够将语音压缩至这些低比特率，但是用于处理大量话音呼叫的网络基础设施在许多市场中，尤其在发展中国家的新兴市场中是有限的。这使得无线电话的费用很高。

借助蜂窝电话的一种可替代的并且日益流行的通信方法是发送文本消息。由于话音呼叫的高费用，基于文本的所谓SMS(短消息服务)的移动到移动消息发送已经在一些市场中广泛应用，尤其在年轻人群中。SMS使得用户能够在任意时刻发送和接收短文本消息，而与话音

呼叫是否正在进行无关。用户通常通过设在设备上的小键盘输入消息文本。这些消息在硬件方面限制在 160 个字符，并且通过低带宽的带外消息传递信道以数据包的形式发送。这允许在无线网络负担最小的情况下进行便捷通信。

大部分传统的无线网络系统例如 **GSM**、**TDMA** 和 **CDMA** 具有能够发送和接收 **SMS** 的文本/数据信道，因此针对这项服务的基础设施即使在发展中国家的新兴市场中也已经存在。一些评估表明目前发出的全球 **SMS** 消息数量为大约每月 400 亿条消息。人们认为 **SMS** 是目前全球无线网络运营商的基于非话音收入的最主要来源。因此，通信公司对促进 **SMS** 的使用非常感兴趣。实际上，在发展中市场中的网络运营商由于与发送文本消息相关的巨大收入而可能限制采用更先进的话音网络基础设施。

## 发明内容

在一些市场中，蜂窝电话呼叫的费用相对昂贵，从而使得发送文本消息(例如，借助 **SMS**)成为理想的通信替代方案。但是，在这些市场的一部分中，使用文本可能存在其它障碍。发送者和接收者必须都能够阅读和/或书写。但是在新兴市场例如具有非常多人口的印度中，成年人识字率大约为 60%，因此许多人没有足够的文化将文本消息输入进蜂窝电话中。因此，对于在这些市场中不能写作和阅读的许多消费者而言，作为一种通信模式的 **SMS** 文本消息发送不是一种有效的替代方案。在这里所述的至少一些实施方案提供了一种代替文本消息发送的机制，利用这种机制，这些消费者能够使用费用更低的非话音无线通信信道来进行口头通信。

一般来说，在一个方面，本发明提出了一种借助移动通信设备发送话音消息的方法。该方法涉及：接收来自移动通信设备用户的讲话；生成所接收的讲话的非文本表示；将非文本表示插入到文本消息主体中；并且通过无线消息发送信道将文本消息从移动通信设备发送到接收者的设备。

实施方案包括以下特征中的一个或多个。移动通信设备为蜂窝电话。生成所接收讲话的非文本表示涉及对从所接收的讲话中获得的信号进行识别以生成符号串，其中符号串为非文本表示。符号串中的符号选自音素（phoneme）、双音素和三音素(更具体地说，这些符号为音素)构成的组。无线消息发送信道为 SMS 信道，并且文本消息为 SMS 消息。该方法还涉及包括文本消息所带的指示符，用来标识该文本消息包含讲话的非文本表示。非文本表示为所接收的讲话的压缩形式。

一般来说，在另一个方面中，本发明提供了一种在移动通信设备上接收代表讲话的消息的方法。该方法涉及：在无线消息发送信道上接收文本消息，其中文本消息包含讲话的非文本表示；从文本消息中提取非文本表示；并且由非文本表示合成口头讲话的音频表示；并且通过在移动通信设备上的音频输出装置播放合成的音频表示。

实施方案包括以下特征中的一个或多个。移动通信设备为蜂窝电话。生成所接收讲话的非文本表示涉及对从所接收的讲话中获得的信号进行识别以生成符号串，其中符号串为非文本表示。在符号串中的符号选自音素、双音素和三音素(更具体地说，这些符号为音素)构成的组。无线消息发送信道为 SMS 信道，并且文本消息为 SMS 消息。该方法还涉及包括文本消息所带的指示符，用来标识该文本消息包含有讲话的非文本表示。非文本表示为所接收到的讲话的压缩形式。

一般来说，在另一个方面中，本发明提供了一种用于发送话音消息的移动通信设备。该移动通信设备包括：处理器系统；麦克风，用于接收来自移动通信设备用户的讲话；收发器；以及存储有代码的存储器，该代码当在处理器系统上执行时使得移动通信设备：生成所接收讲话的非文本表示；将非文本表示插入到文本消息的主体中；并且经由收发器，通过无线消息发送信道将文本消息从移动通信设备发送到接收者的设备。

一般来说，在另一个方面中，本发明提供了一种用于接收话音消息的移动通信设备。该移动通信设备包括：处理器系统；收发器，用于接收包含有讲话的非文本表示的文本消息；音频输出装置；以及存

储有代码的存储器，该代码当在处理器系统上执行时使得移动通信设备：从接收到的文本消息中提取非文本表示；由非文本表示合成口头讲话的音频表示；并且通过音频输出装置播放合成的音频表示。

在下面的附图和说明中阐述了本发明的一个或多个实施方案的细节。从该说明书和附图中以及权利要求书中将了解本发明的其它特征、目的和优点。

### 附图简要说明

图 1 示出音素识别系统的方框图。

图 2 示出音素合成系统的方框图。

图 3 示出包含音素识别和合成系统的智能话机的高级方框图。

### 具体实施方式

上述实施方案为一种通过在蜂窝电话中可用的SMS网络发送和接收口头或音频信息的方法。用户向蜂窝电话说出所期望的消息或讲话。在电话中的音素识别算法生成讲话的非文本表示。电话中的SMS应用程序通过SMS网络将在SMS消息主体中的该非文本表示发送给接收者的电话。在接收者的电话处，另一个SMS应用程序从SMS消息主体中提取非文本表示。然后，合成器从非文本表示中合成出音频消息，并且向接收者播放该合成消息。

图1显示出高级方框图，更详细地显示出所实施的功能性。用户向蜂窝电话100讲话110，并且在电话内的识别引擎120的前端中的特征提取器130处理该讲话以提取其声学特征。典型地，特征提取器130包括数字转换器102，用来将接收到的模拟信号转换成数字表示。数字转换器102将输入的信号分成一系列重叠帧，然后在每个帧内输出信号的数字表示。滤波器104对该信号的频谱进行过滤，从而尤其降低非语音噪声对语音信号的影响，并且对由用来接收该讲话的信道的频谱特征引起的各种损害进行校正。滤波过程保留了讲话的主要语言内容，同时消除了可能没有携带明显有用信息的各种频率，例如非常高和非常低

的频率。分析器106分析经过滤的数字信号以提取这些帧的相关声学特征即特征矢量。特征提取器130的输出为所接收的讲话的声学表示140。在所述的实施方案中，特征提取器130采用MEL对数逆谱(cepstrum)编码技术来提取相关特征。

在存储器中的数据库160中，电话存储有一组音素，它们是构建口头语言的声音的基本音素单位。它还存储有每个音素的声学模型和识别该音素的索引或指针。声学模型实际上为统计模型，表示在出现一组特定声学特征的情况下说出特定音素的可能性。

在所述实施方案中，识别引擎120采用无约束的音素识别器150来确定在给出表征用户讲话的特征矢量顺序的情况下最可能的音素顺序(即，音素串)。识别器150是无约束的，因为它认为每个候选音素具有相同的权重，而没有针对用户所说的顺序或语言作出假设。换句话说，音素识别器150是相对粗的识别器，它没有采用允许识别说话的语言模型。

识别器150以统计的方式将讲话的声学表示与存储在音素数据库160中的音素的声学表示进行比较。音素数据库160包含足够大的一组音素及其声学表示，以便有效描绘出在用户语言中找到的声音。音素识别器150将所接收讲话的声学表示与音素的声学表示进行统计比较以识别最佳匹配。它是采用被称为隐藏马尔可夫模型(HMM)的公知技术来实现的，也可以采用其它将语音特征与所存储的音素单位进行比较的统计或非统计技术或模型。

音素识别器150将识别出的音素序列以索引或指针序列的形式输出到其音素数据库中。也就是说，对于在识别出的音素串中的每个音素，音素识别器150在其音素数据库中找出表示该音素的特定索引或指针，并且它输出该索引或指针。该输出信号为口头讲话的非文本表示，在该情况下为音素串。该音素串的作用在于，在通信链路的接收端上的合成器能够再现出构成该讲话的声音序列，即它能够如此再现该讲话，从而它通常对于在另一端上的用户是可识别的。但是，典型地，由于没有进行单词识别，所以音素串不如文本那么易读。

音素识别器150将该音素串存储在用于也在蜂窝电话上运行的SMS应用程序180的缓冲器175中。SMS应用程序180生成用于接收非文本表示的文本消息包，并且在其地址域寄存接收者电话的地址。在缓冲器175满了或者讲话完成时，SMS应用程序180将所存储的音素串域表示该消息包含有打算用于在接收端上的合成器的非文本音素串的标识符一起插入到SMS消息的主体中。实际上，该标识符警告了在另一端上的SMS应用程序不要如通常显示给用户的文本消息一样除了该SMS的内容。SMS应用程序180然后通过SMS信道将SMS消息无线传送给接收者的蜂窝电话。

音素识别器150在SMS缓冲器175中还存储有其它信息，这些信息用来改善由接收者的蜂窝电话合成的声音的质量和/或可理解性。例如，它还规定了每个音素的时间长度、其容量以及其它可以用来控制由在接收电话中的合成器所生成的声音质量。在所述的实施方案中，由于音素识别器150还识别出停顿，所以它截去那些识别出的停顿以节省表示该讲话所需的空间。

采用音素识别算法，讲话110通常能够以大约每秒200-700比特或更小的速率压缩成非文本表示170。在通过SMS网络发送时，该网络在许多区域中其单个消息信息量限制为1200比特，这对应于当在接收蜂窝电话上播放时的大约10秒长度的讲话。

要指出的是，上述方案的可选方案可以是在讲话上进行完全语音识别并且在SMS消息主体中发送经识别的文本消息。但是，这需要在电话上具有全容量识别器，并且具有包含有由这种电话所销售的消费者所说的那种文字的字典的词典。这是不切实际的，尤其在该电话打算在例如存在超过350种不同语言的印度这种市场中销售的情况下是不切实际的。在这种环境种进行语音识别的算法将完善和复杂；而且，进行该语音识别所需的资源通常会超过在打算销售给普通人群的廉价蜂窝电话上可用的资源。另一方面，与经识别文本相反，只需要生成所说的话的音素串表示的识别器构造简单的多，并且所需的机上计算和存储资源明显更少。另外，尤其与进行全语音识别所需的文字词汇

相比，支持音素识别所需的音素组较少。实际上，使用通用音素组将使得识别器能够针对在这里所述的目的处理大多数语言。

还应该指出的是，在音素识别器150以统计的方式将讲话170的声学表示的若干片段与音素的声学表示匹配时，最佳匹配音素偶尔会与讲话不正确匹配。例如，识别器可能将“d”声音解译为“t”，因为特征提取器130针对这两个声音所获得的特征是类似的，使得两个声音中没有一个在音素识别器150中比另一个声音明显更匹配。这些错误在语音-文本识别上具有更大的不利影响，但是在里所述的应用中通常具有很小的不利影响。对于收听该合成音频消息的人而言，在正在合成的音素串中存在这些错误不会使播放内容难以理解。实际上，它们甚至不会被注意到。

图2显示出高级方框图，显示出在SMS信道的接收者侧上实现的功能。由接收者操作的蜂窝电话200接收包含有讲话的非文本表示的SMS消息，并且SMS应用程序280处理该消息以便呈现给用户。在所接收到的SMS消息内的标识符将该SMS消息的内容识别为必须由合成器处理以生成音频信号的音素串。换句话说，该标识符使得SMS应用程序与仅仅将消息内容显示给用户的正常文本消息不同的方式处理该消息。SMS应用程序280将该音素串与任意所存储的提供用来控制合成器220以及它产生声音的方式的参数一起传送给在蜂窝电话内的合成器220。与发送者的蜂窝电话一样，接收者的蜂窝电话也包含音素及其声学表示的数据库。构成所接收的音素串的索引或指针识别出数据库中的哪些音素将要被合成，以将音素串组成到音频消息中。合成器通过蜂窝电话扬声器播放代表音素串的声音序列。这样，口头讲话通过SMS消息设施非实时地传送给接收者。

在适当的情况下，可以将SMS应用程序编程以生成多个SMS消息序列以处理更长的讲话，对于该讲话而言非文本表示将不能配合到单个消息主体中。其实，SMS应用程序将音素串“打包”并且将多个SMS消息(或数据包)发送给接受者的蜂窝电话，每个消息包含有整个讲话的一部分。给每个消息加上索引或标签，从而在接收者上的SMS可以

精确重构讲话的完整表示。在该连接的接收者上的SMS应用程序也需要编程以识别出所接收到的非文本表示将要通过将一个以上SMS消息的内容串起来而构成。

在上述实施方案中，用来表示讲话的语音单元为音素。但是，可以采用音素以外的多种其它符号组的任一种。例如，这些符号可以为双音素、三音素、音节、半音节（**demisyllable**），或者任意其它用来有效表示包含在所期望用户的口头讲话内的声音的字符集。

对于目标定位于特定市场的应用而言，可以选择用来最优表示在那个市场中所使用的语言的定制的音素单元词汇。可选的是，可以使用通用音素集，这将使得电话能够识别并且表示大部分语言。

如上所述，音素识别器没有包括全部语言模型，实际上也许根本没有包括任何语言模型。其功能在于简单地识别出在讲话内的声音。为此，不必识别出其含义。因此由识别器生成的所得到的表示是用户很可能无法阅读的非文本表示。但是，这并不意味所选择的符号集可能不包括该符号集当中的词语。例如短的单音节词可能是所选符号集当中的符号或单元。

结果，音素识别算法生成口头讲话的压缩形式。在所述实施方案中，该压缩形式为非文本表示(即，音素串)。实际上，也可以采用其它算法，这些算法仅仅进行压缩而没有进行任何识别。例如，不使用音素识别器，可以使用声音合成机（**vocoder**）来生成压缩表示，然后将该压缩表示插入到SMS消息的主体中。换句话说，可以采用用来生成适于通过SMS或另一条非话音信道发送的非文本表示的任意算法。理想的是，所选的算法能够充分压缩语音，从而可以发送足够长以输送有意义的信息的讲话。在接收端上，需要应用适当的解压缩算法来重建口头讲话的音频形式。

可以将各种特征加入到该系统中以提高可用性。如上所述，一个这种特征在于给予发送用户选择“话音”的选项，在该话音中接收电话给接收用户播放音频消息。通过将表示“话音”参数的附加字符串加入到讲话的非文本表示中来实现该特征，该字符串向合成算法给出指令。

用户能够通过在电话上的菜单驱动界面来选择和/或调节这些参数。这些参数可以用来将合成算法定制为产生所期望的效果。同样，可以包括针对播放速度或对使之听起来更自然的音频消息的其它改变或者发送用户的更多表示的参数。可以应用在该系统中的另一个特征是针对发送和接收用户的音频提示向导，这能够更好地使得没有文化用户操作该系统。在该情况中，两个电话存储有许多预记录消息例如“请在哔哔声之后记录”，“输入你想要发送的人的电话号码”等。电话响应于用户输入以音频的方式播放适当的消息。

能够应用在该系统中的另一个特征是允许用户在讲话长度和再现质量之间区分优先次序。在该情况中，电话可以存储多个算法，这些算法允许改变讲话的非文本表示的长度和质量。在用户记录讲话之前，电话给用户提供长度/质量选择。用户口头地或者通过电话键盘输入其应答；然后电话采用与用户指令对应的算法来处理该讲话。电话然后加入一系列字符，从而针对如何从讲话的非文本表示合成消息向接收电话给出指令。

在所述实施方案中，蜂窝电话为智能电话300，例如由图3的高级功能框图所示的那种。智能电话300为Microsoft PocketPC支持的电话，其核心包括用于处理蜂窝通信功能的基带DSP302(例如包括话音频带和信道编码功能)以及PocketPC操作系统在其上运行的应用处理器304(例如，Intel StrongArm SA-1110)。该电话支持GSM话音呼叫，SMS(短消息服务)发送文本消息、无线电子邮件和桌面式网页浏览以及更多的传统PDA特征。

通过RF合成器306和RF无线电收发器308以及后面的用来通过天线312处理最终阶段RF发射负载的功率放大器模块310来实现发送和接收功能。接口ASIC(专用集成电路)314和音频编解码器316给扬声器、麦克风和其它设在电话中的输入/输出装置例如用于输入指令和信息的数字或字母键盘(未示出)提供接口。DSP302使用闪存318来进行代码存储。Li离子(锂离子)电池320给电话供电，并且与DSP302连接的电源管理模块322管理了在电话内的电源消耗。用于应用程序处理器114的

易失性和非易失性存储器分别以SDRAM324和闪存326形式提供。存储器的这种布置用来保持用于操作系统的代码、用于可定制特征例如电话号码簿的代码以及用于可以包括在该智能电话中的任意应用程序软件的代码，例如上述音素识别、合成器和SMS应用程序代码。它还存储有音素数据库，该数据库包括音素、音素的声学表示和代表音素的符号。

用于该智能电话的视觉显示装置包括用来驱动LCD显示器330的LCD驱动器芯片328。还有时间模块332，用来为在电话内的其它器件提供时钟信号并且提供实时指示符。

所有上述部件包装在适当设计的外壳334内。由于上述智能收集代表许多不同市售电话的总体内部结构并且由于那些电话的内部电路设计通常对于本领域普通技术人员是已知的，所以有关在图3中所示的那些部件及其操作的其它细节并没有提供，并且对于理解本发明而言不是必须的。

一般来说，该设备根本不必是蜂窝电话，而是具有接收讲话、将它转换成讲话的非文本表示并且通过SMS或另一条非话音信道发送它的功能。例如，也可以应用具有麦克风的膝上型计算机、用于生成讲话的非文本表示的适当软件以及利用SMS协议和频率的无线发射器或任意其它具有类似功能的设备。

虽然在上面实施例中给出了SMS网络，但是也可以采用人们通过它可以发送除了话音之外的文本、数据和/或媒体的任意网络。例如，人们也可以采用MMS(多媒体服务)消息发送信道。

还有，上述MEL对数逆谱编码技术仅仅是许多已知用于提取并且表示所接收到的讲话的特征的可选方案的一个例子。代替MEL对数逆谱编码技术，可以采用其它已知技术中的任一种例如LPC对数逆谱系数。可以用来产生非文本表示的两种编码技术的例子为：(1)Takashi Masuko, Keiichi Tokuda, Takao Kobayashi, 在1998 ICASSP处发表的论文“**A Very Low Bit Rate Speech Coder Using HMM with Speaker Adaptation**”以及在2001年12月出版的**Systems and Computers in**

Japan , 卷 32 中 出 现 的 版 本 ; 以 及 (2) 在 下 面 URL :  
<http://www.geocities.com/virag81/docs.html> 处 的 网 页 上 找 到 的 M.  
Habibullah Pagarkar, Lakshmi Gopalakrishnan, Nimish Sheth,  
Rizwana Shaikh, Virag Shah , “Language Independent Speech  
Compression Using Devanagari Phonetics”, 这 两 篇 文 献 在 这 里 被 引 用  
作 为 参 考 。

其 它 方 面 、 变 体 和 实 施 方 案 都 在 所 附 权 利 要 求 的 范 围 内 。

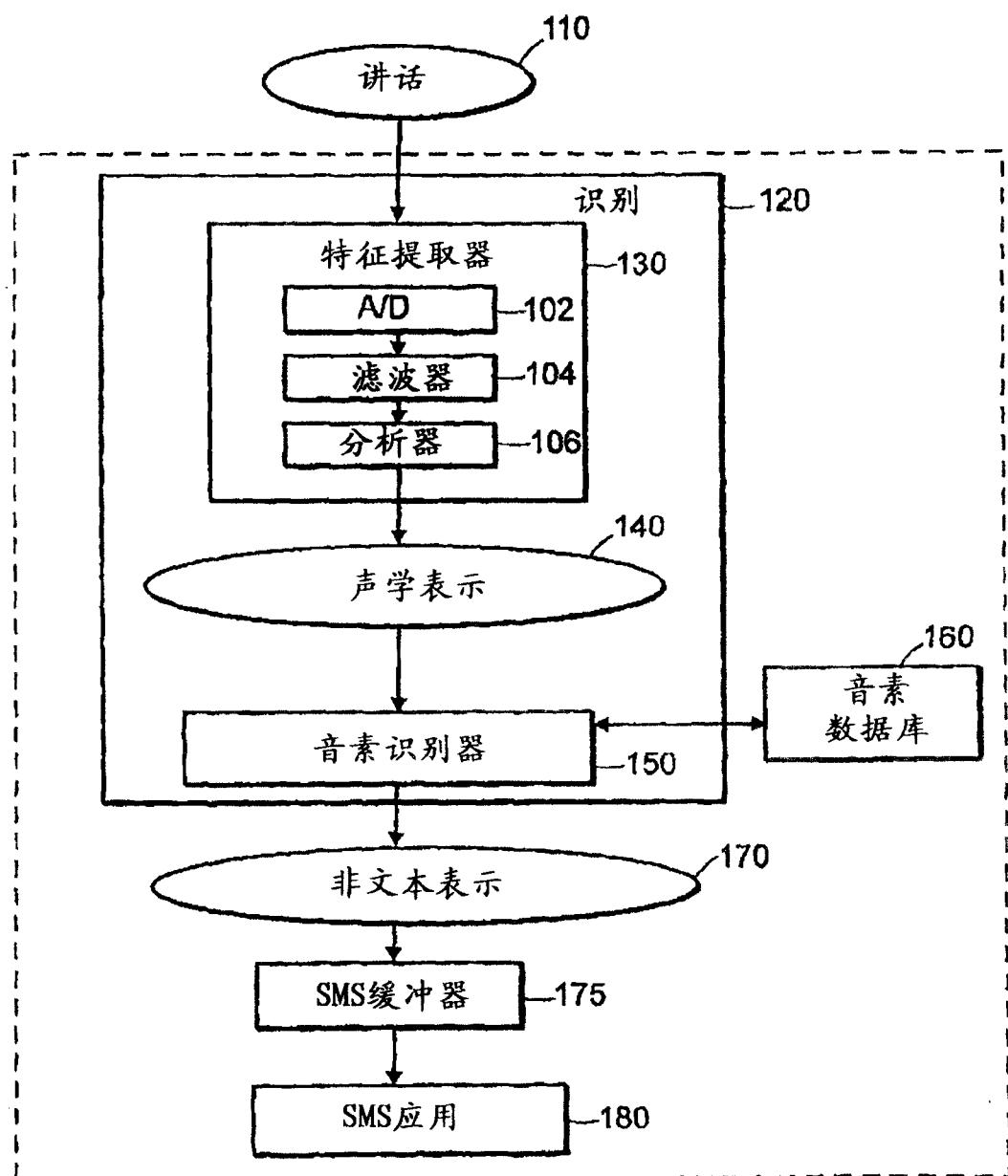


图 1

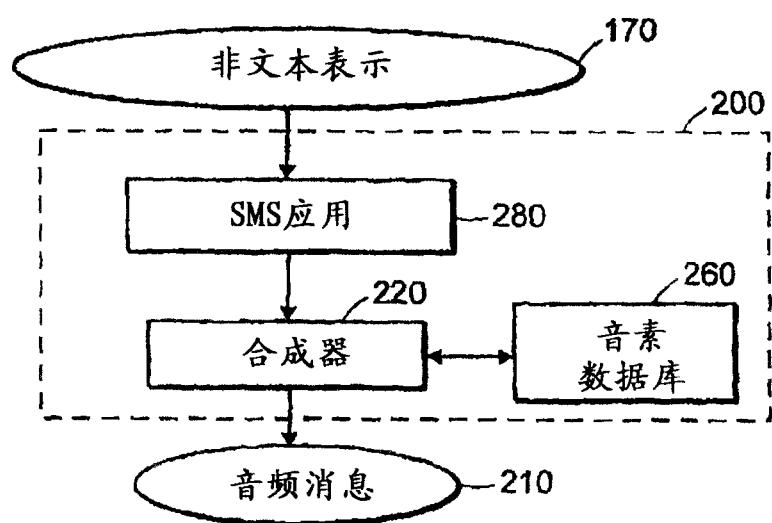


图 2

图 3

