



[12] 发明专利说明书

专利号 ZL 00803589. X

[45] 授权公告日 2005 年 7 月 27 日

[11] 授权公告号 CN 1212604C

[22] 申请日 2000.2.4 [21] 申请号 00803589. X

[30] 优先权

[32] 1999. 2. 8 [33] US [31] 09/246,605

[86] 国际申请 PCT/US2000/002900 2000.2.4

[87] 国际公布 WO2000/046795 英 2000.8.10

[85] 进入国家阶段日期 2001.8.8

[71] 专利权人 高通股份有限公司

地址 美国加利福尼亚州

[72] 发明人 张承纯

审查员 杨 叁

[74] 专利代理机构 上海专利商标事务所有限公司

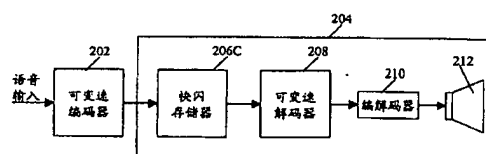
代理人 吴蓉军

权利要求书 2 页 说明书 9 页 附图 2 页

[54] 发明名称 基于可变速语音编码的语音合成器

[57] 摘要

提供一种根据可变速声编码进行语音合成的装置和方法。由可变速声码器(202)编码输入的语音信号,并且将语音信号参数存在存储器中。为了合成语音,可变速解码器(208)解码这些参数以产生语音采样。编码解码器(210)将语音采样从数字信号转换成模拟信号,并通过扬声器(212)将它广播。



1. 一种在无线通信系统中合成预选词汇表的装置，所述词汇表由可变速编码器按一组可变速率编码，其特征在于包括：

存储器，用于存储一组语音参数，所述一组语音参数表示所述编码的预选词汇表；

可变速编码器，结构成从用户接受口头输入，并根据所述用户的口头输入选择所述一组语音参数的一个语音参数子集；

可变速解码器，用于解码所述语音参数子集以产生经解码的语音采样；和
数字-模拟变换器，用于将所述语音采样转换成模拟信号以作为合成语音广播。

2. 如权利要求 1 所述的装置，其特征在于，所述可变速编码器是基于线性预测的。

3. 如权利要求 1 所述的装置，其特征在于，所述可变速解码器是基于线性预测的。

4. 如权利要求 1 所述的装置，其特征在于，所述可变速编码器以一组可变速率编码所述一组语音参数，其中所述可变速率组包括全速、半速、四分之一速率和八分之一速率。

5. 如权利要求 4 所述的装置，其特征在于，所述可变速编码器以一组可变速率编码所述一组语音参数，其中所述可变速率组包括 13.2kbps, 6.2kbps、2.7kbps 和 1.0kbps。

6. 如权利要求 4 所述的装置，其特征在于，所述可变速编码器以响应于一个或多个测定模式准则固定的速率编码所述一组语音参数。

7. 如权利要求 4 所述的装置，其特征在于，所述可变速编码器以所述半速固定的速率编码所述一组语音参数。

8. 如权利要求 4 所述的装置，其特征在于，所述可变速编码器根据对语音质量和所述存储器大小的要求，选择编码速率。

9. 如权利要求 1 所述的装置，其特征在于，所述无线通信系统是 CDMA 系统。

10. 如权利要求 1 所述的装置，其特征在于，所述可变速编码器包括增强型可变速编码器。

11. 一种在无线通信系统中合成预选词汇表的方法，所述词汇表由可变速率编码器按一组可变速率编码，其特征在于，包括下列步骤：

接收用户的口头输入；

检索存储在存储器中的一组语音参数，所述一组语音参数是对应于所述用户的口头输入的一部分所述编码的预选词汇表；

运用可变速编码方案解码所述一组语音参数以产生经解码的语音采样；和将所述语音采样转换成模拟信号以作为合成语音广播。

12. 如权利要求 11 所述的方法，其特征在于，所述可变速编码方案是基于线性预测的。

13. 如权利要求 11 所述的方法，其特征在于，所述可变速解码方案是基于线性预测的。

14. 如权利要求 11 所述的方法，其特征在于，以可变速率组编码所述一组语音参数，其中所述可变速率组包括全速、半速、四分之一速率和八分之一速率。

15. 如权利要求 14 所述的方法，其特征在于，所述全速是 13.2kbps，所述半速约 6.2kbps、所述四分之一速率约 2.7kbps 和所述八分之一速率约 1.0kbps。

16. 如权利要求 14 所述的方法，其特征在于，以响应于一个或多个测定模式准则固定的速率编码所述一组语音参数。

17. 如权利要求 14 所述的方法，其特征在于，以所述半速固定的速率编码所述一组语音参数。

18. 如权利要求 14 所述的方法，其特征在于，根据对语音质量和所述存储器尺寸的要求，选择编码速率。

19. 如权利要求 11 所述的方法，其特征在于，所述无线通信系统包括 CDMA 系统。

20. 如权利要求 11 所述的方法，其特征在于，还包括将所述用户的口头输入编码以及将所述编码的用户的口头输入加到所述存储器作为所述一组语音参数的一部分的步骤。

基于可变速语音编码的语音合成器

技术领域

本发明涉及语音合成。特别是，本发明涉及对由可变速声码器编码的话音的合成。本发明还涉及语音合成与无线通信设备的合用。

背景技术

电子语音合成在许多应用中都十分有用。越来越多的计算机和其它电子设备提供有声提示的选项作为用户接口。例如，可利用语音来读取电子邮件消息、产生在语音响应系统中的口头提示或者向汽车上的司机指明方向。

一般有两种语音合成器或技术用来产生语音。第一种被称为文本语音(TTS)语音合成器，而且是根据语法的。基于TTS的系统将普通文本转换成可理解且自然发声的语音。它对于需要将任意输入的文本自动转换成可理解且自然发声的语音输出的应用十分有用。它对于包含大量词汇和/或动态改变数据的情况特别有用。TTS系统在诸如提供自动语音警报和提示、校读(proofreading)、接入数据库的电话入口和将电子邮件转换成语音邮件或音频输出的应用十分有用。由于TTS是灵活和有利的，所以它可用于多种应用。然而，实施TTS系统可能需要庞大的存储和处理能力资源。如果合成器没有近似地模拟人语音语调，那么它还可能包括机器音调(machine tone)。因此，对具有有限存储和处理资源的应用(诸如，小型便携式无线装置、远程通信装置或计算机等)，TTS不是实用的选择。

第二种语音合成器是基于语音编码器(声码器)的。声码器通过提取与人语音生成模型相关的参数来压缩发出的语音或者音频信号。声码器已发展到压缩已被数字变换到速率为64千比特/秒(kbps)的输入语音降至13kbps、8kbps或更低速率。基于声码器的语音合成器产生对于或用于要合成的语音的某些参数。将参数存储在某些类型的存储器中，最好是快闪存储器，而且一旦语音合成就对其进行解码。由于要合成的所有字的参数都需要存储在存储器中，所以基于声码器的合成器更加适于不需要大量词汇的应用。它们特别适于具有有限存储和处理资源的系统。

对于基于声码器的语音合成器，在保持可接受的语音质量时需要有最优化存储用途。对于一些应用，理想的是，对于给定的存储器规模，使词汇量最大。此外，也是理想的是，将在给定通信系统设计内已获得的信号处理资源用来实现语音合成。本发明通过下列方法提供处理这些和其它特征的语音合成器。

发明内容

本发明是一种根据可变速声编码(vocoding)的语音合成的装置和方法。由可变速声码器对要合成的语音进行编码。可变速声码器根据在语音帧内发生的语音活动，以一组预定速率之一，编码语音帧。在一个实施例中，可变速声码器是码激励线性预测(code excited linear prediction)(CELP)声码器，它具有4种比特率。于是，以4种速率之一，运用对于所选速率的CELP编码方案，将输入语音信号编码成语音参数。一般，向解码器提供语音参数，其中解码器根据所用的可变速编码方案，执行可变速解码方案。解码器提供语音采样，向编码器-解码器或编解码器提供用于数字-模拟转换。于是，通过扬声器或其它已知的音频输出装置广播由编解码器产生的所得模拟信号作为合成语音。

本发明的语音合成器特别适合在无线通信系统中使用，其中在该系统中已执行可变速声编码。在这些系统中，将现有的声编码资源用于语音合成。另一方面，可结合少量的存储器使用DSP元件(现有的或容易结合的)，以提供语音合成器功能。此外，基于可变速声编码的语音合成器能够提供良好的语音质量，而无需大量内存。由可变速声码器提供的压缩程度使得它适于具有有限的内存的应用。

附图说明

结合附图，从下面的详细描述，本发明的特征、目的和优点将显而易见，在附图中相同标号做相应表示，其中

图1是可变速声码器的方框图；和

图2是本发明的语音合成器的方框图。

具体实施方式

本发明提供当结合无线通信设备一起使用时特别有用的合成语音的装置和方法。本发明利用在无线通信设备中的现有信号处理资源或者最少量的附加

硬件以提供高语音质量并要求小存储量的方式来合成语音。

本发明在结合多种已知的通信装置或系统使用时是十分有用的，并在下面参照 CDMA 无线通信系统对它进行描述。此外，应理解，它特别适于特定应用，诸如用来安装和操作车辆内无线装置的免提(hands-free)汽车配套设备。然而，熟悉本领域的人员容易理解这并不是对本发明进行限制，而且它还可结合其它类型的通信设备(包括，那些通过有线的、电缆或光缆类系统通信以及那些运用其它信号调制技术的装置)一起使用。

示例无线通信系统利用码分多址(CDMA)调制技术。虽然已知其它技术(诸如，时分多址(TDMA)、频分多址(FDMA)和诸如幅度压扩单边带的调幅(AM)方案)，但是 CDMA 具有优于其它这些技术的显著优点。在美国专利号 4,901,307(题为“运用卫星或地面中继站的扩展频谱多址通信系统(Spread Spectrum Multiple Access Communication System Using Satellite Or Terrestrial Repeaters)”，已转让给本发明的受让人并作为参考资料在此引入)中揭示了在多址通信系统中 CDMA 技术的运用。

出于多个理由，在无线通信装置或设备中可以实施语音合成器。例如，语音合成可以是在无线电话或用于支持在汽车中操作的“免提”汽车配套设备中的一部分语音识别系统。当装置用户或操作者不能用肉眼观察输出屏幕或装置上的指示符时，语音合成器可用音频形式提供信息。例如，可提供信息以允许当车辆司机或机器操作者不近距离不能安全地察看通信装置时操作或输出。语音合成器还允许通过提供用于要执行的操作的语音提示，来装置免提操作。例如，语音合成器可能要求呼叫人名，允许装置自动拨打电话号码或者要求执行诸如拨号、存储、打开邮件、终止尝试打电话或关闭等的命令。

在一个实施例中，本发明的语音合成器使用已经出现在多种无线装置(诸如无线电话和由通信服务用户用来产生发音语音的其它产品)中的声码器电路。特别是，基于可变速声码器的语音合成器。可变速声码器运用语音活动来改变它的瞬间数据速率。在现行说话期间，声码器编码器用大量的位来编码语音采样。在静默期间，声码器编码器几乎不用或且较少量的位来编码背景噪声。在美国专利 5,414,796 中(题为“可变速声码器(Variable Rate Vocoder)”，已转让给本发明的受让人，并作为参考资料在此引入)中描述了一种可变速声码器的示例实施例。

通常在 CDMA 类通信系统中使用可变速声码器来通过减小每个通信信号所

用的比特数来增加系统容量。例如，可在上述专利 4,901,307 的 CDMA 通信系统中执行可变速声码器。在 CDMA 通信系统中，不同用户运用相同的带宽但用不同的码信道来进行通信。在 CDMA 通信系统中的可变速声码器利用用户仅在任一给定信道上的 40% 时间内说话的这一事实。当用户静默时通过发送较少的位，可变速声码器允许更多用户共享相同带宽。

图 1 示出典型的可变速声码器的示意方框图，而且概括地用 100 表示。如图 1 所示的声码器运用了 4 个不同的数据速率，虽然应理解可用不同数量的数据速率(如现有技术中已知的那样)。在这组 4 个速率中，如果峰值速率是 13.2kbps，那么全速对应于 13.2kbps、1/2 速率对应于约 6.2kbps、1/4 速率对应于约 2.7kbps 和 1/8 速率对应于约 1.0kbps。注意，如现有技术中已知的那样，由于运用附加位，所以对于除全速之外的其它速率的实际比特率都是近似的。

仍然参考图 1，可见可变速声码器 100 包括编码器 102 和解码器 104。编码器 102 接收用于语音数据帧的语音采样作为输入，例如，在 64kbps 数据速率时以 μ -定律(μ -Law)或 a 定律(a-law)格式的 8 位 PCM 采样。编码器 102 根据语音活动，以 4 个数据速率之一将这些语音采样编码成语音参数。还向速率确定元件 106 提供输入语音采样。

速率确定元件 106 可执行多个速率判定算法中的任一个算法。在一个实施例中，用与背景噪声能量级相关的阈值来确定语音活动，以及编码输入采样的速率。如果语音采样的当前帧的能量远高于背景噪声能量，那么速率确定元件 106 将确定以全速编码帧。如果当前帧的能量接近背景噪声能量，那么如已知的那样，速率确定元件 106 将确定以八分之一速率编码帧，等等。

在待批美国专利申请号 08/286,842(发明名称为“执行减速可变速声编码的方法和装置(Method And Apparatus For Performing Reduced Rate Variable Rate Vocoding)”，转让给本发明的受让人并作为参考资料在此引入)中揭示了另一种速率确定技术。该技术提供被称为模式测量(mode measure)的一组速率确定准则。第一模式测量是来自前一编码帧的目标匹配信噪比(TMSNR)，它通过将合成的语音信号与输入语音信号相比较提供关于编码模型执行得有多好的信息。第二模式测量是归一化自相关函数(NACF)，它测量在语音帧内的周期性。第三模式测量是零交叉(ZC)参数，它测量在输入语音帧内的高频率内容。第四测量，预定增益差分(PGD)，确定编码器是否保持它的预定效率。第

五测量是能量差分(ED)，它将在当前帧中的能量与平均帧能量作比较。

运用上述模式测量，速率确定逻辑选择用于每个输入语音数据帧的编码速率。对于各模式的值例如从4个或更多模式中选择要操作的一个模式。即，根据预定模式或分级，对于与门限或其它准则相关的每个模式测量检测到的值确定选择哪个编码速率。例如，如果对于NACF的值小于预选门限和ZC大于第二预选门限，可选择一速率。然而，如果没有满足这些条件，但是ED低于第三门限，那么可选择四分之一速率。如果对于TSNR的值更大、PGD更少而且NACF分别大于第四、第五和第六门限，那么可选择半速。熟悉本技术领域的人员可采用各种这样的组合和门限来选择编码速率。

应理解，速率确定元件106还可采用其它速率确定技术。

仍然参照图1，一个由速率确定元件106确定的数据速率的指示信号提供给开关108。开关108从全速编码元件110、半速编码元件112、四分之一速率编码元件114和八分之一速率编码元件116之间选择一个元件，用于编码输入语音采样帧，如数据速率信号所指定的那样。所选编码元件对语音采样进行编码以产生编码数据分组的信号。速率确定元件106还向开关118提供指示数据速率的信号，上述开关选择与开关108相同的编码元件，从而可向可变速声码器的输出提供所选编码元件产生的编码数据分组的信号。

将编码元件110、112、114和116中的每一个元件配置成运用预定编码方案编码语音。在较佳实施例中使用基于线性预测的编码方案(诸如，代码激励线性预测(Code Excited Linear Predictive)(CELP)编码器。在Thomas E. Tremain等人所著的论文“4.8Kbps代码激励线性预测编码器(A 4.8 Kbps Code Excited Linear Predictive Coder)”(移动卫星会议进程(Proceedings of the Mobile Satellite Conference)，1998)中描述了CELP编码器。通过去除在语音中固有的自然冗余，基于线性预测的编码器压缩语音。语音一般呈现由于嘴唇和舌头的机械动作所致的短期冗余和由于声带振动所致的长期冗余。线性预测方案模拟这些操作为滤波器、去除冗余并模拟所得剩余信号作为白高斯噪声。因此，通过发送滤波器系数和量化噪声而不是全带宽语音信号，线性预测编码器获得减小了的比特率。

采用可变速的线性预测编码方案进一步减小比特率，而不影响语音质量。在图1中，全速编码元件110运用更多位来对输入语音信号的参数进行编码以便更好地保留输入的特征。对于没有检测到任何语音的期间内，由于几乎没有

获得细节或有用的信息，所以八分之一速率编码元件 116 运用较少位对参数进行编码。由半速编码元件 112 和四分之一编码元件 114 对在现行语音期间和没有检测到语音期间之间的过渡进行编码。

现在参照可变速声码器的解码元件，解码器 104 接收编码语音参数的信号以及指示用来编码语音的速率的信号。速率提取元件 128 接收该输入信号并确定语音的数据速率。还向开关 130 提供数据速率的信号，所述开关从一组解码元件中选择解码元件来正规地解码输入参数。在图 1 中，提供 4 个解码元件（全速解码元件 120、半速解码元件 122、四分之一速率解码元件 124 和八分之一速率解码元件 126）在这 4 个可能的速率下解码语音参数。所选解码元件根据数据速率解码输入参数以产生解码采样信号，所述采样一般是 64kpbs 脉冲编码调制 (PCM) 采样。还向开关 132 提供由速率提取元件 128 确定的数据速率的信号。开关 132 选择与开关 130 相同的解码元件，从而向声码器的输出提供解码采样。

现在，参照图 2，示出了根据本发明的原理操作的语音合成系统的方框图，所述系统结合了可变速声码器。语音合成系统包括可变速编码器 202 和语音合成器 204。可变速编码器 202 的例子是图 1 的编码器 102。可变速编码器 202 接收语音信号作为输入并以一组预定速率之一编码语音。在较佳实施例中，可变速编码器 202 是 CELP 编码器，它根据输入语音段中的语音活动，以一个速率产生语音参数。

本发明运用如上述美国专利 5,414,796 中所述的可变速声码器，该可变速声码器在市场上有售，例如高通公司生产的 13kpbs 声码器产品。在较佳实施例中，可变速解码器是诸如根据 IS127 标准所述的增强型可变速解码器。

在本发明的一个实施例中，编码速率判定是根据上述“模式测量”。熟悉本技术领域的人员会理解，用来作出速率选择的不同的准则组合来产生所谓“减速的速率模式”或“模式”，并简称为模式 0、模式 1、模式 2，等等。本发明可利用这种模式来作语音合成。

由可变速编码器 202 接收到的语音可以是来自一预选词汇表的字或短语，其中设计诸如无线电话、车辆配套设备或其它通信装置的通信装置来合成上述预选词汇表。该词汇表可包括向装置用户提供的提示和警报。例如，通过提取和合成 5 个词汇字：“呼叫”、“重拨”、“程序”、“或”和“退出”，可将语音合成器设计成在从用户的响应请求中提供提示“呼叫、重拨、程序或

退出”。另一方面，可将语音合成器设计成响应于各种装置输入(包括，音频)，向装置用户提供先前所存的信息(诸如在电话簿、查询表或数据库中)。把由可变速编码器 202 接收到的语音编码，并向语音合成器 204 的存储元件或电路 206 提供已编码的参数以存储。

在一段时间内，存储器 206 是用来保持或存储参数以操作所需装置。然而，通常理想的是，以使它们可更新或可替代(诸如，当需要改变词汇表以改变条件或升级到装置特征时)的方式存储参数。因此，以非易失但可重写存储器的形式构成存储器 206，其中如现有技术已知的那样，可运用快闪类存储元件来实现上述非易失但可重写存储器。

如人们所认识到的那样，可在根据本发明制造通信装置期间执行负载参数的操作。由于可预定要合成的提示和警报，所以可在使用之前制造和存储在快闪存储器 206 期间，对这些进行编码。在装置服务期间，可改变或替换参数，或者通过用于无线装置的新发展的空中编程技术来实现。

另一方面，在操作通信装置期间，可变速编码器 202 可接收语音信号输入。例如，响应于来自语音合成器的提示，用户可提供口说的响应。可变速速率编码器 202 将编码用户的语音，而且可向快闪存储器 206 提供经编码的参数用于存储，和/或向语音识别器(未图示)提供用于语音识别。通过这种方法，在制造之后(post manufacture)(诸如在装置进入实用服务的即时或超时)就输入参数，诸如通过为每个装置的(声码器)用户建立与该用户的要求相关的词汇库来实现。

快闪存储器 206 应具有足以存储预选词汇表的参数以及用户预期的参数的规模。于是，根据特定应用的要求可改变快闪存储器 206 的大小。制造后的存储器可具有减小存储要求的优点，因为与制造商为了覆盖整个较大装置市场而必需安装的词汇表相比，每个装置用户不要求如此大的词汇表。语音合成器可通过检测目标或所需短语或语音的端点、去除静默或冗余并对其编码来记录名字或其它字，诸如“Fred Smith”。因此，可在线记录语音并随后用于合成语音输出。

应注意，可根据可用的存储器和所需语音质量来配置可变速编码器 202。在具有 4 个速率的系统中，其中全速是 13kbps，根据 40%语音活动，平均速率一般是 5.88kbps。对可变速的运用提供了高语音质量。然而，如果存储器大小是被限定的，那么可变速编码器 202 被配置成假设以固定的半速(约 800 字节/

秒)操作。否则,可从一预定速率组的子集中选择速率,而不是从整个速率组中选择。例如,可用上述的已减速的速率模式来选择各速率。在本发明的一个实施例中,将速率分成一组4个模式,标为模式0、1、2和3。运用根据模式的固定速率,分别可采用大约1800字节/秒、1540字节/秒、1400字节/秒和1100字节/秒的速率。对这些固定的已减小速率的运用允许以给定的预定数据速率传递质量很高的语音,该质量达到了陆上通讯线的质量。这4种模式在合成语音质量和存储器必要的规格之间提供最佳折衷。

此外,根据应用的瞬间要求,可变速编码器202可在不同的操作模式之间(可变速、所有半速、可变速的子集,等等)转换。由于在语音质量和存储器大小之间存在折衷,所以将采用的配置应根据要实施的应用。

当需要语音合成时,向可变速解码器208提供存储在快闪存储器206中的语音参数。将可变速解码器208配置成对通过与对应的可变速编码器202所产生的参数解码。可变速解码器208的例子是图1的解码器104。

一般,将可变速解码器208作为在通信装置中使用的一部分数字信号处理器(DSP)来实施。将这些DSP用作或形成用于信号编码/解码、组合、CDMA编码、功率调节,等等的处理元件。由于一般在可采用本发明的无线装置和多种其它装置中使用这些元件,所以利用它们的存在能够十分经济地实施本发明。

为了要本发明实施解码功能,在DSP中只要具有少量的存储器或者耦合到DSP。在DSP中或运用DSP的独立解码器要求很少量的内存(程序和数据)来获得语音合成器能力。运用诸如可从模拟装置(Analog Devices)和高通公司(Qualcomm Inc.)购得的著名的DSP电路和装置,可实施语音合成器。

向编码解码器210提供一般以脉冲编码调制(PCM)采样形式的经解码的参数。编码解码器210将PCM采样从数字格式转换成模拟信号。向扬声器或其它已知音频输出装置212提供模拟信号,其中输出装置212将合成语音投向或广播到能听到它的周围的装置环境中。

因此,本发明提供基于可变速率声编码的语音合成器。语音合成器特别适用于已包括可变速声码器的无线通信装置。换句话说,通过在程序或操作命令中的适当变换或运用控制硬件,语音合成器可采用现有的可变速声码器。此外,通过运用可变速声编码,获得的压缩允许将预定词汇表存储在与它所连接的无线装置或其它设备相关的大小限定的存储器中。此外,在配置可变速声码器以向语音合成器提供所需的语音质量和存储器大小过程中,可考虑在语音质

量和存储器大小之间进行折衷。

本发明可用于多种通信装置和接口设备。参照无线通信设备(诸如(但不限于)通常被称为用户终端、订户单元、移动站或简称为“用户”、“移动”或“订户”的蜂窝和卫星电话),讨论上述示例实施例。此外,还可考虑其它装置,诸如消息接收机和数据传递装置(例如,便携式计算机、个人数据助理、调制解调器、机器控制器),或者可考虑用于公共电话交换网或专用通信信道的接口。

运用以专用元件或用途特定集成电路(ASIC)形式的分立电路实施本发明以形成可安装在所需装置中的语音合成器。另一方面,通过运用少量的附加存储器来与现有数字信号处理元件一起工作,可将它加入其它 ASIC 和装置中。

向熟悉本技术领域的人员提供较佳实施例的描述以制作或运用本发明。对这些实施例的各种变换对于熟悉这些现有技术的人员而言是显而易见的,而且可将这里定义的一般原理用于其它实施例,而不用进行创造性劳动。于是,本发明并不限于这里所示的实施例,而是根据这里揭示的原理和新颖性特征一致的最宽范围。

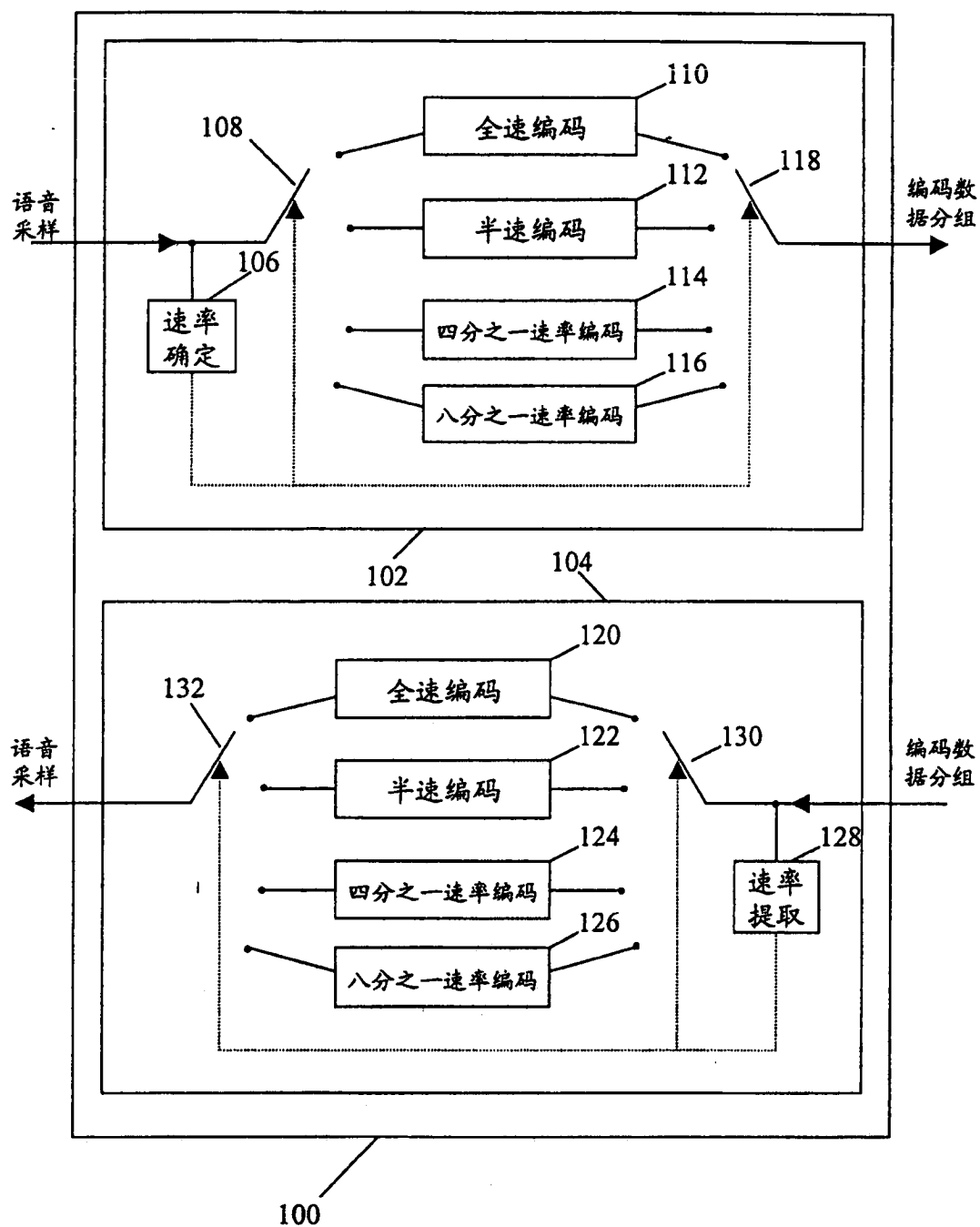


图 1

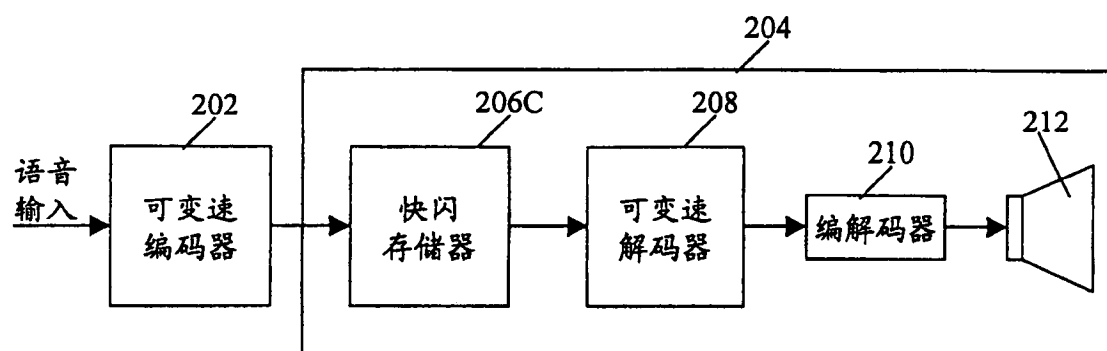


图 2