[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 15/02 (2006.01)
G10L 15/08 (2006.01)
G10L 19/14 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200910096638.9

[43] 公开日 2009 年 8 月 19 日

[11] 公开号 CN 101510424A

[22] 申请日 2009.3.12

[21] 申请号 200910096638.9

[71] 申请人 孟智平

地址 646006 四川省泸州市茜草坝长起厂一生活区 26 号楼 1 号

[72] 发明人 孟智平 郭海锋

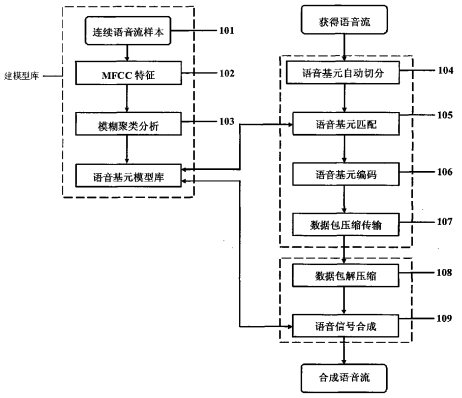
权利要求书 6 页 说明书 18 页 附图 2 页

[54] 发明名称

基于语音基元的语音编码与合成方法及系统

[57] 摘要

本发明公开了一种基于语音基元的语音编码与合成方法及系统，可用于低带宽高音质的语音传输。 本发明在数字语音传输的基础上，以构建的语音基元为编码对象，通过对日常语音的分析，采用聚类算法构建语音基元模型库；然后，利用语音基元自动切分算法，对获得的连续语音流进行语音基元的自动切分，并提取语音基元 MFCC 特征，通过与语音基元模型库中的语音基元进行匹配识别，获得语音基元所对应的编号，以编号代替语音基元进行编码。 在合成语音过程中，通过编号从语音基元模型库中取出该编号所对应的语音基元，并通过数学变换对语音基元的频谱包络进行插值拟合等处理，形成平滑过度的语音。



1、一种生成语音基元模型库的方法，其特征在于，包括以下步骤：

获取语音流样本数据，并对所述语音流数据进行切分，以获取由不同音素或不同波形为单位所构成的语料库，其中，所述构成语料库的基本单元称为语音基元；

提取所述语音基元的特征，构成特征向量；

对所述语音基元特征向量样本进行模糊聚类，将所有数据样本分为 N 类，得到对应的聚类中心和隶属度函数；

分析各类语音基元的特征，进而确定拟建语音基元模型库所需的基本语音基元；

对各类语音基元的语音特性进行分析处理，以获得每一类音素的频谱包络特征，并将所述频谱包络特征存储于语音基元模型库中，构成语音基元模型库。

2、如权利要求 1 所述生成语音基元模型库的方法，其特征在于，所述对语音流数据进行切分为：以音素或者帧为单位，对连续语音流进行切分；

所述以音素为单位进行切分是指采用音素自动切分算法，将连续的语音流自动地切分成由不同的音素所构成的音素集合；

所述以帧为单位进行切分是指以某一时间帧为单位，将连续的语音流切分成由不同波形所构成的波形集合；

所述语音基元模型库是指构成可理解的语音流所需的最小的音素样本库或最小的语音波形样本库。

3、如权利要求 1 所述生成语音基元模型库的方法，其特征在于，所述音素自动切分算法包括以下步骤：

将获得的连续语音流自动切分成以音节为单位的音节序列；

对每一个音节进一步分析音素的构成；

如果该音节为单个音素构成，则将所述音节切分为对应的音素；

如果该音节为多个音素构成，则对所述音节进一步细致切分，最终切分成几个独立的单个音素；

采用 AMDF、AC、CC、SHS 基频提取算法中的任何一种，提取每个音素基频 F0；

采用 Mel 频率倒谱系数 MFCC 作为语音信号特征参数，提取每个音素的频谱包络；

采用隐马尔可夫模型对语音特征参数样本集进行训练、识别，最终确定模型中的相关参数，训练测试后的隐马尔可夫模型，用于对连续语音流中所包含的音素进行自动切分。

4、如权利要求 1 所述生成语音基元模型库的方法，其特征在于，所述切分语音流获取不同波形的方法还包括：

以相同时间帧为切分点，对连续语音流的波形进行切分，获取等时间帧情况下不同的波形集合；

以不同的时间帧为切分点，对连续语音流的波形进行切分，获取不同时间帧情况下的不同波形集合；

采用 AMDF、AC、CC、SHS 基频提取算法中的任何一种，提取切分后每一段波形的语音基频 F0；

采用 Mel 频率倒谱系数 MFCC 作为语音信号特征参数，提取每段波形的频谱包络。

5、如权利要求 1 所述生成语音基元模型库的方法，其特征在于，生成语音基元模型库的过程还包括以下步骤：

采用模糊聚类的方法对音素集合或波形集合进行聚类分析，将音素或波形划分为 N 类；

对每一类音素或波形的语音特征进行分析，以聚类中心点或其他点的相应组合为对象，替代该类音素集或波形集，即同一类音素或波形集中抽取出一个音素或一个波形以代表该类，最终抽取出 N 个音素或 N 个波形；

确定取出的 N 个音素或 N 个波形的基频 F0 和频谱包络；

将上述 N 个音素或 N 个波形赋予其相应的编号，以编号为顺序

将 N 个音素或 N 个波形的相关信息进行存储，以构成语音基元模型库。

6、一种基于语音基元模型库的语音编码方法，其特征在于，包括以下步骤:

对连续的语音流进行自动切分，获取语音基元及其基频 F0，并提取语音基元的频谱包络; 所述语音基元是指音素或等时间帧的语音波形或不同时间帧的语音波形;

将提取的语音基元与语音基元模型库中的语音基元进行匹配，如果匹配成功，则返回该语音基于在语音基元模型库中所对应的语音基元编号;

将返回的语音基元编号、语音基元的基频 F0 和相关信息按照预设格式进行编码;

采用压缩算法对已编码的数据进一步压缩，以分组或电路交换的形式通过IP网络或电话通信系统将该语音压缩数据包传输到目的地。

7、如权利要求 6 所述基于语音基元模型库的语音编码方法，其特征在于，所述语音基元匹配包括以下步骤:

采集连续的语音流信息;

对获得的连续语音流进行分析，并采用语音基元自动切分算法将连续语音流分割成语音基元序列，即音素序列或波形序列;

将分割的语音基元直接或通过变换或误差处理操作后，与语音基元模型库中的语音基元进行模式匹配;

如果匹配成功则返回语音基元所对应的编号及相关信息;

如果匹配不成功则采用相应容错处理方法。

8、如权利要求 7 所述基于语音基元模型库的语音编码方法，其特征在于，所述语音基元变换是指通过曲线拟合、噪声误差处理的方式对语音基元的异常情形进行分析处理，以便与语音基元模型库中的语音基元进行匹配;

所述语音基元的曲线拟合是指通过最小二乘法或 B 样条或三次样条插值法，对信息不完整的语音基元波形曲线进行拟合，以复原该

语音基元的原本波形;

所述语音基元误差处理是指通过采用语音增强算法,对语音基元进行处理,以消除噪声、增强语音清晰度,提高语音自然度;

所述容错处理方法是指通过容错算法,对匹配不成功的语音基元进行处理,使语音编码过程具有较强的鲁棒性和健壮性。

9、如权利要求 6 所述基于语音基元模型库的语音编码方法,其特征在于,所述编码过程包括以下步骤:

获得语音基元编号、语音基元的基频 F0 和相关信息;

对语音基元编号、语音基元的基频 F0 和相关信息进行分析,以确定合适的编码方法;

采用霍夫曼 Huffman、LZW、曼彻斯特、单极性码等编码方法之一对上述信息进行编码;

将编码后的字符串称为语音基元编码串。

10、如权利要求 6 所述基于语音基元模型库的语音编码方法,其特征在于,所述对已编码的数据进一步压缩包括以下步骤:

接收语音基元编码串;

采用压缩分析算法对语音基元编码串进行分析,如果该语音基元编码串有进一步压缩的空间,则采用压缩算法对其进行压缩,然后对压缩后的语音基元数据包进行打包传输;

如果该语音基元编码串没有可压缩的空间,则不进行压缩,直接对压缩后的语音基元数据包进行打包传输;

所述打包传输是指采用 IP 网络协议或电路交换中的相关协议,对压缩的数据包以分组或电路交换的形式通过 IP 网络或电话系统进行传输,送至目的地。

11、一种基于语音基元模型库的语音解码方法,其特征在于,包括以下步骤:

接收方接收语音基元压缩数据包;

按照与压缩算法相对应的解压缩算法对该数据包进行解压处理;

从解压的数据包中获得语音基元编码串;

按照语音基元编码算法，对语音基元编码串进行逆向的解码操作，以获得原始语音基元数据串；

从语音基元数据串中获得语音基元编号、语音基元基频 F0 和相关信息；

根据语音基元编号，查找语音基元模型库，取出该编号对应的语音基元的语音特征，并进行语音合成；

通过语音合成方法，将发过来的语音基元还原为可理解的、清晰的语音信息。

12、如权利要求 11 所述基于语音基元模型库的语音解码方法，其特征在于，所述语音合成方法还包括以下步骤：

分析接收到的语音基元编号，如果该数值正常则根据该数值查询语音基元模型库，否则进行容错处理或忽略该语音基元；

以语音基元编号为检索条件，从语音基元模型库中取出该编号所对应的语音基元，即音素或波形；

根据取出的语音基元的语音特征、接收到的该语音基元的基频 F0 和相关信息对语音进行合成。

13、一种基于语音基元的语音编码与合成方法，其特征在于，包括以下步骤：

获取大量语音流样本数据，通过对所述样本数据进行处理，构成语音基元模型库；

对获取到的连续语音流进行切分，获取语音基元及其基频 F0，然后将该语音基元与语音基元模型库中的语音基元进行匹配，获得相对应的语音基元编号，采用编码方法按照一定的格式对语音基元编号、语音基元基频 F0 和语音特征附属信息进行编码，将编码后的数据包进一步压缩，通过 IP 网络或电话网络将该语音压缩数据包传输到目的地；

接收方收到语音压缩数据包后，采用相应的解压缩算法解压数据包，根据语音基元编号查找语音基元模型库，取出该语音基元所对应的语音特征，并根据基频 F0 和附属信息还原为语音。

14、一种基于语音基元的语音编码与合成系统，其特征在于，包括以下模块：预处理模块、语音编码模块和语音解码模块；

所述预处理模块，负责采集分析连续语音流，将语音流切分成语音基元序列，并通过聚类算法对大量的语音基元进行聚类分析，构建语音基元模型库，以供语音编码模块和语音解码模块调用；

所述语音编码模块，以预处理模块构建的语音基元模型库为基础，对接收到的语音流进行切分以获取语音基元及其基频 F0，根据语音基元匹配算法从语音基元模型库中获得该语音基元所对应的编号，然后将语音基元编号、基频 F0 和附属信息按照相应编码算法进行编码，并采用压缩算法对其进一步压缩，然后将其打包发送；

所述语音解码模块，负责接收所述语音编码模块传送过来的语音数据包，对其进行解压缩，获取语音基元编号，以该编号为检索条件，查询语音基元模型库，提取该编号对应的语音基元信息，最终通过语音合成算法还原语音。

15、如权利要求 14 所述基于语音基元的语音编码与合成系统，其特征在于，包括语音发送端和语音接收端；

所述语音发送端，包括语音基元模型库、语音编码模块，发送端语音编码模块对接收到的语音流进行切分，并根据语音基元匹配算法从语音基元模型库中获得该语音基元所对应的编号，将语音基元编号、基频 F0 和附属信息按照相应编码算法进行编码，并采用压缩算法对其进一步压缩，然后将其打包发送；

所述语音接收端，包括语音基元模型库、语音解码模块，接收端语音解码模块负责接收所述语音编码模块传送过来的语音数据包，对其进行解压缩，获取语音基元编号，以该编号为检索条件，查询语音基元模型库，提取该编号对应的语音基元信息，最终通过语音合成算法还原语音。

基于语音基元的语音编码与合成方法及系统

技术领域

本发明涉及语音编码、语音传输、语音电话等领域，尤其涉及一种基于语音基元的语音编码与合成方法及系统。

背景技术

随着现代网络技术的发展，通过因特网传送语音信号的应用越来越多，尤其是在线聊天工具的迅速普及，已使网络电话成为一种受人喜爱的沟通工具。目前大部分的网络电话都采用 G.711、G.723、G.726、G.729 等通用的编码技术，网络传送中的语音多采用压缩比较高的中、低速率语音编码。低速率的语音压缩编码虽然给信道的传输带来了方便，也节省了存储空间，但是由于大部分语音编码都是有损压缩，语音质量势必会受到损失。这些技术的共同点都是利用人耳感知的先验知识对语音进行有损压缩。专利号 00126112.6 公开了一种采用单帧、变帧长、帧内比特自适应的低速语音压缩编码方法，可使编码压缩的能力进一步提高，进而提高了数据传输效率。这些编码方式都是针对人耳听觉特点，设计人耳能容忍的有损压缩方案来达到减小编码速率的目的。实际上，如果只是针对人的语音进行编码，不涉及音乐等其他问题，压缩率还可以进一步改进。

语音学研究表明，音素是从音质角度划分的最小的语音单位，从发音特征上看，人们发出的语音都是由不同的音素构成的，一个音素或者多个音素的组合，形成了不同的音节，如每一个汉字的发音即是一个音节。经过统计分析发现，人发音的音素个数其实是有限的，而且有一些音素是可以由其他一些音素组合而成，由此可知，每一种语言便可统计出构成该语言发音特征的基本音素。根据国际音标协会组织 2005 年最近公布结果，世界上已知的发音中，肺部气流音有 59 个，非肺部气流音有 14 个，其他辅音 12 个，单元音 28 个，其他的发音，不外乎这些音的组合。

网络语音传输或电话语音通信时，通常收听方所关心的仅是说话方发出的语音信息，如果传输或通信的内容只有人说话的语音信息，没有其他声音或者滤掉其他声音，则语音传输在已有方法基础上还可以进一步压缩。

此外，通过对连续语音流的波形及频谱包络分析发现，无论是一次连续的语音流所生成的同一波形中，还是不同语音流所生成的不同波形中，很多波形是相同或非常相似的，如果在编码之前能够对这些波形进行处理，对具有共同特征的波形段进行分析，建立波形模型库，为不同的波形赋予编号，便可以改进已有的以帧为单位进行采样的编码方式，而是仅对波形对应的编号进行编码，从而极大地提高编码的效率。

本发明以语音基元为编码单位，设计了一种更优的语音编码方案。该方案根据获得的连续语音流数据，提取相应的语音基元，构建

语音基元模型库，通过对获得的连续语音流进行切分，将切分的语音基元与模型库中的语音基元进行匹配，获得当前语音的语音基元编号。于是原先需要上百维的频谱信号或者十几维的倒谱信号来描述的语音信号，现在仅用一个整数编号就可以描述。在解码的时候，根据此整数，从库中获得真正的谱信号重建语音，从而大大提高语音的压缩率。

发明内容

为了对语音流数据进行压缩编码，使语音数据在低带宽或网络性能较差情况下进行有效传输，本发明首先公开了一种生成语音基元模型库的方法，包括以下步骤：

获取语音流样本数据，并将所述语音流数据进行切分，以获取由不同音素或不同波形为单位所构成的语料库，其中，所述构成语料库的基本单元称为语音基元；

提取所述语音基元的特征，构成特征向量；

对所述语音基元特征向量样本进行模糊聚类，将所有数据样本分为 N 类，得到对应的聚类中心和隶属度函数；

分析各类语音基元的特征，进而确定拟建语音基元模型库所需的最少语音基元；

对各类语音基元的语音特性进行分析处理，以获得每一类语音基元的频谱包络特征，并将所述频谱包络特征存储于语音基元模型库中，构成语音基元模型库；

所述对语音流数据进行切分，是以音素或者帧为单位，对连续语音流进行切分；

所述以音素为单位进行切分是指采用音素自动切分算法，将连续的语音流自动地切分成由不同的音素所构成的音素集合；

所述以帧为单位进行切分是指以某一时间帧为单位，将连续的语音流切分成由不同波形所构成的语音波形集合；

所述语音基元模型库是指构成可理解的语音流所需的最小的音素样本库或最小的语音波形样本库；

所述音素自动切分算法包括以下步骤：

将获得的连续语音流自动切分成以音节为单位的音节序列；

对每一个音节进一步分析音素的构成；

如果该音节为单个音素构成，则将所述音节切分为对应的音素；

如果该音节为多个音素构成，则对所述音节进一步细致切分，最终切分成几个独立的单个音素；

采用 AMDF、AC、CC、SHS 基频提取算法中的任何一种，提取每个音素基频 F0；

采用 Mel 频率倒谱系数 MFCC）作为语音信号特征参数，提取每个音素的频谱包络；

采用隐马尔可夫模型对音素特征参数样本集进行训练、识别，最终确定模型中的相关参数，训练测试后的隐马尔可夫模型，用于对连续语音流中所包含的音素进行自动切分。

所述切分语音流获取不同波形的方法还包括：

以相同时间帧为切分点，对连续语音流的波形进行切分，获取等时间帧情况下不同的语音波形集合；

或以不同的时间帧为切分点，对连续语音流的波形进行切分，获取不同时间帧情况下的不同语音波形集合；

采用 AMDF、AC、CC、SHS 基频提取算法中的任何一种，提取切分后每一段波形的语音基频 F0；

采用 Mel 频率倒谱系数（MFCC）作为语音信号特征参数，提取每段波形的频谱包络。

所述生成语音基元模型库的过程还包括以下步骤：

采用模糊聚类的方法对音素集合或波形集合进行聚类分析，将音素或波形划分为 N 类；

对每一类音素或波形的语音特征进行分析，以聚类中心点或其他点的相应组合为对象，替代该类音素集或波形集，即从同一类音素或波形集中抽取出一个音素或一个波形以代表该类，最终抽取出 N 个音素或 N 个波形；

确定取出的 N 个音素或 N 个波形的基频 F0 和频谱包络；

将上述 N 个音素或 N 个波形赋予其相应的编号，以编号为顺序将 N 个音素或 N 个波形的相关信息进行存储，以构成语音基元模型库。

本发明还公开了一种基于语音基元模型库的语音编码方法，包括以下步骤：

对连续的语音流进行自动切分，获取语音基元及其基频 F0，并

提取语音基元的频谱包络；所述语音基元是指音素或等时间帧的语音波形或不同时间帧的语音波形；

将提取的语音基元与语音基元模型库中的语音基元进行匹配，如果匹配成功，则返回该语音基于在语音基元模型库中所对应的语音基元编号；

将返回的语音基元编号、语音基元的基频 F0 和相关信息按照预设格式进行编码；

采用压缩算法对已编码的数据进一步压缩，以分组或电路交换的形式通过IP网络或电话通信系统将该语音压缩数据包传输到目的地；

所述语音基元匹配包括以下步骤：

采集连续的语音流信息；

对获得的连续语音流进行分析，并采用语音基元自动切分算法将连续语音流分割成语音基元序列，即音素序列或波形序列；

将分割的语音基元直接或通过变换或误差处理操作后，与语音基元模型库中的语音基元进行模式匹配；

如果匹配成功则返回语音基元所对应的编号及相关信息；

如果匹配不成功则采用相应容错处理方法；

所述语音基元变换是指通过曲线拟合、噪声误差处理的方式对语音基元的异常情形进行分析处理，以便与语音基元模型库中的语音基元进行匹配；

所述语音基元的曲线拟合是指通过最小二乘法或 B 样条或三次样条插值法，对信息不完整的语音基元波形曲线进行拟合，以复原该

语音基元的原本波形;

所述语音基元误差处理是指通过采用语音增强算法,对语音基元进行处理,以消除噪声、增强语音清晰度,提高语音自然度;

所述容错处理方法是指通过容错算法,对匹配不成功的语音基元进行处理,使语音编码过程具有较强的鲁棒性和健壮性。

所述编码过程包括以下步骤:

获得语音基元编号、语音基元的基频 F0 和相关信息;

对语音基元编号、语音基元的基频 F0 和相关信息进行分析,以确定合适的编码方法;

采用 LZW、霍夫曼 Huffman、曼彻斯特、单极性码等编码方法之一对上述信息进行编码;

将编码后的字符串称为语音基元编码串。

所述对已编码的数据进一步压缩包括以下步骤:

接收语音基元编码串;

采用压缩分析算法对语音基元编码串进行分析,如果该语音基元编码串有进一步压缩的空间,则采用压缩算法对其进行压缩,然后对压缩后的语音基元数据包进行打包传输;

如果该语音基元编码串没有可压缩的空间,则不进行压缩,直接对压缩后的语音基元数据包进行打包传输;

所述打包传输是指采用 IP 网络协议或电路交换中的相关协议,对压缩的数据包以分组或电路交换的形式通过 IP 网络或电话系统进行传输,送至目的地。

本发明还提供了一种基于语音基元模型库的语音解码方法，包括以下步骤：

接收方接收语音基元压缩数据包；

按照与压缩算法相对应的解压缩算法对该数据包进行解压缩处理；

从解压缩的数据包中获得语音基元编码串；

按照语音基元编码算法，对语音基元编码串进行逆向的解码操作，以获得原始语音基元数据串；

从语音基元数据串中获得语音基元编号、语音基元基频 F0 和相关信息；

根据语音基元编号，查找语音基元模型库，取出该编号对应的语音基元的语音特征，并进行语音合成；

通过语音合成方法，将发过来的语音基元还原为可理解的、清晰的语音信息；

所述语音合成方法还包括以下步骤：

分析接收到的语音基元编号，如果该数值正常，则根据该数值查询语音基元模型库，否则进行容错处理或忽略该语音基元；

以语音基元编号为检索条件，从语音基元模型库中取出该编号所对应的语音基元，即音素或波形；

根据取出的语音基元的语音特征、接收到的该语音基元的基频 F0 和相关信息对语音进行合成。

本发明还提供了一种基于语音基元的语音编码与合成方法，包括

以下步骤:

获取大量语音流样本数据，通过对所述样本数据进行处理，构成语音基元模型库;

对获取到的连续语音流进行切分，获取语音基元及其基频 F0，然后将该语音基元与语音基元模型库中的语音基元进行匹配，获得相对应的语音基元编号，采用编码方法按照一定的格式对语音基元编号语音基元基频 F0 和语音特征附属信息进行编码，将编码后的数据包进一步压缩，通过 IP 网络或电话网络将该语音压缩数据包传输到目的地;

接收方收到语音压缩数据包后，采用相应的解压缩算法解压数据包，根据语音基元编号查找语音基元模型库，取出该语音基元所对应的语音特征，并根据基频 F0 和附属信息还原为语音。

本发明还公开了一种基于语音基元的语音编码与合成系统，包括以下模块：预处理模块、语音编码模块和语音解码模块;

所述预处理模块，负责采集分析连续语音流，将语音流切分成语音基元序列，并通过聚类算法对大量的语音基元进行聚类分析，构建语音基元模型库，以供语音编码模块和语音解码模块调用;

所述语音编码模块，以预处理模块构建的语音基元模型库为基础，对接收到的语音流进行切分以获取语音基元及其基频 F0，根据语音基元匹配算法从语音基元模型库中获得该语音基元所对应的编号，然后将语音基元编号、基频 F0 和附属信息按照相应编码算法进行编码，并采用压缩算法对其进一步压缩，然后将其打包发送;

所述语音解码模块，负责接收所述语音编码模块传送过来的语音数据包，对其进行解压缩，获取语音基元编号，以该编号为检索条件，查询语音基元模型库，提取该编号对应的语音基元信息，最终通过语音合成算法还原语音。

所述基于语音基元的语音编码与合成系统，包括语音发送端和语音接收端；

所述语音发送端，包括语音基元模型库、语音编码模块，发送端语音编码模块对接收到的语音流进行切分，并根据语音基元匹配算法从语音基元模型库中获得该语音基元所对应的编号，将语音基元编号、基频 F0 和附属信息按照相应编码算法进行编码，并采用压缩算法对其进一步压缩，然后将其打包发送；

所述语音接收端，包括语音基元模型库、语音解码模块，接收端语音解码模块负责接收所述语音编码模块传送过来的语音数据包，对其进行解压缩，获取语音基元编号，以该编号为检索条件，查询语音基元模型库，提取该编号对应的语音基元信息，最终通过语音合成算法还原语音。。

通过本发明提供的方法，在进行语音传输的时候，只需传输语音基元模型库中语音基元的编号、基频信号和音素声调编码即可。也就是说，如果采用 256 个聚类来描述人类的语音，而基频信号采用一个字节来记录的话，每帧语音信号（通常是 25 毫秒的语音，采用 16K16BitsPCM 格式需要 800 字节）只需要 2 个字节来表示。

当语音数据包被传输到目的地后，由语音解码模块对收到的语音

数据进行解码，并由语音合成方法完成语音合成工作。

语音合成过程是根据语音基元编号从语音基元模型库中获得语音谱包络特征。由于模板匹配分类过程可能产生错误，需要对取出来的特征进行平滑，如果相邻的模板之间距离过大，人耳就会听到刺激性的噪声，因此，从模板序号到特征的映射的过程，不仅仅是将模板均值取出来这么简单。模板库中，还需要保存各个特征的一阶差分和二阶差分信息，在解码的时候，利用最小二乘法求解出匹配误差最小，一阶差分误差也最小，二阶差分也最小的动态谱包络。

最后，用基频 F0 生成具有平坦谱包络的激励源，再用谱包络滤波此信号，合成相应的语音。

本发明的有益效果主要包括:

（1）与以往的以帧为单位，对每一帧的语音进行采样、编码的方法相比，本发明以语音基元为单位进行编码，由于每一种语言所构成的语音基元个数有限，因此，以语音基元为单位进行编码降低了编码空间;

（2）本发明通过建立语音基元模型库，在对语音基元进行编码时以语音基元模型对应的编号数值，取代以往编码方法中的采样点，即以一个数值替代多个数值，降低了编码字符串的长度，提高了编码的效率;

（3）在以语音基元编号数值进行编码的基础上，本发明采用相应压缩算法对其可压缩性进行分析，从而进一步压缩，以便在网络性能较差、带宽较小的情况下，能够对语音信息进行可靠

（4）本发明是一种在网络性能处于极限状态下的极限语音编码、传输和合成方法，可用于一些特殊情况下对语音通信的需求。

附图说明

图 1 是本发明中系统总体框架图；

图 2 是本发明中提取 MFCC 特征图；

图 3 是本发明中音素切分流程图。

具体实施方式

本发明中的语音基元可以是音素，也可以是等帧或变帧截取的波形，采用不同的语音基元便能够建立不同的语音基元模型库。在具体实施时，可以以其中的一种模型库为基础，在此之上对传输的语音进行编码和解码；也可以将几种模型库组合使用，对一些特殊情况下的复杂的语音进行编码。

本发明的基本构思为：采集大量的语音流数据样本，对连续的语音流进行语音基元的自动切分，形成语音基元集，提取语音基元的特征，并采用模糊聚类的方法对语音基元集进行聚类，从而建立语音基元模型库；以建立的语音基元模型库为基础，当获得连续语音流时，则对语音流进行语音基元的自动切分，然后在语音基元模型库中搜索出与当前语音基元最接近的模型，采用此模型的编号及其他相关信息通过语音编码后传输给接收方，接收方收到该语音数据包后由语音解码处理模块根据收到的语音基元编号，查找语音基元模型库，并根据上下文重估出语音包络，结合基频合成语音。

图 1 是本发明系统总体框架图。

首先在 101 处，采用隐马尔可夫模型（HMM）对连续语音流样本进行语音基元的自动切分，构成语料库；

在 102 处，通过图 2 Mel 频率倒谱系数（Mel-Frequency Cepstrum Coefficients）的方法从每一个语音基元中提取出 MFCC 特征；

MFCC 定义为语音信号经过快速傅立叶变换后所得到的加窗短时信号的实倒谱。与实倒谱的不同之处在于，加窗短时信号的实倒谱使用了非线性频率刻度，以与人的听觉系统相接近。

通过 MFCC 算法对语音基元的特征进行提取后，每个语音基元便可表示为相应的特征矢量，语料库便转换为相对应的语音基元特征矢量库。

在 103 处，通过模糊聚类的方法，根据语音基元的 MFCC 特征，对构成的语音基元集进行聚类，根据所使用的语言的特征，将语音基元聚为 N 类，进而构造包含有 N 类语音基元的模型库，具体的聚类过程为：

首先准备采集的语音基元特征集，$X=\{x_i, i=1,2,\ldots,n\}$是 n 个语音基元样本组成的样本集合，c 为预定的类别数目，m_j，j=1,2,...c 为每个聚类的中心，$\mu_j(x_i)$是第 i 个样本对于第 j 类的隶属度函数。用隶属度函数定义的聚类损失函数见下式（1）。

$$J=\sum_{j=1}^{c}\sum_{i=1}^{n}[\mu_j(x_i)]^b\|\mathbf{x_i}-\mathbf{m_j}\|^2 \qquad (1)$$

其中，b>1 是一个可以控制聚类结果的模糊指数。

在不同的隶属度定义方法下最小化式(1)的损失函数，且要求一

个样本对于各个类聚类的隶属度之和为 1，即：

$$\sum_{j=1}^{c}\mu_j(x_i)=1,\ i=1,2,\ldots,\ n \tag{2}$$

在条件式(2)下求式(1)的极小值，令J对 m_j 和 $\mu_j(x_i)$ 的偏导数为0，可得必要条件：

$$m_j=\frac{\sum_{i=1}^{n}[\mu_j(x_i)]^b x_i}{\sum_{i=1}^{n}[\mu_j(x_i)]^b},\ j=1,2,\ldots,\ c \tag{3}$$

$$\mu_j(x_i)=\frac{\left(1/\|x_i-m_j\|^2\right)^{\frac{1}{b-1}}}{\sum_{k=1}^{c}\left(1/\|x_i-m_k\|^2\right)^{\frac{1}{b-1}}} \tag{4}$$

用迭代方法求解式(3)和式(4)，当算法收敛时，就得到了各类音素的聚类中心和各个样本对于各类的隶属度值，从而完成了模糊聚类的划分，对每一类语音基元进一步处理，抽取出能够代表该类的语音基元，从而构建语音基元模型库。

语音基元模型库建立后，便可以基于该语音基元模型库，对获得的连续语音流进行分析。在 104 处，对获得的的语音流进行语音基元的自动切分，并采用 Mel 频率倒谱系数为语音信号特征参数，提取语音基元的特征：

$$c_n=\sum_{m=0}^{M-1}S_2[m]*\cos\left(\frac{2\pi mn}{2M}\right)\ \ n=0,1,\cdots,N-1 \tag{5}$$

$$m=\frac{1000\cdot\ln\left(1+\frac{f}{700}\right)}{\ln\left(1+\frac{1000}{700}\right)}\approx 1127\ln\left(1+\frac{f}{700}\right) \tag{6}$$

在 105 处，通过如下公式判断当前 MFCC 特征对应的最佳模型：

$$P(M_i \mid X) = \frac{P(X \mid M_i)P(M_i)}{\sum_j P(X \mid M_j)P(M_j)} \quad (7)$$

$$P(X \mid M_i) = \frac{1}{\sqrt{2\pi}\,|\Sigma|}\exp\{-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)\} \quad (8)$$

最终获得最佳模型序号为 $n = \arg\max_i\{P(M_i \mid X)\}$

在 106 处，将音素模型对应的序号 n、基频及其他相关信息，按照一定的格式进行编码;

在 107 处，根据 106 处发过来的编码信息，采用压缩算法对其进一步压缩，并按照网络协议将其打包传输;

在 108 处，根据最佳模型序号 n，取出对应模型的均值、一阶差分、二阶差分，联合前面 N 帧的知识，采用最小二乘法，以误差总和最小为原则，求出最佳的谱包络特征。

在 109 处，根据基频 F0，生成频谱均匀的激励源信号，对此信号进行滤波，使得其谱包络为 104 处提取的包络，则此语音为恢复出来的结果。

下面以音素为例，进一步阐述音素的自动切分过程、聚类、建模型库和编码解码过程:

获得连续语音流后，便可对连续语音流进行分析，如图 3 所示，先以音节为单位对连续的语音流进行切分，如汉语发音中每一个字即是一个音节，这一切分过程实际上是将连续语音流中的每一个字的发音切分出来;

切分出音节后，再对每一个音节进行分析，如果该音节是由单个音素构成，则将该音素存入语料库;

如果该音节不是由单个音素构成，则对该音节进一步切分，将该音节切分成由多个单个音素构成，并将这些音素存入语料库；

参考郑鸿的《基于 HMM 的普通话连续语流中音素的自动切分》，如果将连续语音流中出现的语音数据看作是一个随机过程，则语音序列可看作是一个随机序列，进而建立马尔可夫链以及隐马尔可夫模型（HMM）；

为 HMM 模型分配累计器，并将累计器清零；

获得含有大量音素的语料库，然后将语音序列样本对应的描述符号相应的 HMM 连接起来组成一个组合 HMM；

计算组合 HMM 的前向和后向概率；

使用计算所得的前向和后向概率计算每一时间帧的状态占有概率，更新相应的累计器；

对所有语音数据样本中的数据进行上述过程，完成对语音样本的训练；

使用累计器的值计算 HMM 的新估计参数；

将每一个 HMM 的状态 θ_i 拥有的每一个令牌的拷贝传递到所有相邻的状态 θ_j，并增加该令牌拷贝的对数概率 $\log\{a_{ij}\}+\log\{b_j(O_t)\}$；

每一个后续状态检查前面状态传递过来的所有令牌，保留最高概率的令牌，其余的丢弃；

经过上述过程后，便可对连续的语音流进行自动识别切分，获得连续的音素序列。

完成上述音素的自动切分后，便可对音素集进行模糊聚类，可根

据不同语言音素的构成特征设定模糊聚类的聚类个数，如汉语的语音可由 29 个基本音素及其组合构成，具体参见黄中伟等的《普通话语音识别中的基本音素分析》，因此，本实施例中在对音素进行聚类时，将聚类的个数设为 30，模糊指数 b 设为 2，完成聚类后，以每一类的类心作为该类的特征音素：

$$m_j = \frac{\sum_{i=1}^{n}[\mu_j(x_i)]^b x_i}{\sum_{i=1}^{n}[\mu_j(x_i)]^b}, \quad j = 1,2,\ldots, \quad c$$

因此，便可生成一个由 30 个音素所构成的语音基元模型库，该语音基元模型库的结构如下：

语音基元编号	语音基元	语音基元基频	语音基元波形

采用 Mel 频率倒谱系数，提取收到的连续语音流中每一个音素的频谱包络特征，并将该频谱包络特征与语音基元模型库中的语音基元的波形进行匹配，从而获得当前音素的编号。

将连续获得的音素编号、音素的基频，进行编码，并可通过压缩算法，如 LZW 数据压缩算法进一步压缩，然后将压缩后的数据包通过网络或电话通信网络传输到目的地。

接收端收到数据包解压缩后，取出数据包中的音素编号序列，并根据最佳模型编号 n，取出对应模型的均值、一阶差分、二阶差分，联合前面 N 帧的知识，采用最小二乘法，以误差总和最小为原则，求出最佳的谱包络特征。

最后，根据基频 F0，生成频谱均匀的激励源信号，对此信号进行滤波，使得其谱包络为 104 处提取的包络，将语音还原。

以上公开的仅为本发明的一个具体实施例，但是，本发明并非局限于此，任何按照本专利发明内容所描述方法而设计的变化都应落入本发明的保护范围。

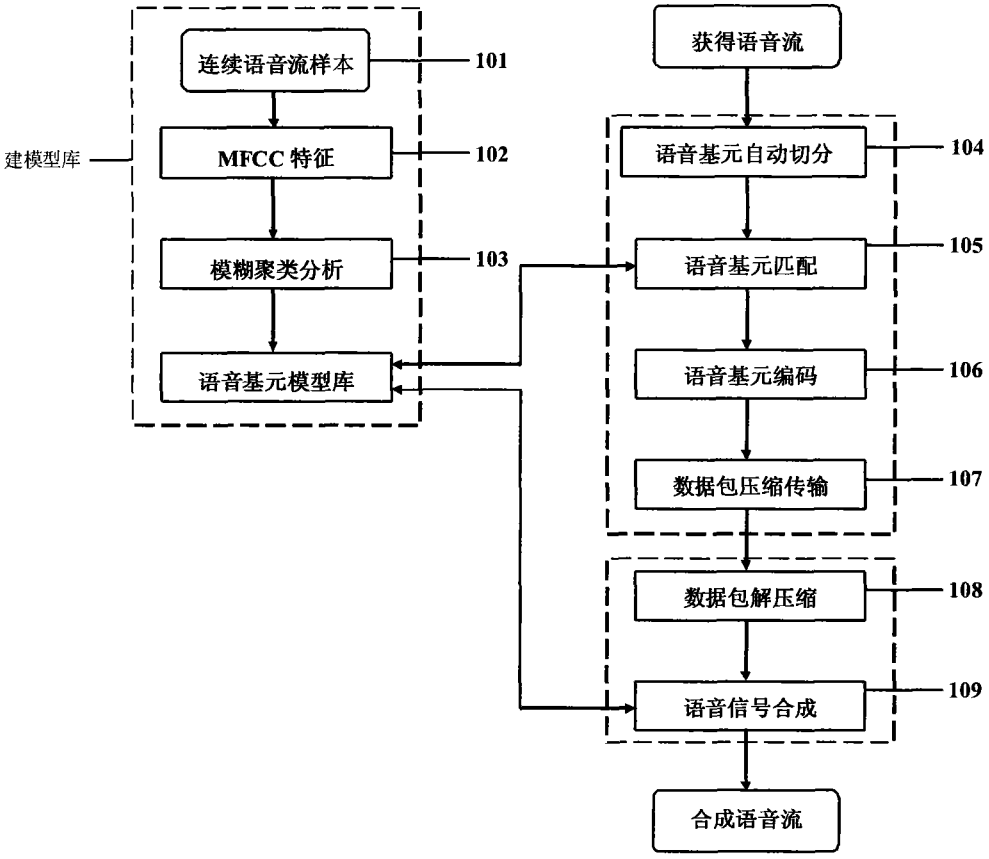

连续语音流样本
101
建模型库
MFCC 特征
102
模糊聚类分析
103
语音基元模型库
获得语音流
语音基元自动切分
104
语音基元匹配
105
语音基元编码
106
数据包压缩传输
107
数据包解压缩
108
语音信号合成
109
合成语音流

图 1

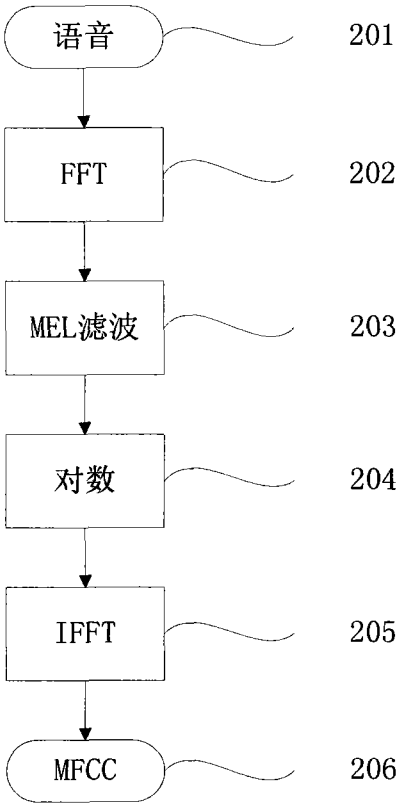

语音
201
FFT
202
MEL滤波
203
对数
204
IFFT
205
MFCC
206

图 2

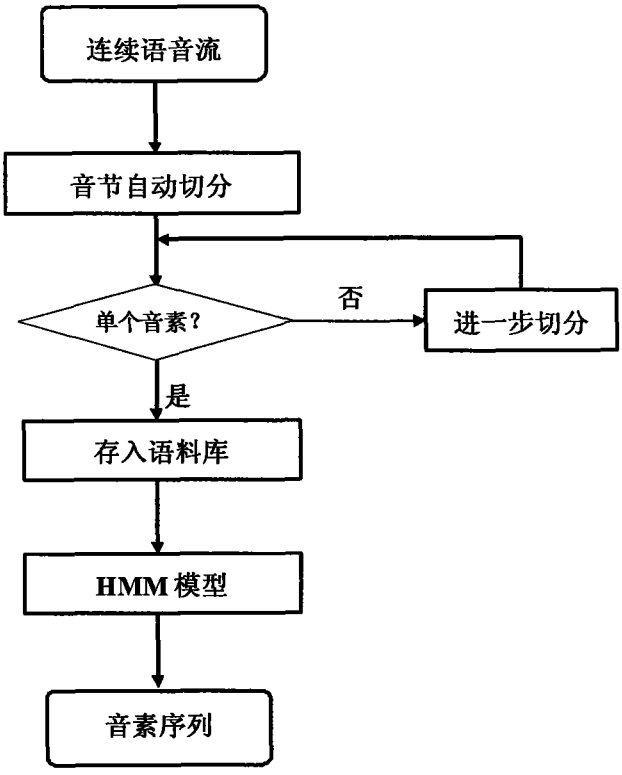

连续语音流
音节自动切分
单个音素?
否
进一步切分
是
存入语料库
HMM 模型
音素序列

图 3