



[12] 发 明 专 利 说 明 书

[21] ZL 专利号 98811825.4

[45] 授权公告日 2003 年 9 月 3 日

[11] 授权公告号 CN 1120469C

[22] 申请日 1998.12.30 [21] 申请号 98811825.4

[30] 优先权

[32] 1998. 2. 3 [33] DE [31] 19804189.6

[86] 国际申请 PCT/DE98/03812 1998.12.30

[87] 国际公布 WO99/40568 德 1999.8.12

[85] 进入国家阶段日期 2000.6.5

[71] 专利权人 西门子公司

地址 德国慕尼黑

[72] 发明人 K·许利希 W·弗拉尔斯

审查员 杨 叁

[74] 专利代理机构 中国专利代理(香港)有限公司

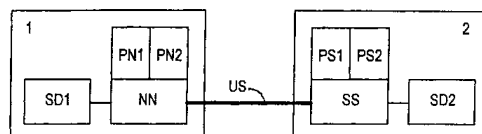
代理人 程天正 张志醒

权利要求书 3 页 说明书 6 页 附图 1 页

[54] 发明名称 传输语音数据的方法

[57] 摘要

为了传输语音数据, 语音数据流被分解成音素, 在一个可选的具有语音和/或说话人特征的音素目录(PN1、PN2)中, 为每个音素分配一个代码字符, 然后将该代码字符传输到传输目的地(SD2)的语音合成装置(SS)中, 这样, 待传输的数据量被大大缩减。语音数据流分解成音素是由神经网络(NN)来实施的, 此神经网络被训练用来识别存放于所选择具有语音和/或说话人特征的音素目录(PN1、PN2)中的音素。所接收的代码字符流又由语音合成装置(SS)转换成音素序列, 然后输出。



1. 运用具有语音和/或说话人特征的音素目录(PN1、PN2、PS1、PS2)将语音数据从语音数据源(SD1)传输到传输目的地(SD2)的方法,在这些目录中,存放有与语音数据模型相对应的音素,并且每个音素均被分配了一个单值的代码字符,它具有以下步骤:

a) 根据发送语音数据的用户的标识,选择一个给该用户分配的、具有说话人和/或语音特征的音素目录(PN1),

b) 待传输的语音数据被传送到一个被训练用来识别音素的神经网络(NN)中,该音素是存放在被选择出的音素目录(PN1)中的,神经网络就所含音素方面对语音数据作出分析,

c) 为语音数据中被识别的音素确定其在所选择音素目录(PN1)中所分别分配的代码字符,

d) 代码字符被传输到传输目的地(SD2)处的语音合成装置(SS)中,

e) 所接收的代码字符流由语音合成装置(SS)转换成在音素目录(PS1)中均分配了音素的代码字符序列,以及

f) 输出这个序列。

2. 运用具有语音和/或说话人特征的音素目录(PN1、PN2、PS1、PS2)将语音数据从语音数据源(SD1)传输到传输目的地(SD2)的方法,在这些目录中,存放有与语音数据模型相对应的音素,并且每个音素均被分配了一个单值的代码字符,该方法具有以下步骤:

a) 待传输的语音数据被传输到一个被训练用来识别不同语音和/或说话人的神经网络中,此神经网络识别待传输语音数据所属的语音,和/或待传输语音数据的来源-说话人,此神经网络还促使选择一种分配给语音和/或说话人的、具有语音和/或说话人特征的音素目录(PN1),

b) 待传输的语音数据被传送到一个被训练用来识别音素的神经网络(NN)中,该音素是存放在被选择出的音素目录(PN1)中的,神经网络就所含音素方面对语音数据作出分析,并被训练用来识别待传输的语音数据,

c) 为语音数据中被识别的音素确定其在所选择音素目录(PN1)中所分别分配的代码字符,

- d) 代码字符被传输到传输目的地 (SD2) 处的语音合成装置 (SS) 中,
- e) 所接收的代码字符流由语音合成装置 (SS) 转换成在音素目录 (PS1) 中均分配了音素的代码字符序列, 以及
- f) 输出这个序列.
- 5 3. 根据权利要求 1 或 2 的方法,
其特征在在于,
从语音数据源 (SD1) 传输到传输目的地 (SD2) 的语音数据作为
代码字符临时存储在一个存储器中.
- 10 4. 根据权利要求 1 或 2 的方法,
其特征在在于,
信息经过语音合成装置 (SS) 的音素目录 (PN1) - 在给待传输语
音数据进行音素分析时采用 - 被传输到传输目的地 (SD2), 以便在
那儿用来重建传输语音数据.
- 15 5. 根据权利要求 1 或 2 的方法,
其特征在在于,
在语音传输过程中, 音素目录 (PN1、PN2、PS1、PS2) 利用待传
输语音数据进行更新.
- 20 6. 根据权利要求 1 或 2 的方法,
其特征在在于,
在语音传输过程中, 神经网络 (NN) 被训练用来识别待传输的语
音数据.
- 25 7. 根据权利要求 1 或 2 的方法,
其特征在在于,
待传输的代码字符流在传输之前通过使用另一种压缩方法进行压
缩, 并在传输目的地 (SD2) 被再次解压.
- 30 8. 根据权利要求 1 或 2 的方法,
其特征在在于,
语音合成装置 (SS) 输出的音素其长度、响度、重音、音域和/或
声音可调.
9. 根据权利要求 8 的方法,
其特征在在于,
声响特征和/或声音特征在压缩之前从待传输的语音数据中调出

来，然后传输到语音合成装置（SS），以便在那儿被用来对长度、响度、重音、音域和/或声音-音素是利用它们进行输出的-进行调整。

10. 根据权利要求8的方法，

其特征在于，

- 5 发送语音数据的用户的标识被传输到语音合成装置（SS），并在那儿根据标识对长度、响度、重音、音域和/或声音-音素是利用它们进行输出的-进行调整。

传输语音数据的方法

技术领域

- 5 本发明涉及一种传输语音数据的方法，其中，语音数据在传输之前被压缩，在到达传输目的地后又被解压。在此，压缩是以把语音数据分解成音素为基础的。音素是感觉说话语言的基本声学语言元素。

背景技术

- 10 大家已知道，语音数据在通信网内传输之前需要被压缩，以尽可能少地占用通信网的传输带宽。在这种情况下，若想在传输目的地再现语音，被压缩的语音数据就要通过解压恢复到原来的状态或与之等效的状态。由于通过这种方法实现的传输带宽缩减直接取决于所使用的压缩方法压缩率，所以争取一个尽可能高的压缩率是很有必要的。

- 15 在语音传输中，通常采用预测方法来进行压缩，它们利用了语音数据中产生的数据模型在统计上的非均匀分布，来减少语音数据所特有的高冗余。在解压时，可以由被压缩的语音数据重建最初的语音数据，并且只有微小的、方法所固有的损失，甚至毫无改变。在此，所达到的压缩率大约在 1: 10 这个数量级。这种方法已经示例性地在“信息与编码”一书中描述过，其作者为 Richard W.Hamming, Weinhein
20 VCH 出版社，1987，第 81 页至 97 页。

- 在典型的语音数据中，纯内容的信息仅占全部语音信息的很小一部分。通常，语音信息的绝大部分是由说话人的特征信息组成，该特征信息譬如用说话人的声音细微差别或音域细微差别来表达。语音数据的传输基本上只取决于其内容信息 - 比如在纯信息通知、自动报告
25 中等，因此，通过对说话人的特征信息进行缩减，也可以获得一个比较高的压缩率，该压缩率比采用维持语音数据的信息内容完整或接近完整的方法要高得多。

- 表达说话人语音以及内容信息 - 所说出的词 - 的最小声学单位是音素。专利文献 EP 71716 B1、DE 3513243 C2 和 EP 423800 B1 中曾
30 公布过一些装置和方法，其中，对语音数据流所包含的音素进行分析，然后将其转换成一种分配给可识别音素的代码字符流，以便由此在传输之前对语音数据进行压缩。

在此，存在一个重要的问题，即任何待传输的语音数据流其组成

- 成分-音素-需要可靠地识别。尤其困难的是,按照说话人及其说话习惯,同一个音素可能产生很不相同的表征。如果语音数据流中的音素不能被识别或分配给了错误的声音,则语音的传输质量就会变坏-可能导致语音不清晰。因此,对这种语音传输方法的质量以及应用范围来说,可靠的音素分析是一个重要的标准。

发明内容

本发明的任务在于,提供一种传输语音数据的灵活有效方法,在该方法中,通过使用一种经过改善的音素分析,语音数据能够在传输之前进行压缩。

- 在本发明方法中,对于从语音数据源传输到传输目的地的语音数据,它们要在真正传输之前进行音素分析。为了使用该方法,语音数据可以以各种形式出现;比如以模拟或数字的形式,或者为描述成语音信号的特征向量,它们均为时间分辨和/或频率分辨的表示法。根据本发明,音素分析由一种神经网络来实现,此神经网络被训练用来对音素进行识别。借助神经网络进行语音或音素识别的原理已在诸如“用于语音识别的神经网络回顾(Review of Neural Networks for Speech Recognition)”中讲述过,其作者为 R. P.Lippmann,此文出自《神经计算》1989年1月,1至38页。

- 语音数据流是按照音素进行分析的,神经网络则训练成与音素相适配,这种音素存放在具有语音特征和/或说话人特征的音素目录中,在此,它们均被分配了单值的代码字符-如一个索引或一个数字。在这种或其它类的关系中,作为语音的土语、方言以及具有语音特征的表达都能够被理解。在本发明方法所提供的步骤中,通过对属于各种不同语音和/或源自不同说话人的有代表性语音数据进行分析,可以构建具有语音或说话人特征的音素目录。在此,这种声学语言元素-对于每个有代表性的语音数据,该语言元素被证明是有特征的,或者为该语言元素的有代表性理想形式-作为音素被存放在相关的音素目录中。分析有代表性的语音数据尤其还可以利用神经网络来实现。此外,可以选择具有代表性的语音数据或其特征语言元素,以优选地适应待传输语音数据的特性-如背景噪声或其它类似物等。

音素目录比如可以作为一种常规的存储器和/或作为受训神经网络的一部分来实现,其中,在后一种情况下,所存放的音素可以由受训

神经网络的一个内部状态来体现。此外，由神经网络学习状态所确定的所谓计权矩阵也可以理解为音素目录。此种计权矩阵可以从受训的神经网络中读出，并存放到一个存储器中，然后在需要时读入到任意一个神经网络中去，由此使其成为一个相应的受训神经网络。

- 5 在语音数据的音素分析之前，首先通过一个神经网络把具有语音和/或说话人特征的音素目录选择出来，接着，将语音数据输至神经网络中进行音素分析-该神经网络被训练为对存放在音素目录中的音素进行识别。在此，音素目录的选择可以根据用来发送待传输语音数据的用户的标识来进行，或者由区别语音和/或说话人的神经网络来促成
- 10 -此神经网络被训练用来识别待传输语音数据所属的语音，和/或识别待传输的语音数据的来源，亦即说话人。

- 对于在音素分析时从语音数据内识别的音素，为它们查明其在选择音素目录中被分配的代码字符，然后用该代码字符代表同每个音素相对应的语音数据模型而传送到目的地。因为传输的代码字符通常比
- 15 由此所代表的语音数据模型要短得多，所以待传送的数据量通常被大大缩减了，从而获得一个高的压缩率。在传输目的地，接收的代码字符被输入一种语音合成装置，该装置把代码字符流再次转换为音素序列，并将其输出。在此，为了把所接收的代码字符替换为输出的音素，语音合成装置同样也对音素目录进行访问。压缩率还有可能通过下述
- 20 方法再度提高，即在传输之前，代码字符流通过另一种压缩方法进行压缩，而在传输目的地输送给语音合成装置之前，又再次被解压。

- 通过使用具有说话人语音特征的音素目录，可提高音素分析的可靠性，尤其是在传输源自不同说话人的语音数据时，它还可以提高语音合成质量。此外，通过选择作为语音合成根据的音素目录，语音数
- 25 据再现还可适合预定的要求，比如，模拟某个说话人或语音的特征，或者在特殊情况下通过改变声音来引起注意等。

- 按照本发明方法的一种优选扩展，在语音数据传输期间，音素目录可通过分析具体待传输的语音数据-有时是多次-来进行更新，以及/或者神经网络借助待传输语音数据继续进行训练。这样，音素分析
- 30 或语音合成也能适应语音特征的短时改变。

 根据本发明的一种扩展，对于语音合成装置在音素分析中所使用的音素目录，信息能够经过它进行传输。利用这种方式，比如，语音

合成装置可以示例性地根据说话人或语音的改变而指示出音素目录的变化,或者,如果需要要用到的音素目录还不存在,就把信息全部或部分地传送出去。这样,音素目录的改变或更新也可以传输给语音合成装置,并由此能够被使用。

- 5 本发明的另一个优选扩展规定:在语音合成装置中,各种用于产生音素的相关参数-如长度、响度、重音、音域、声音等-是利用音素来再现和输出的,并且是可调整的;特别是对于每个独立的音素。这样,所传输数据的再现能够适合不同的要求。例如,再现质量可通过下述方法得以改善,即响度和/或声音特征-如音节或与其相类似的语言元素的响度和音高-在压缩前从待传输的语音数据中调出来,然后传送给语音合成装置以调整再现参数。再现参数也能够根据用来发送语音数据的用户的标识进行调整,以让用户特征元素在语音再现中起作用-比如,该特征元素能够使用户借助再现的声音而被认出。

- 10 另外,还有一种可能性,即从语音数据源传输到传输目的地的语音数据可作为代码字符临时存放在存储器中,以待以后一次或多次调用。为此,通过在代表音素的代码字符基础上对语音数据进行缩减,通常会比临时存储原样不动的语音数据要占用更少的存储空间。

- 此外,代表音素的代码字符形成了一个可能的输出点,以用于语音数据多种多样的继续处理。例如,代码字符能够转换成与分配音素相对应的字母群,因而语音数据可以以文本的形式输出。还有许多用于自动语音识别的流行方法,它们是以把语音分解成音素为基础的,从而可以与本发明的方法联合使用。

附图说明

结合附图,下面将借助一个优选实施例来详细阐述本发明。

- 25 附图1简略地示出了两个通过传输链路连接起来的通信设备。

具体实施方式

- 图1简略地示出了作用为语音数据发送机的通信设备1和与之通过传输链路US相连的、作用为语音数据接收机的通信设备2。通信设备1包括作为功能部件的语音数据源SD1和与之相连的神经网络NN,此神经网络可访问具有说话人特征的音素目录PN1和PN2。在通信设备2中,描绘了语音合成装置SS和一个与之相连的语音数据接收器SD2,语音合成装置SS与具有说话人特征的音素目录PS1和PS2

连接在一起。

各通信设备只限制有两个具有说话人特征的音素目录，在此只作为例子来看待，以简化如下的阐述。

5 在具有说话人特征的音素目录 PN1 和 PN2 中，均存放有音素，这些音素是通过分析有代表性的语音数据、并作为特征声学语言元素来进行识别的。在此，对于每个音素目录，用于创建音素目录 PN1、PN2 的有代表性的语音数据均源自特定的说话人。所存放的音素均被分配了一个代码字符，该代码字符对相关音素目录来说，是单值可调用的。

10 神经网络 NN 被训练用来从语音数据流中识别存放于音素目录 PN1、PN2 中的音素，其中，音素目录 PN1、PN2 中具体哪一个作为音素分析的根据，是可以调整的。这种调整比如可以这样来进行：为神经网络印制一个与每个音素目录 PN1 或 PN2 相对应的所谓计权矩阵。

在实施例中，由语音数据源 SD1 发出的语音数据流从通信设备 1
15 传输到通信设备 2 的语音数据接收器 SD2，其中，语音数据在真正传输之前要进行压缩。对此，用来发送语音数据的用户其标识首先从语音数据源 SD1 输送到神经网络 NN，此神经网络根据传输的标识，选出一个给用户分配的具有说话人特征的音素目录 - 此处为 PN1 - 以用于此后的音素分析。此标识随后继续传送到通信设备 2。

20 随即，待传送的语音数据流被输入神经网络 NN，根据放置在选择语音数据目录 PN1 中的音素，该网络 NN 把语音数据流中产生的语音数据模型进行分类。对于能够有效分配给存放音素的每一个语音数据模型，在音素目录中给存放音素所分配的代码字符将代替该语音数据模型经传输链路 US 而传输到通信设备 2。不能有效分配给存放音素的
25 的语音数据模型，则原样不动地传输 - 用一个与分配给音素的代码字符不相同的特殊代码字符进行表征。

在通信设备 2 中，首先根据传输标识选出一个作为随后语音合成基础的音素目录 - 此处为 PS1。接着，随后接收的代码字符流被传送到语音合成装置 SS，此装置为每个代码字符生成一个其在所选择音素
30 目录 PS1 中所分配的音素，然后传送到语音数据接收器 SD2。那些用特殊代码字符标记的、原样不动传输的语音数据模型则直接被传送到语音数据接收器 SD2。这样，语音数据流再次近似地得到重建和解压。

音素目录 PS1 和 PS2 在实施例中是作为一种常规存储器实现的，其中，每个所包含的代码字符都被分配了一个信息，该信息是语音合成设备 SS 中生成分配音素时所必需的。音素目录 PS1 和 PS2 可通过调用或读出音素目录 PN1 或 PN2 来进行创建。

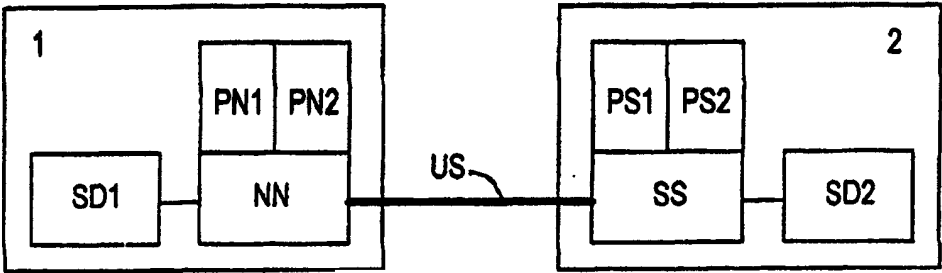


图 1