

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁷

G10L 15/02

G10L 15/08

G10L 19/00

[12] 发明专利说明书

[21] ZL 专利号 94190055.X

[45]授权公告日 2000 年 2 月 2 日

[11]授权公告号 CN 1049062C

[22]申请日 1994.2.10 [24]颁证日 1999.11.13

[21]申请号 94190055.X

[30]优先权

[32]1993.2.12 [33]FI [31]930629

[86]国际申请 PCT/FI94/00054 1994.2.10

[87]国际公布 WO94/18669 英 1994.8.18

[85]进入国家阶段日期 1994.10.11

[73]专利权人 诺基亚电信公司

地址 芬兰埃斯波

[72]发明人 马科·万斯卡

审查员 何越峰

[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

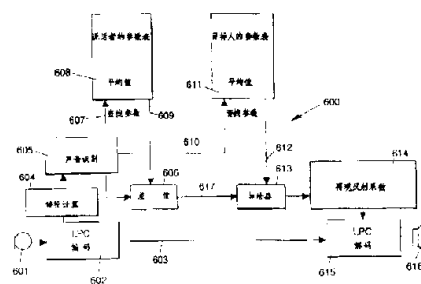
代理人 杨国旭

权利要求书 1 页 说明书 7 页 附图页数 5 页

[54]发明名称 转换语音的方法

[57]摘要

本发明涉及语音转换的方法,其中根据说话人的语音信号计算反射系数,从中计算模拟说话人声管的无损耗管的圆柱部分的横截面积的特征,从所述特征中识别声音并提供识别符。而后,计算代表所述声音的特征与代表同一声音的后序特征之差,根据识别符在存储器中查寻对同一声音模拟那个说话者的声道的第二说话人的说话者特定特征,通过把所述差值和第二说话者的特定特征相加得到和值,算出新的反射系数,产生新的语音信号。



ISSN 1008-4274

权 利 要 求 书

1. 一种转换语音的方法,其中由第一说话人产生的语音信号(IN)被取样,以计算反射系数(r_k),该方法的特征在于下列步骤:

从反射系数(r_k)中计算(16;51;604)模拟第一说话人声道的无损耗管(图1和2)的截面积的特征,

第一说话者的无损耗管(图1和2)的圆柱部分的截面积(图2; A_k)的所述特征与用于声音识别的模拟说话者的声道的无损耗管的圆柱部分的截面积(A_k)的至少一个以前说话者的各个存储的特定声音特征相比较(17;52;605),以给所识别的声音提供各自的识别符,

对所述声音计算模拟说话者声道的无损耗管的圆柱部分的截面积(图2; A_k)的存储的特征与同一声音的后序各个特征的差;

根据所识别的声音的识别符在存储器(611)中查找对同一声音的模拟说话者的声道的无损耗管的圆柱部分的截面积(图2; A_k)的第二说话者的特定说话的特征,

通过把所述差值(617)与对同一声音的模拟那个说话者的声道的无损耗管的圆柱部分的截面积的第二说话者的说话者特定特征(612)相加构成(20;613)一个和值;

根据该和值计算新的反射系数,以及

从所述新的反射系数中产生(615)一个新的语音信号(616)。

2. 根据权利要求1的方法,其特征在于对代表第一说话人的同一声音的无损耗管的物理尺寸计算一个特征,并存在存储器(608)中。

说明书

转换语音的方法

本发明涉及转换语音的方法,该方法中,由第一说话者产生的语音信号被取样用于计算反射系数。

有语音障碍的人的语音通常不清楚,而且听起来很难识别。语音障碍人的语音质量会产生问题,尤其是当用通信装置或通信网把语音障碍人产生的语音信号发送并转换到接收机。由于通信网有限的传输容量和声学特性,所以语音障碍人产生的语音对听者来说仍较难识别和听懂。另一方面,无论是否用通信装置或通信网传送语音信号,对听者来说总是很难识别并明白语音障碍人的语音。

此外,在需要改变说话者产生的语音时,即校正语言声音为更好的声音格式或把说话者产生的语音转换成另一说话者的声音,这样,第一说话者的语音听起来就和第二说话者的语音一样。

本发明的目的是提供一种方法,使说话者的语音能以这样的方式被改变或校正,即使听者听到的语音或校正或改变后的语音信号对应于另一说话者产生的语音,或以某一希望的方式校正的同一说话者的语音。

这种新颖的转换语音的方式就是根据本发明所提供的方法,其特征包括下列步骤:根据反射系数计算模拟第一说话者声道(vocal tract)的无损耗管的圆柱部分的截面积的特征,第一说话者的无损耗管的圆柱部分的截面积的所述特征与至少一个以前的讲话者各个存储的模拟该说话者声道的无损耗管的圆柱部分的截面积的特定声音特征相比较,以识别声音。为给所识别的声音提供各自的识别符,计算对所述声音模拟说话者声道的无损耗管的圆柱部分的截面积地存储的特性和同一声音的各自的后序特性之间的差值,根据所识别的声音的识别符,在存储器中检索模拟同样声音的那个说话者的声道的无损耗管的圆柱部分的截面积的第二说话者的特定说话特性,通过把所述差值和模拟同样声音的那个说话者的声道的无损耗管的圆柱部分的截面积的第二

说话者的特定说话特征相加构成一个和值,根据该和值计算出新的反射系数,从新的反射系数中产生新的语音信号。

本发明基于下列思想,通过 LPC(线性预测编码)的方法分析语音信号,产生一套模拟说话者声道的参数,这些参数通常为反射系数的特征。根据本发明,声音是通过比较从要转换的声音的反射系数中计算出的无损耗管的圆柱部分的截面积和几个说话者以前接收的各自对同一声音计算的圆柱的截面积,从要转变的语音中识别的。此后,对每个说话者,每个声音的截面积计算某个特征,通常是平均值。然后,从这个特征中减去对应于每个声音的声音参数,即说话者的无损耗声道的圆柱的截面积,提供要传送到下一转换步的差值以及该声音的识别符。在此之前,对应于要模拟的说话者(即,目标人)的每个声音识别符的声音参数的特征已经确定,从而通过把所述差值和存储在存储器中检索到的目标人的同一声音的声音参数的特征相加,原始声音就可以被再现,就好象是目标人发出的声音。通过把那个差值相加,语言的声音,即不包括在基于识别符的声音中的声音间的信息被产生,其中对应于那些声音,即通常为说话者声道的无损耗管的圆柱的截面积的平均值的特征已在存储器中被找到。

这种转换语音的方法的优点是有可能以一种方式校正语言声音中发生的以及由说话者的物理特性引起的错误和不准确,使语音能更容易地被听者听懂。

而且,根据本发明的方法,有可能把一个说话者的语音转换成听起来象另一说话人的语音。

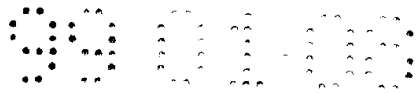
本发明所使用的无损耗管模型的圆柱部分的截面积可很容易地以常规的语音编码规则从所谓反射系数中计算出来。自然,该面积的其它截面尺寸,如半径或直径也可确定为一个参考参数。另一方面,除圆形之外,管的横截面也可以是其它形状。

下面将参照附图详细说明本发明。其中

图 1 和 2 通过包括模拟说话者声管的无损耗管的顺序的圆柱部分的无损耗管说明说话者声道的模型;

图 3 说明了说话期间无损耗管是如何改变的,

图 4 示出了声音是如何被识别并转换到与所希望的参数一致的



流程图；

图 5a 是说明在语言转换器的声级 (sound level) 上根据本发明的语音编码的框图，

图 5b 是通过语音信号转换方法说明根据本发明在一个声级上语音信号的再现步骤的处理图，

图 6 是实现根据本发明的方法的一个实施例的语音转换器的功能和简化的框图。

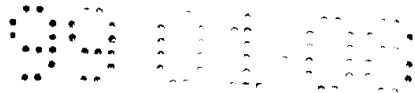
图 1 示出了包括顺序的圆柱部分 C1 到 C8 且构成人的声道的粗模型的无损耗管模型的透视图。图 1 的无损耗管模型的侧视图示于图 2。人的声道通常指由人的声带、喉、咽和唇限定的声音通道，通过该通道人产生语音声音。在图 1 和 2 中，圆柱部分 C1 说明了紧接在声带间声门后的声道部分的形状，圆柱部分 C8 指示出在唇部声道的形状，圆柱部分 C2 到 C7 说明声门和唇间离散的声道部分的形状。当产生不同的声音时，声道的形状在说话时通常是变化的。同样，代表声道的各个部分的离散的圆柱 C1 到 C8 的直径和面积在说话期间也是变化的。但是，同一发明人以前的国际申请 Wo. 92/20064 中公开了这种特征，从大量的瞬时声道形状计算出的声道的平均形状是每个说话者的一个不变的特征，该不变的特征用于在电信系统中声音更紧凑地传输，用于识别说话者或用于转换说话者的语音。因此，长时间内从声道的无损耗管模型的圆柱 C1 到 C8 的截面积的瞬时值计算的圆柱部分 C1 到 C8 的截面积的平均值也是相当准确的常数。而且，圆柱的截面尺寸的值也是由实际声道的值确定的，因此是说话者相当准确的常数特征。

根据本发明的方法利用在现有技术中线性预测编码 (LPC) 中作为临时结果产生的所谓反射系数，即具有与声道的形状和结构有特定联系的所谓 PARCOR—系数 r_k 。声道的无损耗管模型的圆柱部分 C_k 的反射系数 r_k 和面积 A_k 间的联系遵循公式(1)

$$-r(k) = \frac{A(k+1) - A(k)}{A(k+1) + A(k)} \quad (1)$$

其中 $K=1, 2, 3, \dots$

产生用于本发明的反射系数的 LPC 分析也被用在许多已知的语音编码方法中。



接下来将参阅图 4 总的说明对理解本发明很重要的那部分方法的步骤。在图 4 中,输入信号 IN 在方框 10 中以 8KHz 的取样频率被取样,形成一个 8 比特的取样序列 S_0 。在方框 11 中,从样值中抽取直流(DC)分量,以消除编码中可能出现的干扰侧音。此后,取样信号在方框 12 中由第一阶 FIR(有限脉冲响应)滤波器通过加权高信号频率进行预加重。在方框 13 中,样值被分割成每帧有 160 个样值的帧,每帧的持续期为 20ms。

在方框 14 中,通过自相关的方法在每帧上进行 LPC 分析来模拟语音信号谱,特征级为 $P=8$,然后,用下列公式(2)从帧中计算自相关函数 ACF 的 $P+1$ 值:

$$ACF(k) = \sum_{i=1}^{160} s(i)s(i-k) \quad (2)$$

其中 $K=0,1,\dots,8$

除自相关函数外,也可能用其它合适的函数,如协方差函数。用舒尔递推或其它合适的递推方法从所得到的自相关函数的值中计算出语音编码器中使用的短期分析滤波器的 8 个所谓反射系数 r_k 的值。舒尔递推每 20ms 产生新的反射系数。在本发明的一个实施例中,所述系数包括 16 比特,它们的数目为 8。如果需要,可通过较长时间地使用舒尔递推,使反射系数的数目增加。

在第 16 步,通过圆柱部分模拟说话者声道的无损耗管的每个圆柱部分 C_k 的截面积 A_k 从每帧算出的反射系数 r_k 中得出。由于舒尔递推每 20ms 产生新的反射系数,所以对每个圆柱部分 C_k 每秒将获得 50 个截面积。算出无损耗管的圆柱的截面积后,在第 17 步通过比较这些算出的圆柱体的截面积和参数存储器中所存的圆柱的截面积的值,识别语音信号的声音。这一比较操作在结合图 5a 中的标号 60、60A、61A 将详细给出。在第 18 步,在存储器中查找第一说话者对同一声音的以前的参数的平均值,从这些平均值中减去从同一说话者收到的样值的瞬时参数,从而产生一个差值,存在存储器中。

然后,在第 19 步,在存储器中查找预存储的目标人的有关声音的几个样值的圆柱体截面积的平均值,目标人是其语音将被模仿的人。目标人也可以是例如第一个说话者,但用这种方法说话者的发音错误通过用这种转换步骤,以及新的更精确的参数被校正,通过上述参数,

说话者的语音可被转换成例如更清楚或更容易识别。

此后在第 20 步,在上述第 18 步中计算出的差值加到该目标人的同一声音的圆柱体截面积的平均值上。在第 21 步,从该和值中计算反射系数,该反射系数在步 22 中进行 LPC 解码,解码产生的电语音信号加到例如麦克风或数据通信系统上。

在图 5a 所示的本发明的实施例中,以这样的方式描述用于在一个声级上编码语音的分析,即根据在要分析的语音信号的预定的声音中产生的瞬时无损耗管模型的圆柱部分的面积计算模拟声道的无损耗管的圆柱部分的截面积的平均值。一个声音的持续时间很长,所以从语音信号中出现的单个声音可计算出几个甚至数十个暂时连续的无损耗管模型。这示于图 3,示出了 4 个暂时连续的瞬时无损耗管模型 S1 到 S4。从图 3 中可清楚地看出无损耗管的单个圆柱体的半径和截面积随时间的变化。例如,瞬时模型 S1、S2 和 S3 可在同一声音期间被粗分类产生,从中计算出平均值。而模型 S4 明显不同且与另一声音相联,因此在平均时不予考虑。

接下来,将参照图 5a 的框图描述在一个声级上的语音转换。虽然语音可借助于单个声音被编码并转换,但在转换中用所有这些声音也是合理的,这种转换希望以这样的方式进行,即这些声音听起来是新的声音。例如,把语音转换成听起来象是另一个说话者说的,而不是实际说话者说的,或者例如以这样的方式提高语音质量,和原始的未转换的语音相比,使听者更清楚地区别被转换的语音的声音。在语音中,转换可用于例如所有的远音和辅音。

如果瞬时无损耗管 59 的每个圆柱部分的截面的尺寸在一个已知的说话者的各个声音的预定存储的有限值内,语音信号中产生的瞬时无损耗管模型 59(图 5a)可在方框 52 中被识别以对应一个特定声音。这些特定声音和特定圆柱有限值被存在产生一个所谓声音伪装(mask)的所谓量化表 54 中。在图 5a 中,标号 60 和 61 说明所述特定声音和特定圆柱有限值是如何在要识别的瞬时声道模型 59 需适合的允许区域 60A 和 61A(无阴影区域)中为每个声音产生一个伪装或模型。在图 5a 中,瞬时声道模型 59 适合声音伪装 60,而显然不适合声音伪装 61。这样,方框 52 用作一类声滤波器,它把声道模型分成正确的声音组 a、e、i

等。声音被识别后,根据在图 5a 的框 52 中识别的声音的识别符 53 在参数存储器 55 中查找对应于每个声音,如 a、e、i、k 的参数,这些参数是无损耗管的圆柱的截面积的特定声音特征,如平均值。在声音识别方框 52 中,通过在参数存储器 55 中查找对应于每个瞬时声音的参数,已可能对每个要识别的声音提供一个识别符 53。这些参数可输入到一个减法装置,根据图 5a 的方框 56,借助声音识别符,即无损耗管圆柱的截面积的特征,通常是平均值,计算在参数存储器中查找的声音参数和所述声音的瞬时值之间的差值。这个差值进而被送去以图 5b 的方式相加并解码,这将结合所述附图详细描述。

图 5b 是说明在本发明的语音转换方法中,在一个声级上再现语音信号的处理图,接收被识别的声音的识别符 500,并根据该声音参数 500 在参数存储器 501 中查寻与该声音相应的参数,并在 502 处提供给求和器 503,通过对该差值和参数求和产生新的反射系数。通过解码新的反射系数计算新的语音信号。这种通过相加产生语音信号将在图 6 和相应的解释中更详细地介绍。

图 6 是实现根据本发明的方法的一个实施例的语音转换器 600 的功能的和简化的框图。第一个说话者(即将被模仿的说话者)的语音通过麦克风 601 到达语音转换器 601。转换器也可连接到某个数据通信系统,从而使要被转换的语音信号作为一个电信号输入。由麦克风 601 转换的语音信号在 602 处进行 LPC 编码(编码),并从中为每个声音计算反射系数。信号的其它部分在 603 处被发出以供后面在 615 处的解码。所计算的反射系数被发送到一个单元 604,用于特征计算,该单元根据反射系数为每个声音计算模拟说话者的声道的无损耗管的圆柱的截面积的特征,声音识别单元 605 通过把根据第一说话者(即要被模仿的说话者)产生的声音的反射系数计算的说话声道的无损耗管模型的圆柱部分的截面积与存在某存储器中的至少一个以前的说话者的各个以前识别的特定声音值相比较来识别声音。比较的结果得到所识别的声音的识别符、借助所识别的声音的识别符,在 607 和 609 处,在说话者的参数表 608 中寻找参数,在表中已存储了该第一说话人(要被模仿的)对同一声音的各个参数,如平均值,减法装置 606 从中减去仅从同一说话人得到的样值的瞬时参数。这样就产生了一个差值,被存在

存储器中。

此外,借助在方框 605 中识别的声音的识别符,在 610、612 处从目标人,即第一说话人的语音要转换成的第二人的参数表 611 中查寻对应于那个识别的声音的特征,例如根据反射系数计算的模拟说话人声道的无损耗管的截面积的特定声音平均值,并提供给加法器 613。对于加法器也已在 617 处从减法装置 606 输入由减法装置计算的差值,该差值通过加法器 617 加到在目标人的参数表 611 中查寻到的特征上,例如加到根据说话者声道的反射系数计算的模拟说话者声道的无损耗管的圆柱的截面积的声音特定平均值。然后产生一个总和,在反射系数的再现方框 614 中计算反射系数。而且,从反射系数中产生一个信号,其中第一说话者的语音信号以这样的方式被转换成声结构,使听者相信他听到了第二说话者的语音,虽然实际的说话者是其语音已被转换的第一说话者,所以其语音听起来和第二说话者一样。这个语音信号进一步加到 LPC 解码器 615,其中这一语音信号被 LPC 解码并与该语音信号的 LPC 未编码部分相加,产生最终的语音信号,该信号在扬声器 616 中被转换成声结构。在这一步,该语音信号也可仅保留电结构,并被转送到某个数据或电信系统,以进一步发送或转送。

根据本发明的上述方法,在实际中例如可以用软件通过常规信号处理器来实现。

附图和相关的解释仅仅意在说明本发明的思想。致于具体细节,根据本发明的转换语音的方法可在权利要求的范围内变化。显然,上述本发明是主要结合语音模仿进行描述的,但所述语音转换器也可用于某种语音转换。

说明书附图

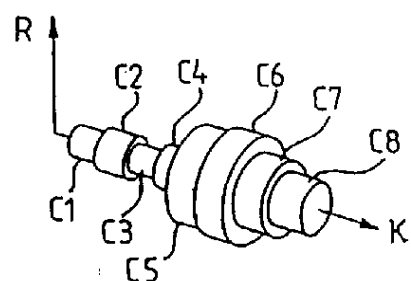


图1

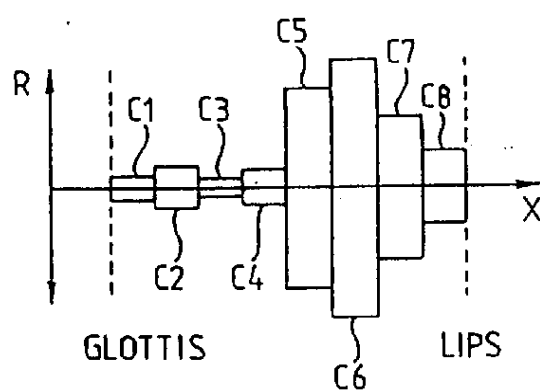


图2

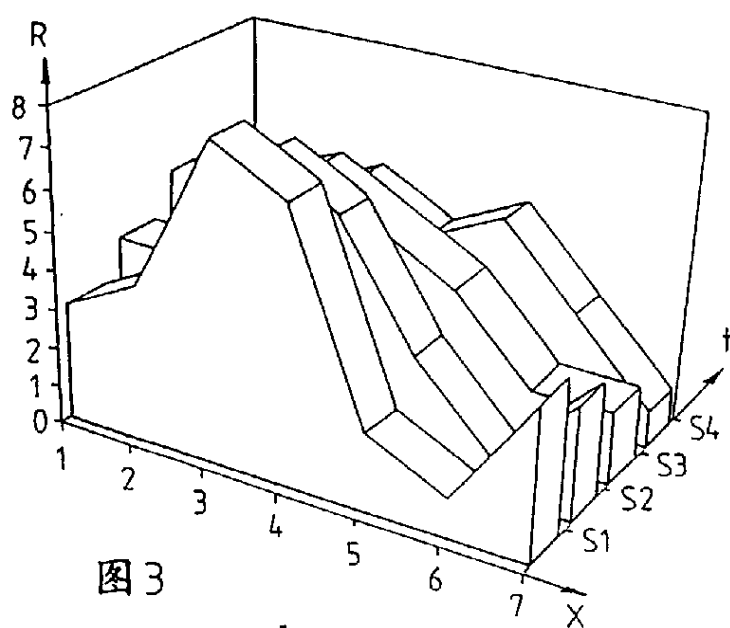
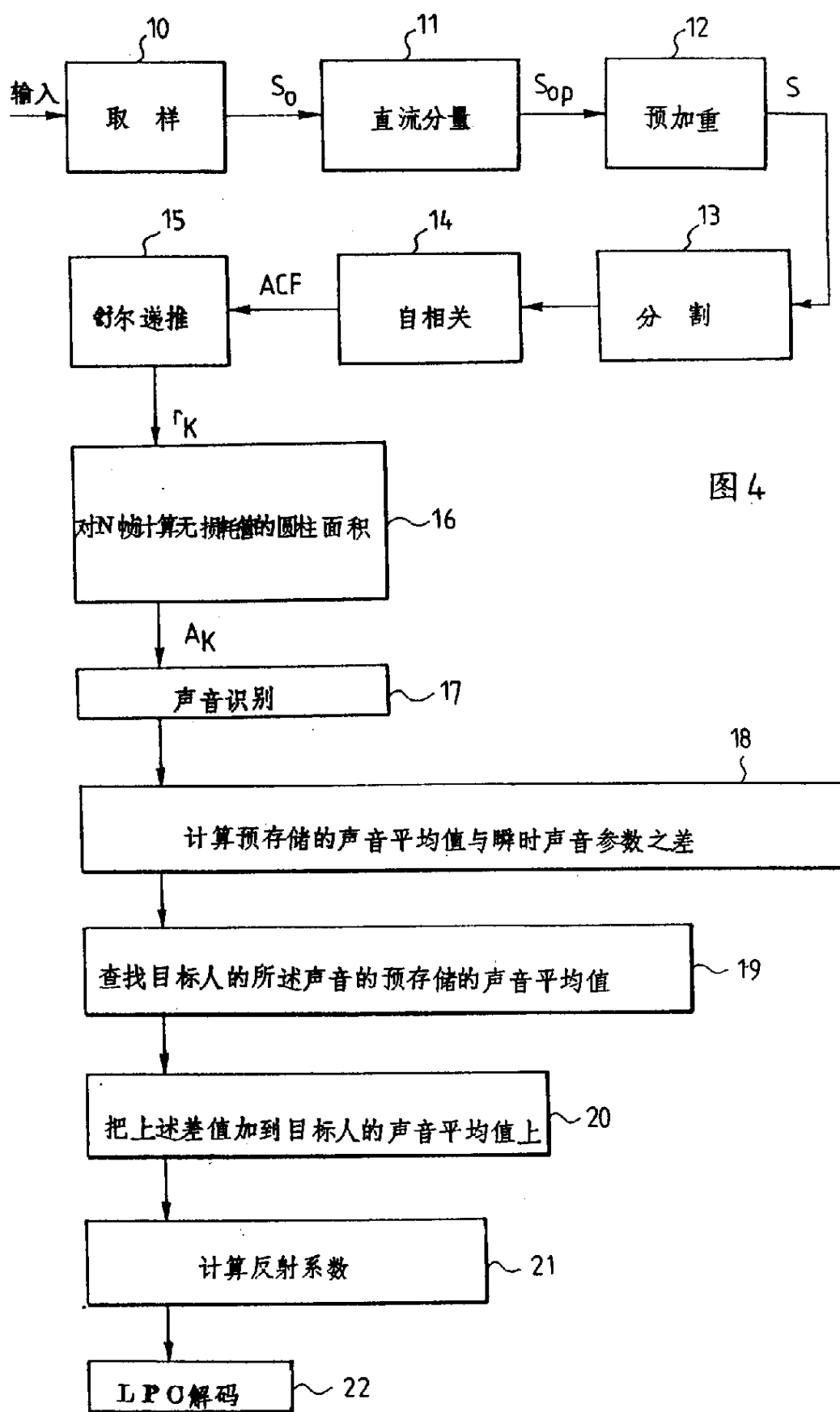
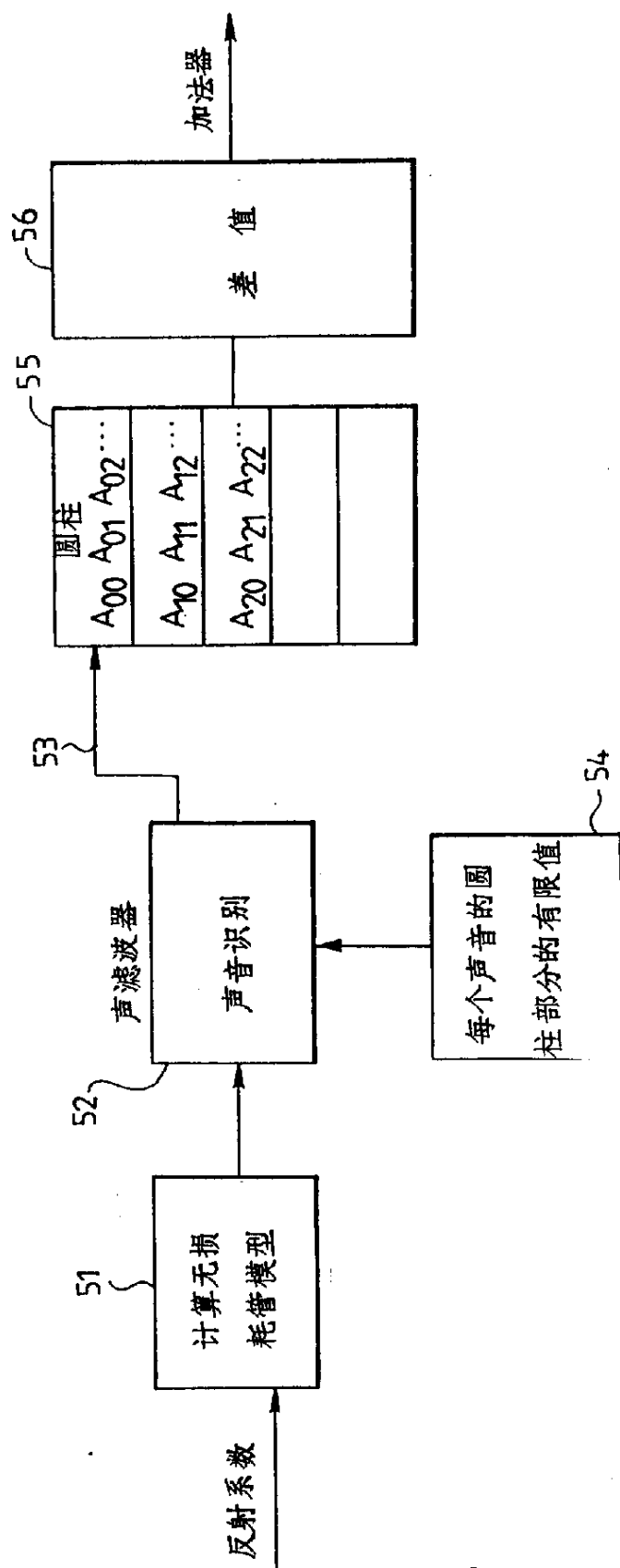


图3





声道59的
瞬时形状



声音伪装



图5a

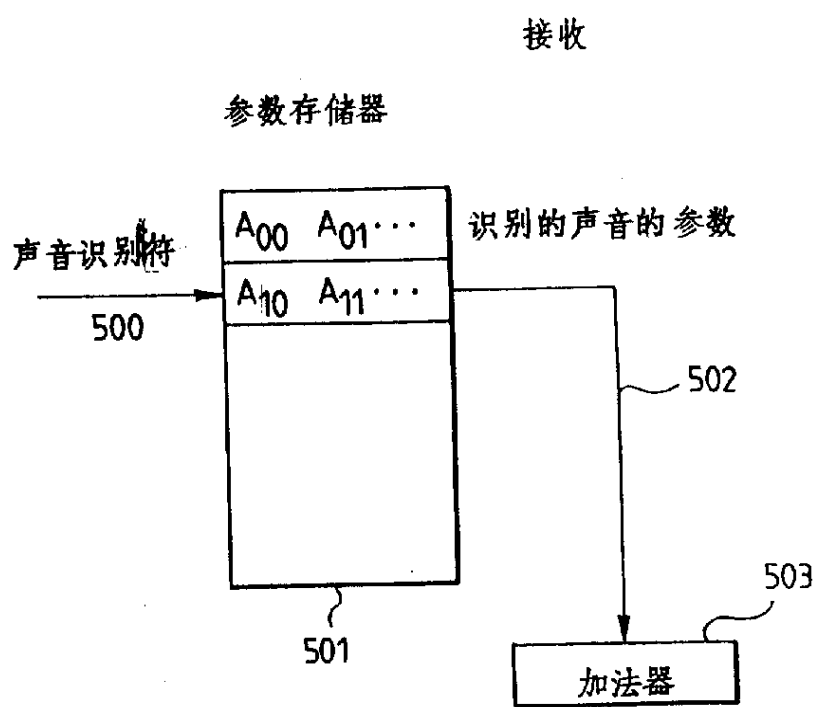


图 5b

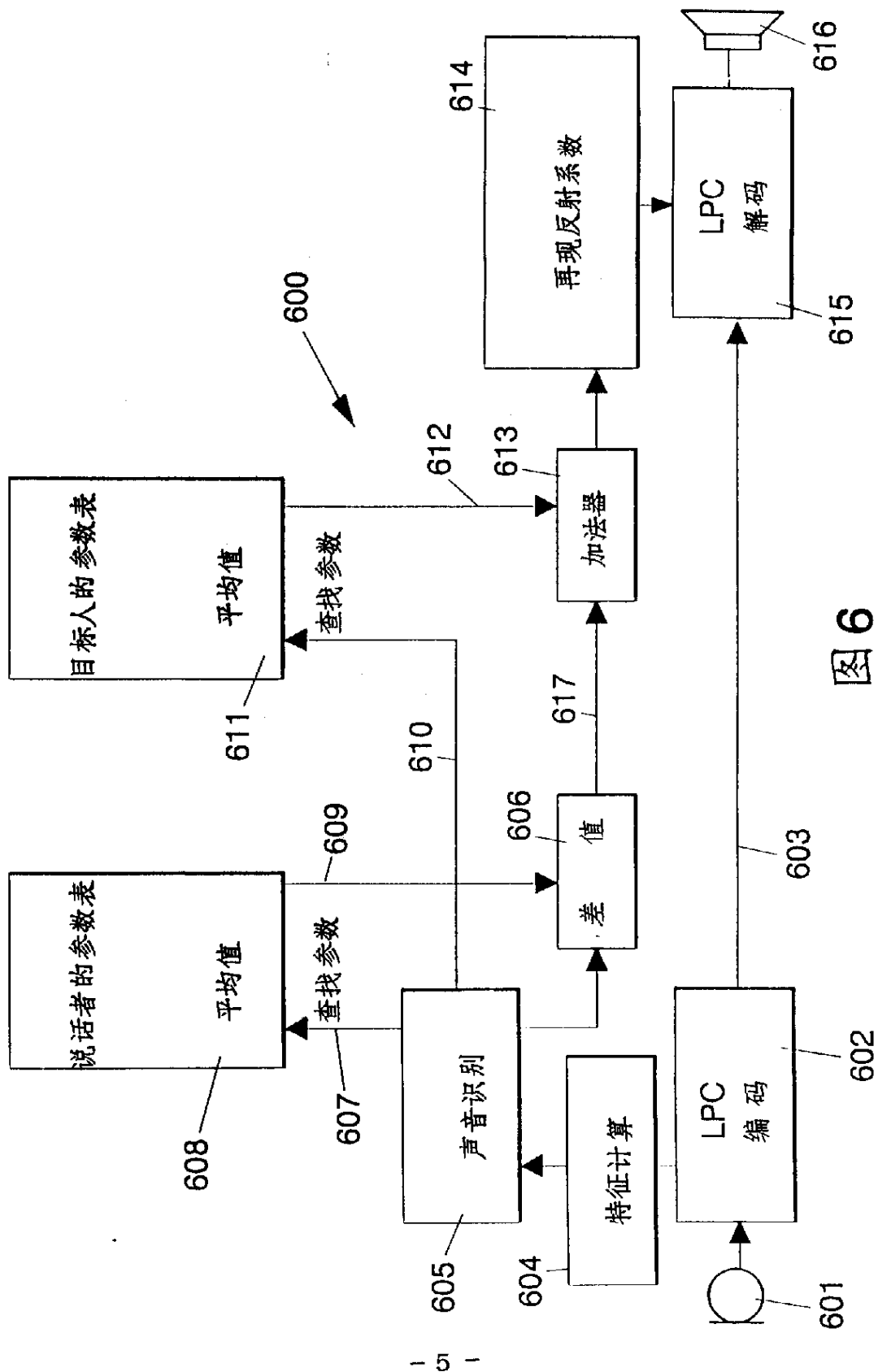


图 6