[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 21/04 (2006.01)
G10L 19/02 (2006.01)

[12] 发明专利说明书

专利号 ZL 200410056699.X

[45] 授权公告日 2007 年 3 月 28 日

[11] 授权公告号 CN 1307614C

[22] 申请日 1996.10.26
[21] 申请号 200410056699.X
分案原申请号 96121905.X
[30] 优先权
[32] 1995.10.26 [33] JP [31] 279410/95
[32] 1995.10.27 [33] JP [31] 280672/95
[32] 1996.10.11 [33] JP [31] 270337/96
[73] 专利权人 索尼公司
地址 日本东京都
[72] 发明人 饭岛和幸 西口正之 松本淳 大森士郎
[56] 参考文献
US4435832A 1984.3.6
EP0279451A 1988.8.24

HARMONCI AND NOISE CODING OF OLPC RESIDUALSWITH CLASSIFIED VECTOR QUANTIZATION NASHICUCHI M ET AL, PROCEEDINGS OF THE INTERNATIONAL CONFERNCE ON ACOUSTICS, SPEECHAND SIG. 1995

SHAPE INVARIANT TIME – SCALE AND PITCHMODIFICATION OF APEECH QUATIERI T F ET AL, IEEE TRANSACTIONS ON SIGNAL PROCESSING, Vol. 40 1992

审查员 刘红梅
[74] 专利代理机构 中国专利代理(香港)有限公司
代理人 张志醒

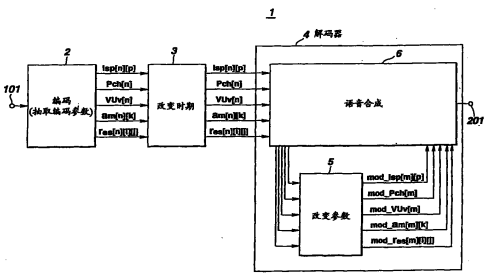
权利要求书 1 页 说明书 24 页 附图 14 页

[54] 发明名称

合成语音的方法和装置

[57] 摘要

一种以控制的速度复制语音的方法，其编码单元鉴别输入语音信号是浊音还是清音信号。依鉴别结果，编码单元对判定为浊音的信号部分执行正弦合成和编码，对判定为清音的信号部分由对最优矢量的检索使用综合分析方法执行矢量量化、解码单元产生与对应于预先设定祯的不同时间点相关的修改的已编码参数。语音合成单元合成浊音部分和清音部分。用编码的数据输出单元输出编码的位流或编码的数据。波形合成单元合成语音波形。



1. 一种语音合成方法，包括:

按照一个预先设定的用于输出谐波的振幅数据的规则而合成常规语音的常规语音合成步骤，

变换输入数据的谐波的基本频率和在一个预先设定的频带内的振幅数目的数据数目变换步骤，

为修改合成语音的音调，对每一个输入谐波中的谱分量的大小的数据进行插值的步骤。

2. 按照权利要求 1 所述的语音合成方法，其中，使用一个频带限制型过采样滤波器执行所述插值。

3. 一种语音合成装置，包括:

按照一个用于输出谐波的振幅数据的文本而合成常规语音的常规语音合成装置，变换输入数据的谐波的基本频率和在一个预先设定的频带内的振幅数目的数据数目变换装置，

为修改合成语音的音调，对每一个谐波中的谱分量的大小的数据进行插值的装置。

4. 按照权利要求 3 所述的语音合成装置，其中，使用一个频带限制型过采样滤波器执行所述插值。

合成语音的方法和装置

本发明申请是申请日为1996年10月26日、申请号为96121905.X的名为“复制语音信号的方法和装置，解码语音的方法和装置，合成语音的方法和装置以及便携式无线电终端设备”的专利申请的一个分案申请。

技术领域

本发明涉及的是以一受控速度复制语音信号的方法和装置，解码语音信号的方法和装置和合成语音信号的方法和装置，其中音调变换可以由简化的结构实现。本发明还涉及发射和接收音调变换的语音信号的便携式无线电终端设备。

背景技术

迄今为止已知各种编码音频信号（包括语音和声学信号）的编码方法，它们使用这些信号在时域和在频域的统计特性和人耳的心理声学特征压缩信号。这些编码方法可以粗略地分为时域编码、频域编码和分析／合成编码。

语音信号高效编码的例子包括正弦分析编码，例如谐波编码、多频带激励（MBE）编码、子频带编码（SBC）、线性预测编码（LPC）、离散余弦变换（DCT）、修改DCT（MDCT）和快速富里叶变换（FFT）。

同时，通过时轴处理的高效语音编码方法，像典型的码激励线性预测（CELP）编码，在快速时轴变换（修改）上遇到困难，因为在解码操作后需要执行大量的处理操作。此外，由于速度控制是在解码后在时域执行，因此该方法不能用于位速率变换。

另一方面，如果打算解码用上述编码方法编码的语音信号，常常希望只改变语音的音调而不改变其音素。然而，使用通常的语音解码方法，解码的语音必须使用音调控制变换音调，使结构变得复杂，同时增加成本。

发明内容

因此，本发明的一个目的是提供复制语音信号的一种方法和装置，其中可以使速度在宽的范围内控制到一个希望的速率得到高的声音质量而不改变音素或音调。

本发明的另一目的是提供解码语音信号的方法和装置和合成语音的方法和装置，其中可以使用简化的结构实现音调变换或音调控制。

本发明的再一目的是提供发射和接收语音信号的便携式无线电终端设备，其中可以使用一个简化的结构发射和接收音调变换的或音调受控的语音信号。

使用按照本发明的语音信号复制方法，输入语音信号在时轴上根据预先设定的编码单元切分来产生编码的参数，将其插值，为希望的时间点产生修改的已编码参数，并根据这些修改的已编码参数复制语音信号。

使用按照本发明的语音信号复制装置，输入语音信号在时轴上根据预先设定的编码单元切分来产生编码的参数，将其插值，为希望的时间点产生修改的已编码参数，然后根据这些修改的已编码参数复制语音信号。

使用该语音信号复制方法，用不同于编码的块长，使用根据预先设定的块作为单元在时轴切分输入语音信号而得到的编码的参数，并根据编码块编码切分的语音信号复制语音。

使用根据本发明的语音解码方法和装置，变换在输入编码语音数据的谐波的一个预先设定的频带内的基本频率和数目，并插值说明在每一输入谐波中谱分量振幅的数据数目来修改音调。

编码时使用大小变换修改音调频率，其中把谐波数设定为一个预先设定值。在这种情况下，语音压缩解码器可以同时用作文本语音合成的语音合成器。对于日常语音发声，通过压缩和扩展得到清晰的回放语音，而对于特殊的语音合成，使用文本合成或按照预定的规则合成来构成高效语音输出系统。

使用根据本发明的语音信号复制方法和装置，输入语音信号在时轴上根据预先设定的编码单元切分，并根据该编码单元编码以便寻找已编码参

数，然后将其插值，为希望的时间点寻找修改的已编码参数。然后根据修改的已编码参数复制该语音信号，从而在宽的范围内容易地实现速度控制，不改变音素或音调而具有高品质。

使用根据本发明的语音信号复制方法和装置，用不同于编码的块长，使用根据预先设定的块作为单元在时轴切分输入语音信号而得到的已编码参数，和根据该编码块编码切分的语音信号来复制语音。结果是，在宽的范围内容易地实现速度控制，不改变音素或音调而具有高品质。

使用根据本发明的语音解码方法和装置，变换在输入编码语音数据的谐波中一个预先设定的频带内的基本频率和数目，并插值说明在每一输入谐波中谱分量振幅的数据数来修改音调。结果是，可以使用一个简化的结构改变音调为一个希望的值。

在这种情况下，语音压缩解码器可以同时用作文本语音合成的语音合成器。对于日常语音发声，通过压缩和扩展得到清晰的回放语音，而对于特殊的语音合成，使用文本合成或按照预先规定的规则合成来构成高效语音输出系统。

使用便携无线电终端设备，可以用一个简化的结构发射和接收音调变换到音调受控的语音信号。

附图说明

图 1 是表示语音信号复制方法和实现按照本发明的语音信号复制方法的一个语音信号复制装置的基本结构的框图；

图 2 是表示图 1 所示的语音信号复制装置的编码单元的原理框图；

图 3 是表示编码单元的详细结构的框图；

图 4 是表示图 1 所示的语音信号复制装置的解码单元的原理框图；

图 5 是表示该解码单元的详细结构的框图；

图 6 是说明为计算解码单元修改的编码参数的单元的操作流程图；

图 7 原理说明由修改的编码参数计算单元在时轴上得到的修改的编码参数；

图 8 是说明由修改的编码参数计算单元执行的详细的插值操作的流程

图;

图 9A 到 9D 说明插值操作;

图 10A 到 10C 说明由修改的编码参数计算单元执行的典型操作;

图 11A 到 11C 说明由修改的编码参数计算单元执行的其它典型操作;

图 12 说明在使帧长变化而由解码单元快速控制速度的情况下的一个操作;

图 13 说明在使帧长变化而由解码单元慢速控制速度的情况下的一个操作;

图 14 是表示解码单元另一详细结构的框图;

图 15 是表示语音合成设备应用实例的框图;

图 16 是表示文本语音合成装置应用实例的框图;

图 17 是表示应用编码单元的一个便携式终端的发射器结构的框图;

图 18 是表示应用编码单元的一个便携式终端的接收器结构的框图。

具体实施方式

参考附图，下面叙述根据本发明的最佳实施例的语音信号复制方法和装置。本实施例是关于根据编码参数复制语音信号的语音信号复制装置 1，这些编码参数是在时轴上根据预先规定的帧数作为编码单元切分输入语音信号，并将该切分的输入语音信号编码而得到的，如图 1 所示。

语音信号复制装置 1 包括根据作为单元的帧而编码在输入端子 101 进入的语音信号的编码单元 2，它输出例如线性预测编码（LPC）参数、线谱对（LSP）参数、音调、浊音（V）/清音（UV）或谱振幅 Am 等已编码参数，并且包括通过时轴压缩为修改编码参数的输出时期的时期修改单元 3。语音信号复制装置还包括解码单元 4，其插值在由时期修改单元 3 为希望的时间点寻找修改的已编码参数而修改的时期中输出的已编码参数，并根据修改的已编码参数合成语音信号以便在输出端子 201 输出合成的语音信号。

参考图 2 和图 3 解释编码单元 2。编码单元 2 根据鉴别结果判定输入语音信号是浊音信号还是清音信号，并对判定为浊音的信号部分执行正弦合成编码，而对判定为清音的信号部分通过使用综合分析方法进行的最优

矢量的闭环检索执行矢量量化。也就是说，编码单元 2 包括第一编码单元 110，它为寻找输入语音信号的短期预测残数，例如线性预测编码（LPC）残数，执行正弦分析编码，例如谐波编码，编码单元 2 还包括第二编码单元 120，其通过传输输入语音信号的相位分量执行波形编码。第一编码单元 110 和第二编码单元 120 分别用于编码浊音（V）部分和清音（UV）部分。

在图 2 的实施例中，供给输入端子 101 的语音信号送往第一编码单元 110 的逆 LPC 滤波器 111 和 LPC 分析量化单元 113。从 LPC 分析／量化单元 113 或所谓的 α 参数得到的 LPC 系数送往逆 LPC 滤波器 111 以通过该逆 LPC 滤波器 111 取出输入语音信号的线性预测残数（LPC 残数）。从 LPC 分析／量化单元 113 取出线性谱对的量化输出，其在后面叙述，并送往输出端子 102。从逆 LPC 滤波器 111 来的 LPC 残数送往正弦分析编码单元 114，正弦分析编码单元 114 由浊音（V）／清音（UV）鉴别单元 115 执行音调检测、谱包络线振幅计算和 V／UV。从正弦分析编码单元 114 来的谱包络线振幅数据送往矢量量化单元 116。从矢量量化单元 116 来的码表指数作为谱包络线的矢量量化输出经由开关 117 送往输出端子 103，而正弦分析编码单元 114 的输出通过开关 118 送往输出端子 104。从浊音/清音鉴别单元 115 来的浊音／清音鉴别输出送往输出端子 105 和开关 117 和 118 作为开关控制信号。对于浊音（V）信号，选择指数和音调以便在输出端子 103、104 取出。对在矢量量化器 116 的矢量量化，将一个用于对频率轴上的有效频带块的振幅数据进行插值的适当数目的哑数据附在该块的尾端和前端，该哑数据为从在该块中最后一个振幅数据直到块中第一振幅数据的哑数据，或者为延伸块中最后数据和第一数据的哑数据，以增大数据数目到 N_F。然后通过频带限制型 Os 元组采样，例如 8 元组过采样，寻找振幅数据的 Os 元组数。振幅数据的 Os 元组数（$(m_{MX}+1)\times$ 数据的 Os 数）通过线性插值进一步扩展到一个更大的数 N_M 数，例如 21048。该数据通过多中取一而变换为预先规定的数 M（例如 44），然后在该预先规定的数目的数据上执行矢量量化。

在本实施例中，第二编码单元 120 有一个码激励的线性预测（CELP）编码配置，并且该编码单元通过使用综合分析方法的闭环检索在时域波形上执行矢量量化。具体地说，噪声码表 121 的输出由加权合成滤波器 122

合成产生一个加权合成的语音，送往减法器123，这里寻找加权合成语音和供给输入端子101、继而由感性加权滤波器125处理的语音之间的误差。距离计算电路124计算距离，并在噪声码表121中检索使误差最小的矢量。这一CELP编码用于编码上述清音部分，从噪声码表121来的作为UV数据的码表指数通过开关127在输出端子107取出，开关107在从浊音/清音鉴别单元115的浊音／清音鉴别结果指示一个清音（UV）声音时打开。

参考图3，现在解释图1所示的语音信号编码器的更详细的结构。在图3中，相似于图1中所示的元部件用同样的参考数字表示。

在图3所示的语音信号编码器2中，供给输入端子101的语音信号由高通滤波器109滤波，以滤除不需要的范围的信号，然后供给LPC分析／量化单元113的LPC分析电路132和逆LPC滤波器111。LPC分析／量化单元113的LPC分析电路132应用一个哈明（Hamming）窗，其输入信号波形的长度以256个样本为一块，并通过自相关方法寻找线性预测系数，亦即所谓的α参数。作为数据输出单元的帧区间设定为大约160个样本。如果采样频率fs例如是8kHz的话，则一帧的区间是20 毫秒或 160个样本。

从LPC分析电路132得到的α参数送往α－LSP变换电路133变换为线性谱对（LSP）参数。它把α参数作为直接型滤波器系数变换为例如10，也就是说，5对LSP参数。这种变换例如可以用牛顿－拉普森（Newton－Rhapson）方法实现。把α参数变换成LSP参数的理由是LSP参数在插值特征上高于α参数。

从α－LSP变换电路133来的LSP参数由LSP量化器134矩阵量化或矢量量化。有可能在矢量量化或收集多帧到一起之前取帧对帧的差别执行矩阵量化。在现在的例子中，每20毫秒计算的LSP参数以每帧20毫秒矢量量化。

在端子102取出量化器134的量化输出，亦即LSP量化的索引数据到解码单元103，而量化了的LSP矢量送往一个LSP插值电路136。

LSP插值电路136插值每20毫秒或40毫秒量化的LSP矢量，以提供一个8元组速率。也就是说，LSP矢量每2.5毫秒更新。理由是，如果残数波形通过谐波编码／解码方法用分析／合成处理，则合成波形的包络线描

述一个极度平静的波形，以致，如果LPC系数每20毫秒突然改变，则可能产生一个外部噪声。也就是说，如果LPC系数每2.5毫秒逐渐改变有可能阻止产生这样的外部噪声。

对于使用每过2.5毫秒产生的插值的LSP矢量的输入语音的逆滤波，LSP参数由一个LSP到α的变换电路137变换为作为例如10阶直接型滤波器的系数的α参数。LSP到α的变换电路137的输出送往LPC逆滤波器电路111，然后它执行逆滤波，以产生一个使用α参数每2.5毫秒更新的平滑输出。逆LPC滤波器111的输出送往正弦分析编码单元114，例如一个谐波编码电路的正交变换电路145，例如一个DCT电路。

从LPC分析／量化单元113的LPC分析电路132来的α参数送往一个感性加权滤波器计算电路139，这里寻找感性加权数据。这些加权数据送往第二编码单元120的感性加权矢量量化器116，感性加权滤波器125和感性加权的合成滤波器122。

谐波编码电路的正弦分析编码单元114用谐波编码方法分析逆LPC滤波器111的输出。亦即，执行音调检测，代表谐波的振幅Am的计算和浊音（V）／清音（UV）辨别，以及用大小变换维持由音调改变的代表谐波的振幅（Am）后包络线的数目。

在图3所示的正弦分析编码电路114的例子中，使用平常的谐波编码。特别在多频带激励（MBE）编码中，提取模型时假定浊音部分和清音部分在同一时间点（在同一块或帧）出现在频域或频带中。在其它谐波编码技术中，唯一判别在一块或一帧中的语音是否是浊音或清音。在下面的叙述中，如果整个频带是UV，则判定一个给定帧为UV，只要涉及到MBE编码的话。

图3的正弦分析编码单元141的开环音调检索单元141和过零计数器142由分别由从输入端子101来的输入语音信号和从高通滤波器（HPF）109来的信号供给。正弦分析编码单元114的正交变换电路145由从逆LPC滤波器111来的LPC残数或线性预测残数供给。开环音调检索单元141取输入信号的LPC残数通过开环检索执行相对粗音调检索。提取的粗音调数据由闭环检索送往细音调检索单元146，其在后面叙述。从开环音调检索单元141，通过正规化LPC残数的自相关的最大值与粗音调数据得到的正规化自

相关的最大值 r（p）连同粗音调数据取出，以便送往浊音／清音鉴别单元 115。

正交变换电路 145 执行正交变换，例如离散富里叶变换（DFT），为变换时间轴上的 LPC 残数为频率轴上的谱振幅数据。正交变换电路 145 的输出送往细音调检索单元 146 和谱评测单元 148 来评估谱振幅或包络线。

细音调检索单元 146 用由开环音调检索单元 141 提取的相对粗音调数据和由正交变换电路 145 得到的频域数据供给。细音调检索单元 146 围绕粗音调数据为中心以 0.2 到 0.5 的速率以 ± 几个样本摆动音调数据，以便最终达到具有最优十进数点（浮点）的细音调数据的值。使用综合分析方法作为选择音调的细检索技术使得功率谱接近于原来信号的功率谱。从闭环细音调检索单元 146 来的音调数据通过开关 118 送往输出端子 104。

在谱评测单元 148 中，根据谱振幅和作为 LPC 残数的正交变换输出评估每一谐波的振幅和作为这些谐波的和的谱包络线并送往细音调检索单元 146、浊音／清音鉴别单元 115 和感性加权矢量量化单元 116。

浊音／清音鉴别单元 115 根据正交变换电路 145 的输出、从细音调检索单元 146 来的最优音调、从谱评测单元 148 来的谱振幅数据、从开环音调检索单元 141 来的正规化自相关的最大值 r（p）和从过零计数器来的过零计数值鉴别一帧的浊音／清音。另外，对于 MBE，也可以利用基于频带的浊音／清音鉴别的边界位置为浊音／清音鉴别的条件。浊音／清音鉴别单元 115 的鉴别输出在输出端子 105 取出。

把一些数据变化单元（执行一种采样速率变换的一个单元）供给谱评测单元 148 的输出单元或矢量量化单元 116 的输入单元。考虑到在频率轴上分解的频带数和数据数目与音调不同这一事实，使用数据数目变换单元设定一个包络线的振幅数据。也就是说，如果有效频带到 3400 千赫，则根据音调可以分解该有效频带为 8 到 63 个频带。从频带到频带得到的振幅数据 $|Am|$ 的 $m_{MX}+1$ 的数目在 8 到 63 的范围内变化。于是数据数目变换单元 119 变换变化的数目 $m_{MX}+1$ 的振幅数据为数据的一个预先设定数 M，例如 44 个数据

从数据数目变换单元供给到谱评测单元 148 的输出单元或矢量量化单

元 116 的输入的比如 44 个的预先设定数 M 个振幅数据或比如 44 个的包络线数据根据预先设定数收集为单元，并由感性加权滤波器计算单元 139 矢量量化。从矢量量化器 116 来的包络线指数在输出端子 103 经由开关 117 取出。在对加权的矢量量化之前，建议为由预先设定数目的数据组成的矢量使用合适的泄漏系数取帧间差。下面说明第二编码单元 120。第二编码单元 120 具有码激励线性预测（CELP）编码结构，并特别用于输入语音信号的清音部分编码。在为清音部分的 CELP 编码结构中，相应于清音部分的 LPC 残数的噪声输出为噪声码表亦即所谓的随机码表 121 的代表输出通过增益电路 126 送往感性加权的滤波器 122。从输入端子 101 经由高通滤波器（HPF）109 提供并由感性加权滤波器 125 感性加权的语音信号供给减法器 123，在这里得到感性加权的语音信号与从合成滤波器 122 来的信号之间的差或误差。这一误差供给距离计算电路 124 以找出距离，并由噪声码表 121 检索使误差最小的代表值矢量。上述即是使用闭环检索接着使用综合分析方法的时域波形矢量量化的总结。

作为从第二编码器 120 来的使用 CELP 编码结构的清音（UV）部分的数据，从噪声码表 121 来的码表的形状指数、从增益电路 126 码表来的增益指数被取出。作为从噪声码表 121 来的 UV 数据的形状指数通过开关 127s 送往输出端子 107s，而作为增益电路 126 的 UV 数据的增益指数通过开关 127g 送往输出端子 107g。

根据从 V／UV 鉴别单元 115 得到的 V／UV 鉴别结果打开或关闭这些开关 127s、127g 和开关 117、118。具体地说，要传输的语音信号帧的 V／UV 鉴别结果指示为浊音（V）时，打开开关 117、118；而如果要传输的语音信号帧为清音（UV）时，打开开关 127s、127g。

由编码单元 2 输出的编码参数供给时期修改单元 3。时期修改单元 3 通过时间轴的压缩／扩展修改输出时期。在由时期修改单元 3 修改的时期中输出的编码的参数送往解码单元 4。

解码单元 4 包括一个为插值编码参数的参数修改单元 5，其由时期修改单元 3 沿时间轴使用示例的方法压缩，产生与预先设定的帧的时间点关联的修改过的编码参数，还包括一个为根据修改的编码参数合成浊音信号

部分和清音信号部分的语音合成单元6。

参考图4和图5说明解码单元4。在图4中，码表指数数据，作为从时期修改单元3来的线性谱对（LSPs）的量化输出数据供给输入端子202。时期修改单元3的输出，也就是说指数数据，作为量化包络线数据、音调数据和V／UV鉴别输出数据分别供给输入端子203、204和205。从时期修改单元3来的指数数据作为清音部分的数据还供给输入端子207。

从输入端子203来的指数数据作为量化了的包络线输出送往逆矢量量化器212矢量量化以寻找LPC残数的谱包络线。在送往浊音合成单元211之前，LPC残数的谱包络线在靠近图4中的用箭头P1指示的点临时取出，由参数处理器5进行参数修改，其在后面说明。然后指数数据送往浊音合成单元211。

浊音合成单元211使用正弦合成方法合成浊音信号部分的LPC残数。音调和V／UV鉴别数据分别进入输入端子204、205，并在图4中的点P2和P3处临时取出，由参数修改单元5修改参数，其类似地供给浊音合成单元211。从浊音合成单元211来的浊音的参数送往LPC合成滤波器214。

从输入端子207来的UV数据的指数数据送往清音合成单元220。UV数据的指数数据由浊音合成单元220参考噪声码表变为清音部分的LPC残数。UV数据的指数数据从清音合成单元220临时取出，由在图4中的点P4指示的参数修改单元5修改参数。这样用参数修改处理的LPC残数也送往LPC合成滤波器214。

LPC合成滤波器214在浊音信号部分的LPC残数上和在清音信号部分的LPC残数上执行独立的合成。可选的另外方案为可以在浊音信号部分的LPC残数和清音信号部分的LPC残数加在一起之上执行LPC合成。

从输入端子202来的LSP指数数据送往LPC参数再生单元213。虽然LPC的α参数最终由LPC参数再生单元213产生，但是LSP的逆矢量量化的数据由箭头P5指示的参数修改单元5部分地取出进行参数修改。

用参数修改如此处理的去量化数据返回到这一LPC参数再生单元213进行LPC插值。然后去量化数据转变为LPC的α参数供给LPC合成滤波器214。由LPC合成通过LPC合成滤波器214得到的语音信号在输出端子201

取出。图 4 中所示的语音合成单元 6 接收修改的已编码参数，由参数修改单元 5 如上所述计算，而输出合成的语音。语音合成单元的实际结构示于图 5，其中相应于图 4 所示的元部件由同样的数字表示。

参考图 5，进入输入端子 202 的 LSP 指数数据送往 LPC 参数再生单元 213 的 LSPs 的逆矢量量化器 231，以便逆矢量量化为 LSPs（线性谱对），其供给参数修改单元 5。

从输入端子来的谱包络线 Am 的矢量量化指数数据送往逆矢量量化器 212 进行逆矢量量化并转变为谱包络线数据送往参数修改单元 5。

从输入端子 204、205 来的音调数据和浊音／清音鉴别数据也送往参数修改单元 5。

从图 3 的输出端子 107s 和 107g 通过时期修改单元 3 供给图 5 的输入端子 207s 和 207g 作为 UV 数据的形状指数数据和增益指数数据。然后形状指数数据和增益指数数据供给清音合成单元 220 从端子 207s 来的形状指数数据和从端子 207g 来的增益指数数据分别供给清音合成单元 220 的噪声码表 221 和增益电路 222。从噪声码表 221 读出的代表值输出是相应于清音的 LPC 残数的噪声信号分量，并成为增益电路 222 的预先设定增益的振幅。结果信号供给参数修改单元 5。

参数修改单元 5 插值由编码单元 2 输出并使其输出时期由时期修改单元 3 修改的已编码参数，以产生修改的已编码参数，供给语音合成单元 6。参数修改单元 3 修改已编码参数的速度。这消除了解码器输出后的速度修改操作，并允许语音信号再生设备 1 以与类似算法不同的固定速率处理。

参考图 6 和图 8 的流图，解释时期修改单元 3 和参数修改单元 5。

在图 6 的步骤 S1 时期修改单元 3 接收编码参数，例如 LSPs、音调、浊音／清音（V／UV）、谱包络线 Am 和 LPC 残数。LSPs、音调、（V／UV）、Am 和 LPC 残数分别表示为 Lsp[n][p]、Peh[n]、VUv[n]/a_m[n][k]和 res[n][i][j]。

最终由参数修改单元 5 计算的修改的编码参数表示为 mod_lsp[m][p]、mod_Pch[m]、mod_UVv[m]、mod_a_m[m][k]和 mod_r_{es}[m][i][j]，其中 k 和 p 分别表示谐波数和 LSP 的阶数。每一个 n 和 m 分别表示相应于

时间轴变换前和后的时域指数数据的帧数。同时，每一个 n 和 m 表示具有间隔为 20 毫秒的帧的指数，而 i 和 j 分别表示子帧数和采样。

然后时期修改单元 3 分别设定代表原始时间区间的帧数为 N1，代表修改以后的时间区间的帧数为 N2，如步骤 S2 所示。然后时期修改单元进行语音 N1 到语音 N2 的时间轴压缩，如步骤 S3 所示。也就是说，在时期修改单元 3 的时间轴压缩比率是 spd = N2 / N1，限制条件为 0 ≤n < N1 和 0≤m < N2。

然后参数修改单元 5 设定相应于帧数，依次相应于时间轴修改后的时间轴的指数 m 为 2。

然后参数修改单元 5 找两个帧 fr_0 和 fr_1 和在两帧 fr_0 和 fr_1 之间的左差和右差以及比率 m / spd。

如果参数 lsp、P_{ch}、UVv、a_m 和 r_{es} 表示为 *，则 $^*[m]$ 可以由生成公式

$$mod\_^*[m] = {}^*[m/spd]$$

其中 0≤m < N。然而，由于 m / spd 不是一个整数，所以在 m / spd 处的修改的已编码参数从下列两帧

$$fr_0 = [m/spd]$$

和

$$fr_1 = fr_0 + 1$$

的插值产生。

在帧 fr_0，即 m/spd 和帧 fr_1 之间，图 7 所示的关系式，即

$$左 = m/spd - fr_0$$
$$右 = fr_1 - m/spd$$

成立。

对在图 7 中的 m／spd 的已编码参数，亦即修改的已编码参数可以由内插找到，如步骤 S6 所示。

可以简单地通过线性插值找到修改的编码参数:

$$mod\_^{*}[m] = {}^{*}[fr_0] \times 右 + {}^{*}[fr_1] \times 左$$

然而，对于在两帧 fr_0 和 fr_1 之间的插值，如果两帧不同于 V／UV，也就是说，如果其中之一是 V，而另一个是 UV，则不能使用上面的一般的公式。因此，参数修改单元 5 改变该方法，根据两帧 fr_0 和 fr_1 的浊音（V）和清音（UV）特征来寻找编码参数，其在图 8 的步骤 11 等指出。

首先，如步骤 11 所示，确定两帧 fr_0 和 fr_1 的浊音（V）和清音（UV）特征。如果发现这两帧 fr_0 和 fr_1 都是浊音（V）处理转移到步骤 S12，这里所有的参数都线性插值，并由下式表示:

$$mod\_Pch[m] = Pch[fr_0] \times 右 + Pch[fr_1] \times 左$$

$$mod\_a_m[m][k] = a_m[fr_0][k] \times 右 + a_m[fr_1][k] \times 左$$

式中 $0 \leqslant k < 1$，L 是谐波的最大可能数。对于 $a_m[fr_1][k]$，0 插在无谐波的位置。如果谐波数在帧 fr_0 和 fr_1 之间不同，则凡在空的位置均插入 0。另一方案为，在通过解码器侧一些数据转换器之前，可能使用一个固定的数，例如 $0 \leqslant k < L$，这里 $L = 43$。

$$mod\_lsp[m][p] = lsp[fr_0][p] \times 右 + lsp[fr_1][p] \times 左$$

式中 $0 \leqslant p < P$，其中 P 表示 LSP 的阶数，通常为 10。

$$mod\_VUv[m] = 1$$

在 V／UV 鉴别中，1 和 0 分别表示浊音（V）和清音（UV）。

如果在步骤 S11，判定两帧 fr_0 和 fr_1 均不是浊音（V），则在步骤 S13

判定两帧 fr_0 和 fr_1 是否都是清音（UV）。如果在步骤 S13 的判定结果为是，也就是说，如果两帧都是清音，则插值单元 5 以 m／spd 作为中心和以 pch 作为最大值在 res 的前和后切分 80 个样本，如步骤 S14 所示。

结果是，如果在步骤 S14 左＜右，则以 m／spd 为中心在 res 的前和后切分 80 个样本，并插入 res 的模中，如图 9A 所示。也就是说，

对于(j＝0; j<FRM×(1/2－m/spd＋fr0); j++) {mod res[m][o][j]
＝res[fr_0][o][j＋(m/spd－fr_0)×FRM]; }

对于(j＝FRM×(1/2－m/spd＋fr0); j<FRM/2; j++) {mod res[m][o][j]
＝res[m][o][j]＝res[fr_0][1][j－FRM×(1/2－m/spd＋fr_0)]; };

对于(j＝0; j<FRM×(1/2－m/spd＋fr_0); j++) {mod res[m][1][j]
＝res[fr_0][1][j＋m/spd－fr_0)×FRM]; };

对于(j＝FRM×(1/2－m/spd＋fr_0); j＝FRM/2; j++) {mod res[m][1][j]
＝res[fr_0][o][j＋FRM×(1/2－m/spd＋fr_0)]; };

式中 FRM 例如取 160。

另一方面，如果在步骤 S14，左≥右，则插值单元 5 以 m／spd 为中心在 res 的前和后切分 80 个样本，以产生 mod_res，如图 9B 所示。

如果在步骤 S13 条件不满足，处理转移到步骤 S15，这里判定帧 fr_0 是否是浊音（V）和帧 fr_1 是否是清音（UV），如果判定的结果为是，也就是说，如果帧 fr_0 是浊音（V）和帧 fr_1 是清音（UV），处理转移到步骤 S16。如果判定结果为否，也就是说，如果帧 fr_0 是清音(UV)，帧 fr1 是浊音(V)，处理转移到步骤 S17。

在步骤 S15 等的向下处理中，两帧 fr_0 和 fr_1 不同于浊音／清音，也就是说浊音（V）和清音（UV）。这考虑了下述事实，如果在不同于 V／UV 的两帧之间插值参数的话，则插值结果没有意义。

在步骤 S16，比较左尺寸（＝m／spd－fr_0）和右尺寸（＝fr_1－m/spd）以判定帧 fr_0 是否接近 m／spd。

如果帧 fr_0 接近 m／spd，使用帧 fr_0 的参数设定修改的已编码参数，使得

$mod\_Pch[m] = Pch[fr_0]$

$mod\_a_m[m][k] = a_m[fr_0][k]$，其中 $0 \leqslant k \leqslant L$;

$mod\_lsp[m][p] = lsp[fr_0][p]$，其中 $0 \leqslant p \leqslant I$；和

$mod\_UVv[m] = 1$

如步骤 S18 所示。

如果在步骤 S16 的判定结果为否，即左≥右，使得帧 fr_1 更接近，则处理转移到步骤 S19 使音调最大。同时，直接使用帧 fr_1 的 r_{es} 如图 9C 所示，并设定为 mod_r_{es}。亦即 $mod\_r_{es}[m][i][j] = r_{es}fr_1[i][j]$。理由是，对于浊音帧 fr_0 不传输 LPC 残数 r_{es}。

在步骤 S17，根据在步骤 S15 给出的判定，即两帧 fr_0 和 fr_1 分别为清音（UV）和浊音（V），给出类似于步骤 S16 的判定。也就是说，比较左大小（$=m/spd-fr_0$）和右大小（$=fr_1-m/spd$）以便判定 fr_0 是否接近 m／spd。

如果帧 fr_0 接近 m／spd，处理转移到步骤 S18 使音调最大。同时，直接使用帧 fr_0 的 r_{es} 并设定为模 r_{es}。也就是说 $mod\_r_{es}[m][i][j] = r_{es}fr_0[i][j]$ 理由是，对于浊音帧 fr_1，不传输 LPC 残数 r_{es}。

如果在步骤 S17 的判定结果为否，左≥右，因此帧 fr0 接近 m／spd，处理前进到步骤 S21，并使用帧 fr_1 的参数设定修改的已编码参数，使得

$mod\_P_{ch}[m] = P_{ch}[fr_1]$

$mod\_a_m[m][k] = a_m[fr_1][n]$，其中 $0 \leqslant k \leqslant L$;

$mod\_lsp[m][p] = lsp[fr_1][p]$，其中 $0 \leqslant p \leqslant I$;

$mod\_Vuv[m] = 1$

以这种方式，插值单元 5 根据两帧 fr_0 和 fr_1 的浊音／清音特征在图 6

的步骤S6（更详细地示于图8）提供不同的插值操作。在步骤S6的插值结束后，处理转移到步骤S7使m增量。步骤S5和S6的操作重复，直到m的值等于N2。

参考图10集中解释时期修改单元3和参数修改单元5的操作。参考图10，由编码单元2每20毫秒提取编码参数的时期通过时期修改单元5执行的时间轴压缩修改为15毫秒，如图10A所示。在响应两帧fr_0和fr_1的V／UV状态执行的插值操作中，参数修改单元每20毫秒计算修改的已编码参数，如图10C所示。

时期修改单元3和参数修改单元5的操作顺序可以倒过来，也就是说首先对在图11A中所示的已编码参数进行如图11B中的插值，接着如图11C进行压缩以计算修改的已编码参数。

返回到图5，LSP上的数据的修改的已编码参数lsp[m][p]由参数计算单元5计算，送往LSP插值电路232v、232u进行LSP插值。结果数据由LSP到α变换电路234v、234u变换为用于线性预测编码（LPC）的α参数，送往LPC合成滤波器214。LSP插值插电路232v和LSP到α变换电路234v用于浊音（V）信号部分，而LSP插值电路234u和LSP到α变换电路234u用于清音（UV）信号部分。LPC合成滤波器214由一个用于浊音部分的LPC合成滤波器236和一个用于清音部分的LPC合成滤波器237组成。也就是说，LPC系数的插值对于浊音部分和清音部分独立地进行，以防止在具有完全不同特征的插值时在从浊音部分到清音部分的过渡区域或在从浊音部分到清音部分的过渡区域可能产生的不良影响。

由参数修改单元5找到的谱包络线数据上的修改的已编码参数$mod_a_m[m][k]$，送往浊音合成单元211的正弦合成电路215。由参数修改单元5计算的音调mod_pch[m]上的修改的已编码参数和在V／UV判定数据上的修改的已编码参数mod_UVv[m]也供给浊音合成单元211。相应于图3的LPC逆滤波器111的输出的LPC残数数据从正弦合成电路215取出送往加法器218。

由参数修改单元5找到的谱包络线数据上的修改的已编码参数$mod_a_m[m][k]$、音调上的修改的已编码参数起$mod_P_{ch}[m]$和浊音／清音判定数据

上的修改的已编码参数 mod_UV v [m]送往噪声合成电路 216 为浊音（V）部分进行噪声相加。噪声合成电路 216 的输出通过加权叠加电路 217 送往加法器 218。特别地说，考虑了来自已编码语音数据的参数的噪声，例如音调谱包络线振幅、在帧或残数信号级中的最大振幅，加在 LPC 合成滤波器输入的 LPC 残数信号的浊音部分，它是激励信号，考虑到如果对 LPC 合成滤波器的浊音的输入，它是激励信号，是由正弦合成产生的话，则在低音调声音中，例如男人的语音，产生“窒息的”感觉，而在声音质量在 V 和 UV 部分之间迅速改变时，就会产生不自然的感觉。

加法器 218 的和输出送往用于浊音的合成滤波器 236，在这里由 LPC 合成产生时间波形数据。另外，结果时间波形数据由一个后滤波器 238 v 滤波，然后供给加法器 239。

注意如前所述，LPC 合成滤波器 214 分为为 V 用的合成滤波器 236 和为 UV 用的合成滤波器 237。如果合成滤波器不按这种方式分开，亦即如果在 V 和 UV 信号部分之间不加区别连续地每 20 个样本或每 2.5 毫秒对 LSPs 进行插值的话，则在 V 到 UV 和 UV 到 V 的过渡部分完全不同的特征的 LSPs 插值，于是产生外部声音。为防止这种不良的效应，分开 LPC 合成滤波器为 V 的滤波器和为 UV 的滤波器以便独立地对 V 和 UV 插值 LPC 系数。

由参数修改单元 5 计算的 LPC 残数上的修改的编码参数 $mod\_res[m][i][j]$ 送往窗口电路 223 以便用浊音部分平滑连接部分。

窗电路 223 的输出送往 LPC 合成滤波器 214 的为 UV 的合成滤波器 237 作为清音合成单元 220 的输出。合成滤波器 237 执行数据的 LPC 合成，为清音部分提供时间波形，其由为清音的后滤波器 238 u 滤波，然后供给加法器 239。

加法器 239 把从为浊音的后滤波器 238 v 来的浊音部分的时间波形加到从为清音部分的后滤波器 238 u 来的清音部分的时间波形数据上并把结果数据在输出端子 201 上输出。

使用现在的语音信号再生设备 1，代替固有矩阵*[^]，其中 $0 \leqslant n < N1$，以这种方式解码一个修改的已编码参数矩阵 mod_*[m]，式中 $0 \leqslant m < N2$，。解码时期的帧区间可以固定为例如通常的 20 毫秒。在这种情况下，时间轴

压缩和因而得到的再生率的加速可能在N2＜N1下实现，而时间轴的扩展和因而得到的再生率的减速可能在N2＞N1下实现。

使用本系统，最终得到的参数串放在一个固有间隔为20毫秒的矩阵中供解码，以便可以容易地实现最优加速。此外，加速和减速的实现使用同一处理操作不需任何区别。其结果是，可以以两倍于实时速度的速度复制固态纪录的内容。由于不管增加的放音速度音调和音素保持不变，所以可以容易地辨别纪录的内容而不管以显著增加的放音速度复制。

如果N2＜N1，也就是说如果放音速度降低，则由于在清音的场合复数参数 mod_res 从同一LPC残数 res 产生。所以放出的声音可能不自然。在这种情况下，在参数 mod_res 上可以加一个适当数量的噪声以消除这种不自然到一定程度。也可以应用适当产生的高斯噪声或从码表中随机选择的激励矢量代替参数 mod_res 而不用加噪声。

使用上述语音信号复制设备1，从编码单元2来的编码参数的输出时期的时间轴由时期修改单元3为加速复制速度而压缩。不过，帧长度可以由解码单元4改变以控制复制速度。

在这种情况下，由于帧长度是可变的，而帧数目n在解码单元4的参数修改单元5产生参数前和后不变。

参数修改单元5还分别修改参数，lsp[n][p]和UVv[m]为mod_lsp[n][p]和mod_UVv[n]，而不管所涉及的帧是浊音还是清音。

如果mod_UVv[m]是1，也就是说如果所涉及的帧是浊音（V），则参数 $P_{ch}[n]$ 和 $a_m[n][k]$ 分别修改为 $mod\_P_{ch}[n]$ 和 $mod\_a_m[n][k]$。

如果mod_UVv[m]是0，也就是说如果所涉及的帧是清音（V），则参数 $res[n][i][j]$ 修改为 $mod\_res[n][i][j]$。

参数修改单元5直接修改lsp[n][p]、$P_{ch}[n]$、UVv[n]和 $a_m[n][k]$ 为 $mod\_lsp[n][p]$、$mod\_P_{ch}[n]$、mod_UVv[m]和 $mod\_a_m[n][k]$。不过参数修改单元根据速度spd改变残数信号 $mod\_res[n][i][j]$。

如果速度spd＜1.0，也就是说，如果速度较快，则原始信号的残数信号在中间部分切分，如图12所示。如果原始帧长度是OrgFrmL，则从原始帧长度 $res[n][j]$ 切下 $(OrgFrmL-FrmL)/2 \leqslant j \leqslant (OrgFrmL+frmL)/2$

给 $mod\_res[n][j]$。从原始帧的前端切分也是可以的。

如果速度 spd＞1.0，也就是说，如果速度比较慢，则使用原始帧而对任何短缺的部分使用加有噪声分量的原始帧。也可以使用加有适当产生的噪声的解码激励矢量。可以产生高斯噪声并用作激励矢量以减少由同样波形的帧的连续而产生的不相容的感觉。上面的噪声分量也可以加在原始帧的两端。

于是，在速度信号复制设备 1 配置为通过改变帧的长度而改变速度控制的场合，速度合成单元 6 构造和设计为使 LSP 插值单元 232v 和 232u、正弦合成单元 215 和窗单元 223 执行不同的操作来通过时间轴压缩控制速度。

如果所涉及的帧是浊音帧（V），则 LSP 插值单元 232v 寻找满足关系 frmL/P≤20 的最小整数 p。如果所涉及的帧是清音帧（UV），则 LSP 插值单元 232u 寻找满足关系 frmL/P≤80 的最小整数 p。为 LSP 插值的子帧的范围 sub1[i][j]由下式确定:

$$nint(frm\ L/p\times i)\leqslant j\leqslant nint(frm\ L/P\times(j+1)),\ 其中\ 0\leqslant i\leqslant p-1$$

在上式中，nint（x）是一个函数，它通过圆整十分位而返回一个最接近 x 的整数。对于浊音和清音两者，如果 frmL 小于 20 或 80，则 p=1。

例如，对于第 i 个子帧，由于该子帧的中心是 frmL×（2i+1）/2p，LSPs 以 frmL×（2p-2i-1）/（20: frmL×(2i+1)/2p 的速率插值，如同在我们的待审日本专利申请 6-198451 中公开的。

另一方案为，子帧数目可以固定，而每一子帧的 LSPs 可以以同一速率在任何时间插值。正弦合成单元 223 修改窗长度以与帧长度 frmL 匹配。

使用上述语音信号复制设备 1，对于输出时期在时间轴上被压缩的已编码参数，使用时期修改单元 3 和参数修改单元 5 修改，以改变复制速度而不改变音调和音素。不过也可以省略时期修改单元 3 而由编码单元 2 通过在图 14 中所示解码单元 8 的一些数据变换单元 270 处理这些已编码数

据，改变音调而不改变音素。在图 14 中，用同样的数字指示相应于图 4 所示的元部件。

解码单元 8 所基于的基本概念是在一个预先设定的频带中变换从编码单元 2 来的编码语音数据的谐波的基本频率和振幅数据的数目，它使用作为数据变换设备的一定数目的数据变换单元 270 执行只变换音调而不改变音素的操作。数据数目变换单元 270 通过修改说明在每一输入谐波中的谱分量大小的数据数目改变音调。

参考图 14，相应于图 2 和图 3 的输出端子 102 的输出的一个 LSPs 的矢量量化输出，或码表指数，供给输入端子 202。

LSP 指数数据送往 LPC 参数复制单元 213 的逆矢量量化器 231 为逆矢量量化为线性谱对（LSPs）。LSPs 送往 LSP 插值电路 232、233 进行插值，然后供给 LSP 到 α 变换电路 234、235 以变换为线性预测码的 α 参数。这些 α 参数送往 LPC 合成滤波器 214。LSP 插值电路 232 和 α 变换电路 234 用于浊音（V）信号部分，而 LSP 插值电路 233 和 LSP 到 α 变换电路 235 用于清音（UV）信号部分。LPC 合成滤波器 214 由用于浊音部分的一个 LPC 合成滤波器 236 和用于清音部分的一个 LPC 合成滤波器 237 组成。也就是说，LPC 系数插值对浊音部分和清音部分独立地执行，以防止在插值从浊音部分到清音部分的过渡区域和从清音部分到浊音部分的过渡区域完全不同特征的 LSPs 可能引起的不良影响。

在图 14 的输入端子 203 上供有相应于在图 2 和图 3 所示的编码器的端子 103 的输出的谱包络线 Am 的加权矢量量化代码指数数据。在输入端子 205 供有从图 2 和图 3 的端子 105 来的浊音／清音判定数据。

从输入端子 203 来的谱包络线 Am 的矢量量化指数数据送往逆矢量量化器进行逆矢量量化。逆矢量量化的包络线的振幅数据的数目固定为一个预先设定的值例如 44。基本上，变换数据数目为给出相应于音调数据的谐波数目。如果希望改变音调，例如在本实施例中这样，从逆矢量量化器 212 来的包络线数据送往数据数目变换单元 270 为例如通过插值改变振幅数据的数目，取决于希望的音调值。

数据数目变换单元 270 还由从输入端子 204 来的音调数据供给使得在

编码时期的音调改变为一个希望的音调输出。振幅数据和修改的音调数据送往浊音合成电路211的正弦合成电路215。供给合成电路215的振幅数据的数目相应于从数据数目变换单元270来的LPC残数的谱包络线的修改后的音调。

有多种插值方法用于使用数据数目变换单元270来改变LPC残数的谱包络线的振幅数据数目。例如，为插值在频率轴上有效频带块的振幅数据从在该块中最后一个振幅数据直到块中第一振幅数据的一适当数目的哑数据或延伸块中最左端（第一数据）到最右端（最后数据）的哑数据附在块中的振幅数据上以增大数据数目到 N_F。然后通过频带限制型 Os 元组过采样，例如8元组过采样寻找一个振幅的Os元组数。振幅数据的Os元组数（$(m_{MX}+1)\times$数据的Os数）通过插值进一步扩展到一个更大的数 N_M，例如2048。这一NM数目数据通过多中取一而变换为预先设定的数M（例如44），然后对该预先设定的数目的数据执行矢量量化。

作为对数据数目变换单元270的操作示例，说明对音调延迟的频率为F0＝fs/L的情况，式中fs是采样频率，为fs＝8千赫兹＝8000赫兹。

在这种情况下，音调频率F＝8000／L，而有n＝L／2的谐波设定到4000赫兹。在3400Hz的一般语音范围中，谐波数目为（L／2）×（3400／4000）。它们在进行矢量量化之前由上述数据数目变换或大小变换变换到例如44。如果只是要改变音调，则没有必要进行量化。

在逆矢量量化后，谐波数目44可以由数据数目变换单元270通过大小变换改变为一个希望的数目，也就是说变为一个希望的音调频率Fx。相应于音调频率Fx（Hz）的音调延迟Lx为Lx＝8000／Fx，这样，设定到3400赫兹的谐波的数目为（Lx/2）×（3400/4000）＝（4000/Fx）×(3400/4000) ＝ 3400/Fx 即 3400／Fx。也就是说通过大小变换或在数据数目变换单元270中的数据数目变换执行从44到3400／Fx就足够了。

如果在执行谱数据的矢量量化之前的编码时期发现帧与帧的差，则在逆矢量量化后解码帧与帧的差。然后执行数据数目的变换以产生谱包络线数据。

正弦合成电路215不只由音调数据和从数据数目变换单元270来的

LPC 残数的谱包络线振幅数据供给，还由从输入端子 205 来的浊音／清音判定数据供给。从正弦合成电路 215 取出 LPC 残数数据并送往加法器 218。

从逆矢量量化器 212 来的包络线数据、从输入端子 204 来的音调数据和从输入端子 205 来的浊音／清音判定数据送往噪声加法电路 216 为浊音（V）部分进行噪声相加。具体地说，考虑了从已编码语音数据来的参数的噪声，例如音调谱包络线振幅，帧或残数信号级中的最大振幅，加到 LPC 残数信号的浊音部分作为 LPC 合成滤波器的输入，它是激励信号，考虑到如果对浊音的 LPC 合成滤波器的输入，它是激励信号，是由正弦合成产生的话，则在低音调声音中，例如男人的语音，产生“窒息的”感觉，而在声音质量在 V 和 UV 语音部分之间迅速改变时，就会产生不自然的感觉。

加法器 218 的和输出送往为浊音的合成滤波器 236，这里由 LPC 合成产生时间波形数据。另外，结果时间波形数据为由浊音数据的一个后滤波器 238 v 滤波，然后供给加法器 239。

在图 14 的输入端子 207s 和 207g 上供给形状指数数据和增益指数数据作为通过时期修改单元 3 从图 3 的输出端子 107s 和 107g 来的 UV 数据。然后形状指数数据和增益指数数据供给清音合成单元 220。从端子 207s 来的形状指数数据和从端子 207g 来的增益指数数据分别供给清音合成单元 220 的噪声码表 221 和增益电路 222。从噪声码表 221 读出的代表值输出是相应于清音的 LPC 残数的噪声信号分量而成为增益电路 222 中一个预先设定的增益的振幅。预先设定的增益振幅的代表值输出送往窗电路 223 以平滑到浊音信号部分的连接部分。

窗电路 223 的输出作为清音合成单元 220 的输出送往 LPC 合成滤波器 214 的为清音（UV）部分的合成滤波器 237。窗电路 223 的输出由合成滤波器 237 用 LPC 合成处理而给出清音信号部分的时域波形信号，然后由为清音部分的后滤波器 238 u 滤波再供给加法器 239。

加法器 239 把从为浊音的后滤波器 238 v 来的浊音信号部分的时域波形信号加到从为清音信号部分的后滤波器 238 u 来的清音信号部分的时域波形数据上。结果和信号在输出端子 201 上输出。

从上面可以看到，通过改变谐波的数目而不改变谱包络线的形状可以

改变音调而不改变语音的音素。于是，如果一个语音模式的已编码数据，亦即编码位流可用的话，则可以为合成有选择地改变音调。

参考图 15，由图 2 和图 3 的编码器编码而获得的已编码数据的编码位流通过编码数据输出单元 301 输出。在这些数据中，至少音调数据和谱包络线数据通过数据变换单元 302 送往波形合成单元 303。与音调变换无关的数据，例如浊音／清音（V／UV）判定数据直接送往波形合成单元 303。

波形合成单元 303 根据谱包络线数据或音调数据合成语音波形。自然，在图 4 和图 5 所示的合成设备的场合下，LSP 数据和 CELP 数据也从输出单元 301 中取出而如上述供给。

在图 15 的配置中，至少音调数据和谱包络线数据根据希望的音调由数据变换单元 302 如上所述变换，然后供给波形合成单元 303，在这里从变换的数据合成语音波形。于是，音调改变而音素未变的语音信号可以在输出端子 304 取出。

上述技术可以应用于由规则或文本的语音的合成。

图 16 表示本发明应用到语音文字合成的一个例子。在本实施例中，上述用于压缩语音编码的解码器可以同时用作文字语音合成器。在图 16 的例子中，语音数据的再生联合使用。

在图 16 中，语音规则合成器和带上述为修改音调的数据变换的语音合成器结合在一个按照规则的语音合成单元 300 中。从文字分析单元 310 来的数据供给按照规则的语音合成单元 300，从它输出具有希望音调的合成语音并送往转换开关 330 的一个固定触点 a。语音复制单元 320 读出偶而压缩的语音数据并存储在例如只读存储器的存储器中并为扩展而解码这些数据。解码的数据送往转换开关 330 的另一固定触点 b。一个合成语音信号和复制语音信号由转换开关 330 选择并在输出端子 340 上输出。

图 16 所示设备用于例如车辆导行系统。在这种场合下，从语音再生器 320 来的高质量高清晰度的复制语音可用于日常语音，例如提供指示的“请向右转”，而从按照规则的语音合成生成器 300 来的合成语音可以用于特殊指示物的语音，例如建筑物或地界，它数量大，不能作为语音信息存储在只读存储器中。

本发明有另外的优点，即同样的硬件可以用于计算机语音合成器 300 和语音复制器 320。

本发明不限于上述实施例。例如，上述作为硬件叙述的图 1 和图 3 的语音分析侧（编码器）或在图 14 中的语音合成一侧（解码器）的构建可以通过使用例如数字信号处理器（DSP）的一个软件程序实现。多个帧的数据可以一起处理并由矩阵量化而代替矢量量化。本发明也可以应用于大量语音分析／合成方法。本发明也不限于传输或纪录／复制而可能应用于各种用途，例如音调转换速度或速率变换按照规则的语音合成或噪声抑制。

上述信号编码和信号解码设备可以用作示于图 14 中的用于例如便携式通信终端或便携电话机中的语音编码器。

图 17 表示使用在图 2 和图 3 中所示配置的语音编码单元 160 的便携式终端的发送一侧。由受话器 161 接收的语音信号由放大器 162 放大并由模／数（A／D）转换器 163 变换为数字信号，其送往在图 1 和图 3 中所示配置的语音编码单元 160。从换／数变换器 163 来的数字信号供给输入端子 101。语音编码单元 160 执行编码，其已结合图 1 和图 3 叙述。图 1 和图 2 的输出端子的输出信号作为语音编码单元 160 的输出信号送往发送频道编码单元 164，它随之对供给的信号执行频道编码。发送频道编码单元 164 的输出信号送往调制电路 165 进行调制，然后通过数／模变换器 166 和一个 RF 放大器 167 供给天线 168。

图 18 表示使用在图 5 和图 14 中所示配置的语音解码单元 260 的便携式终端的接收一侧。由图 18 的天线 261 接收的语音信号由 RF 放大器 262 放大并通过模／数变换器 263 送往解调电路 264，解调的信号从这里送往发送频道解码单元 265。解码单元 265 的输出信号供给在图 5 和图 14 中所示配置的语音解码单元 260。语音解码单元 260 解码这一信号，其已结合图 5 和图 14 叙述。在图 2 和图 4 的输出端子 201 上的输出信号作为语音解码单元 260 的信号送往数／模（D/A）变换器 266。从模／数变换器 266 来的模拟语音信号送往扬声器 268。

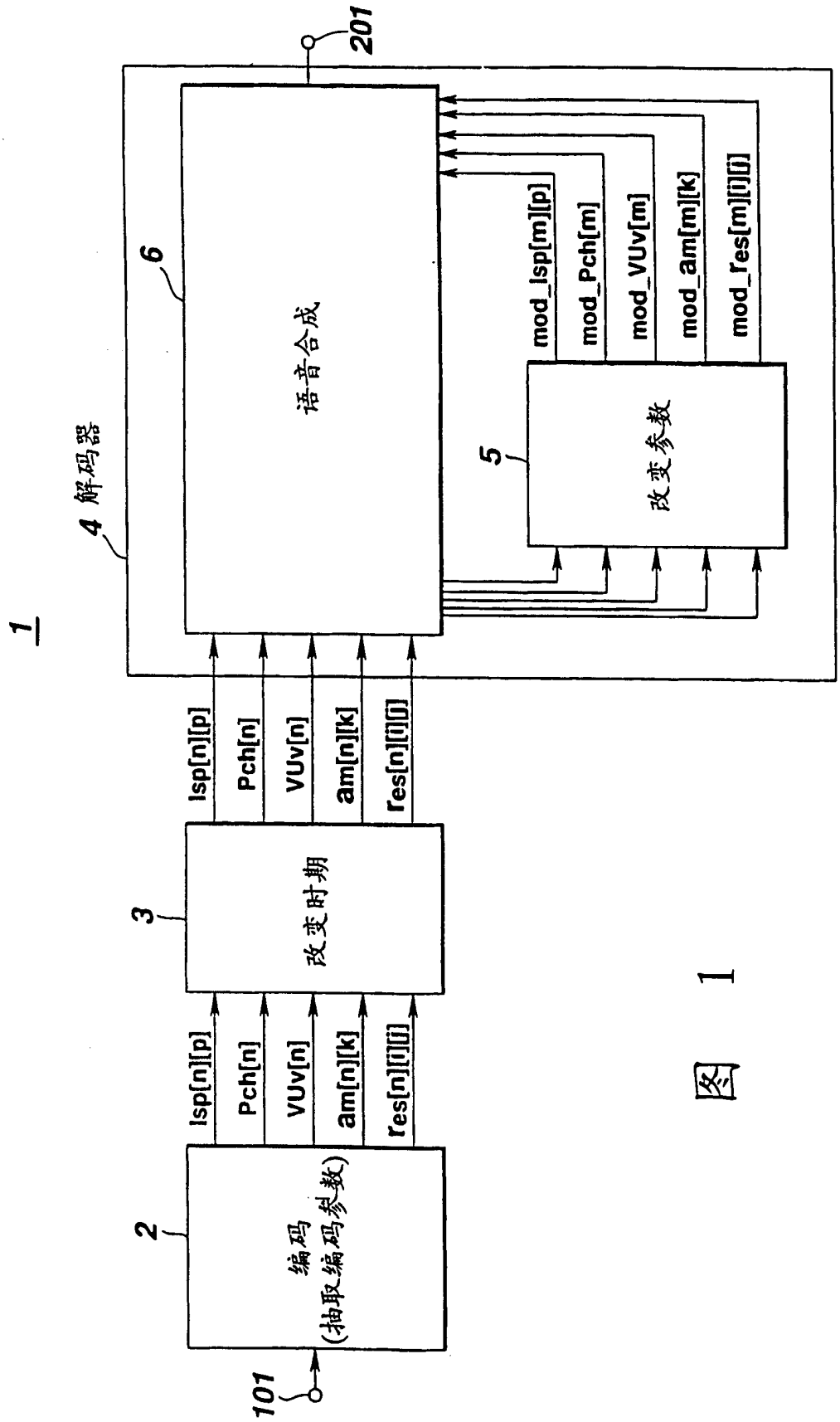

1
2
编码
(抽取编码参数)
101
3
改变时期
lsp[n][p]
Pch[n]
VUv[n]
am[n][k]
res[n][i][j]
4 解码器
5
改变参数
6
语音合成
201
mod_lsp[m][p]
mod_Pch[m]
mod_VUv[m]
mod_am[m][k]
mod_res[m][i][j]

图 1

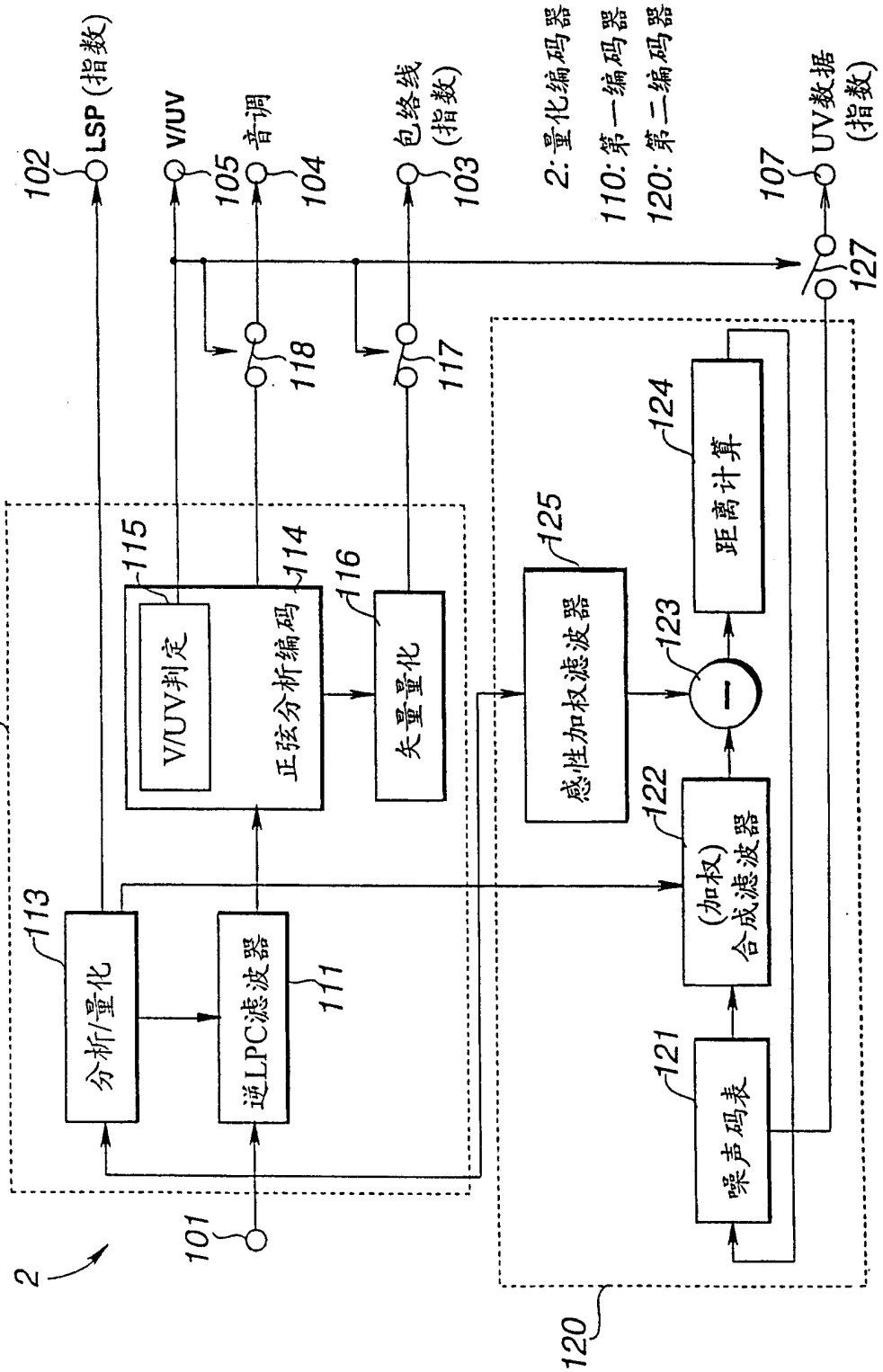

2
101
110
113
分析/量化
逆LPC滤波器
111
V/UV判定
115
正弦分析编码
114
矢量量化
116
感性加权滤波器
125
(加权)
合成滤波器
122
噪声码表
121
距离计算
124
123
120
102
LSP (指数)
105
V/UV
音调
104
包络线
(指数)
103
118
117
2: 量化编码器
110: 第一编码器
120: 第二编码器
107
UV数据
(指数)
127

图 2

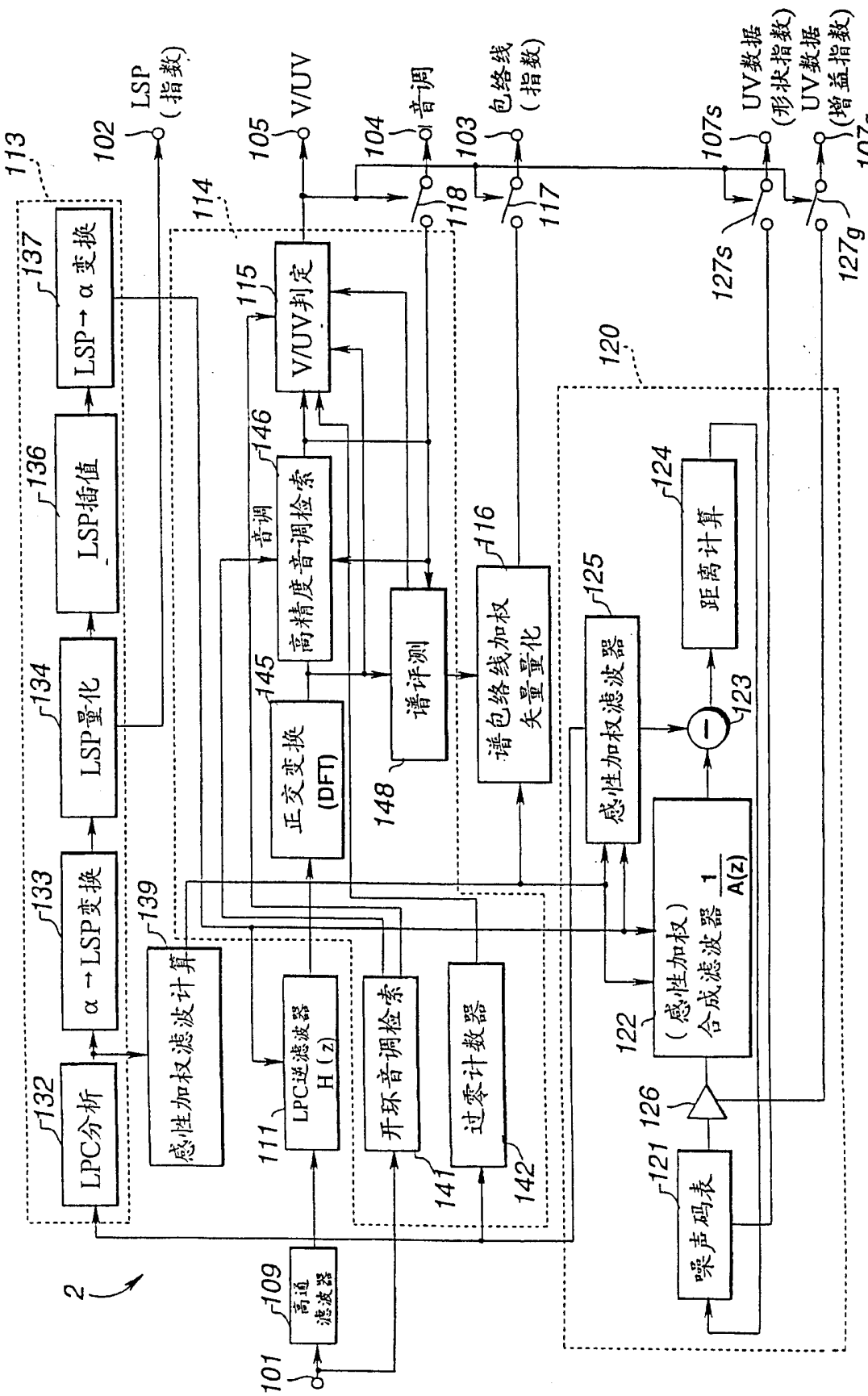

LSP (指数)
V/UV
音调
包络线 (指数)
UV数据 (形状指数)
UV数据 (增益指数)
LPC分析
α→LSP变换
LSP量化
LSP插值
LSP→α变换
感性加权滤波计算
LPC逆滤波器 H(z)
开环音调检索
过零计数器
高通滤波器
正交变换 (DFT)
高精度音调检索
V/UV判定
谱评测
谱包络线加权失量量化
感性加权滤波器
距离计算
(感性加权) 合成滤波器 1/A(z)
噪声码表

图 3

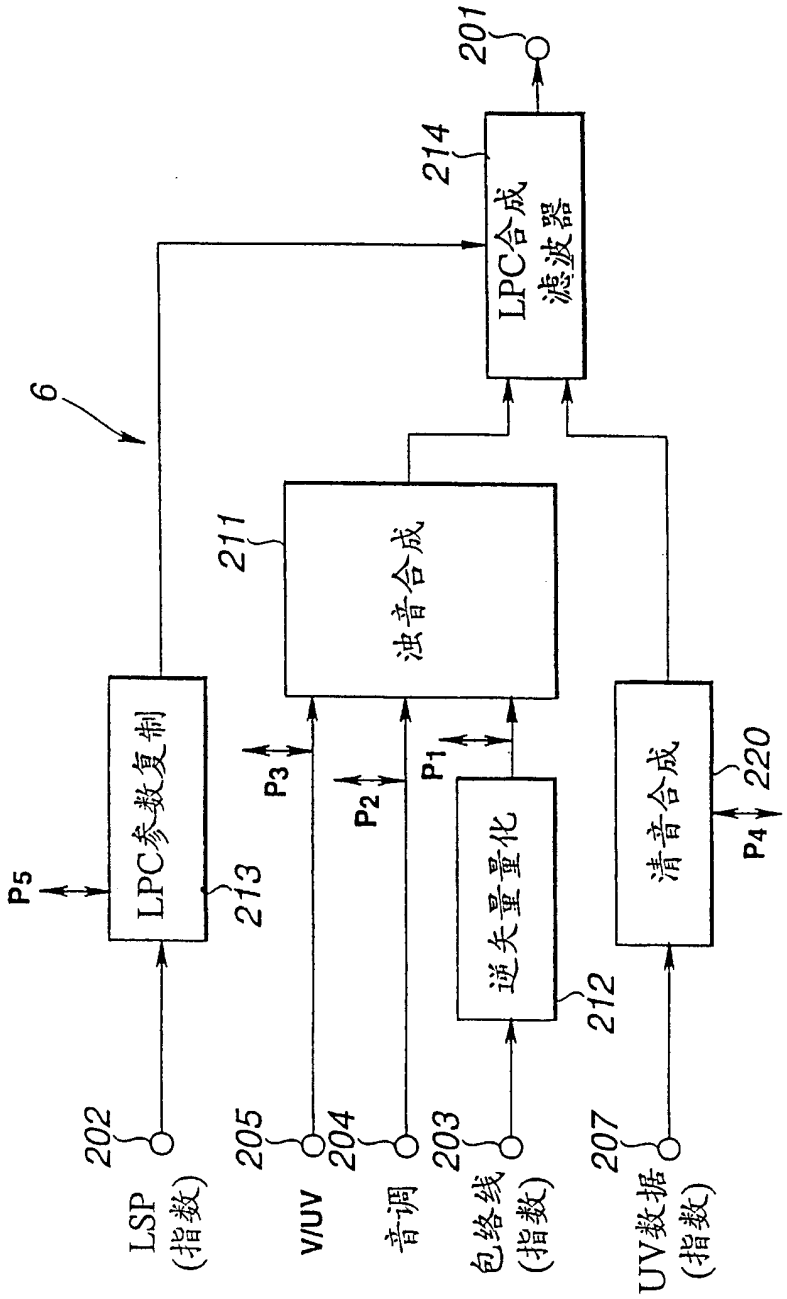

6
201
214
LPC合成
滤波器
211
浊音合成
P5
LPC参数复制
213
P3
P2
P1
逆矢量量化
212
清音合成
220
P4
202
205
204
203
207
LSP
(指数)
V/UV
音调
包络线
(指数)
UV数据
(指数)

图 4

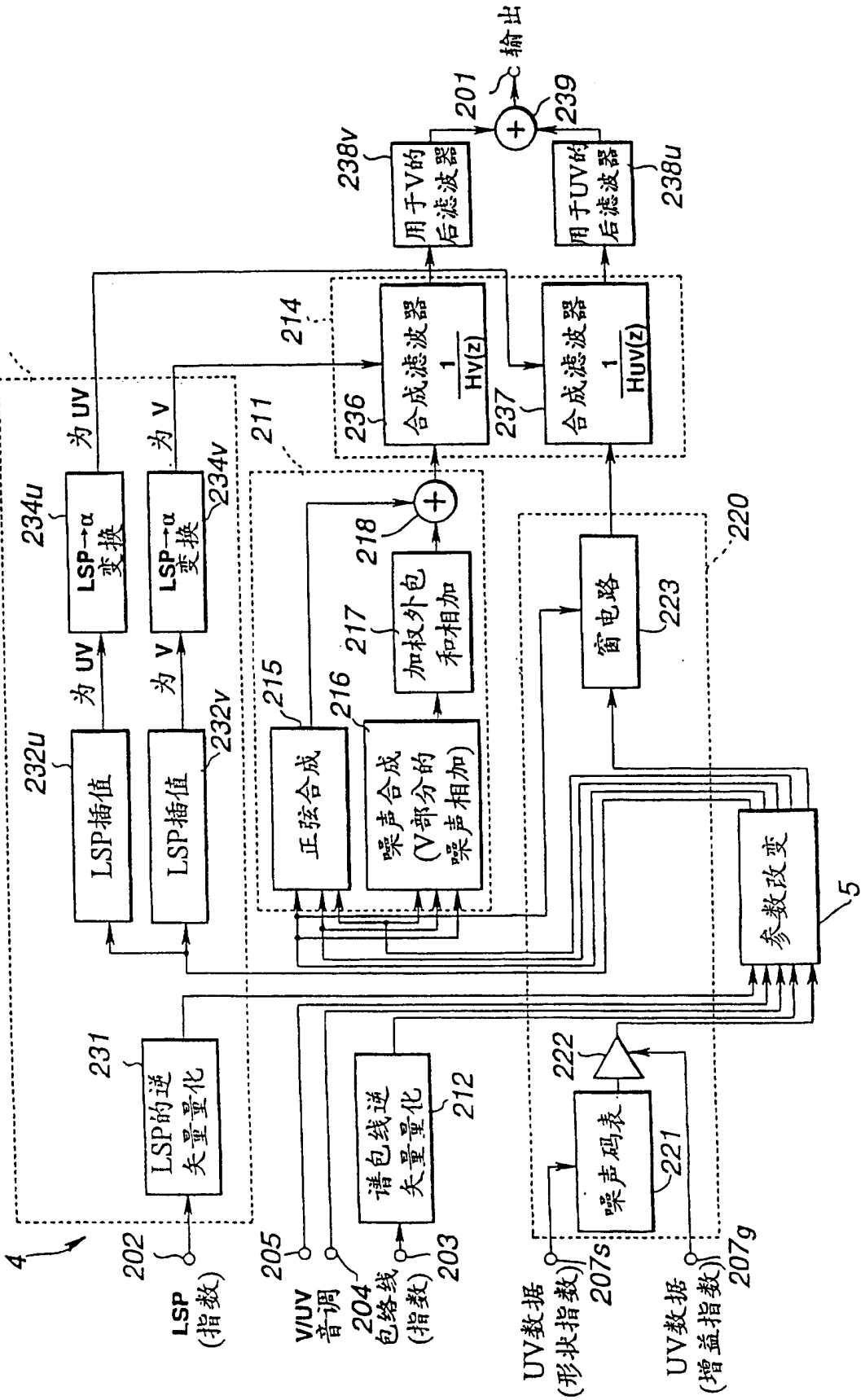

4
LSP (指数)
202
V/UV 音调
205
204 包络线 (指数)
203
UV数据 (形状指数)
207s
UV数据 (增益指数)
207g
LSP的逆矢量量化
231
谱包线逆矢量量化
212
噪声码表
221
222
参数改变
5
LSP插值
232u
LSP插值
232v
为 UV
为 V
LSP→α变换
234u
LSP→α变换
234v
为 UV
为 V
213
正弦合成
215
噪声合成 (V部分的噪声相加)
216
加权外包和相加
217
218
211
窗电路
223
220
合成滤波器 1/Hv(z)
236
合成滤波器 1/Huv(z)
237
214
用于V的后滤波器
238v
用于UV的后滤波器
238u
201
239
输出

图 5

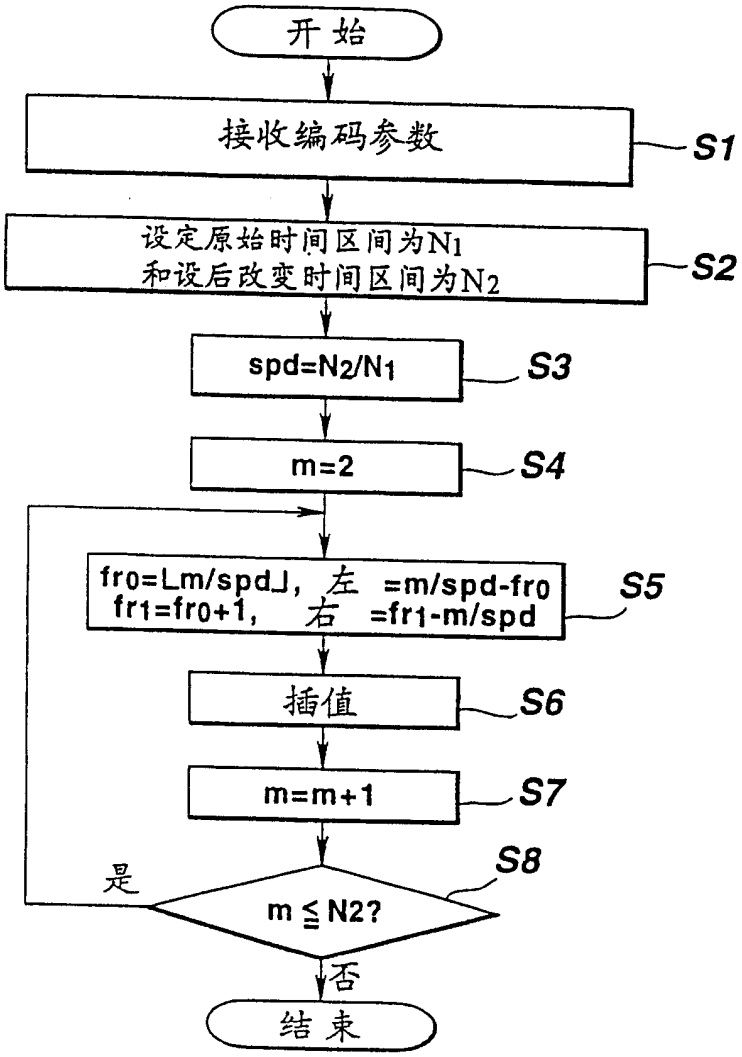

开始
接收编码参数
S1
设定原始时间区间为N1
和设后改变时间区间为N2
S2
spd=N2/N1
S3
m=2
S4
fr0=⌊m/spd⌋, 左 =m/spd-fr0
fr1=fr0+1, 右 =fr1-m/spd
S5
插值
S6
m=m+1
S7
是
m ≦ N2?
S8
否
结束

图 6

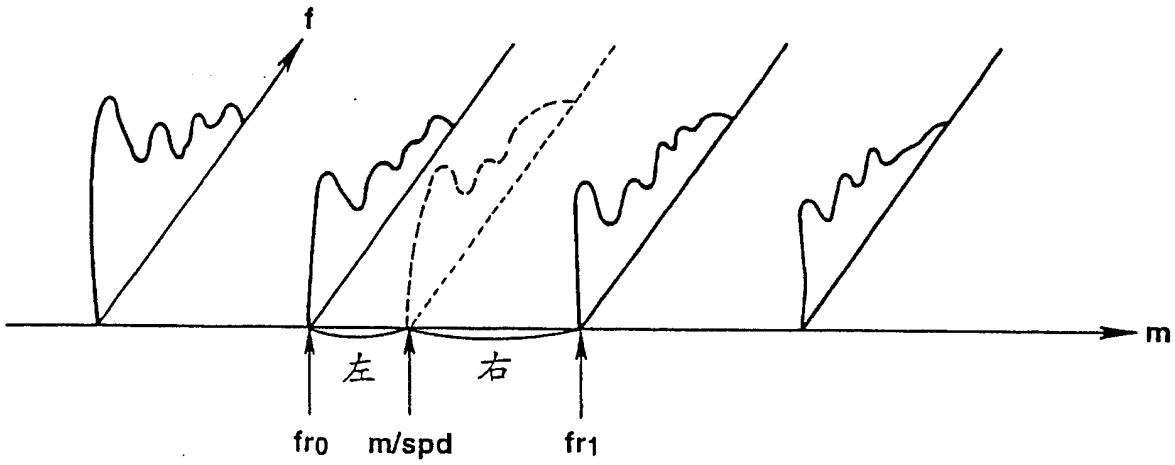


图 7

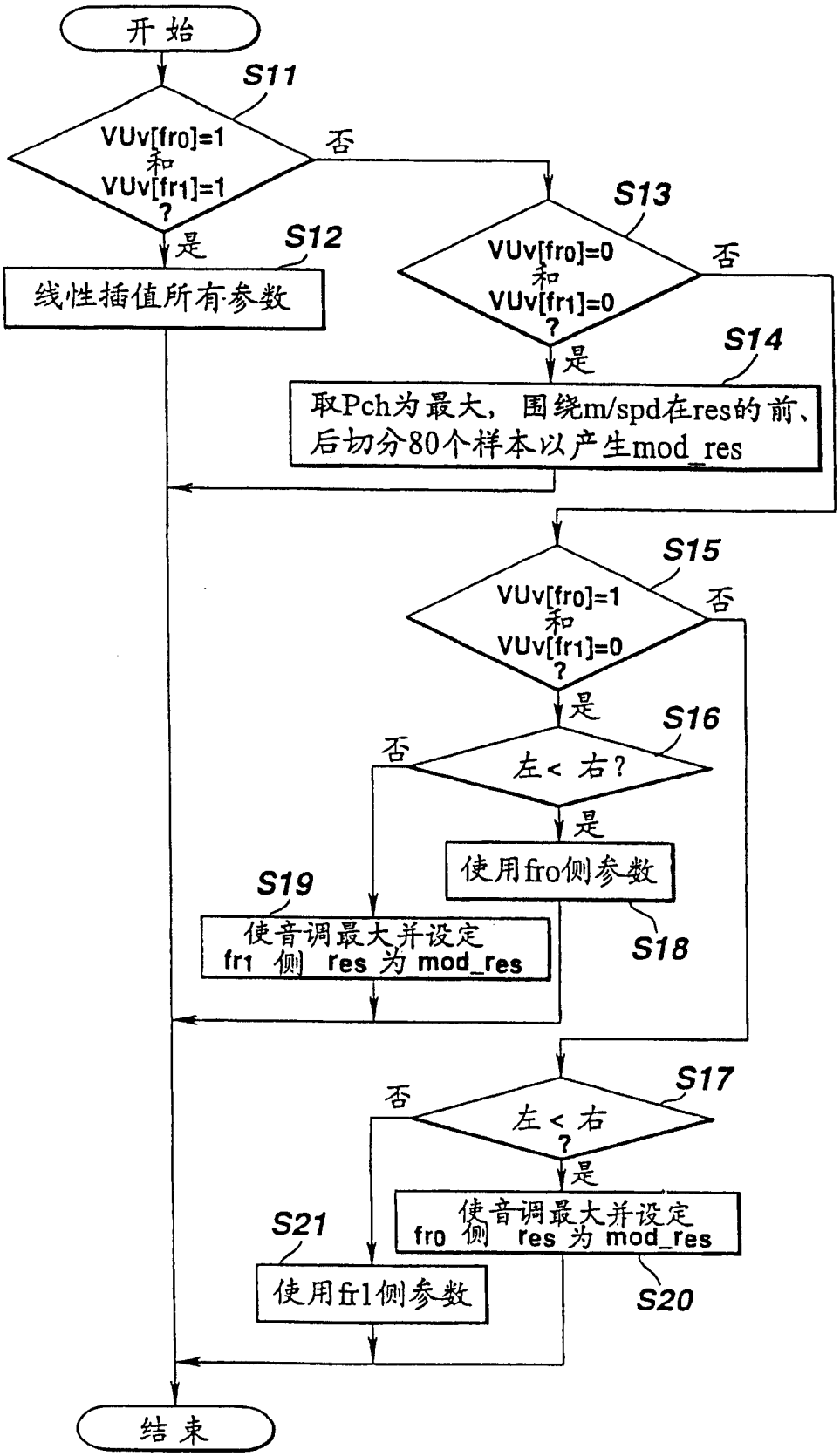

开始
S11
VUv[fr0]=1 和 VUv[fr1]=1 ?
否
是
S12
线性插值所有参数
S13
VUv[fr0]=0 和 VUv[fr1]=0 ?
否
是
S14
取Pch为最大，围绕m/spd在res的前、后切分80个样本以产生mod_res
S15
VUv[fr0]=1 和 VUv[fr1]=0 ?
否
是
S16
左< 右?
否
是
使用fro侧参数
S18
S19
使音调最大并设定 fr1 侧 res 为 mod_res
S17
左 < 右 ?
否
是
使音调最大并设定 fr0 侧 res 为 mod_res
S20
S21
使用fr1侧参数
结束

图 8

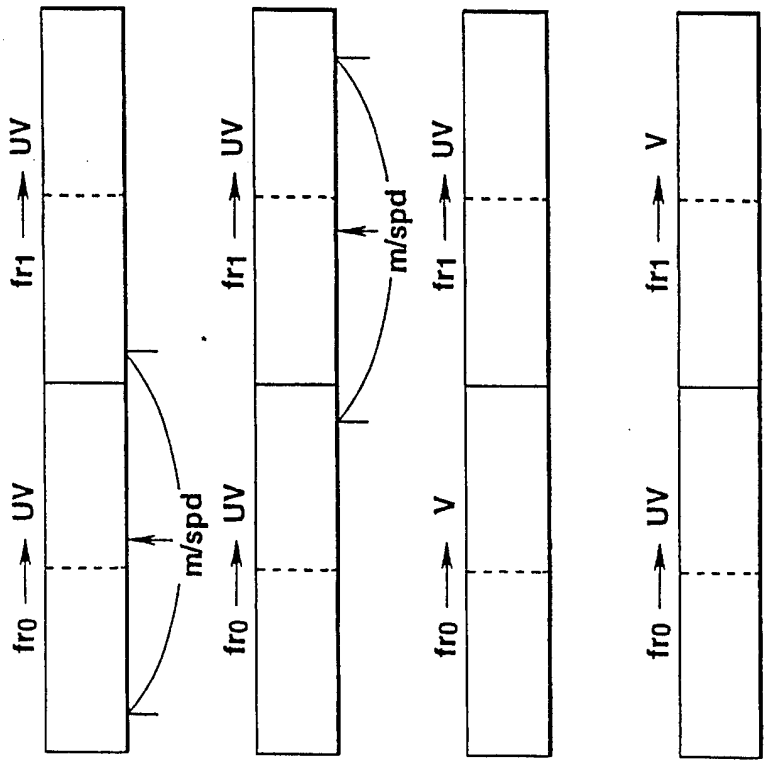


图 9A 图 9B 图 9C 图 9D

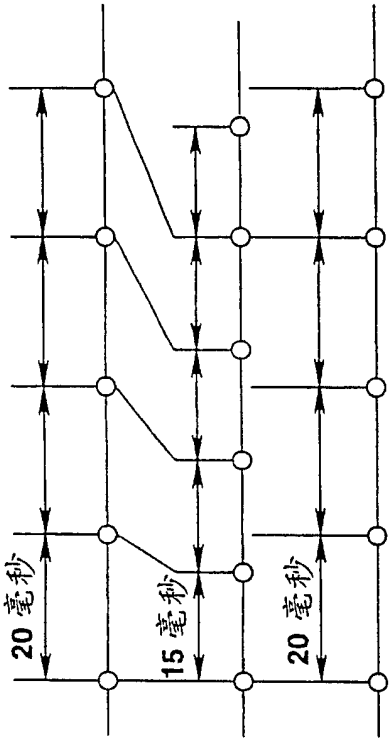


图 10A 图 10B 压缩 图 10C 插值

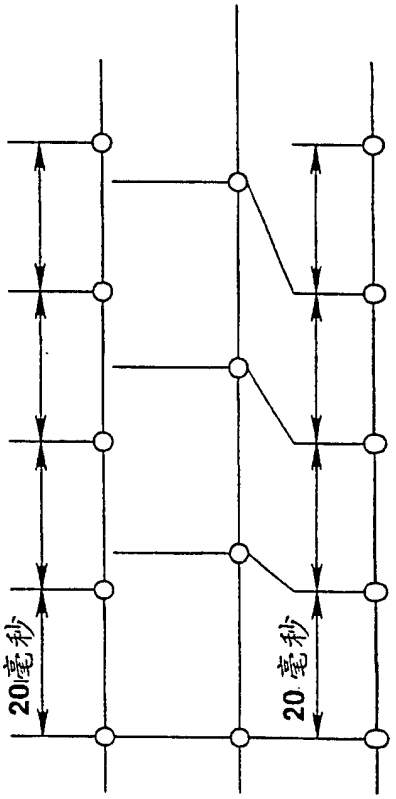


图 11A 图 11B 插值 图 11C 压缩

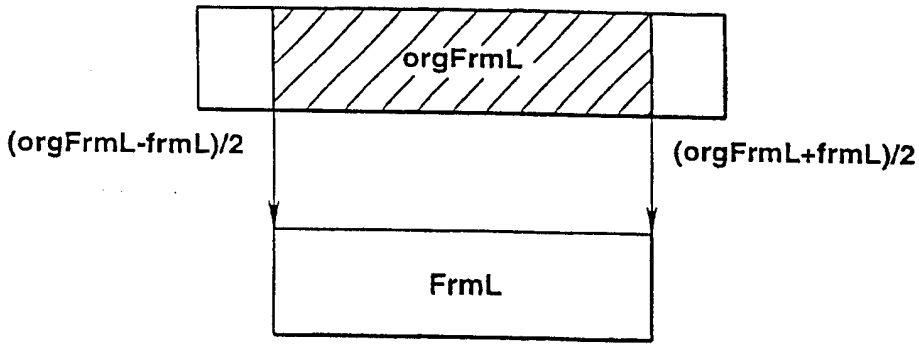


图 12

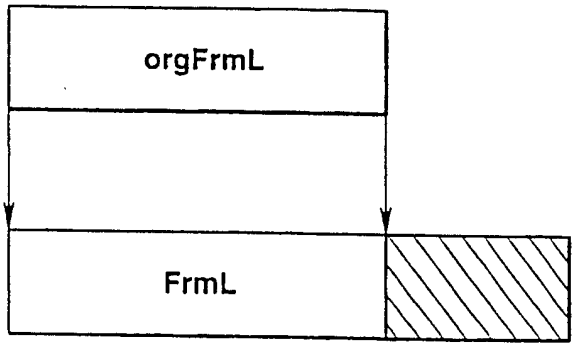


图 13

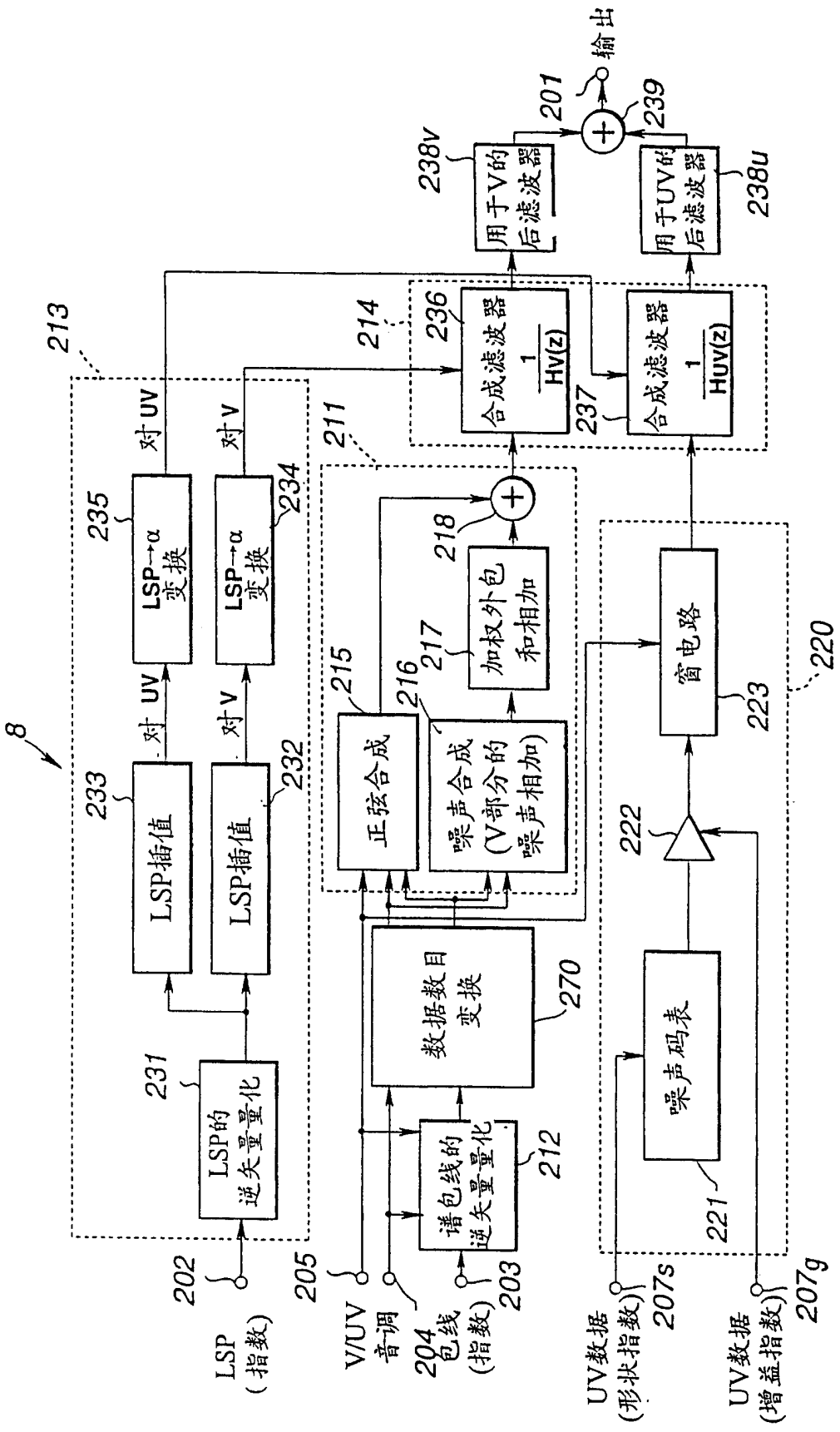

8
213
LSP的逆矢量量化
231
LSP插值
233
LSP插值
232
对 UV
对 V
LSP→α变换
235
LSP→α变换
234
对 UV
对 V
LSP（指数）
202
V/UV音调
205
包线（指数）
204
谱包线的逆矢量量化
212
数据数目变换
270
211
正弦合成
215
噪声合成（V部分的噪声相加）
216
加权外包和相加
217
218
214
合成滤波器 1/Hv(z)
236
合成滤波器 1/Huv(z)
237
用于V的后滤波器
238v
用于UV的后滤波器
238u
201
239
输出
220
窗电路
223
222
噪声码表
221
UV数据（形状指数）
207s
UV数据（增益指数）
207g

图 14

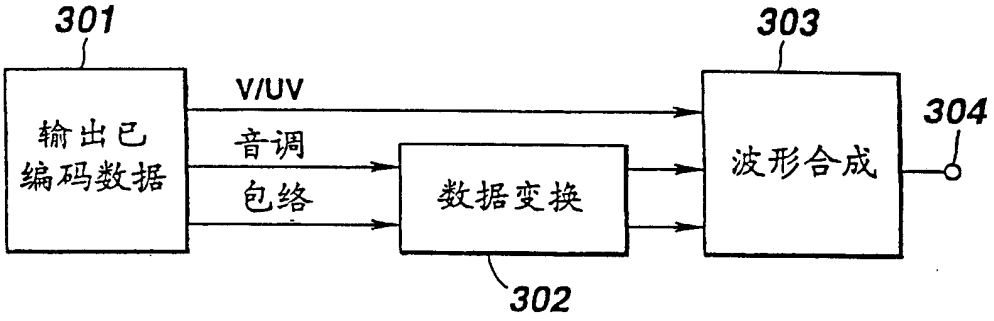

301
303
输出已
编码数据
V/UV
音调
包络
数据变换
波形合成
304
302

图 15

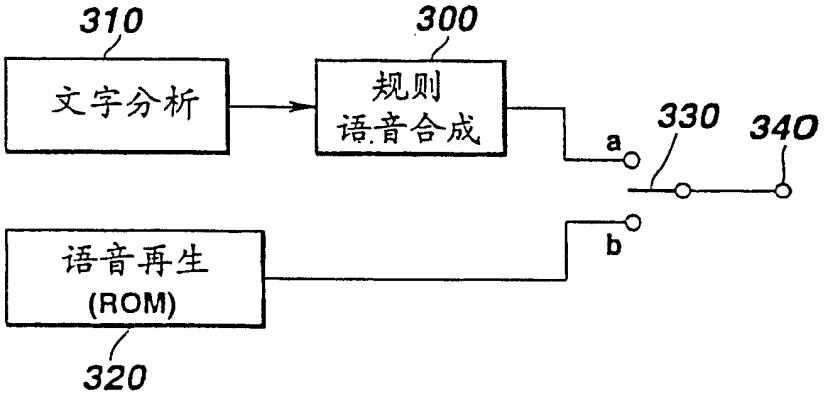

310
300
文字分析
规则
语音合成
a
330
340
b
语音再生
(ROM)
320

图 16

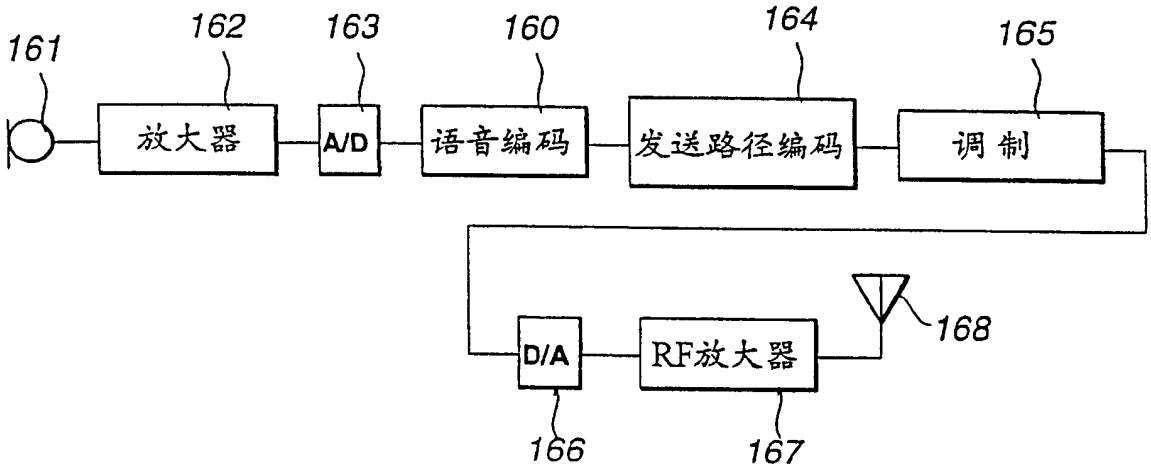

161
162
163
160
164
165
放大器
A/D
语音编码
发送路径编码
调 制
168
D/A
RF放大器
166
167

图 17

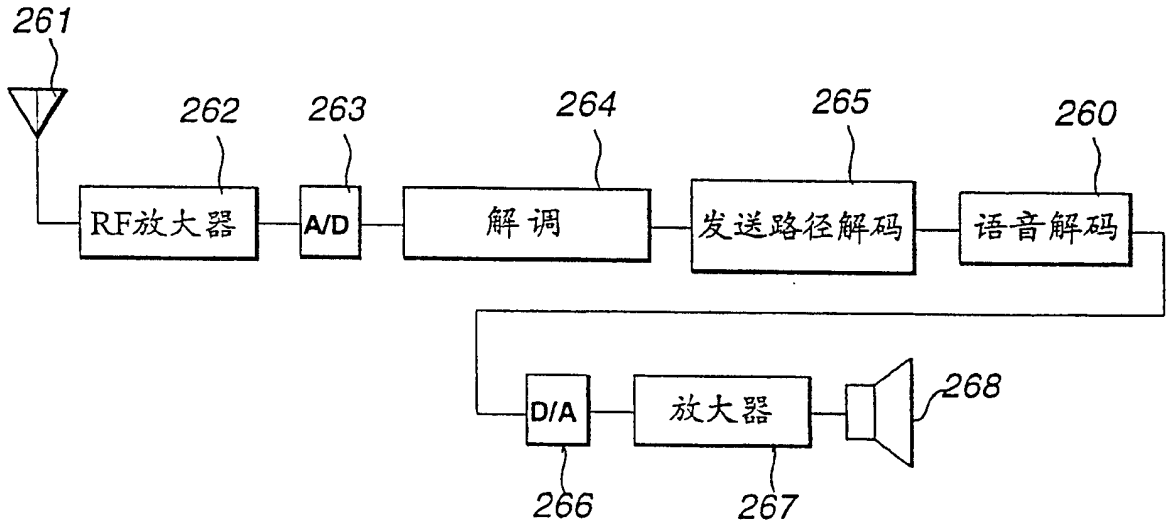

261
262
263
264
265
260
RF放大器
A/D
解 调
发送路径解码
语音解码
268
D/A
放大器
266
267

图 18