

[19]中华人民共和国国家知识产权局

[51]Int. Cl<sup>7</sup>

G10L 19/04

//G10L101: 12

# [12] 发明专利申请公开说明书

[21] 申请号 00801770.0

[43]公开日 2001年11月7日

[11]公开号 CN 1321297A

[22]申请日 2000.8.23 [21]申请号 00801770.0

[30]优先权

[32]1999.8.23 [33]JP [31]235050/1999

[32]1999.8.24 [33]JP [31]236728/1999

[32]1999.9.2 [33]JP [31]248363/1999

[86]国际申请 PCT/JP00/05621 2000.8.23

[87]国际公布 WO01/15144 日 2001.3.1

[85]进入国家阶段日期 2001.4.23

[71]申请人 松下电器产业株式会社

地址 日本国大阪府门真市

[72]发明人 安永和敏 森井利幸

[74]专利代理机构 上海专利商标事务所

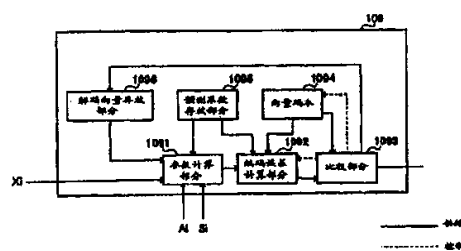
代理人 孙敬国

权利要求书 8 页 说明书 32 页 附图页数 14 页

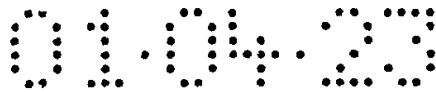
[54]发明名称 音频编码装置以及音频编码方法

[57]摘要

预先作成存放了多个量化对象向量的代表性采样的向量码本。各向量由 3 个部分即 AC 增益、SC 增益的对数值所对应的值、SC 的预测系数的调整系数组成。在预测系数存放部分中存放用于进行预测编码的系数。在参数计算部分中，从输入的听觉加权输入音频、听觉加权 LPC 合成后的自调声源、听觉加权 LPC 合成后的概率性声源、存放在解码向量存放部分的解码向量、存放在预测系数存放部分的预测系数而来计算间距计算所必要的参数。



ISSN 1008-4274



# 权 利 要 求 书

1. 一种音频编码装置，其特征在于，

具备 LPC 合成手段、增益运算手段以及参数编码手段，所述 LPC 合成手段对于自调码本以及概率性码本中存放的自调声源以及概率性声源通过使用从输入音频中求出的 LPC 系数进行滤波而获得合成音，所述增益运算手段求出所述自调声源以及所述概率性声源的增益并且使用所述增益而获得所述输入音频与所述合成音之间的编码误差来搜索自调声源以及概率性声源的代码，所述参数编码手段使用求得的代码所对应的自调声源以及概率性声源进行增益的预测编码。

所述参数编码手段具备根据以前子帧的状态调整使用于所述预测编码中的预测系数的预测系数调整手段。

2、如权利要求 1 所述音频编码装置，其特征在于，

所述预测系数调整手段当以前的子帧的状态为极大值或极小值时，调整所述预测系数使得减少影响。

3. 如权利要求 1 所述音频编码装置，其特征在于，

所述参数编码手段具有码本，所述码本包含自调声源的增益向量、概率性声源的增益向量以及调整预测系数的系数。

4. 如权利要求 3 所述音频编码装置，其特征在于，

在预测编码中，在求取状态与预测系数的积时，乘以所述状态对应的预测系数调整系数。

5. 如权利要求 1 所述音频编码装置，其特征在于，

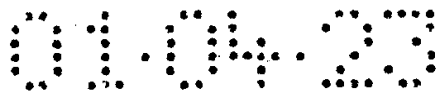
具备按每个状态对应地存放所述自调声源、概率性声源以及预测系数调整系数的存放手段。

6. 如权利要求 5 所述向量量化装置，其特征在于，

在更新存放在所述存放手段中的所述自调声源以及所述概率性声源的状态时，也更新所述预测系数调整系数。

7. 一种音频编码装置，其特征在于，

所述音频编码是将一帧分解为多个子帧进行编码的 CELP 型音频编码装置，具备 LPC 合成手段、增益运算手段以及参数编码手段，所述 LPC 合成手段对于自调码本以及概率性码本中存放的自调声源以及概率性声源通过使用从



输入音频中求出的 LPC 系数进行滤波而获得合成音，所述增益运算手段求出所述自调声源以及所述概率性声源的增益，所述参数编码手段对于使用所述输入音频与所述合成音之间的编码误差而求得自调声源及概率性声源并且进行所述增益的向量量化。

所述音频编码装置还具备音调分析手段，所述音调分析手段是在进行最初子帧的自调码本搜索之前分析构成帧的多个子帧的音调并求出相关值，使用该相关值算出最近似于音调周期的值。

8. 如权利要求 7 所述的音频编码装置，其特征在于，

还具备搜索范围决定手段，所述搜索范围决定手段根据所述音调分析手段获得的相关值以及最近音调周期的值，决定多个子帧的滞后的搜索范围。

9. 如权利要求 8 所述的音频编码装置，其特征在于，

所述搜索范围设定手段使用由所述音调分析手段获得相关值以及最近似于音调周期的值而求得成为搜索范围的中心的假设音调。

10. 如权利要求 9 所述音频编码装置，其特征在于，

所述搜索范围设定手段在假设音调周围的指定范围中设定滞后的搜索区间。

11. 如权利要求 8 所述的音频编码装置，其特征在于，

所述搜索范围设定手段使得滞后为较短候补的较少，并且设定滞后的搜索区间。

12. 如权利要求 8 所述的音频编码装置，其特征在于，

所述搜索范围设定手段在进行自调码本搜索时，在设定的范围中进行滞后搜索。

13. 一种由计算机可读的记录媒体，其特征在于，

所述记录媒体中存放了音频编码程序、存放了以前合成的声源信号的自调码本、存放了多个声源向量的概率性码本，

所述音频编码程序包含：对于存放在所述自调码本以及所述概率性码本中的自调声源以及概率性声源使用从输入音频中求得的 LPC 系数进行滤波而由此获得合成音的步骤；求取所述自调声源以及所述概率性声源的增益的步骤；使用所述输入音频与所述合成音之间的编码误差对于求得的自调声源、概率性声源以及所述增益进行向量量化的步骤，

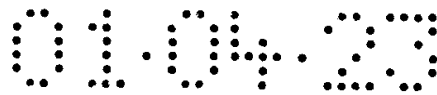
在进行向量量化步骤中包含：根据多个量化对象向量以及使用于预测编码

中的预测系数之间的编码误差而求得的量化对象向量的步骤；根据以前子帧的状态调整所述预测系数的步骤。

14. 一种计算机可读的记录媒体，其特征在于，

所述记录媒体中存放了音频编码程序、存放以前合成声的声源信号的自调码本、存放了多个声源向量的概率性码本，

所述音频编码程序包含：对于存放在所述自调码本以及所述概率性码本中的自调声源以及概率性声源使用从输入音频中求得的LPC系数进行滤波而由此获得合成音的步骤；求取所述自调声源以及所述概率性声源的增益的步骤；使用所述输入音频与所述合成音之间的编码误差对于求得的自调声源、概率性声源以及所述增益进行向量量化的步骤；在最初子帧的自调码本搜索之前分析构成帧的多个子帧的音调且求出相关值，并且使用所述相关值求出最近似于音调周期的值的步骤。



# 说明书

## 音频编码装置以及音频编码方法

### 技术领域

本发明涉及使用于数字通信系统的音频编码装置以及音频编码方法。

### 背景技术

在移动电话等数字通信系统的领域中，为了能够解决参加者增加的情况，寻求低位速率的音频压缩编码的方法，各研究机关正在继续开发该项研究。

在日本国内，采用作为数字移动电话用标准编码方法的

摩托罗拉公司开发的位速率 11.2kbps 称为 VSELP 的编码方法，采用相同方式的数字移动电话 1994 年秋季在日本国内开始发售。

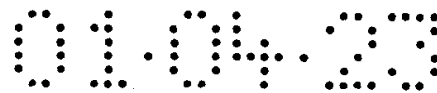
又，NTT 移动通信网股份公司开发的位速率 5.6kbps 的称为 PSI-CELP 编码方式正在制造中。这些方式中任意一个都是将称为 CELP(记载在 Code Excited Linear Prediction: M. R. Schroeder “High Quality Speech at Low Rates Bates” Proc. ICASSP’ 85pp. 937-940)方式改良后的方式。

该 CELP 方式将音频分离为声源信息以及声道信息，其特点在于，对于声源信息由存放在码本中的多个声源采样的指数进行编码，对于声道信息，采用将 LPC(线性预测系数)编码以及在声源信息编码时加入声道信息并与输入音频进行比较的方法(A-b-S:Analysis by synthesis)。

在该 CELP 方法中，首先对于输入的音频数据(输入音频)进行相关分析以及 LPC 分析而获得 LPC 系数，将获得的 LPC 系数编码并且获得 LPC 码。并且，将获得 LPC 码进行解码而获得解码 LPC 系数。另一方面，输入音频使用采用 LPC 系数的听觉加权滤波器来进行听觉加权。

对于自调码本与概率性码本存放的的声源采样(分别称为自调代码向量(或者自调声源))、概率代码向量(或者概率声源))各自的代码向量，根据获得解码 LPC 系数进行滤波并且获得 2 个合成音。

然后，分析获得的 2 个合成音与加权听觉后的输入音频的关系，求取 2 个合成音的最佳的值(最佳增益)，根据所求得的最佳增益调整合成音的功率，并将各自的合成音进行加法运算而获得综合合成音。之后，求出所获得的综合合成音与



输入音频之间的编码误差。如此，对于全体的声源采样，求得总和合成音与输入音频之间的编码误差，求取编码误差最小时声源采样的指数。

将这样获得增益以及声源采样的指数进行编码，将编码后的增益以及声源采样与 LPC 码一同传送到传送通道。又，从增益码与声源采样的指数所对应的的 2 个声源作成实际的声源信号，在将它存放在自调码本中的同时废除以前的声源采样。

又，一般地，对于自调码本以及概率码所进行的声源搜索是以将分析区进行细分后的区间(称为: subframe, 子帧)来进行。

增益的编码(增益量化)使用声源采样的指数所对应的 2 个合成音并根据评价增益的量化误差的向量量化(VQ)而进行。

在该算法中，预先作成存放了多个参数向量的代表性采样(代码向量)的向量码本。然后，对于听觉加权的输入音频、将自调声源与概率声源进行听觉加权 LPC 合成后的音频，使用存放在向量本码中的增益代码向量根据下述 1 式来计算编码误差。

$$E_n = \sum_{i=0}^I (X_i - g_n \times A_i - h_n \times S_i)^2 \quad \text{式 1}$$

这里，

$E_n$ : 使用了  $n$  个增益代码向量时的编码误差

$X_i$ : 听觉加权音频

$A_i$ : 听觉加权 LPC 合成后的自调声源

$S_i$ : 听觉加权 LPC 合成后的概率声源

$g_n$ : 代码向量的部分(自调声源侧的增益)

$h_n$ : 代码向量的部分(概率声源侧的增益)

$n$ : 代码向量的序号

$i$ : 声源数据的指数

$I$ : 子帧的长度(输入声频的编码单位)

其次，通过控制向量码本来比较使用了各代码向量时的误差  $E_n$ ，将最小误差的代码向量的序号作为向量的编码。又，求得存放在向量码本中的所有的代码向量中最小误差的代码向量的序号，并且将它作为向量的代码。

参照上述式 1 可以看到，对于每个  $n$  必须要进行较多的计算，而由于可以预先对于  $i$  进行求和计算，因此能够以较少量的计算来求得  $n$ 。



另一方面，在音频解码装置(decoder)中，根据传送来的向量的代码，通过求得代码向量而将编码的数据进行解码并获得代码向量。

又，以上述的算法为基础，进行了基于以往的改良。例如，利用人们声压的听觉特性为对数这一点，取功率的对数并进行量化，使该功率下标准化的 2 个增益为 VQ。该方法是使用日本 PDC 半速率编码(half rate coding)的标准方式的方法。此外，有利用增益参数的帧间相关进行编码的方法(预测编码)。该方法是使用了 ITU-T 国际标准 G.729 的方法。但是，通过这些改良也不能够获得非常好的性能。

至今，人们开发了利用了人们听觉特性以及帧间相互关系的增益信息编码方法，可以进行效率较高的编码。特别地，由于预测量化而大大地提高了性能，而在以往的方法中，作为状态的值使用以往子帧的值并进行预测量化。但是，在作为表示状态而被存放的值中，有时会获取其中最大(小)的值，当将该值使用于下一个子帧时，并不能很好地进行子帧的量化，有时在局部位置上会有杂音。

### 发明内容

本发明的目的是提供一种利用预测量化而能够进行局部不会产生杂音的音频编码的 CELP 型音频编码装置以及方法。

本发明的主要内容是在预测量化中当以前子帧中的状态值为极大值或极小值时通过自动地调制预测系数能够防止产生局部的杂音。

### 附图简述

图 1 是表示具备本发明的音频编码装置的无线通信装置的构造的框图。

图 2 是表示本发明实施形态 1 的音频编码装置的构造的框图。

图 3 是表示图 2 所示的音频编码装置的增益运算部分的构造的框图。

图 4 是表示图 2 所示的音频编码装置参数编码部分的框图。

图 5 是表示本发明实施形态 1 的音频编码装置中将编码后的音频数据进行解码的音频解码装置构造的框图。

图 6 用于说明自调码本搜索。

图 7 是表示本发明实施形态 2 的音频编码装置构造的框图。

图 8 用于说明脉冲扩散码本的框图。

图 9 是表示脉冲扩散码本的详细构造一示例的框图。

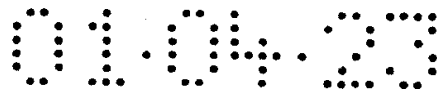


图 10 是表示脉冲扩散码本的详细构造一示例的框图。

图 11 是表示本发明实施形态 3 的音频编码装置构造的框图。

图 12 是表示在本发明实施形态 3 的音频编码装置中将编码后的音频数据进行解码的音频解码装置构造的框图。

图 13A 表示在本发明实施形态 3 的音频编码装置中使用的脉冲扩散码本的一示例。

图 13B 表示在本发明实施形态 3 的音频解码装置中使用的脉冲扩散码本一示例。

图 14A 表示在本发明实施形态 3 的音频编码中使用的脉冲扩散码本的一示例。

图 14B 表示在本发明实施形态 3 的音频解码装置中使用的脉冲扩散码本的一示例。

### 最佳实施形态

以下，参照附图对于本发明的实施形态进行详细的说明。

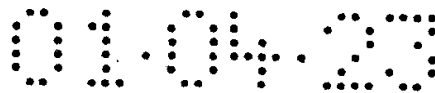
#### (实施形态 1)

图 1 是表示具备本发明实施形态 1~3 的音频编码装置的无线通信装置构造的框图。

在该无线通信装置中，发送侧通过麦克风等的音频输入装置 11 将音频变换为电的模拟信号，并且输出到 A/D 变换器 12。模拟音频信号通过 A/D 变换器 12 而变换为数字音频信号并且输出到音频编码部分 13。音频编码部分 13 对于数字音频信号进行音频编码处理并且将编码后的信息输出到调制解调部分 14。调制解调部分 14 将编码后的音频信号进行数字调制并送入到无线发送部分 15。在无线发送部分 15 中，对于调制后的信号进行规定的无线发送处理。该信号通过天线 16 被发送。又，信息处理器 21 使用合适的存放在 RAM22 以及 ROM23 中的数据进行处理。

另一方面，在无线通信装置的接收侧，由天线 16 接收的信号在无线接收部分 17 受到规定的无线接收处理并送到调制解调部分 14。在调制解调部分 14 中，对于接收信号进行解调处理并将解码后的信号输出到音频解码部分 18。音频解码部分 18 对于解调后的信号进行解码处理而获得数字解码音频信号，并且将该数字解码音频信号输出到 D/A 变换器 19。D/A 变换器 19 将由音频解码部分 18 输出的





数字解码音频信号变换为模拟解码音频信号并且输出到扬声器等的音频输出装置 20。最后，音频输出装置将电的模拟解码音频信号变换为解码音频而输出。

这里，音频编码部分 13 以及音频解码部分 18 使用存放在 RAM22 以及 ROM23 中的码本并通过 DSP 等的信息处理器 21 进行动作。又，这些动作程序存放在 ROM23 中。

图 2 表示本发明实施形态 1 的 CELP 型音频编码装置的构造的框图。该音频编码装置包含在图 1 所示音频编码部分 13 中。又，图 2 所示自调的码本 103 存放在图 1 所示 RAM22 中，图 2 所示概率性码本 104 存放在图 1 所示 ROM23 中。

在图 2 所示音频编码装置中，在 LPC 分析部分 102 中，对于输入的音频数据 101 进行自相关分析以及 LPC 分析而获得 LPC 系数。又，在 LPC 分析部分 102 中，将获得的 LPC 系数编码并获得 LPC 码。而且在 LPC 分析部分 102 将得到的 LPC 码进行解码并获得解码 LPC 系数。将输入的音频数据 101 送到听觉加权部分 107，这里采用利用了上述 LPC 系数的听觉加权滤波器来进行听觉加权。

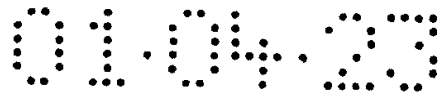
其次，在声源作成部分 105 中，取出存放于自调码本 103 中的音源采样（自调代码向量或自调音源）与存放于概率性码本 104 中的音源采样（概率性代码向量或概率性音源），将各自的代码向量送到听觉加权 LPC 合成部分 106。而且在听觉加权 LPC 合成部分 106 中对于由音源作成部分获得的 2 个音源，根据由 PLC 分析部分 102 获得的解码 LPC 系数进行滤波，获得 2 个合成音。

再在听觉加权 LPC 合成部分 106 中一并使用采用 LPC 系数、高频加强滤波器或长期预测系数（通过对输入音频的长期预测分析得到）的听觉加权滤波器，对各合成音进行听觉加权 LPC 合成。

听觉加权 LPC 合成部 106 将 2 个合成音输出到增益运算部分 108。增益运算部分 108 具有图 3 所示的构造。在增益运算部分 108 中，将在听觉加权 LPC 合成部分 106 获得的 2 个合成音以及听觉加权的输入音频送到分析部分 1081 并且分析 2 个合成音与输入音频之间的关系，求得 2 个合成音的最佳值（最佳增益）。将该最佳增益输出到功率调整部分 1082。

在功率调整部分 1082 中，根据求得的最佳增益调整 2 个合成音的功率。将进行功率调整后的合成音输出到合成部分 1083，在合成部分 1083 进行加法运算并形成综合合成音。该综合合成音被输出到编码误差运算部分 1084。在编码误差运算部分 1084 中，求得获得的综合合成音与输入音频之间的编码误差。

编码误差运算部分 1084 控制声源作成部分 105，使得输出自调码本 103 以及



概率性码本 104 的所有的音频采样, 对于所有的声源采样求出综合合成音与输入间频之间的编码误差, 求出编码误差最小时的声源采样的指数。

其次, 分析部分 1081 将声源采样的指数、对应于该指数的 2 个听觉加权 LPC 合成的声源以及输入音频发送到参数编码部分 109。

在参数编码部分 109 中, 利用将增益码而获得增益码并且将 LPC 码、声源采样的指数总和起来发送到传送通道。又, 从增益码与指数所对应的 2 个声源作成实际声源的信号, 并在将它存放在自调码本 103 中的同时废除以往的声源采样。又, 一般地, 对于自调码本与概率码本所对应的声源搜寻是以将分析区间进一步细分而获得区间(称为子帧)来进行的。

这里, 对于具有上述构造的音频编码装置的参数编码部分 109 的增益码的动作进行说明。图 4 是表示本发明音频编码装置的参数编码部分构造的框图。

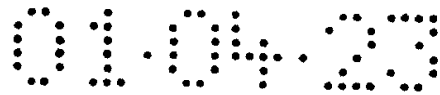
在图 4 中, 将听觉加权输入音频( $X_i$ )、听觉加权 LPC 合成后的自调声源( $A_i$ )以及听觉加权 LPC 合成后的概率声源( $S_i$ )发送到参数计算部分 1091。在参数计算部分 1091 中, 算出编码误差计算所必须的参数。在参数计算部分 1091 计算出的参数被输出到编码误差计算部分 1092 并且在此计算编码误差。该编码误差被输出到比较部分 1093。在比较部分 1093 中, 控制编码误差计算部分 1092 以及向量码本 1094, 从获得编码误差中求得最佳的编码(解码向量), 根据该编码将从向量码本 1094 中获得的代码向量输出到解码向量存放部分 1096 并且更新解码向量存放部分 1096。

预测系数存放部分 1095 存放用于预测编码的预测系数。由于该预测系数是用于参数计算以及编码误差计算中, 故将它输出到参数计算部分 1091 以及编码误差计算部分 1092。解码向量存放部分 1096 为进行预测编码而存放状态。由于该状态用于参数计算, 故将该状态输出到参数计算部分 1091。向量码本 1094 存放代码向量。

其次, 对于本发明的增益码方法的算法进行说明。

首先, 作成存放了多个量化对象向量代表性采样(代码向量)的向量码本 1094。各个向量由 AC 增益、SC 增益的指数值所对应的值以及 SC 预测系数的调整系数这 3 个部分形成。

该调整系数是根据以前子帧的状态来调整预测系数的系数。具体地, 当以前的子帧的状态为最大值或者最小值时, 设定该调整系数使得它们的影响变小。能够利用由本发明者所提出的使用了多个向量采样的研究算法而求出该调整系数。



这里，省略对于学习算法的说明。

例如，使用于音频的次数多的代码向量设定调整系数为较大。即，当相同波形并排时，因先前的子帧状态的可靠性高而使得调整系数较大，可以继续利用先前的子帧预测系数。由此，能够进行更有效的预测。

另一方面，对于使用于语音首部等的使用频率较小的代码向量使得调整系数较小。即，当与前次波形完全不相同时，因先前的子帧的状态可靠性低（考虑自调码本没有起作用），则使得调整系数变小，减小先前子帧的预测系数的影响。由此，能够防止下次预测的失败并且能够实现良好的预测编码。

如此，通过根据各代码向量(状态)来控制预测系数，则能够进一步提高预测编码的性能。

又，在预测系数存放部分 1095 中，预先存放了用于进行预测编码的预测系数。该预测系数为 MA(moving average, 移动平均数)的预测系数并且按预测次数存放 AC 与 SC 的 2 个种类。又，一般地通过预先使用了大量数据的研究求得这些数据。又，在解码向量存放部分 1096 中，作为初值预先存放了表示无声状态的值。

其次，对于编码方法进行详细地说明。首先，向参数计算部分 1091 送入听觉加权输入音频( $X_i$ )、听觉加权 LPC 合成后的自调音源( $A_i$ )、听觉加权 LPC 合成后的概率音源( $S_i$ )，并且送入存放在解码的向量存放部分 1096 中的解码向量(AC、SC、调整系数)、存放在预测系数存放部分 1095 中的预测系数(AC、SC)。使用这些数据计算出编码误差计算所必要的参数。

编码误差计算部分 1092 的编码误差计算按照下式 2 进行。

$$E_n = \sum_{i=0}^I (X_i - G_{an} \times A_i - G_{sn} \times S_i)^2 \quad \text{式 2}$$

这里，

$G_{an}$ ,  $G_{sn}$ : 解码增益

$E_n$ : 使用  $n$  个增益代码向量时的编码误差

$X_i$ : 听觉加权音频

$A_i$ : 听觉加权 LPC 合成后的自调声源

$S_i$ : 听觉加权 LPC 合成后的概率声源

$n$ : 代码向量的序号

$i$ : 音源向量的指数

$I$ : 子帧的长度(输入音频的编码单位)

此时，由于运算量较少，在参数计算部分 1091 中，进行不依赖于代码向量部分的计算。计算好的数据是上述预测向量与 3 个合成音的  $(X_i, A_i, S_i)$  间的相关值、功率。该计算按照下述式 3 进行。

$$\begin{aligned}
 D_{xx} &= \sum_{i=0}^I X_i \times X_i \\
 D_{xa} &= \sum_{i=0}^I X_i \times A_i \times 2 \\
 D_{xs} &= \sum_{i=0}^I X_i \times S_i \times 2 \\
 D_{aa} &= \sum_{i=0}^I A_i \times A_i \\
 D_{as} &= \sum_{i=0}^I A_i \times S_i \times 2 \\
 D_{ss} &= \sum_{i=0}^I S_i \times S_i
 \end{aligned}
 \tag{式 3}$$

$D_{xx}, D_{xa}, D_{xs}, D_{aa}, D_{as}, D_{ss}$ : 合成音之间的相关值、功率

$X_i$ : 听觉加权音频

$A_i$ : 听觉加权 LPC 合成后的自调声源

$S_i$ : 听觉加权 LPC 合成后的概率声源

$n$ : 代码向量的序号

$i$ : 声源向量的指数

$I$ : 子帧的长度(输入音频的编码单位)

又，在参数计算部分 1091 中，使用存放在解码向量存放部分 1096 中的以前的代码向量、存放在预测系数存放部分 1095 中的预测系数而预先进行下述式 4 所示的 3 个预测值的计算。

$$\begin{aligned}
 P_{ra} &= \sum_{m=0}^M \alpha_m \times S_{am} \\
 P_{rs} &= \sum_{m=0}^M \beta_m \times S_{cm} \times S_{sm} \\
 P_{sc} &= \sum_{m=0}^M \beta_m \times S_{cm}
 \end{aligned}
 \tag{式 4}$$

这里，

$P_{ra}$ : 预测值(AC 增益)

$P_{rs}$ : 预测值(SC 增益)

$P_{sc}$ : 预测值(预测系数)

$\alpha_m$ : 预测系数 (AC 增益、固定值)

$\beta_m$ : 预测系数 (SC 增益、固定值)

$S_{am}$ : 状态 (以前的代码向量部分、AC 增益)

$S_{sm}$ : 状态 (以前的代码向量部分、SC 增益)

$S_{cm}$ : 状态 (以前的代码向量部分、SC 预测系数调整系数)

$m$ : 预测指数

$M$ : 预测次数

从上述 4 式可知, 对于  $Pr_s$ 、 $Psc$ , 乘以与以往不同的调整系数。因此, 对于 SC 增益的预测值以及预测系数, 根据调整系数, 当先前的子帧的状态值为最大或最小时, 能够减缓它们 (减小影响)。即, 根据状态, 能够合适地改变 SC 增益的预测值以及预测系数。

其次, 在编码误差运算部分 1092 中, 使用参数计算部分 1091 所计算的参数、预测系数存放部分 1095 中存放的预测系数、向量码本 1094 中存放的代码向量, 根据下述式 5 计算出编码误差。

$$E_n = D_{xx} + (G_{an})^2 \times D_{aa} + (G_{sn})^2 \times D_{ss} - G_{an} \times D_{xa} - G_{sn} \times D_{xs} + G_{an} \times G_{sn} \times D_{as}$$

$$G_{an} = P_{ra} + (1 - P_{ac}) \times C_{an}$$

$$G_{sn} = 10^{\{Pr_s + (1 - P_{sc}) \times C_{sn}\}} \quad \text{式 5}$$

这里

$E_n$ : 使用  $n$  号的增益代码向量时的编码误差

$D_{xx}$ ,  $D_{xa}$ ,  $D_{xs}$ ,  $D_{aa}$ ,  $D_{as}$ ,  $D_{ss}$ : 合成音间的相关值、功率

$G_{an}$ ,  $G_{sn}$ : 解码增益

$P_{ra}$ : 预测值 (AC 增益)

$P_{rs}$ : 预测值 (SC 增益)

$P_{ac}$ : 预测系数的和 (固定值)

$P_{sc}$ : 预测系数的和 (按上述式 4 算出)

$C_{an}$ ,  $C_{sn}$ ,  $C_{cn}$ : 代码向量、 $C_{cn}$  预测系数调整系数而这里不使用

$n$ ; 代码向量的序号

又, 由于实际上  $D_{xx}$  不依赖于代码向量的序号  $n$ , 能够省略它的加法运算。

其次, 比较部分 1093 控制向量码本 1094 与编码误差计算部分 1092, 在存放在向量码本 1094 中的多个代码向量中, 求出利用编码误差计算部分 1092 计算出的编码误差的最小的代码向量的序号, 将它作为增益的代码。又, 使用获得增益



的代码来更新解码向量存放部分 1096 的内容。根据下述 6 式进行更新。

$$S_{am} = S_{am-1} (M=M \sim 1), S_{a0} = C_a J$$

$$S_{sm} = S_{sm-1} (M=M \sim 1), S_{s0} = C_s J$$

$$S_{cm} = S_{cm-1} (M=M \sim 1), S_{c0} = C_c J$$

式 6

这里,

$S_{am}$ ,  $S_{sm}$ ,  $S_{cm}$ : 状态向量 (AC、SC、预测系数调整系数)

m: 预测指数

M: 预测次数

J: 在比较部分求出的编码

从式 4 到式 6 可知, 在本实施形态中, 在解码向量存放部分 1096 中预先存放状态向量  $S_{cm}$ , 使用该预测系数调整系数来合适地控制预测系数。

图 5 是表示本发明实施形态的音频解码装置的构造的框图。该音频解码装置包含在图 1 所示的音频解码部分 18 中。又, 图 5 所示的自调码本 202 存放在图 1 所示的 RAM22 中, 图 5 所示的概率码本 203 存放在图 1 所示的 ROM23 中。

在图 5 所示的音频解码装置中, 参数解码部分 201 在获得编码音频信号的同时从传送通道获得各音源码本 (自调码本 202、概率性码本 203) 的音源采样的代码、LPC 代码以及增益代码。然后, 从 LPC 码中获得解码后的 LPC 系数, 从增益码中获得解码增益。

然后, 音源作成部分 204 对各自的音源采样乘上解码后的增益并且进行加法运算, 由此获得解码后的音源信号。此时, 将获得的解码后的音源信号作为音源采样存放在自调码本 204 中, 同时废除旧的音源采样。这样, 在 LPC 合成部分 205 中, 对于解码后的音源信号进行根据解码后的 LPC 系数的滤波, 由此获得合成音。

又, 2 个音源码本与图 2 所示的音频编码装置中所含有的码本 (图 2 的参照符号 103, 104) 是相同的, 用于取出音源采样的采样序号 (输入自调码本的代码以及概率性码本的代码) 都是由参数解码部分 201 提供的。

如此, 在本实施形态的音频编码装置中, 根据各代码向量能够控制预测系数, 能够根据音频局部的特征进行合适有效的预测并且能够防止非稳态部位预测的失败, 能够获得前所未有的良好效果。

(实施形态 2)

在音频编码装置中, 如上所述, 在增益运算部分中, 对于从音源作成部分获得自调码本、概率码本的所有音源进行合成音与输入音频之间的比较。此时, 在

运算量上，通常开环搜索 2 个音源(自调码本与概率性码本)。以下，参照图 2 进行说明。

在该开环搜寻中，首先，声源作成部分 105 仅从自调码本 103 中一个接一个地选出候补音源，使得听觉加权 LPC 合成部分 106 进行工作而获得合成音并且送入增益运算部分算 108，比较合成音与输入音频并选择最佳的自调码本 103 的代码。

其次，固定上述自调码本 103 的代码，从自调码本 103 选择相同的音源，从概率性码本 104 中一个接一个地选择运算部分 108 的代码所对应的声源并且传送到听觉加权 LPC 合成部分 106。在增益运算部分 108 中，比较两合成音的和与输入音频，决定概率性码本 104 的代码。

当运用该算法时，分别地搜索所有的码本的代码，由此会引起个别代码性能的劣化，而大幅度地消减地计算量。因此，一般使用这种开环搜索。

这里，对于以往开环的音源搜索中代表性的算法进行说明。这里，对于 1 个分析区间(帧)，由 2 个子帧构成时的音源搜索的顺序进行说明。

首先，接受增益运算部分 108 的指示，音源作成部分 105 从自调码本 103 中引出音源并且送听觉加权 LPC 合成部分 106。在增益运算部分 108 中，反复进行合成的音源与第 1 子帧的输入音频之间的比较并且求得最佳代码。这里，表示自调码本的特征。自调码本是以往使用于合成中的音源。这样，代码对应于图 6 所示的时间滞后。

其次，当决定了自调码本 103 的代码之后，进行概率性码本的搜索。音源作成部分 105 取出通过搜索自调码本 103 而获得的编码的音源以及由增益运算部分 108 所指定的概率性码本 104 的音源并且送到听觉加权 LPC 合成部分 106。然后，在增益运算部分 108 中，计算出听觉加权后的合成音与听觉加权后的输入音频之间的编码误差，决定最适当的(二乘误差为最小的)概率音源 104 的代码。以下表示一个分析区间(子帧为 2 时)中的音源代码的搜索顺序。

- 1) 决定第 1 子帧的自调码本的代码
- 2) 决定第 1 子帧的概率性码本的代码
- 3) 在参数编码部分 109 中将增益代码，利用解码的增益作成第 1 子帧的音源并且更新自调码本 103。

- 4) 决定第 2 子帧的自调码本的代码
- 5) 决定第 2 子帧的概率码本的代码

6) 在参数编码部分 109 中将增益码并且利用解码的增益作成第 2 子帧的音源并且更新自调码本 103。

根据上述的算法，能够进行更有效的音源编码。但是，最近，还希望更低的位速率并且使得音源的位数更少。特别注目的是利用与自调码本的滞后非常相关的这一点，该算法是保持第 1 子帧的代码不变，压缩第 2 子帧的搜索范围接近第 1 子帧的滞后(减少输入端)并且使得位数变少。

在该算法中，考虑到了当分析区间(帧)的途中音频发生变化的情况以及 2 个子帧大小不同的情况时局部地区会引起劣化的情况。

在本实施形态中提供了一种实现搜索方法的音频编码装置，该搜索方法是在编码前对于 2 个子帧双方进行间距分析算出相关值，根据获得相关值决定 2 个子帧的滞后的搜索范围。

具体地，本实施形态的音频编码装置是将 1 帧分解成多个子帧并且分别将它们编码的 CELP 型编码装置，其特点在于，它具备在最初的子帧的自调码本搜索之前对构成一帧的多个子帧进行间距分析并且算出相关值的间距分析部分、在上述间距分析部分算出构成一帧的多个子帧的相关值的同时从其相关值的大小求出各子帧中最小间距周期值(称为代表间距)并且根据间距分析部分获得相关值与代表间距来决定多个子帧滞后的搜索范围的搜索范围设定部分。并且，对于该音频编码装置，在搜索范围设定部分中，利用由间距分析部分获得多个子帧的代表间距与相关值而求得作为搜索范围中心假设的间距(称为假设音调)，在搜索范围设定部分，在求得的假设音调的周围指定范围中设定滞后的搜索区间并且在设定滞后的搜索区间时设定向假设音调前后的搜索范围。此时，滞后较短部分的候补较少，设定滞后为更长的范围并且在自调码本搜索时在由上述搜索范围设定部分设定的范围中进行滞后的搜索。

以下，对于本实施形态的音频编码装置参照附图进行详细地说明。这里，将 1 帧分为 2 个子帧。即使分割为 3 帧以上的情况下，也能够按照相同的顺序进行编码。

在该音频编码装置中，即在根据 $\Delta$ 滞后方式的间距搜索中，对于分割后的子帧求出所有的间距，并且求出各间距间存在多大程度的相关，根据该相关结果决定搜索范围。

图 7 表示本发明实施形态 2 的音频编码装置的构造的框图。首先，在 LPC 分析部分 302 中，对于输入的音频数据(输入音频)301 进行自相关分析与 LPC 分析，



由此获得 LPC 系数。又，在 LPC 分析部分 302 中，将获得 LPC 系数编码并且获得 LPC 代码。而且，在 LPC 分析部分 302 中，将获得 LPC 代码解码并求得解码 LPC 系数。

其次，在间距分析部分 310 中，进行 2 个子帧份额的输入音频的间距分析，求得间距候补以及参数。1 个子帧所对应的算法如下所示。根据下述式 7 可以求得 2 个相关系数。又，此时对于  $C_{pp}$  首先求得  $P_{min}$ ，对于后面的  $P_{min+1}$ 、 $P_{min+2}$  能够利用帧端部的值来有效地进行计算。

$$V_p = \sum_{i=0}^L X_i \times X_{i-P} - P \quad (P=P_{min}-P_{max})$$

$$C_{pp} = \sum_{i=0}^L X_i - P \times X_{i-P} - P \quad (P=P_{min}-P_{max}) \quad \text{式 7}$$

这里，

$X_i, X_{i-P}$ : 输入音频

$V_p$ : 自相关函数

$C_{pp}$ : 功率成分

$i$ : 输入音频的采样序号

$L$ : 子帧的长度

$P$ : 间距

$P_{min}, P_{max}$ : 进行间距搜索的最小值与最大值

这样，由上述式 7 求得的自相关函数与功率成分存储在存储器中，接着求出代表间距  $P_1$ 。这是求得使得  $V_p$  为正并且  $V_p \times V_p / C_{pp}$  为最大的间距  $P$  的处理。然而，由于除法计算一般需要较大的计算量，要存放分子与分母两者，将它变换为乘法则能够提高效率。

这里，寻找输入音频与从输入音频起经过间距的自调音源的差分的平方和为最小的间距。该处理与求出使得  $V_p \times V_p / C_{pp}$  为最大的间距  $P$  的处理等价。具体的处理如下所示。

- 1) 初始化 ( $P=P_{min}$ 、 $VV=C=0$ 、 $P_1=P_{min}$ )
- 2) 若 ( $V_p \times V_p \times C < VV \times C_{pp}$ ) 或者 ( $V_p < 0$ )，则转向 4)。否则，转向 3)。
- 3)  $VV = V_p \times V_p$ 、 $C = C_{pp}$ 、 $P_1 = P$  转向 4)
- 4)  $P = P + 1$ 。此时，若  $P > P_{max}$  则结束，否则转向 2)。

对于 2 个子帧分别进行上述工作，求出代表间距  $P_1$ 、 $P_2$  以及相关系数  $V_{1p}$ 、 $V_{2p}$ 、功率成分  $C_{1pp}$ 、 $C_{2pp}$  ( $P_{min} < P < P_{max}$ )。

接着，在搜索范围设定部分 311 中设定自调码本的滞后的搜索范围。首先，求得作为该搜索范围的轴的间距。假设音调使用由间距分析部分 310 所求得的代表间距与参数。

按照下以下顺序求出假设音调  $Q_1$ 、 $Q_2$ 。又，在以下的说明中，作为滞后的范围使用常数  $Th$  (具体地相当于 6 的程度)。又，相关值采用由上述式 7 求得值。

首先，在固定了  $P_1$  的状态下  $P_1$  的附近 ( $\pm Th$ ) 寻找相关的最大假设音调 ( $Q_2$ )。

- 1) 初始化 ( $p=P_1-Th$ 、 $C_{max}=0$ 、 $Q_1=P_1$ 、 $Q_2=P_1$ )
- 2) 若  $(V_{1p1} \times V_{1p1}/C_{1p1p1} + V_{2p} \times V_{2p}/C_{2pp} < C_{max})$  或者  $(V_{2p} < 0)$ ，则转向 4)。否则，转向 3)。
- 3)  $C_{max}=V_{1p1} \times V_{1p1}/C_{1p1p1} + V_{2p} \times V_{2p}/C_{2pp}$ 、 $Q_2=p$  转向 4)
- 4)  $P=P+1$  转向 2)。但是，若此时  $p > P_1+Th$  则转向 5)。

如此，进行 2)~4) 的处理直到  $P_1-Th \sim P_1+Th$ ，求得相关的最大的  $C_{max}$  与假设音调  $Q_2$ 。

其次，在固定了  $P_2$  的状态下，在  $P_2$  的附近 ( $\pm Th$ ) 求得最大的假设音调 ( $Q_1$ )。此时， $C_{max}$  不需要进行初始化。包含求得  $Q_2$  时的  $C_{max}$  并且求得相关为最大的  $Q_1$ ，由此，能够求得在第 1，第 2 子帧之间带有最大的相关的  $Q_1$ 、 $Q_2$ 。

- 5) 初始化 ( $p=P_2-Th$ )
- 6) 若  $(V_{1p} \times V_{1p}/C_{1pp} + V_{2p2} \times V_{2p2}/C_{2p2p2} < C_{max})$  或者  $(V_{1p} < 0)$ ，则转向 8)。否则，转向 7)。
- 7)  $C_{max}=V_{1p} \times V_{1p}/C_{1pp} + V_{2p2} \times V_{2p2}/C_{2p2p2}$ 、 $Q_1=p$ 、 $Q_2=P_2$  转向 8)
- 8)  $P=P+1$  转向 6)。但是，此时，若  $p > P_2+Th$  转向 9)。
- 9) 结束。

如此，进行 6)~8) 的处理直到  $P_2-Th \sim P_2+Th$ ，求得相关的最大值的  $C_{max}$  与假设音调  $Q_1$ 、 $Q_2$ 。此时的  $Q_1$ 、 $Q_2$  是第 1 子帧与第 2 子帧的假设音调。

根据上述的算法，同时地评价 2 个子帧的相关并且能够选择 2 个大小上没有较大相差 (相差最大为  $Th$ ) 的假设音调。通过利用该假设音调，在搜索第 2 子帧的自调码本时，即使设定搜索范围较窄，也能够防止编码性能较大的劣化。例如，当从第 2 子帧起音质急剧发生变化等的情况下，第 2 子帧的相关较强时通过利用反映第 2 子帧的自调码本的相关的  $Q_1$  而能够避免第 2 子帧的劣化。

而且，搜索范围设定部分 311 使用求得的假设音调  $Q_1$  如下述式 8 设定进行自调码本搜索的范围 ( $L_{ST} \sim L_{EN}$ )。

## 第 1 子帧

$$L\_ST=Q1-5 \text{ (而 } L\_ST < L_{min} \text{ 时 } L\_ST=L_{min})$$

$$L\_EN=L\_ST+20 \text{ (而 } L\_ST > L_{max} \text{ 时 } L\_ST=L_{max})$$

## 第 2 子帧

$$L\_ST=T1-10 \text{ (而 } L\_ST < L_{min} \text{ 时 } L\_ST=L_{min})$$

$$L\_EN=L\_ST+21 \text{ (而 } L\_ST > L_{max} \text{ 时 } L\_ST=L_{max})$$

这里，

$L\_ST$ : 最小搜索范围

$L\_EN$ : 最大搜索范围

$L_{min}$ : 滞后的最小值(例: 20)

$L_{max}$ : 滞后的最大值(例: 143)

$T_1$ : 第 1 子帧的自调码本滞后

在上述设定中，不必使得第 1 子帧的搜索范围很小。然而，本发明者们通过实验确认，将根据输入音频间距的值的附近作为搜索区间性能则更好，在本实施形态中，采用压缩到 26 个采样进行搜索的算法。

又，第 2 子帧将第 1 子帧所求得的滞后  $T_1$  为中心，设定此附近的搜索范围。然而，总共 32 个记录下，能够以 5 个比特将第 2 子帧的自调码本的滞后进行编码。又，本发明者们通过设定此时滞后小的候补较少而滞后大的候补较大，通过实验确认能够获得更加好的性能。然而，在本实施形态中，为了使得清楚地理解本发明而没有使用假设音调  $Q_2$ 。

这里，对于本实施形态的效果进行说明。在由搜索范围设定部分 311 获得的第 1 子帧的假设音调附近还存在第 2 子帧的假设音调(因由常数  $T_h$  进行了限制)。又，在第 1 子帧缩小搜索范围而进行了搜索，则通过搜索结果而能获得滞后并没有从第 1 子帧的假设音调中分离出来。

因此，在进行第 2 子帧的搜索时，通过能够搜索第 2 子帧的假设音调的附近范围，对于第 1，第 2 子帧两者能够搜索适当的滞后。

作为示例，研究第 1 子帧为无声而从第 2 子帧器发出声音的情况。在以往的方法中，因缩小搜索范围使得第 2 子帧的间距不包含在搜索区间之中，则音质会发生较大的劣化。在本实施形态的方法中，在间距分析部分的假设音调分析中，代表间距  $P_2$  的相关变强。因此，第 1 子帧的假设音调为  $P_2$  附近的值。因此，当根据  $\Delta$  滞后进行搜索时，能够在发出声音的部分使得附近部分为假设音调。即，

在搜索第 2 子帧的自调码本时，能够搜索 P2 附近的值，即使在当中产生声音也不会发生劣化，能够根据 $\Delta$ 滞后进行第 2 子帧的自调码本的搜索。

其次，在音源作成部分 305 中，取出存放在自调码本 303 中音源采样(自调代码向量或者自调音源)以及存放在概率性码本 304 中的音源采样(概率代码向量或者概率音源)，并且将它们分别送入听觉加权 LPC 合成部分 306。而且，在听觉加权 LPC 合成部分 306 中，对于音源作成部分 305 获得 2 个音源，按照 LPC 分析部分 302 获得解码 LPC 系数进行滤波并且合成 2 个合成音。

并且，在增益运算部分 308 中，分析由听觉加权 LPC 合成部分 306 获得 2 个合成音与输入音频的关系，求出 2 个合成音的最佳值(最佳增益)。又，在增益运算部分 308 中，根据该最佳增益将分别调整了功率的合成音进行加法运算而获得总和合成音。然后，增益运算部分 308 计算该总和合成音与输入音频的编码误差。又，在增益运算部分 308 中，对于自调码本 303 与概率性码本 304 所有的声源采样，计算通过使得声源作成部分 305、听觉加权 LPC 合成部分 306 作用而获得多个合成音与输入音频之间的编码误差，求得在获得的结果中编码误差最小时的声源采样的指数。

其次，将获得音源采样的指数、该指数所对应的 2 个声源以及输入音频送入到参数编码部分 309。在参数编码部分 309 中，通过将增益进行编码而获得增益码，将 LPC 码、声源采样的指数一同送入传送通道。

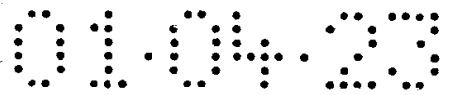
又，参数编码部分 309 从增益码与声源采样的指数所对应的 2 个声源中作成实际的声源信号，在将它存放在自调码本 303 的同时废除旧的声源采样。

又，在听觉加权 LPC 合成部分 306 中，采用使用了 LPC 系数、高频滤波器以及长期预测系数(通过进行输入音频的长期预测分析而获得)的听觉加权滤波器。

上述增益运算部分 308 将从声源作成部分 305 获得自调码本 303 与概率性码本 304 所有声源的输入音频之间进行比较，而为了减少计算量，对于 2 个声源(自调码本 303 与概率性码本 304)利用上述的开环进行搜索。

如此，根据本实施形态的间距搜索方法，在最初的采样的自调码本搜索之前，对于构成帧的多个子帧的间距进行分析而计算出相关值，由此能够同时地把握一帧内的所有子帧的相关值。

这样，在计算出各子帧的相关值的同时，从该相关值的大小中求得子帧中近似间距周期的值(称为代表间距)，根据间距分析获得相关值与代表间距，设定多个子帧的滞后的搜索范围。在该搜索范围的设定中，利用间距分析所获得的多个



子帧的代表间距与相关值并且求得搜索范围中心相差较小的适当的假设音调(称为假设音调)。

而且, 由于在所述搜索范围的设定中所求得的假设音调的前后的指定范围中限定了滞后的搜索区间, 则能够进行更加高效的自调码本的搜索。此时, 由于使得滞后短的部分的候补较少并且设定滞后为更加长的范围, 故能够设定可以获得良好性能的适当的搜索范围。又, 在进行自调码本搜索时, 在上述搜索范围设定中设定的范围中由于进行滞后的搜索, 能够进行可获得良好解码后声音的解码。

如此, 根据本实施形态, 在由搜索范围设定部分 311 获得第 1 子帧的假设音调的附近, 也存在第 2 子帧的假设音调, 在第 1 子帧中, 由于缩小了搜索范围, 作为搜索结果而获得的滞后没有远离假设音调。因此, 在进行第 2 子帧的搜索时, 能够搜索第 2 子帧的假设音调的附近, 即使对于从帧的后半部分发出声音等的不稳定子帧, 在第 1、第 2 子帧中也能够进行适当的搜索, 能够获得前所未有的良好效果。

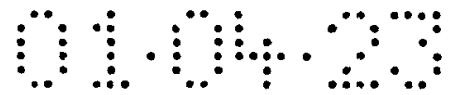
### (实施形态 3)

在初期的 CELP 方式中, 将随机数列作为概率性声源向量使用登记了多种类型的概率性码本, 即使用直接记录了多个种类型的随机数的概率编码。另一方面, 对于在近年的低位速率 CELP 编码·解码装置, 大量开发了概率性码本部分具备代数性码本的装置, 该代数性码本生成含有少数振幅为 +1 或 -1 的非零部分(非零部分以外的部分振幅为零)的概率性声源向量。

又, 代数性码本如“Fast CELP Coding based on Algebraic codes”, J. Adoul et al, Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, 1987, pp. 1957-1960 以及“Comparison of some Algebraic Structure for CELP Coding of Speech”, J. Adoul et al, Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, 1987, pp. 1953-1956 等所揭示的。

上述文献所揭示的代数性码本具有下述优点, (1) 当适用于位速率为 8kb/s 程度的 CELP 方式的情况下, 能够生成高质量的合成音, (2) 以比较少的运算量能够搜索概率声源码本, (3) 不需要直接存放概率性声源向量的数据 ROM。

这样, 将代数性码本作为概率性码本使用的 CS-ACELP(位速率为 8kbs/s) 以及 ACELP(位速率为 5.3kb/s) 作为 G. 729、g723.1, 分别在 1996 从 ITU-T 中被推崇。又, 关于 CS-ACELP, 在“Design and Description of CS-ACELP: A Toll Quality 8 kb/s Speech coder”, Redwan Salami et al, IEEE trans. SPEECH AND AUDIO



PROCESSING , vol. 6, no 2, March 1988 等中详细揭示了该技术。

代数性码本是具有上述优点的码本。然而，当将代数性码本使用于 CELP 编码·解码装置的概率性码本中时，由于通常由概率声源的目标仅含有很少几个比零部分概率性声源向量进行编码(向量量化)，会产生不能忠实地表现概率声源目标编码的问题。这样，在处理帧相当于无声子音区间以及背景噪音区间等时，这个问题会更加显著。

这是由于通常在无声子音区间以及背景噪音区间概率声源目标形成复杂形状。而且，当对于位速率比 8kb/s 更低的 CELP 编码·解码装置采用代数性码本时，由于使得概率性声源向量中的比零部分较少，仅仅因概率声源目标容易形成脉冲状的有声区间，就会发生上述问题。

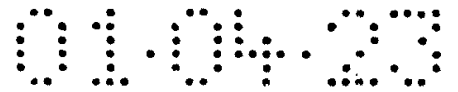
作为解决具有代数性码本的上述问题的方法，提出一种使用脉冲扩散码本的方法，该脉冲扩散码本使得含有比代数性码本更少非零部分(非零部分以外的部分具有零的值)的向量与称为扩散图形的固定波形重叠而将获得向量作为合成滤波器的驱动声源。脉冲扩散编码如特开平 10-232696 号公报、“兼用脉冲扩散构造声源的 ACELP 编码”安永他，电子信息通信学会平成 9 年度春季全国大会发表予稿集，D-14-11, p. 253, 1997-03、“使用了脉冲扩散声源的低速率音频编码”安永他，日本音乡学会平成 10 年秋期研究发表演讲论文集，pp. 281-282, 1988-10 等所揭示的。

这里，接着参照图 8 以及图 9 对于上述文献中所揭示的脉冲扩散码本的概要进行说明。又，图 9 更详细地表示了图 8 的脉冲扩散码本的一示例。

在图 8 以及图 9 的脉冲扩散编码中，代数性码本 4011 是生成由少数非零部分(振幅为+1 或-1)形成的脉冲向量的码本。在上述文献所记载的 CELP 编码装置·解码装置中，将作为代数性码本 4011 的输出的脉冲向量(由少数个非零部分构成)原封不动地作为概率声源的向量使用。

在扩散图案存放部分 4012 中，每一通道存放一种类型以上的称为扩散图案的固定波形。又，对于各通道中存放的上述扩散图案研究每个通道中存放不同形状的扩散图案的情况以及各通道中存放同一形状(共通的)扩散图案的情况。当存放在各通道中的扩散图案为共通时，由于存放了在各通道中存放的扩散图案的情况相当于简化的情况，在本说明书的以下说明中，对于存放在每一通道中的扩散图案形状分别不同的情况，逐步进行说明。

脉冲扩散码本 401 不将代数性码本 4011 的输出向量原封不动地作为概率性



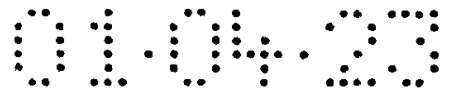
声源向量而输出，而是将从代数性码本 4011 输出的向量与从扩散图案存放部分 4012 读出的扩散图案在脉冲扩散部分 4013 中按每一通道进行叠加，将经过叠加运算而获得向量进行加法运算，并将由此获得的向量作为概率声源的向量而利用。

又，在上述文献中所揭示的 CELP 编码·解码装置的特点在于采用由编码装置与解码装置同一构成(代数性码本部分的通道数、扩散图案存放部分所登录的扩散图案的种类数目以及形状等在编码装置侧与解码装置侧是共通的)的脉冲扩散码本。这样，预先登录在扩散图案存放部分 4012 的扩散图案的形状、种类数、登录了多个种类以上的情况下，通过有效地设定它们的选择方法，由此提高合成声源的品质。

又，这里关于脉冲扩散码本的说明是作为生成由少数非零部分形成的脉冲向量的码本，对于使用了将非零部分的振幅限于+1 或-1 的代数性码本的情况进行了说明，而作为生成该脉冲向量的码本，也可能使用没有限定非零部分的振幅的多脉冲码本以及标准脉冲码本，此时，将脉冲向量与扩散图案叠加的部分作为概率性声源向量而利用，由此也能够实现合成音的质量提高。

到此，提出了将多数的概率声源目标的形状进行统计，并且在每一从代数性码本输出的声源向量中的非零部分(通道)预先登录一个以上的种类图案，所述图案是在概率声源目标中统计上高频率所含有的形状的扩散图案、用于有效地表现无声子音区间与噪音区间的随机形状的扩散图案、用于有效地表现有声稳定区间的脉冲的形状的扩散图案、使得具有从代数性码本输出的脉冲向量的能量(非零部分的位置上集中了能量)分散到周围的作用的形状的扩散图案、对于适当准备的数个扩散图案候补将音频信号反复进行编码、解码、合成音的视听评价并且使得输出高质量的合成音而选择的扩散图案、根据声学知识作成的扩散图案等，按照每一通道将已登录的扩散图案与由代数性码本生成的向量(由几个非零部分构成)叠加，将各通道的叠加结果进行加法运算后的结果作为概率性声源向量使用，由此能够有效地提高合成音的质量。

又，特别地，提出了下述的两种方法，对于在每个通道登录了多个种类(2 个种类以上)的扩散图案的情况，作为这些多个扩散图案的选择方法，扩散图案存放部分 4012 对于登录的扩散图案的全部组合实际地进行编码·解码并且闭合地选择该结果生成的编码误差为最小的扩散图案的选择方法以及在进行概率性码本搜索时利用已知的音频信息(这里所谓的音频信息，例如，利用增益码的动态变动信息或者增益值(与预先设定的阈值)的大小关系信息等来进行判定的有声性的强弱



信息或者利用线性预测编码的变动来进行判定的有声性的强弱信息等。)开放地选择扩散图案的方法等等。

又，在以下的说明中，为了简化说明，限定于特点在于图 9 的脉冲扩散码本内的扩散图案存放部分 4013 每一通道仅登录一种类的扩散图案的图 10 的脉冲扩散码本来进行说明。

这里，接着与将代数性码本使用于 CELP 编码装置时的概率性码本搜索处理进行比较而来说明将脉冲扩散码本使用于 CELP 编码装置时的概率性码本搜索处理。首先，在概率性码本部分使用代数性码本时的码本搜索处理进行说明。

将由代数性码本输出的向量内的非零部分作为 N(将代数目标的通道数目作为 N)、将仅含有 1 个每个通道输出的振幅为 +1 或 -1 的非零部分的向量(非零部分以外的部分的振幅为 0)作为  $d_i$  ( $i$  是通道序号:  $0 \leq i \leq N-1$ )、将子帧长度作为 L 时，由代数性码本输出的登录序号 k 的概率性声源向量  $C_k$  可由下式 9 求得。

$$C_k = \sum_{i=0}^{N-1} d_i$$

$C_k$ : 根据代数性码本的登录序号 K 的概率性声源向量

$d_i$ : 非零部分向量( $d_i = \pm \delta(n-p_i)$ )而  $p_i$ : 非零部分位置)

N: 代数性码本的通道数目 (= 概率性声源向量中的非零部分数目) 式 9

然后，将式 9 代入式 10，如此可得到下式 11。

$$DK = \frac{(V^t H c_k)^2}{\|H c_k\|^2}$$

$V^t$ : v(概率声源目标)的转置向量

$H^t$ : H(合成滤波的脉冲响应行列)的转置行列

$c_k$ : 登录序号第 k 个的概率性声源向量

式 10

$$DK = \frac{(V^t H (\sum_{i=0}^{N-1} d_i))^2}{\|H (\sum_{i=0}^{N-1} d_i)\|^2}$$

v: 概率声源目标向量

H: 合成滤波器的脉冲响应卷积行列

$d_i$ : 非零部分向量( $d_i = \pm \delta(n-p_i)$ )而  $p_i$ : 非零部分位置

N: 代数性码本的通道数目 (= 概率性声源向量的非零部分数目)

$x^t = v^t H$

$M = H^t H$

式 11



使得整理该式 10 所得的式 12 为最大, 特定登录序号  $k$  的处理成为概率性码本搜索处理。

$$DK = \frac{\left( \sum_{i=0}^{N-1} x^t d_i \right)^2}{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} d_i^t M d_j} \quad \text{式 12}$$

然而, 在式 12 中,  $x^t = v^t H$ 、 $M = H^t H$  ( $V$  为概率性音源目标)。这里, 对于各登录序号  $k$  计算式 12 的值时, 在此前的处理阶段中计算  $x^t = v^t H$  以及  $M = H^t H$ , 并且将该计算结果存储在存储器中。通过进行该前置处理, 能够大幅度地削减作为概率性声源向量对每个登录的各候补进行式 12 计算时的运算量, 作为该结果, 能够控制概率性码本搜索需要的运算量为较少, 而在少数文献等中已有揭示且为一般已知。

下面说明将脉冲扩散码本用于概率性码本时的概率性码本探索处理。

将作为脉冲扩散码本构成一部分由代数性码本输出的非零部分作为  $N$  (将代数目标的通道数目作为  $N$ )、将仅含有 1 个每个通道输出的振幅为 +1 或 -1 的非零部分的向量 (非零部分以外的部分的振幅为 0) 作为  $d_i$  ( $i$  是通道序号:  $0 \leq i \leq N-1$ )、将扩散图案存放部分存放的通道序号  $i$  用扩散图案作为  $w_i$ 、将子帧长度作为  $L$  时, 由脉冲扩散码本输出的登录序号  $k$  的概率性声源向量  $C_k$  可由下式 13 求得。

$$C_k = \sum_{i=0}^{N-1} w_i d_i$$

$C_k$ : 根据脉冲扩散码本的登录序号  $K$  的概率性声源向量

$w_i$ : 扩散图案 ( $w_i$ ) 叠加行列

$d_i$ : 代数性码本部分输出的非零部分向量

( $d_i = \pm \delta(n - p_i)$  而  $p_i$ : 非零部分位置)

$N$ : 代数性码本部分的通道数目

式 13

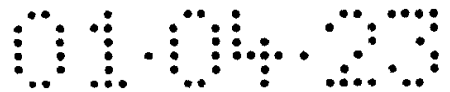
然后, 将式 13 代入式 10, 如此可得到下式 14。

$$DK = \frac{\left( v^t H \left( \sum_{i=0}^{N-1} w_i d_i \right) \right)^2}{\left\| H \left( \sum_{i=0}^{N-1} w_i d_i \right) \right\|^2}$$

$v$ : 概率声源目标向量

$H$ : 合成滤波器的脉冲响应卷积行列

$w_i$ : 扩散图案 ( $w_i$ ) 叠加行列



$d_i$ : 代表码本部分输出的非零部分向量

( $d_i = \pm \delta(n - p_i)$ ) 而  $p_i$ : 非零部分位置

$N$ : 代数性码本的通道数目 (= 概率性声源向量的非零部分数目)

$$H_i = H W_i$$

$$x^t = v^t H_i$$

$$R = H_i H_j$$

式 14

特定整理该式 14 所得的式 15 为最大的概率性声源向量的登录序号  $k$  的处理成为使用了脉冲扩散码本时的概率性码本搜索处理。

$$DK = \frac{\left( \sum_{i=0}^{N-1} x_i^t d_i \right)^2}{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} d_i^t R d_j} \quad \text{式 15}$$

然而，在式 15 中， $x^t = v^t H_i$  (而  $H_i = H W_i$ ;  $W_i$ : 扩散图案叠加行列)。这里，对于各登录序号  $k$  计算式 15 的值时，在此前的处理中，可以计算  $H_i = H W_i$ 、 $x^t = v^t H_i$  以及  $R = H_i H_j$  并且将该计算结果存储在存储器中。如此，能够使得作为概率性声源向量对每个登录的各候补进行式 15 计算时的运算量与使用了代数性码本时计算式 12 时的运算量相同 (明显地式 12 与式 15 形式相同)，即使采用扩散码本的情况下，也能够以较少的运算量来进行概率性码本的搜索。

在上述技术中，表示了将脉冲扩散码本使用于 CELP 编码装置·解码装置的概率性码本部分的效果以及将脉冲扩散码本使用于概率性码本部分时、以及与将代数性码本使用于概率性码本部分时相同的方法进行概率性码本搜索。将代数性码本使用于概率性码本部分时概率性码本搜索所必要的运算量与将脉冲扩散码本使用于概率性码本部分时概率性码本搜索所必要的运算量的差别为式 12 与式 15 各自前置处理阶段所必要的运算量的差别，即是前置处理 ( $x^t = v^t H_i$ 、 $M = H^t H$ ) 与前置处理 ( $H_i = H W_i$ 、 $x^t = v^t H_i$ 、 $R = H_i H_j$ ) 所必要的运算量的差别。

一般地，在 CELP 编码装置·解码装置中，该位速率越低概率性码本部分能分配到的位数也越少。这样这种倾向在将代数性码本以及脉冲扩散码本使用于概率性码本部分时构成概率性声源向量时非零部分数目也随之减少。因此，CELP 编码装置·解码装置的位速率越低，使用代数性码本时与使用脉冲扩散码本时的运算量的相差也越小。但是，当位速率较高时以及即使位速率较低而必须极力控制运算量时，由于使用脉冲扩散码本，有时不能够忽视产

生的前置处理阶段运算量的增加。

在本实施形态中，对于在概率性码本部分使用了脉冲扩散码本的 CELP 方式的音频编码装置、音频解码装置以及音频编码解码系统，在将比代数性码本使用于概率编码部分时增加的、编码搜索处理中前置处理部分的运算量增加份额控制在较小程度的同时，在解码侧获得高品质的合成音这一点进行说明。

具体地，本实施形态的技术是用于解决将脉冲扩散码本使用在 CELP 编码·解码装置的概率性码本部分时所产生的上述问题，其特点在于在编码装置侧与解码装置侧采用不同的扩散图案。即，在本实施形态中，在音频解码装置侧的扩散图案存放部分中登录了上述的扩散图案，通过使用这些图案而生成比采用代数性码本时更高质量的合成音频。另一方面，在音频编码装置侧，登录了简化了登录在解码装置侧的扩散图案存放部分的扩散图案的扩散图案（例如，以一定间隔拉开扩散图案或者以一定长度截断的扩散图案）并且采用它来进行概率性码本的搜索。

由此，将脉冲扩散码本使用于概率性码本部分时，在编码侧，能够抑制使得比将代数性码本使用于概率性码本部分时所增加的、前置阶段编码搜索时的运算量为较少，并且在解码侧能够获得高质量的合成音。

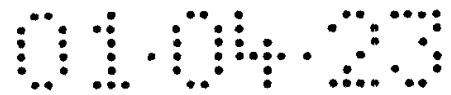
在编码装置侧与解码装置侧采用不同扩散图案是将预先准备的（解码装置用的）扩散向量保留特性而进行变形，由此获得解码用扩散向量。

这里，作为预先准备解码装置用扩散向量的方法，本发明者们研究了以往提出的申请（特开平 10-63300 号公报）中揭示的方法、即研究音源搜索用目标线路的统计性倾向来进行准备的方法、实际地将音源目标进行编码并且反复进行向此时产生编码误差总和变小的方向变形的操作而来进行准备的方法、以及提高合成音的质量并根据声学知识进行设计的方法等，以随机化脉冲声源的高频相位成分为目的进行设计的方法。这些内容都包含于此。

如此获得扩散向量其特点在于，任何的扩散向量的前部采样附近的采样其振幅比后部的采样的振幅要大。即使在中部，前部的采样的振幅经常是扩散向量内全部采样之中最大的（大多数情况下）。

作为将解码装置用扩散向量保留特性地进行变形而获得解码用扩散向量的具体方法，可以列举下述的方法。

1) 每隔适当间隔将解码装置用扩散向量的采样值置换为 0，由此获得解码



用扩散向量。

2) 通过将某长度的解码装置用扩散向量以适当长度截断而获得解码用扩散向量。

3) 预先设定振幅的阈值并对于解码装置用扩散向量将比设定的阈值振幅要小的采样置换为 0，由此获得解码用扩散向量。

4) 对于某长度的解码装置用扩散向量，每适当间隔保存含有前部采样的采样值并且将此外采样值置换为 0，由此获得编码装置用扩散向量。

这里例如上述 1) 的方法，即使采用了扩散向量前部起的多个采样，也能够保存了扩散向量的大致形状(大致特性)并且能够获得新的编码装置用扩散向量。

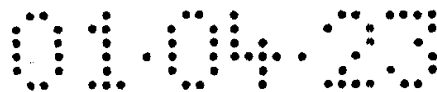
又，例如上述 2) 的方法，即使每适当间隔将采样值置换为 0，也能够保存原来的扩散向量的大致形状(大致特性)并且可以获得新的编码装置用扩散向量。特别地，在上述 4) 方法的情况下由于限定必须保持通常振幅最大的前部采样的振幅，因此能够更可靠地保存原来的扩散向量的大致形状。

又，例如 3) 方法，原封不动地保存具有特定值以上振幅的采样，即使将具有所述特定值以下振幅的采样其振幅置换为 0，也能够保持扩散向量的大致形状(大致特性)，能够获得编码装置用的扩散向量。

以下，对于本实施形态的音频编码装置以及音频解码装置参照附图进行详细地说明。又，附图所示的 CELP 音频编码装置(图 11)以及 CELP 音频解码装置(图 12)在以往的 CELP 音频装置以及 CELP 音频解码装置的概率性码本部分中具有采用上述脉冲扩散码本这一特征。因此，在以下说明中，记载了概率性码本、概率性声源向量、概率声源增益的部分分别能够替代为脉冲扩散码本、脉冲扩散声源向量、脉冲扩散声源增益。又，CELP 音频编码装置以及 CELP 音频解码装置的概率性码本因具有噪声码本或者存放多个种类的固定波形的作用而有时也被称为固定码本。

在图 11 的 CELP 音频编码装置中，首先，线性预测分析部分 501 对于输入音频进行线性预测分析并计算出线性预测系数，将算出的线性预测系数输入线性预测系数编码部分 502。其次，线性预测系数编码部分 502 将线性预测系数编码(向量量化)，将由向量量化获得量化指数(以下，称为线性预测编码)输出到编码输出部分 513 以及线性预测码解码部分 503。

其次，线性预测码解码部分 503 将由线性预测系数编码部分 502 获得线



性预测码进行解码(反量化)并输出到合成滤波器 504。合成滤波器 504 构成以由线性预测码解码部分 503 获得解码线性预测码为系数的全极型模式合成滤波器。

然后, 将从自调码本 506 选出的自调声源向量乘上自调声源增益 509 而获得向量与将从脉冲扩散码本 507 选出的概率性声源向量乘上概率声源增益 510 而获得向量在向量加法运算部分 511 进行加法运算并生成驱动声源向量。然后, 误差计算部分 505 按照下式 16 计算由该驱动声源向量驱动合成滤波器 504 时的输出向量与输入音频的误差, 将误差 ER 输出到编码特定部分 512。

$$ER = \|u - (g_a H p + g_c H c)\|^2$$

u: 输入音频 (向量)

H: 合成滤波器的脉冲响应行列

p: 自调声源向量

c: 概率性声源向量

$g_a$ : 自调声源增益

$g_c$ : 概率声源增益

式 16

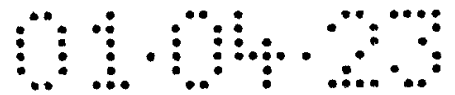
然而, 在式 16 中, u 表示处理帧内的输入音频向量、H 表示合成滤波器的脉冲响应行列、 $g_a$  表示自调声源增益、 $g_c$  表示概率声源增益、p 表示自调声源向量、c 表示概率性声源向量。

这里, 自调码本 506 是存放了以往数帧份额的驱动声源向量的缓冲器(动态存储器), 使用从上述自调码本 506 选出的自调声源向量是为了表现将输入音频通过合成滤波器的反滤波器而获得线性预测残差向量中的周期成分。

另一方面, 使用从脉冲扩散码本 507 选出的声源向量是为了表现在现处理帧中向线性预测残差向量新添加的非周期成分(从线性预测残差向量中去除周期性(自调声源向量成分)的成分)。

然后, 自调声源向量增益乘法运算部分 509 以及概率性声源向量增益乘法运算部分 510 相对于从自调码本 506 中选出的自调声源向量以及从扩散编码 507 中选出的概率性声源向量具有乘以从增益码本中读出的自调声源增益以及概率声源增益的功能。又, 所谓增益码本 508 是多个种类存放乘以自调声源向量的自调声源增益以及乘以概率性声源向量的概率声源增益组合的静态存储器。

代码特定部分 512 选择使得由误差计算部分 505 计算的式 16 的误差 ER



为最小的上述 3 个码本(自调码本、脉冲扩散码本、增益码本)指数的最佳组合。然后, 代码特定部分 512 将上述误差为最小时所选择的各码本的指数分别作为自调声源码、概率声源码、增益码而输出到代码输出部分 513。

最后, 代码输出部分 513 将线性预测系数编码部分 502 获得线性预测码、由代码特定部分 512 特定的自调声源码、概率声源码以及增益码进行汇总并且作为表现当前处理帧内地输入音频的代码(位信息)而输出到解码装置侧。

又, 有时代码特定部分 512 所进行的自调声源码、概率声源码、增益码的特定是在将一定时间间隔的帧分割为称为子帧的更短时间间隔之后而进行的。然而, 在本说明书中, 帧与子帧没有特别的区别(统一称为帧), 并且在以下进行说明。

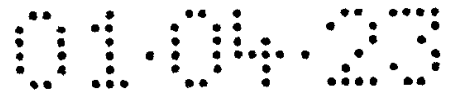
其次, 参照图 12 对于 CELP 音频解码装置的概要进行说明。

在图 12 的 CELP 解码装置中, 首先代码输入部分 601 接受由 CELP 音频编码装置(图 11)所特定的代码(用于代码表现帧区间的音频信号的位信息), 并且将接受的代码分解为线性预测代码。自调声源码、概率声源码以及增益码这 4 种类型的代码。然后, 将线性预测代码、自调声源码、概率声源码、增益码分别输出到线性预测系数解码部分 602、自调码本 603、脉冲扩散码本 604、增益码本 605。

其次, 线性预测系数解码部分 602 将从代码输入部分 601 输入的线性预测码解码并获得解码的线性预测码, 将该解码的线性预测码输出到合成滤波器 609。

合成滤波器 609 构成将线性预测系数解码部分 602 获得解码的线性预测码作为系数的全极型模式合成滤波器。又, 自调码本 603 输出从代码输入部分 601 输入的自调声源码所对应的自调声源向量。又, 脉冲扩散码本 604 输出从代码输入部分 601 输入的概率声源码所对应的概率性声源向量。又, 增益码本 605 读出从代码输入部分输入的增益码所对应自调声源增益以及概率声源增益并且分别输出到自调声源增益乘法运算部分 606 以及概率声源增益乘法运算部分 607。

然后, 自调声源增益乘法运算部分 606 在从自调码本 603 输出的自调声源向量上乘上从增益码本 605 输出的自调声源增益, 概率声源增益乘法运算部分 607 在从脉冲扩散码本 604 输出的概率性声源向量上乘以由增益码本 605 输出的概率声源增益。然后, 向量加法运算部分 608 加上自调声源增益乘法



运算部分 606 以及概率声源增益乘法运算部分 607 各自的输出向量并且生成驱动声源向量。此后，由该驱动声源向量驱动合成滤波器 609 并且输出接收到的帧区间的合成音。

在如此的 CELP 方式的音频编码装置·音频解码装置中，为了获得高质量的合成音，必须抑制式 16 的误差 ER 为较小。因此，为了使得式 16 的 ER 最小，希望在闭环下特定自调声源码、概率声源码、增益码的组合。然而，由于在闭环下特定式 16 的误差 EG 的运算处理量过大，一般在开环下特定上述 3 种代码。

具体地，首先进行自调码本搜索。这里所谓的自调码本搜索处理是由从存放了先前帧的驱动声源向量的自调码本中输出的自调声源向量而将输入音频通过反滤波器而获得的预测残差向量中的周期成分来进行向量量化的处理。然后，将具有线性预测残差向量内的周期成分与近似周期成分的自调声源向量的登录序号作为自调声源码进行特定。又，通过自调码本搜索，同时暂时确认理想自调声源增益。

其次，进行脉冲扩散码本搜索。脉冲扩散码本搜索是将从处理帧的线性预测残差向量中去除了周期成分的成分，即从线性预测残差向量中减去了自调声源向量成分的成分(以下，也称为概率声源目标)使用存放在脉冲扩散码本中的多个概率性声源向量候补而进行向量量化的处理。然后，通过该脉冲扩散码本搜索处理，将最小误差进行编码的概率性声源向量的登录序号作为概率声源码来特定概率声源目标。又，通过脉冲扩散码本搜索，同时暂时区定理想的概率目标。

此后，进行增益目标搜索。增益码本搜索是如下所述的处理，将由在自调码本搜索时暂时获得理想的自调增益与脉冲扩散码本搜索时暂时获得理想的概率增益这 2 部分构成的向量由存放在增益码本的增益候补向量(由自调声源增益候补与概率声源增益候补这 2 部分形成的向量候补)进行编码(向量量化)而使得误差为最小。然后，将这里所选的增益后补向量的登录序号作为增益码输出到代码输出部分。

这里，接着，在 CELP 音频编码装置中上述一般的代码搜索处理中，对于脉冲扩散码本搜索处理(特定了自调声源码之后，特定概率声源码的处理)进行更详细地说明。

如上所述，对于一般的 CELP 编码装置，在进行脉冲扩散码本搜索时，已

经特定了线性预测码以及自调声源码。这里，如将由已经特定的线性预测码构成的合成滤波器的脉冲响应行列作为  $H$ 、将与自调声源码对应的自调声源向量作为  $p$ 、将在特定自调声源码的同时所求得理想自调声源增益(暂定值)作为  $g_a$ ，则式 16 的误差  $ER$ ，可以变形为下式 17。

$$ER_k = \|v - g_c H c_k\|^2$$

$v$ : 概率声源目标(而  $v = u - g_a H p$ )

$g_c$ : 概率声源增益

$H$ : 合成滤波器的脉冲响应行列

$c_k$ : 概率性声源向量( $k$  登录编码)

式 17

然而，式 17 内的向量  $v$  是使用了帧区间的输入音频信号  $u$ 、合成滤波器的脉冲响应行列  $H$ (已知)、自调声源向量  $p$ (已知)、理想自调声源增益  $g_a$ (暂定值)的下式 18 的概率声源目标。

$$v = u - g_a H p$$

$u$ : 输入音频(向量)

$g_c$ : 概率声源增益(暂定值)

$H$ : 合成滤波器的脉冲响应行列

$p$ : 自调声源向量

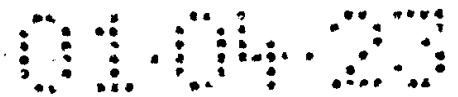
式 18

又，式 16 中以  $c$  表示概率性声源向量，另一方面，式 17 中以  $c_k$  表示概率性声源向量。这是由于，在式 16 中没有表示概率性声源向量的登录序号( $k$  这点)不同，相对于此，在式 17 中表示了登录序号，虽表示上的不同，而所指的对象是相同的。

因此，作为脉冲扩散码本搜索，是求得使得式 17 的  $ER_k$  为最小的概率性声源向量  $c_k$  的登录序号  $k$  的处理。然后，在特定使得式 17 的误差最小的概率性声源向量  $c_k$  的登录序号  $k$  时，概率声源增益  $g_c$  能够假设为任意的值。因此，取得使得式 17 的误差为最小的登录序号的处理可以置换为特定使得上式 10 中分数  $D_k$  为最大的概率性声源向量  $c_k$  的登录序号  $k$  的处理。

然后，脉冲扩散码本搜索进行下述 2 个阶段的处理，对于每个概率性声源向量  $c_k$  的登录序号  $k$  由误差计算部分 505 计算出式 10 的分数  $D_k$  并将该值输出到代码特定部分 512，在代码特定部分 512 中比较每个登录序号  $k$  的式 10 的值并将该值为最大时登录序号  $k$  作为概率声源代码输出到代码输出部分 513。





以下，对于本实施形态音频编码装置以及音频解码装置的动作进行说明。

图 13A 表示图 11 所示的音频编码装置的脉冲扩散码本 507 的构造，图 13B 表示图 12 所示的音频解码装置的脉冲扩散码本 604 的构造，比较图 13A 所示的脉冲扩散码本 507 与图 13B 所示的脉冲扩散码本 604 时，构造上的不同在于登录在扩散图案存放部分中的扩散图案形状有所不同。

在图 13B 的音频解码装置中，在扩散图案存放部分 4012 中在每一通道分别登录一种图案，所述图案如下：(1)统计多数概率声源目标的形状并且概率声源目标中以统计上高频率地所含有的形状的扩散图案，(2)用于有效地表现无声子音区间及噪声区间的随机形状的扩散图案，(3)用于有效地表现有声稳定区间的脉冲形状的扩散图案，(4)发挥作用而使得从代数性码本输出的声源向量的能量(在非零部分的位置上集中了能量)分散到周围的形状的扩散图案，(5)对于适当准备的数个扩散图案候补，将音频信号编码、解码，反复进行合成音的视听评价，并且为了输出高质量的合成音而选择的扩散图案，(6)根据声学知识而作成的扩散图案中任意的扩散图案。

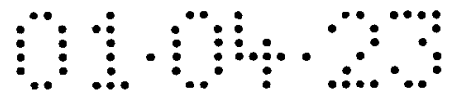
另一方面，在图 13A 的音频编码装置侧，在扩散图案存放部分 4012 中登录将在图 13B 的音频解码装置侧的扩散图案存放部分 4012 中登录的的扩散图案每隔一个采样置换为 0 的扩散图案。

然后，对于如此构成的 CELP 音频编码装置/音频解码装置中，没有注意到登录了在编码装置侧与解码装置侧不同的扩散图案，以上述相同的方法将音频信号编码·解码。

在编码装置中，能够减少在概率性码本部分采用脉冲扩散码本时的概率性码本搜索时的前置处理运算量(能够减去约一半的  $H_i=H_t W_i$  以及  $x_{it}=v_t H_i$  的运算量)，在解码装置侧通过在脉冲向量上重叠与以往相同的扩散图案，能够将集中在非零部分位置上的能量分散到周围，能够提高合成音的质量。

又，在本实施形态中，如图 13A 以及图 13B 所示，已经说明了在音频编码装置侧采用将使用于音频解码装置侧的扩散图案每隔 1 个采样置换为 0 的扩散图案的情况进行了说明，而在音频编码装置侧采用将音频解码装置侧使用的扩散图案的部分每隔  $N$  个( $N \geq 1$ )采样置换为 0 而获得扩散图案的情况，也能够照样地适用本实施形态，此时也能够获得同样的效果。

又，在本实施形态中，对于扩散图案存放部分按每 1 通道登录 1 种类型的扩散图案情况下的实施形态进行了说明，而对于每个通道登录 2 种类型以



上的扩散图案并且选择使用这些扩散图案为特征的将脉冲扩散码本用于概率码本部分的 CELP 音频编码装置·解码装置，也能够适用本发明，此时也能够取得同样的效果。

又，在本实施形态中，对于使用输出代数性码本部分含有 3 个非零部分的向量的脉冲扩散码本的情况，说明了实施的情况，而对于代数性码本部分输出的向量中非零部分数目为  $M$  个 ( $M \geq 1$ ) 的情况，也能够适用本实施形态，此时也能够获得同样的作用·效果。

又，在本实施形态中，对于生成由少数个非零部分构成的脉冲向量的码本而采用了代数性码本的情况进行了说明，而作为生成该向量的码本，当采用多脉冲码本、标准脉冲码本等其他码本的情况下，也能够适用本实施形态，在这种场合也能获得同样的作用·效果。

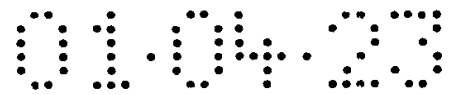
其次，图 14A 表示图 11 所示的音频编码装置的脉冲扩散码本的构造，图 14B 表示图 12 所示的音频解码装置的脉冲扩散码本的构造。

比较图 14A 所示的脉冲扩散码本与图 14B 所示的脉冲扩散码本的构造时，构造上的不同在于登录在扩散图案存放部分的扩散图案的长度有所不同。在图 14B 的音频解码装置中，在扩散图案存放部分 4012 中在每一通道分别登录一种与上述扩散图案相同的扩散图案，即 (1) 统计多数概率声源目标的形状并且概率声源目标中以统计上高频率地所含有的形状的扩散图案，(2) 用于有效地表现无声子音区间及噪声区间的随机形状的扩散图案，(3) 用于有效地表现有声稳定区间的脉冲形状的扩散图案，(4) 发挥作用而使得从代数性码本输出的声源向量的能量(在非零部分的位置上集中了能量)分散到周围的形状的扩散图案，(5) 对于适当准备的数个扩散图案候补，将音频信号编码、解码，反复进行合成音的视听评价，并且为了输出高质量的合成音而选择的扩散图案，(6) 根据声学知识而作成的扩散图案中任意的扩散图案。

另一方面，在图 14A 的音频编码装置侧，在扩散图案存放部分 4012 中登录了将在图 14B 的音频解码装置侧的扩散图案存放部分 4012 中登录的的扩散图案以一半长度截断的扩散图案。

然后，对于如此构成的 CELP 音频编码装置/音频解码装置中，没有注意到登录了在编码装置侧与解码装置侧不同的扩散图案，以上述相同的方法将音频信号编码·解码。

在编码装置中，能够减少在概率性码本部分采用脉冲扩散码本时的概率



性码本搜索时的前置处理运算量(能够减去约一半的  $H_i=H_tW_i$  以及  $x_{it}=v_tH_i$  的运算量), 在解码装置侧, 能够利用与以往相同的扩散图案, 由此能够提高合成音的质量。

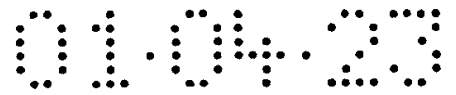
又, 在本实施形态中, 如图 14A 以及图 14B 所示, 已经说明了在音频编码装置侧采用将使用于音频解码装置侧的扩散图案以一半长度截断的扩散图案的情况进行了说明, 而在音频编码装置侧采用将音频解码装置侧使用的扩散图案以更短的长度  $N(N \geq 1)$  截断后的扩散图案的情况, 能够进一步地减少概率性码本搜索时前置处理运算量。然而, 这里将使用于音频编码装置侧的扩散图案以长度 1 截断时相当与没有使用扩散图案的音频编码装置(在音频解码装置中适用扩散图案)。

又, 在本实施形态中, 对于扩散图案存放部分按每 1 通道登录 1 种类型的扩散图案情况下的实施形态进行了说明, 而对于每个通道登录 2 种类型以上的扩散图案并且选择使用这些扩散图案为特征的脉冲扩散码本用于概率码本的音频编码装置/解码装置, 也能够适用本实施形态, 此时也能够取得同样的效果·作用。

又, 在本实施形态中, 对于使用输出代数性码本部分含有 3 个非零部分的向量的脉冲扩散码本的情况进行了说明, 而对于代数性码本部分输出的向量中非零部分数目为  $M$  个 ( $M \geq 1$ ) 的情况, 也能够适用本实施形态, 此时也能够获得同样的作用·效果。

又, 在本实施形态中, 对于在音频编码装置侧采用将使用于音频解码装置侧的扩散图案以一半长度截断的扩散图案的情况进行了说明, 而也可能在音频编码装置侧将使用于音频解码装置侧的扩散图案以长度  $N(N \geq 1)$  截断并且将截断后的扩散图案每隔  $M(M \geq 1)$  个采样置换为 0, 此时能够进一步地减少编码搜索运算量。

如此, 根据本实施形态, 对于在概率性码本部分采用脉冲扩散码本的 CELP 方式音频编码装置与解码装置以及音频编码解码系统, 将在研究获得概率声源目标中频繁包含的固定波形作为扩散图案进行登录, 在脉冲向量上叠加该扩散图案(反映), 由此能够利用比概率声源目标更近的概率性声源向量, 所以能够提高解码侧合成音的质量, 而且在编码侧可以获得能够将概率性码本搜索的运算量抑制得比以往更少的有利效果, 所述概率性码本搜索有时会在将脉冲扩散码本使用于概率性码本部分产生问题。



又，作为生成由少数个非零部分形成的脉冲向量的码本，即使使用多脉冲码本、标准脉冲码本等的其他码本的情况下，也能够获得同样的作用·效果。

上述实施形态 1~3 中音频的编码/解码以音频编码装置/音频解码装置进行了说明，而也可以作为软件而构成这些音频编码/音频解码。例如，也可以这样构成，即在 ROM 中存放上述音频编码/解码的程序并且根据该程序按照 CPU 的指示进行动作。又，也可以将程序、自调码本以及概率性码本(脉冲扩散码本)存放在计算机中能够读取的存储媒体中，将该存储媒体的程序、自调码本以及概率码本(脉冲扩散码本)记录在计算机的 RAM 中而使得根据程序来进行动作。即使在这种情况下之下，也能够实现与上述实施形态 1~3 相同的作用、效果。而且，可以在通信终端下载实施形态 1~3 的程序而使得在该通信终端实行程序。

又，对于上述实施形态 1~3，可以个别地实施，也可以组合起来实施。

本说明书是根据 1999 年 8 月 23 日申请的特愿平 11-235050 号、1999 年 8 月 24 日申请的特愿平 11-236728 以及 1999 年 9 月 2 日申请的特愿平 11-248363。它们的内容也全部包含在本说明书中。

### 工业利用性

本发明能够适用于数字通信系统的基地局以及通信终端装置。

说明书附图

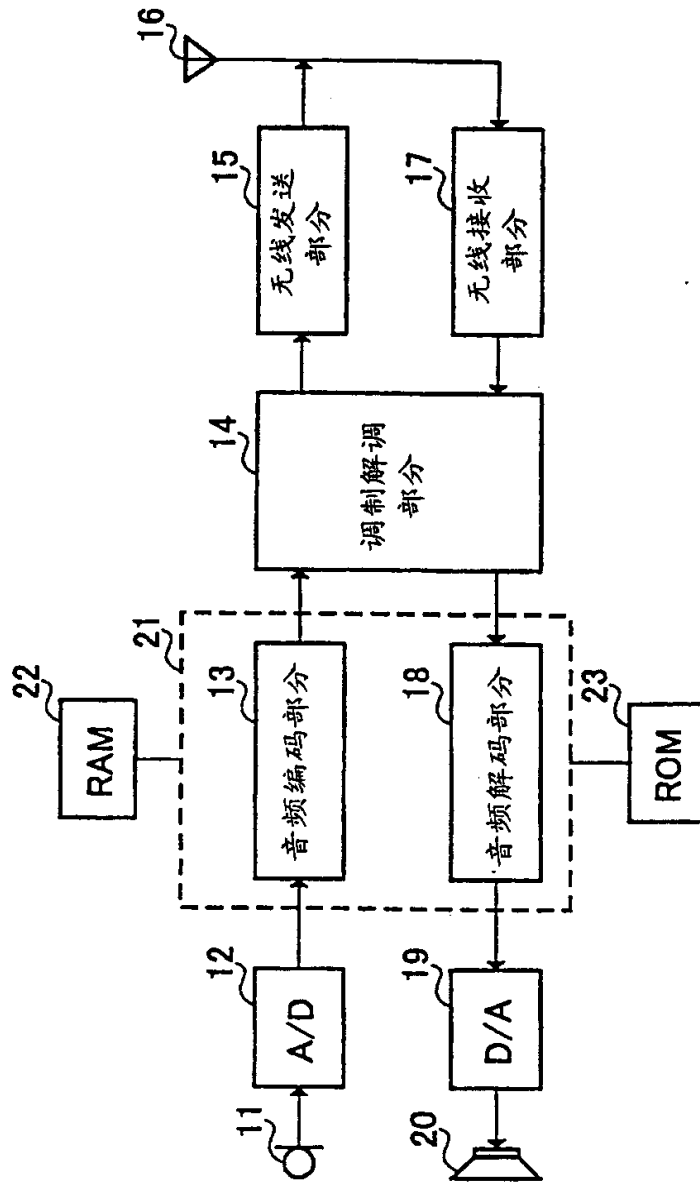


图 1

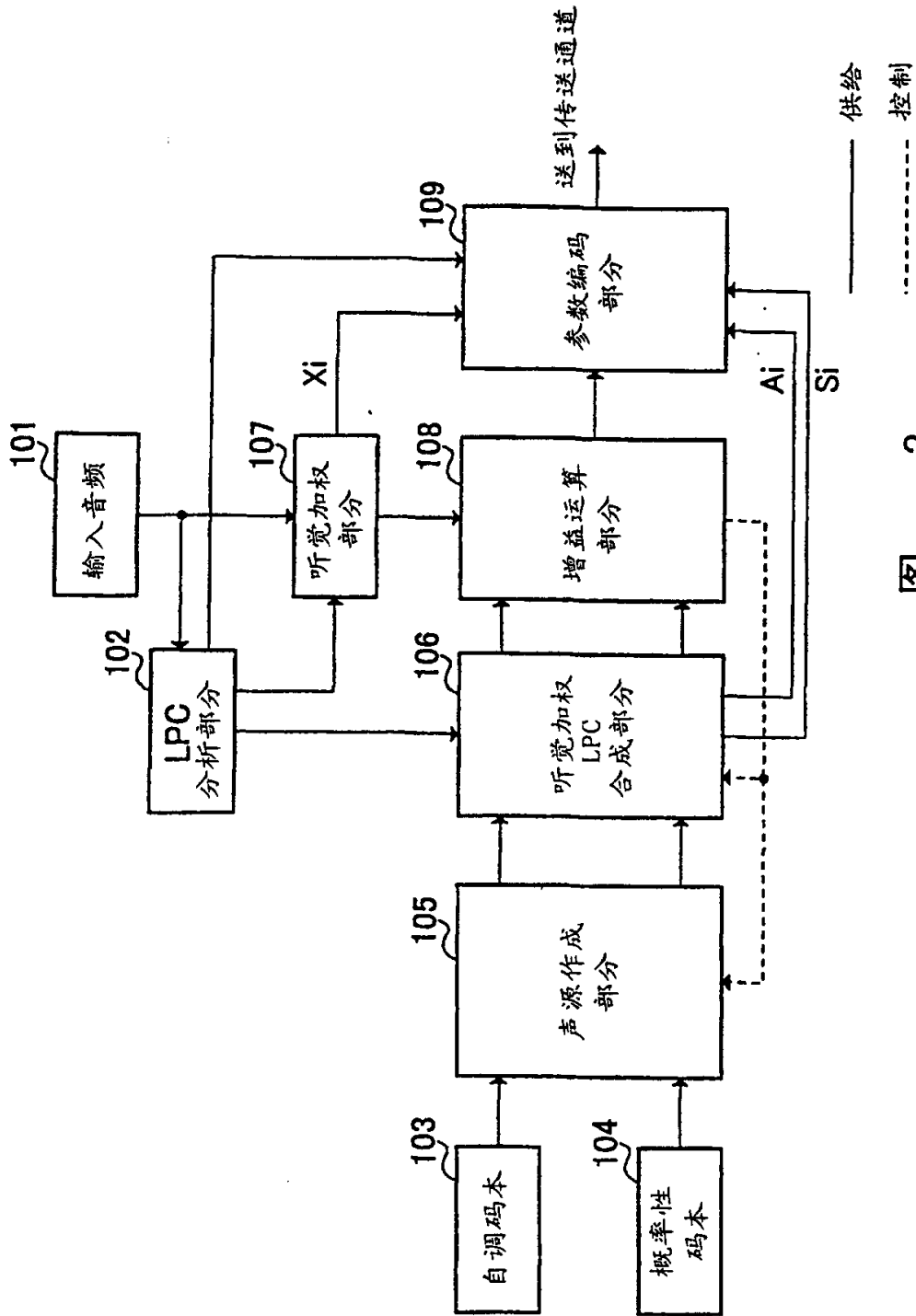


图 2

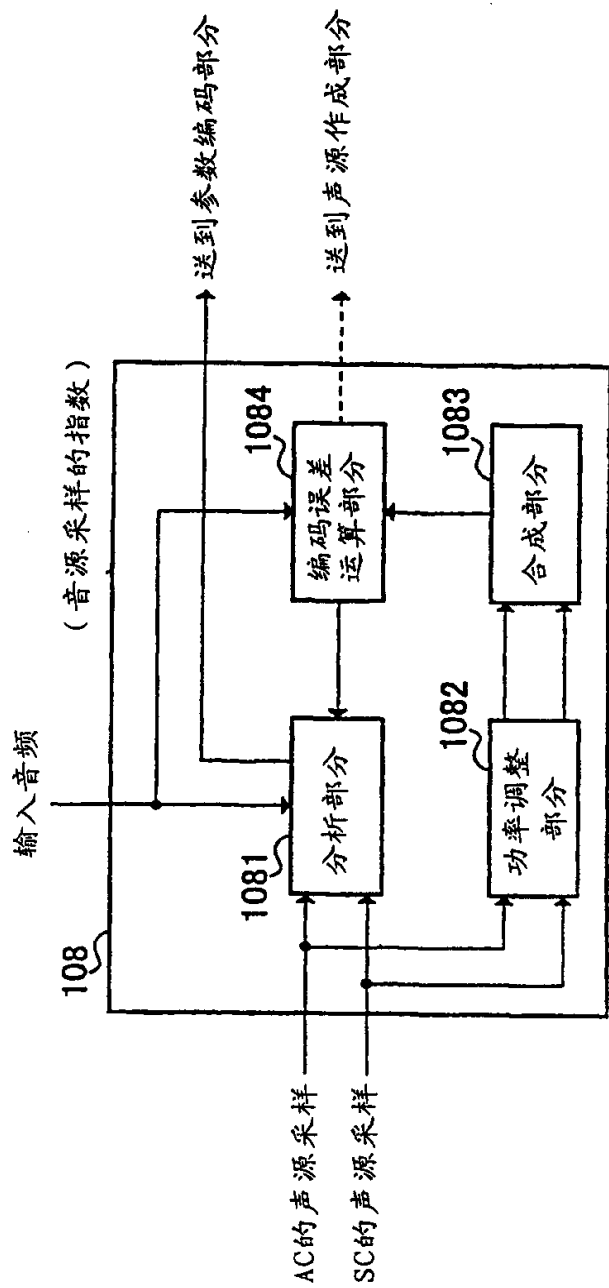


图 3

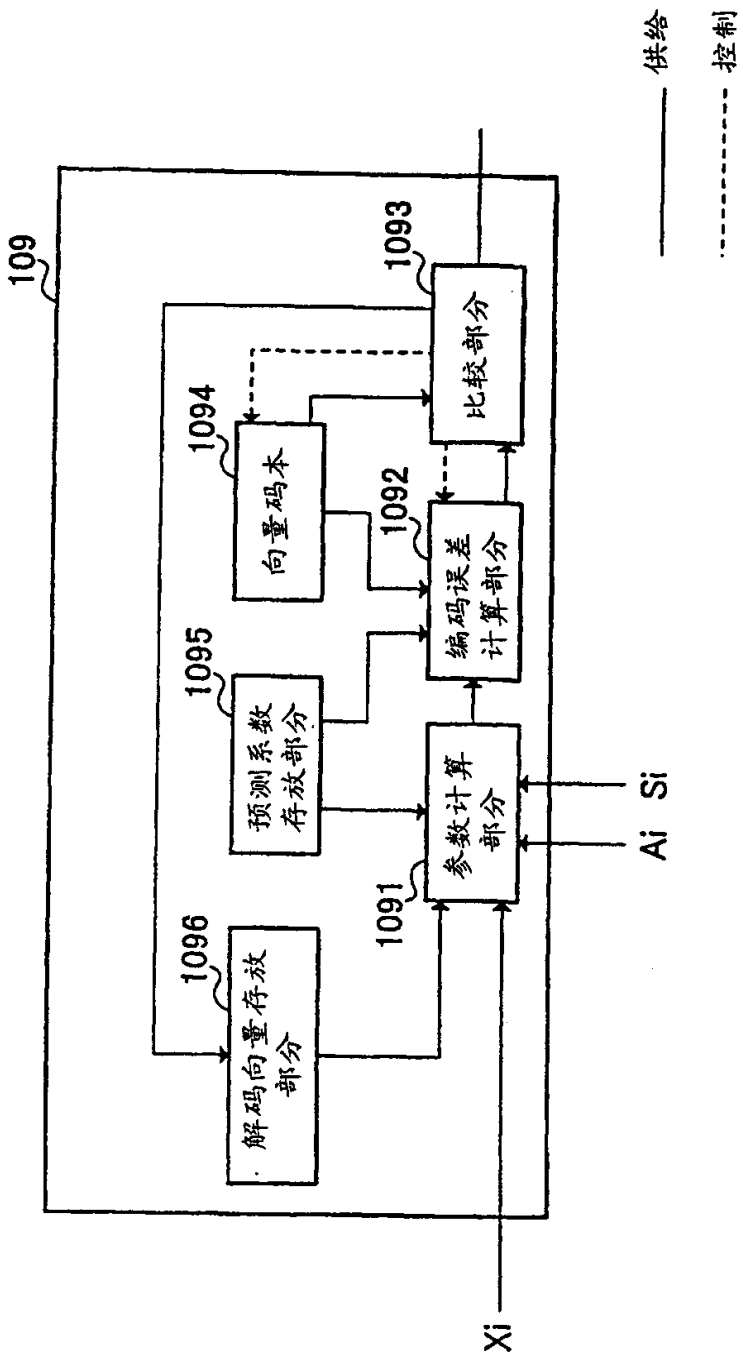


图 4



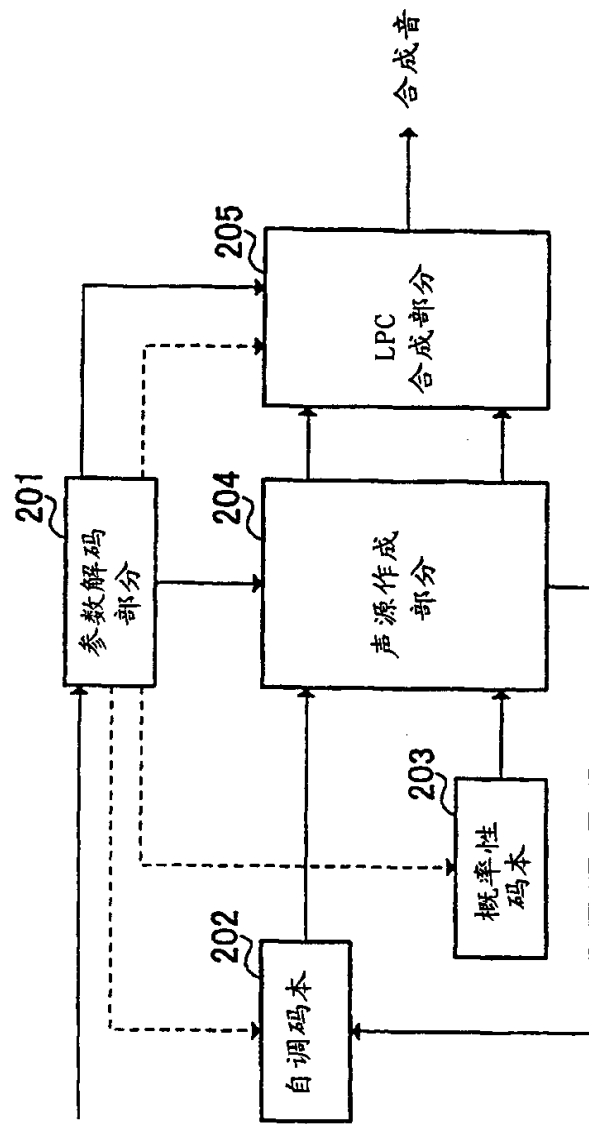


图 5

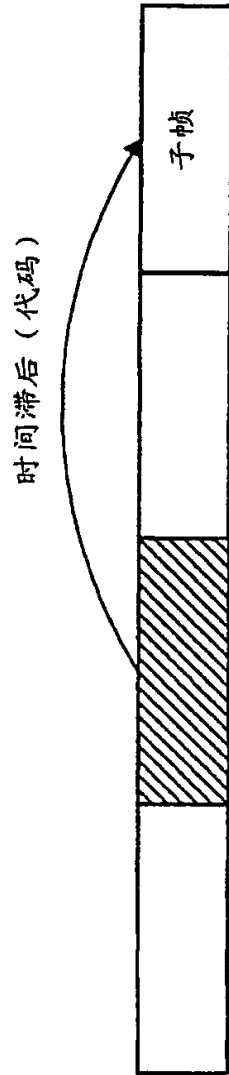


图 6

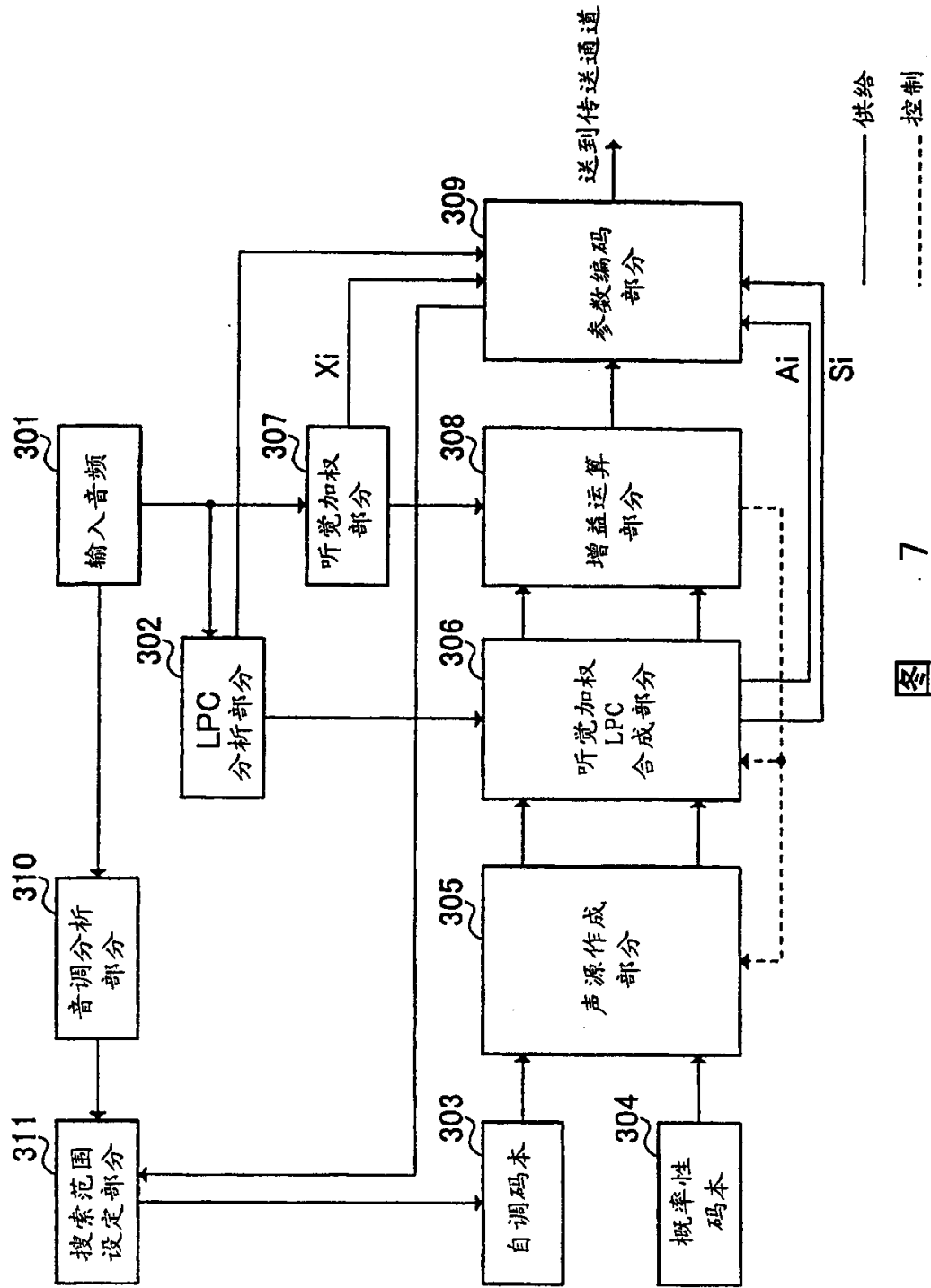
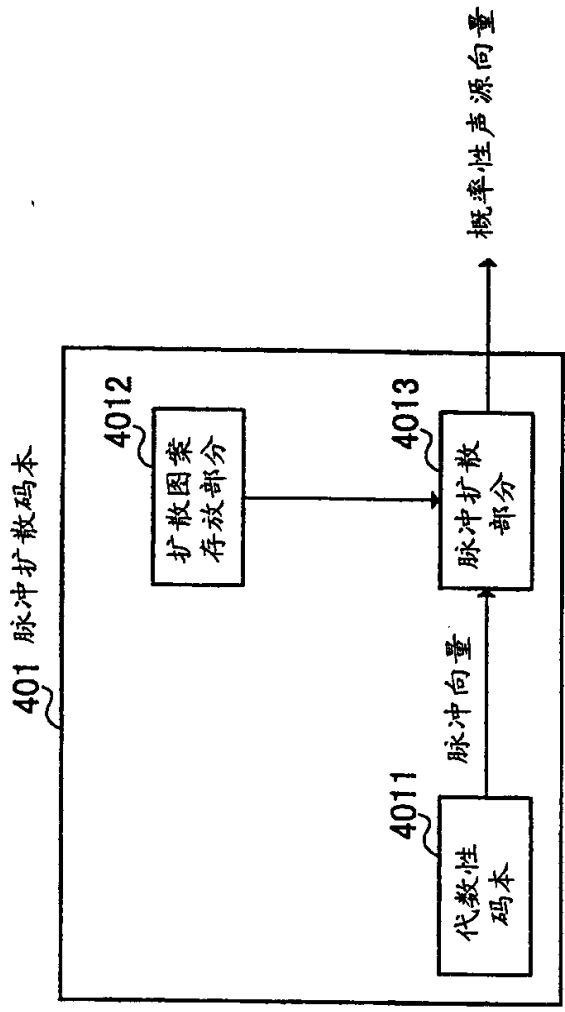


图 7



图

8

01000

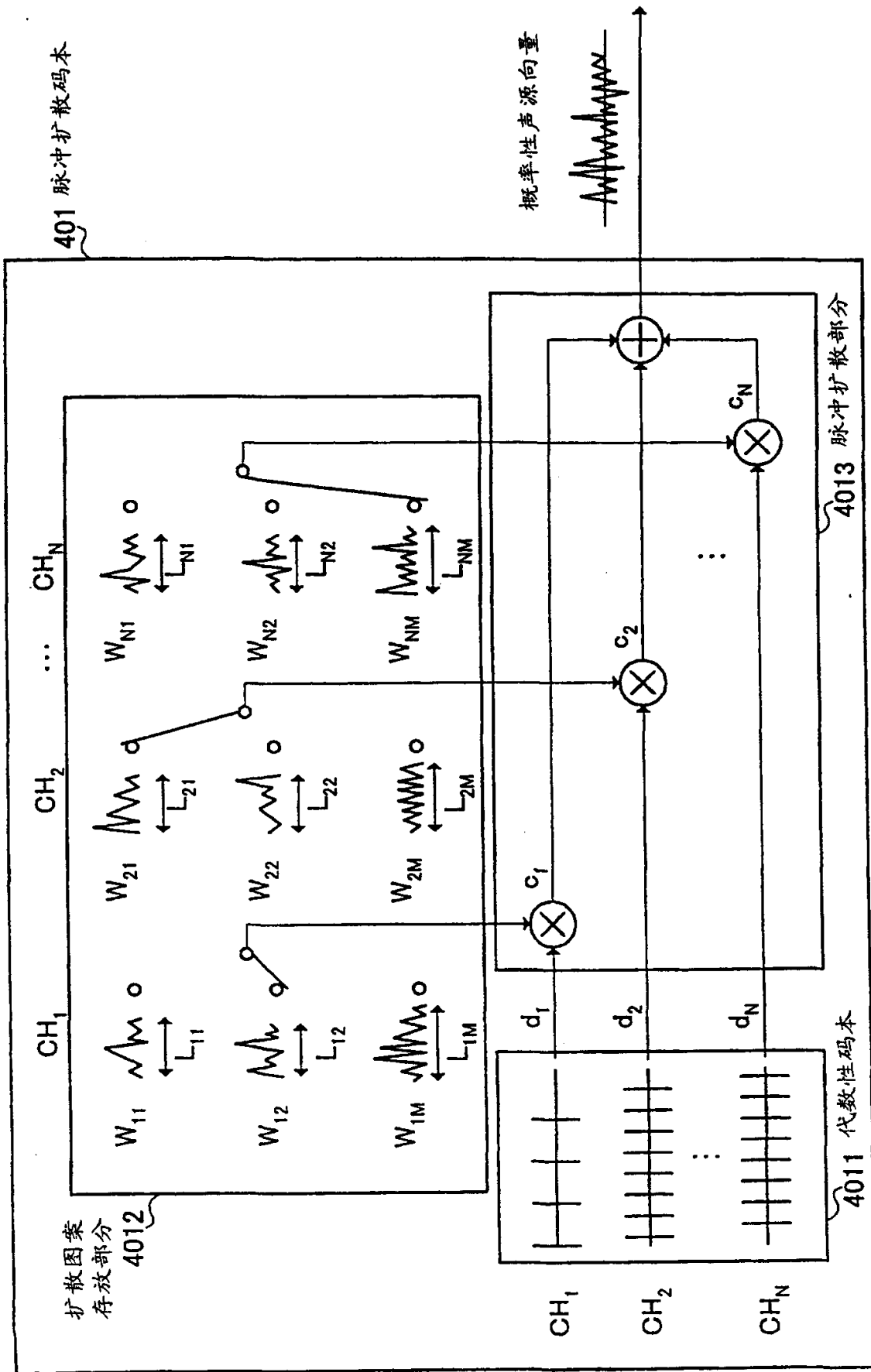


图 9

401 脉冲扩散码本

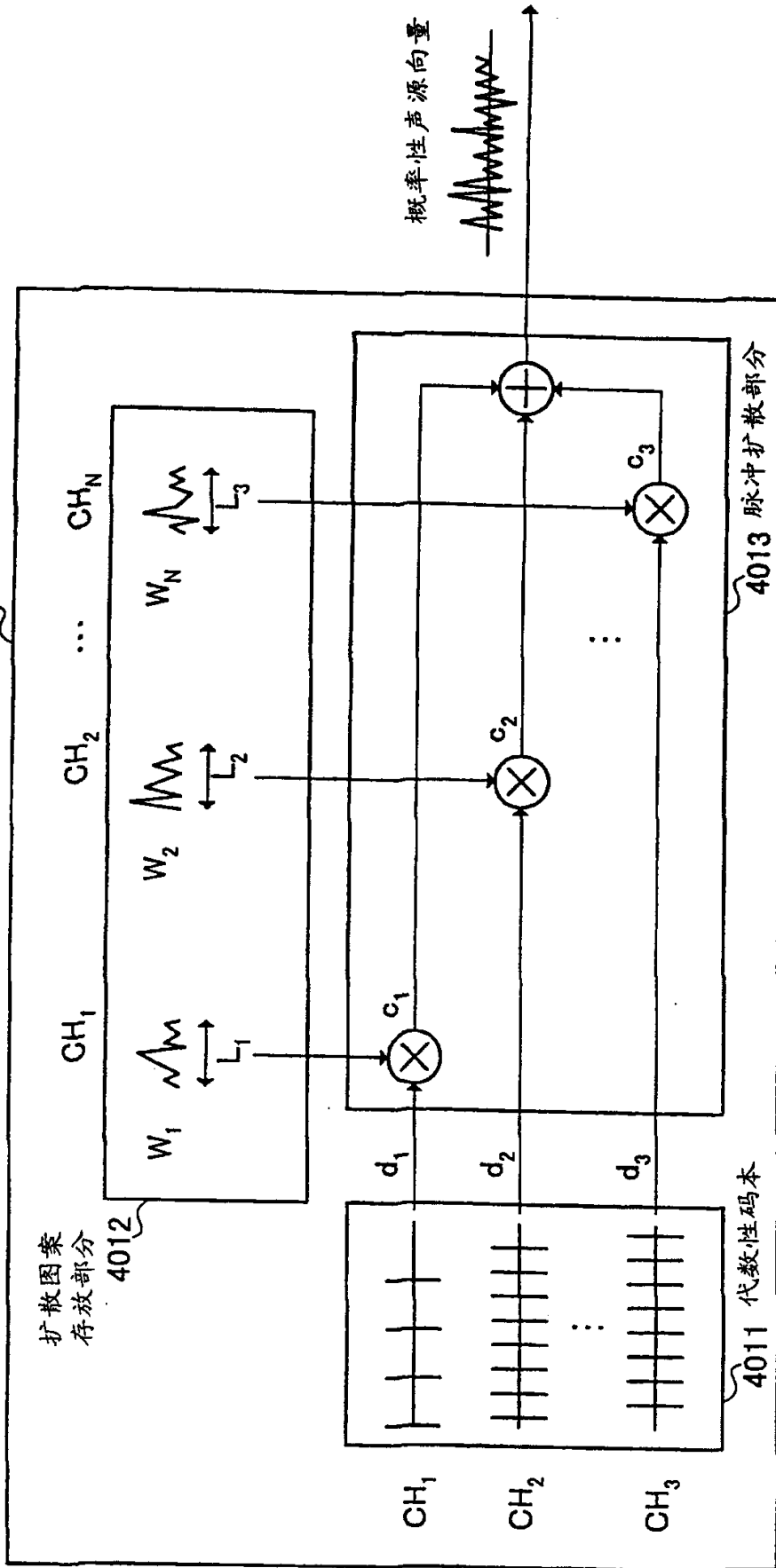


图 10



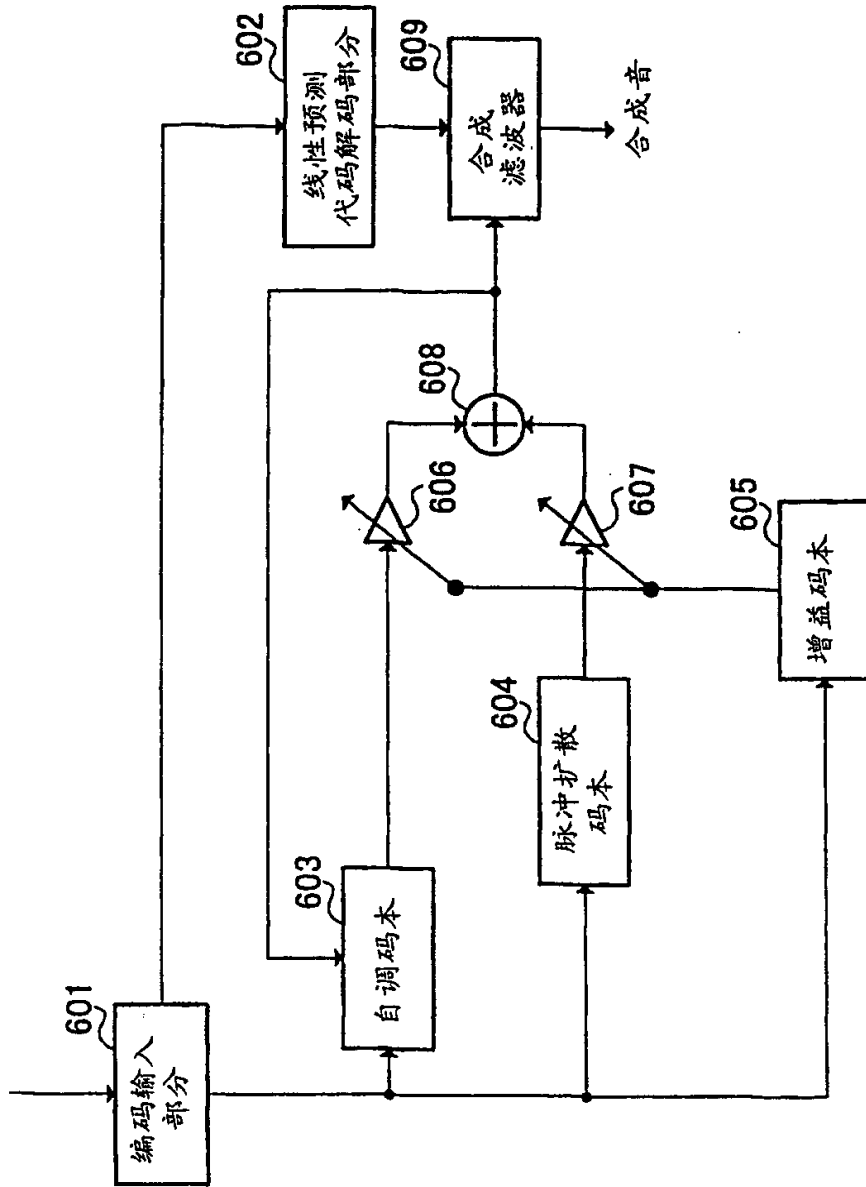


图 12



604 脉冲扩散码本

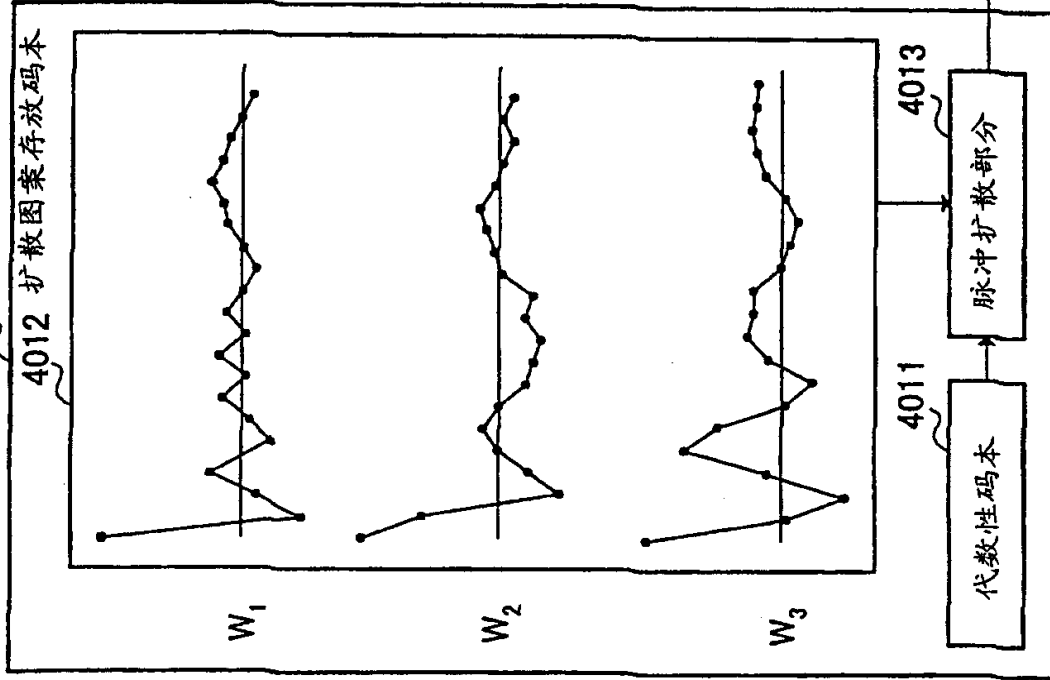


图 13B

507 脉冲扩散码本

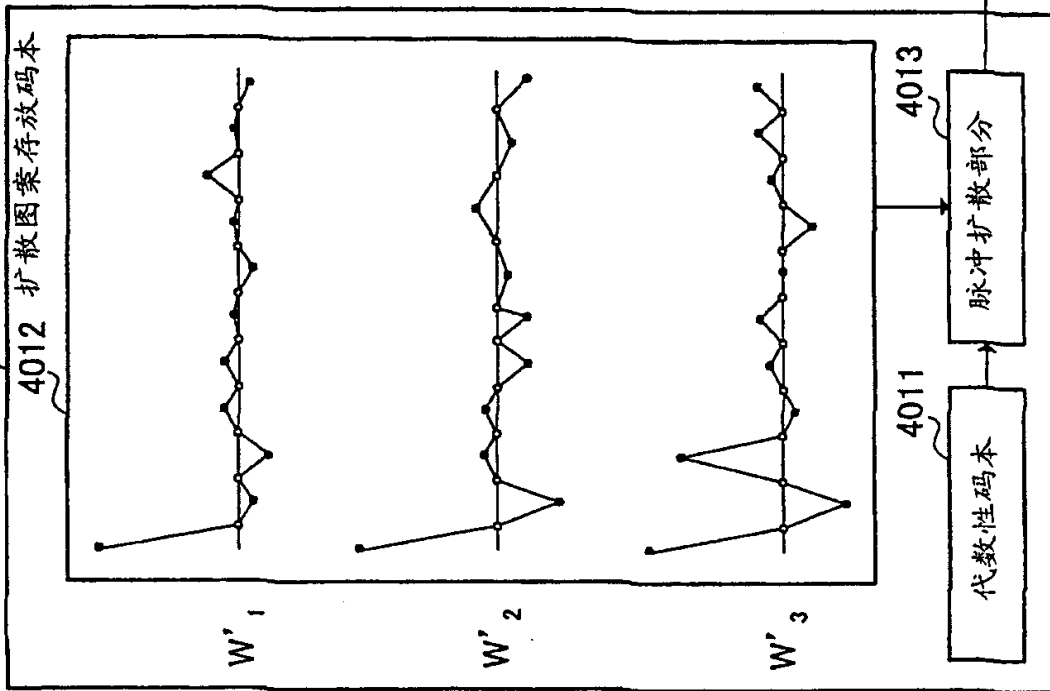


图 13A

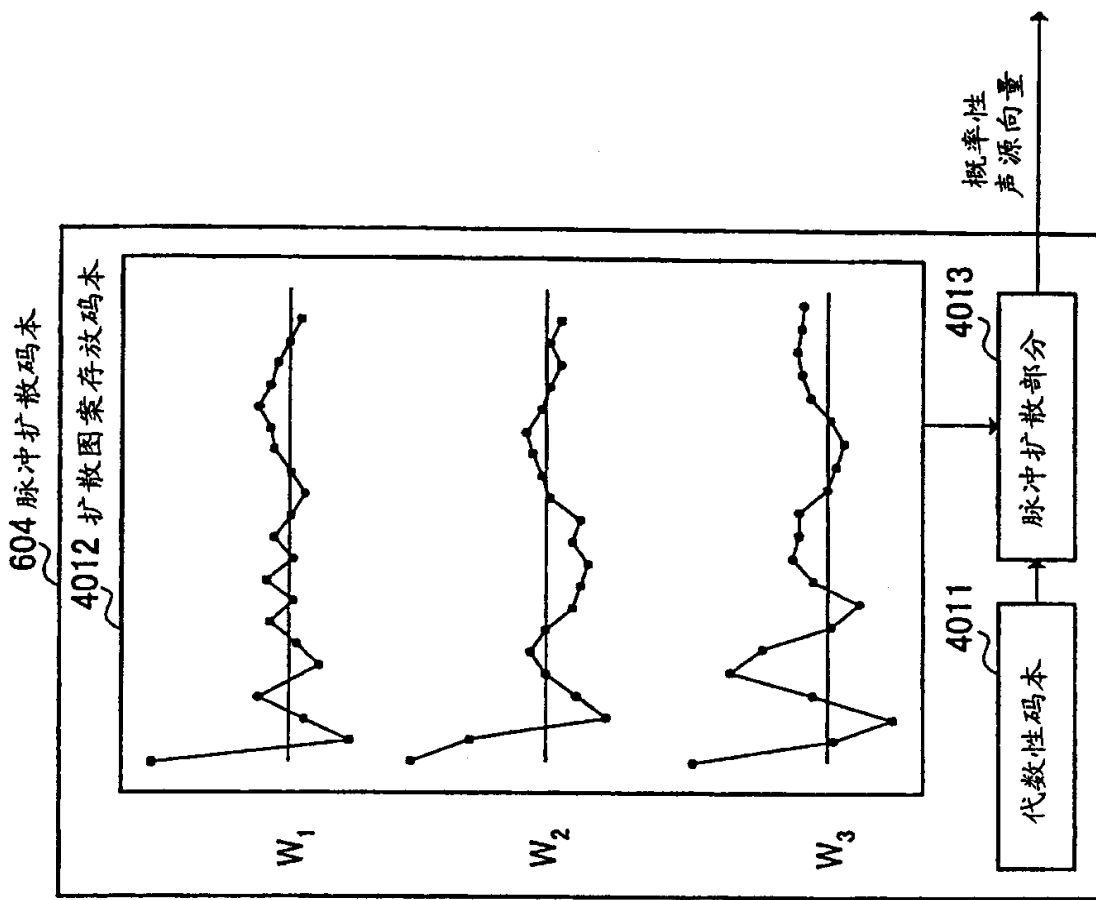


图 14B

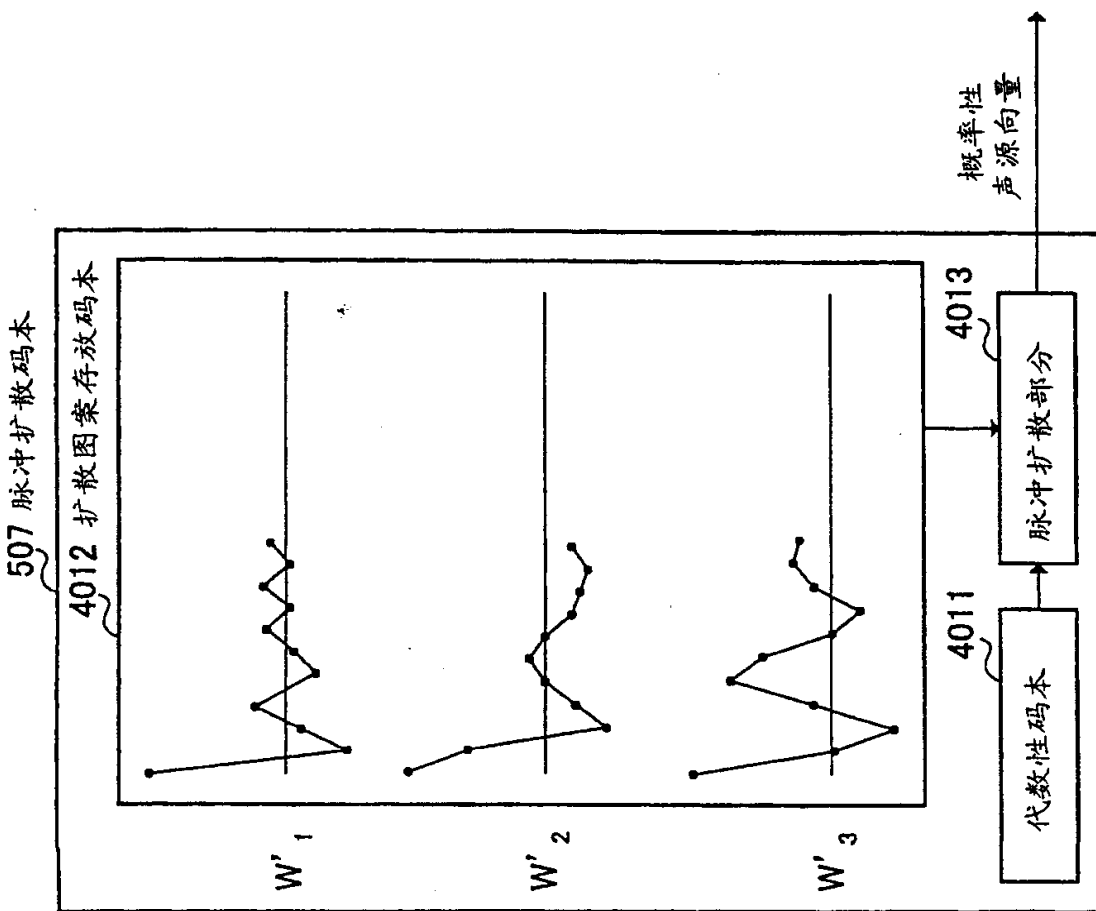


图 14A

权 利 要 求 书  
按照条约第 19 条的修改

---

1. 一种音频编码装置，其特征在于，

具备 LPC 合成手段、增益运算手段以及参数编码手段，所述 LPC 合成手段对于自调码本以及概率性码本中存放的自调声源以及概率性声源通过使用从输入音频中求出的 LPC 系数进行滤波而获得合成音，所述增益运算手段求出所述自调声源以及所述概率性声源的增益并且使用所述增益而获得所述输入音频与所述合成音之间的编码误差来搜索自调声源以及概率性声源的代码，所述参数编码手段使用求得的代码所对应的自调声源以及概率性声源进行增益的预测编码。

所述参数编码手段具备根据以前子帧的状态调整使用于所述预测编码中的预测系数的预测系数调整手段。

2、如权利要求 1 所述的音频编码装置，其特征在于，

所述预测系数调整手段当以前的子帧的状态为极大值或极小值时，调整所述预测系数使得减少影响。

3. 如权利要求 1 所述的音频编码装置，其特征在于，

所述参数编码手段具有码本，所述码本包含自调声源的增益向量、概率性声源的增益向量以及调整预测系数的系数。

4. 如权利要求 3 所述的音频编码装置，其特征在于，

在预测编码中，在求取状态与预测系数的积时，乘以所述状态对应的预测系数调整系数。

5. 如权利要求 1 所述的音频编码装置，其特征在于，

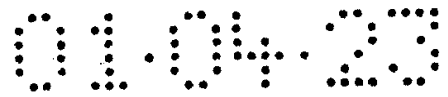
具备按每个状态对应地存放所述自调声源、概率性声源以及预测系数调整系数的存放手段。

6. 如权利要求 5 所述的音频编码装置，其特征在于，

在更新存放在所述存放手段中的所述自调声源以及所述概率性声源的状态时，也更新所述预测系数调整系数。

7. 一种音频编码装置，其特征在于，

所述音频编码是将一帧分解为多个子帧进行编码的 CELP 型音频编码装置，具备 LPC 合成手段、增益运算手段以及参数编码手段，所述 LPC 合成手段



对于自调码本以及概率性码本中存放的自调声源以及概率性声源通过使用从输入音频中求出的 LPC 系数进行滤波而获得合成音，所述增益运算手段求出所述自调声源以及所述概率性声源的增益，所述参数编码手段对于使用所述输入音频与所述合成音之间的编码误差而求得自调声源及概率性声源并且进行所述增益的向量量化。

所述音频编码装置还具备音调分析手段，所述音调分析手段是在进行最初子帧的自调码本搜索之前分析构成帧的多个子帧的音调并求出相关值，使用该相关值算出最近似于音调周期的值。

8. 如权利要求 7 所述的音频编码装置，其特征在于，  
还具备搜索范围决定手段，所述搜索范围决定手段根据所述音调分析手段获得的相关值以及最近音调周期的值，决定多个子帧的滞后的搜索范围。

9. 如权利要求 8 所述的音频编码装置，其特征在于，  
所述搜索范围设定手段使用由所述音调分析手段获得相关值以及最近似于音调周期的值而求得成为搜索范围的中心的假设音调。

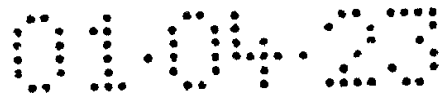
10. 如权利要求 9 所述的音频编码装置，其特征在于，  
所述搜索范围设定手段在假设音调周围的指定范围中设定滞后的搜索区间。

11. 如权利要求 8 所述的音频编码装置，其特征在于，  
所述搜索范围设定手段使得滞后为较短候补的较少，并且设定滞后的搜索区间。

12. 如权利要求 8 所述的音频编码装置，其特征在于，  
所述搜索范围设定手段在进行自调码本搜索时，在设定的范围中进行滞后搜索。

13. 一种由计算机可读取的记录媒体，其特征在于，  
所述记录媒体中存放了音频编码程序、存放了以前合成的声源信号的自调码本、存放了多个声源向量的概率性码本，

所述音频编码程序包含：对于存放在所述自调码本以及所述概率性码本中的自调声源以及概率性声源使用从输入音频中求得的 LPC 系数进行滤波而由此获得合成音的步骤；求取所述自调声源以及所述概率性声源的增益的步骤；使用所述输入音频与所述合成音之间的编码误差对于求得的自调声源、概率性声源以及所述增益进行向量量化的步骤，



在进行向量量化步骤中包含：根据多个量化对象向量以及使用于预测编码中的预测系数之间的编码误差而求得的量化对象向量的步骤；根据以前子帧的状态调整所述预测系数的步骤。

14. 一种计算机可读的记录媒体，其特征在于，

所述记录媒体中存放了音频编码程序、存放以前合成声的声源信号的自调码本、存放了多个声源向量的概率性码本，

所述音频编码程序包含：对于存放在所述自调码本以及所述概率性码本中的自调声源以及概率性声源使用从输入音频中求得的LPC系数进行滤波而由此获得合成音的步骤；求取所述自调声源以及所述概率性声源的增益的步骤；使用所述输入音频与所述合成音之间的编码误差对于求得的自调声源、概率性声源以及所述增益进行向量量化的步骤；在最初子帧的自调码本搜索之前分析构成帧的多个子帧的音调且求出相关值，并且使用所述相关值求出最近似于音调周期的值的步骤。

15. 一种音频编码装置，其特征在于，

具备叠加包含至少一个非零部分的向量(非零部分以外的部分值为0)与称作为扩散图案的固定波形而生成向量的脉冲扩散码本，

所述脉冲扩散码本具有与音频解码装置侧的脉冲扩散码本不同的构造。

16. 如权利要求15所述的音频编码装置，其特征在于，

作为所述脉冲扩散编码的构造部位的扩散图案存放部分存放了不同于音频解码装置侧的扩散图案存放部分存放的扩散图案的扩散图案。

17. 如权利要求16所述的音频编码装置，其特征在于，

所述扩散图案存放部分存放了将音频解码装置侧的扩散图案存放部分存放的扩散图案进行简易化选择获得的扩散图案。

18. 如权利要求16所述的音频编码装置，其特征在于，

所述扩散图案存放部分存放了将音频解码装置侧的扩散图案存放部分存放的扩散图案的构造部分以每规定间隔置换为0而获得的扩散图案。

19. 如权利要求16所述的音频编码装置，其特征在于，

所述扩散图案存放部分存放了将音频解码装置侧的扩散图案部分存放的扩散图案的构造部分以每N个采样(N为自然数)置换为0而获得的扩散图案。

20. 如权利要求19所述的音频编码装置，其特征在于，

所述扩散图案存放部分存放了将音频解码装置侧的扩散图案部分存放的

扩散图案的构造部分以每 1 个采样置换为 0 而获得的扩散图案。

21. 如权利要求 16 所述的音频编码装置, 其特征在于,  
所述扩散图案存放部分存放了将音频解码装置侧的扩散图案部分存放的扩散图案的构造部分以适当长度截断而获得的扩散图案。

22. 如权利要求 16 所述的音频编码装置, 其特征在于,  
所述扩散图案存放部分存放了将音频解码装置侧的扩散图案部分存放的扩散图案的构造部分以 N 个采样(N 为自然数)的长度截断而获得的扩散图案。

23. 如权利要求 16 所述的音频编码装置, 其特征在于,  
所述扩散图案存放部分存放了将音频解码装置侧的扩散图案部分存放的扩散图案的构造部分以一半的长度截断而获得的扩散图案。

24. 一种音频解码装置, 其特征在于,  
将具有由权利要求 15 所述的编码装置生成的音频编码的音频信号进行解码。

25. 一种信号处理用处理程序, 其特征在于,  
记载了实现权利要求 15 所述的音频编码装置的软件程序。

26. 一种信号处理用处理程序, 其特征在于,  
记载了实现权利要求 24 所述的音频编码装置的软件程序。

27. 一种音频编码解码系统, 其特征在于,  
音频编码装置侧具有的脉冲扩散码本的构造与音频解码装置侧具有的脉冲扩散码本的构造不同。

28. 如权利要求 27 所述的音频编码解码系统, 其特征在于,  
音频编码装置侧具有的脉冲扩散码本的构造与音频解码装置侧具有的脉冲扩散码本的构造不同, 而所述构造是指各自的脉冲扩散码本所具备的扩散图案的形状。

29. 如权利要求 28 所述的音频编码解码系统, 其特征在于,  
所述音频编码装置侧的扩散图案的形状是将音频解码装置侧的扩散图案的形状简化后的扩散图案的形状。

30. 如权利要求 27 所述的音频编码解码系统, 其特征在于,  
所述音频编码装置侧的扩散图案的形状是将音频解码装置侧的扩散图案的构造部分以每隔适当间隔置换为 0 而获得的形状。

31. 如权利要求 27 所述的音频编码解码系统, 其特征在于,

所述音频编码装置侧的扩散图案的形状是将音频解码装置侧的扩散图案的构造部分以每隔  $N$  个采样( $N$  为自然数)置换为 0 而获得的形状。

32. 如权利要求 27 所述的音频编码解码系统, 其特征在于,  
所述音频编码装置侧的扩散图案的形状是将音频解码装置侧的扩散图案的构造部分以每隔一个采样置换为 0 而获得的形状。

33. 如权利要求 27 所述的音频编码解码系统, 其特征在于,  
所述音频编码装置侧的扩散图案的形状是将音频解码装置侧的扩散图案的构造部分以适当长度截断获得的形状。

34. 如权利要求 27 所述的音频编码解码系统, 其特征在于,  
所述音频编码装置侧的扩散图案的形状是将音频解码装置侧的扩散图案的构造部分以  $N$  个采样( $N$  为自然数)的长度截断获得的形状。

35. 如权利要求 27 所述的音频编码解码系统, 其特征在于,  
所述音频编码装置侧的扩散图案的形状是将音频解码装置侧的扩散图案的构造部分以一半长度截断获得的形状。

36. 一种通信用基地局, 其特征在于,  
具备权利要求 25 所述的信号处理用处理程序。

37. 一种通信用终端, 其特征在于,  
具备权利要求 25 所述的信号处理用处理程序。

38. 一种无线通信系统, 其特征在于,  
将权利要求 36 所述的通信用基地局以及权利要求 27 所述的通信用终端以无线网络连接。