



(12) 发明专利

(10) 授权公告号 CN 101009096 B

(45) 授权公告日 2011.01.26

(21) 申请号 200610165246.X

(22) 申请日 2006.12.15

(73) 专利权人 清华大学

地址 100084 北京市 100084 信箱 82 分箱清华大学专利办公室

(72) 发明人 崔慧娟 唐昆 李晔 洪侃

(51) Int. Cl.

G10L 19/00(2006.01)

G10L 19/08(2006.01)

G10L 19/14(2006.01)

G10L 11/06(2006.01)

(56) 对比文件

US 6687667 B1, 2004.02.03, 全文.

CN 1614686 A, 2005.05.11, 全文.

CN 1173690 A, 1998.02.18, 全文.

US 2002/0062209 A1, 2002.05.23, 全文.

CN 1338095 A, 2002.02.27, 全文.

周志杰, 胡光锐. 采用非线性网络实

现清浊音判决. 南京航空航天大学学报 30

1. 1998, 30(1), 47-50.

毛六平, 唐艳, 王耀南. 基于模糊 MLP 网络的清浊音判决. 湖南师范大学自然科学学报 26 4. 2003, 26(4), 27-30.

李军林 杜松 崔慧娟 唐昆. 0.8kb/s 高质量声码器算法. 清华大学学报(自然科学版) 43 1. 2003, 43(1), 12-15.

朱益厅, 李永明, 陈弘毅. 一种多带清浊音判决方法. 微电子学与计算机 5. 1999, (5), 1-4.

于俊凤, 曹俊兴. 基于组合参数的清浊音判决方法. 太原理工大学学报 35 4. 2004, 35(4), 467-469.

审查员 游晓梅

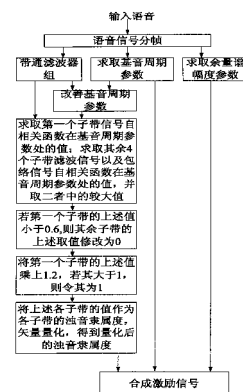
权利要求书 1 页 说明书 5 页 附图 2 页

(54) 发明名称

子带清浊音模糊判决的方法

(57) 摘要

子带清浊音模糊判决方法, 属于低速率语音压缩编码技术领域. 其特征在于, 在利用目前通用的带通浊音度参数提取方法得到除第一子带以外的其余四个子带的浊音隶属度后, 将第一个子带滤波后信号的自相关函数在基音周期参数处的值乘以一个设定的增益因子, 作为第一个子带的浊音隶属度, 如果结果超过 1, 则令其等于 1; 将各子带的浊音隶属度一起进行矢量量化, 得到量化后各子带的浊音隶属度, 和基音周期参数等一起合成激励信号. 这种方法可以提高合成语音的自然度. 该方法适合低速率参数语音编码.



CN 101009096 B

1. 子带清浊音模糊判决的方法,其特征在于该方法包括以下步骤:

(1) 对输入语音信号样点按时间顺序分帧;

(2) 对当前帧提取基音周期参数;

(3) 对当前帧提取余量谱幅度参数;

(4) 将当前帧语音信号经过 5 个 6 阶的巴特沃思滤波器进行滤波,得到通带分别为 0-500,500-1000,1000-2000,2000-3000 和 3000-4000Hz 的 5 个子带信号;

(5) 根据第一个子带的语音信号对步骤 (2) 中求取的基音周期参数进一步改善,改善后的基音周期参数为 \hat{P}_n ;

(6) 求取第一个子带语音信号的自相关函数在步骤 (5) 中求取的基音周期参数 \hat{P}_n 位置上的值 $Vb_{p_{n,1}}$;

(7) 求取其余 4 个子带语音信号的自相关函数以及其时间包络的自相关函数在步骤 (5) 中求取的基音周期参数 \hat{P}_n 位置上的值,并分别对每个子带取两者中的较大值:

$$Vb_{p_{n,i}} = \max \{Vbp'_{n,i}, Vbp''_{n,i}\} \quad i = 2, 3, 4, 5$$

其中, $Vbp'_{n,i}$ 表示当前帧,即第 n 帧的第 i 子带的自相关函数在 \hat{P}_n 位置上的值; $Vbp''_{n,i}$ 表示当前帧,即第 n 帧的第 i 子带的时间包络的自相关函数在 \hat{P}_n 位置上的值;

(8) 如果步骤 (6) 中求取的值 $Vb_{p_{n,1}}$ 小于 0.6,则将步骤 (7) 中的值全部修改为 0,即 $Vb_{p_{n,i}} = 0 \quad i = 2, 3, 4, 5$,否则保持不变;

(9) 将步骤 (6) 中求取的值 $Vb_{p_{n,1}}$ 乘以一个设定的增益因子,该增益因子设定为 1.2,如果相乘后结果大于 1,则令其等于 1,即:

$$Vb_{p_{n,1}} = \begin{cases} 1 & \text{if } Vb_{p_{n,1}} \times 1.2 > 1 \\ Vb_{p_{n,1}} \times 1.2 & \text{if } Vb_{p_{n,1}} \times 1.2 \leq 1 \end{cases}$$

(10) 将步骤 (8) 与步骤 (9) 中求取的值作为各子带的浊音隶属度,合并成为一个矢量 $\overline{Vbp} = (Vb_{p_{n,1}}, Vb_{p_{n,2}}, Vb_{p_{n,3}}, Vb_{p_{n,4}}, Vb_{p_{n,5}})$, 一起进行矢量量化; 矢量量化采用对码本中的码字进行全搜索的方法获得最优量化码字:

$$\overline{Vbp} = \arg \min_{\overline{Vbp}_i \in C} Er(\overline{Vbp}, \overline{Vbp}_i)$$

其中 \overline{Vbp} 表示待量化的输入矢量, \overline{Vbp}_i 表示码本中的码字矢量, C 表示码本, i 是码字矢量在码本中的索引值, $Er()$ 函数表示特定的失真测度,这里采用最小加权均方误差,即 $Er(\overline{Vbp}, \overline{Vbp}_i) = \sum_{k=1}^5 W_k (Vb_{p_k} - Vb_{p_{i,k}})^2$, 其中 Vb_{p_k} 为待量化矢量的第 k 个分量, $Vb_{p_{i,k}}$ 为码本中第 i 个码字矢量的第 k 个分量, W 为加权因子向量,这里取 $W = [16, 8, 4, 2, 1]$, 即 $W_1 = 16$, $W_2 = 8$, $W_3 = 4$, $W_4 = 2$, $W_5 = 1$; 量化后得到量化的各子带的浊音隶属度 \overline{Vbp} ; (11) 将步骤 (10) 中得到量化的各子带的浊音隶属度、步骤 (3) 中的余量谱幅度参数以及步骤 (5) 中的基音周期参数一起用来合成激励信号。

2. 按权利要求 1 所述的子带清浊音模糊判决方法,其特征在于,所述步骤 (1) 中每一帧包含 180 个或 200 个语音样点。

子带清浊音模糊判决的方法

技术领域

[0001] 本发明属于语音编码技术领域，特别涉及低速率参数语音编码技术。

背景技术

[0002] 语音编码在通信系统、语音存储回放系统、具有语音功能的消费类产品中有广泛的应用。近些年来国际电信联盟 (ITU)、一些区域组织和一些国家相继制定了一系列语音压缩编码标准，在编码速率为 1.2kb/s 到 16kb/s 上得到了令人满意的语音质量。目前国内外的研究主要集中在 1.2kb/s 速率以下高质量语音压缩编码上，主要用于无线通信、保密通信、大容量语音存储回放等。低速率语音编码中激励信号的合成非常重要，而多带激励信号合成是目前激励信号合成的重要手段。多带激励信号合成主要依靠于带通浊音度参数，目前带通浊音度参数求取的主要步骤如下：

[0003] (1) 对输入语音信号样点按时间顺序分帧；

[0004] (2) 对当前帧提取基音周期参数；

[0005] (3) 对当前帧提取余量谱幅度参数；

[0006] (4) 将当前帧语音信号经过 5 个 6 阶的巴特沃思滤波器进行滤波，其通带分别为 0-500, 500-1000, 1000-2000, 2000-3000 和 3000-4000Hz；

[0007] (5) 根据第一个子带滤波后的语音信号对步骤 (2) 中求取的基音周期参数进一步改善，改善后的基音周期值为 \hat{P}_n ；

[0008] (6) 求取第一个子带滤波后语音信号的自相关函数在步骤 (5) 中求取的基音周期参数位置上的值 $Vb_{p_{n,1}}$ ；

[0009] (7) 求取其 4 个子带滤波后语音信号的自相关函数及其时间包络的自相关函数在步骤 (5) 中求取的基音周期参数 \hat{P}_n 位置上的值，并对每个子带取两者中的较大值；

[0010] $Vb_{p_{n,j}} = \max \{Vb_{p_{n,j}'}, Vb_{p_{n,j}''}\} \quad i = 2, 3, 4, 5$

[0011] 其中， $Vb_{p_{n,i}'}$ 表示当前帧 (第 n 帧) 的第 i 子带的自相关函数在 \hat{P}_n 位置上的值； $Vb_{p_{n,i}''}$ 表示当前帧 (第 n 帧) 的第 i 子带的时间包络的自相关函数在 \hat{P}_n 位置上的值。

[0012] (8) 如果步骤 (6) 中求取的值 $Vb_{p_{n,1}}$ 小于 0.6，则将步骤 (7) 中的值全部修改为 0，即， $Vb_{p_{n,j}} = 0 \quad i = 2, 3, 4, 5$ ，否则保持不变；

[0013] (9) 将步骤 (7) 与步骤 (8) 中求取的值做二值化，即分别与阈值 0.6 作比较，如果大于 0.6，则认为当前子带为浊音，用 1 表示其带通浊音度，否则为清音，用 0 表示其带通浊音度，即：

[0014]
$$Vb_{p_{n,i}} = \begin{cases} 1 & \text{if } Vb_{p_{n,i}} > 0.6 \text{ 浊音} \\ 0 & \text{if } Vb_{p_{n,i}} \leq 0.6 \text{ 清音} \end{cases} \quad i = 1, 2, 3, 4, 5$$

[0015] (10) 将步骤 (9) 中得到的各子带的带通浊音度、步骤 (3) 中的余量谱幅度参数以及步骤 (5) 中的基音周期参数一起用来合成激励信号。

[0016] 上述已有技术对于子带浊音度的表示采用了 0、1 判决，即当前子带不是浊音就是

清音。实际上子带的浊音性与清音性之间并没有明确的分界,简单的将子带划分为清音或者浊音会导致语音帧过渡的不自然。

[0017] 如图 1 所示,在表示带通浊音度时,原有技术采用简单的 0、1 判决,这会使低码率参数语音编码合成感增强,自然度下降。

发明内容

[0018] 本发明的目的是为克服已有技术的不足之处,改变旧有的不是清音就是浊音的简单判决方法,提出带通清浊音模糊判决的方法,增强语音的自然度。

[0019] 本发明提出的子带清浊音模糊判决的方法,包括以下步骤:

[0020] (1) 对输入语音信号样点按时间顺序分帧;

[0021] (2) 对当前帧提取基音周期参数;

[0022] (3) 对当前帧提取余量谱幅度参数;

[0023] (4) 将当前帧语音信号经过 5 个 6 阶的巴特沃思滤波器进行滤波,得到通带分别为 0-500, 500-1000, 1000-2000, 2000-3000 和 3000-4000Hz 的 5 个子带信号;

[0024] (5) 根据第一个子带的语音信号对步骤 (2) 中求取的基音周期参数进一步改善,改善后的基音周期值为 \hat{P}_n ;

[0025] (6) 求取第一个子带语音信号的自相关函数在步骤 (5) 中求取的基音周期参数 \hat{P}_n 位置上的值 $V_{bp_{n,1}}$;

[0026] (7) 求取其 4 个子带语音信号的自相关函数以及其时间包络的自相关函数在步骤 (5) 中求取的基音周期参数 \hat{P}_n 位置上的值,并分别对每个子带取两者中的较大值:

[0027] $V_{bp_{n,i}} = \max \{ V_{bp_{n,i}}', V_{bp_{n,i}}'' \} \quad i = 2, 3, 4, 5$

[0028] 其中, $V_{bp_{n,i}}'$ 表示当前帧,即第 n 帧的第 i 子带的自相关函数在 \hat{P}_n 位置上的值; $V_{bp_{n,i}}''$ 表示当前帧,即第 n 帧的第 i 子带的时间包络的自相关函数在 \hat{P}_n 位置上的值;

[0029] (8) 如果步骤 (6) 中求取的值 $V_{bp_{n,1}}$ 小于 0.6,则将步骤 (7) 中的值全部修改为 0,即 $V_{bp_{n,i}} = 0 \quad i = 2, 3, 4, 5$,否则保持不变;

[0030] (9) 将步骤 (6) 中求取的值 $V_{bp_{n,1}}$ 乘以一个设定的增益因子,该增益因子设定为 1.2,如果相乘后结果大于 1,则令其等于 1,即:

[0031]
$$V_{bp_{n,1}} = \begin{cases} 1 & \text{if } V_{bp_{n,1}} \times 1.2 > 1 \\ V_{bp_{n,1}} \times 1.2 & \text{if } V_{bp_{n,1}} \times 1.2 \leq 1 \end{cases}$$

[0032] (10) 将步骤 (8) 与步骤 (9) 中求取的值作为各子带的浊音隶属度,合并成为一个矢量 $\overline{Vbp} = (V_{bp_{n,1}}, V_{bp_{n,2}}, V_{bp_{n,3}}, V_{bp_{n,4}}, V_{bp_{n,5}})$, 一起进行矢量量化;矢量量化采用对码本中的码字进行全搜索的方法获得最优量化码字:

[0033]

$$\widehat{\overline{Vbp}} = \arg \min_{\overline{Vbp}_i \in C} Er(\overline{Vbp}, \overline{Vbp}_i)$$

[0034] 其中 \overline{Vbp} 表示待量化的输入矢量, \overline{Vbp}_i 表示码本中的码字矢量, C 表示码本, i 是码字矢量在码本中的索引值, $Er()$ 函数表示特定的失真测度,这里采用最小加权均方误差,即

$Er(\overline{Vbp}, \overline{Vbp_i}) = \sum_{k=1}^5 W_k (Vbp_k - Vbp_{i,k})^2$, 其中 Vbp_k 为待量化矢量的第 k 个分量, $Vbp_{i,k}$ 为码本中第 i 个码字矢量的第 k 个分量, W 为加权因子向量, 这里取 $W = [16, 8, 4, 2, 1]$; 量化后得到量化的各子带的浊音隶属度 \widehat{Vbp} ;

[0035] (11) 将步骤 (10) 中得到的各子带的浊音隶属度、步骤 (3) 中的余量谱幅度参数以及步骤 (5) 中的基音周期参数一起用来合成激励信号。

[0036] 本发明的特点是对低速率参数语音编码中的带通浊音度参数采用了模糊判决的方法。原有的技术对带通浊音度参数采用了 0、1 判决, 增加了语音的合成感, 降低了自然度。本发明采用模糊数学的相关原理, 用自相关函数来决定当前子带的浊音隶属度。并考虑到第一个子带往往受噪声的影响比较大, 因此其浊音隶属度要乘上一个增益因子, 更精确合理地描述了各个子带的浊音度。

[0037] 本方法可以提高合成语音的自然度。该方法最适合 600 ~ 800b/s 低速率参数语音编码, 将在信号处理器芯片 DSP 上实现。

附图说明

[0038] 图 1 为已有技术的带通浊音度判决方法流程框图。

[0039] 图 2 为本发明提出的带通浊音度模糊判决方法流程框图。

具体实施方式

[0040] 本发明提出的带通浊音度模糊判决方法结合附图及实施例进一步说明如下:

[0041] 本发明的方法流程如图 2 所示, 包括以下步骤:

[0042] (1) 对输入语音信号样点按时间顺序分帧;

[0043] (2) 对当前帧提取基音周期参数;

[0044] (3) 对当前帧提取余量谱幅度参数;

[0045] (4) 将当前帧语音信号经过 5 个 6 阶的巴特沃思滤波器进行滤波, 得到通带分别为 0-500, 500-1000, 1000-2000, 2000-3000 和 3000-4000Hz 的 5 个子带信号;

[0046] (5) 根据第一个子带的语音信号对步骤 (2) 中求取的基音周期参数进一步改善, 改善后的基音周期值为 \hat{P}_n ;

[0047] (6) 求取第一个子带语音信号的自相关函数在步骤 (5) 中求取的基音周期参数 \hat{P}_n 位置上的值 $Vbp_{n,1}$;

[0048] (7) 求取其余 4 个子带语音信号的自相关函数及其时间包络的自相关函数在步骤 (5) 中求取的基音周期参数 \hat{P}_n 位置上的值, 并分别对每个子带取两者中的较大值:

[0049] $Vbp_{n,j} = \max \{Vbp_{n,j}', Vbp_{n,j}''\} i = 2, 3, 4, 5$

[0050] 其中, $Vbp_{n,i}'$ 表示当前帧, 即第 n 帧的第 i 子带的自相关函数在 \hat{P}_n 位置上的值; $Vbp_{n,i}''$ 表示当前帧, 即第 n 帧的第 i 子带的时间包络的自相关函数在 \hat{P}_n 位置上的值;

[0051] (8) 如果步骤 (6) 中求取的值 $Vbp_{n,1}$ 小于 0.6, 则将步骤 (7) 中的值全部修改为 0, 即 $Vbp_{n,i} = 0 i = 2, 3, 4, 5$, 否则保持不变;

[0052] (9) 将步骤 (6) 中求取的值 $Vbp_{n,1}$ 乘以一个设定的增益因子, 该增益因子设定为

1.2, 如果相乘后结果大于 1, 则令其等于 1, 即 :

$$[0053] \quad V_{bp_{n,1}} = \begin{cases} 1 & \text{if } V_{bp_{n,1}} \times 1.2 > 1 \\ V_{bp_{n,1}} \times 1.2 & \text{if } V_{bp_{n,1}} \times 1.2 \leq 1 \end{cases}$$

[0054] (10) 将步骤 (8) 与步骤 (9) 中求取的值作为各子带的浊音隶属度, 合并成为一个矢量 $\overline{Vbp} = (V_{bp_{n,1}}, V_{bp_{n,2}}, V_{bp_{n,3}}, V_{bp_{n,4}}, V_{bp_{n,5}})$, 一起进行矢量量化; 矢量量化采用对码本中的码字进行全搜索的方法获得最优量化码字 :

[0055]

$$\widehat{Vbp} = \arg \min_{\overline{Vbp}_i \in C} Er(\overline{Vbp}, \overline{Vbp}_i)$$

[0056] 其中 \overline{Vbp} 表示待量化的输入矢量, \overline{Vbp}_i 表示码本中的码字矢量, C 表示码本, i 是码字矢量在码本中的索引值, $Er()$ 函数表示特定的失真测度, 这里采用最小加权均方误差, 即 $Er(\overline{Vbp}, \overline{Vbp}_i) = \sum_{k=1}^5 W_k (V_{bp_k} - V_{bp_{i,k}})^2$, 其中 V_{bp_k} 为待量化矢量的第 k 个分量, $V_{bp_{i,k}}$ 为码本中第 i 个码字矢量的第 k 个分量, W 为加权因子向量, 这里取 $W = [16, 8, 4, 2, 1]$; 量化后得到量化的各子带的浊音隶属度 \widehat{Vbp} ;

[0057] (11) 将步骤 (10) 中得到的各子带的浊音隶属度、步骤 (3) 中的余量谱幅度参数以及步骤 (5) 中的基音周期参数一起用来合成激励信号。

[0058] 本发明上述方法各步骤的具体实施例分别详细说明如下 :

[0059] 上述方法步骤 (1) 对输入语音信号样点按时间顺序分帧的实施例是按 8kHz 频率采样、已经过高通滤波去除工频干扰的语音样点。每 25ms, 也就是 200 个语音样点构成一帧 ;

[0060] 上述方法步骤 (2) 的实施例为 : 按美国政府 2400b/s 混合激励的线性预测 (MELP) 语音编码算法标准所描述的方法求取当前帧的基音周期参数 p_n ;

[0061] 上述方法步骤 (3) 的实施例为 : 按美国政府 2400b/s 混合激励的线性预测 (MELP) 语音编码算法标准所描述的方法求当前帧的余量谱幅度参数, 记为矢量 R , 其维数为 k , $R = [r_1, r_2, \dots, r_k]$ $k = 10$;

[0062] 上述方法步骤 (4) 的实施例为 : 按美国政府 2400b/s 混合激励的线性预测 (MELP) 语音编码算法标准所描述的方法对当前帧语音信号进行带通滤波 ;

[0063] 上述方法步骤 (5) 中的实施例为 : 按美国政府 2400b/s 混合激励的线性预测 (MELP) 语音编码算法标准所描述的方法对当前帧语音信号的基音周期参数进行进一步改善, 改善后的基音周期参数为 \hat{P}_n ;

[0064] 上述方法步骤 (6) 的实施例为 : 按美国政府 2400b/s 混合激励的线性预测 (MELP) 语音编码算法标准所描述的方法求取当前帧语音信号第一个子带语音信号自相关函数及其在 \hat{P}_n 处的值, 作为该子带的浊音隶属度 V_{bp_1} ;

[0065] 上述方法步骤 (7) 的实施例为 : 按美国政府 2400b/s 混合激励的线性预测 (MELP) 语音编码算法标准所描述的方法求取当前帧语音信号后 4 个子带滤波信号以及包络信号自相关函数在 \hat{P}_n 处的值, 并取两者中的较大值作为该子带的浊音隶属度 V_{bp_i} ;

[0066] 上述方法步骤 (8) 的实施例为 : 若 $V_{bp_1} < 0.6$, 则 $V_{bp_i} = 0$, $i = 2, 3, 4, 5$;

[0067] 上述方法步骤 (9) 的实施例为 : $V_{bp_1} = V_{bp_1} \times 1.2$, 如果 $V_{bp_1} > 1$, 则令 $V_{bp_1} = 1$;

[0068] 上述方法步骤 (10) 的实施例的具体做法为 :对带通浊音度参数 \overline{Vbp} 进行矢量量化, $\overline{Vbp} = [Vbp_1, Vbp_2, Vbp_3, Vbp_4, Vbp_5]$ 。矢量量化的码本需要单独训练, 训练算法采用 LBG 算法迭代生成。码字训练与搜索时, 失真测度采用最小加权均方误差, 加权因子 W , 即

$$Er = \sum_{i=1}^5 W_i (Vbp_i - \hat{Vbp}_i)^2$$

其中, Er 为失真测度, Vbp_i 为训练矢量或者待量化矢量的第 i 个分量, \hat{Vbp}_i 为码本中矢量的第 i 个分量, W 为加权因子向量, $W = [16, 8, 4, 2, 1]$;

码字搜索采用全搜索的方法获得最优量化码字, 即遍历码本中所有的码字取失真最小的码字作为最后量化的结果。

[0069] 上述方法步骤 (11) 的实施例的具体做法为 :按美国政府 2400b/s 混合激励的线性预测 (MELP) 语音编码算法标准所描述的方法, 利用带通浊音度参数、基音周期参数、余量谱幅度参数进行激励信号合成。

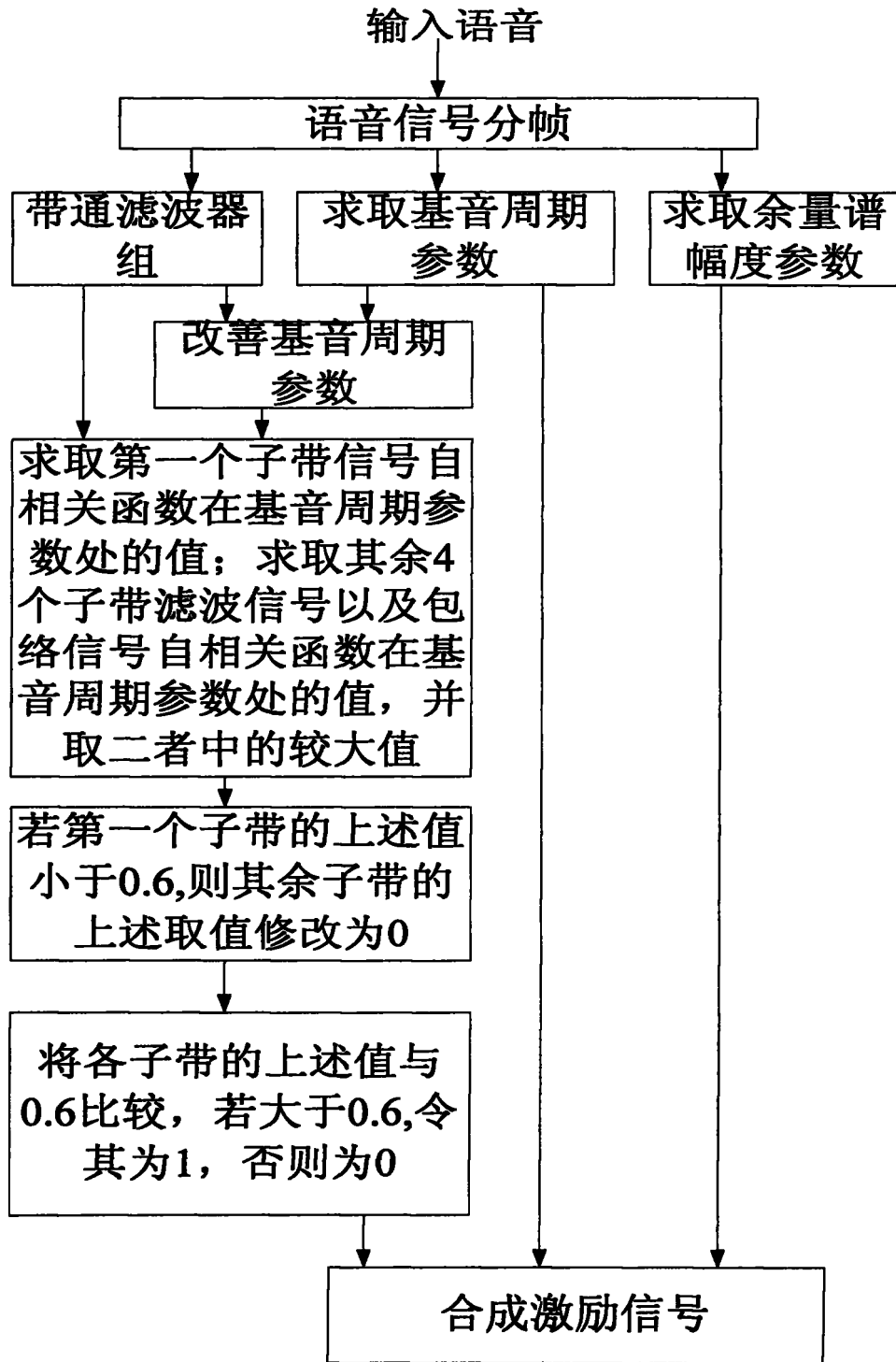


图 1

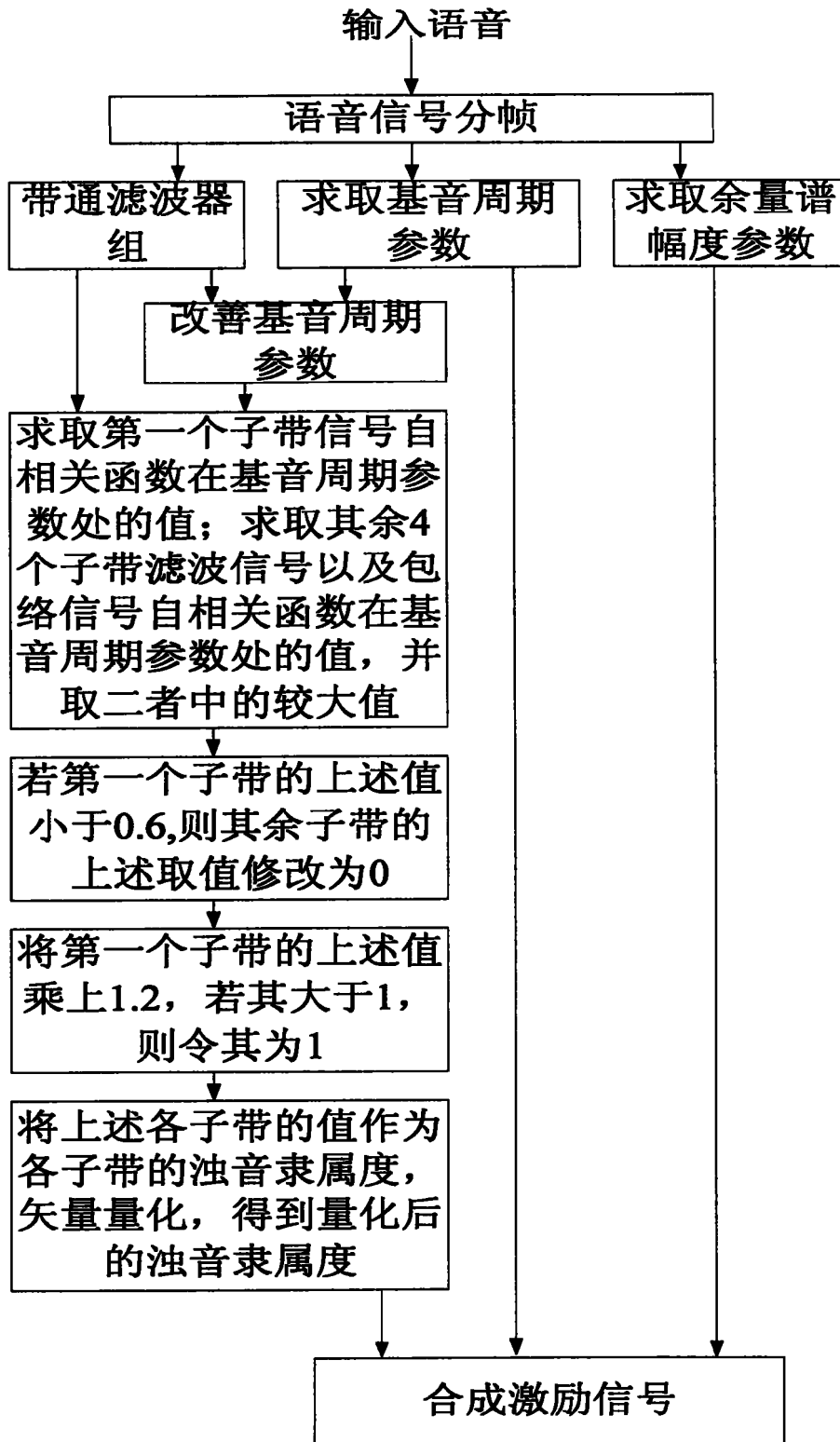


图2