

(12) 发明专利

(10) 授权公告号 CN 101025918 B

(45) 授权公告日 2011.06.29

(21) 申请号 200710062862.7

US 2006/0173675 A1, 2006.08.03, 全文.

(22) 申请日 2007.01.19

审查员 王玥

(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 张树华 窦维蓓 杨华中 张斌

(51) Int. Cl.

G10L 19/14 (2006.01)

G10L 19/12 (2006.01)

(56) 对比文件

EP 0673014 A2, 1995.09.20, 全文.

EP 1278184 A2, 2003.01.22, 全文.

CN 1116022 A, 1996.01.31, 全文.

EP 0932141 A2, 1999.07.28, 全文.

US 5982817 A, 1999.11.09, 全文.

CN 1326269 A, 2001.12.12, 全文.

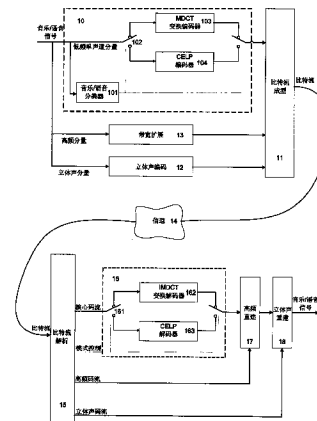
权利要求书 4 页 说明书 10 页 附图 5 页

(54) 发明名称

一种语音 / 音乐双模编解码无缝切换方法

(57) 摘要

一种语音 / 音乐双模编解码无缝切换方法。当双模编解码器从基于 CELP 的语音模式切换到基于 MDCT 的音乐模式时,对切换前的最后一个 CELP 帧时域音频信号的尾部采用加窗和折叠处理,切换的连续性由 MDCT 变换的重叠相加性质保证;当双模编解码器从基于 MDCT 的音乐模式切换到基于 CELP 的语音模式时,切换前最后一个 MDCT 帧采用一种新的窗型使其与切换后第一个 CELP 帧无时域重叠,切换的连续性由预编码技术保证。该方法对采用 CELP 模式的音频时域信号进行特定的降采样处理,以适应两种编解码模式对输入音频信号采样率的不同要求以及音频信号动态的时频域特性。



1. 一种语音 / 音乐双模编码无缝切换方法,其特征在于依次含有以下步骤:

步骤 (1),在数字集成电路中构成一个编码器,在该编码器至少有四个子模块:语音 / 音乐分类子模块,区分语音帧和音乐帧;语音 / 音乐切换模块,对语音帧和音乐帧作无缝切换;修正余弦变换的音乐编码子模块,采用先进音频编码 AAC 对音乐编码;语音编码子模块,采用码激励线性预测 CELP 对语音编码;

步骤 (2),该语音 / 音乐分类子模块把输入的音频信号划分为语音帧和音乐帧;

步骤 (3),把步骤 (2) 所得到的结果送入语音 / 音乐切换模块,判断是否为语音帧;

步骤 (4),在步骤 (3) 中,若判断为语音帧,则执行步骤 (5),若判断为音乐帧,则执行步骤 (6);

步骤 (5),判断步骤 (4) 得到的语音帧的上一帧是否为语音帧:若其上一帧是音乐帧,则执行步骤 (5.1);若其上一帧是语音帧,则执行步骤 (5.2);

步骤 (5.1),若步骤 (5) 判断的结果,所选语音帧的上一帧是音乐帧,则进行预编码:首先从上一帧经过编码量化后的整数修正余弦变换 MDCT 频谱数据中恢复第 L 帧的前 $N/2$ 个样点, L 是当前帧的帧序号, N 是一帧的采样点数,为偶数,各帧相同;再从这 $N/2$ 个点中截取最后的 P 个样点, P 为一个 CELP 编码帧的样点数,降采样后作为语音编码器 CELP 的输入;编码该帧前 CELP 的内部状态清 0,该帧的编码数据不写入最终的编码码流;然后,CELP 保持其内部状态进入步骤 (5.2);

步骤 (5.2),判断步骤 (4) 得到的语音帧的下一帧是否为语音帧:若其下一帧是语音帧,则执行步骤 (5.3);若其下一帧不是语音帧,执行步骤 (5.4);

步骤 (5.3),若步骤 (5.2) 判断的结果,所选语音帧的下一帧是音乐帧,则进行加窗折叠:取当前语音帧的最后 $M/2$ 个样点和与当前语音帧相邻的下 $M/2$ 个时域样点, M 为偶数,组成时间上连续的 M 个样点,即加窗折叠区,做如下的加窗折叠处理:

$$y_L(N+i) = x_L(N+i)*h_2(M/2-1-i)+x_L(N-1-i)*h_2(M/2+i) \quad i = -M/2, \dots, -1,$$

N 是一帧的采样点数,为偶数,

$x_L(N+i)$, $i = -M/2, \dots, -1$ 是当前帧输入的最后 $M/2$ 个时域信号,

$x_L(N-1-i)$, $i = -M/2, \dots, -1$ 是紧邻当前帧之后的 $M/2$ 个时域信号,

$h_2(M/2-1-i)$, $i = -M/2, \dots, -1$ 是窗函数 h_2 的后 $M/2$ 个样点,

$h_2(M/2+i)$, $i = -M/2, \dots, -1$ 是窗函数 h_2 的前 $M/2$ 个样点,

h_2 是满足 MDCT 完全重建条件的单调递增函数:

$$h_2(i)^2+h_2(M-1-i)^2 = 1, \quad i = 0, \dots, M-1,$$

得到 $y_L(N-M/2), \dots, y_L(N-1)$ 共 $M/2$ 个样点,与当前帧的前 $N-M/2$ 个样点 $x_L(0), \dots, x_L(N-M/2+1)$ 组成新的 N 样点序列:

$$x_L(0), \dots, x_L(N-M/2+1), y_L(N-M/2), \dots, y_L(N-1),$$

以这 N 个样点作为输入,进入步骤 (5.4);

步骤 (5.4),将来自步骤 (5.2) 原 N 点时域信号或 (5.3) 经处理的 N 点时域信号进行 k 倍降采样处理,得到 $\lfloor k*N \rfloor$ 个样点, $\lfloor k*N \rfloor$ 表示不超过 $k*N$ 最大整数, $k = f_s/f_a$ 为 CELP 的采样率 f_s 与输入音频的采样率 f_a 的比值,以这 $\lfloor k*N \rfloor$ 个样点为输入,进入步骤 (5.5);

步骤 (5.5),对步骤 (5.4) 得到的 $\lfloor k*N \rfloor$ 点序列进行 CELP 编码,得到 CELP 码流;

步骤 (6), 判断步骤 (4) 得到的音乐帧的上一帧是否为音乐帧, 若上一帧是语音帧则执行步骤 (6.1); 若上一帧是音乐帧则执行步骤 (6.2);

步骤 (6.1), 对由作为上一帧的语音帧向音乐帧切换的第一个 MDCT 变换的作用区域第 I 帧和第 (I+1) 帧, I 是当前帧的帧序号, 做如下的加窗处理:

$$\begin{cases} y_I(i) = x_I(i) * w_2(i), & i = 0, 1, \dots, N-1 \\ y_{(I+1)}(i) = x_{(I+1)}(i) * w_2(i+N), & i = 0, 1, \dots, N-1 \end{cases}$$

$x_I(i)$, $i = 0, 1, \dots, N-1$ 为第 I 帧的 N 点时域信号,

$x_{(I+1)}(i)$, $i = 0, 1, \dots, N-1$ 为第 (I+1) 帧的 N 点时域信号,

w_2 为窗函数, 由下式决定:

$$w_2(i) = \begin{cases} 0, & 0 \leq i < (N-M)/2 \\ h_2(i - (N-M)/2) & (N-M)/2 \leq i < (N+M)/2 \\ 1, & (N+M)/2 \leq i < N \\ h_1(2N-1-i) & N \leq i < 2N \end{cases}$$

其中 h_2 与步骤 (5.3) 中的 h_2 一致, h_1 为满足 MDCT 完全重建条件的窗函数:

$$h_1(j)^2 + h_1(M-1-j)^2 = 1, \quad j = 0, \dots, M-1,$$

得到的时域上连续的 2N 点加窗信号

$$y_I(0), \dots, y_I(N-1), y_{(I+1)}(0), \dots, y_{(I+1)}(N-1),$$

作为输入, 进入步骤 (6.3);

步骤 (6.2), 若上一帧是音乐帧, 则判断当前帧的下一帧是否为音乐帧; 若下一帧不是音乐帧, 执行步骤 (6.2.1); 若下一帧是音乐帧, 执行步骤 (6.2.2);

步骤 (6.2.1), 当前帧的下一帧是语音帧, 对当前帧和下一帧进行如下的加窗处理:

$$\begin{cases} y_I(i) = x_I(i) * w_1(i), & i = 0, 1, \dots, N-1 \\ y_{(I+1)}(i) = x_{(I+1)}(i) * w_1(i+N), & i = 0, 1, \dots, N-1 \end{cases}$$

I 是当前帧的帧序号,

(I+1) 是下一帧的帧序号,

$x_I(i)$, $i = 0, 1, \dots, N-1$ 为第 I 帧的 N 点时域信号,

$x_{(I+1)}(i)$, $i = 0, 1, \dots, N-1$ 为第 (I+1) 帧的 N 点时域信号,

w_1 为窗函数, 由下式决定:

$$w_1(i) = \begin{cases} h_1(i), & 0 \leq i < N \\ 1, & N \leq i < 3N/2 \\ 0, & 3N/2 \leq i < 2N \end{cases}$$

其中 h_1 与步骤 (6.1) 中的 h_1 一致, 得到的时域上连续的 2N 点加窗信号

$$y_I(0), \dots, y_I(N-1), y_{(I+1)}(0), \dots, y_{(I+1)}(N-1),$$

作为输入, 进入步骤 (6.3);

步骤 (6.2.2), 当前帧的下一帧是语音帧, 对当前帧和下一帧进行如下的加窗处理:

$$\begin{cases} y_I(i) = x_I(i) * w_0(i), & i = 0, 1, \dots, N-1 \\ y_{(I+1)}(i) = x_{(I+1)}(i) * w_0(i+N), & i = 0, 1, \dots, N-1 \end{cases}$$

I 是当前帧的帧序号,

(I+1) 是下一帧的帧序号,

$x_I(i)$, $i = 0, 1, \dots, N-1$ 为第 I 帧的 N 点时域信号,

$x_{(I+1)}(i)$, $i = 0, 1, \dots, N-1$ 为第 (I+1) 帧的 N 点时域信号,

w_0 为窗函数, 由下式决定:

$$w_0(i) = \begin{cases} h_1(i), & 0 \leq i < N \\ h_1(2N-1-i), & N \leq i < 2N \end{cases}$$

其中 h_1 与步骤 (6.1) 中的 h_1 一致, 得到的时域上连续的 2N 点加窗信号

$y_I(0), \dots, y_I(N-1), y_{(I+1)}(0), \dots, y_{(I+1)}(N-1)$,

作为输入, 进入步骤 (6.3);

步骤 (6.3), 对步骤 (6.1) 或者 (6.2.1) 或者 (6.2.2) 得到的点序列进行 AAC 编码, 得到 AAC 码流。

2. 一种语音 / 音乐双模解码器, 能够解码权利要求 1 所述的语音 / 音乐双模编码无缝切换方法生成的 CELP/AAC 混合码流, 其特征在于依次含有以下步骤:

步骤 (1), 在数字集成电路中构成一个解码器, 在该解码器至少有三个子模块: 语音 / 音乐切换模块, 对语音帧和音乐帧作无缝切换; 修正余弦变换的音乐解码子模块, 对 AAC 码流解码; 语音解码子模块, 对 CELP 码流解码;

步骤 (2), 根据码流的类别选择不同的解码子模块: 当前的码流为 AAC 时, 执行步骤 (3); 当前的码流为 CELP 时, 执行步骤 (4);

步骤 (3), 把步骤 (2) 所得到的码流输入到修正余弦变换的音乐解码子模块, 得到输出时域音频信号;

步骤 (4), 判断当前帧的上一帧是否为 CELP 码流, 如果上一帧是 AAC 码流, 执行步骤 (4.1); 如果上一帧是 CELP 码流, 执行步骤 (4.2);

步骤 (4.1), 当前帧是 CELP 码流, 当前帧的上一帧是 AAC 码流, 进行预编码处理, 即根据量化后的 MDCT 谱系数反量化获得当前帧的前 N/2 个样点, 然后截取这 N/2 个样点的最后 P 个样点降采样后作为内部状态清 0 的 CELP 编码器的一帧输入, 预编码后, 将 CELP 编码器的内部状态传送给 CELP 解码器, CELP 解码器保持这个状态进入步骤 (4.2);

步骤 (4.2), 把步骤 (2) 得到的码流输入到 CELP 解码器, 得到时域语音信号, 执行步骤 (4.3);

步骤 (4.3), 将步骤 (4.2) 得到的时域语音信号进行 k 倍升采样处理, 这里的 k 是升采样的倍数; 执行步骤 (4.4);

步骤 (4.4), 判断当前帧的下一帧是否为 AAC 码流, 如果下一帧是 AAC 码流, 执行步骤 (4.5); 如果下一帧不是 AAC 码流, 执行步骤 (4.6);

步骤 (4.5), 如果当前帧的下一帧是 AAC 码流, 将步骤 (4.3) 得到的当前帧时域样点中的最后 M/2 个样点进行镜像和加窗:

$$x'(N-M/2+i) = \begin{cases} y'(N-M/2+i) * h_2(M-1-i), & i = 0, 1, \dots, M/2-1 \\ y'(N+M/2-1-i) * h_2(M-1-i), & i = M/2, \dots, M-1 \end{cases}$$

N 是一帧的时域样点数,

M 是重叠区的样点数,

$y'(N-M/2+i)$, $i = 0, 1, \dots, M/2-1$ 表示步骤 (4.3) 得到语音信号的最后 M/2 个时域

样点,

$y'(N+M/2-1-i)$, $i = M/2, \dots, M-1$ 表示步骤 (4.3) 得到的语音信号的最后 $M/2$ 个样点的时域镜像,

h_2 是满足 MDCT 完全重建条件的单调递增函数:

$$h_2(i)^2 + h_2(M-1-i)^2 = 1, i = 0, \dots, M-1$$

得到 M 个镜像加窗后的样点 $x'(N-M/2), \dots, x'(N+M/2-1)$, 与下一帧 AAC 解码得到的有效时域信号的前 M 个时域样点 $x''(N-M/2), \dots, x''(N+M/2-1)$ 逐点相加, 得到完整的 M 点语音 / 音乐过渡信号 $y''(N-M/2), \dots, y''(N+M/2-1)$, 与步骤 (4.3) 得到的语音信号的前 $N-M/2$ 个样点 $y'(0), \dots, y'(N-M/2-1)$ 构成时域上连续的 $N+M/2$ 个样点:

$$y'(0), \dots, y'(N-M/2-1), x'(N-M/2), \dots, x'(N+M/2-1),$$

得到输出时域音频信号。

3. 根据权利要求 1 所述的一种语音 / 音乐双模编码无缝切换方法, 其特征在于:

$$h_1(i) = \sin(2\pi i/4N + \pi/4N), i = 0, 1, \dots, N-1.$$

4. 根据权利要求 1 所述的一种语音 / 音乐双模编码无缝切换方法, 其特征在于:

$$h_2(i) = \sin(2\pi i/4M + \pi/4M), i = 0, 1, \dots, M-1.$$

5. 根据权利要求 2 所述的一种语音 / 音乐双模解码器, 其特征在于:

$$h_1(i) = \sin(2\pi i/4N + \pi/4N), i = 0, 1, \dots, N-1.$$

6. 根据权利要求 2 所述的一种语音 / 音乐双模解码器, 其特征在于:

$$h_2(i) = \sin(2\pi i/4M + \pi/4M), i = 0, 1, \dots, M-1.$$

一种语音 / 音乐双模编解码无缝切换方法

技术领域

[0001] 本发明涉及一种可用于移动通信中低码率语音 / 音乐双模编解码器的设计。特别的,当语音模式采用码激励线性预测编码 CELP,且音乐模式采用基于修正余弦变换 MDCT 的变换编码时,两种模式的无缝切换和降采样处理。

背景技术

[0002] 语音信号和一般的音乐信号在时频统计特性上有很大的不同。时域语音信号表现为准周期特性,其频谱较平坦且带宽在 7KHz 以下;一般的音乐信号在时域和频域都有很大的动态特性,其频谱带宽主要受限于采样率,可以达到 16KHz 以上。因此语音编码通常采用线性预测结合长时预测(基音)的编码方法,如码激励线性预测编码 CELP(Code Excited Linear Prediction);而音乐编码器则通常采用时频变换结合心理声学的编码方法,如 MPEG-2/4AAC(Advanced Audio Coding)。但是当语音编码器的输入是一般的音乐或者音乐编码器的输入是语音时,编码音质将明显低于编码器与输入音频类型匹配时的情况。

[0003] 随着无线通信, IPTV 等信息技术的发展,能在低码率下同时高质量处理语音和音乐的编解码器成为一个重要的应用需求。一个可能的解决方案就是将两种编码方法结合起来,根据音频信号的类型动态切换成语音或音乐编码模式的语音 / 音乐双模编码器。例如在第三代移动通信 3GPP(the 3rd Generation Partnership Project)中使用的自适应多速率长宽带编码器 AMR-WB+(Adaptive Multi-Rate Wide Band Plus),它的核心就是语音编码器 AMR-WB 和音乐编码器 TCX(Transform Coded Excitation)。

[0004] 语音 / 音乐双模编码器的一个核心问题就是如何实现语音 / 音乐模式的无缝切换,即切换本身不引起附加的编解码失真,同时避免切换带来的码率增加。无论是音乐编码器还是语音编码器,他们的基本处理单元都是帧,即某一有限长度且时间上连续信号片断。单纯基于时频变换的音乐编码器帧间的平滑过渡采用重叠相加技术,而单纯的基于线性预测的语音编码器帧间的连续性依赖于线性预测滤波器的记忆性(零输入响应)。对于语音 / 音乐双模编解码器的模式切换,特别的,当音乐模式采用基于 MDCT 的编码器时,文献中存在如下几种切换方法:

[0005] (1) 切换帧共 20ms 中的前 5ms 延续切换前的编码模式,后 5ms 采用切换后的编码模式,中间 10ms 在解码端通过线性预测合成滤波器外插得到,外插的基础是先前的解码输出(见欧洲专利 EP932141)。该方法的问题是外插的 10ms 全部来自基于过去信息的预测,如果切换帧具有很大动态变化,预测将和实际信号有很大出入。因此这种切换方法本身就会带来附加的解码失真。

[0006] (2) 首先对输入的时域音频信号进行线性预测滤波得到激励信号,切换方法工作在滤波输出的激励信号上。对于音乐到语音的切换,不对上一帧离散余弦变换 DCT 延伸到当前语音帧的部分做重叠相加处理,而对语音帧直接采用 CELP 编解码;对于语音到音乐切换,音乐帧的 DCT 变换采用的起始重叠长度为 0 的窗函数(见美国专利 US6658383)。该方法中没有对两种模式的切换做过渡处理,因此本身无法实现无缝切换。最终解码输出的时

域音频信号的连续性是通过激励信号线性预测合成滤波器的平滑作用得到。另外,由于在音乐到语音的切换时采取直接舍弃重叠信号的方法,也意味着这种切换有额外的码率增加。

[0007] (3) 首先对输入的时域音频信号进行线性预测滤波得到激励信号,切换方法工作在滤波输出的激励信号上。该方法采用一种特殊的窗型的 MDCT 变换处理语音到音乐模式的切换,避免了 MDCT 重叠相加部分在切换帧的混叠(见美国专利申请 20060173605)。与(2)类似,最终输出时域音频信号的连续性是通过线性预测合成滤波器的平滑作用得到。

[0008] 以上三种方法都没有有效解决两种模式无缝切换的问题。

发明内容

[0009] 本发明需要解决的技术问题是:在不降低编码效率的前提下,如何实现语音/音乐双模编解码中模式的无缝切换。该切换方法直接处理未经线性预测滤波的时域音频信号,并且语音模式采用 CELP,音乐模式采用基于 MDCT 变换编码。

[0010] 本发明的特征在于,在编码端依次包含以下步骤:

[0011] 步骤(1),在数字集成电路中构成一个编码器,在该编码器至少有四个子模块:语音/音乐分类子模块,区分语音帧和音乐帧;语音/音乐切换模块,对语音帧和音乐帧作无缝切换;修正余弦变换的音乐编码子模块,采用先进音频编码 AAC 对音乐编码;语音编码子模块,采用码激励线性预测 CELP 对语音编码;

[0012] 步骤(2),该语音/音乐分类子模块把输入的音频信号划分为语音帧和音乐帧;

[0013] 步骤(3),把步骤(2)所得到的结果送入语音/音乐切换模块,判断是否为语音帧;

[0014] 步骤(4),在步骤(3)中,若判断为语音帧,则执行步骤(5),若判断为音乐帧,则执行步骤(6);

[0015] 步骤(5),判断步骤(4)得到的语音帧的上一帧是否为语音帧:若其上一帧是音乐帧,则执行步骤(5.1);若其上一帧是语音帧,则执行步骤(5.2);

[0016] 步骤(5.1),若步骤(5)判断的结果,所选语音帧的上一帧是音乐帧,则进行预编码:首先从上一帧经过编码量化后的整数修正余弦变换 MDCT 频谱数据中恢复第 L 帧的前 $N/2$ 个样点, L 是当前帧的帧序号, N 是一帧的采样点数,为偶数,各帧相同;再从这 $N/2$ 个点中截取最后的 P 个样点, P 为一个 CELP 编码帧的样点数,降采样后作为语音编码器 CELP 的输入;编码该帧前 CELP 的内部状态清 0,该帧的编码数据不写入最终的编码码流;然后, CELP 保持其内部状态进入步骤(5.2);

[0017] 步骤(5.2),判断步骤(4)得到的语音帧的下一帧是否为语音帧:若其下一帧是语音帧,则执行步骤(5.3);若其下一帧不是语音帧,执行步骤(5.4);

[0018] 步骤(5.3),若步骤(5.2)判断的结果,所选语音帧的下一帧是音乐帧,则进行加窗折叠:取当前语音帧的最后 $M/2$ 个样点和与当前语音帧相邻的下 $M/2$ 个时域样点, M 为偶数,组成时间上连续的 M 个样点,即加窗折叠区,做如下的加窗折叠处理:

[0019] $y_L(N+i) = x_L(N+i)*h_2(M/2-1-i)+x_L(N-1-i)*h_2(M/2+i) \quad i = -M/2, \dots, -1,$

[0020] N 是一帧的采样点数,为偶数,

[0021] $x_L(N+i), i = -M/2, \dots, -1$ 是当前帧输入的最后 $M/2$ 个时域信号,

[0022] $x_L(N-1-i)$, $i = -M/2, \dots, -1$ 是紧邻当前帧之后的 $M/2$ 个时域信号,

[0023] $h_2(M/2-1-i)$, $i = -M/2, \dots, -1$ 是窗函数 h_2 的后 $M/2$ 个样点,

[0024] $h_2(M/2+i)$, $i = -M/2, \dots, -1$ 是窗函数 h_2 的前 $M/2$ 个样点,

[0025] h_2 是满足 MDCT 完全重建条件的单调递增函数:

[0026] $h_2(i)^2 + h_2(M-1-i)^2 = 1$, $i = 0, \dots, M-1$,

[0027] 得到 $y_L(N-M/2), \dots, y_L(N-1)$ 共 $M/2$ 个样点, 与当前帧的前 $N-M/2$ 个样点 $x_L(0), \dots, x_L(N-M/2+1)$ 组成新的 N 样点序列:

[0028] $x_L(0), \dots, x_L(N-M/2+1), y_L(N-M/2), \dots, y_L(N-1)$,

[0029] 以这 N 个样点作为输入, 进入步骤 (5.4);

[0030] 步骤 (5.4), 将来自步骤 (5.2) 原 N 点时域信号或 (5.3) 经处理的 N 点时域信号进行 k 倍降采样处理, 得到 $\lfloor k*N \rfloor$ 个样点, $\lfloor k*N \rfloor$ 表示不超过 $k*N$ 最大整数, $k = f_s/f_a$ 为 CELP 的采样率 f_s 与输入音频的采样率 f_a 的比值, 以这 $\lfloor k*N \rfloor$ 个样点为输入, 进入步骤 (5.5);

[0031] 步骤 (5.5), 对步骤 (5.4) 得到的 $\lfloor k*N \rfloor$ 点序列进行 CELP 编码, 得到 CELP 码流;

[0032] 步骤 (6), 判断步骤 (4) 得到的音乐帧的上一帧是否为音乐帧, 若上一帧是语音帧则执行步骤 (6.1); 若上一帧是音乐帧则执行步骤 (6.2);

[0033] 步骤 (6.1), 对由作为上一帧的语音帧向音乐帧切换的第一个 MDCT 变换的作用区域第 I 帧和第 $(I+1)$ 帧, I 是当前帧的帧序号, 做如下的加窗处理:

$$[0034] \begin{cases} y_I(i) = x_I(i) * w_2(i), & i = 0, 1, \dots, N-1 \\ y_{(I+1)}(i) = x_{(I+1)}(i) * w_2(i+N), & i = 0, 1, \dots, N-1 \end{cases}$$

[0035] $x_I(i)$, $i = 0, 1, \dots, N-1$ 为第 I 帧的 N 点时域信号,

[0036] $x_{(I+1)}(i)$, $i = 0, 1, \dots, N-1$ 为第 $(I+1)$ 帧的 N 点时域信号,

[0037] w_2 为窗函数, 由下式决定:

$$[0038] w_2(i) = \begin{cases} 0, & 0 \leq i < (N-M)/2 \\ h_2(i - (N-M)/2), & (N-M)/2 \leq i < (N+M)/2 \\ 1, & (N+M)/2 \leq i < N \\ h_1(2N-1-i), & N \leq i < 2N \end{cases}$$

[0039] 其中 h_2 与步骤 (5.3) 中的 h_2 一致, h_1 为满足 MDCT 完全重建条件的窗函数:

[0040] $h_1(j)^2 + h_1(M-1-j)^2 = 1$, $j = 0, \dots, M-1$,

[0041] 得到的时域上连续的 $2N$ 点加窗信号

[0042] $y_I(0), \dots, y_I(N-1), y_{(I+1)}(0), \dots, y_{(I+1)}(N-1)$,

[0043] 作为输入, 进入步骤 (6.3);

[0044] 步骤 (6.2), 若上一帧是音乐帧, 则判断当前帧的下一帧是否为音乐帧; 若下一帧不是音乐帧, 执行步骤 (6.2.1); 若下一帧是音乐帧, 执行步骤 (6.2.2);

[0045] 步骤 (6.2.1), 当前帧的下一帧是语音帧, 对当前帧和下一帧进行如下的加窗处理:

$$[0046] \begin{cases} y_I(i) = x_I(i) * w_1(i), & i = 0, 1, \dots, N-1 \\ y_{(I+1)}(i) = x_{(I+1)}(i) * w_1(i+N), & i = 0, 1, \dots, N-1 \end{cases}$$

[0047] I 是当前帧的帧序号，

[0048] (I+1) 是下一帧的帧序号，

[0049] $x_I(i)$, $i = 0, 1, \dots, N-1$ 为第 I 帧的 N 点时域信号，

[0050] $x_{(I+1)}(i)$, $i = 0, 1, \dots, N-1$ 为第 (I+1) 帧的 N 点时域信号，

[0051] w_1 为窗函数，由下式决定：

$$[0052] \quad w_1(i) = \begin{cases} h_1(i), & 0 \leq i < N \\ 1, & N \leq i < 3N/2 \\ 0, & 3N/2 \leq i < 2N \end{cases}$$

[0053] 其中 h_1 与步骤 (6.1) 中的 h_1 一致，得到的时域上连续的 2N 点加窗信号

[0054] $y_I(0), \dots, y_I(N-1), y_{(I+1)}(0), \dots, y_{(I+1)}(N-1)$,

[0055] 作为输入，进入步骤 (6.3)；

[0056] 步骤 (6.2.2)，当前帧的下一帧是语音帧，对当前帧和下一帧进行如下的加窗处理：

$$[0057] \quad \begin{cases} y_I(i) = x_I(i) * w_0(i), & i = 0, 1, \dots, N-1 \\ y_{(I+1)}(i) = x_{(I+1)}(i) * w_0(i+N), & i = 0, 1, \dots, N-1 \end{cases}$$

[0058] I 是当前帧的帧序号，

[0059] (I+1) 是下一帧的帧序号，

[0060] $x_I(i)$, $i = 0, 1, \dots, N-1$ 为第 I 帧的 N 点时域信号，

[0061] $x_{(I+1)}(i)$, $i = 0, 1, \dots, N-1$ 为第 (I+1) 帧的 N 点时域信号，

[0062] w_0 为窗函数，由下式决定：

$$[0063] \quad w_0(i) = \begin{cases} h_1(i), & 0 \leq i < N \\ h_1(2N-1-i), & N \leq i < 2N \end{cases}$$

[0064] 其中 h_1 与步骤 (6.1) 中的 h_1 一致，得到的时域上连续的 2N 点加窗信号

[0065] $y_I(0), \dots, y_I(N-1), y_{(I+1)}(0), \dots, y_{(I+1)}(N-1)$,

[0066] 作为输入，进入步骤 (6.3)；

[0067] 步骤 (6.3)，对步骤 (6.1) 或者 (6.2.1) 或者 (6.2.2) 得到的点序列进行 AAC 编码，得到 AAC 码流。

[0068] 在解码端依次含有以下步骤：

[0069] 步骤 (1)，在数字集成电路中构成一个解码器，在该解码器至少有三个子模块：语音 / 音乐切换模块，对语音帧和音乐帧作无缝切换；修正余弦变换的音乐解码子模块，对 AAC 码流解码；语音解码子模块，对 CELP 码流解码；

[0070] 步骤 (2)，根据码流的类别选择不同的解码子模块：当前的码流为 AAC 时，执行步骤 (3)；当前的码流为 CELP 时，执行步骤 (4)；

[0071] 步骤 (3)，把步骤 (2) 所得到的码流输入到修正余弦变换的音乐解码子模块，得到输出时域音频信号；

[0072] 步骤 (4)，判断当前帧的上一帧是否为 CELP 码流，如果上一帧是 AAC 码流，执行步骤 (4.1)；如果上一帧是 CELP 码流，执行步骤 (4.2)；

[0073] 步骤 (4.1)，当前帧是 CELP 码流，当前帧的上一帧是 AAC 码流，进行预编码处理，即

根据量化后的 MDCT 谱系数反量化获得当前帧的前 $N/2$ 个样点,然后截取这 $N/2$ 个样点的最后 P 个样点降采样后作为内部状态清 0 的 CELP 编码器的一帧输入,预编码后,将 CELP 编码器的内部状态传送给 CELP 解码器,CELP 解码器保持这个状态进入步骤 (4.2);

[0074] 步骤 (4.2),把步骤 (2) 得到的码流输入到 CELP 解码器,得到时域语音信号,执行步骤 (4.3);

[0075] 步骤 (4.3),将步骤 (4.2) 得到的时域语音信号进行 k 倍升采样处理,这里的 k 是升采样的倍数;执行步骤 (4.4)

[0076] 步骤 (4.4),判断当前帧的下一帧是否为 AAC 码流,如果下一帧是 AAC 码流,执行步骤 (4.5);如果下一帧不是 AAC 码流,执行步骤 (4.6);

[0077] 步骤 (4.5),如果当前帧的下一帧的是 AAC 码流,将步骤 (4.3) 得到的当前帧时域样点中的最后 $M/2$ 个样点进行镜像和加窗:

$$[0078] \quad x'(N-M/2+i) = \begin{cases} y'(N-M/2+i) * h_2(M-1-i), & i = 0, 1, \dots, M/2-1 \\ y'(N+M/2-1-i) * h_2(M-1-i), & i = M/2, \dots, M-1 \end{cases}$$

[0079] N 是一帧的时域样点数,

[0080] M 是重叠区的样点数,

[0081] $y'(N-M/2+i)$, $i = 0, 1, \dots, M/2-1$ 表示步骤 (4.3) 得到语音信号的最后 $M/2$ 个时域样点,

[0082] $y'(N+M/2-1-i)$, $i = M/2, \dots, M-1$ 表示步骤 (4.3) 得到的语音信号的最后 $M/2$ 个样点的时域镜像,

[0083] h_2 是满足 MDCT 完全重建条件的单调递增函数:

$$[0084] \quad h_2(i)^2 + h_2(M-1-i)^2 = 1, \quad i = 0, \dots, M-1$$

[0085] 得到 M 个镜像加窗后的样点 $x'(N-M/2), \dots, x'(N+M/2-1)$,与下一帧 AAC 解码得到的有效时域信号的前 M 个时域样点 $x''(N-M/2), \dots, x''(N+M/2-1)$ 逐点相加,得到完整的 M 点语音/音乐过渡信号 $y''(N-M/2), \dots, y''(N+M/2-1)$,与步骤 (4.3) 得到的语音信号的前 $N-M/2$ 个样点 $y'(0), \dots, y'(N-M/2-1)$ 构成时域上连续的 $N+M/2$ 个样点:

$$[0086] \quad y'(0), \dots, y'(N-M/2-1), x'(N-M/2), \dots, x'(N+M/2-1),$$

[0087] 得到输出时域音频信号。

[0088] 本发明的特征还在于上述语音/音乐双模编码方法和上述语音/音乐解码器使用窗函数

$$[0089] \quad h_1(i) = \sin(2\pi i/4N + \pi/4N), \quad i = 0, 1, \dots, N-1,$$

[0090] 以及窗函数

$$[0091] \quad h_2(i) = \sin(2\pi i/4M + \pi/4M), \quad i = 0, 1, \dots, M-1。$$

[0092] 本发明可以在语音→音乐和音乐→语音两种切换模式下实现平滑的过渡。并且与专利 EP93214, US-6658383, 及专利申请 US20060173675 相比,没有切换引起的额外编码样点数的增加或重复编码,实现了高效的切换。

附图说明

[0093] 图 1 语音/音乐双模编码器框图

[0094] 图 2 切换帧的窗型和加窗折叠处理

- [0095] 图 3 窗型示意图
[0096] 图 4 具体实施例编码流程图
[0097] 图 5 具体实施例解码流程图

具体实施例

[0098] 本发明的技术解决方案是：当从语音切换到音乐时，对切换前最后一个语音帧的尾部进行加窗和折叠处理，此时连续性由 MDCT 变换的重叠相加特性保证；当从音乐切换到语音时，对切换前的最后一个音乐帧采用新的 MDCT 窗型使其与相继的语音帧间无时域重叠，此时的连续性由 CELP 非线性预测合成滤波器的记忆性保证。另一方面，为了使语音编码和音乐编码的采样率相匹配，对语音帧进行特定的降采样处理。下面使结合附图 1, 2, 3 给予详细说明。

[0099] 图 1 表示语音 / 音频双模编码器的总体结构，分为核心双模编码器 10，立体声编码 13，带宽扩展 12，和比特成型 11 四个模块。输入的时域音频信号的立体声信息（如果输入为立体声）和低频信息分别由 13 和 12 进行编码；输入信号的低频单声道部分由 10 进行编码，最后由 11 将 13, 12, 10 输出的比特流打包，生成最后的编码码流。

[0100] 带宽扩展模块 12 的输入是语音 / 音乐信号中 1/4 采样率到 1/2 采样率的高频成分。首先通过时频分析得到高频成分的频谱包络，并以参数表示。然后根据音质和比特率要求量化频谱包络参数并进行熵编码（如 Huffman 编码）。最后数据输出到比特流成型模块 11。

[0101] 立体声模块 13 提取输入语音 / 音乐信号中的立体声分量的参数信息，包括声道间强度差 IID (Interchannel Intensity Difference)、声道间相位差 IPD (Interchannel Phase Difference)、声道间相关度 IC (Interchannel Coherence)。该模块进一步将 IID, IPD, 以及 IC 参数根据音质和比特率要求进行量化及熵编码，最后输出到比特流成型模块 11。

[0102] 比特流成型模块 11 的输入是 10, 12, 和 13 的输出。根据定义好的码流格式，将这三部分数据以一定的格式封装起来，加上信号的采样率，编码帧长等必要的辅助信息，形成完整的编码码流。

[0103] 图 1 的 10 由四个子模块组成：语音 / 音乐鉴别 101, 语音 / 音乐切换 102, MDCT 变换编码 103, 和 CELP 编码 104。首先 101 给出当前帧类型的判断，然后 102 根据当前帧和上一帧类型的是否不一致启动切换处理，103 和 104 分别处理对应类型的音频信号帧。

[0104] 语音 / 音乐鉴别模块 101 首先计算反映信号类型的统计量，如低频单声道分量平均能量、能量的变化率、以及谱线分布等。然后根据预先设定的统计量的权重对各个统计量加权，得出一个表征当前帧为语音的可能性的指数。当这个指数在高于某一域值时，将当前帧标识为语音帧；否则标识为音乐帧。

[0105] 如果当前帧被标识为音乐，调用 MDCT 变换编码器 103 进行编码。首先对输入信号进行 MDCT 变换和心理声学分析，分别得到 MDCT 频谱和按心理声学划分的 MDCT 各个子带在给定比特率下的允许失真（量化误差）。然后在每个子带允许的失真和总的比特率这两个约束下，对 MDCT 频谱进行量化。为进一步降低码率，量化数据进行 Huffman 编码以去除冗余。音频编码数据输出到比特流成型模块 11。

[0106] 如果当前帧被标识为语音,调用 CELP 编码器 104 进行编码。首先对输入信号进行线性预测 (LP) 分析,得到线性预测滤波器的系数。然后对输入信号进行 LP 滤波,得到预测残差信号或激励信号。将当前的激励信号与之前帧的激励信号进行相关度分析,得到基音周期 (Pitch Lag) 和基音增益 (Pitch Gain),并从当前的残差信号中减去一个基音周期前进行基音增益调整的残差信号,得到新的残差信号。这个过程也被称为长时预测 (Long Term Prediction)。将这个新的残差信号作为目标,在给定的码本中寻找与之最匹配的码本,并求出相应的码本增益。CELP 将上面过程中得到的线性预测滤波器系数、基音周期、基音增益、码本标号、及码本增益进行量化,输出到比特成型模块 11。

[0107] 经过信道 14,比特流被传送到语音 / 音乐双模解码器。

[0108] 解码是编码的逆过程。顺次的,比特流解析模块 15 从比特流中分离出核心码流、模式控制参数、高频码流、以及立体声码流。核心码流传送到核心解码器 16,并在模式控制参数的控制下,通过切换器 161 选择 IMDCT 变换解码器 162 或者 CELP 解码器 163 进行解码。核心解码器输出的是单声道低频信号,经过高频重建模块 17 输出单声道全频带信号,再经过立体声重建模块 18 输出立体声语音 / 音乐信号。

[0109] 核心编码器 10 和核心解码器 16 的切换完全对应。当连续处于语音模式时,核心编解码器都不需要加窗操作,CELP 编解码器能自动实现帧间的平滑过渡;当连续处于音乐模式的时候,编解码器都采用图 2(a) 所示的窗型,结合 MDCT/IMDCT 的时域混叠特性,实现帧间的平滑过渡。

[0110] 图 2 表示音乐向语音和语音向音乐两种互补切换对输入音频信号时序关系及相应的处理。21 是音乐到语音切换的最后一个 MDCT 变换;22 是语音片断范围;23 是语音到音乐切换时 CELP 和 MDCT 的重叠区域;24 是语音到音乐切换后第一个 MDCT 变换;25 表示语音片断末尾经加窗和重叠得到样点;27 表示预编码部分的样点;26 (包含 25) 表示降采样处理后需要进行 CELP 编码的一个连续语音片断。

[0111] 将输入音频信号表示为 $x(1*N+i)$,其中常数 N 为每帧的样点数, $l = 0, 1, 2, \dots$ 为帧序号, $i = 0, 1, \dots, N-1$ 是帧内相对帧起始的样点序号。图 2 中的音乐模式结束于 $l = L$ 帧的样点 $x(L*N+N/2-1)$;语音模式起始于相邻的下一个样点 $x(L*N+N/2)$ 。语音模式持续到第 $l = L+K$ 帧的样点 $x((L+K)*N+N/2+M/2-1)$,其中 M 为重叠区 23 的样点数,音乐模式起始于下一个样点 $x((L+K)*N+N/2-M/2)$ 。这里我们要求 N, M 均为偶数。 L 帧是音乐 \rightarrow 语音的切换帧; $(L+K)$ 是语音 \rightarrow 音乐的切换帧。

[0112] $(L-1)$ 和 L 帧是切换前最后一个 MDCT 作用的区域。由于后续的 CELP 变换不提供消除 MDCT 变换混叠效应的信号,第 $(L-1)$ 和 L 帧的 MDCT 变换采用如下的末端无重叠的图 3(b) 的过渡窗:

$$[0113] \quad w_1(i) = \begin{cases} h_1(i), & 0 \leq i < N \\ 1, & N \leq i < 3N/2 \\ 0, & 3N/2 \leq i < 2N \end{cases} \quad (1)$$

[0114] 其中 $h_1(i)$ 为满足 MDCT 变换完全重建条件

$$[0115] \quad h_1(i)^2 + h_1(N-1-i)^2 = 1, 0 \leq i < N \quad (2)$$

[0116] 的单调递增函数,一个典型的例子就是正弦窗 $h_1(i) = \sin(2\pi i/4N + \pi/4N)$ 。于是输入 MDCT 变换的加窗音频信号 $y_l(i)$:

$$[0117] \quad y_L(i) = w_1(i) * x((L-1)N+i), 0 \leq i < 2N \quad (3)$$

[0118] 再经过 MDCT 变换得到相应的 N 个 MDCT 谱线 $Y_L(k)$

$$[0119] \quad Y_L(k) = \frac{1}{N} \sum_{i=0}^{2N} y_L(i) * \cos\left(\frac{\pi}{2N}(2i+1+N)(2k+1)\right), 0 \leq k < N \quad (4)$$

[0120] 第 $(L+K)$ 和 $(L+K+1)$ 是语音向音乐切换的第一个 MDCT 变换作用的区域。为了利用 MDCT 变换本身的重叠相加去混叠特性,这里 MDCT 和 CELP 有长度为 M 的重叠,既图 3(c) 过渡窗:

$$[0121] \quad w_2(i) = \begin{cases} 0, & 0 \leq i < (N-M)/2 \\ h_2(i-(N-M)/2) & (N-M)/2 \leq i < (N+M)/2 \\ 1, & (N+M)/2 \leq i < N \\ h_1(2N-1-i) & N \leq i < 2N \end{cases} \quad (5)$$

[0122] 其中 $h_2(i)$ 为满足 MDCT 变换完全重建条件

$$[0123] \quad h_2(i)^2 + h_2(M-1-i)^2 = 1, 0 \leq i < M \quad (6)$$

[0124] 的单调递增函数,一个典型的例子就是正弦窗 $h_2(i) = \sin(2\pi i/4M + \pi/4M)$ 。与音乐到语音的切换不同的地方在于,对 23 的 M 个时域样点 $x(S), \dots, x(S+M-1)$, 作如下的加窗和折叠处理,其中 $S = (L+K)*N + (N-M)/2$ 为重叠区 23 的第一个样点的序号,得到 $y_w(S+i)$:

$$[0125] \quad y_w(S+i) = x(S+i) * h_2(M-1-i) + x(S+M-i) * h_2(i) \quad 0 \leq i < M/2 \quad (7)$$

[0126] 而 23 前的 $K*N - M/2$ 个语音样点与 23 经加窗和折叠后的 $M/2$ 个 $y_w(S+i)$ 顺次连接,构成一个 $K*N$ 点语音片断。

[0127] 由于语音和音乐有不同的带宽,他们的要求的采样率通常不一致。例如 3GPP 的 enhAacPlus 编码器中,核心 AAC 编码器的输入采样率为 16KHz,而语音编码器 AMRWB 的核心采样率为 12.8KHz。由于高采样率意味着更多的样点,进而更高的码率,考虑语音信号的窄带特性,需要对语音分进行降采样处理。降采样处理要同时考虑 CELP 帧的尺寸,保证一个语音片断降采样后的样点数是 CELP 帧的整数倍。例如取 $N = 1024$,经末端加窗和折叠的 k 帧语音共 $1024k$ 个时域样点,其原始采样率为 16KHz。而 CELP 帧的尺寸若为 256 个样点,那么 $3/4$ 降采样后有 $768k$ 个时域样点或 $3k$ 个 CELP 帧。

[0128] 下面说明第 L 帧和第 $(L+K)$ 帧实现平滑切换的预编码技术和加窗折叠技术。

[0129] 由于第 L 帧采用矩形窗,根据 MDCT 的数学性质,经过反变换 IMDCT 后可直接得到第 L 帧的前 $N/2$ 个样点的时域信号。为了使这 $N/2$ 个样点和相继的语音样点平滑过渡,采用预编码技术。所谓预编码技术就是首先从编码端量化后的整数 MDCT 频谱数据恢复出 L 帧的前 $N/2$ 个样点,这 $N/2$ 个样点与解码端输出的是严格一致的。从这 $N/2$ 个样点的后面截取一个 CELP 帧长度的样点序列 27 也作为 CELP 编码器输入,但是这部分编码数据不写入码流,其目的在于建立线性预测滤波器和基音搜索器的状态。解码端,对 IMDCT 变换得到的预编码时域信号进行同样的 CELP 编码,这样当 CELP 的当前帧是 26 的第一个语音帧时,解码器可以获得和编码器一致的 27 的激励信号。这个激励信号结合码流中传送的基音周期和增益信息,再通过线性预测合成滤波器,音乐到语音实现了无缝切换。预编码技术避免了把 26 的第一帧作为 CELP 输入起始帧,由于其算法延时和内部初始状态的不稳定造成的第一帧音频信号严重失真问题。预编码使 CELP 在进行实际编码前建立稳定的内部状态,保证

切换帧间的平滑过渡。

[0130] 对于语音到音乐的切换,经 CELP 解码和与编码端降采样对应的上采样得到与 25 对应的 $M/2$ 个样点 $y'_w(S+i)$, $0 \leq i < M/2$, 与 (7) 中的 $y_w(S+i)$ 一一对应, 两者的差异就是编解码误差。将这 $M/2$ 个样点进行镜像并加窗, 得到

$$\begin{aligned}
 [0131] \quad x'_w(S+i) &= \begin{cases} y'_w(S+i) * h_2(M-1-i), & 0 \leq i < M/2 \\ y'_w(S+M-1-i) * h_2(M-1-i), & M/2 \leq i < M \end{cases} \\
 [0132] \quad &\approx h_2(M-1-i) * (h_2(M-1-i) * x(S+i) + h_2(i) * x(S+M-1-i)), \quad 0 \leq i < M
 \end{aligned} \tag{8}$$

[0133] 根据 MDCT 的数学性质, 对应 23 的 IMDCT 输出并加窗信号:

$$[0134] \quad x'_v(S+i) \approx h_2(i) * (h_2(i) * x(S+i) - h_2(M-1-i) * x(S+M-1-i)), \quad 0 \leq i < M \tag{9}$$

[0135] 将 (8), (9) 相加, 得到去回混叠的 23 解码输出

$$\begin{aligned}
 [0136] \quad x'(S+i) &= x'_w(S+i) + x'_v(S+i) \\
 [0137] \quad &\approx (h_2(i)^2 + h_2(M-1-i)^2) * x(S+i). \tag{10} \\
 [0138] \quad &= x(S+i), \quad 0 \leq i < M
 \end{aligned}$$

[0139] 如果没有 CELP 和 MDCT 的编解码失真, $x'(S+i) = x(S+i)$, 即切换帧完全重建。同时由于 M 点过渡区加窗和重叠相加的存在, 两个分段连续的区域, 语音解码帧和音乐解码帧被粘合成一个连续的区域。

[0140] 下面以 3GPP AMR-WB 和 MPEG-2AAC 为例说明语音 / 音乐双模编解码无缝切换方法的实现, 但是该方法不限于 AMR-WB 和 AAC 混合的情况。这里, AMR-WB 的内部采样率是 12KHz, 而不是标准的 12.8KHz, 并且进行了延时补偿; AAC 的采样率是 16KHz, 并且内部只采用长正弦窗。

[0141] 输入的音频信号为 16KHz 采样的语音 / 音乐混合信号。信号被划分为连续的帧, 每帧 1024 样点, 或 64ms。401 判断当前帧是否为语音, 如果是, 则对当前帧应用 AMR-WB 编码 412, 否则对当前帧采用 AAC 编码 413。在进行 412 或 413 的编码前, 需要进行模式切换处理, 既 402 ~ 411。由于实际的语音或音乐片断的长度都远大于 100ms, 因此我们假定没有“语音 - 音乐 - 语音 - 音乐”这种两个模式间隔出现的频繁的切换情况。考虑到语音的带宽和样点数的约束, 对语音帧进行 16KHz \rightarrow 12KHz 的降采样处理, 对音乐帧保持 16KHz 采样率。AMR-WB 每帧长度为 256 个样点, AAC 每帧长度为 1024 个样点。因此一个语音帧处理后有 768 个样点, 既 3 个 AMR-WB 帧; 一个音乐帧对应一个 AAC 帧。

[0142] 如果当前帧被 401 标识为语音, 402 给出的上一帧是否也为语音帧。如果是, 这意味着当前帧没有音乐 \rightarrow 语音切换发生, 并且如果 405 给出当前帧不是语音尾帧, 既连续的语音帧序列中的最后一帧, 直接进行 16KHz \rightarrow 12KHz 的降采样 407 处理; 如果 405 给出当前帧是语音尾帧时, 这说明当前帧是语音 \rightarrow 音乐的切换帧, 对当前帧的最后 128 个样点加余弦窗并进行偶折叠 406, 然后进行降采样 407 处理。如果 402 给出上一帧是音乐时, 这说明当前帧是语音 \rightarrow 音乐的切换帧, 怎进行预编码 404 处理。所谓预编码 404 就是将上一个 AAC 帧进行解码, 并进行 16KHz \rightarrow 12KHz 的降采样处理, 得到当前帧前的 256 个样点, 既一个 AMR-WB 帧。然后对这个 AMR-WB 帧进行编码, 使 AMR-WB 内部状态达到稳定。

[0143] 如果当前帧被 401 标识为音乐帧, 403 给出上一帧是否为音乐。如果是, 这说明当前帧不是语音 \rightarrow 音乐的切换帧, 进一步, 如果 408 给出当前帧是音频尾帧, 既连续的音乐帧

序列中的最后一帧,则当前帧是音乐→语音的切换帧,对当前帧加图 3(b) 的窗型 410,否则,加图 3(a) 中的窗型 411。如果 403 给出上一帧不是音乐帧,那么当前帧是语音→音乐切换帧,对当前帧应用图 3(c) 的窗型 409。

[0144] 解码是编码的逆过程,根据当前帧的类型,轮流由 AMR-WB 解码器 505 和 AAC 解码器 504 输出语音 / 音乐混合信号。解码端切换处理模块 502,503,506,507,508 保真切换的平滑过渡。

[0145] 如果 501 给出当前帧是语音帧,则 502 进一步给出上一帧的是否为语音帧。如果上一帧也是语音,则直接进行 AMR-WB 解码 505;否则当前帧是音乐→语音的切换帧,需要进行预编码 503 处理。所谓预编码 503 和编码端 404 完全一致,既对上一个 AAC 帧的解码信号进行 16KHz → 12KHz 的降采样,然后对当前帧之前的一个 AMR-WB 帧 (256 个样点) 进行编码,编码器的内部状态被 AMR-WB 解码器继承,使它在解当前帧 (连续的语音帧序列中的第一帧) 时具有稳定的内部状态。AMR-WB 解码器 505 的输出直接进行 12KHz → 16KHz 的升采样处理 506。如果 507 给出当前帧是语音尾帧,那么对当前的语音帧进行镜像加窗处理 508。所谓镜像加窗 508 就是将当前解码输出的 1024 个样点的最后 64 个样点变换成 128 个样点,这 128 样点的前 64 个就是原来的 64 个样点,而后 64 个和前 64 个是镜像对称的关系,然后在对这 128 个样点家余弦窗。508 是编码端加窗折叠 406 的逆过程。这 128 个点与下一帧 AAC 的解码重叠相加得到最后的时域输出。

[0146] 如果当前帧是 AAC 的帧,则直接进行 AAC 解码,得到时域音频信号。

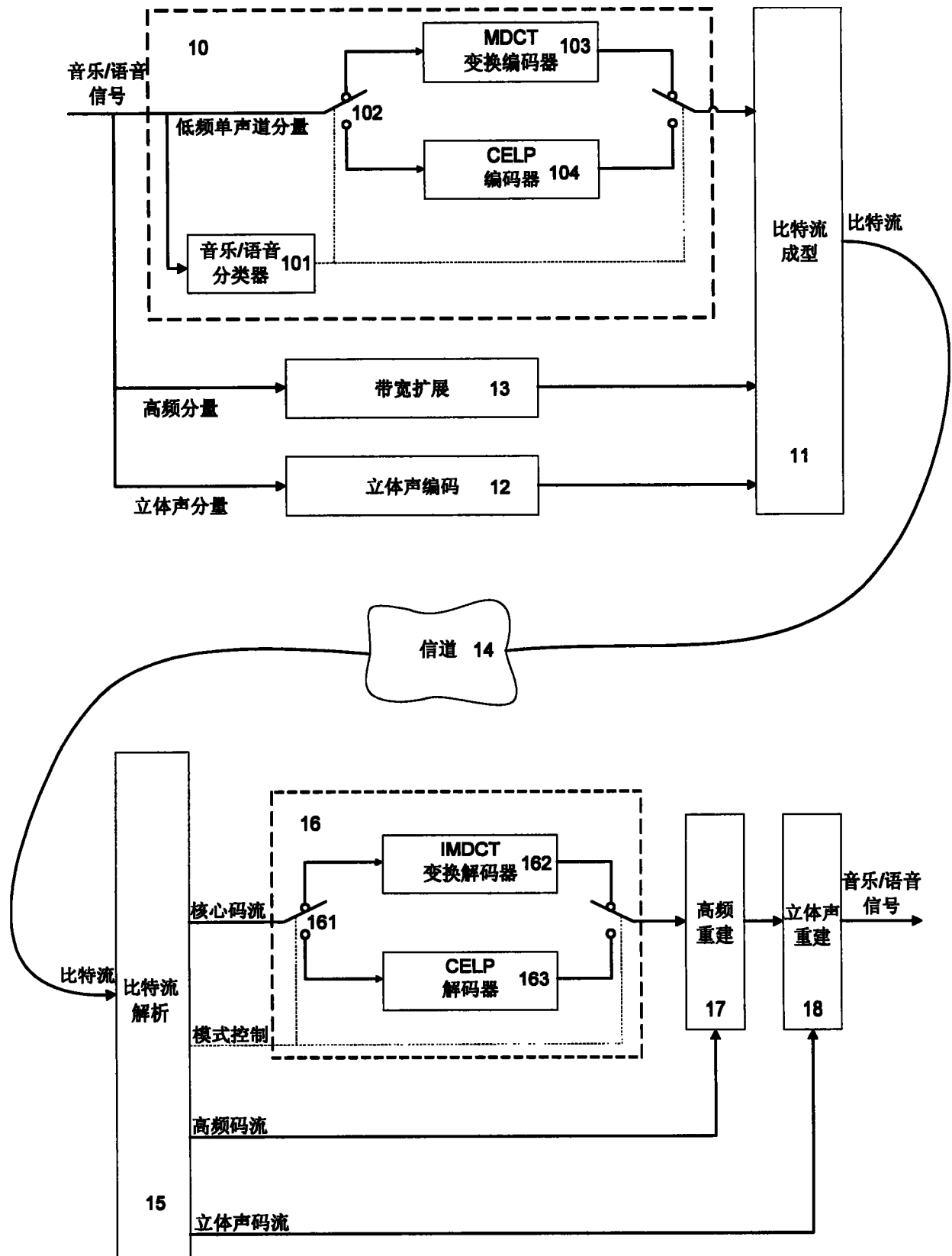


图 1

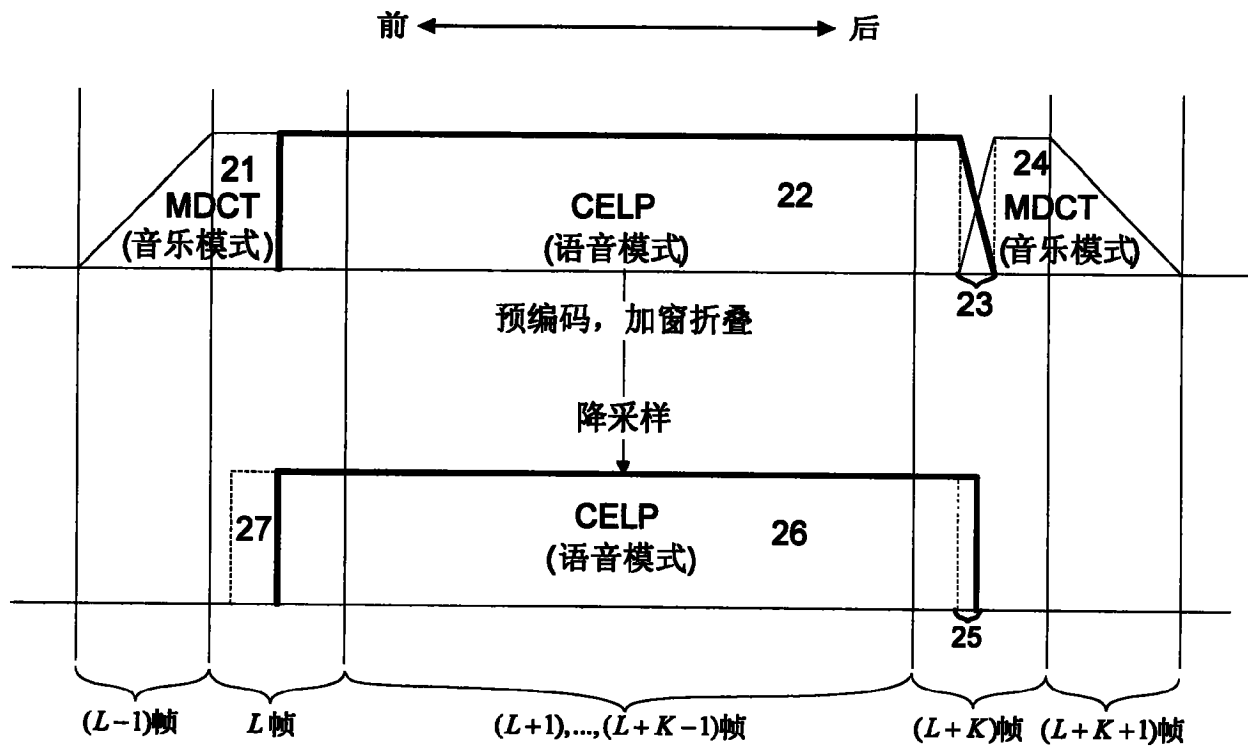


图 2

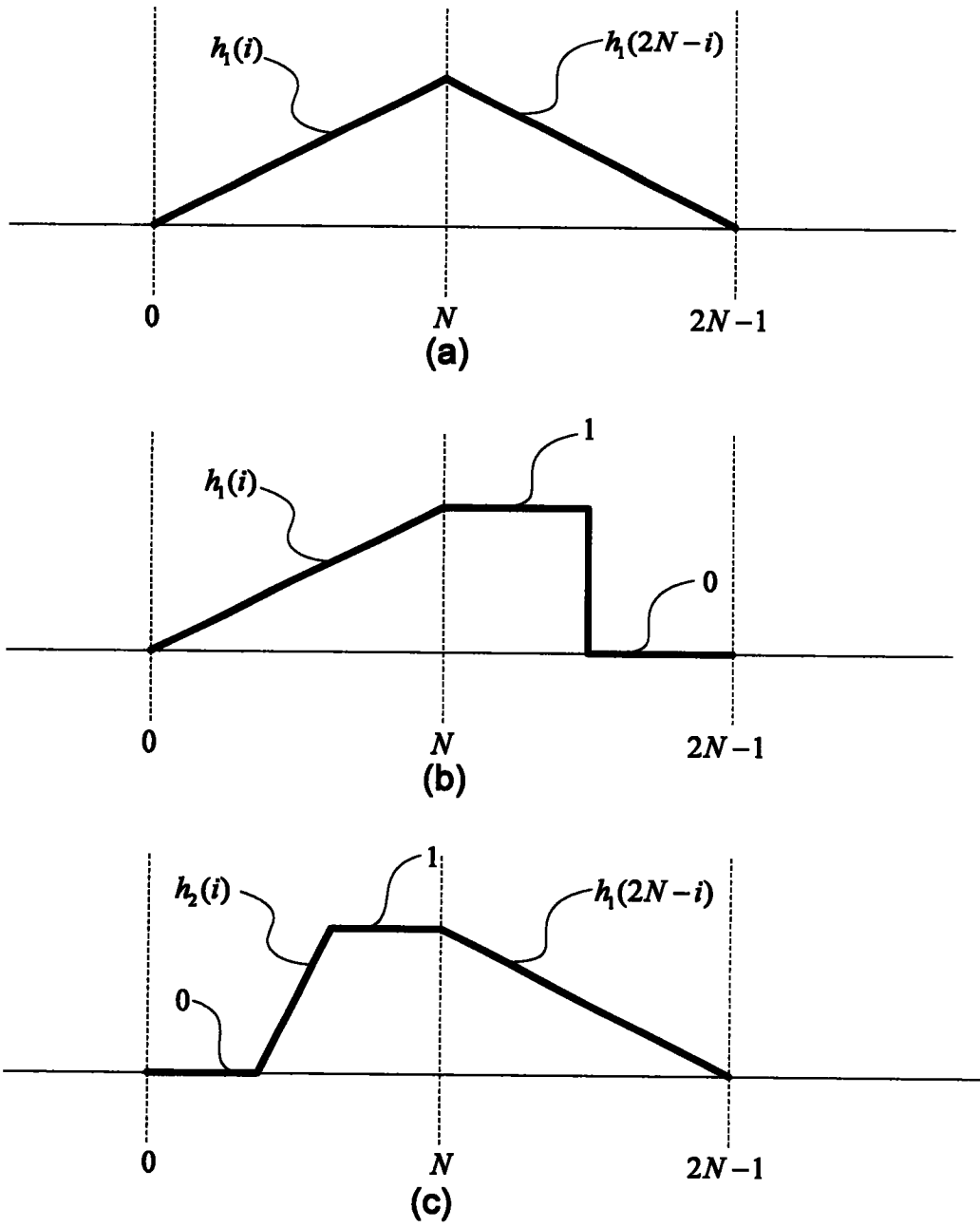


图 3

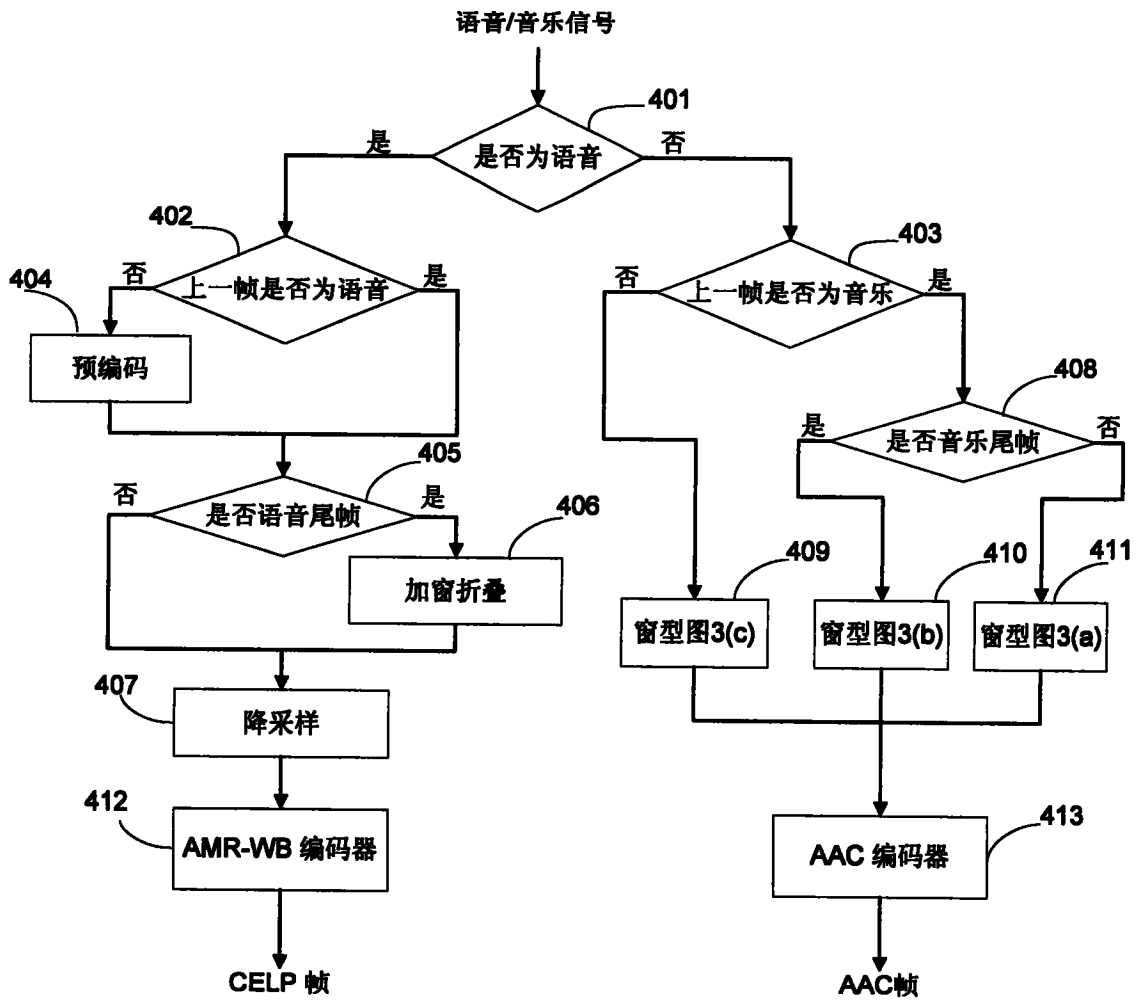


图 4

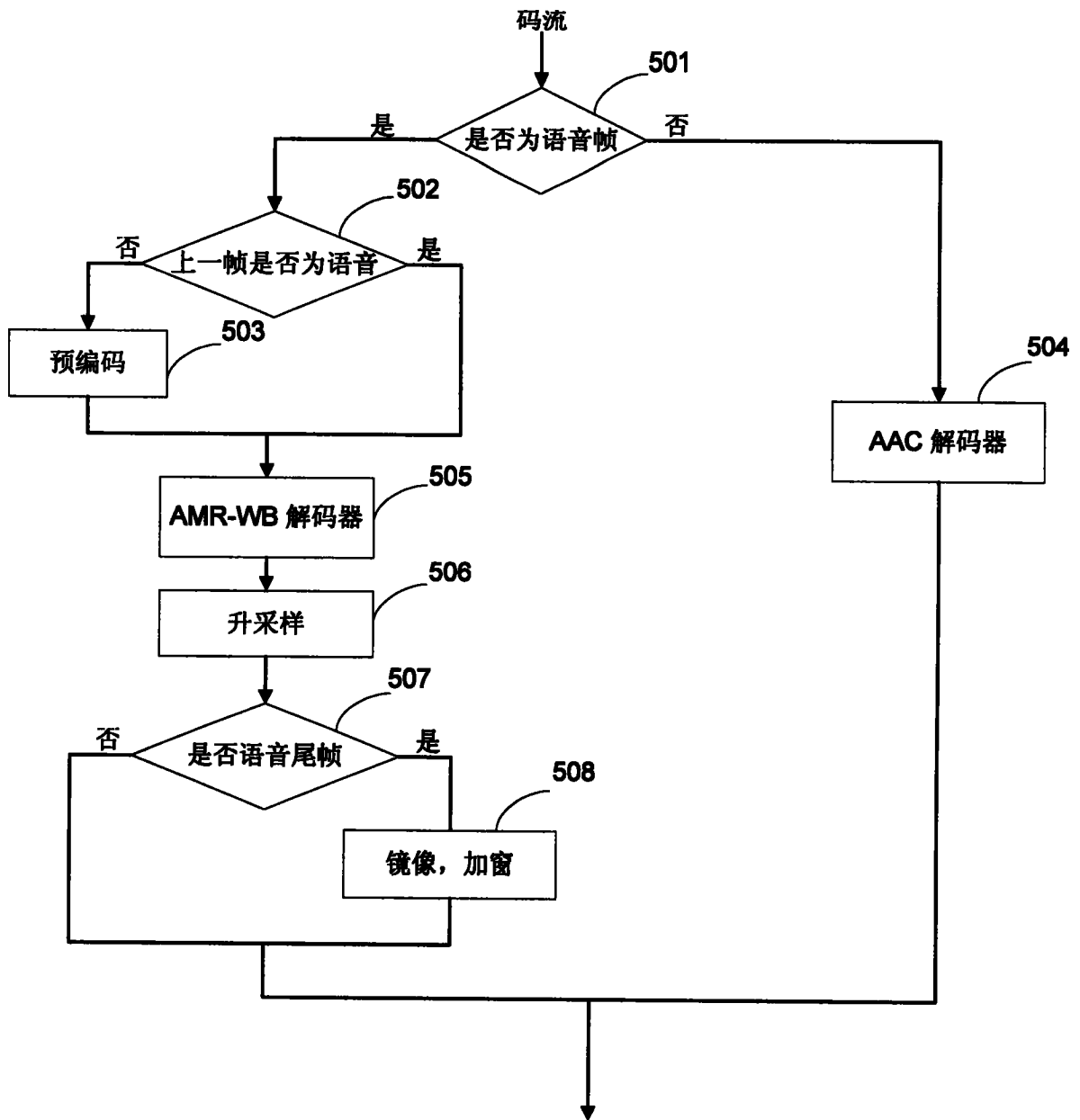


图 5