



(12) 发明专利

(10) 授权公告号 CN 102074242 B

(45) 授权公告日 2012. 03. 28

(21) 申请号 201010606009. 9

审查员 王馨宁

(22) 申请日 2010. 12. 27

(73) 专利权人 武汉大学

地址 430072 湖北省武汉市武昌区珞珈山武汉大学

(72) 发明人 胡瑞敏 杨玉红 高丽 杨裕才
曾琦 陈先念 王国英

(74) 专利代理机构 武汉科皓知识产权代理事务所(特殊普通合伙) 42222
代理人 张火春

(51) Int. Cl.

G10L 19/14(2006. 01)

G10L 19/02(2006. 01)

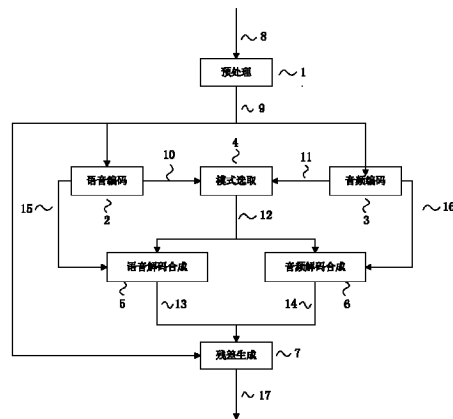
权利要求书 2 页 说明书 5 页 附图 2 页

(54) 发明名称

语音音频混合分级编码中核心层残差提取系统及方法

(57) 摘要

本发明涉及音频编码技术领域,尤其涉及一种语音音频混合分级编码中核心层残差提取系统及方法。本发明提供的系统包括预处理模块(1)、语音编码模块(2)、音频编码模块(3)、模式选取模块(4)、语音解码合成模块(5)、音频解码合成模块(6)、残差生成模块(7);本发明提供的用于获得语音音频混合编码模式下分级核心层与原始信号的准确残差的方法是本发明的主要内容,编码端的语音/音频编码器生成的合成信号用于进行最优编码模式的选择,利用语音/音频编码器输出的编码参数得到与解码端一致的合成信号,从而得到准确的残差用于分级增强层。本发明克服了现有语音音频混合编码模式下可分级核心层残差信号无法准确提取的问题。



1. 一种语音音频混合分级编码框架中核心层残差提取系统,其特征在于,包括:

预处理模块(1):来自音频输入设备的单声道或多声道的输入信号(8)进行预处理,得到当前帧的音频信号(9),并将所获得的当前帧的音频信号(9)做三路输出:一路输出给语音编码模块(2),一路输出给音频编码模块(3),一路输出给残差生成模块(7);

语音编码模块(2):采用通用的语音编码算法对输入的当前帧的音频信号(9)进行编码,编码后作两路输出,其中一路输出为语音编码合成码流(10),传递给模式选取模块(4),用于模式判决,另一路输出为编码提取的各类编码参数(15),传递给语音解码合成模块(5);

音频编码模块(3):采用通用的音频编码算法对输入的当前帧的音频信号(9)进行编码,编码后作两路输出,其中一路输出为音频编码合成码流(11),传递给模式选取模块(4),用于模式判决,另一路输出为编码提取的各类编码参数(16),传递给音频解码合成模块(6);

模式选取模块(4):根据语音编码模块(2)和音频编码模块(3)输出的编码合成码流(10)和(11),选取最优的编码模式,得到模式位标识(12)分两路输出,一路进入语音解码合成模块(5),一路进入音频解码合成模块(6);

语音解码合成模块(5):如果模式选取模块的输出(12)是语音编码模式,则将语音编码模块输出的各类编码参数(15)输入给语音解码合成模块(5),输出最终的核心层语音解码合成信号(13);

音频解码合成模块(6):如果模式选取模块的输出(12)是音频编码模式,则将音频编码模块输出的各类编码参数(16)输入给音频解码合成模块(6),输出最终的核心层音频解码合成信号(14);

残差生成模块(7):预处理模块输出的预处理后的当前帧的音频信号(9)、语音解码合成模块(5)或音频解码合成模块(6)输出的解码合成信号为本模块的两路输入,根据模式位标识不同,选择语音解码合成信号(13)或音频解码合成信号(14),计算得到残差信号(17),本模块的输出为残差信号(17)。

2. 一种语音音频混合分级编码中核心层残差提取方法,其特征在于,包括以下步骤:

①来自音频输入设备的输入信号首先经过预处理,输入信号可以是单声道或多声道信号,预处理可以包括高通滤波、分帧、预加重,得到预处理后的信号 $s(n)$;

②由①所得的预处理后的信号,进行语音编码后一路输出为语音编码合成码流 $x_1(n)$,另一路输出为语音编码参数;

③由①所得的预处理后的信号,进行音频编码后一路输出为音频编码合成码流 $x_2(n)$,另一路输出为音频编码参数;

④由②和③得到的合成码流 $x_1(n)$ 和 $x_2(n)$,进行模式选取,选择最优的编码方式,输出模式位标识;

⑤由④所得的模式位标识,如果选择的是语音编码模式,则转入步骤⑥,执行语音解码合成;如果选择的是音频编码模式,则转入步骤⑦,执行音频解码合成;

⑥由步骤②输出的编码参数,进行语音解码合成得到语音解码合成信号 $x_1'(n)$;

⑦由步骤③输出的编码参数,进行音频解码合成得到音频解码合成信号 $x_2'(n)$;

⑧由①得到预处理后的信号 $s(n)$ 和由⑥或⑦得到的解码后的合成信号 $x_1'(n)$ 或

$x_2'(n)$ 得到残差信号 $r(n)$ 。

3. 根据权利要求 2 所述的语音音频混合分级编码中核心层残差提取方法,其特征在
于:

所述步骤⑥进一步包括以下子步骤:

解码 LP 滤波器参数,通过接收到的 ISP 量化索引合成已量化的 ISP 矢量,插值后的 ISP
矢量被转换到 LP 滤波器系数域,用于合成滤波器重建语音;

解码自适应码本矢量和固定码本矢量及两者的增益,合成语音;

进行白噪声特性增强和基音增强的后处理;

得到最终的合成语音信号并更新公共缓存。

4. 根据权利要求 2 或 3 所述的语音音频混合分级编码中核心层残差提取方法,其特征
在于:

所述步骤⑦进一步包括以下子步骤:

读取量化频率样值,进行基于分裂表的反矢量量化;

增益平衡,去除不同缩放因子的影响;

峰值逆整形;

逆时频变换,信号由频域变换到时域,得到的时域信号与全局增益相乘;

加窗和 TVC 内的重叠相加;

通过逆感知加权滤波器得到合成音频信号;

如果输入信号的前一帧采用的是 ACELP 模式编码,那么将合成音频信号的当前帧起始
的交迭部分和上一帧 ACELP 合成信号的最后一个子帧做加窗交迭,得到最终的合成音频信
号并更新公共缓存。

语音音频混合分级编码中核心层残差提取系统及方法

技术领域

[0001] 本发明涉及音频编码技术领域,尤其涉及一种语音音频混合分级编码中核心层残差提取系统及方法。

背景技术

[0002] 可分级增强层编码方法中,编码端通过计算原始信号与核心层合成信号的残差,同时接收核心层的输出参数,对残差域信号做分级编码。解码端解码出残差域信号,与核心层信号相加,得到恢复的合成信号。各个增强层的残差域信号逐级加到核心层信号上,从而渐次改善重建质量。残差域信号的准确提取与增强层编码所能提供的音质增益有直接关系,解码端解码得到的残差信号越接近编码端计算的残差信号,则合成信号越接近原始信号,解码音质越高。

[0003] 由于语音编码和音频编码的技术和方法存在差异,现有的语音音频的混合编码器,采用不同的模式分别对语音/音频进行编码。系统会根据当前声音信号的类型选择采用何种模式,或者分别采用两种模式进行编码,然后根据编码生成的合成信号选出编码效果好的作为最终的编码模式。由于该合成信号的生成只是用于模式选择,并不等同于解码端生成合成信号的所有操作,再加上两种编码模式的交叠,往往会导致编、解码端得到的合成信号不一致,因而编码端得到的原始信号与合成信号的残差信号并非准确的残差信号。对于语音音频混合编码模式下核心层编码残差的准确提取,成为提高语音音频混合编码器分级编码质量的关键。

发明内容

[0004] 针对上述存在的技术问题,本发明的目的是提供一种语音音频混合分级编码中核心层残差提取系统及方法,以解决语音音频混合分级编码的需求。

[0005] 为达到上述目的,本发明采用如下的技术方案:

[0006] 一种语音音频混合分级编码框架中核心层残差提取系统,包括:

[0007] 预处理模块:来自音频输入设备的单声道或多声道的输入信号进行预处理,得到当前帧的音频信号,并将所获得的当前帧的音频信号做三路输出:一路输出给语音编码模块,一路输出给音频编码模块,一路输出给残差生成模块;

[0008] 语音编码模块:采用通用的语音编码算法对输入的当前帧的音频信号进行编码,编码后作两路输出,其中一路输出为语音解码合成码流,用于模式判决,另一路输出为编码提取的各类编码参数,传递给语音解码合成模块;

[0009] 音频编码模块:采用通用的音频编码算法对输入的当前帧的音频信号进行编码,编码后作两路输出,其中一路输出为音频解码合成码流,用于模式判决,另一路输出为编码提取的各类编码参数,传递给音频解码合成模块;

[0010] 模式选取模块:选取最优的编码模式,得到模式位标识分两路输出,一路进入语音解码合成模块,一路进入音频解码合成模块;

[0011] 语音解码合成模块:如果模式选取模块的输出是语音编码模式,则进入语音解码合成模块,利用语音编码模块输出的各类编码参数作为输入,输出最终的核心层语音解码合成信号;

[0012] 音频解码合成模块:如果模式选取模块的输出是音频编码模式,则进入音频解码合成模块,利用音频编码模块输出的各类编码参数作为输入,输出最终的核心层音频解码合成信号;

[0013] 残差生成模块:预处理模块输出的预处理后的当前帧的音频信号和语音/音频解码合成模块输出的解码后合成信号为本模块的两路输入,根据模式位不同,选择语音解码合成信号或音频解码合成信号,计算得到残差信号,本模块的输出为残差信号。

[0014] 一种语音音频混合分级编码中核心层残差提取方法,包括:

[0015] ①输入语音/音频信号首先经过预处理,输入信号可以是单声道或多声道信号,预处理可以包括高通滤波、分帧、预加重等处理,得到预处理后的信号 $s(n)$;

[0016] ②由①所得的预处理后的音频信号,进行语音编码后一路输出为语音编码合成码流 $x_1(n)$,另一路输出为语音编码参数;

[0017] ③由①所得的预处理后的音频信号,进行音频编码后一路输出为音频编码合成码流 $x_2(n)$,另一路输出为音频编码参数;

[0018] ④由②和③得到的合成码流 $x_1(n)$ 和 $x_2(n)$,进行模式选取,选择最优的编码方式,输出模式位标识;

[0019] ⑤由④所得的模式位,如果选择的是语音编码模式,则转入步骤⑥,执行语音解码合成;如果选择的是音频编码模式,则转入步骤⑦,执行音频解码合成;

[0020] ⑥由步骤②输出的编码参数,进行语音解码合成得到语音解码合成信号 $\tilde{x}_1(n)$;

[0021] ⑦由步骤③输出的编码参数,进行音频解码合成得到音频解码合成信号 $\tilde{x}_2(n)$;

[0022] ⑧由①得到预处理后的信号 $s(n)$ 和由⑥或⑦得到的解码后的合成信号 $\tilde{x}_1(n)$ 或 $\tilde{x}_2(n)$ 得到残差信号 $r(n)$ 。

[0023] 所述步骤⑥进一步包括以下子步骤:

[0024] 解码 LP 滤波器参数,通过接收到的 ISP 量化索引合成已量化的 ISP 矢量,插值后的 ISP 矢量被转换到 LP 滤波器系数域,用于合成滤波器重建语音;

[0025] 解码自适应码本矢量和固定码本矢量及两者的增益,合成语音;

[0026] 进行白噪声特性增强和基音增强的后处理;

[0027] 得到最终的合成音频信号并更新公共缓存。

[0028] 所述步骤⑦进一步包括以下子步骤:

[0029] 读取量化频率样值,进行基于分裂表的反矢量量化;

[0030] 增益平衡,去除不同缩放因子的影响;

[0031] 峰值逆整形;

[0032] 逆时频变换,信号由频域变换到时域,得到的时域信号与全局增益相乘;

[0033] 加窗和 TVC 内的重叠相加;

[0034] 通过逆感知加权滤波器得到合成音频信号;

[0035] 如果前一帧采用的是 ACELP 模式编码,那么将当前帧起始的交迭部分和上一帧 ACELP 合成信号的最后一个子帧做加窗交迭,得到最终的合成音频信号并更新公共缓存。

[0036] 本发明具有以下优点和积极效果：

[0037] 本发明克服了现有语音音频混合编码模式下可分级核心层残差信号无法准确提取的问题。

附图说明

[0038] 图 1 是本发明提供的语音音频混合分级编码中核心层残差提取系统的框架示意图。

[0039] 图 2 是本发明提供的语音音频混合分级编码中核心层残差提取方法的流程图。

[0040] 1- 预处理模块, 2- 语音编码模块, 3- 音频编码模块, 4- 模式选取模块, 5- 语音解码合成模块, 6- 音频解码合成模块, 7- 残差生成模块, 8- 音频原始信号, 9- 预处理后信号, 10- 语音编码合成信号, 11- 音频编码合成信号, 12- 编码模式标识位, 13- 语音解码合成信号, 14- 音频解码合成信号, 15- 语音编码参数, 16- 音频编码参数, 17- 残差信号。

具体实施方式

[0041] 下面以具体实施例结合附图对本发明作进一步说明：

[0042] 本发明提供的语音音频混合分级编码中核心层残差提取系统分为两部分, 一部分为核心编码层, 另一部分为残差增强层, 其中核心层采用通用的语音 / 音频混合编码技术, 残差增强层可以分为多个增强层, 如图 1 所示, 包括：

[0043] 预处理模块 1、语音编码模块 2、音频编码模块 3、模式选取模块 4、语音解码合成模块 5、音频解码合成模块 6、残差生成模块 7；

[0044] 预处理模块 1：对输入信号进行预处理, 其输入为音频原始信号, 输出为预处理后的信号, 预处理后的信号做三路输出：一路输出给语音编码模块, 一路输出给音频编码模块, 一路输出给残差生成模块；

[0045] 语音编码模块 2：该模块是可分级编码框架中的核心层编码模块, 采用通用的语音编码算法对预处理模块的输入信号进行编码, 编码后作两路输出, 其中一路输出为语音编码合成码流, 用于模式判决, 另一路输出为编码提取的各类编码参数, 传递给语音解码合成模块, 用于计算准确的合成信号；

[0046] 音频编码模块 3：该模块是可分级编码框架中的核心层编码模块, 采用通用的音频编码算法对预处理模块的输入信号进行编码, 编码后作两路输出, 其中一路输出为音频编码合成码流, 用于模式判决, 另一路输出为编码提取的各类编码参数, 传递给音频解码合成模块, 用于计算准确的合成信号；

[0047] 模式选取模块 4：对于语音编码模块和音频编码模块的两路输出, 选取最优的模式作为编码器, 输出为模式位标识。如果选择了语音编码模式, 则进入语音解码合成模块, 如果选择了音频编码模式, 则进入音频解码合成模块；

[0048] 语音解码合成模块 5：如果模式选取模块的输出是语音编码模式, 则进入语音解码合成模块, 利用语音编码模块输出的各类编码参数作为输入, 输出最终的核心层语音编码合成信号；

[0049] 音频解码合成模块 6：如果模式选取模块的输出是音频编码模式, 则进入音频解码合成模块, 利用音频编码模块输出的各类编码参数作为输入, 输出最终的核心层音频编

码合成信号；

[0050] 残差生成模块 7：预处理模块输出的预处理后的信号和语音 / 音频解码合成模块输出的解码后合成信号为本模块的两路输入，得到残差信号，本模块的输出为残差信号。

[0051] 本发明提供的语音音频混合分级编码中核心层残差提取方法包括以下步骤：

[0052] 步骤 1：对采样率为 16kHz 输入信号进行预处理，预处理具体包括高通滤波和感知加权两个过程，输出信号为 $s(n)$ ；

[0053] 将输入信号送入高通滤波器，滤除 50Hz 以下的低频信号；

[0054] 将高通滤波后的信号送入感知加权滤波器 $W_{LB}(z)$ ，同时 γ'_1, γ'_2 以及 γ'_3 ($0 < \gamma'_1, \gamma'_2, \gamma'_3 < 1$) 三个系数也相应调整以缓和量化噪声谱：

$$[0055] \quad W_{LB}(z) = \frac{\hat{A}(z/\gamma'_1)}{\hat{A}(z/\gamma'_2)} \left(1 + \sum_{i=1}^2 a_i \gamma_3^i z^{-i} \right)$$

[0056] 其中 $\gamma'_1, \gamma'_2, \gamma'_3$ 为调整参量， a_i 为线性预测分析系数， i 为线性预测的阶数， $\hat{A}(z) = \hat{a}_0 + \hat{a}_1 z^{-1} + \dots + \hat{a}_{10} z^{-10}$ 。

[0057] 步骤 2：对经过预处理过后的信号进行 12kbps 模式的 ACELP 编码器编码，进行编码后一路输出为语音编码合成码流 $x_1(n)$ ，另一路输出为语音编码参数，包括 ISF 索引值、VQ 增益索引值、码本索引值、滤波索引值、基因延迟索引值；

[0058] 步骤 3：对经过预处理过后的信号进行 12kbps 模式的 TVC 编码器（变换域音频编码器）编码，进行编码后一路输出为音频编码合成码流 $x_2(n)$ ，另一路输出为音频编码参数，包括量化频率样值、缩放因子、全局增益；

[0059] 步骤 4：由步骤 2 和步骤 3 得到的合成码流 $x_1(n)$ 和 $x_2(n)$ ，分别计算与步骤 1 的输出信号 $s(n)$ 的感知加权分段信噪比 SNR1 和 SNR2，如果 SNR1 > SNR2，则模式位标识 mod 置 0，选择 ACELP 编码器编码，否则 mod 置 1，选择 TVC 编码器编码；

[0060] 步骤 5：如果 mod 为 0，则转入步骤 6，执行语音解码合成模块；如果 mod 为 1，则转入步骤 7，执行音频解码合成模块；

[0061] 步骤 6：由步骤 2 输出的编码参数，通过 ACELP 解码器合成得到解码合成信号 $\tilde{x}_1(n)$ ，本步骤具体实施方式包括以下子步骤：

[0062] ①. 解码 LP 滤波器参数，通过接收到的 ISP 量化索引合成已量化的 ISP 矢量，插值后的 ISP 矢量被转换到 LP 滤波器系数域，用于合成滤波器重建语音；

[0063] ②. 解码自适应码本矢量和固定码本矢量及两者的增益，合成语音；

[0064] ③. 后处理（白噪声特性增强和基音增强）；

[0065] ④. 得到最终的合成音频信号并更新公共缓存。

[0066] 步骤 7：由步骤 3 输出的编码参数，通过 TVC 解码器合成得到音频解码合成信号 $\tilde{x}_2(n)$ ，本步骤具体实施方式包括以下子步骤：

[0067] ①. 读取量化频率样值，进行基于分裂表的反矢量量化；

[0068] ②. 增益平衡，去除不同缩放因子的影响；

[0069] ③. 峰值逆整形；

[0070] ④. 逆时频变换，信号由频域变换到时域，得到的时域信号与全局增益相乘；

[0071] ⑤. 加窗和 TVC 内的重叠相加；

[0072] ⑥. 通过逆感知加权滤波器得到合成音频信号；

[0073] ⑦. 如果前一帧采用的是 ACELP 模式编码, 那么将当前帧起始的交迭部分和上一帧 ACELP 合成信号的最后一个子帧做加窗交迭, 得到最终的合成音频信号并更新公共缓存。

[0074] 步骤 8 : 由步骤 1 得到预处理后的信号 $s(n)$ 减去解码后的合成信号, 如果采用语音编码模式, 则与 $\tilde{x}_1(n)$ 相减, 如采用音频编码模式则与 $\tilde{x}_2(n)$ 相减, 得到残差信号 $r(n)$ 。

[0075] 本发明的分级核心层语音音频编码方法是通用的编码算法, 获得语音音频混合编码模式下分级核心层与原始信号的准确残差的方法是本发明的主要内容, 编码端的语音 / 音频编码器生成的合成信号用于进行最优编码模式的选择, 利用语音 / 音频编码器输出的编码参数得到与解码端一致的合成信号, 从而得到准确的残差用于分级增强层。

[0076] 以上实施例仅供说明本发明之用, 而非对本发明的限制, 有关技术领域的技术人员, 在不脱离本发明的精神和范围的情况下, 还可以作出各种变换或变型, 因此所有等同的技术方案, 都落入本发明的保护范围。

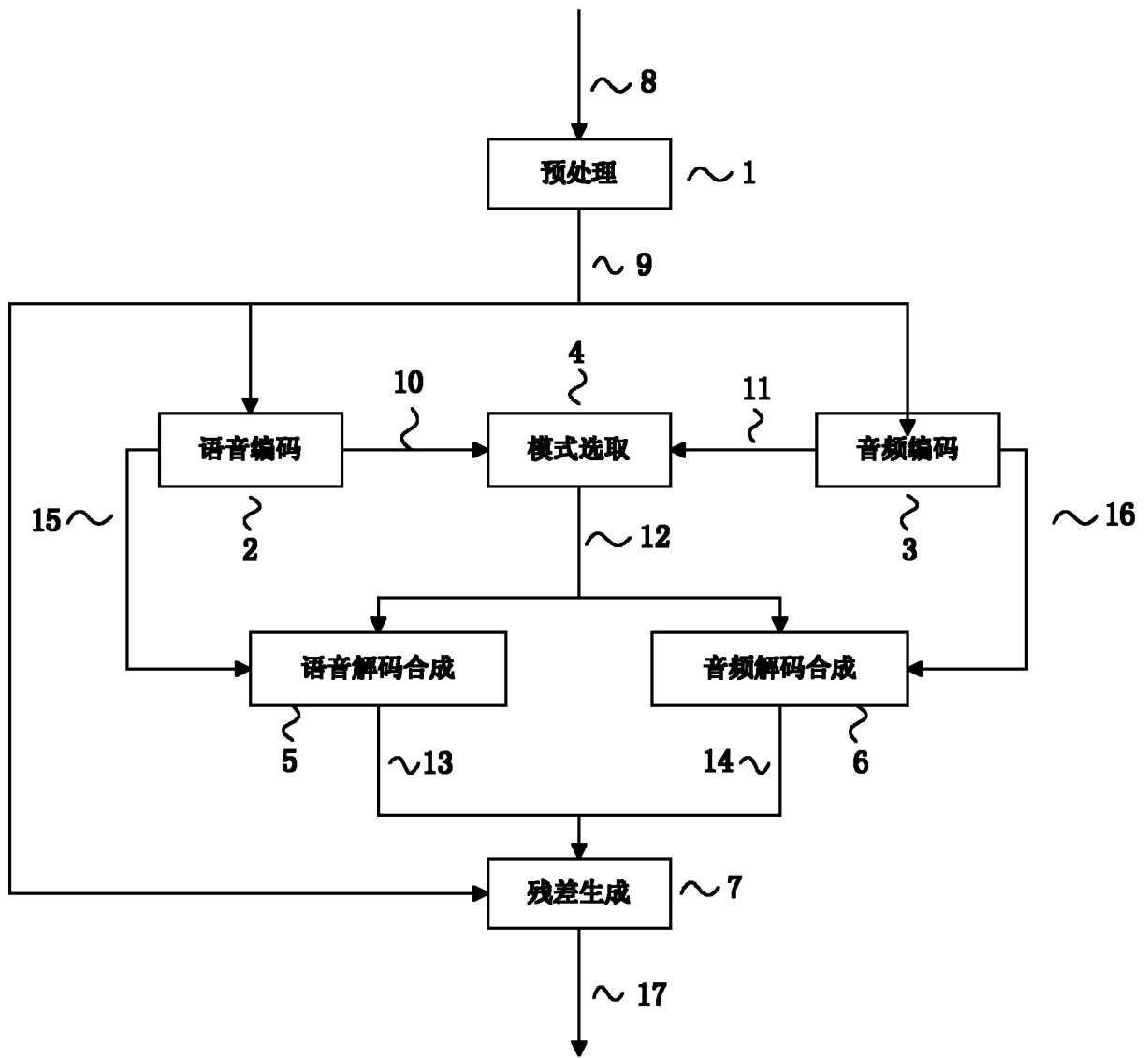


图 1

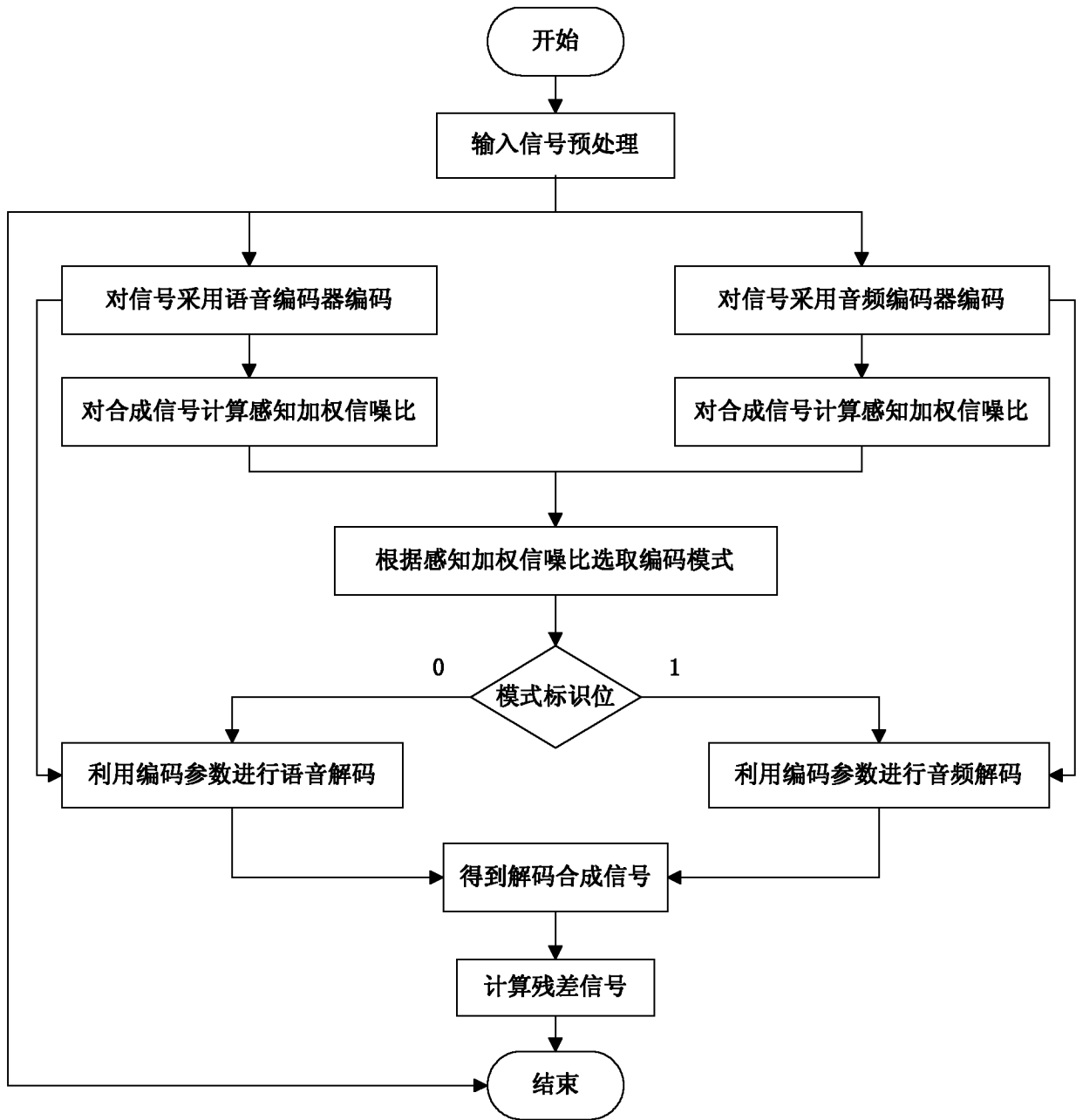


图 2