



(12) 发明专利

(10) 授权公告号 CN 107742521 B

(45) 授权公告日 2021.08.13

(21) 申请号 201610652507.4

H04S 3/00 (2006.01)

(22) 申请日 2016.08.10

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 107742521 A

US 2006206323 A1, 2006.09.14

US 2009119111 A1, 2009.05.07

CN 104205211 A, 2014.12.10

(43) 申请公布日 2018.02.27

AU 2011357816 A1, 2013.08.15

WO 2013029225 A1, 2013.03.07

(73) 专利权人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

CN 101601217 A, 2009.12.09

CN 103065636 A, 2013.04.24

CN 103854649 A, 2014.06.11

(72) 发明人 李海婷 刘泽新 张兴涛 苗磊

CN 103280222 A, 2013.09.04

US 2006206323 A1, 2006.09.14

(74) 专利代理机构 北京龙双利达知识产权代理有限公司 11329

审查员 张鹤

代理人 王君 张欣

(51) Int. Cl.

G10L 19/008 (2013.01)

G10L 19/02 (2013.01)

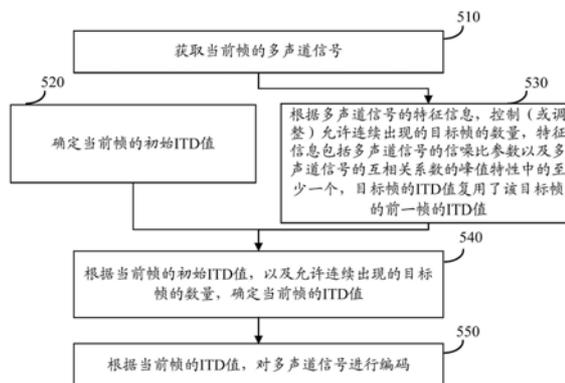
权利要求书4页 说明书27页 附图4页

(54) 发明名称

多声道信号的编码方法和编码器

(57) 摘要

本发明提供一种多声道信号的编码方法和编码器,该编码方法包括:获取当前帧的多声道信号;确定当前帧的初始ITD值;根据多声道信号的特征信息,控制允许连续出现的目标帧的数量,特征信息包括多声道信号的信噪比参数以及多声道信号的互相关系数的峰值特性中的至少一个,目标帧的ITD值复用了目标帧的前一帧的ITD值;根据当前帧的初始ITD值,以及允许连续出现的目标帧的数量,确定当前帧的ITD值;根据当前帧的ITD值,对多声道信号进行编码。本发明能够提高多声道信号的编码质量。



1. 一种多声道信号的编码方法,其特征在于,包括:

获取当前帧的多声道信号;

确定所述当前帧的初始声道间时间差ITD值;

根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量,所述特征信息包括所述多声道信号的信噪比参数以及所述多声道信号的互相关系数的峰值特性,所述目标帧的ITD值复用了所述目标帧的前一帧的ITD值;

根据所述当前帧的初始ITD值,以及所述允许连续出现的目标帧的数量,确定所述当前帧的ITD值;

根据所述当前帧的ITD值,对所述多声道信号进行编码。

2. 如权利要求1所述的方法,其特征在于,在所述根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量之前,所述方法还包括:

根据所述多声道信号的互相关系数的峰值的幅度和所述多声道信号的互相关系数的峰值位置的索引,确定所述多声道信号的互相关系数的峰值特性。

3. 如权利要求2所述的方法,其特征在于,所述根据所述多声道信号的互相关系数的峰值的幅度和所述多声道信号的互相关系数的峰值位置的索引,确定所述多声道信号的互相关系数的峰值特性,包括:

根据所述多声道信号的互相关系数的峰值的幅度,确定峰值幅度可信度参数,所述峰值幅度可信度参数表征所述多声道信号的互相关系数的峰值幅度的可信度;

根据所述多声道信号的互相关系数的峰值位置的索引对应的ITD值,以及所述当前帧的前一帧的ITD值,确定峰值位置波动性参数,所述峰值位置波动性参数表征所述多声道信号的互相关系数的峰值位置的索引对应的ITD值与所述当前帧的前一帧的ITD值的差异;

根据所述峰值幅度可信度参数和所述峰值位置波动性参数,确定所述多声道信号的互相关系数的峰值特性。

4. 如权利要求3所述的方法,其特征在于,所述根据所述多声道信号的互相关系数的峰值的幅度,确定峰值幅度可信度参数,包括:

将所述多声道信号的互相关系数中的峰值的幅度值和次大值的幅度值之差与所述峰值的幅度值的比值确定为所述峰值幅度可信度参数。

5. 如权利要求3或4所述的方法,其特征在于,所述根据所述多声道信号的互相关系数的峰值位置的索引对应的ITD值,以及所述当前帧的前一帧的ITD值,确定峰值位置波动性参数,包括:

将所述多声道信号的互相关系数的峰值位置的索引对应的ITD值与所述当前帧的前一帧的ITD值之差的绝对值确定为所述峰值位置波动性参数。

6. 如权利要求3或4所述的方法,其特征在于,所述根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量,包括:

根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量,在所述多声道信号的互相关系数的峰值特性满足预设条件的情况下,通过调整目标帧计数值和所述目标帧计数值的阈值中的至少一个,减少允许连续出现的目标帧的数量,其中,所述目标帧计数值用于表征当前已连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

7. 如权利要求6所述的方法,其特征在于,所述通过调整目标帧计数值和所述目标帧计数值的阈值中的至少一个,减少允许连续出现的目标帧的数量,包括:

通过增加所述目标帧计数值,减少允许连续出现的目标帧的数量。

8. 如权利要求7所述的方法,其特征在于,所述通过调整目标帧计数值和所述目标帧计数值的阈值中的至少一个,减少允许连续出现的目标帧的数量,包括:

通过减小所述目标帧计数值的阈值,减少允许连续出现的目标帧的数量。

9. 如权利要求8所述的方法,其特征在于,所述根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量,包括:

在所述多声道信号的信噪比参数不满足预设的信噪比条件的情况下,才根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量;

所述方法还包括:

在所述多声道信号的信噪比满足所述信噪比条件的情况下,停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值。

10. 如权利要求1-4中任一项所述的方法,其特征在于,所述根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量,包括:

确定所述多声道信号的信噪比参数是否满足预设的信噪比条件;

在所述多声道信号的信噪比参数不满足所述信噪比条件的情况下,根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量;

在所述多声道信号的信噪比满足所述信噪比条件的情况下,停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值。

11. 如权利要求9所述的方法,其特征在于,所述停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值,包括:

增加目标帧计数值,使得所述目标帧计数值的取值大于或等于所述目标帧计数值的阈值,其中,所述目标帧计数值用于表征当前已经连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

12. 如权利要求1-4中任一项所述的方法,其特征在于,所述根据所述当前帧的初始ITD值,以及所述允许连续出现的目标帧的数量,确定所述当前帧的ITD值,包括:

根据所述当前帧的初始ITD值,目标帧计数值,所述目标帧计数值的阈值,确定所述当前帧的ITD值,其中,所述目标帧计数值用于表征当前已连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

13. 如权利要求1-4中任一项所述的方法,其特征在于,所述信噪比参数为所述多声道信号的修正的分段信噪比。

14. 一种编码器,其特征在于,包括:

获取单元,用于获取当前帧的多声道信号;

第一确定单元,用于确定所述当前帧的初始声道间时间差ITD值;

控制单元,用于根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量,所述特征信息包括所述多声道信号的信噪比参数以及所述多声道信号的互相关系数的峰值特性,所述目标帧的ITD值复用了所述目标帧的前一帧的ITD值;

第二确定单元,用于根据所述当前帧的初始ITD值,以及所述允许连续出现的目标帧的

数量,确定所述当前帧的ITD值;

编码单元,用于根据所述当前帧的ITD值,对所述多声道信号进行编码。

15.如权利要求14所述的编码器,其特征在于,所述编码器还包括:

第三确定单元,用于根据所述多声道信号的互相关系数的峰值的幅度和所述多声道信号的互相关系数的峰值位置的索引,确定所述多声道信号的互相关系数的峰值特性。

16.如权利要求15所述的编码器,其特征在于,所述第三确定单元具体用于根据所述多声道信号的互相关系数的峰值的幅度,确定峰值幅度可信度参数,所述峰值幅度可信度参数表征所述多声道信号的互相关系数的峰值幅度的可信度;根据所述多声道信号的互相关系数的峰值位置的索引对应的ITD值,以及所述当前帧的前一帧的ITD值,确定峰值位置波动性参数,所述峰值位置波动性参数表征所述多声道信号的互相关系数的峰值位置的索引对应的ITD值与所述当前帧的前一帧的ITD值的差异;根据所述峰值幅度可信度参数和所述峰值位置波动性参数,确定所述多声道信号的互相关系数的峰值特性。

17.如权利要求16所述的编码器,其特征在于,所述第三确定单元具体用于将所述多声道信号的互相关系数中的峰值的幅度值和次大值的幅度值之差与所述峰值的幅度值的比值确定为所述峰值幅度可信度参数。

18.如权利要求16或17所述的编码器,其特征在于,所述第三确定单元具体用于将所述多声道信号的互相关系数的峰值位置的索引对应的ITD值与所述当前帧的前一帧的ITD值之差的绝对值确定为所述峰值位置波动性参数。

19.如权利要求16或17所述的编码器,其特征在于,所述控制单元具体用于根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量,在所述多声道信号的互相关系数的峰值特性满足预设条件的情况下,通过调整目标帧计数值和所述目标帧计数值的阈值中的至少一个,减少允许连续出现的目标帧的数量,其中,所述目标帧计数值用于表征当前已连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

20.如权利要求19所述的编码器,其特征在于,所述控制单元具体用于通过增加所述目标帧计数值,减少允许连续出现的目标帧的数量。

21.如权利要求20所述的编码器,其特征在于,所述控制单元具体用于通过减小所述目标帧计数值的阈值,减少允许连续出现的目标帧的数量。

22.如权利要求21所述的编码器,其特征在于,所述控制单元具体用于在所述多声道信号的信噪比参数不满足预设的信噪比条件的情况下,才根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量;所述编码器还包括:停止单元,用于在所述多声道信号的信噪比满足所述信噪比条件的情况下,停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值。

23.如权利要求14-17中任一项所述的编码器,其特征在于,所述控制单元具体用于确定所述多声道信号的信噪比参数是否满足预设的信噪比条件;在所述多声道信号的信噪比参数不满足所述信噪比条件的情况下,根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量;在所述多声道信号的信噪比满足所述信噪比条件的情况下,停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值。

24.如权利要求22所述的编码器,其特征在于,所述停止单元具体用于增加目标帧计数

值,使得所述目标帧计数值的取值大于或等于所述目标帧计数值的阈值,其中,所述目标帧计数值用于表征当前已经连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

25.如权利要求14-17中任一项所述的编码器,其特征在于,所述第二确定单元具体用于根据所述当前帧的初始ITD值,目标帧计数值,所述目标帧计数值的阈值,确定所述当前帧的ITD值,其中,所述目标帧计数值用于表征当前已连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

26.如权利要求14-17中任一项所述的编码器,其特征在于,所述信噪比参数为所述多声道信号的修正的分段信噪比。

多声道信号的编码方法和编码器

技术领域

[0001] 本发明涉及音频信号编码领域,并且更为具体地,涉及一种多声道信号的编码方法和编码器。

背景技术

[0002] 随着生活质量的提高,人们对高质量音频的需求不断增大。相对于单声道信号,立体声具有各声源的方位感和分布感,能够提高声音的清晰度、可懂度及临场感,因而备受人们青睐。

[0003] 立体声处理技术主要有和差(Mid/Sid,MS)编码、强度立体声(Intensity Stereo,IS)编码以及参数立体声(Parametric Stereo,PS)编码。

[0004] MS编码基于声道间相关性将两路信号作和、差变换,各声道能量主要集中在和声道,使声道间冗余得以去除。在MS编码技术中,码率的节省依赖于输入信号的相关性,当左右声道信号的相关性差时,需分别传输左声道信号和右声道信号。

[0005] IS编码基于人耳听觉系统对声道的高频成分(例如,大于2kHz的成分)的相位差异不敏感的特性,将左右两路信号的高频分量进行简化处理。但IS编码技术仅对高频成分有效,如将IS编码技术扩展到低频,将会引起严重的人为噪声。

[0006] PS编码是基于双耳听觉模型的编码方式。如图1所示(图1中的 x_L 为左声道时域信号, x_R 为右声道时域信号),在PS编码过程中,编码端会将立体声信号转换成单声道信号和少量描述空间声场的空间参数(或称空间感知参数)。如图2所示,解码端得到单声道信号和空间参数之后,会结合空间参数恢复立体声信号。相对于MS编码,PS编码压缩比高,因此,PS编码可以在保持较好音质的前提下,获得更高的编码增益。此外,PS编码可以工作在全音频带宽中,能够很好地还原立体声的空间感知效果。

[0007] PS编码中,空间参数包括声道间相关性(Inter-channel Coherent,IC)、声道间电平差(Inter-channel Level Difference,ILD)、声道间时间差(Inter-channel Time Difference,ITD)以及声道间相位差(Inter-channel Phase Difference,IPD)。IC描述了声道间的互相关或相干性,该参数决定了声场范围的感知,可以提高音频信号的空间感和声响稳定性。ILD用于分辨立体声源的水平方向角度,描述了声道间的能量差别,该参数将影响整个频谱的频率成分。ITD和IPD为表示声源水平方位的空间参数,描述了声道间的时间和相位的差别。ILD、ITD和IPD能够决定人耳对声源位置的感知,可以有效确定声场位置,对立体声信号的恢复具有重要作用。

[0008] 在立体声的录音过程中,受到背景噪声、混响、多人同时讲话等因素的影响,按照现有的PS编码方式计算出的ITD经常会出现不稳定(ITD的取值来回跳变)的现象。如果基于这样的ITD计算下混合信号,就会导致下混合信号不连续,从而导致解码端得到的立体声质量差,如解码端播放的立体声的声像会频繁晃动,甚至出现听感上的卡顿。

发明内容

[0009] 本申请提供一种多声道信号的编码方法和编码器,以提升PS编码中的ITD的稳定性,从而提升多声道信号的编码质量。

[0010] 第一方面,提供一种多声道信号的编码方法,包括:获取当前帧的多声道信号;确定所述当前帧的初始ITD值;根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量,所述特征信息包括所述多声道信号的信噪比参数以及所述多声道信号的互相关系数的峰值特性中的至少一个,所述目标帧的ITD值复用了所述目标帧的前一帧的ITD值;根据所述当前帧的初始ITD值,以及所述允许连续出现的目标帧的数量,确定所述当前帧的ITD值;根据所述当前帧的ITD值,对所述多声道信号进行编码。

[0011] 结合第一方面,在第一方面的某些实现方式中,在所述根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量之前,所述方法还包括:根据所述多声道信号的互相关系数的峰值的幅度和所述多声道信号的互相关系数的峰值位置的索引,确定所述多声道信号的互相关系数的峰值特性。

[0012] 结合第一方面,在第一方面的某些实现方式中,所述根据所述多声道信号的互相关系数的峰值的幅度和所述多声道信号的互相关系数的峰值位置的索引,确定所述多声道信号的互相关系数的峰值特性,包括:根据所述多声道信号的互相关系数的峰值的幅度,确定峰值幅度可信度参数,所述峰值幅度可信度参数表征所述多声道信号的互相关系数的峰值幅度的可信度;根据所述多声道信号的互相关系数的峰值位置的索引对应的ITD值,以及所述当前帧的前一帧的ITD值,确定峰值位置波动性参数,所述峰值位置波动性参数表征所述多声道信号的互相关系数的峰值位置的索引对应的ITD值与所述当前帧的前一帧的ITD值的差异;根据所述峰值幅度可信度参数和所述峰值位置波动性参数,确定所述多声道信号的互相关系数的峰值特性。

[0013] 结合第一方面,在第一方面的某些实现方式中,所述根据所述多声道信号的互相关系数的峰值的幅度,确定峰值幅度可信度参数,包括:将所述多声道信号的互相关系数中的峰值的幅度值和次大值的幅度值之差与所述峰值的幅度值的比值确定为所述峰值幅度可信度参数。

[0014] 结合第一方面,在第一方面的某些实现方式中,所述根据所述多声道信号的互相关系数的峰值位置的索引对应的ITD值,以及所述当前帧的前一帧的ITD值,确定峰值位置波动性参数,包括:将所述多声道信号的互相关系数的峰值位置的索引对应的ITD值与所述当前帧的前一帧的ITD值之差的绝对值确定为所述峰值位置波动性参数。

[0015] 结合第一方面,在第一方面的某些实现方式中,所述根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量,包括:根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量,在所述多声道信号的互相关系数的峰值特性满足预设条件的情况下,通过调整目标帧计数值和所述目标帧计数值的阈值中的至少一个,减少允许连续出现的目标帧的数量,其中,所述目标帧计数值用于表征当前已连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

[0016] 结合第一方面,在第一方面的某些实现方式中,所述通过调整目标帧计数值和所述目标帧计数值的阈值中的至少一个,减少允许连续出现的目标帧的数量,包括:通过增加所述目标帧计数值,减少允许连续出现的目标帧的数量。

[0017] 结合第一方面,在第一方面的某些实现方式中,所述通过调整目标帧计数值和所述目标帧计数值的阈值中的至少一个,减少允许连续出现的目标帧的数量,包括:通过减小所述目标帧计数值的阈值,减少允许连续出现的目标帧的数量。

[0018] 结合第一方面,在第一方面的某些实现方式中,所述根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量,包括:在所述多声道信号的信噪比参数不满足预设的信噪比条件的情况下,才根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量;所述方法还包括:在所述多声道信号的信噪比满足所述信噪比条件的情况下,停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值。

[0019] 结合第一方面,在第一方面的某些实现方式中,所述根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量,包括:确定所述多声道信号的信噪比参数是否满足预设的信噪比条件;在所述多声道信号的信噪比参数不满足所述信噪比条件的情况下,根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量;在所述多声道信号的信噪比满足所述信噪比条件的情况下,停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值。

[0020] 结合第一方面,在第一方面的某些实现方式中,所述停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值,包括:增加目标帧计数值,使得所述目标帧计数值的取值大于或等于所述目标帧计数值的阈值,其中,所述目标帧计数值用于表征当前已经连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

[0021] 结合第一方面,在第一方面的某些实现方式中,所述根据所述当前帧的初始ITD值,以及所述允许连续出现的目标帧的数量,确定所述当前帧的ITD值,包括:根据所述当前帧的初始ITD值,目标帧计数值,所述目标帧计数值的阈值,确定所述当前帧的ITD值,其中,所述目标帧计数值用于表征当前已连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

[0022] 结合第一方面,在第一方面的某些实现方式中,所述信噪比参数为所述多声道信号的修正的分段信噪比。

[0023] 第二方面,提供一种编码器,包括用于执行第一方面中的方法的单元。

[0024] 第三方面,提供一种编码器,包括存储器和处理器,所述存储器用于存储程序,所述处理器用于执行程序,当所述程序被执行时,所述处理器执行第一方面中的方法。

[0025] 第四方面,提供一种计算机可读介质,所述计算机可读介质存储用于编码器执行的程序代码,所述程序代码包括用于执行第一方面中的方法的指令。

[0026] 本申请能够降低背景噪声、混响、多说话人同时讲话等环境因素对ITD值计算结果的准确性以及稳定性的影响,在存在噪声、混响以及多说话人同时讲话或者信号谐波特征不明显的情况下,改善PS编码中的ITD值的稳定性,尽量减少ITD值的不必要的跳变,从而避免下混信号的帧间不连续以及解码信号的声像不稳定,同时,本发明实施例能够更好地保持立体声信号的相位信息,提升听觉质量。

附图说明

[0027] 为了更清楚地说明本发明实施例的技术方案,下面将对本发明实施例中所需要使

用的附图作简单地介绍,显而易见地,下面所描述的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0028] 图1是现有技术中的PS编码的流程图。

[0029] 图2是现有技术中的PS解码的流程图。

[0030] 图3是现有技术中的基于时域的ITD参数提取方法的示例性流程图。

[0031] 图4是现有技术中的基于频域的ITD参数提取方法的示例性流程图。

[0032] 图5是本发明实施例的多声道信号的编码方法的示意性流程图。

[0033] 图6是本发明实施例的多声道信号的编码方法的示意性流程图。

[0034] 图7是本发明实施例的编码器的示意性结构图。

[0035] 图8是本发明实施例的编码器的示意性结构图。

具体实施方式

[0036] 需要说明的是,立体声信号也可称为多声道信号。上文简单介绍了多声道信号的ILD、ITD以及IPD的作用和含义,为了便于理解,下文以第一个麦克拾取到的信号为第一声道信号,第二个麦克拾取到的信号为第二声道信号为例,对ILD、ITD以及IPD进行更为详细的说明。

[0037] ILD描述了第一声道信号和第二声道信号之间的能量差别。例如,如果ILD大于0,表示第一声道信号的能量高于第二声道信号的能量;如果ILD等于0,表示第一声道信号的能量等于第二声道信号的能量;如果ILD小于0,表示第一声道信号的能量小于第二声道信号的能量。又如,如果ILD小于0,表示第一声道信号的能量高于第二声道信号的能量;如果ILD等于0,表示第一声道信号的能量等于第二声道信号的能量;如果ILD大于0,表示第一声道信号的能量小于第二声道信号的能量。应理解,以上数值仅是举例,ILD的取值与第一声道信号和第二声道信号之间的能量差别的关系可以根据经验或实际需要定义。

[0038] ITD描述了第一声道信号和第二声道信号之间的时间差别,即声源产生的声音到达第一个麦克和第二个麦克的时间差异。例如,如果ITD大于0,表示声源产生的声音到达第一个麦克的时间早于声源产生的声音到达第二个麦克的时间;如果ITD等于0,表示声源产生的声音同时到达第一个麦克和第二个麦克;如果ITD小于0,表示声源产生的声音达到第一个麦克的时间晚于声源产生的声音到达第二个麦克的时间。又如,如果ITD小于0,表示声源产生的声音到达第一个麦克的时间早于声源产生的声音到达第二个麦克的时间;如果ITD等于0,表示声源产生的声音同时到达第一个麦克和第二个麦克;如果ITD大于0,表示声源产生的声音达到第一个麦克的时间晚于声源产生的声音到达第二个麦克的时间。应理解,以上数值仅是举例ITD的取值与第一声道信号和第二声道信号之间的时间差别的关系可以根据经验或实际需要定义。

[0039] IPD描述了第一声道信号和第二声道信号的相位差别,该参数通常和ITD结合在一起,用于解码端恢复多声道信号的相位信息。

[0040] 由上文可知,现有的ITD值计算方式会引起ITD值不连续的现象,为了便于理解,下文结合图3和图4,以多声道信号为左右声道信号为例,详细描述现有ITD值的计算方式及其缺点。

[0041] 在现有技术中,ITD值大多基于多声道信号的互相关系数进行计算,具体的计算方式可以有多种,例如,可以在时域进行ITD值的计算,也可以在频域进行ITD值的计算。

[0042] 图3是基于时域的ITD值计算方法的示例性流程图。图3的方法包括:

[0043] 310、基于左右声道时域信号计算ITD值。

[0044] 具体而言,可以基于左右声道时域信号,采用时域互相关函数计算ITD值,例如:在 $0 \leq i \leq T_{\max}$ 范围内,计算:

$$[0045] \quad c_n(i) = \sum_{j=0}^{Length-1-i} x_R(j) \cdot x_L(j+i) \quad (1)$$

$$[0046] \quad c_p(i) = \sum_{j=0}^{Length-1-i} x_L(j) \cdot x_R(j+i) \quad (2)$$

[0047] 如果 $\max_{0 \leq i \leq T_{\max}} (c_n(i)) > \max_{0 \leq i \leq T_{\max}} (c_p(i))$, 则 T_1 取 $\max(C_n(i))$ 对应的索引值的相反数; 否则 T_1 取 $\max(C_p(i))$ 对应的索引值; 其中, i 为计算互相关函数的索引值, x_L 为左声道时域信号, x_R 为右声道时域信号, T_{\max} 对应于不同采样率下ITD取值的最大值, $Length$ 为帧长。

[0048] 320、对ITD值进行量化处理。

[0049] 图4是基于频域的ITD值计算方法的示例性流程图。图4的方法包括:

[0050] 410、对左右声道时域信号进行时频变换,得到左右声道频域信号。

[0051] 具体而言,时频变换可以采用离散傅里叶变换(Discrete Fourier Transformation, DFT)、修正的离散余弦变换(Modified Discrete Cosine Transform, MDCT)等技术,将时域信号变换为频域信号。

[0052] 例如,对于输入的左右声道时域信号,可以采用如下公式(3)进行DFT变换。

$$[0053] \quad X(k) = \sum_{n=0}^{Length-1} x(n) \cdot e^{-j \frac{2\pi \cdot n \cdot k}{L}}, 0 \leq k < L \quad (3)$$

[0054] 其中, n 为时域信号的样点的索引值, k 为频域信号的频点的索引值, L 为时频变换长度。 $x(n)$ 为左声道时域信号或右声道时域信号。

[0055] 420、基于左右声道频域信号提取ITD值。

[0056] 具体地,可以将左右声道频域信号中的每个频域信号的 L 个频点(Frequency Bin)划分为 N 个子带,对于该 N 个子带中的第 b 个子带,其包含的频点的取值范围可以定义为 $A_{b-1} \leq k \leq A_b - 1$ 。在搜索范围 $-T_{\max} \leq j \leq T_{\max}$, 可以采用如下公式计算幅度值:

$$[0057] \quad mag(j) = \sum_{k=A_{b-1}}^{A_b-1} X_L(k) * X_R(k) * \exp\left(\frac{2\pi * k * j}{L}\right) \quad (4)$$

[0058] 则第 b 个子带的ITD值可以为 $T(k) = \arg \max_{-T_{\max} \leq j \leq T_{\max}} (mag(j))$, 即公式(4)计算出的最大值对应的样点的索引值。

[0059] 430、对ITD值进行量化处理。

[0060] 现有技术中,如果当前帧中的多声道信号的互相关系数峰值较小,计算出的ITD值被认为是不准确的,在这种情况下,当前帧的ITD值将被置零。

[0061] 受到背景噪声、混响、多人同时讲话等因素的影响,按照现有的PS编码方式计算出的ITD值会出现被频繁置零的情况,从而导致ITD值来回跳变,利用这样的ITD值计算出的下

混合信号会出现帧间不连续的现象,同时解码得到的多声道信号会出现声像不稳定的现象,从而导致多声道信号的听觉质量差。

[0062] 为了解决ITD值来回跳变的问题,一种可行的处理方式如下:当计算出的当前帧的ITD值被认为不准确时,当前帧可以复用当前帧的前一帧(某一帧的前一帧具体是指与该帧紧邻的前一帧)的ITD值,即将当前帧的前一帧的ITD值作为当前帧的ITD值。这种处理方式可以很好地解决ITD值来回跳变的问题,但是,这种处理方式可能会引起如下问题:多声道信号的信号质量较好时,许多当前帧也会不恰当地舍弃已经计算出的比较准确的ITD值,而去复用当前帧的前一帧的ITD值,从而引起多声道信号的相位信息的丢失。

[0063] 为了避免ITD值来回跳变的问题,同时更好地保留多声道信号的相位信息,下文结合图5,详细描述根据本发明实施例的多声道信号的编码方法。需要说明的是,为了便于描述,下文将ITD值复用前一帧的ITD值的帧称为目标帧。

[0064] 图5的方法包括:

[0065] 510、获取当前帧的多声道信号。

[0066] 520、确定当前帧的初始ITD值。

[0067] 例如,可以按照图3所示的基于时域的方式计算当前帧的初始ITD值。又如,可以按照图4所示的基于频域的方式计算当前帧的初始ITD值。

[0068] 530、根据多声道信号的特征信息,控制(或调整)允许连续出现的目标帧的数量,特征信息包括多声道信号的信噪比参数以及多声道信号的互相关系数的峰值特性中的至少一个,目标帧的ITD值复用了该目标帧的前一帧的ITD值。

[0069] 应理解,本发明实施例中,会先计算出当前帧的初始ITD值,然后基于当前帧的初始ITD值确定当前帧的ITD值(或称当前帧的实际ITD值,或称当前帧的最终ITD值)。当前帧的初始ITD值与当前帧的ITD值可以是同一ITD值,也可以是不同ITD值,这要视具体的计算规则而定。例如,在初始ITD值准确的情况下,可以将初始ITD值作为当前帧的ITD值;又如,在初始ITD值不准确的情况下,可以舍弃当前帧的初始ITD值,而将当前帧的前一帧的ITD值作为当前帧的ITD值。

[0070] 应理解,当前帧的多声道信号的互相关系数的峰值特性可以指当前帧的多声道信号的互相关系数的峰值(或称最大值)的幅度值(或称大小)与次大值的幅度值的差异特性,也可以指当前帧的多声道信号的互相关系数的峰值的幅度值与某个阈值的差异特性,也可以指当前帧的多声道信号的互相关系数的峰值位置索引对应的ITD值与前N帧的ITD值的差异特性,也可以指当前帧的多声道信号的互相关系数的峰值位置的索引与前N帧的多声道信号的互相关系数的峰值位置的索引的差异特性(或称波动特性),N为大于等于1的正整数,也可以是上述各种特性的组合。当前帧的多声道信号的互相关系数的峰值位置的索引可表征:在当前帧中,多声道信号的第几个互相关系数的取值为峰值。同理,前一帧的多声道信号的互相关系数的峰值位置的索引可表征:在前一帧中,多声道信号的第几个互相关系数的取值为峰值。例如,当前帧的多声道信号的互相关系数的峰值位置的索引为5表示:在当前帧中,多声道信号的第5个互相关系数的取值为峰值。又如,前一帧的多声道信号的互相关系数的峰值位置的索引为4表示:在前一帧中,多声道信号的第4个互相关系数的取值为峰值。

[0071] 步骤530中的控制允许连续出现的目标帧的数量可以通过设置目标帧计数值和/

或目标帧计数值的阈值实现。例如,可以通过强制改变目标帧计数值的方式达到控制允许连续出现的目标帧的数量的目的,也可以通过强制改变目标帧计数值的阈值的方式达到控制允许连续出现的目标帧的数量的目的,当然也可以通过既强制改变目标帧计数值的方式,也强制改变目标帧计数值的阈值的方式来达到控制允许连续出现的目标帧的数量的目的。其中,目标帧计数值可用于指示当前已连续出现的目标帧的数量,目标帧计数值的阈值可用于指示允许连续出现的目标帧的数量。

[0072] 540、根据当前帧的初始ITD值,以及允许连续出现的目标帧的数量,确定当前帧的ITD值。

[0073] 550、根据当前帧的ITD值,对多声道信号进行编码。

[0074] 例如,可以执行图1所示的单声道音频编码、空间参数编码、比特流复用等操作,具体编码方式可以参照现有技术。

[0075] 本发明实施例能够降低背景噪声、混响、多说话人同时讲话等环境因素对ITD值计算结果的准确性以及稳定性的影响,在存在噪声、混响以及多说话人同时讲话或者信号谐波特征不明显的情况下,改善PS编码中的ITD值的稳定性,尽量减少ITD值的不必要的跳变,从而避免下混信号的帧间不连续以及解码信号的声像不稳定,同时,本发明实施例能够更好地保持立体声信号的相位信息,提升听觉质量。

[0076] 需要说明的是,除非特别指明多声道信号是前一帧或前N帧的多声道信号,下文中出现多声道信号均指当前帧的多声道信号。

[0077] 在步骤530之前,图5的方法还可包括:根据多声道信号的互相关系数的峰值的幅度,确定多声道信号的互相关系数的峰值特性。

[0078] 具体地,可以根据多声道信号的互相关系数的峰值的幅度,确定峰值幅度可信度参数,峰值幅度可信度参数可用于表征多声道信号的互相关系数的峰值幅度的可信度。进一步地,步骤530可包括:在峰值幅度可信度参数满足预设条件的情况下,减少允许连续出现的目标帧的数量;在峰值幅度可信度参数不满足预设条件的情况下,允许连续出现的目标帧的数量保持不变。峰值幅度可信度参数满足预设条件例如可以是峰值幅度可信度参数的取值大于某个阈值,也可以是峰值幅度可信度参数的取值在预设范围内。

[0079] 本发明实施例中,峰值幅度可信度参数的定义方式可以有多种。

[0080] 例如,峰值幅度可信度参数可以是:多声道信号的互相关系数的峰值的幅度值与次大值的幅度值之间的差值。具体地,差值越大,则说明峰值幅度的可信度越高。

[0081] 又如,峰值幅度可信度参数可以是:多声道信号的互相关系数的峰值的幅度值与次大值的幅度值之间的差值与该峰值的幅度值的比值。具体地,比值越大,则说明峰值幅度的可信度越高。

[0082] 又如,峰值幅度可信度参数可以是:多声道信号的互相关系数的峰值的幅度值与目标幅度值之间的差值。具体地,差值的绝对值越大,则说明峰值幅度的可信度越高。其中,该目标幅度值可以根据经验或实际情况选取,例如,可以是固定值,也可以当前帧的某个预设位置(位置可以通过互相关系数的索引表示)的互相关系数的幅度值。

[0083] 又如,峰值幅度可信度参数可以是:多声道信号的互相关系数的峰值的幅度值与目标幅度值之间的差值与该峰值的幅度值之间的比值。具体地,比值越大,则说明峰值幅度的可信度越高。该目标幅度值可以根据经验或实际情况选取,例如,可以是固定值,也可以

当前帧的某个预设位置的互相关系数的幅度值。

[0084] 可选地,在一些实施例中,在步骤530之前,图5的方法还可包括:根据多声道信号的互相关系数的峰值位置的索引,确定当前帧的多声道信号的互相关系数的峰值特性。

[0085] 例如,可以根据多声道信号的互相关系数的峰值位置的索引对应的ITD值,以及当前帧的前N帧的ITD值,确定峰值位置波动性参数,峰值位置波动性参数可用于表征多声道信号的互相关系数的峰值位置的索引对应的ITD值与当前帧的前一帧的ITD值之间的差异。N为大于等于1的正整数。

[0086] 又如,可以根据多声道信号的互相关系数的峰值位置的索引,以及当前帧的前N帧的多声道信号的互相关系数的峰值位置的索引,确定峰值位置波动性参数,峰值位置波动性参数可用于表征多声道信号的互相关系数的峰值位置的索引与当前帧的前N帧的多声道信号的互相关系数的峰值位置的索引的差异。

[0087] 进一步地,步骤530可包括:在峰值位置波动性参数满足预设条件的情况下,可以减少允许连续出现的目标帧的数量;在峰值位置波动性参数不满足预设条件的情况下,允许连续出现的目标帧的数量保持不变。峰值位置波动性参数满足预设条件例如可以是峰值位置波动性参数的取值大于某个阈值,也可以是峰值位置波动性参数的取值在预设范围内。例如,峰值位置波动性参数是根据多声道信号的互相关系数的峰值位置索引对应的ITD值以及当前帧的前一帧的ITD值确定时,峰值位置波动性参数满足预设条件例如可以是峰值位置波动性参数的取值大于某个阈值,该阈值可以设置为4,5,6或其他经验值,也可以是峰值位置波动性参数的取值在预设范围内,该预设范围可以设置为[6,128]或其他经验值。具体的阈值/取值范围可以根据不同的参数计算方法,不同的需要,不同的应用场景等进行设置。

[0088] 本发明实施例中,峰值位置波动性参数的定义方式可以有多种。

[0089] 例如,峰值位置波动性参数可以是:当前帧的多声道信号的互相关系数的峰值位置索引对应的ITD值与当前帧的前一帧的多声道信号的互相关系数的峰值位置索引对应的ITD值之差的绝对值。

[0090] 又如,峰值位置波动性参数可以是:当前帧的多声道信号的互相关系数的峰值位置索引对应的ITD值与当前帧的前一帧的ITD值之差的绝对值。

[0091] 又如,峰值位置波动性参数可以是:当前帧的多声道信号的互相关系数的峰值位置索引对应的ITD值与前N帧的ITD值之差的方差,N为大于或等于2的整数。

[0092] 可选地,在一些实施例中,在步骤530之前,图5的方法还可包括:根据多声道信号的互相关系数的峰值的幅度和多声道信号的互相关系数的峰值位置的索引,确定多声道信号的互相关系数的峰值特性。

[0093] 具体地,可以根据多声道信号的互相关系数的峰值的幅度,确定峰值幅度可信度参数;并根据多声道信号的互相关系数的峰值位置的索引对应的ITD值,以及前一帧的ITD值,确定峰值位置波动性参数;根据峰值幅度可信度参数和峰值位置波动性参数,确定多声道信号的互相关系数的峰值特性。峰值幅度可信度参数和峰值位置波动性参数的定义方式可以参照上述实施例,此处不再详述。

[0094] 进一步地,在该实施例中,步骤530可包括:在峰值幅度可信度参数和峰值位置波动性参数均满足预设条件的情况下,控制允许连续出现的目标帧的数量。

[0095] 例如,在峰值幅度可信度参数大于预设的峰值幅度可信度阈值,峰值位置波动性参数大于预设的峰值位置波动性阈值,则减少允许连续出现的目标帧的数量。具体地,例如,峰值幅度可信度参数为多声道信号的互相关系数的峰值的幅度值与次大值的幅度值之间的差值与该峰值的幅度值的比值时,峰值幅度可信度阈值可以设置为0.1,0.2,0.3或其他经验值。峰值位置波动性参数为当前帧中的多声道信号的互相关系数的峰值位置索引对应的ITD值与当前帧的前一帧的多声道信号的互相关系数的峰值位置索引对应的ITD值之差的绝对值时,峰值位置波动性阈值可以设置为4,5,6或其他经验值。具体的阈值/取值范围可以根据不同的参数计算方法,不同的需要,不同的应用场景等进行设置。

[0096] 又如,在峰值幅度可信度参数的取值位于两个阈值之间,且峰值位置波动性参数大于预设的峰值位置波动性阈值,则减少允许连续出现的目标帧的数量。

[0097] 又如,在峰值幅度可信度参数的取值大于预设的峰值幅度可信度阈值,且峰值位置波动性参数位于两个阈值之间,则减少允许连续出现的目标帧的数量。

[0098] 需要说明的是,在某些实施例中,可以将上文描述的峰值幅度可信度参数和/或峰值位置波动性参数称为表征多声道信号的互相关系数的峰值位置的稳定程度的参数。此时,步骤530可包括:在多声道信号的互相关系数的峰值位置的稳定程度满足预设条件的情况下,减少允许连续出现的目标帧的数量。

[0099] 需要说明的是,本发明实施例对表征多声道信号的互相关系数的峰值位置的稳定程度的参数满足预设条件的定义方式不作具体限定。

[0100] 可选地,多声道信号的互相关系数的峰值位置的稳定程度满足预设条件可以是指:表征多声道信号的互相关系数的峰值位置的稳定程度的参数中的一个或多个参数的取值位于预先设定的取值范围之内,或者,表征多声道信号的互相关系数的峰值位置的稳定程度的参数中的一个或多个参数的取值位于预先设定的取值范围之外。例如,多声道信号的互相关系数的峰值位置的稳定程度为峰值位置波动性参数,且峰值位置波动性参数的计算方法为当前帧中的多声道信号的互相关系数的峰值位置索引对应的ITD值与当前帧的前一帧的多声道信号的互相关系数的峰值位置索引对应的ITD值之差的绝对值时,预先设定的取值范围可以设置为峰值位置波动性参数大于5或其他经验值。又如,多声道信号的互相关系数的峰值位置的稳定程度为峰值位置波动性参数和峰值幅度可信度参数,且峰值位置波动性参数的计算方法为当前帧中的多声道信号的互相关系数的峰值位置索引对应的ITD值与当前帧的前一帧的多声道信号的互相关系数的峰值位置索引对应的ITD值之差的绝对值,峰值幅度可信度参数为多声道信号的互相关系数的峰值的幅度值与次大值的幅度值之间的差值与该峰值的幅度值的比值时,预先设定的取值范围可以设置为峰值位置波动性参数大于5且峰值幅度可信度参数大于0.2或其他经验取值范围。具体的取值范围可以根据不同的参数计算方法,不同的需要,不同的应用场景等进行设置。

[0101] 下文详细描述如何根据多声道信号的信噪比参数,控制允许连续出现的目标帧的数量。

[0102] 上述多声道信号的信噪比参数可用于表征多声道信号的信噪比。

[0103] 应理解,多声道信号的信噪比参数可以由一个或多个参数表示,本发明实施例对参数的具体选取方式不作限定。例如,多声道信号的信噪比参数可以用子带信噪比、修正的子带信噪比、分段信噪比、修正的分段信噪比、全带信噪比、修正的全带信噪比以及可以表

征多声道信号的信噪比特性的其他参数中的至少一种来表示。

[0104] 还应理解,本发明实施例对多声道信号的信噪比参数的确定方式不作具体限定。例如,可以采用多声道信号整体计算该多声道信号的信噪比参数。又如,可以采用多声道信号中的部分信号计算该多声道信号的信噪比参数,即利用部分信号的信噪比表征该多声道信号的信噪比。又如,可以自适应选择多声道信号中的任意一个声道的信号进行计算,即利用该一个声道的信号的信噪比表征该多声道信号的信噪比。又如,可以先对表征多声道信号的数据进行加权平均,形成新的信号,然后利用新的信号的信噪比表征该多声道信号的信噪比。

[0105] 下面以多声道信号包括左右声道信号为例,对多声道信号的信噪比的计算方式进行举例说明。

[0106] 例如,可以先对左右声道时域信号进行时频变换,得到左右声道频域信号;然后,将左声道频域信号的幅度谱与右声道频域信号的幅度谱进行加权平均,得到左右声道频域信号的平均幅度谱;然后,根据该平均幅度谱计算修正的分段信噪比,作为表征多声道信号的信噪比特性的参数。

[0107] 又如,可以先对左声道时域信号进行时频变换,得到左声道频域信号;然后,根据左声道频域信号的幅度谱计算左声道频域信号的修正的分段信噪比。同样地,对右声道时域信号进行时频变换,得到右声道频域信号;根据右声道时域信号的幅度谱计算右声道信号的修正的分段信噪比。然后根据左声道频域信号的修正的分段信噪比和右声道频域信号的修正的分段信噪比,计算左右声道频域信号的修正的分段信噪比的平均值,作为表征多声道信号的信噪比特性的参数。

[0108] 上述根据多声道信号的信噪比参数,控制允许连续出现的目标帧的数量,可包括:在多声道信号的信噪比参数满足预设条件的情况下,减少允许连续出现的目标帧的数量;在多声道信号的信噪比参数不满足预设条件的情况下,允许连续出现的目标帧的数量保持不变。例如,在多声道信号的信噪比参数的取值大于预设阈值的情况下,减少允许连续出现的目标帧的数量;又如,在多声道信号的信噪比参数的取值位于预先设定的取值范围之内,的情况下,减少允许连续出现的目标帧的数量;又如,在多声道信号的信噪比参数的取值位于预先设定的取值范围之外的情况下,减少允许连续出现的目标帧的数量。例如,当多声道信号的信噪比参数为分段信噪比时,预设阈值可以是6000或其他经验值,预先设定的取值范围可以是大于6000且小于3000000或其他经验取值范围。具体的阈值/取值范围可以根据不同的参数计算方法,不同的需要,不同的应用场景等进行设置。

[0109] 上文主要描述的是如何根据多声道信号的互相关系数的峰值特性或多声道信号的信噪比参数,控制允许连续出现的目标帧的数量。下文详细描述如何根据多声道信号的信噪比参数和多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量。

[0110] 具体地,可以在多声道信号的信噪比参数满足预设条件,且多声道信号的互相关系数的峰值幅度可信度参数和/或峰值位置波动性参数也满足预设条件的情况下,减少当前允许连续出现的目标帧的数量。

[0111] 例如,在多声道信号的信噪比参数的取值大于第一阈值并且小于等于第二阈值、峰值幅度可信度参数大于第三阈值、峰值位置波动性参数大于第四阈值的情况下,则减少允许连续出现的目标帧的数量。例如,当多声道信号的信噪比参数为分段信噪比时,第一阈

值可以是5000,6000,7000或其他经验值,第二阈值可以是2900000,3000000,3100000或其他经验值。当峰值幅度可信度参数为多声道信号的互相关系数的峰值的幅度值与次大值的幅度值之间的差值与该峰值的幅度值的比值时,第三阈值可以设置为0.1,0.2,0.3或其他经验值。当峰值位置波动性参数为当前帧中的多声道信号的互相关系数的峰值位置索引对应的ITD值与当前帧的前一帧的多声道信号的互相关系数的峰值位置索引对应的ITD值之差的绝对值时,第四阈值可以设置为4,5,6或其他经验值。具体的阈值可以根据不同的参数计算方法,不同的需要,不同的应用场景等进行设置。

[0112] 又如,在多声道信号的信噪比参数的取值大于等于第一阈值并且小于等于第二阈值,且峰值幅度可信度参数小于第五阈值的情况下,则减少允许连续出现的目标帧的数量。例如,当多声道信号的信噪比参数为分段信噪比时,第一阈值可以是5000,6000,7000或其他经验值,第二阈值可以是2900000,3000000,3100000或其他经验值。当峰值幅度可信度参数为多声道信号的互相关系数的峰值的幅度值与次大值的幅度值之间的差值与该峰值的幅度值的比值时,第五阈值可以设置为0.3,0.4,0.5或其他经验值。具体的阈值可以根据不同的参数计算方法,不同的需要,不同的应用场景等进行设置。

[0113] 应理解,减少允许连续出现的目标帧的数量的方式有很多,在一些实施例中,可以预先配置用于指示允许连续出现的目标帧的数量的数值,通过减少该数值可以达到减少允许连续出现的目标帧的数量的目的。

[0114] 在另一些实施例中,可以预先配置目标帧计数值和目标帧计数值的阈值,目标帧计数值可用于指示当前已连续出现的目标帧的数量,目标帧计数值的阈值可用于指示允许连续出现的目标帧的数量。具体地,通过调整目标帧计数值和所述目标帧计数值的阈值中的至少一个,减少允许连续出现的目标帧的数量。例如,可以通过增加(或称强制增加)目标帧计数值,减少允许连续出现的目标帧的数量;又如,可以通过减小目标帧计数值的阈值,减少允许连续出现的目标帧的数量;又如,可以通过增加目标帧计数值并减少目标帧计数值的阈值,减少允许连续出现的目标帧的数量。

[0115] 上文描述了根据多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量的方式。在某些实施例中,在根据多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量之前,可以先判断多声道信号的信噪比参数是否满足预设的信噪比条件。

[0116] 如果多声道信号的信噪比参数不满足预设的信噪比条件,根据多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量;如果多声道信号的信噪比满足信噪比条件,可以直接停止复用当前帧的前一帧的ITD值作为当前帧的ITD值。

[0117] 或者,如果多声道信号的信噪比参数满足预设的信噪比条件,根据多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量;如果多声道信号的信噪比不满足信噪比条件,可以直接停止复用当前帧的前一帧的ITD值作为当前帧的ITD值。

[0118] 下面对多声道信号的信噪比是否满足信噪比条件的判断方式,以及如何停止复用当前帧的前一帧的ITD值作为当前帧的ITD值进行详细描述。

[0119] 首先,多声道信号的信噪比参数可以由一个或多个参数表示,本发明实施例对参数的具体选取方式不作限定。例如,多声道信号的信噪比参数可以用子带信噪比、修正的子带信噪比、分段信噪比、修正的分段信噪比、全带信噪比、修正的全带信噪比以及可以表征

多声道信号的信噪比特性的其他参数中的至少一种来表示。

[0120] 其次,本发明实施例对多声道信号的信噪比参数的确定方式不作具体限定。例如,可以采用多声道信号整体计算该多声道信号的信噪比参数。又如,可以采用多声道信号中的部分信号计算该多声道信号的信噪比参数,即利用部分信号的信噪比表征该多声道信号的信噪比。又如,可以自适应选择多声道信号中的任意一个声道的信号进行计算,即利用该一个声道的信号的信噪比表征该多声道信号的信噪比。又如,可以先对表征多声道信号的数据进行加权平均,形成新的信号,然后利用新的信号的信噪比表征该多声道信号的信噪比。

[0121] 下面以多声道信号包括左右声道信号为例,对多声道信号的信噪比的计算方式进行举例说明。

[0122] 例如,可以先对左右声道时域信号进行时频变换,得到左右声道频域信号;然后,将左声道频域信号的幅度谱与右声道频域信号的幅度谱进行加权平均,得到左右声道频域信号的平均幅度谱;然后,根据该平均幅度谱计算修正的分段信噪比,作为表征多声道信号的信噪比特性的参数。

[0123] 又如,可以先对左声道时域信号进行时频变换,得到左声道频域信号;然后,根据左声道频域信号的幅度谱计算左声道频域信号的修正的分段信噪比。同样地,对右声道时域信号进行时频变换,得到右声道频域信号;根据右声道频域信号的幅度谱计算右声道频域信号的修正的分段信噪比。然后根据左声道频域信号的修正的分段信噪比和右声道频域信号的修正的分段信噪比,计算左右声道频域信号的修正的分段信噪比的平均值,作为表征多声道信号的信噪比特性的参数。

[0124] 在多声道信号的信噪比满足信噪比条件的情况下,停止复用当前帧的前一帧的ITD值作为当前帧的ITD值,可包括:在多声道信号的信噪比参数的取值大于预设阈值的情况下,停止复用当前帧的前一帧的ITD值作为当前帧的ITD值;又如,在多声道信号的信噪比参数的取值位于预设的取值范围之内的情况下,停止复用当前帧的前一帧的ITD值作为当前帧的ITD值;又如,在多声道信号的信噪比参数的取值位于预设的取值范围之外的情况下,停止复用当前帧的前一帧的ITD值作为当前帧的ITD值。

[0125] 进一步地,在一些实施例中,停止复用当前帧的前一帧的ITD值,可包括:增加(或称强制增加)目标帧计数值,使得目标帧计数值的取值大于或等于目标帧计数值的阈值。在另一些实施例中,停止复用当前帧的前一帧的ITD值作为当前帧的ITD值,可包括:设置停止标志位,使得该停止标志位的某些取值表征停止复用当前帧的前一帧的ITD值作为当前帧的ITD值,例如,如果将停止标志位置1,表示停止复用当前帧的前一帧的ITD值作为当前帧的ITD值;如果将停止标志位置0,表示允许复用当前帧的前一帧的ITD值作为当前帧的ITD值。

[0126] 下面结合具体的例子,对停止复用当前帧的前一帧的ITD值作为当前帧的ITD值的方式进行详细描述。

[0127] 例如,当多声道信号的信噪比参数的取值小于某个阈值,则强制修改目标帧计数值的取值,使其大于或等于该目标帧计数值的阈值。

[0128] 又如,当多声道信号的信噪比参数的取值大于某个阈值,则强制修改目标帧计数值的取值,使其大于或等于该目标帧计数值的阈值。

[0129] 又如,无论多声道信号的信噪比参数的取值小于某个阈值还是大于另一阈值,均强制修改目标帧计数值的取值,使其大于或等于该目标帧计数值的阈值。

[0130] 又如,当多声道信号的信噪比参数的取值小于某个阈值或者大于另一阈值,则将停止标志位置1。

[0131] 需要说明的是,步骤540中描述的当前帧的ITD值的确定方式可以有多种,本发明实施例对此不作具体限定。

[0132] 可选地,在一些实施例中,可以综合考虑当前帧的初始ITD值的准确性、允许连续出现的目标帧的数量(允许连续出现的目标帧的数量可以是基于步骤530进行控制或调整之后得到的数量)等因素确定当前帧的ITD值。

[0133] 可选地,在另一些实施例中,可以综合考虑当前帧的初始ITD值的准确性、允许连续出现的目标帧的数量(允许连续出现的目标帧的数量可以是基于步骤530进行调制之后得到的数量)以及当前帧是否为连续话音帧等因素确定当前帧的ITD值。例如,如果当前帧的初始ITD值的可信度高,可以直接将当前帧的初始ITD值作为当前帧的ITD值。又如,当前帧的初始ITD值的可信度低,且当前帧满足复用当前帧的前一帧的ITD值的条件,则当前帧可以复用当前帧的前一帧的ITD值。

[0134] 应理解,计算当前帧的初始ITD值的可信度的方式可以有多种,本发明实施例对此不作具体限定。

[0135] 例如,如果多声道信号的互相关系数中的与初始ITD值对应的互相关系数的取值大于预先设定的阈值,则可以认为该初始ITD值的可信度高。

[0136] 又如,如果多声道信号的互相关系数中的与初始ITD值对应的互相关系数的取值与多声道信号的互相关系数中的次大值之差大于预先设定的阈值,则可以认为该初始ITD值的可信度高。

[0137] 又如,如果多声道信号的互相关系数的峰值的幅度值大于预设阈值,则可以认为该初始ITD值的可信度高。

[0138] 应理解,判断当前帧是否满足复用当前帧的前一帧的ITD值的条件的方式可以有多种。

[0139] 可选地,在一些实施例中,当前帧满足复用当前帧的前一帧的ITD值的条件可以是:目标帧计数值小于该目标帧计数值的阈值。

[0140] 可选地,在一些实施例中,当前帧满足复用当前帧的前一帧的ITD值的条件可以是:当前帧的语音激活检测结果表明当前帧与当前帧的前N(N为大于1的正整数)帧形成了连续话音帧,在这种情况下,如果当前帧的前一帧的ITD值不等于第一预设值(如果某一帧的ITD值为第一预设值,可以认为计算出的该帧的ITD值由于不准确而被强制置为该第一预设值,该第一预设值例如可以是0),且当前帧的ITD值等于该第一预设值,且目标帧计数值小于该目标帧计数值的阈值。例如,当前帧的语音激活检测结果与当前帧的前N(N为大于1的正整数)帧的语音激活检测结果均为话音帧,若当前帧的前一帧的ITD值不等于零,当前帧的ITD值被强制置为零,且目标帧计数值小于该目标帧计数值的阈值,则可以将当前帧的前一帧的ITD值作为当前帧的ITD值,并增加目标帧计数值的取值。需要说明的是,当前帧的ITD值被强制置为零的方式有多种,例如,可以更改当前帧的ITD值的取值,使其变为零;或者,可以设置一个标志位,表征当前帧的ITD值已被强制置为零;或者,可以是上述两种方式

的结合。

[0141] 下面结合具体例子,更加详细地描述本发明实施例。应注意,图6的例子仅仅是为了帮助本领域技术人员理解本发明实施例,而非要将本发明实施例限于所例示的具体数值或具体场景。本领域技术人员根据所给出的图6的例子,显然可以进行各种等价的修改或变化,这样的修改或变化也落入本发明实施例的范围内。

[0142] 图6是本发明实施例的多声道信号的编码方法的示意性流程图。应理解,图6示出的处理步骤或操作仅是示例,本发明实施例还可以执行其它操作或者图6中的各种操作的变形。此外,图6中的各个步骤可以按照与图6呈现的不同的顺序来执行,并且有可能并非要执行图6中的全部操作。图6是以多声道信号包括左声道信号和右声道信号为例进行说明的。还应理解,图6实施例中的表征多声道信号的互相关系数的峰值位置的稳定程度的参数可以是上文中的峰值幅度可信度参数和/或峰值位置波动性参数。

[0143] 图6的方法包括:

[0144] 602、对左声道时域信号和右声道时域信号进行时频变换。

[0145] 具体地,当前帧的第 m 子帧的左声道时域信号可以通过 $x_{m,\text{left}}(n)$ 表示,该第 m 子帧的右声道时域信号可以通过 $x_{m,\text{right}}(n)$ 表示,其中, $m=0,1,\dots,\text{SUBFR_NUM}-1$, SUBFR_NUM 为一个音频帧所包含的子帧的个数, n 为样点的索引值, $n=0,1,\dots,N-1$, N 为第 m 个子帧的左声道时域信号或右声道时域信号包含的样点的数量。以多声道信号的采样率为16KHz,一个音频帧的长度为20ms为例,一个音频帧的左声道时域信号和右声道时域信号分别包括320个采样点,如果一个音频帧被分为两个子帧,每个子帧的左声道时域信号和右声道时域信号分别包括160个采样点,此时, $N=160$ 。

[0146] 分别对 $x_{m,\text{left}}(n)$ 和 $x_{m,\text{right}}(n)$ 进行 L 点快速傅里叶变换,得到第 m 子帧的左声道频域信号 $X_{m,\text{left}}(k)$ 以及第 m 子帧的右声道频域信号 $X_{m,\text{right}}(k)$,其中 $k=0,1,\dots,L-1$, L 为快速傅里叶变换长度,例如, L 可以取400,800等。

[0147] 604-605、根据左声道频域信号和右声道频域信号,计算修正的分段信噪比,并基于修正的分段信噪比进行语言激活检测。

[0148] 具体地,根据 $X_{m,\text{left}}(k)$ 和 $X_{m,\text{right}}(k)$ 计算修正的分段信噪比的方式有多种,下面给出一种具体的计算方式。

[0149] 步骤一、根据 $X_{m,\text{left}}(k)$ 和 $X_{m,\text{right}}(k)$,计算第 m 子帧的左右声道频域信号的平均幅度谱 $\text{SPD}_m(k)$ 。

[0150] 例如,可以根据公式(5)计算 $\text{SPD}_m(k)$:

$$[0151] \quad \text{SPD}_m(k) = A * \text{SPD}_{m,\text{left}}(k) + (1-A) \text{SPD}_{m,\text{right}}(k) \quad (5)$$

[0152] 其中:

$$[0153] \quad \text{SPD}_{m,\text{left}}(k) = (\text{real}\{X_{m,\text{left}}(k)\})^2 + (\text{imag}\{X_{m,\text{left}}(k)\})^2,$$

$$[0154] \quad \text{SPD}_{m,\text{right}}(k) = (\text{real}\{X_{m,\text{right}}(k)\})^2 + (\text{imag}\{X_{m,\text{right}}(k)\})^2,$$

[0155] 其中, $k=1,\dots,L/2-1$, A 为预先设定的左右声道幅度谱混合比例因子, A 一般可以取0.5,0.4,0.3或取其他经验值。

[0156] 步骤二、根据第 m 子帧的左右声道频域信号的平均幅度谱 $\text{SPD}_m(k)$,计算子带能量 $E_{\text{band}_m}(i)$,其中, $i=0,1,\dots,\text{BAND_NUM}-1$, BAND_NUM 为子带个数。

[0157] 例如,可以通过公式(6)计算 $E_{\text{band}}(i)$:

$$[0158] \quad E_band(i) = \frac{1}{band_rb[i+1] - band_rb[i]} \sum_{k=band_rb[i]}^{band_rb[i+1]-1} SPD_m(k) \quad (6)$$

[0159] 其中band_tb为预先设定的用于子带划分的表格,band_tb[i]为第i个子带下限频点,band_tb[i+1]-1为第i个子带上限频点。

[0160] 步骤三、根据子带能量E_band(i)以及子带噪声能量估计E_band_n(i),计算修正的分段信噪比mssnr。

[0161] 例如,可以通过公式(7)和公式(8)计算mssnr:

$$[0162] \quad msnr(i) = \max\left(0, \frac{E_band(i)}{E_band_n(i)} - 1\right) \quad (7)$$

[0163] 如果msnr(i) < G, 则msnr(i) = msnr(i)²/G

$$[0164] \quad mssnr = \sum_{i=0}^{BAND_NUM-1} msnr(i) \quad (8)$$

[0165] 其中,msnr(i)为修正的子带信噪比,G为预先设定的子带信噪比修正门限,一般G可以取5,6,7或其他经验值。应理解,计算修正的分段信噪比的方法有多种,这里仅是一个示例。

[0166] 步骤四、根据修正的分段信噪比以及子带能量E_band(i)对子带噪声能量估计E_band_n(i)进行更新。

[0167] 具体地,可以先根据公式(9)计算子带平均能量energy。

$$[0168] \quad energy = \frac{1}{BAND_NUM} \sum_{i=0}^{BAND_NUM-1} E_band(i) \quad (9)$$

[0169] 如果VAD计数vad_fm_cnt小于预先设定的噪声初始设定帧长度,则可以增加VAD计数。预先设定的噪声初始设定长度,一般为预先设定的经验值,例如可以取29,30,31或其他经验值。

[0170] 如果VAD计数vad_fm_cnt小于预先设定的噪声初始设定帧长度并且子带平均能量小于噪声能量阈值ener_th,则可以对子带噪声能量E_band_n(i)进行更新,并将噪声能量更新标志设置为1。噪声能量阈值一般为预先设定的经验值,例如可以取35000000,40000000,45000000或其他经验值。

[0171] 具体地,可以采用公式(10)对子带噪声能量进行更新:

$$[0172] \quad E_band_n(i) = \frac{E_band_n_{n-1}(i) * vad_fm_cnt + E_band(i)}{vad_fm_cnt + 1} \quad (10)$$

[0173] 其中E_band_n_{n-1}(i)为历史子带噪声能量,例如,可以是更新前的子带噪声能量。

[0174] 否则,如果修正的分段信噪比小于噪声更新门限th_UPDATE,仍然可以对子带噪声能量E_band_n(i)进行更新,并将噪声能量更新标志设置为1。噪声更新门限th_UPDATE可以为4,5,6或其他经验值。

[0175] 具体地,可以通过公式(11)对子带噪声能量进行更新:

$$[0176] \quad E_band_n(i) = (1 - update_fac) E_band_n_{n-1}(i) + update_fac * E_band(i) \quad (11)$$

[0177] 其中,update_fac为设定的噪声更新速率,可以是0-1之间的常数,例如,可以取0.03,0.04,0.05或其他经验值。E_band_n_{n-1}(i)为历史子带噪声能量,例如,可以是更新前

的子带噪声能量。

[0178] 此外,为了保证子带信噪比计算的有效性,可以对更新后的子带噪声能量的取值进行限制,例如,可以将 $E_{band_n}(i)$ 的最小值限定1。

[0179] 需要说明的是,根据修正的分段信噪比以及 $E_{band}(i)$ 对 $E_{band_n}(i)$ 进行更新的方法有很多种,本发明实施例对此不作具体限定,这里仅是一个示例。

[0180] 接下来,可以根据修正的分段信噪比进行第 m 子帧的语音激活检测。具体地,如果修正的分段信噪比大于语音激活检测阈值 th_{VAD} ,则第 m 子帧为语音帧,此时,第 m 子帧的语音激活检测标志 $vad_flag[m]$ 设置为1,否则第 m 子帧为背景噪音帧,此时,第 m 子帧的语音激活检测标志 $vad_flag[m]$ 可以设置为0。语音激活检测阈值 th_{VAD} 可以取3500,4000,4500或其他经验值。

[0181] 606-608、根据左声道频域信号和右声道频域信号,计算左右声道频域信号的互相关系数,并基于左右声道频域信号的互相关系数,计算当前帧的初始ITD值。

[0182] 根据 $X_{m,left}(k)$ 和 $X_{m,right}(k)$ 计算左右声道频域信号的互相关系数 $Xcorr(t)$ 的方式可以有多种,下面给出一种具体的实现方式。

[0183] 首先,根据公式(12),计算第 m 子帧中的左右声道频域信号的互相关功率谱 $Xcorr_m(k)$ 。

$$[0184] \quad Xcorr_m(k) = X_{m,left}(k) * X_{m,right}^*(k) \quad (12)$$

[0185] 然后,根据公式(13),对左右声道频域信号的互相关功率谱进行平滑处理,得到平滑后的互相关功率谱 $Xcorr_smooth(k)$ ：

$$[0186] \quad Xcorr_smooth(k) = smooth_fac * Xcorr_smooth(k) + (1 - smooth_fac) * Xcorr_m(k) \quad (13)$$

[0187] 其中 $smooth_fac$ 为平滑因子,该平滑因子可以取0-1中的任意正数,例如,可以取0.4,0.5,0.6或其他经验值。

[0188] 接着,可以根据 $Xcorr_smooth(k)$,通过公式(14),计算 $Xcorr(t)$ 。

$$[0189] \quad Xcorr(t) = IDFT\left(\frac{Xcorr_smooth(k)}{|Xcorr_smooth(k)|}\right), \quad (14)$$

[0190] 其中, $IDFT(*)$ 表示傅里叶变换的逆变换,参与计算的ITD值的取值范围可以选取为 $[-ITD_MAX, ITD_MAX]$,根据ITD值的取值范围对 $Xcorr(t)$ 进行截取重排后得到用于确定当前帧的初始ITD值的左右声道频域信号的互相关系数 $Xcorr_itd(t)$,此时, $t=0, \dots, 2*ITD_MAX$ 。

[0191] 然后,可以根据 $Xcorr_itd(t)$,通过公式(15),估计当前帧的初始ITD值。

$$[0192] \quad ITD = \operatorname{argmax}(Xcorr_itd(t)) - ITD_MAX \quad (15)$$

[0193] 610-612、判断当前帧的初始ITD值的可信度,若初始ITD值的可信度高,则可以将目标帧计数值设为预设的初始值。

[0194] 具体地,可以先对当前帧的初始ITD值的可信度进行判断,具体的判断方式可以有多种,下面进行举例说明。

[0195] 例如,可以将左右声道频域信号的互相关系数中的与初始ITD值对应的互相关系数的幅度值与预先设定的门限值进行比较。若该幅度值大于预先设定的门限值,则可以认

为当前帧的初始ITD值的可信度高。

[0196] 又如,可以先按照幅度值从大到小,将左右声道频域信号的互相关系数进行排列;然后从排列后的互相关系数中选取位于预设位置(位置可以通过互相关系数的索引值表示)的目标互相关系数;接着,将左右声道频域信号的互相关系数中的与初始ITD值对应的互相关系数的幅度值与该目标互相关系数的幅度值进行比较:如果二者的差值大于预先设定的门限值,则可以认为当前帧的初始ITD值的可信度高,或者,如果二者的比值大于预先设定的门限值,则可以认为当前帧的初始ITD值的可信度高,或者,如果左右声道频域信号的互相关系数中的与初始ITD值对应的互相关系数的幅度值大于目标互相关系数的幅度值,则可以认为当前帧的初始ITD值的可信度高。

[0197] 此外,还可以在得到目标互相关系数后,先对目标互相关系数进行修正,接着,将左右声道频域信号的互相关系数中的与初始ITD值对应的互相关系数的幅度值与修正后的目标互相关系数的幅度值进行比较:如果左右声道频域信号的互相关系数中的与初始ITD值对应的互相关系数的幅度值大于修正后的目标互相关系数的幅度值,则可以认为当前帧的初始ITD值的可信度高。

[0198] 若当前帧的初始ITD值的可信度高,可以将该初始ITD值作为当前帧的ITD值。进一步地,可以预设ITD值准确计算标志位:itd_cal_flag,如果当前帧的初始ITD值的可信度高,可以将itd_cal_flag置为1,如果当前帧的初始ITD值的可信度低,可以将itd_cal_flag置为0。

[0199] 进一步地,如果当前帧的初始ITD值的可信度高,可以将目标帧计数值置为预设的初始值,例如,可以将目标帧计数值置为0,或置为1。

[0200] 614、若当前帧的初始ITD值的可信度低,可以对初始ITD值进行ITD值修正。ITD值修正的方式可以有很多种,例如,可以对ITD值进行拖尾处理,或者,可以根据前后帧相关性对ITD值进行修正等,本发明实施例对此不作具体限定。

[0201] 616-618、判断当前帧是否复用了前一帧的ITD值,如果当前帧复用了前一帧的ITD值,增加目标帧计数值的取值。

[0202] 620-622、判断修正的分段信噪比是否满足预设的信噪比条件,如果修正的分段信噪比满足预设的信噪比条件,则停止复用前一帧的ITD值作为当前帧的ITD值。例如,可以修改目标帧计数值的取值,使其大于或等于该目标帧计数值的阈值(该阈值可以指示允许连续出现的目标帧的数量),从而停止复用当前帧的前一帧的ITD值作为当前帧的ITD值。

[0203] 判断修正的分段信噪比是否满足预设的信噪比条件的方式可以有多种,可选地,在一些实施中,当修正的分段信噪比小于第一阈值或者大于第二阈值时,可以认为修正的分段信噪比满足预设的信噪比条件,在这种情况下,可以修改目标帧计数值的取值,使其大于或等于该目标帧计数值的阈值。

[0204] 例如,假设预先设定了高信噪比语音门限值HIGH_SNR_VOICE_TH为10000,可以将上述第一阈值设定为 $A_1 * \text{HIGH_SNR_VOICE_TH}$,并将上述第二阈值设定为 $A_2 * \text{HIGH_SNR_VOICE_TH}$, A_1 、 A_2 为正实数,且 $A_1 < A_2$,这里 A_1 可以取0.5,0.6,0.7或其他经验值, A_2 可以取290,300,310或其他经验值。目标帧计数值的阈值可以等于9,10,11或其他经验值。

[0205] 624、如果修正的分段信噪比不满足预设的信噪比条件,计算表征左右声道频域信号的互相关系数中的峰值位置的稳定程度的参数。

[0206] 具体地,如果修正的分段信噪比大于等于第一阈值且小于等于第二阈值,可以认为修正的分段信噪比不满足预设的信噪比条件,在这种情况下,计算表征左右声道频域信号的互相关系数中的峰值位置的稳定程度的参数。

[0207] 本实施例中,表征左右声道频域信号的互相关系数中的峰值位置的稳定程度的参数可以是一组参数,该组参数可以包括互相关系数的峰值幅度可信度参数 $peak_mag_prob$ 以及峰值位置波动性参数 $peak_pos_fluc$ 。

[0208] 具体地, $peak_mag_prob$ 可以采用如下方式计算:

[0209] 首先,对左右声道频域信号的互相关系数 $Xcorr_itd(t)$ 按照幅度值从大到小或者从小到大的顺序进行排序,根据排序后的左右声道频域信号的互相关系数 $Xcorr_itd(t)$,通过公式(16),计算 $peak_mag_prob$:

$$[0210] \quad peak_mag_prob = \frac{Xcorr_itd(X) - Xcorr_itd(Y)}{Xcorr_itd(X)} \quad (16)$$

[0211] 其中,X表征排序后的左右声道频域信号的互相关系数中的峰值位置的索引,Y表征排序后的左右声道频域信号的互相关系数的预设位置的索引。例如,按照幅度值从小到大的顺序对左右声道频域信号的互相关系数 $Xcorr_itd(t)$ 进行排序,X的位置为 $2*ITD_MAX$,Y的位置可以选取为 $2*ITD_MAX-1$,这样一来,本发明实施例就将左右声道频域信号的互相关系数中的峰值的幅度值与次大值的幅度值之间的差值与该峰值的幅度值之间的比值作为互相关系数的峰值幅度可信度参数,即 $peak_mag_prob$,当然,这仅是 $peak_mag_prob$ 的一种选取方式。

[0212] 进一步地, $peak_pos_fluc$ 的计算方式也可以有多种。可选地,在一些实施例中, $peak_pos_fluc$ 可以是根据左右声道频域信号的互相关系数中的峰值位置的索引对应的ITD值以及该当前帧的前N帧的ITD值计算得到的,其中,N为大于等于1的整数。可选地,在一些实施例中, $peak_pos_fluc$ 可以是根据左右声道频域信号的互相关系数中的峰值位置的索引与当前帧的前N帧的左右声道频域信号的互相关系数中的峰值位置的索引计算得到,其中,N为大于等于1的整数。

[0213] 例如,参见公式(17), $peak_pos_fluc$ 可以选取左右声道频域信号的互相关系数中的峰值位置的索引对应的ITD值与当前帧的前一帧的ITD值之差的绝对值:

$$[0214] \quad peak_pos_fluc = abs(\operatorname{argmax}(Xcorr(t)) - ITD_MAX - prev_itd) \quad (17)$$

[0215] 其中, $prev_itd$ 表征当前帧的前一帧的ITD值, $abs(*)$ 表征取绝对值操作, argmax 表征搜索最大值位置的操作。

[0216] 626-628、判断左右声道频域信号的互相关系数中的峰值位置的稳定程度是否满足预设条件,如果满足预设条件,增加目标帧计数值。

[0217] 换句话说,就是当左右声道频域信号的互相关系数中的峰值位置的稳定程度满足预设条件时,则减少允许连续出现的目标帧的数量。

[0218] 例如,若 $peak_mag_prob$ 大于峰值幅度可信度阈值 th_{prob} ,并且 $peak_pos_fluc$ 大于峰值位置波动性阈值 th_{fluc} ,则增加目标帧计数值。本发明实施例中,峰值幅度可信度阈值 th_{prob} 可以设置为0.1,0.2,0.3或其他经验值,峰值位置波动性阈值 th_{fluc} 可以设置为4,5,6或其他经验值。

[0219] 应理解,增加目标帧计数值的方式可以有多种。

[0220] 可选地,在一些实施例中,可以是直接将目标帧计数值加1。

[0221] 可选地,在一些实施例中,可以根据修正的分段信噪比和/或表征不同声道间互相关系数中的峰值位置的稳定程度的一组参数中的一个或多个,控制目标帧计数值的增加量。

[0222] 例如,若 $R_1 \leq \text{mssnr} < R_2$,目标帧计数值加1;若 $R_2 \leq \text{mssnr} < R_3$,目标帧计数值加2;若 $R_3 \leq \text{mssnr} \leq R_4$,目标帧计数值加3,其中, $R_1 < R_2 < R_3 < R_4$ 。

[0223] 又如,若 $U_1 < \text{peak_mag_prob} < U_2$ 且 $\text{peak_pos_fluc} > \text{th}_{\text{fluc}}$,目标帧计数值加1;若 $U_2 < \text{peak_mag_prob} < U_3$ 且 $\text{peak_pos_fluc} > \text{th}_{\text{fluc}}$,目标帧计数值加2;若 $U_3 \leq \text{peak_mag_prob}$ 且 $\text{peak_pos_fluc} > \text{th}_{\text{fluc}}$,目标帧计数值加3。此处的 U_1 可以为上述峰值幅度可信度阈值 th_{prob} ,且 $U_1 < U_2 < U_3$ 。

[0224] 630-634、判断当前帧是否满足复用当前帧的前一帧的ITD值的条件。若满足,则将当前帧的前一帧的ITD值作为当前帧的ITD值,并增加目标帧计数值;否则,当前帧的ITD值不复用当前帧的前一帧的ITD值,执行下一帧处理。

[0225] 需要说明的是,本发明实施例对当前帧是否满足复用当前帧的前一帧的ITD值的条件不作具体限定,该条件的设置可以考虑初始ITD值的准确性、目标帧的计数值是否达到阈值、当前帧是否为连续的话音帧等因素中的一个或多个因素。

[0226] 例如,如果当前帧的第 m 子帧的语音激活检测结果与前一帧语音激活检测的结果均为话音帧,若前一帧的ITD值不等于零,当前帧的初始ITD值等于零,且当前帧的初始ITD值的可信度低(初始ITD值的可信度可以通过 itd_cal_flag 的取值进行标识,例如, itd_cal_flag 不等于1表示初始ITD值的可信度低,具体参见步骤612的描述),且目标帧数计数值小于该目标帧计数值的阈值,则可以将当前帧的前一帧的ITD值作为当前帧的ITD值,并增加目标帧计数值。

[0227] 进一步地,如果当前帧和当前帧的前一帧的第 m 子帧的语音激活检测结果均为话音帧,则可以将该前一帧的语音激活检测结果的标志位 pre_vad 更新为话音帧标志,即 pre_vad 等于1,否则将前一帧语音激活检测的结果 pre_vad 更新为背景噪音帧标志,即 pre_vad 等于0。

[0228] 上文结合步骤604,详细描述了修正的分段信噪比的一种计算方式,但本发明实施例不限于此,下文给出修正的分段信噪比的其他实现方式。

[0229] 可选地,在一些实施例中,可以按照如下方式计算修正的分段信噪比:

[0230] 步骤一,根据第 m 子帧的左声道频域信号 $X_{m,\text{left}}(k)$ 以及第 m 子帧的右声道频域信号 $X_{m,\text{right}}(k)$,通过公式(18)和(19),计算第 m 子帧的左声道频域信号的平均幅度谱 $\text{SPD}_{m,\text{left}}(k)$ 以及第 m 子帧的右声道频域信号的平均幅度谱 $\text{SPD}_{m,\text{right}}(k)$ 。

$$[0231] \quad \text{SPD}_{m,\text{left}}(k) = (\text{real}\{X_{m,\text{left}}(k)\})^2 + (\text{imag}\{X_{m,\text{left}}(k)\})^2 \quad (18)$$

$$[0232] \quad \text{SPD}_{m,\text{right}}(k) = (\text{real}\{X_{m,\text{right}}(k)\})^2 + (\text{imag}\{X_{m,\text{right}}(k)\})^2 \quad (19)$$

[0233] 其中, $k=1, \dots, L/2-1$, L 为快速傅里叶变换长度,例如, L 可以取400、800等。

[0234] 步骤二、根据 $\text{SPD}_{m,\text{left}}(k)$ 和 $\text{SPD}_{m,\text{right}}(k)$,通过公式(20)和(21),计算当前帧的左、右声道频域信号的平均幅度谱 $\text{SPD}_{\text{left}}(k)$ 和 $\text{SPD}_{\text{right}}(k)$ 。

$$[0235] \quad \text{SPD}_{\text{left}}(k) = \frac{1}{\text{SUBFR_NUM}} \sum_{m=0}^{\text{SUBFR_NUM}-1} \text{SPD}_{m,\text{left}}(k) \quad (20a)$$

$$[0236] \quad SPD_{right}(k) = \frac{1}{SUBFR_NUM} \sum_{m=0}^{SUBFR_NUM-1} SPD_{m,right}(k) \quad (21a)$$

[0237] 或者,也可以

$$[0238] \quad SPD_{left}(k) = \sum_{m=0}^{SUBFR_NUM-1} SPD_{m,left}(k) \quad (20b)$$

$$[0239] \quad SPD_{right}(k) = \sum_{m=0}^{SUBFR_NUM-1} SPD_{m,right}(k) \quad (21b)$$

[0240] 其中, SUBFR_NUM 表征一个音频帧所包含的子帧的个数。

[0241] 步骤三、根据 $SPD_{left}(k)$ 、 $SPD_{right}(k)$, 通过公式 (22), 计算当前帧左右声道频域信号的平均幅度谱 $SPD(k)$:

$$[0242] \quad SPD(k) = A * SPD_{left}(k) + (1-A) * SPD_{right}(k) \quad (22)$$

[0243] 其中, A 为预先设定的左右声道幅度谱混合比例因子, A 可以取 0.4, 0.5, 0.6 或其他经验值。

[0244] 步骤四、根据 $SPD(k)$, 通过公式 (23), 计算子带能量 $E_band(i)$, $i=0, 1, \dots, BAND_NUM-1$, BAND_NUM 表征子带个数。

$$[0245] \quad E_band(i) = \frac{1}{band_rb[i+1] - band_rb[i]} \sum_{k=band_rb[i]}^{band_rb[i+1]-1} SPD(k) \quad (23)$$

[0246] 其中 $band_tb$ 表征预先设定用于子带划分的表格, $band_tb[i]$ 表征第 i 个子带下限频点, $band_tb[i+1]-1$ 表征第 i 个子带上限频点。

[0247] 步骤五、根据 $E_band(i)$ 以及子带噪声能量估计 $E_band_n(i)$, 计算修正的分段信噪比 $mssnr$ 。具体可以采用公式 (7) 和公式 (8) 描述的实现方式计算 $mssnr$, 此处不再详述。

[0248] 步骤六、根据 $E_band(i)$ 对 $E_band_n(i)$ 进行更新。具体可以采用公式 (9) 至公式 (11) 描述的实现方式对 $E_band_n(i)$ 进行更新, 此处不再详述。

[0249] 可选地, 在另一些实施例中, 可以按照如下方式计算修正的分段信噪比:

[0250] 步骤一、根据第 m 子帧的左声道频域信号 $X_{m,left}(k)$ 以及第 m 子帧的右声道频域信号 $X_{m,right}(k)$, 通过公式 (24) 和公式 (25), 计算第 m 子帧的左声道频域信号的平均幅度谱 $SPD_{m,left}(k)$ 和第 m 子帧的右声道频域信号的平均幅度谱 $SPD_{m,right}(k)$ 。

$$[0251] \quad SPD_{m,left}(k) = (\text{real}\{X_{m,left}(k)\})^2 + (\text{imag}\{X_{m,left}(k)\})^2 \quad (24)$$

$$[0252] \quad SPD_{m,right}(k) = (\text{real}\{X_{m,right}(k)\})^2 + (\text{imag}\{X_{m,right}(k)\})^2 \quad (25)$$

[0253] 其中, $k=1, \dots, L/2-1$, L 为快速傅里叶变换长度, 例如, L 可以取 400、800 等。

[0254] 步骤二、根据 $SPD_{m,left}(k)$ 和 $SPD_{m,right}(k)$, 通过公式 (26), 计算第 m 子帧的左右声道频域信号的平均幅度谱 $SPD_m(k)$ 。

$$[0255] \quad SPD_m(k) = A * SPD_{m,left}(k) + (1-A) * SPD_{m,right}(k) \quad (26)$$

[0256] 其中, A 为预先设定的左右声道幅度谱混合比例因子, A 可以取 0.4, 0.5, 0.6 或其他经验值。

[0257] 步骤三、根据 $SPD_m(k)$, 通过公式 (27), 计算当前帧的左右声道频域信号的平均幅度谱 $SPD(k)$ 。

[0258] 一种可选的计算方式如下:

$$[0259] \quad SPD(k) = \frac{1}{SUBFR_NUM} \sum_{m=0}^{SUBFR_NUM-1} SPD_m(k) \quad (27a)$$

[0260] 另一种可选的计算方式如下:

$$[0261] \quad SPD(k) = \sum_{m=0}^{SUBFR_NUM-1} SPD_m(k) \quad (27b)$$

[0262] 步骤四、根据SPD(k),通过公式(28),计算子带能量E_band(i), $i=0,1,\dots,BAND_NUM-1$,BAND_NUM为子带个数。

$$[0263] \quad E_band_m(i) = \frac{1}{band_rb[i+1] - band_rb[i]} \sum_{k=band_rb[i]}^{band_rb[i+1]-1} SPD_m(k) \quad (28)$$

[0264] 其中band_tb表征预先设定用于子带划分的表格,band_tb[i]表征第i个子带下限频点,band_tb[i+1]-1表征第i个子带上限频点。

[0265] 步骤五、根据E_band_m(i)以及子带噪声能量估计E_band(i),计算修正的分段信噪比mssnr。具体可以采用公式(7)和公式(8)描述的实现方式计算mssnr,此处不再详述。

[0266] 步骤六、根据E_band(i)对E_band_n(i)进行更新。具体可以采用公式(9)至公式(11)描述的实现方式对E_band_n(i)进行更新,此处不再详述。

[0267] 可选地,在另一些实施例中,可以按照如下方式计算修正的分段信噪比:

[0268] 步骤一、根据第m子帧的左声道频域信号 $X_{m,left}(k)$ 以及第m子帧的右声道频域信号 $X_{m,right}(k)$,通过公式(29),计算第m子帧的左右声道频域信号的平均幅度谱 $SPD_m(k)$:

$$[0269] \quad SPD_m(k) = A * SPD_{m,left}(k) + (1-A) SPD_{m,right}(k) \quad (29)$$

[0270] 其中:

$$[0271] \quad SPD_{m,left}(k) = (\text{real}\{X_{m,left}(k)\})^2 + (\text{imag}\{X_{m,left}(k)\})^2$$

$$[0272] \quad SPD_{m,right}(k) = (\text{real}\{X_{m,right}(k)\})^2 + (\text{imag}\{X_{m,right}(k)\})^2$$

[0273] $k=1,\dots,L/2-1$,L为快速傅里叶变换长度,例如,L可以取400、800等。A为预先设定的左右声道幅度谱混合比例因子,A可以取0.4,0.5,0.6或其他经验值。

[0274] 步骤二、根据 $SPD_m(k)$,通过步骤(30),计算第m子帧的子带能量E_band_m(i), $i=0,1,\dots,BAND_NUM-1$,BAND_NUM为子带个数。

$$[0275] \quad E_band_m(i) = \frac{1}{band_rb[i+1] - band_rb[i]} \sum_{k=band_rb[i]}^{band_rb[i+1]-1} SPD_m(k) \quad (30)$$

[0276] 其中band_tb表征预先设定用于子带划分的表格,band_tb[i]表征第i个子带下限频点,band_tb[i+1]-1表征第i个子带上限频点。

[0277] 步骤三、根据第m子帧的子带能量E_band_m(i),通过公式(31),计算当前帧的子带能量E_band(i)。

$$[0278] \quad E_band(i) = \frac{1}{SUBFR_NUM} \sum_{m=0}^{SUBFR_NUM-1} E_band_m(i) \quad (31a)$$

[0279] 或者,也可以

$$[0280] \quad E_band(i) = \sum_{m=0}^{SUBFR_NUM-1} E_band_m(i) \quad (31b)$$

[0281] 步骤四、根据 $E_{\text{band}}(i)$ 以及子带噪声能量估计 $E_{\text{band}_n}(i)$,计算修正的分段信噪比 $mssnr$ 。具体可以采用公式(7)和公式(8)描述的实现方式计算 $mssnr$,此处不再详述。

[0282] 步骤五、根据 $E_{\text{band}}(i)$ 对 $E_{\text{band}_n}(i)$ 进行更新。具体可以采用公式(9)至公式(11)描述的实现方式对 $E_{\text{band}_n}(i)$ 进行更新,此处不再详述。

[0283] 上文结合步骤605,详细描述了语音激活检测的一种实现方式,但本发明实施例不限于此,下文给出了语音激活检测的另一种实现方式。

[0284] 具体地,如果修正的分段信噪比大于语音激活检测阈值 th_{VAD} ,则当前帧为语音帧,当前帧的语音激活检测标志 vad_flag 设置为1,否则当前帧为背景噪声帧,当前帧的语音激活检测标志 vad_flag 设置为0。语音激活检测阈值 th_{VAD} 一般为经验值,这里可以3500,4000,4500等。

[0285] 相应地,步骤630-634的实现方式可以修改成如下实现方式:

[0286] 当前帧的语音激活检测结果与前一帧语音激活检测的结果 pre_vad 均为语音帧时,若前一帧的ITD值不等于零,当前帧的ITD值等于零,且当前帧的ITD值的可信度低(初始ITD值的可信度可以通过 itd_cal_flag 的取值进行标识,例如, itd_cal_flag 不等于1表示初始ITD值的可信度低,具体参见步骤612的描述),且目标帧计数值小于该目标帧计数值的阈值,则将前一帧的ITD值作为当前帧的ITD值,并增加目标帧计数值。

[0287] 若当前帧的语音激活检测结果为语音帧时,将前一帧语音激活检测的结果 pre_vad 更新为语音帧标志,即 pre_vad 等于1,否则将前一帧语音激活检测的结果 pre_vad 更新为背景噪声帧标志,即 pre_vad 等于0。

[0288] 上文结合步骤626-628,详细描述了允许连续出现的目标帧的数量的一种调整或控制方式,但本发明实施例不限于此,下文给出允许连续出现的目标帧的数量的其他调整或控制方式。

[0289] 可选地,在一些实施例中,首先,判断左右声道频域信号的互相关系数中的峰值位置的稳定程度是否满足预设条件,如果满足预设条件,减小目标帧计数值的阈值。换句话说,本发明实施例通过减小目标帧计数值的阈值的方式,减少允许连续出现的目标帧的数量。

[0290] 需要说明的是,判断左右声道频域信号的互相关系数中的峰值位置的稳定程度是否满足预设条件的方式可以有多种,本发明实施例对此不作具体限定。例如,该预设条件可以是:左右声道频域信号的互相关系数的峰值幅度可信度参数大于预设的峰值幅度可信度阈值,且峰值位置波动性参数大于预设的峰值位置波动性阈值,其中,峰值幅度可信度阈值可以取0.1,0.2,0.3或其他经验值,峰值位置波动性阈值可以取4,5,6或其他经验值。

[0291] 需要说明的是,减小目标帧计数值的阈值的方式可以有多种,本发明实施例对此不作具体限定。

[0292] 可选地,在一些实施例中,可以直接将目标帧计数值的阈值减1。

[0293] 可选地,在另一些实施例中,可以根据修正的分段信噪比以及表征左右声道频域信号的互相关系数中的峰值位置的稳定程度的一组参数中的一个或多个,控制目标帧计数值的阈值的减少量。

[0294] 例如,若 $R_1 \leq mssnr < R_2$,可以将目标帧计数值的阈值减1;若 $R_2 \leq mssnr < R_3$,可以将目标帧计数值的阈值减2;若 $R_3 \leq mssnr \leq R_4$,可以将目标帧计数值的阈值减3,其中, R_1 、 R_2 、

R_3 、 R_4 满足 $R_1 < R_2 < R_3 < R_4$ 。

[0295] 又如,若 $U_1 < \text{peak_mag_prob} < U_2$ 且 $\text{peak_pos_fluc} > \text{th}_{\text{fluc}}$,可以将目标帧计数值的阈值减1;若 $U_2 < \text{peak_mag_prob} < U_3$ 且 $\text{peak_pos_fluc} > \text{th}_{\text{fluc}}$,可以将目标帧计数值的阈值减2;若 $U_3 \leq \text{peak_mag_prob}$ 且 $\text{peak_pos_fluc} > \text{th}_{\text{fluc}}$,可以将目标帧计数值的阈值减3,其中, U_1 、 U_2 、 U_3 可以满足 $U_1 < U_2 < U_3$,此外, U_1 可以是上文描述的峰值幅度可信度阈值 th_{prob} 。

[0296] 上文结合步骤624,详细描述了表征左右声道频域信号的互相关系数中的峰值位置的稳定程度的参数的计算方式。其中,在步骤624中,表征左右声道频域信号的互相关系数中的峰值位置的稳定程度的参数主要包括峰值幅度可信度参数 peak_mag_prob 以及峰值位置波动性参数 peak_pos_fluc 两种,但本发明实施例不限于此。

[0297] 可选地,在一些实施例中,表征左右声道频域信号的互相关系数中的峰值位置的稳定程度的参数可以仅包括 peak_pos_fluc 。相应地,步骤626可以修改为:如果 peak_pos_fluc 大于峰值位置波动性阈值 th_{fluc} ,则增加目标帧计数值。

[0298] 可选地,在另一些实施例中,表征不同声道间互相关系数中的峰值位置的稳定程度的参数可以是将 peak_mag_prob 以及 peak_pos_fluc 进行线性和/或非线性运算得到的峰值位置稳定度参数 peak_stable 。

[0299] 例如, peak_stable 可以与 peak_mag_prob 和 peak_pos_fluc 的关系可以通过公式(32)表示:

$$[0300] \quad \text{peak_stable} = \text{peak_mag_prob} / (\text{peak_pos_fluc})^P \quad (32)$$

[0301] 又如, peak_stable 可以与 peak_mag_prob 和 peak_pos_fluc 的关系可以通过公式(33)表示:

$$[0302] \quad \text{peak_stable} = \text{diff_factor}[\text{peak_pos_fluc}] * \text{peak_mag_prob} \quad (33)$$

[0303] 其中, diff_factor 表征预设的相邻帧的ITD值的差异影响因子序列, diff_factor 可以包含 peak_pos_fluc 的所有可能取值对应的相邻帧的ITD值的差异影响因子。 diff_factor 可以通过经验设定,也可以通过大量数据训练得到。 P 可以表示左右声道频域信号的互相关系数的峰值位置波动影响斜度, P 可以取大于或等于1的正整数,例如, P 可以为1,2,3或其他经验值。

[0304] 相应地,步骤626可以修改为:如果 peak_stable 大于预先设定的峰值位置稳定度阈值,则增加目标帧计数值。这里,预先设定的峰值位置稳定度阈值可以选取大于或等于0的正实数,或者选取其他经验值。

[0305] 进一步地,在一些实施例中,可以对 peak_stable 进行平滑处理,得到平滑处理后的峰值位置稳定度参数 lt_peak_stable ,并基于 lt_peak_stable 进行后续判断。

[0306] 具体地, lt_peak_stable 可以通过公式(34)计算得到:

$$[0307] \quad \text{lt_peak_stable} = (1 - \alpha) * \text{lt_peak_stable} + \alpha * \text{peak_stable} \quad (34)$$

[0308] 其中, α 表征长时平滑因子,一般可以取大于等于0,且小于等于1的正实数,例如, α 取0.4,0.5,0.6或其他经验值。

[0309] 相应地,步骤626可以修改为:若 lt_peak_stable 大于预先设定的峰值位置稳定度阈值,则增加目标帧计数值。这里,预先设定的峰值位置稳定度阈值可以选取大于或等于0的正实数,或者选取其他经验值。

[0310] 下面对本发明的装置实施例进行描述,由于装置实施例可以执行上述方法,因此

未详细描述的部分可以参见前面各方法实施例。

[0311] 图7是本发明实施例的编码器的示意性框图。图7的编码器700包括：

[0312] 获取单元710,用于获取当前帧的多声道信号；

[0313] 第一确定单元720,用于确定所述当前帧的初始ITD值；

[0314] 控制单元730,用于根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量,所述特征信息包括所述多声道信号的信噪比参数以及所述多声道信号的互相关系数的峰值特性中的至少一个,所述目标帧的ITD值复用了所述目标帧的前一帧的ITD值；

[0315] 第二确定单元740,用于根据所述当前帧的初始ITD值,以及所述允许连续出现的目标帧的数量,确定所述当前帧的ITD值；

[0316] 编码单元750,用于根据所述当前帧的ITD值,对所述多声道信号进行编码。

[0317] 本发明实施例能够降低背景噪声、混响、多说话人同时讲话等环境因素对ITD值计算结果的准确性以及稳定性的影响,在存在噪声、混响以及多说话人同时讲话或者信号谐波特征不明显的情况下,改善PS编码中的ITD值的稳定性,尽量减少ITD值的不必要的跳变,从而避免下混信号的帧间不连续以及解码信号的声像不稳定,同时,本发明实施例能够更好地保持立体声信号的相位信息,提升听觉质量。

[0318] 可选地,在一些实施例中,所述编码器700还包括:第三确定单元,用于根据所述多声道信号的互相关系数的峰值的幅度和所述多声道信号的互相关系数的峰值位置的索引,确定所述多声道信号的互相关系数的峰值特性。

[0319] 可选地,在一些实施例中,所述第三确定单元具体用于根据所述多声道信号的互相关系数的峰值的幅度,确定峰值幅度可信度参数,所述峰值幅度可信度参数表征所述多声道信号的互相关系数的峰值幅度的可信度;根据所述多声道信号的互相关系数的峰值位置的索引对应的ITD值,以及所述当前帧的前一帧的ITD值,确定峰值位置波动性参数,所述峰值位置波动性参数表征所述多声道信号的互相关系数的峰值位置的索引对应的ITD值与所述当前帧的前一帧的ITD值的差异;根据所述峰值幅度可信度参数和所述峰值位置波动性参数,确定所述多声道信号的互相关系数的峰值特性。

[0320] 可选地,在一些实施例中,所述第三确定单元具体用于将所述多声道信号的互相关系数中的峰值的幅度值和次大值的幅度值之差与所述峰值的幅度值的比值确定为所述峰值幅度可信度参数。

[0321] 可选地,在一些实施例中,所述第三确定单元具体用于将所述多声道信号的互相关系数的峰值位置的索引对应的ITD值与所述当前帧的前一帧的ITD值之差的绝对值确定为所述峰值位置波动性参数。

[0322] 可选地,在一些实施例中,所述控制单元730具体用于根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量,在所述多声道信号的互相关系数的峰值特性满足预设条件的情况下,通过调整目标帧计数值和所述目标帧计数值的阈值中的至少一个,减少允许连续出现的目标帧的数量,其中,所述目标帧计数值用于表征当前已连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

[0323] 可选地,在一些实施例中,所述控制单元730具体用于通过增加所述目标帧计数值,减少允许连续出现的目标帧的数量。

[0324] 可选地,在一些实施例中,所述控制单元730具体用于通过减小所述目标帧计数值的阈值,减少允许连续出现的目标帧的数量。

[0325] 可选地,在一些实施例中,所述控制单元730具体用于在所述多声道信号的信噪比参数不满足预设的信噪比条件的情况下,根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量;所述编码器700还包括:停止单元,用于在所述多声道信号的信噪比满足所述信噪比条件的情况下,停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值。

[0326] 可选地,在一些实施例中,所述控制单元730具体用于确定所述多声道信号的信噪比参数是否满足预设的信噪比条件;在所述多声道信号的信噪比参数不满足所述信噪比条件的情况下,根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量;在所述多声道信号的信噪比满足所述信噪比条件的情况下,停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值。

[0327] 可选地,在一些实施例中,所述停止单元具体用于增加目标帧计数值,使得所述目标帧计数值的取值大于或等于所述目标帧计数值的阈值,其中,所述目标帧计数值用于表征当前已经连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

[0328] 可选地,在一些实施例中,所述第二确定单元740具体用于根据所述当前帧的初始ITD值,目标帧计数值,所述目标帧计数值的阈值,确定所述当前帧的ITD值,其中,所述目标帧计数值用于表征当前已连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

[0329] 可选地,在一些实施例中,所述信噪比参数为所述多声道信号的修正的分段信噪比。

[0330] 图8是本发明实施例的编码器的示意性框图。图8的编码器800包括:

[0331] 存储器810,用于存储程序;

[0332] 处理器820,用于执行程序,当所述程序被执行时,所述处理器820用于获取当前帧的多声道信号;确定所述当前帧的初始ITD值;根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量,所述特征信息包括所述多声道信号的信噪比参数以及所述多声道信号的互相关系数的峰值特性中的至少一个,所述目标帧的ITD值复用了所述目标帧的前一帧的ITD值;根据所述当前帧的初始ITD值,以及所述允许连续出现的目标帧的数量,确定所述当前帧的ITD值;根据所述当前帧的ITD值,对所述多声道信号进行编码。

[0333] 本发明实施例能够降低背景噪声、混响、多说话人同时讲话等环境因素对ITD值计算结果的准确性以及稳定性的影响,在存在噪声、混响以及多说话人同时讲话或者信号谐波特征不明显的情况下,改善PS编码中的ITD值的稳定性,尽量减少ITD值的不必要的跳变,从而避免下混信号的帧间不连续以及解码信号的声像不稳定,同时,本发明实施例能够更好地保持立体声信号的相位信息,提升听觉质量。

[0334] 可选地,在一些实施例中,所述编码器800还用于根据所述多声道信号的互相关系数的峰值的幅度和所述多声道信号的互相关系数的峰值位置的索引,确定所述多声道信号的互相关系数的峰值特性。

[0335] 可选地,在一些实施例中,所述编码器800具体用于根据所述多声道信号的互相关

系数的峰值的幅度,确定峰值幅度可信度参数,所述峰值幅度可信度参数表征所述多声道信号的互相关系数的峰值幅度的可信度;根据所述多声道信号的互相关系数的峰值位置的索引对应的ITD值,以及所述当前帧的前一帧的ITD值,确定峰值位置波动性参数,所述峰值位置波动性参数表征所述多声道信号的互相关系数的峰值位置的索引对应的ITD值与所述当前帧的前一帧的ITD值的差异;根据所述峰值幅度可信度参数和所述峰值位置波动性参数,确定所述多声道信号的互相关系数的峰值特性。

[0336] 可选地,在一些实施例中,所述编码器800具体用于将所述多声道信号的互相关系数中的峰值的幅度值和次大值的幅度值之差与所述峰值的幅度值的比值确定为所述峰值幅度可信度参数。

[0337] 可选地,在一些实施例中,所述编码器800具体用于将所述多声道信号的互相关系数的峰值位置的索引对应的ITD值与所述当前帧的前一帧的ITD值之差的绝对值确定为所述峰值位置波动性参数。

[0338] 可选地,在一些实施例中,所述编码器800具体用于根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量,在所述多声道信号的互相关系数的峰值特性满足预设条件的情况下,通过调整目标帧计数值和所述目标帧计数值的阈值中的至少一个,减少允许连续出现的目标帧的数量,其中,所述目标帧计数值用于表征当前已连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

[0339] 可选地,在一些实施例中,所述编码器800具体用于通过增加所述目标帧计数值,减少允许连续出现的目标帧的数量。

[0340] 可选地,在一些实施例中,所述编码器800具体用于通过减小所述目标帧计数值的阈值,减少允许连续出现的目标帧的数量。

[0341] 可选地,在一些实施例中,所述编码器800具体用于在所述多声道信号的信噪比参数不满足预设的信噪比条件的情况下,才根据所述多声道信号的特征信息,控制允许连续出现的目标帧的数量;所述编码器800还用于在所述多声道信号的信噪比满足所述信噪比条件的情况下,停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值。

[0342] 可选地,在一些实施例中,所述编码器800具体用于确定所述多声道信号的信噪比参数是否满足预设的信噪比条件;在所述多声道信号的信噪比参数不满足所述信噪比条件的情况下,根据所述多声道信号的互相关系数的峰值特性,控制允许连续出现的目标帧的数量;在所述多声道信号的信噪比满足所述信噪比条件的情况下,停止复用所述当前帧的前一帧的ITD值作为所述当前帧的ITD值。

[0343] 可选地,在一些实施例中,所述编码器800具体用于增加目标帧计数值,使得所述目标帧计数值的取值大于或等于所述目标帧计数值的阈值,其中,所述目标帧计数值用于表征当前已经连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

[0344] 可选地,在一些实施例中,所述编码器800具体用于根据所述当前帧的初始ITD值,目标帧计数值,所述目标帧计数值的阈值,确定所述当前帧的ITD值,其中,所述目标帧计数值用于表征当前已连续出现的目标帧的数量,所述目标帧计数值的阈值用于指示允许连续出现的目标帧的数量。

[0345] 可选地,在一些实施例中,所述信噪比参数为所述多声道信号的修正的分段信噪比。

[0346] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0347] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统、装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0348] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统、装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0349] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0350] 另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。

[0351] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0352] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应所述以权利要求的保护范围为准。

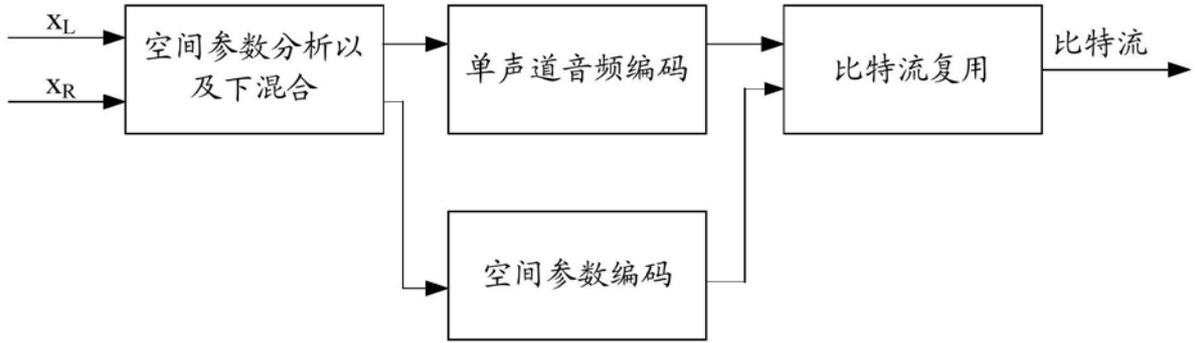


图1

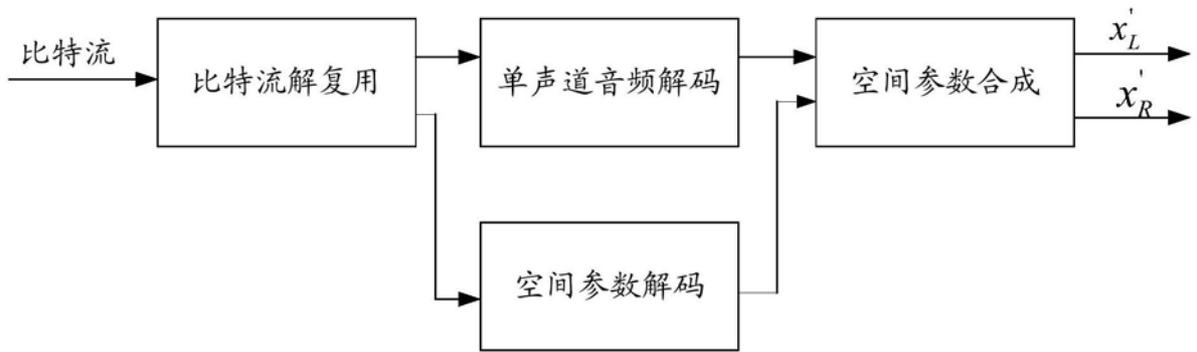


图2

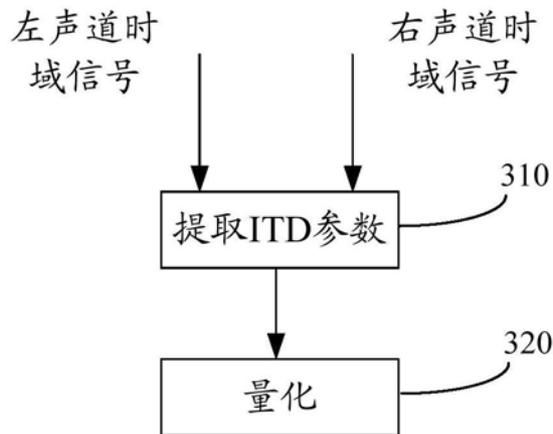


图3

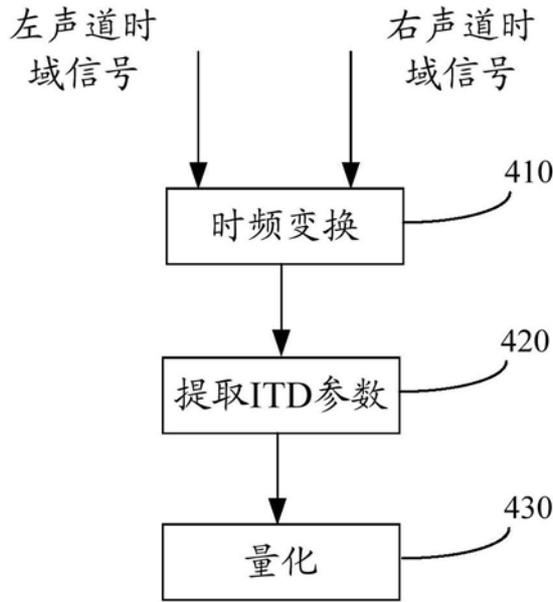


图4

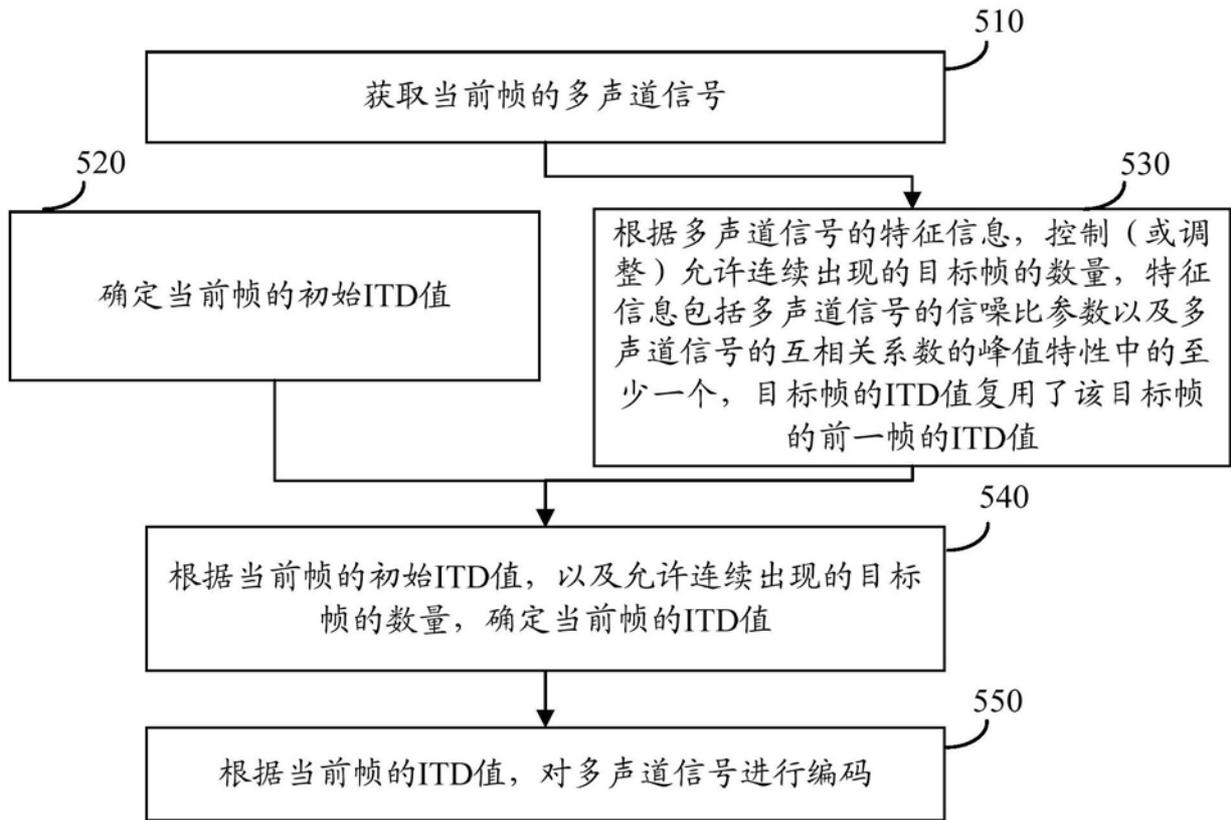


图5

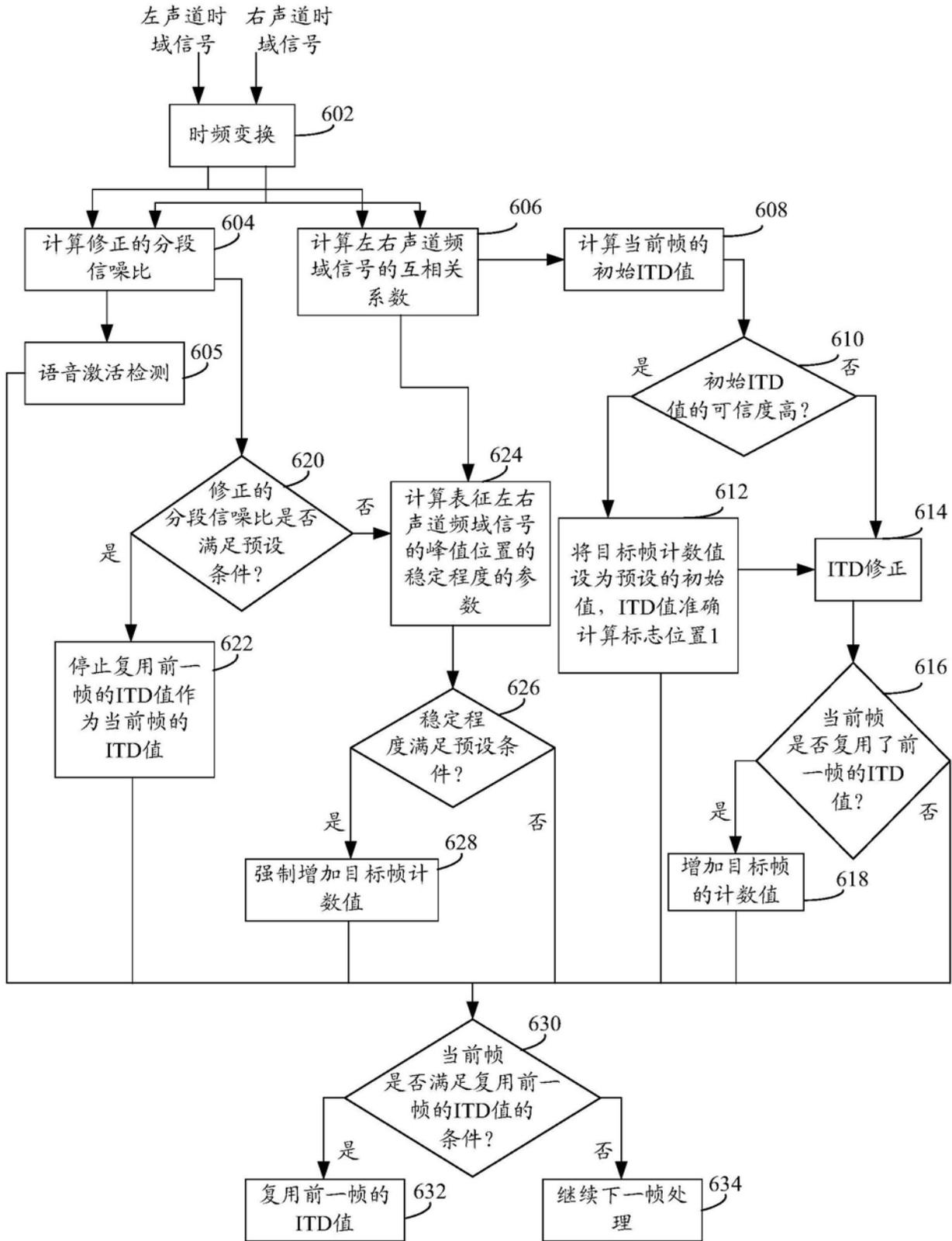


图6

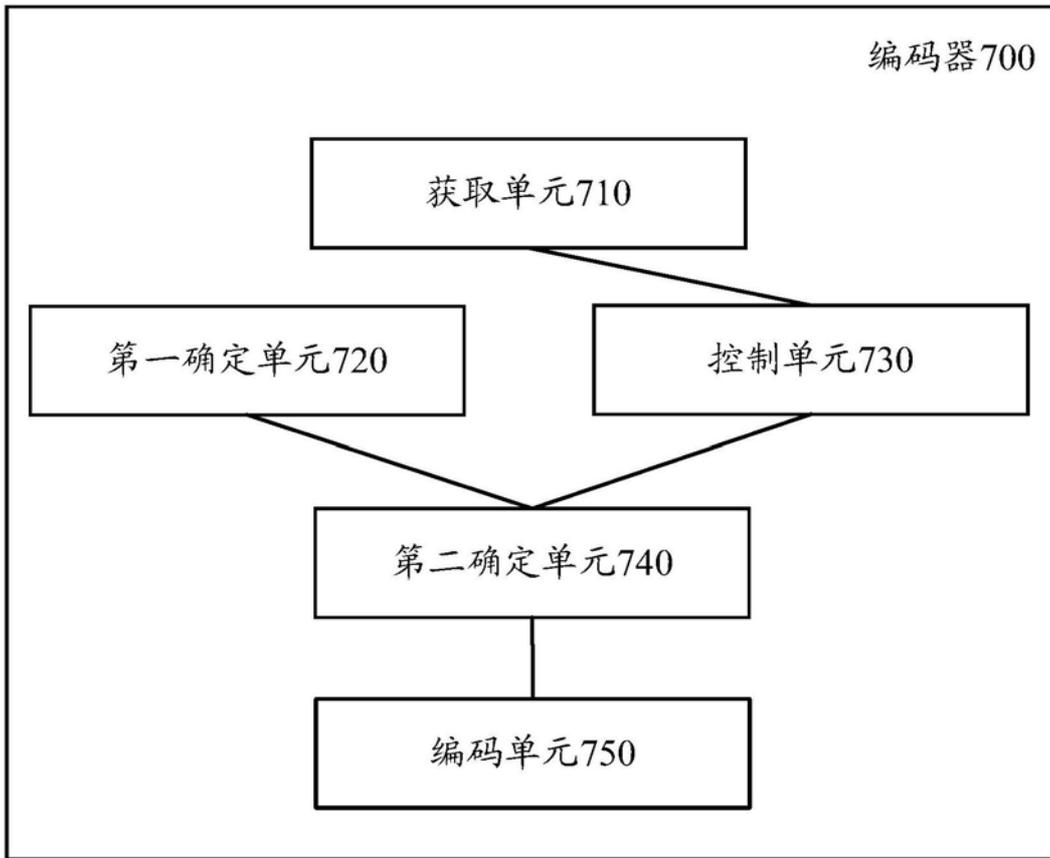


图7



图8