

一种立体声音频信号时延估计方法及装置

5 本申请要求于 2020 年 07 月 17 日提交中国专利局、申请号为 202010700806.7、申请名称为“一种立体声音频信号时延估计方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

10 本申请涉及音频编解码领域，特别涉及一种立体声音频信号时延估计方法及装置。

背景技术

15 在日常的音视频通信系统中，人们不仅追求高质量的图像，而且也追求高质量的音频。在语音与音频通信系统中，单通道音频越来越无法满足人们的需求，而立体声音频携带了各个声源的位置信息，提高了音频的清晰度、可懂度、真实感，因此越来越受到人们的青睐。

在立体声音频编解码技术中，参数立体声编解码技术是一种常见的音频编解码技术，常用的空间参数包含通道间相干性（inter-channel coherence, IC），通道间幅度差（inter-channel level difference, ILD），声道间时间差（inter-channel time difference, ITD），通道间相位差（inter-channel phase difference, IPD）等。其中 ILD 和 ITD 蕴含声源的位置信息，准确估计 ILD 和 ITD 信息对编码后立体声声像及声场的重建至关重要。

20 目前，最常用的一类 ITD 估计方法为广义互相关法，这是因为这类算法复杂度低，实时性好，易于实现，而且不依赖立体声音频信号的其它先验信息。但是在噪声环境下，现有的几种广义互相关算法的性能下降严重，导致对立体声音频信号的 ITD 估计精度偏低，使得参数编解码技术中解码后的立体声音频信号出现声像不准确、不稳定、空间感差、头中效应明显等问题，严重影响编码后立体声音频信号的音质。

发明内容

30 本申请提供了一种立体声音频信号时延估计方法及装置，以提高对立体声音频信号的声道间时间差的估计精度，进而提高解码后立体声音频信号声像的准确性和稳定性，提高音质。

第一方面，本申请提供一种立体声音频信号时延估计方法，该方法可以应用于一音频编码装置，该音频编码装置可以用于涉及立体声及多声道的音视频通信系统中的音频编码部分，也可以用于虚拟现实（virtual reality, VR）应用程序中的音频编码部分。该方法可以包括：音频编码装置获得立体声音频信号的当前帧，当前帧包括第一声道音频信号和第二声道音频信号；如果当前帧所包含的噪声信号的信号类型为相关性噪声信号类型，则采用第一算法估计第一声道音频信号和第二声道音频信号的声道间时间差（inter-channel time difference, ITD）；如果当前帧所包含的噪声信号的信号类型为弥散性噪声信号类型，则采用第二算法估计第一声道音频信号和所述第二声道

音频信号的ITD；其中，第一算法包括采用第一加权函数对当前帧的频域互功率谱加权，第二算法包括采用第二加权函数对当前帧的频域互功率谱加权，第一加权函数与第二加权函数的构造因子不同。

上述立体声音频信号可以是原始的立体声音频信号（包括左声道音频信号和右声道音频信号），也可以是多声道音频信号中的两路音频信号组成的立体声音频信号，还可以是由多声道音频信号中的多路音频信号联合产生的两路音频信号组成的立体声信号。当然，立体声音频信号还可以存在其他形式，本申请实施例不做具体限定。

可选的，上述音频编码装置具体可以为立体声编码装置，该装置可以构成独立的立体声编码器；也可以为多声道编码器中的核心编码部分，旨在对由多声道音频信号中的多路信号联合产生的两路音频信号所组成的立体声音频信号进行编码。

在一些可能的实施方式中，音频编码装置获得的立体声信号中的当前帧可以是频域音频信号或者时域音频信号。如果当前帧为频域音频信号，则音频编码装置可以直接在频域中对当前帧进行处理；而如果当前帧为时域音频信号，则音频编码装置可以先对时域中的当前帧进行时频变换，以得到频域中的当前帧，进而在频域中对当前帧进行处理。

在本申请中，音频编码装置通过对包含不同类型噪声的立体声音频信号采用不同的ITD估计算法，大幅提高了弥散性噪声和相关性噪声条件下对立体声音频信号的ITD估计的精度和稳定性，减少了立体声下混信号之间的帧间不连续，同时更好地保持了立体声信号的相位，编码后的立体声的声像更加准确和稳定，真实感更强，提高了编码后立体声信号的听觉质量。

在一些可能的实施方式中，在获得立体声音频信号的当前帧之后，上述方法还包括：获得当前帧的噪声相干值；如果噪声相干值大于或者等于预设阈值，则确定当前帧所包含的噪声信号的信号类型为相关性噪声信号类型；如果噪声相干值小于预设阈值，则确定当前帧所包含的噪声信号的信号类型为弥散性噪声信号类型。

可选的，上述预设阈值为经验值，可以设定为如0.20、0.25、0.30。

在一些可能的实施方式中，上述获得当前帧的噪声相干值，可以包括：对当前帧进行语音端点检测；如果检测结果表示当前帧的信号类型为噪声信号类型，则计算当前帧的噪声相干值；或者，如果检测结果表示当前帧的信号类型为语音信号类型，则将立体声音频信号中的当前帧的前一帧的噪声相干值确定为当前帧的噪声相干值。

可选的，音频编码装置可以以时域、频域或者时域频域结合的方式计算语音端点检测的值，对此不做具体限定。

在本申请中，音频编码装置计算当前帧的噪声相干值之后，还可以对其进行平滑处理，以减小噪声相干值估计的误差，提高噪声类型的识别准确率。

在一些可能的实施方式中，第一声道音频信号为第一声道时域信号，第二声道音频信号为第二声道时域信号；所述采用第一算法估计第一声道音频信号和第二声道音频信号的声道间时间差，包括：对第一声道时域信号和第二声道时域信号进行时频变换，以获得第一声道频域信号和第二声道频域信号；根据所述第一声道频域信号和所述第二声道频域信号，计算当前帧的频域互功率谱；采用第一加权函数对频域互功率谱进行加权；根据加权后的频域互功率谱，获得声道间时间差的估计值；其中，第一

加权函数的构造因子包括：第一声道频域信号对应的维纳增益因子、第二声道频域信号对应的维纳增益因子、幅值加权参数和当前帧的相干平方值。

5 在一些可能的实施方式中，第一声道音频信号为第一声道频域信号，第二声道音频信号为第二声道频域信号；所述采用第一算法估计第一声道音频信号和第二声道音频信号的声道间时间差，包括：根据第一声道频域信号和第二声道频域信号，计算当前帧的频域互功率谱；采用第一加权函数对频域互功率谱进行加权；根据加权后的频域互功率谱，获得声道间时间差的估计值；其中，第一加权函数的构造因子包括：第一声道频域信号对应的维纳增益因子、第二声道频域信号对应的维纳增益因子、幅值加权参数和当前帧的相干平方值。

10 在一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式：

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \frac{\Gamma^2(k)}{(1.0-\Gamma^2(k))};$$

其中， β 为幅值加权参数， $W_{x1}(k)$ 为第一声道频域信号对应的维纳增益因子； $W_{x2}(k)$ 为第二声道频域信号对应的维纳增益因子； $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

在一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式：

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

20 其中， β 为幅值加权参数， $W_{x1}(k)$ 为第一声道频域信号对应的维纳增益因子； $W_{x2}(k)$ 为第二声道频域信号对应的维纳增益因子； $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

可选的， $\beta \in [0,1]$ ，例如， $\beta=0.6$ 、 0.7 、 0.8 等。

25 在一些可能的实施方式中，第一声道频域信号对应的维纳增益因子可以为第一声道频域信号的第一初始维纳增益因子和/或第一改进维纳增益因子；第二声道频域信号对应的维纳增益因子可以为第二声道频域信号的第二初始维纳增益因子和/或第二改进维纳增益因子。

30 例如，第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一初始维纳增益因子，第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二初始维纳增益因子；那么，在获得立体声音频信号中的当前帧之后，上述方法还包括：根据第一声道频域信号，获得第一声道噪声功率谱的估计值；根据第一声道噪声功率谱的估计值，确定第一初始维纳增益因子；根据第二声道频域信号，获得第二声道噪声功率谱的估计值；根据第二声道噪声功率谱的估计值，确定第二初始维纳增益因子。

35 在本申请中，经过维纳增益因子加权后，立体声音频信号的频域互功率谱中的相关性噪声成分的权重大幅降低，残留噪声成分的相关性也会大幅减小，在大部分情况下，残留噪声的相干平方值会比立体声音频信号中的目标信号（如语音信号）的相干平方值小很多，这样目标信号对应的互相关峰值会更加突出，立体声音频信号的ITD估计的精

度和稳定性会大幅提高。

在一些可能的实施方式中，上述第一初始维纳增益因子 $W_{x1}^A(k)$ 满足以下公式：

$$W_{x1}^A(k) = \frac{|X_1(k)|^2 - |\hat{N}_1(k)|^2}{|X_1(k)|^2};$$

上述第二初始维纳增益因子 $W_{x2}^A(k)$ 满足以下公式：

$$W_{x2}^A(k) = \frac{|X_2(k)|^2 - |\hat{N}_2(k)|^2}{|X_2(k)|^2};$$

其中， $|\hat{N}_1(k)|^2$ 为第一声道噪声功率谱的估计值， $|\hat{N}_2(k)|^2$ 为第二声道噪声功率谱的估计值； $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， k 为频点索引值， $k=0,1,\dots,N_{\text{DFT}}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

又如，第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一改进维纳增益因子，第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二改进维纳增益因子；

在获得立体声音频信号中的当前帧之后，上述方法还包括：获得上述第一初始维纳增益因子和上述第二维纳增益因子；对第一初始维纳增益因子构建二值掩蔽函数，获得第一改进维纳增益因子；对第二初始维纳增益因子构建二值掩蔽函数，获得第二改进维纳增益因子。

在本申请中，通过对第一声道频域信号对应的第一初始维纳增益因子和第二声道频域信号对应的第二初始维纳增益因子构造二值掩蔽函数，筛选出受噪声影响比较小的频点，以提高 ITD 估计的精度。

在一些可能的实施方式中，上述第一改进维纳增益因子 $W_{x1}^B(k)$ 满足以下公式：

$$W_{x1}^B(k) = \begin{cases} 1 & \text{if } W_{x1}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x1}^A(k) < \mu_0; \end{cases}$$

上述第二改进维纳增益因子 $W_{x2}^B(k)$ 满足以下公式：

$$W_{x2}^B(k) = \begin{cases} 1 & \text{if } W_{x2}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x2}^A(k) < \mu_0; \end{cases}$$

其中， μ_0 为维纳增益因子的二值掩蔽门限， $W_{x1}^A(k)$ 为所述第一初始维纳增益因子； $W_{x2}^A(k)$ 为所述第二初始维纳增益因子。

可选的， $\mu_0 \in [0.5, 0.8]$ ，例如， $\mu_0 = 0.5$ 、0.66、0.75、0.8 等。

在一些可能的实施方式中，第一声道音频信号为第一声道时域信号，第二声道音频信号为第二声道时域信号；采用第二算法估计第一声道频域信号和所述第二声道频域信号的声道间时间差，包括：对第一声道时域信号和第二声道时域信号进行时频变换，以获得第一声道频域信号和第二声道频域信号；根据所述第一声道频域信号和所述第二声道频域信号，计算当前帧的频域互功率谱；采用第二加权函数对频域互功率谱进行加权，得到第一声道频域信号和第二声道频域信号的声道间时间差的估计值；其中，第二加权函数的构造因子包括：幅值加权参数和当前帧的相干平方值。

在一些可能的实施方式中，第一声道音频信号为第一声道频域信号，第二声道音频信号为第二声道频域信号；采用第二算法估计第一声道音频信号和第二声道音频信号的声道间时间差，包括：根据第一声道频域信号和第二声道频域信号，计算当前帧的频域互功率谱；采用第二加权函数对频域互功率谱进行加权；根据加权后的频域互功率谱，获得声道间时间差的估计值；其中，第二加权函数的构造因子包括：幅值加

权参数和当前帧的相干平方值。

在一些可能的实施方式中，第二加权函数 $\Phi_{new_2}(k)$ 满足以下公式：

$$\Phi_{new_2}(k) = \frac{1}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

其中， β 为幅值值加权参数， $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值，

- 5 $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

可选的， $\beta \in [0,1]$ ，例如， $\beta=0.6$ 、 0.7 、 0.8 等。

- 10 第二方面，本申请提供一种立体声音频信号时延估计方法，该方法可以应用于一音频编码装置，该音频编码装置可以用于涉及立体声及多声道的音视频通信系统中的音频编码部分，也可以用于VR应用程序中的音频编码部分。该方法可以包括：当前帧包括第一声道音频信号和第二声道音频信号；根据第一声道音频信号和第二声道音频信号，计算当前帧的频域互功率谱；采用预设加权函数对频域互功率谱进行加权；根据加权后的频域互功率谱，得到第一声道频域信号和第二声道频域信号的声道间时间差的估计值。

其中，预设加权函数包括第一加权函数或者第二加权函数，第一加权函数与第二加权函数的构造因子不同。

- 20 可选的，第一加权函数的构造因子包括：第一声道频域信号对应的维纳增益因子、第二声道频域信号对应的维纳增益、幅值加权参数和当前帧的相干平方值；第二加权函数的构造因子包括：幅值加权参数和当前帧的相干平方值。

- 25 在一些可能的实施方式中，第一声道音频信号为第一声道时域信号，第二声道音频信号为第二声道时域信号；根据所述第一声道音频信号和第二声道音频信号，计算当前帧的频域互功率谱，包括：对第一声道时域信号和第二声道时域信号进行时频变换，以获得第一声道频域信号和第二声道频域信号；根据所述第一声道频域信号和所述第二声道频域信号，计算所述当前帧的频域互功率谱。

在一些可能的实施方式中，第一声道音频信号为第一声道频域信号，第二声道音频信号为第二声道频域信号。

在一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式：

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \frac{\Gamma^2(k)}{(1.0-\Gamma^2(k))};$$

- 30 其中， β 为幅值加权参数， $W_{x1}(k)$ 为第一声道频域信号对应的维纳增益因子； $W_{x2}(k)$ 为第二声道频域信号对应的维纳增益因子； $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

- 35 在一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式：

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

其中， β 为幅值加权参数， $W_{x1}(k)$ 为第一声道频域信号对应的维纳增益因子； $W_{x2}(k)$ 为第二声道频域信号对应的维纳增益因子； $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值，

$\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$, $X_1(k)$ 为第一声道频域信号, $X_2(k)$ 为第二声道频域信号, $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数 k 为频点索引值, $k=0,1,\dots,N_{\text{DFT}}-1$, N_{DFT} 为当前帧在进行时频变换后的频点总数。

可选的, $\beta \in [0,1]$, 例如, $\beta=0.6$ 、 0.7 、 0.8 等。

- 5 在一些可能的实施方式中, 第一声道频域信号对应的维纳增益因子可以为第一声道频域信号的第一初始维纳增益因子和/或第一改进维纳增益因子; 第二声道频域信号对应的维纳增益因子可以为第二声道频域信号的第二初始维纳增益因子和/或第二改进维纳增益因子。

10 例如, 第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一初始维纳增益因子, 第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二初始维纳增益因子; 在获得立体声音频信号中的当前帧之后, 上述方法还包括: 根据第一声道频域信号, 获得第一声道噪声功率谱的估计值; 根据第一声道噪声功率谱的估计值, 确定第一初始维纳增益因子; 根据第二声道频域信号, 获得第二声道噪声功率谱的估计值; 根据第二声道噪声功率谱的估计值, 确定第二初始维纳增益因子。

- 15 在一些可能的实施方式中, 第一初始维纳增益因子 $W_{x1}^A(k)$ 满足以下公式:

$$W_{x1}^A(k) = \frac{|X_1(k)|^2 - |\hat{N}_1(k)|^2}{|X_1(k)|^2};$$

第二初始维纳增益因子 $W_{x2}^A(k)$ 满足以下公式:

$$W_{x2}^A(k) = \frac{|X_2(k)|^2 - |\hat{N}_2(k)|^2}{|X_2(k)|^2};$$

20 其中, $|\hat{N}_1(k)|^2$ 为第一声道噪声功率谱的估计值, $|\hat{N}_2(k)|^2$ 为第二声道噪声功率谱的估计值; $X_1(k)$ 为第一声道频域信号, $X_2(k)$ 为第二声道频域信号, k 为频点索引值, $k=0,1,\dots,N_{\text{DFT}}-1$, N_{DFT} 为当前帧在进行时频变换后的频点总数。

又如, 第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一改进维纳增益因子, 第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二改进维纳增益因子; 在获得立体声音频信号中的当前帧之后, 上述方法还包括: 获得上述第一
25 初始维纳增益因子和上述第二初始维纳增益因子; 对第一初始维纳增益因子构建二值掩蔽函数, 获得第一改进维纳增益因子; 对第二初始维纳增益因子构建二值掩蔽函数, 获得第二改进维纳增益因子。

在一些可能的实施方式中, 第一改进维纳增益因子 $W_{x1}^B(k)$ 满足以下公式:

$$W_{x1}^B(k) = \begin{cases} 1 & \text{if } W_{x1}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x1}^A(k) < \mu_0; \end{cases}$$

- 30 第二改进维纳增益因子 $W_{x2}^B(k)$ 满足以下公式:

$$W_{x2}^B(k) = \begin{cases} 1 & \text{if } W_{x2}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x2}^A(k) < \mu_0; \end{cases}$$

其中, μ_0 为维纳增益因子的二值掩蔽门限, $W_{x1}^A(k)$ 为第一维纳增益因子; $W_{x2}^A(k)$ 为第二维纳增益因子。

可选的, $\mu_0 \in [0.5, 0.8]$, 例如, $\mu_0=0.5$ 、 0.66 、 0.75 、 0.8 等。

- 35 在一些可能的实施方式中, 第二加权函数 $\Phi_{\text{new}_2}(k)$ 满足以下公式:

$$\Phi_{\text{new}_2}(k) = \frac{1}{|X_1(k)X_2^*(k)|^\beta} \Gamma^2(k);$$

其中, β 为幅值加权参数, $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值,

$\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$, $X_1(k)$ 为第一声道频域信号, $X_2(k)$ 为第二声道频域信号, $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数, k 为频点索引值, $k=0,1,\dots,N_{\text{DFT}}-1$, N_{DFT} 为当前帧在进行时频变换后的频点总数。

可选的, $\beta \in [0,1]$, 例如, $\beta=0.6$ 、 0.7 、 0.8 等。

- 5 第三方面, 本申请提供一种立体声音频信号时延估计装置, 该装置可以为音频编码装置中的芯片或者片上系统, 还可以为音频编码装置中用于实现第一方面或第一方面的任一可能的实施方式所述的方法的功能模块。举例来说, 该立体声音频信号时延估计装置, 包括: 第一获得模块, 用于获得立体声音频信号的当前帧, 当前帧包括第一声道音频信号和第二声道音频信号; 第一声道间时间差估计模块, 用于如果当前帧所包含的噪声信号的信号类型为相关性噪声信号类型, 则采用第一算法估计第一声道音频信号和第二声道音频信号的声道间时间差; 如果当前帧所包含的噪声信号的信号类型为弥散性噪声信号类型, 则采用第二算法估计第一声道音频信号和第二声道音频信号的声道间时间差; 其中, 第一算法包括采用第一加权函数对当前帧的频域互功率谱加权, 第二算法包括采用第二加权函数对当前帧的频域互功率谱加权, 第一加权函数与第二加权函数的构造因子不同。
- 10 15

- 在一些可能的实施方式中, 上述装置还包括: 噪声相干值计算模块, 用于在第一获得模块获得当前帧之后, 获得当前帧的噪声相干值; 如果噪声相干值大于或者等于预设阈值, 则确定当前帧所包含的噪声信号的信号类型为相关性噪声信号类型; 或者, 如果噪声相干值小于预设阈值, 则确定当前帧所包含的噪声信号的信号类型为弥散性噪声信号类型。
- 20

- 在一些可能的实施方式中, 上述装置还包括: 语音端点检测模块, 用于对当前帧进行语音端点检测, 获得检测结果; 噪声相干值计算模块, 具体用于如果检测结果显示当前帧的信号类型为噪声信号类型, 则计算当前帧的噪声相干值; 或者, 如果检测结果显示当前帧的信号类型为语音信号类型, 则将立体声音频信号中的当前帧的前一帧的噪声相干值确定为当前帧的噪声相干值。
- 25

在本申请中, 语音端点检测模块可以以时域、频域或者时域频域结合的方式计算语音端点检测的值, 对此不做具体限定。

- 在一些可能的实施方式中, 第一声道音频信号为第一声道时域信号, 第二声道音频信号为第二声道时域信号; 第一声道间时间差估计模块, 用于对第一声道时域信号和第二声道时域信号进行时频变换, 以获得第一声道频域信号和第二声道频域信号; 根据所述第一声道频域信号和所述第二声道频域信号, 计算当前帧的频域互功率谱; 采用第一加权函数对频域互功率谱进行加权; 根据加权后的频域互功率谱, 获得声道间时间差的估计值; 其中, 第一加权函数的构造因子包括: 第一声道频域信号对应的维纳增益因子、第二声道频域信号对应的维纳增益因子、幅值加权参数和当前帧的相干平方值。
- 30 35

在一些可能的实施方式中, 第一声道音频信号为第一声道频域信号, 第二声道音频信号为第二声道频域信号; 第一声道间时间差估计模块, 用于根据第一声道频域信号和第二声道频域信号, 计算当前帧的频域互功率谱; 采用第一加权函数对频域互功率谱进行加权; 根据加权后的频域互功率谱, 获得声道间时间差的估计值; 其中, 第

一加权函数的构造因子包括：第一声道频域信号对应的维纳增益因子、第二声道频域信号对应的维纳增益因子、幅值加权参数和当前帧的相干平方值。

在一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式：

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \frac{\Gamma^2(k)}{(1.0-\Gamma^2(k))};$$

- 5 其中， β 为幅值加权参数， $\beta \in [0,1]$ ， $W_{x1}(k)$ 为第一声道频域信号对应的维纳增益因子； $W_{x2}(k)$ 为第二声道频域信号对应的维纳增益因子； $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

- 10 在一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式：

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

- 其中， β 为幅值加权参数， $\beta \in [0,1]$ ， $W_{x1}(k)$ 为第一声道频域信号对应的维纳增益因子； $W_{x2}(k)$ 为第二声道频域信号对应的维纳增益因子； $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

- 20 在一些可能的实施方式中，第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一初始维纳增益因子，第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二初始维纳增益因子；第一声道间时间差估计模块，具体用于在第一获得模块获得当前帧之后，根据第一声道频域信号，获得第一声道噪声功率谱的估计值；根据第一声道噪声功率谱的估计值，确定第一初始维纳增益因子；根据第二声道频域信号，获得第二声道噪声功率谱的估计值；根据第二声道噪声功率谱的估计值，确定第二初始维纳增益因子。

在一些可能的实施方式中，第一初始维纳增益因子 $W_{x1}^A(k)$ 满足以下公式：

$$25 \quad W_{x1}^A(k) = \frac{|X_1(k)|^2 - |\hat{N}_1(k)|^2}{|X_1(k)|^2};$$

第二初始维纳增益因子 $W_{x2}^A(k)$ 满足以下公式：

$$W_{x2}^A(k) = \frac{|X_2(k)|^2 - |\hat{N}_2(k)|^2}{|X_2(k)|^2};$$

- 30 其中， $|\hat{N}_1(k)|^2$ 为第一声道噪声功率谱的估计值， $|\hat{N}_2(k)|^2$ 为第二声道噪声功率谱的估计值； $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

- 35 在一些可能的实施方式中，第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一改进维纳增益因子，第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二改进维纳增益因子；第一声道间时间差估计模块，具体用于在第一获得模块获得当前帧之后，对第一初始维纳增益因子构建二值掩蔽函数，获得第一改进维纳增益因子；对第二初始维纳增益因子构建二值掩蔽函数，获得第二改进维纳增益因子。

在一些可能的实施方式中，第一改进维纳增益因子 $W_{x1}^B(k)$ 满足以下公式：

$$W_{x1}^B(k) = \begin{cases} 1 & \text{if } W_{x1}^A(k) \geq \mu_0 \\ 0 & \text{if } W_{x2}^A(k) < \mu_0 \end{cases};$$

第二改进维纳增益因子 $W_{x2}^B(k)$ 满足以下公式：

$$W_{x2}^B(k) = \begin{cases} 1 & \text{if } W_{x2}^A(k) \geq \mu_0 \\ 0 & \text{if } W_{x2}^A(k) < \mu_0 \end{cases};$$

其中， μ_0 为维纳增益因子的二值掩蔽门限， $W_{x1}^A(k)$ 为第一初始维纳增益因子；

5 $W_{x2}^A(k)$ 为第二初始维纳增益因子。

在一些可能的实施方式中，第一声道音频信号为第一声道时域信号，第二声道音频信号为第二声道时域信号；第一声道间时间差估计模块，具体用于对第一声道时域信号和第二声道时域信号进行时频变换，以获得第一声道频域信号和第二声道频域信号；根据所述第一声道频域信号和所述第二声道频域信号，计算当前帧的频域互功率谱；采用
10 第二加权函数对频域互功率谱进行加权，获得声道间时间差的估计值；其中，第二加权函数的构造因子包括：幅值加权参数和当前帧的相干平方值。

在一些可能的实施方式中，第一声道音频信号为第一声道频域信号，第二声道音频信号为第二声道频域信号；第一声道间时间差估计模块，具体用于根据第一声道频域信号和第二声道频域信号，计算当前帧的频域互功率谱；采用第二加权函数对频域互功率
15 谱进行加权；根据加权后的频域互功率谱，获得声道间时间差的估计值；其中，第二加权函数的构造因子包括：幅值加权参数和当前帧的相干平方值。

在一些可能的实施方式中，第二加权函数 $\Phi_{new_2}(k)$ 满足以下公式：

$$\Phi_{new_2}(k) = \frac{1}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

其中， β 为幅度加权参数， $\beta \in [0,1]$ ， $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道
20 频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

第四方面，本申请提供一种立体声音频信号时延估计装置，该装置可以为音频编
25 码装置中的芯片或者片上系统，还可以为音频编码装置中用于实现第二方面或第二方面的任一可能的实施方式所述的方法的功能模块。举例来说，该立体声音频信号时延估计装置，包括：第二获得模块，用于获得立体声音频信号中的当前帧，当前帧包括第一声道音频信号和第二声道音频信号；第二声道间时间差估计模块，用于根据第一声道音频信号和第二声道音频信号，计算当前帧的频域互功率谱；采用预设加权函数对频域互功率谱进行加权；根据加权后的频域互功率谱，获得第一声道频域信号和
30 第二声道频域信号的声道间时间差的估计值；其中，预设加权函数为第一加权函数或者第二加权函数，第一加权函数与第二加权函数的构造因子不同；第一加权函数的构造因子包括：第一声道频域信号对应的维纳增益因子、第二声道频域信号对应的维纳增益、幅值加权参数和当前帧的相干平方值；第二加权函数的构造因子包括：幅值加权参数和当前帧的相干平方值。

35 在一些可能的实施方式中，第一声道音频信号为第一声道时域信号，第二声道音频信号为第二声道时域信号；第二声道间时间差估计模块，用于对第一声道时域信号和第二声道时域信号进行时频变换，以获得第一声道频域信号和第二声道频域信号；

根据所述第一声道频域信号和所述第二声道频域信号，计算所述当前帧的频域互功率谱。

在一些可能的实施方式中，第一声道音频信号为第一声道频域信号，第二声道音频信号为第二声道频域信号。

5 在一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式：

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \frac{\Gamma^2(k)}{(1.0-\Gamma^2(k))};$$

其中， β 为幅值加权参数， $\beta \in [0,1]$ ， $W_{x1}(k)$ 为第一声道频域信号对应的维纳增益因子； $W_{x2}(k)$ 为第二声道频域信号对应的维纳增益因子； $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， $\Gamma^2(k)$ 为当前帧第 k 个频点的相

10 干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

在一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式：

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

15 其中， β 为幅值加权参数， $\beta \in [0,1]$ ， $W_{x1}(k)$ 为第一声道频域信号对应的维纳增益因子； $W_{x2}(k)$ 为第二声道频域信号对应的维纳增益因子； $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

20 在一些可能的实施方式中，第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一初始维纳增益因子，第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二初始维纳增益因子；第二声道间时间差估计模块，具体用于在第二获得模块获得当前帧之后，根据第一声道频域信号，获得第一声道噪声功率谱的估计值；根据第一声道噪声功率谱的估计值，确定第一初始维纳增益因子；根据第二声道频域信号，获得第二声道噪声功率谱的估计值；根据第二声道噪声功率谱的估计值，确定第

25 二初始维纳增益因子。

在一些可能的实施方式中，第一初始维纳增益因子 $W_{x1}^A(k)$ 满足以下公式：

$$W_{x1}^A(k) = \frac{|X_1(k)|^2 - |\hat{N}_1(k)|^2}{|X_1(k)|^2};$$

第二初始维纳增益因子 $W_{x2}^A(k)$ 满足以下公式：

$$W_{x2}^A(k) = \frac{|X_2(k)|^2 - |\hat{N}_2(k)|^2}{|X_2(k)|^2};$$

30 其中， $|\hat{N}_1(k)|^2$ 为第一声道噪声功率谱的估计值， $|\hat{N}_2(k)|^2$ 为第二声道噪声功率谱的估计值； $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

35 在一些可能的实施方式中，第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一改进维纳增益因子，第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二改进维纳增益因子；第二声道间时间差估计模块，具体用于在第二获得模块获得当前帧之后，获得上述第一初始维纳增益因子和第二初始维纳增益因子；对第一初始维纳增益因子构建二值掩蔽函数，获得第一改进维纳增益因子；对第二初始维纳增益因子构建二值掩蔽函数，获得第二改进维纳增益因子。

在一些可能的实施方式中，第一改进维纳增益因子 $W_{x1}^B(k)$ 满足以下公式：

$$W_{x1}^B(k) = \begin{cases} 1 & \text{if } W_{x1}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x1}^A(k) < \mu_0; \end{cases}$$

第二改进维纳增益因子 $W_{x2}^B(k)$ 满足以下公式：

$$W_{x2}^B(k) = \begin{cases} 1 & \text{if } W_{x2}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x2}^A(k) < \mu_0; \end{cases}$$

- 5 其中， μ_0 为维纳增益因子的二值掩蔽门限， W_{x1}^A 为第一初始维纳增益因子； $W_{x2}^A(k)$ 为第二初始维纳增益因子。

在一些可能的实施方式中，第二加权函数 $\Phi_{new_2}(k)$ 满足以下公式：

$$\Phi_{new_2}(k) = \frac{1}{|X_1(k)X_2^*(k)|^\beta} \Gamma^2(k);$$

- 10 其中， $\beta \in [0,1]$ ， $X_1(k)$ 为第一声道频域信号， $X_2(k)$ 为第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， k 为频点索引值， $k=0,1,\dots,N_{\text{DFT}}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

第五方面，本申请提供一种音频编码装置，包括：相互耦合的非易失性存储器和处理器，处理器调用存储在存储器中的程序代码以执行如上述第一至二方面及其任一项所述的立体声音频信号时延估计方法。

- 15 第六方面，本申请提供一种计算机可读存储介质，计算机可读存储介质存储有指令，当指令在计算机上运行时，用于执行如上述第一至二方面及其任一项所述的立体声音频信号时延估计方法。

- 第七方面，本申请提供一种计算机可读存储介质，包括编码码流，编码码流包括根据如上述第一至二方面及其任一可能的实施方式中所述的立体声音频信号时延估计方法获得的立体声音频信号的声道间时间差。

- 20 第八方面，本申请提供一种计算机程序或计算机程序产品，当计算机程序或计算机程序产品在计算机上被执行时，使得计算机实现如上述第一至二方面及其任一项所述的立体声音频信号时延估计方法。

- 25 应当理解的是，本申请的第四至十方面与本申请的第一至二方面的技术方案一致，各方面及对应的可行实施方式所取得的有益效果相似，不再赘述。

附图说明

下面将对本申请实施例或背景技术中所需要使用的附图进行说明。

- 30 图1为本申请实施例中的频域中参数立体声编解码方法的流程示意图；
图2为本申请实施例中的广义互相关算法的流程示意图；
图3为本申请实施例中的立体声音频信号时延估计方法的流程示意图一；
图4为本申请实施例中的立体声音频信号时延估计方法的流程示意图二；
图5为本申请实施例中的立体声音频信号时延估计方法的流程示意图三；
图6为申请实施例中的立体声音频信号时延估计装置的结构示意图；
35 图7为本申请实施例中的音频编码装置的结构示意图。

具体实施方式

下面结合本申请实施例中的附图对本申请实施例进行描述。以下描述中，参考形成本申请一部分并以说明之方式示出本申请实施例的具体方面或可使用本申请实施例的具体方面的附图。应理解，本申请实施例可在其它方面中使用，并可包括附图中未描绘的结构或逻辑变化。例如，应理解，结合所描述方法的揭示内容可以同样适用于用于执行所述方法的对应设备或系统，且反之亦然。例如，如果描述一个或多个具体方法步骤，则对应的设备可以包含如功能单元等一个或多个单元，来执行所描述的一个或多个方法步骤（例如，一个单元执行一个或多个步骤，或多个单元，其中每一个都执行多个步骤中的一个或多个），即使附图中未明确描述或说明这种一个或多个单元。另一方面，例如，如果基于如功能单元等一个或多个单元描述具体装置，则对应的方法可以包含一个步骤来执行一个或多个单元的功能性（例如，一个步骤执行一个或多个单元的功能性，或多个步骤，其中每一个执行多个单元中一个或多个单元的功能性），即使附图中未明确描述或说明这种一个或多个步骤。进一步，应理解的是，除非另外明确提出，本文中所描述的各示例性实施例和/或方面的特征可以相互组合。

在语音与音频通信系统中，单通道音频越来越无法满足人们的需求，而立体声音频携带了各个声源的位置信息，提高了音频的清晰度、可懂度，也提高了音频的真实感，因此，越来越受到人们的青睐。

而在语音与音频通信系统中，音频编解码技术是一项非常关键的技术，该技术基于听觉模型，用最小的能量感知失真，以尽可能低的编码速率来表达音频信号，以便于音频信号的传输与存储。那么，为了满足对高质量音频的需求，一系列立体声编解码技术也应运而生

其中，最常用的一项立体声编解码技术为参数立体声编解码技术。该技术的理论基础为空间听觉原理。具体来说，在进行音频编码的过程中，将原立体声音频信号转换为一 路单通道信号和一些空间参数来表示，或者将原立体声音频信号转换为一 路单通道信号、一路残差信号和一些空间参数来表示；在进行音频解码的过程中，通过解码的单通道信号和空间参数来重建立体声音频信号，或者通过解码的单通道信号、残差信号和空间参数来重建立体声音频信号。

图 1 为本申请实施例中的频域中参数立体声编解码方法的流程示意图，参见图 1 所示，该流程可以包括：

S101：编码侧对立体声音频信号中当前帧的第一声道音频信号和第二声道音频信号进行时频变换（如离散傅里叶变换（discrete fourier transform, DFT）），得到第一声道频域信号和第二声道频域信号；

首先，需要说明的是，编码侧获得输入的立体声音频信号可以包括两路音频信号，也就是第一声道音频信号和第二声道音频信号（如左声道音频信号和右声道音频信号）；上述立体声音频信号所包含的两路音频信号还可以多声道音频信号中的某两路音频信号或者由多声道音频信号中的多路音频信号联合产生的两路音频信号，对此不做具体限定。

这里，编码侧在对立体声音频信号进行编码时，会进行分帧处理，得到多个音频帧，并逐帧进行处理。

S102: 编码侧对第一声道频域信号和第二声道频域信号提取空间参数、下混信号和残差信号;

上述空间参数可以包括: 通道间相干性 (inter-channel coherence, IC)、通道间幅度差 (inter-channel level difference, ILD)、声道间时间差 (inter-channel time difference, ITD)、
5 通道间相位差 (inter-channel phase difference, IPD) 等。

S103: 编码侧对空间参数、下混信号和残差信号分别进行编码;

S104: 编码侧根据编码后的空间参数、下混信号和残差信号, 生成频域参数立体声比特流;

S105: 编码侧将频域参数立体声比特流发送给解码侧。

10 S106: 解码侧对接收到的频域参数立体声比特流进行解码, 获得对应的空间参数、下混信号和残差信号;

S107: 解码侧将下混信号和残差信号进行频域上混处理, 得到上混信号;

S108: 解码侧将上混信号与空间参数进行合成, 得到频域音频信号;

15 S109: 解码侧结合空间参数对频域音频信号进行时频逆变换 (如离散傅里叶逆变换 (inverse discrete fourier transform, IDFT)), 获得当前帧的第一声道音频信号和第二声道音频信号;

进一步地, 编码侧对立体声音频信号中的每个音频帧执行上述第一至五步, 解码侧对每一帧执行上述第六至第九步, 如此, 解码侧可以获得多个音频帧的第一声道音频信号和第二声道音频信号, 进而获得立体声音频信号的第一声道音频信号和第二声道音频
20 信号。

在上述参数立体声编解码的过程中, 空间参数中的 ILD 和 ITD 蕴含声源的位置信息, 那么, 准确的估计 ILD 和 ITD 对立体声声像及声场的重建至关重要。

在参数立体声编码技术中, 估计 ITD 的方法中最常用的方法可以为广义互相关法, 其具有复杂度低、实时性好、易于实现、不依赖于立体声音频信号的其它先验信息等优点。图 2 为本申请实施例中的广义互相关算法的流程示意图, 参见图 2 所示, 该方法可以包括:
25

S201: 编码侧将立体声音频信号进行 DFT, 获得第一声道频域信号和第二声道频域信号;

30 S202: 编码侧根据第一声道频域信号和第二声道频域信号, 计算两者的频域互功率谱和频域加权函数;

S203: 编码侧采用频域加权函数对频域互功率谱进行加权;

S204: 编码侧对加权后的频域互功率谱进行 IDFT, 得到频域互相关函数;

S205: 编码侧对频域互相关函数进行峰值检测;

S206: 编码侧根据互相关函数的峰值, 确定 ITD 的估计值。

35 在上述广义互相关算法中, 上述第二步中的频域加权函数可以采用如下几种函数。

第一种、上述第二步中的频域加权函数可以如公式 (1) 所示:

$$\Phi_{\text{PHAT}}(k) = \frac{1}{|X_1(k)X_2^*(k)|} \quad (1)$$

其中, $\Phi_{\text{PHAT}}(k)$ 为 PHAT 加权函数, $X_1(k)$ 为第一声道音频信号 $x_1(n)$ 的频域音频信号, 即第一声道频域信号; $X_2(k)$ 为第二声道音频信号 $x_2(n)$ 的频域音频信号, 即第二声道

频域信号； $X_1(k)X_2^*(k)$ 为第一声道和第二声道的互功率谱； k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为当前帧在进行时频变换后的频点总数。

相应地，加权后的广义互相关函数可以如公式（2）所示：

$$G_{x1x2}(n) = \frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} \frac{X_1(k)X_2^*(k)}{|X_1(k)X_2^*(k)|} e^{i2\pi \frac{kn}{N_{DFT}}} \quad (2)$$

- 5 在实际应用中，采用公式（1）所示的频域加权函数和公式（2）所示的加权后的广义互相关函数进行 ITD 估计，可以称为广义互相关-相位变换方法（generalized cross correlation with phase transformation, GCC-PHAT）算法。由于立体声音频信号在不同频点的能量差异极大，能量低的频点受噪声的影响很大，而能量高的频点受噪声的影响较小，那么，在 GCC-PHAT 算法中，互功率谱经过 PHAT 加权函数加权后，各个频点的加权值在广义互相关函数中所占的权重完全相同，导致 GCC-PHAT 算法对噪声信号很敏感，即使在中高信噪比下，GCC-PHAT 算法的性能也会大幅下降。另外，当空间中存在一个或若干个噪声源，即存在竞争性声源时，立体声音频信号中会存在相关性噪声信号，使得当前帧中的目标信号（如语音信号）对应的峰值就会被弱化。那么，在某些情况下，例如相关性噪声信号的能量大于目标信号的能量或噪声源距离传声器更近，相关性噪声信号的峰值会大于目标信号对应的峰值，此时立体声音频信号的 ITD 估计值就是噪声信号的 ITD 估计值，即在相关性噪声存在的情况下，不仅立体声音频信号的 ITD 估计精度会严重下降，而且立体声音频信号的 ITD 估计值会在目标信号的 ITD 的值与噪声信号的 ITD 的值之间不断切换，从而影响编码后立体声音频信号声像的稳定性。

第二种、上述第二步中的频域加权函数还可以如公式（3）所示：

$$20 \quad \Phi_{PHAT-\beta}(k) = \frac{1}{|X_1(k)X_2^*(k)|^\beta} \quad (3)$$

其中， β 为幅值加权参数， $\beta \in [0,1]$ 。

相应地，加权后的广义互相关函数还可以如公式（4）所示：

$$G_{x1x2}(n) = \frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} \frac{X_1(k)X_2^*(k)}{|X_1(k)X_2^*(k)|^\beta} e^{i2\pi \frac{kn}{N_{DFT}}} \quad (4)$$

- 25 在实际应用中，采用公式（3）所示的频域加权函数和公式（4）所示的加权后的广义互相关函数进行 ITD 估计，可以称为 GCC-PHAT- β 算法。由于不同噪声信号类型下 β 的最优值不同，并且最优值之间差异较大，那么，GCC-PHAT- β 算法在不同噪声信号类型下的性能是不同的。而且，在中高信噪比下，GCC-PHAT- β 算法的性能虽然有一定程度的提高，但并不能满足参数立体声编解码技术对 ITD 估计精度的需求。进一步地，在存在相关性噪声的情况下，GCC-PHAT- β 算法的性能也会严重下降。

- 30 第三种、上述第二步中的频域加权函数还可以如公式（5）所示：

$$\Phi_{PHAT-Coh}(k) = \frac{1}{|X_1(k)X_2^*(k)|} \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2 |X_2(k)|^2} \quad (5)$$

其中， $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2 |X_2(k)|^2}$ 。

相应地，加权后的广义互相关函数还可以如公式（6）所示：

$$G_{x1x2}(n) = \frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} \frac{X_1(k)X_2^*(k)}{|X_1(k)X_2^*(k)|} \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2 |X_2(k)|^2} e^{i2\pi \frac{kn}{N_{DFT}}} \quad (6)$$

- 35 在实际应用中，采用公式（5）所示的频域加权函数和公式（6）所示的加权后的广义互相关函数进行 ITD 估计，可以称为 GCC-PHAT-Coh 算法。在某些条件下，立体声音频信号中相关性噪声中大部分频点的相干平方值会大于当前帧中目标信号的相干平方

值，这样就会导致 GCC-PHAT-Coh 算法的性能严重下降。并且，由于立体声音频信号在不同频点的能量差异极大，GCC-PHAT-Coh 算法并未考虑到不同频点的能量差异对算法性能的影响，导致一些条件下 ITD 估计性能不佳。

由上述可知，噪声对于广义互相关算法的性能影响较为严重，导致 ITD 的估计精度严重下降，进而使得参数编解码技术中解码后的立体声音频信号出现声像不准确、不稳定、空间感差、头中效应明显等问题，严重影响编码后立体声音频信号的音质。

为了解决上述问题，本申请实施例提供一种立体声音频信号时延估计方法，该方法可以应用于一音频编码装置，该音频编码装置可以用于涉及立体声及多声道的音视频通信系统中的音频编码部分，也可以用于虚拟现实（virtual reality，VR）应用程序中的音频编码部分。

在实际应用中，上述音频编码装置可以设置于音视频通信系统中的终端，例如，该终端可以是一种向用户提供语音或者数据连通性的设备，例如也可以称为用户设备（user equipment，UE）、移动台（mobile station）、用户单元（subscriber unit）、站台

（STation）或者终端（terminal equipment，TE）等。终端设备可以为蜂窝电话（cellular phone）、个人数字助理（personal digital assistant，PDA）、无线调制解调器（modem）、手持设备（handheld）、膝上型电脑（laptop computer）、无绳电话（cordless phone）、无线本地环路（wireless local loop，WLL）台或者平板电脑（pad）等。随着无线通信技术的发展，可以接入无线通信系统、可以与无线通信系统的网络侧进行通信，或者通过无线通信系统与其它设备进行通信的设备都可以是本申请实施例中的终端设备，譬如，智能交通中的终端和汽车、智能家居中的家用设备、智能电网中的电力抄表仪器、电压监测仪器、环境监测仪器、智能安全网络中的视频监控仪器、收款机等等。终端设备可以是静态固定的，也可以是移动的。

或者，上述音频编码器还可以设置于具有 VR 功能的设备，例如，该设备可以为支持 VR 应用的智能手机、平板电脑、智能电视、笔记本电脑、个人计算机、可穿戴设备（如 VR 眼镜、VR 头盔、VR 帽子）等，还可以设置于与上述具有 VR 功能的设备进行通信的云端服务器等。当然，上述音频编码装置还可以设置于具有存储和/或传输立体声音频信号功能的其他设备上，本申请实施例不做具体限定。

在本申请实施例中，立体声音频信号可以是原始的立体声音频信号（包括左声道音频信号和右声道音频信号），也可以是多声道音频信号中的两路音频信号组成的立体声音频信号，还可以是由多声道音频信号中的多路音频信号联合产生的两路音频信号组成的立体声信号。当然，立体声音频信号还可以存在其他形式，本申请实施例不做具体限定。在下述实施例中，以立体声音频信号为原始的立体声音频信号为例进行说明，立体声音频信号在时域中可以包含左声道时域信号和右声道时域信号，而立体声音频信号在频域中可以包含左声道频域信号和右声道频域信号。那么，下述实施例中的第一声道音频信号可以为左声道音频信号（既可以在时域也可以在频域），第一声道时域信号可以为左声道时域信号，第一声道频域信号可以为左声道频域信号；类似的，第二声道音频信号可以为右声道音频信号（既可以在时域也可以在频域），第二声道时域信号可以为右声道时域信号，第二声道频域信号可以为右声道频域信号。

可选的，上述音频编码装置具体可以为立体声编码装置，该装置可以构成独立的

立体声编码器；也可以为多声道编码器中的核心编码部分，旨在对由多声道音频信号中的多路信号联合产生的两路音频信号所组成的立体声音频信号进行编码。

下面对本申请实施例提供的立体声音频信号时延估计方法进行说明。

首先，对本申请实施例提供的频域加权函数进行说明。

- 5 在本申请实施例中，为了改善广义互相关算法的性能，可以对上述几种算法中的频域加权函数（如上述公式（1）、（3）、（5）所示）进行改进，改进的频域加权函数可以为且不限于如下几种函数。

10 第一种、改进的频域加权函数（即第一加权函数）的构造因子可以包括：左声道维纳增益因子（即第一声道频域信号对应的维纳增益因子）、右声道维纳增益因子（即第二声道频域信号对应的维纳增益因子）和当前帧的相干平方值。

这里，构造因子是指用于构造目标函数的因子或者因式，那么，当目标函数为改进的频域加权函数时，其构造因子可以为用于构造改进的频域加权函数的一个或者多个函数。

在实际应用中，第一种改进的频域加权函数可以如公式（7）所示：

$$15 \quad \Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \frac{\Gamma^2(k)}{(1-\Gamma^2(k))} \quad (7)$$

其中， $\Phi_{new_1}(k)$ 为第一种改进的频域加权函数， β 为幅值加权参数， $\beta \in [0,1]$ ，例如， $\beta=0.6、0.7、0.8$ 等， $W_{x1}(k)$ 为左声道维纳增益因子； $W_{x2}(k)$ 为右声道维纳增益因子； $\Gamma^2(k)$ 为当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ 。

在一些可能的实施例中，第一种改进的频域加权函数还可以如公式（8）所示：

$$20 \quad \Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k) \quad (8)$$

相应的，采用第一种改进的频域加权函数加权后的广义互相关函数还可以如公式（9）所示：

$$G_{x1x2}(n) = \frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} \Phi_{new_1}(k) X_1(k) X_2^*(k) e^{i2\pi \frac{kn}{N_{DFT}}} \quad (9)$$

- 25 在一些可能的实施方式中，上述左声道维纳增益因子可以包括第一初始维纳增益因子和/或第一改进维纳增益因子；上述右声道维纳增益因子可以包括第二初始维纳增益因子和/或第二改进维纳增益因子。

30 在实际应用中，第一初始维纳增益因子可以通过对 $X_1(k)$ 进行噪声功率谱估计来确定。具体来说，当左声道维纳增益因子包括第一初始维纳增益因子时，上述方法还可以包括：首先，音频编码装置可以根据当前帧的左声道频域信号 $X_1(k)$ ，获得当前帧的左声道噪声功率谱的估计值，再根据该左声道噪声功率谱的估计值，确定第一初始维纳增益因子；类似的，第二初始维纳增益因子也可以为通过对 $X_2(k)$ 进行噪声功率谱估计来确定。具体来说，当右声道维纳增益因子包括第二初始维纳增益因子时，首先，音频编码装置可以根据当前帧的右声道频域信号 $X_2(k)$ ，获得当前帧的右声道噪声功率谱的估计值，并根据该右声道噪声功率谱的估计值，确定第二初始维纳增益因子。

- 35 在上述对当前帧的 $X_1(k)$ 和 $X_2(k)$ 进行噪声功率谱估计的过程中，可以采用如最小值统计算法、最小值跟踪算法等算法计算得到。当然，还可以采用其他算法来计算 $X_1(k)$ 和 $X_2(k)$ 的噪声功率谱的估计值，本申请实施例不做具体限定。

举例来说, 上述第一初始维纳增益因子 $W_{x1}^A(k)$ 可以如公式 (10) 所示:

$$W_{x1}^A(k) = \frac{|X_1(k)|^2 - |\hat{N}_1(k)|^2}{|X_1(k)|^2} \quad (10)$$

上述第二初始维纳增益因子 $W_{x2}^A(k)$ 可以如公式 (11) 所示:

$$W_{x2}^A(k) = \frac{|X_2(k)|^2 - |\hat{N}_2(k)|^2}{|X_2(k)|^2} \quad (11)$$

- 5 其中, $|\hat{N}_1(k)|^2$ 为左声道噪声功率谱的估计值, $|\hat{N}_2(k)|^2$ 为右声道噪声功率谱的估计值。

10 在一些可能的实施方式中, 上述左声道维纳增益因子和右声道维纳增益因子除了直接使用第一初始维纳增益因子和第二初始维纳增益因子构造第一种改进的频域加权函数之外, 还可以基于第一初始维纳增益因子和第二初始维纳增益因子构造对应的二值掩蔽函数, 得到上述第一改进维纳增益因子和第二改进维纳增益因子, 使用第一改进维纳增益因子和第二改进维纳增益因子构造的第一种改进的频域加权函数, 可以筛选出受噪声影响比较小的频点, 进而提高立体声音频信号的 ITD 估计精度。

15 那么, 当左声道维纳增益因子包括第一改进维纳增益因子时, 上述方法还可以包括: 音频编码装置在获得第一初始维纳增益因子后, 对第一初始维纳增益因子构造二值掩蔽函数, 获得第一改进维纳增益因子; 类似的, 音频编码装置在获得第二初始维纳增益因子后, 对第二初始维纳增益因子构造二值掩蔽函数, 获得第二改进维纳增益因子。

例如, 第一改进维纳增益因子 $W_{x1}^B(k)$ 可以如公式 (12) 所示:

$$W_{x1}^B(k) = \begin{cases} 1 & \text{if } W_{x1}^A(k) \geq \mu_0 \\ 0 & \text{if } W_{x1}^A(k) < \mu_0 \end{cases} \quad (12)$$

第二改进维纳增益因子 $W_{x2}^B(k)$ 可以如公式 (13) 所示:

$$20 \quad W_{x2}^B(k) = \begin{cases} 1 & \text{if } W_{x2}^A(k) \geq \mu_0 \\ 0 & \text{if } W_{x2}^A(k) < \mu_0 \end{cases} \quad (13)$$

其中, μ_0 为维纳增益因子的二值掩蔽门限, $\mu_0 \in [0.5, 0.8]$, 例如, $\mu_0 = 0.5$ 、0.66、0.75、0.8 等。

25 那么, 由上述可知, 左声道维纳增益因子 $W_{x1}(k)$ 可以包括 $W_{x1}^A(k)$ 和 $W_{x1}^B(k)$, 右声道维纳增益因子 $W_{x2}(k)$ 可以包括 $W_{x2}^A(k)$ 和 $W_{x2}^B(k)$, 那么, 在构造上述第一种改进的频域加权函数如公式 (7) 或 (8) 的过程中, 可以将 $W_{x1}^A(k)$ 和 $W_{x2}^A(k)$ 代入公式 (7) 或 (8), 也可以将 $W_{x1}^B(k)$ 和 $W_{x2}^B(k)$ 代入公式 (7) 或 (8)。

例如, $W_{x1}^A(k)$ 和 $W_{x2}^A(k)$ 代入公式 (7) 后的第一种改进的频域加权函数可以如公式 (14) 所示:

$$\Phi_{new_1}(k) = \frac{W_{x1}^A(k)W_{x2}^A(k)}{|X_1(k)X_2^*(k)|^\beta} \frac{\Gamma^2(k)}{(1-\Gamma^2(k))} \quad (14)$$

30 将 $W_{x1}^B(k)$ 和 $W_{x2}^B(k)$ 代入公式 (7) 后的第一种改进的频域加权函数可以如公式 (15) 所示:

$$\Phi_{new_1}(k) = \frac{W_{x1}^B(k)W_{x2}^B(k)}{|X_1(k)X_2^*(k)|^\beta} \frac{\Gamma^2(k)}{(1-\Gamma^2(k))} \quad (15)$$

35 在本申请实施例中, 如果采用第一种改进的频域加权函数对当前帧的频域互功率谱进行加权, 经过维纳增益因子加权后, 立体声音频信号的频域互功率谱中的相关性噪声成分的权重大幅降低, 残留噪声成分的相关性也会大幅减小, 在大部分情况下, 残留噪声的相干平方值会比立体声音频信号中的目标信号的相干平方值小很多, 这样目标信号

对应的互相关峰值会更加突出，立体声音频信号的 ITD 估计的精度和稳定性会大幅提高。

第二种、改进的频域加权函数（即第二加权函数）的构造因子可以包括：幅值加权参数 β 、当前帧的相干平方值。

5 在实际应用中，第二种改进的频域加权函数可以为如公式（16）所示：

$$\Phi_{new_2}(k) = \frac{1}{|X_1(k)X_2^*(k)|^\beta} \Gamma^2(k) \quad (16)$$

其中， Φ_{new_2} 为第二种改进的频域加权函数， $\beta \in [0,1]$ ，例如， $\beta=0.6、0.7、0.8$ 等。

10 相应的，采用第二种改进的频域加权函数加权后的广义互相关函数可以如公式（17）所示：

$$G_{x1x2}(n) = \frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} \Phi_{new_2}(k) X_1(k) X_2^*(k) e^{i2\pi \frac{kn}{N_{DFT}}} \quad (17)$$

在本申请实施例中，如果采用第二种改进的频域加权函数对当前帧的频域互功率谱进行加权，能够确保能量大的频点及相关性高的频点有较大的权重，能量小的频点或者相关性较小的频点有较小的权重，从而提高立体声音频信号的 ITD 估计的精度。

15 其次，介绍本申请实施例提供的一种立体声音频信号时延估计方法，该方法为基于上述改进的频域加权函数来估计当前帧的 ITD 值。

图 3 为本申请实施例中的立体声音频信号时延估计方法的流程示意图一，参见图 3 中实线所示，该方法可以包括：

S301：获得立体声音频信号中的当前帧；

20 其中，当前帧包括左声道音频信号和右声道音频信号。

音频编码装置获得输入的立体声音频信号，立体声音频信号中可以包括两路音频信号，这两路音频信号可以为时域音频信号也可以为频域音频信号。

一种情况，立体声音频信号中的两路音频信号为时域音频信号，即左声道时域信号和右声道时域信号（即第一声道时域信号和第二声道时域信号）。在这种情况下，
25 上述立体声音频信号可以通过如麦克风、受话器等声音传感器输入的。参见图 3 中虚线所示，在 S301 之后，该方法还可以包括：S302：对左声道时域信号和右声道时域信号进行时频变换。这里，音频编码装置通过 S301 对该时域音频信号进行分帧处理，获得时域中的当前帧，此时，当前帧可以包括左声道时域信号和右声道时域信号。然后，音频编码装置对时域中的当前帧进行时频变换，得到频域中的当前帧，此
30 时，当前帧可以包括左声道频域信号和右声道频域信号（即第一声道频域信号和第二声道频域信号）。

另一种情况，立体声音频信号中的两路音频信号为频域音频信号，即左声道频域信号和右声道频域信号（即第一声道频域信号和第二声道频域信号）。在这种情况下，上述立体声音频信号本身为两路频域音频信号，那么，音频编码装置可以直接通
35 过 S301 在频域中对该立体声音频信号（即频域音频信号）进行分帧处理，获得频域中的当前帧，该当前帧可以包括左声道频域信号和右声道频域信号（即第一声道频域信号和第二声道频域信号）。

需要说明的是，在后续实施例的描述中，如果立体声音频信号为时域音频信号，

则音频编码装置可以对其进行时频变换, 得到对应的频域音频信号, 再在频域中对其进行处理; 而如果立体声音频信号本身为频域音频信号, 则音频编码装置可以直接在频域中对其进行处理。

在实际应用中, 当前帧中经过分帧处理后的左声道时域信号可以记作 $x_1(n)$, 当前帧中经过分帧处理后的右声道时域信号可以记作 $x_2(n)$, n 为采样点。

在一些可能的实施方式中, 在 S301 之后, 音频编码装置还可以对当前帧进行预处理, 例如, 对 $x_1(n)$ 和 $x_2(n)$ 分别进行高通滤波处理, 得到预处理后的左声道时域信号和右声道时域信号, 预处理后的左声道时域信号记作 $x_1^{hp}(n)$, 预处理后的右声道时域信号记作 $x_2^{hp}(n)$ 。可选的, 高通滤波处理可以是截止频率为 20Hz 的无限冲击响应 (infinite impulse response, IIR) 滤波器, 也可是其他类型的滤波器, 本申请实施例不做具体限定。

可选的, 音频编码装置还可以对 $x_1(n)$ 和 $x_2(n)$ 进行时频变换, 获得 $X_1(k)$ 和 $X_2(k)$; 其中, 左声道频域信号可以记作 $X_1(k)$, 右声道频域信号可以记作 $X_2(k)$ 。

这里, 音频编码装置可以采用如 DFT、快速傅里叶变换 (fast fourier transformation, FFT)、修正离散余弦变换 (modified discrete cosine transform, MDCT) 等时频变换算法, 将时域信号变换为频域信号。当然, 音频编码装置还可以采用其他时频变换算法, 本申请实施例不做具体限定。

假设, 采用 DFT 对左右声道的时域信号进行时频变换。具体地, 音频编码装置可以对 $x_1(n)$ 或者 $x_1^{hp}(n)$ 进行 DFT, 得到 $X_1(k)$; 同样的, 音频编码装置可以为对 $x_2(n)$ 或者 $x_2^{hp}(n)$ 进行 DFT, 得到 $X_2(k)$ 。

进一步的, 为了克服频谱混叠的问题, 相邻两帧的 DFT 之间一般都采用叠接相加的方法进行处理, 有时还会对 DFT 的输入信号进行补零。

S303: 根据 $X_1(k)$ 和 $X_2(k)$, 计算当前帧的频域互功率谱;

这里, 当前帧的频域互功率谱可以为如公式 (18) 所示:

$$C_{x_1x_2}(k) = X_1(k)X_2^*(k) \quad (18)$$

其中, $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数。

S304: 采用预设加权函数对频域互功率谱进行加权;

这里, 预设加权函数可以指上述改进的频域加权函数, 即上述实施例中第一种改进的频域加权函数 Φ_{new_1} 或者第二种改进的频域加权函数 Φ_{new_2} 。

S304 可以理解为音频编码装置将改进后的加权函数与频域功率谱相乘, 那么, 加权后的频域互功率谱就可以表示为: $\Phi_{new_1}(k)C_{x_1x_2}(k)$ 或者 $\Phi_{new_2}(k)C_{x_1x_2}(k)$ 。

在本申请实施例中, 在执行 S305 之前, 音频编码装置还可以采用 $X_1(k)$ 和 $X_2(k)$ 计算改进的频域加权函数 (即预设加权函数)。

S305: 对加权后的频域互功率谱进行时频逆变换, 得到互相关函数;

音频编码装置可以采用 S302 中所采用的时频变换算法对应的时频逆变换算法, 将频域互功率谱由频域变换到时域, 获得互相关函数。

这里, $\Phi_{new_1}(k)C_{x_1x_2}(k)$ 对应的互相关函数可以如公式 (19) 所示:

$$G_{x_1x_2}(n) = \frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} \Phi_{new_1}(k)C_{x_1x_2}(k) e^{i2\pi \frac{kn}{N_{DFT}}} \quad (19)$$

或者, $\Phi_{new_2}(k)C_{x_1x_2}(k)$ 对应的互相关函数可以如公式 (20) 所示:

$$G_{x1x2}(n) = \frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} \phi_{new_2}(k) C_{x1x2}(k) e^{i2\pi \frac{kn}{N_{DFT}}} \quad (20)$$

S306: 对互相关函数进行峰值检测;

5 音频编码装置在通过 S306 获得互相关函数之后, 可以根据预设的采样率和声音传感器 (即麦克风、受话器等) 之间的最大距离确定 ITD 的最大值 Δ_{max} (也可以理解为 ITD 估计的时间范围)。例如, Δ_{max} 设定为 5ms 对应的采样点数, 那么, 如果立体声音频信号的采样率为 32kHz, 则 $\Delta_{max}=160$, 即左右两声道的最大延迟点数为 160 个采样点。接着, 音频编码装置在 $n \in [-\Delta_{max}, \Delta_{max}]$ 的范围内寻找 $G_{x1x2}(n)$ 的最大峰值, 该峰值对应的索引值即为当前帧的 ITD 的备选值。

S307: 根据互相关函数的峰值, 计算当前帧的 ITD 的估计值。

10 音频编码装置根据互相关函数的峰值, 确定当前帧的 ITD 的备选值, 再结合当前帧的 ITD 备选值、前一帧的 ITD 值 (即历史信息)、音频拖尾处理参数、前后帧之间的相关程度等边信息, 确定当前帧的 ITD 的估计值, 从而去除时延估计的异常值。

进一步地, 音频编码装置在通过 S307 确定了 ITD 的估计值后, 可以将其进行编码, 写入立体声音频信号的编码码流中。

15 在本申请实施例中, 如果采用第一种改进的频域加权函数对当前帧的频域互功率谱进行加, 经过维纳增益因子加权后, 立体声音频信号的频域互功率谱中的相关性噪声成分的权重大幅降低, 残留噪声成分的相关性也会大幅减小, 在大部分情况下, 残留噪声的相干平方值会比立体声音频信号中的目标信号的相干平方值小很多, 这样目标信号对应的互相关峰值会更加突出, 立体声音频信号的 ITD 估计的精度和稳定性会大幅提高。
20 如果采用第二种改进的频域加权函数对当前帧的频域互功率谱进行加权, 能够确保能量大的频点及相关性高的频点有较大的权重, 能量小的频点或者相关性较小的频点有较小的权重, 从而提高立体声音频信号的 ITD 估计的精度。

再次, 介绍本申请实施例提供的另一种立体声音频信号时延估计方法, 该方法在上述实施例的基础上针对立体声音频信号中不同类型的噪声信号采用不同的算法进行 ITD
25 估计。

图 4 为本申请实施例中的立体声音频信号时延估计方法的流程示意图二, 参见图 4 所示, 该方法可以包括

S401: 获得立体声音频信号的当前帧;

这里, S401 的实施过程参见对 S301 的描述, 在此不做具体限定。

30 S402: 判断当前帧所包含的噪声信号的信号类型; 如果当前帧所包含的噪声信号的信号类型为相关性噪声信号类型, 则执行 S403; 如果当前帧所包含的噪声信号的信号类型为弥散性噪声信号类型, 则执行 S404;

在噪声环境下, 不同的噪声信号类型对于广义互相关算法的影响是不同的, 那么, 为了充分利用各广义互相关算法的性能, 提升 ITD 估计的准确度, 音频编码装置
35 可以判断当前帧中所包含的噪声信号的信号类型, 进而从多个频域加权函数中, 为当前帧确定合适的频域加权函数。

在实际应用中, 上述相关性噪声信号类型是指立体声音频信号的两路音频信号中的噪声信号的相关性超过一定程度的噪声信号类型, 也就是说, 当前帧所包含的噪声信号可以归类为相关性噪声信号; 上述弥散性噪声信号类型是指立体声音频信号的两

路音频信号中的噪声信号的相关性低于一定程度的噪声信号类型，也就是说，当前帧锁包含的噪声信号可以归类为弥散性噪声信号。

在一些可能的实施方式中，当前帧中可能既包含相关性噪声信号又包含弥散性噪声信号，此时，音频编码装置会将两种噪声信号中的主噪声信号的信号类型确定为当前帧所包含的噪声信号的信号类型。

5 在一些可能的实施方式中，音频编码装置可以通过计算当前帧的噪声相干值来确定当前帧所包含的噪声信号的信号类型，那么，S402可以包括：获得当前帧的噪声相干值；如果噪声相干值大于或者等于预设阈值，则表明当前帧所包含的噪声信号有较强的相关性，那么，音频编码装置可以确定当前帧所包含的噪声信号的信号类型为相关性噪声信号类型；如果噪声相干值小于预设阈值，则表明当前帧所包含的噪声信号的相关性较弱，那么，音频编码装置可以确定当前帧所包含的噪声信号的信号类型为弥散性噪声信号类型。

10 这里，噪声相干值的预设阈值为经验值，可以根据ITD估计性能等因素设定，例如，预设阈值设定为0.20、0.25、0.30等，当然，也可以设定为其它合适的值，本申请实施例对此不做具体限定。

15 在实际应用中，音频编码装置计算当前帧的噪声相干值之后，还可以对其进行平滑处理，以减小噪声相干值估计的误差，提高噪声类型的识别准确率。

S403：采用第一算法估计左声道音频信号和右声道音频信号的ITD的值；

20 这里，第一算法可以包括采用第一加权函数对当前帧的频域互功率谱加权；还可以包括对加权后的互相关函数进行峰值检测，并根据加权后的互相关函数的峰值，估计当前帧的ITD的值。

25 音频编码装置在通过S402确定当前帧所包含的噪声信号的信号类型为相关噪声信号类型之后，可以采用第一算法来估计当前帧的ITD的值，例如，音频编码装置选择采用第一加权函数对当前帧的频域互功率谱加权，进而对加权后的互相关函数进行峰值检测，并根据加权后的互相关函数的峰值估计当前帧的ITD的值。

在一些可能的实施例中，第一加权函数可以为上述一个或者多个实施例中的频域加权函数和/或改进的频域加权函数中在相关性噪声条件下性能较佳的一个或者多个加权函数，如公式（3）所示的频域加权函数、如公式（7）、（8）所示的改进的频域加权函数。

30 优选的，第一加权函数可以为上述实施例中所述的第一种改进的频域加权函数，如公式（7）、（8）所示的改进的频域加权函数。

S404：采用第二算法估计左声道音频信号和右声道音频信号的ITD的值。

35 这里，第二算法包括采用第二加权函数对当前帧的频域互功率谱，还可以包括对加权后的互相关函数进行峰值检测，并根据加权后的互相关函数的峰值，估计当前帧的ITD的值。

相应地，音频编码装置在通过S402确定当前帧所包含的噪声信号的信号类型为弥散噪声信号类型时之后，可以采用第二算法来估计当前帧的ITD的值，例如，音频编码装置可以选择采用第二加权函数对当前帧的频域互功率谱加权，进而对加权后的互相关函数进行峰值检测，并根据加权后的互相关函数的峰值，估计当前帧的ITD的

值。

在一些可能的实施例中，第二加权函数可以为上述一个或者多个实施例中的频域加权函数和/或改进的频域加权函数中在弥散性噪声条件下性能较佳的一个或者多个加权函数，如公式（5）所示的频域加权函数、公式（16）所示的改进的频域加权函数。

5

优选的，第二加权函数可以为上述实施例中所述的第二种改进的频域加权函数，也就是公式（16）所示的改进的频域加权函数。

在一些可能的实施方式中，由于立体声音频信号中既包括语音信号又包括噪声信号，所以，在S401分帧处理得到的当前帧所包含的信号类型可能为语音信号，也可能为噪声信号，那么，为了简化处理、进一步提高ITD估计的精确度，在S402之前，上述方法还可以包括：对当前帧进行语音端点检测，获得检测结果；如果检测结果表示当前帧的信号类型为噪声信号类型，则计算当前帧的噪声相干值；如果检测结果表示当前帧的信号类型为语音信号类型，则将立体声音频信号中的当前帧的前一帧的噪声相干值确定为当前帧的噪声相干值。

15

音频编码装置在获得当前帧之后，可以对当前帧进行语音端点检测（voice activity detection, VAD），以区分当前帧的主要信号是语音信号还是噪声信号。如果检测出当前帧包含的是噪声信号，那么，在S402中计算噪声相干值就可以直接计算当前帧的噪声相干值；而如果检测出当前帧包含的是语音信号，那么，在S402中计算噪声相干值就可以将结合历史帧的噪声相干值，如当前帧的前一帧的噪声相干值确定为当前帧的噪声相干值。这里，当前帧的前一帧可能包含的是噪声信号也可能包含的是语音信号，如果前一帧包含的仍为语音信号，则将历史帧中前一个噪声帧的噪声相干值确定为当前帧的噪声相干值。

20

在具体实施过程中，音频编码装置可以采用多种方法来进行VAD；当VAD的值为1时，则表明当前帧的信号类型为语音信号类型；当VAD的值为0时，则表明当前帧的信号类型为噪声信号类型。

25

需要说明的是，在本申请实施例中，音频编码装置可以以时域、频域或者时域频域结合的方式计算VAD的值，对此不做具体限定。

下面通过具体实例来对上述图4所示的立体声音频信号时延估计方法进行说明。

图5为本申请实施例中的立体声音频信号时延估计方法的流程示意图三，该方法可以包括：

30

S501：对立体声音频信号进行分帧处理，获得当前帧的 $x_1(n)$ 和 $x_2(n)$ ；

S502：对 $x_1(n)$ 和 $x_2(n)$ 进行DFT，得到当前帧的 $X_1(k)$ 和 $X_2(k)$ ；

S503：根据当前帧的 $x_1(n)$ 和 $x_2(n)$ 或者 $X_1(k)$ 和 $X_2(k)$ ，计算当前帧的VAD值；若VAD=1，执行S504；若VAD=0，执行S505；

35

这里，参见图5中的虚线所示，S503可以在S501之后执行，也可以在S502之后执行，对此不做具体限定。

S504：根据 $X_1(k)$ 和 $X_2(k)$ ，计算当前帧的噪声相干值 $\Gamma(k)$ ；

S505：将前一帧的 $\Gamma_{m-1}(k)$ 确认为当前帧的 $\Gamma(k)$ ；

这里，当前帧的 $\Gamma(k)$ 也可以表示为 $\Gamma_m(k)$ ，即第m帧的噪声相干值，m为正整数。

S506: 将当前帧的 $\Gamma(k)$ 与预设阈值 Γ_{thres} 进行比较; 如果 $\Gamma(k)$ 大于或者等于 Γ_{thres} , 则执行 S507, 如果 $\Gamma(k)$ 小于 Γ_{thres} , 则执行 S508;

S507: 采用 $\Phi_{new_1}(k)$ 对当前帧的 $C_{x1x2}(k)$ 进行加权, 此时, 加权后的频域互功率谱就可以表示为: $\Phi_{new_1}(k)C_{x1x2}(k)$;

5 S508: 采用 $\Phi_{PHAT-Coh}(k)$ 对当前帧的 $C_{x1x2}(k)$ 进行加权, 此时, 加权后的频域互功率谱就可以表示为: $\Phi_{PHAT-Coh}(k)C_{x1x2}(k)$;

在实际应用中, 在 S506 之后, 如果确定执行 S507 之前, 可以采用当前帧的 $X_1(k)$ 和 $X_2(k)$ 计算当前帧的 $C_{x1x2}(k)$ 和 $\Phi_{new_1}(k)$; 如果确定执行 S508 之前, 可以采用当前帧的 $X_1(k)$ 和 $X_2(k)$ 计算当前帧的 $C_{x1x2}(k)$ 和 $\Phi_{PHAT-Coh}(k)$ 。

10 S509: 对 $\Phi_{new_1}(k)C_{x1x2}(k)$ 或者 $\Phi_{PHAT-Coh}(k)C_{x1x2}(k)$ 进行 IDFT, 得到互相关函数 $G_{x1x2}(n)$;

其中, $G_{x1x2}(n)$ 可以如公式 (6) 或 (9) 所示。

S510: 对 $G_{x1x2}(n)$ 进行峰值检测;

S511: 根据 $G_{x1x2}(n)$ 的峰值, 计算当前帧的 ITD 的估计值。

15 至此, 便完成了对立体声音频信号的 ITD 估计过程。

在一些可能的实施方式中, 上述 ITD 估计方法除了可以应用于参数立体声编解码技术中, 还可以应用于如声源定位、语音增强、语音分离等技术中。

由上述可知, 在本申请实施例中, 音频编码装置通过对包含不同类型噪声的当前帧采用不同的 ITD 估计算法, 大幅提高了弥散性噪声和相关性噪声条件下立体声音频信号的 ITD 估计的精度和稳定性, 减少了立体声下混信号之间的帧间不连续, 同时更好地保持了立体声信号的相位, 编码后的立体声的声像更加准确和稳定, 真实感更强, 提高了编码后立体声信号的听觉质量。

20 基于相同的发明构思, 本申请实施例提供一种立体声音频信号时延估计装置, 该装置可以为音频编码装置中的芯片或者片上系统, 还可以为音频编码装置中用于实现上述实施例中图4所示的立体声音频信号时延估计方法及其任一可能的实施方式所述的方法的功能模块。举例来说, 图6为申请实施例中的音频解码装置的结构示意图, 参见图6中实线所示, 该立体声音频信号时延估计装置600, 包括: 获得模块601, 用于获得立体声音频信号的当前帧, 当前帧包括第一声道音频信号和第二声道音频信号; 声道间时间差估计模块602, 用于如果当前帧所包含的噪声信号的信号类型为相关性噪声信号类型, 则采用第一算法估计第一声道音频信号和第二声道音频信号的声道间时间差; 如果当前帧所包含的噪声信号的信号类型为弥散性噪声信号类型, 则采用第二算法估计第一声道音频信号和第二声道音频信号的声道间时间差; 其中, 第一算法包括采用第一加权函数对当前帧的频域互功率谱加权, 第二算法包括采用第二加权函数对当前帧的频域互功率谱加权, 第一加权函数与第二加权函数的构造因子不同。

35 在本申请实施例中, 获得模块601获得的立体声信号中的当前帧可以是频域音频信号或者时域音频信号。如果当前帧为频域音频信号, 则获得模块601将当前帧传递给声道间时间差估计模块602, 声道间时间差估计模块602可以直接在频域中对当前帧进行处理; 而如果当前帧为时域音频信号, 则获得模块601可以先对时域中的当前帧

进行时频变换，以得到频域中的当前帧，进而获得模块601将频域中的当前帧传递给声道间时间差估计模块602，声道间时间差估计模块602可以在频域中对当前帧进行处理。

5 在一些可能的实施方式中，参见图6中虚线所示，上述装置还包括：噪声相干值计算模块603，用于在获得模块601获得当前帧之后，获得当前帧的噪声相干值；如果噪声相干值大于或者等于预设阈值，则确定当前帧所包含的噪声信号的信号类型为相关性噪声信号类型；或者，如果噪声相干值小于预设阈值，则确定当前帧所包含的噪声信号的信号类型为弥散性噪声信号类型。

10 在一些可能的实施方式中，参见图6中虚线所示，上述装置还包括：语音端点检测模块604，用于对当前帧进行语音端点检测，获得检测结果；噪声相干值计算模块603，具体用于如果检测结果表示当前帧的信号类型为噪声信号类型，则计算当前帧的噪声相干值；或者，如果检测结果表示当前帧的信号类型为语音信号类型，则将立体声音频信号中的当前帧的前一帧的噪声相干值确定为当前帧的噪声相干值。

15 在本申请实施例中，语音端点检测模块604可以以时域、频域或者时域频域结合的方式计算VAD的值，对此不做具体限定。获得模块601可以将当前帧传递给语音端点检测模块604，以对当前帧进行VAD。

20 在一些可能的实施方式中，第一声道音频信号为第一声道时域信号，第二声道音频信号为第二声道时域信号；声道间时间差估计模块602，用于对第一声道时域信号和第二声道时域信号进行时频变换，以获得第一声道频域信号和第二声道频域信号；根据所述第一声道频域信号和所述第二声道频域信号，计算当前帧的频域互功率谱；采用第一加权函数对频域互功率谱进行加权；根据加权后的频域互功率谱，获得声道间时间差的估计值；其中，第一加权函数的构造因子包括：第一声道频域信号对应的维纳增益因子、第二声道频域信号对应的维纳增益因子、幅值加权参数和当前帧的相干平方值。

25 在一些可能的实施方式中，第一声道音频信号为第一声道频域信号，第二声道音频信号为第二声道频域信号；声道间时间差估计模块602，用于根据第一声道频域信号和第二声道频域信号，计算当前帧的频域互功率谱；采用第一加权函数对频域互功率谱进行加权；根据加权后的频域互功率谱，获得声道间时间差的估计值；其中，第一加权函数的构造因子包括：第一声道频域信号对应的维纳增益因子、第二声道频域信号对应的维纳增益因子、幅值加权参数和当前帧的相干平方值。

30 在一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足上述公式(7)。

在另一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足上述公式(8)。

35 在一些可能的实施方式中，第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一初始维纳增益因子，第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二初始维纳增益因子；声道间时间差估计模块602，具体用于在获得模块获得当前帧之后，根据第一声道频域信号，获得第一声道噪声功率谱的估计值；根据第一声道噪声功率谱的估计值，确定上述第一初始维纳增益因子；根据第二声道频域信号，获得第二声道噪声功率谱的估计值；根据第二声道噪声功率谱的估计值，确定上述第二初始维纳增益因子。

在一些可能的实施方式中，第一初始维纳增益因子 $W_{x1}^A(k)$ 满足上述公式（10），第二初始维纳增益因子 $W_{x2}^A(k)$ 满足上述公式（11）。

在一些可能的实施方式中，第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一改进维纳增益因子，第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二改进维纳增益因子；声道间时间差估计模块602，具体用于在获得模块获得当前帧之后，获得上述第一初始维纳增益因子和第二初始维纳增益因子；对上述第一初始维纳增益因子构建二值掩蔽函数，获得第一改进维纳增益因子；对上述第二初始维纳增益因子构建二值掩蔽函数，获得第二改进维纳增益因子。

在一些可能的实施方式中，第一改进维纳增益因子 $W_{x1}^B(k)$ 满足上述公式（12），第二改进维纳增益因子 $W_{x2}^B(k)$ 满足上述公式（13）。

在一些可能的实施方式中，第一声道音频信号为第一声道时域信号，第二声道音频信号为第二声道时域信号；声道间时间差估计模块602，具体用于对第一声道时域信号和第二声道时域信号进行时频变换，以获得第一声道频域信号和第二声道频域信号；根据所述第一声道频域信号和所述第二声道频域信号，计算当前帧的频域互功率谱；采用第二加权函数对频域互功率谱进行加权，获得声道间时间差的估计值；其中，第二加权函数的构造因子包括：幅值加权参数和当前帧的相干平方值。

在一些可能的实施方式中，第一声道音频信号为第一声道频域信号，第二声道音频信号为第二声道频域信号；声道间时间差估计模块602，具体用于根据第一声道频域信号和第二声道频域信号，计算当前帧的频域互功率谱；采用第二加权函数对频域互功率谱进行加权；根据加权后的频域互功率谱，获得声道间时间差的估计值；其中，第二加权函数的构造因子包括：幅值加权参数和当前帧的相干平方值。

在一些可能的实施方式中，第二加权函数 $\Phi_{new_2}(k)$ 满足上述公式（16）。

需要说明的是，获得模块601、声道间时间差估计模块602、噪声相干值计算模块603以及语音端点检测模块604的具体实现过程可参考图4至图5实施例的详细描述，为了说明书的简洁，这里不再赘述。

本申请实施例中提到的获得模块601可以为接收接口、接收电路或者接收器等；声道间时间差估计模块602、噪声相干值计算模块603以及语音端点检测模块604可以作为一个或者多个处理器。

基于相同的发明构思，本申请实施例提供一种立体声音频信号时延估计装置，该装置可以为音频编码装置中的芯片或者片上系统，还可以为音频编码装置中用于实现上述图3所示的立体声音频信号时延估计方法及其任一可能的实施方式所述的方法的功能模块。举例来说，仍参见图6所示，该立体声音频信号时延估计装置600，包括：获得模块601，用于获得立体声音频信号中的当前帧，当前帧包括第一声道音频信号和第二声道音频信号；声道间时间差估计模块602，用于对根据第一声道音频信号和第二声道音频信号，计算当前帧的频域互功率谱；采用预设加权函数对频域互功率谱进行加权；根据加权后的频域互功率谱，获得第一声道频域信号和第二声道频域信号的声道间时间差的估计值。

其中，预设加权函数为第一加权函数或者第二加权函数，第一加权函数与第二加权函数的构造因子不同；第一加权函数的构造因子包括：第一声道频域信号对应的维

纳增益因子、第二声道频域信号对应的维纳增益、幅值加权参数和当前帧的相干平方值；第二加权函数的构造因子包括：幅值加权参数和当前帧的相干平方值。

5 在一些可能的实施方式中，第一声道音频信号为第一声道时域信号，第二声道音频信号为第二声道时域信号；声道间时间差估计模块602，用于对第一声道时域信号和第二声道时域信号进行时频变换，以获得第一声道频域信号和第二声道频域信号；根据所述第一声道频域信号和所述第二声道频域信号计算当前帧的频域互功率谱。

在一些可能的实施方式中，第一声道音频信号为第一声道频域信号，第二声道音频信号为第二声道频域信号。此时，可以直接根据第一声道音频信号和第二声道音频信号来计算当前帧的频域互功率谱。

10 在一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足上述公式(7)。

在另一些可能的实施方式中，第一加权函数 $\Phi_{new_1}(k)$ 满足上述公式(8)。

15 在一些可能的实施方式中，第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一初始维纳增益因子，第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二初始维纳增益因子；声道间时间差估计模块602，具体用于在获得模块601获得当前帧之后，根据第一声道频域信号，获得第一声道噪声功率谱的估计值；根据第一声道噪声功率谱的估计值，确定第一初始维纳增益因子；根据第二声道频域信号，获得第二声道噪声功率谱的估计值；根据第二声道噪声功率谱的估计值，确定第二初始维纳增益因子。

20 在一些可能的实施方式中，第一初始维纳增益因子 $W_{x1}^A(k)$ 满足上述公式(10)，第二初始维纳增益因子 $W_{x2}^A(k)$ 满足上述公式(11)。

25 在一些可能的实施方式中，第一声道频域信号对应的维纳增益因子为第一声道频域信号的第一改进维纳增益因子，第二声道频域信号对应的维纳增益因子为第二声道频域信号的第二改进维纳增益因子；声道间时间差估计模块602，具体用于在获得模块601获得当前帧之后，获得上述第一初始维纳增益因子和第二初始维纳增益因子；对第一初始维纳增益因子构建二值掩蔽函数，获得第一改进维纳增益因子；对第二初始维纳增益因子构建二值掩蔽函数，获得第二改进维纳增益因子。

在一些可能的实施方式中，第一改进维纳增益因子 $W_{x1}^B(k)$ 满足上述公式(12)，第二改进维纳增益因子 $W_{x2}^B(k)$ 满足上述公式(13)。

在一些可能的实施方式中，第二加权函数 $\Phi_{new_2}(k)$ 满足上述公式(16)。

30 需要说明的是，获得模块601和声道间时间差估计模块602的具体实现过程可参考图3的实施例的详细描述，为了说明书的简洁，这里不再赘述。

本申请实施例中提到的获得模块601可以为接收接口、接收电路或者接收器等；声道间时间差估计模块602可以为一个或者多个处理器。

35 基于相同的发明构思，本申请实施例提供一种音频编码装置，该音频编码装置与上述实施例中所所述的音频编码装置一致。图7为本申请实施例中的音频编码装置的结构示意图，参见图7所示，该音频编码装置700包括：相互耦合的非易失性存储器701和处理器702，处理器702调用存储在存储器701中的程序代码以执行如上述图3至图5的立体声音频信号时延估计方法及其任一可能的实施方式所述的方法的操作步骤。

在一些可能的实施方式中，音频编码装置具体可以为立体声编码装置，该装置可

以构成独立的立体声编码器；也可以为多声道编码器中的核心编码部分，旨在对由多声道频域信号中的多路信号联合产生的两路音频信号所组成的立体声音频信号进行编码。

在实际应用中，上述音频编码装置可以采用如可编程器件，如专用集成电路

- 5 (application specific integrated circuit, ASIC)、寄存器转换级电路(register transfer level, RTL)、现场可编程逻辑门阵列(field programmable gate array, FPGA)等实现，当然，还可以采用其他可编程器件实现，本申请实施例不做具体限定。

- 10 基于相同的发明构思，本申请实施例提供一种计算机可读存储介质，计算机可读存储介质存储有指令，当指令在计算机上运行时，用于执行如上述图3至图5的立体声音频信号时延估计方法及其任一可能的实施方式所述的方法的操作步骤。

基于相同的发明构思，本申请实施例提供一种计算机可读存储介质，包括编码码流，编码码流包括根据如上述图3至图5的立体声音频信号时延估计方法及其任一可能的实施方式所述的方法获得的立体声音频信号的声道间时间差。

- 15 基于相同的发明构思，本申请实施例提供一种计算机程序或计算机程序产品，当计算机程序或计算机程序产品在计算机上被执行时，使得计算机实现如上述图3至图5的立体声音频信号时延估计方法及其任一可能的实施方式所述的方法的操作步骤。

- 20 本领域技术人员能够领会，结合本文公开描述的各种说明性逻辑框、模块和算法步骤所描述的功能可以硬件、软件、固件或其任何组合来实施。如果以软件来实施，那么各种说明性逻辑框、模块、和步骤描述的功能可作为一或多个指令或代码在计算机可读媒体上存储或传输，且由基于硬件的处理单元执行。计算机可读媒体可包含计算机可读存储媒体，其对应于有形媒体，例如数据存储媒体，或包括任何促进将计算机程序从一处传送到另一处的媒体(例如，根据通信协议)的通信媒体。以此方式，计算机可读媒体大体上可对应于(1)非暂时性的有形计算机可读存储媒体，或(2)通信媒体，例如信号或载波。数据存储媒体可为可由一或多个计算机或一或多个处理器存取以检索用于实施
- 25 本申请中描述的技术的指令、代码和/或数据结构的任何可用媒体。计算机程序产品可包含计算机可读媒体。

- 30 作为实例而非限制，此类计算机可读存储媒体可包括RAM、ROM、EEPROM、CD-ROM或其它光盘存储装置、磁盘存储装置或其它磁性存储装置、快闪存储器或可用于存储指令或数据结构的形式的所要程序代码并且可由计算机存取的任何其它媒体。并且，任何连接被恰当地称作计算机可读媒体。举例来说，如果使用同轴缆线、光纤缆线、双绞线、数字订户线(DSL)或例如红外线、无线电和微波等无线技术从网站、服务器或其它远程源传输指令，那么同轴缆线、光纤缆线、双绞线、DSL或例如红外线、无线电和微波等无线技术包含在媒体的定义中。但是，应理解，所述计算机可读存储媒体和数据存储媒体并不包括连接、载波、信号或其它暂时媒体，而是实际上针对于非暂时性有形存储媒体。如本文中所使用，磁盘和光盘包含压缩光盘(CD)、激光光盘、光学光盘、数字多功能光盘(DVD)和蓝光光盘，其中磁盘通常以磁性方式再现数据，而光盘利用激光以光学方式再现数据。以上各项的组合也应包含在计算机可读媒体的范围内。

可通过例如一或多个数字信号处理器(DSP)、通用微处理器、专用集成电路(ASIC)、现场可编程逻辑阵列(FPGA)或其它等效集成或离散逻辑电路等一或多个处理器来执行指

令。因此，如本文中所使用的术语“处理器”可指前述结构或适合于实施本文中所描述的技术的任一其它结构中的任一者。另外，在一些方面中，本文中所描述的各种说明性逻辑框、模块、和步骤所描述的功能可以提供于经配置以用于编码和解码的专用硬件和/或软件模块内，或者并入在组合编解码器中。而且，所述技术可完全实施于一或多个电路或

5 逻辑元件中。

本申请的技术可在各种各样的装置或设备中实施，包含无线手持机、集成电路(IC)或一组 IC(例如，芯片组)。本申请中描述各种组件、模块或单元是为了强调用于执行所揭示的技术的装置的功能方面，但未必需要由不同硬件单元实现。实际上，如上文所描述，各种单元可结合合适的软件和/或固件组合在编码解码器硬件单元中，或者通过互操作硬

10 件单元（包含如上文所描述的一或多个处理器）来提供。

在上述实施例 中，对各个实施例的描述各有侧重，某个实施例中 没有详述的部分，可以参见其他实施例的相关描述。

以上所述，仅为本申请示例性的具体实施方式，但本申请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申请揭露的技术范围内，可轻易想到的变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应该以权利要求的保护范围为准。

15

权 利 要 求 书

- 1.一种立体声音频信号时延估计方法，其特征在于，包括：
获得立体声音频信号的当前帧，所述当前帧包括第一声道音频信号和第二声道音频信号；
- 5 如果所述当前帧所包含的噪声信号的信号类型为相关性噪声信号类型，则采用第一算法估计所述第一声道音频信号和所述第二声道音频信号的声道间时间差；
 如果所述当前帧所包含的噪声信号的信号类型为弥散性噪声信号类型，则采用第二算法估计所述第一声道音频信号和所述第二声道音频信号的声道间时间差；
 其中，所述第一算法包括采用第一加权函数对所述当前帧的频域互功率谱加权，
10 所述第二算法包括采用第二加权函数对所述当前帧的频域互功率谱加权，所述第一加权函数与所述第二加权函数的构造因子不同。
- 2.根据权利要求1所述的方法，其特征在于，在所述获得立体声音频信号的当前帧之后，所述方法还包括：
 获得所述当前帧的噪声相干值；
- 15 如果所述噪声相干值大于或者等于预设阈值，则确定所述当前帧所包含的噪声信号的信号类型为所述相关性噪声信号类型；
 如果所述噪声相干值小于所述预设阈值，则确定所述当前帧所包含的噪声信号的信号类型为所述弥散性噪声信号类型。
- 3.根据权利要求2所述的方法，其特征在于，所述获得所述当前帧的噪声相干值，包括：
20 对所述当前帧进行语音端点检测；
 如果检测结果表示所述当前帧的信号类型为噪声信号类型，则计算所述当前帧的噪声相干值；或者，
 如果检测结果表示所述当前帧的信号类型为语音信号类型，则将所述立体声音频信号中的所述当前帧的前一帧的噪声相干值确定为所述当前帧的噪声相干值。
- 25 4.根据权利要求1至3任一项所述的方法，其特征在于，所述第一声道音频信号为第一声道时域信号，所述第二声道音频信号为第二声道时域信号；
 所述采用第一算法估计所述第一声道音频信号和所述第二声道音频信号的声道间时间差，包括：
30 对所述第一声道时域信号和所述第二声道时域信号进行时频变换，以获得第一声道频域信号和第二声道频域信号；
 根据所述第一声道频域信号和所述第二声道频域信号，计算所述当前帧的频域互功率谱；
 采用所述第一加权函数对所述频域互功率谱进行加权；
- 35 根据加权后的频域互功率谱，获得所述声道间时间差的估计值；
 其中，所述第一加权函数的构造因子包括：所述第一声道频域信号对应的维纳增益因子、所述第二声道频域信号对应的维纳增益因子、幅值加权参数和当前帧的相干平方值。
- 5.根据权利要求1至3任一项所述的方法，其特征在于，所述第一声道音频信号为

第一声道频域信号, 所述第二声道音频信号为第二声道频域信号;

所述采用第一算法估计所述第一声道音频信号和所述第二声道音频信号的声道间时间差, 包括:

5 根据所述第一声道频域信号和所述第二声道频域信号, 计算所述当前帧的频域互功率谱;

采用所述第一加权函数对所述频域互功率谱进行加权;

根据加权后的频域互功率谱, 获得所述声道间时间差的估计值;

10 其中, 所述第一加权函数的构造因子包括: 所述第一声道频域信号对应的维纳增益因子、所述第二声道频域信号对应的维纳增益因子、幅值加权参数和当前帧的相干平方值。

6. 根据权利要求 4 或 5 所述的方法, 其特征在于, 所述第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式:

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \frac{\Gamma^2(k)}{(1.0 - \Gamma^2(k))};$$

15 其中, β 为幅值加权参数, $\beta \in [0, 1]$, $W_{x1}(k)$ 为所述第一声道频域信号对应的维纳增益因子; $W_{x2}(k)$ 为所述第二声道频域信号对应的维纳增益因子; $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数, $\Gamma^2(k)$ 为所述当前帧第 k 个频点的相干平方值, $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$, k 为频点索引值, $k=0, 1, \dots, N_{DFT}-1$, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

20 7. 根据权利要求 4 或 5 所述的方法, 其特征在于, 所述第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式:

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

25 其中, β 为幅值加权参数, $\beta \in [0, 1]$, $W_{x1}(k)$ 为所述第一声道频域信号对应的维纳增益因子; $W_{x2}(k)$ 为所述第二声道频域信号对应的维纳增益因子; $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数, $\Gamma^2(k)$ 为所述当前帧第 k 个频点的相干平方值, $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$, k 为频点索引值, $k=0, 1, \dots, N_{DFT}-1$, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

8. 根据权利要求 4 至 7 任一项所述的方法, 其特征在于, 所述第一声道频域信号对应的维纳增益因子为所述第一声道频域信号的第一初始维纳增益因子, 所述第二声道频域信号对应的维纳增益因子为所述第二声道频域信号的第二初始维纳增益因子;

30 在所述获得立体声音频信号的当前帧之后, 所述方法还包括:

根据所述第一声道频域信号, 获得第一声道噪声功率谱的估计值; 根据所述第一声道噪声功率谱的估计值, 确定所述第一初始维纳增益因子;

根据所述第二声道频域信号, 获得第二声道噪声功率谱的估计值; 根据所述第二声道噪声功率谱的估计值, 确定所述第二初始维纳增益因子。

35 9. 根据权利要求 8 所述的方法, 其特征在于, 所述第一初始维纳增益因子 $W_{x1}^A(k)$ 满足以下公式:

$$W_{x1}^A(k) = \frac{|X_1(k)|^2 - |\hat{N}_1(k)|^2}{|X_1(k)|^2};$$

所述第二初始维纳增益因子 $W_{x2}^A(k)$ 满足以下公式:

$$W_{x2}^A(k) = \frac{|X_2(k)|^2 - |\hat{N}_2(k)|^2}{|X_2(k)|^2};$$

其中, $|\hat{N}_1(k)|^2$ 为所述第一声道噪声功率谱的估计值, $|\hat{N}_2(k)|^2$ 为所述第二声道噪声功率谱的估计值; $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, k 为频点索引值, $k=0,1,\dots,N_{\text{DFT}}-1$, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

- 5 10. 根据所述权利要求 4 至 7 任一项所述的方法, 其特征在于, 所述第一声道频域信号对应的维纳增益因子为所述第一声道频域信号的第一改进维纳增益因子, 所述第二声道频域信号对应的维纳增益因子为所述第二声道频域信号的第二改进维纳增益因子;

在所述获得立体声音频信号的当前帧之后, 所述方法还包括:

- 10 获得所述第一声道频域信号的第一初始维纳增益因子和所述第二声道频域信号的第二初始维纳增益因子;

对所述第一初始维纳增益因子构建二值掩蔽函数, 获得所述第一改进维纳增益因子;

- 15 对所述第二初始维纳增益因子构建二值掩蔽函数, 获得所述第二改进维纳增益因子。

11. 根据权利要求 10 所述的方法, 其特征在于, 所述第一改进维纳增益因子 $W_{x1}^B(k)$ 满足以下公式:

$$W_{x1}^B(k) = \begin{cases} 1 & \text{if } W_{x1}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x1}^A(k) < \mu_0; \end{cases}$$

所述第二改进维纳增益因子 $W_{x2}^B(k)$ 满足以下公式:

20
$$W_{x2}^B(k) = \begin{cases} 1^{x1} & \text{if } W_{x2}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x2}^A(k) < \mu_0; \end{cases}$$

其中, μ_0 为维纳增益因子的二值掩蔽门限, $W_{x1}^A(k)$ 为所述第一初始维纳增益因子; $W_{x2}^A(k)$ 为所述第二初始维纳增益因子。

12. 根据权利要求 1 至 11 任一项所述方法, 其特征在于, 所述第一声道音频信号为第一声道时域信号, 所述第二声道音频信号为第二声道时域信号;

- 25 所述采用第二算法估计所述第一声道音频信号和所述第二声道音频信号的声道间时间差, 包括:

对所述第一声道时域信号和所述第二声道时域信号进行时频变换, 以获得第一声道频域信号和第二声道频域信号;

- 30 根据所述第一声道频域信号和所述第二声道频域信号, 计算所述当前帧的频域互功率谱;

采用所述第二加权函数对所述频域互功率谱进行加权, 获得所述声道间时间差的估计值;

其中, 所述第二加权函数的构造因子包括: 幅值加权参数和所述当前帧的相干平方值。

- 35 13. 根据权利要求 1 至 11 任一项所述方法, 其特征在于, 所述第一声道音频信号为第一声道频域信号, 所述第二声道音频信号为第二声道频域信号;

所述采用第二算法估计第一声道音频信号和第二声道音频信号的声道间时间差,

包括：

根据所述第一声道频域信号和所述第二声道频域信号，计算所述当前帧的频域互功率谱；

采用所述第二加权函数对所述频域互功率谱进行加权；

5 根据加权后的频域互功率谱，获得所述声道间时间差的估计值；

其中，所述第二加权函数的构造因子包括：幅值加权参数和当前帧的相干平方值。

14.根据权利要求 12 或 13 所述的方法，其特征在于，所述第二加权函数 $\Phi_{new_2}(k)$ 满足以下公式：

$$10 \quad \Phi_{new_2}(k) = \frac{1}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

其中， β 为幅度加权参数， $\beta \in [0,1]$ ， $X_1(k)$ 为所述第一声道频域信号， $X_2(k)$ 为所述第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， $\Gamma^2(k)$ 为所述当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

15.一种立体声音频信号时延估计方法，其特征在于，包括：

获得立体声音频信号中的当前帧，所述当前帧包括第一声道音频信号和第二声道音频信号；

根据所述第一声道音频信号和所述第二声道音频信号，计算所述当前帧的频域互功率谱；

20 采用预设加权函数对所述频域互功率谱进行加权，所述预设加权函数为第一加权函数或者第二加权函数；

根据加权后的频域互功率谱，获得所述第一声道频域信号和所述第二声道频域信号的声道间时间差的估计值；

25 其中，所述第一加权函数的构造因子包括：所述第一声道频域信号对应的维纳增益因子、所述第二声道频域信号对应的维纳增益、幅值加权参数和所述当前帧的相干平方值；所述第二加权函数的构造因子包括：幅值加权参数和所述当前帧的相干平方值；所述第一加权函数与所述第二加权函数的构造因子不同。

16.根据权利要求15所述的方法，其特征在于，所述第一声道音频信号为第一声道时域信号，所述第二声道音频信号为第二声道时域信号；

30 所述根据所述第一声道音频信号和所述第二声道音频信号，计算所述当前帧的频域互功率谱，包括：

对所述第一声道时域信号和所述第二声道时域信号进行时频变换，以获得第一声道频域信号和第二声道频域信号；

35 根据所述第一声道频域信号和所述第二声道频域信号，计算所述当前帧的频域互功率谱。

17.根据权利要求15所述的方法，其特征在于，所述第一声道音频信号为第一声道频域信号，所述第二声道音频信号为第二声道频域信号。

18.根据权利要求 15 至 16 任一项所述的方法，其特征在于，所述第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式：

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \frac{\Gamma^2(k)}{(1.0-\Gamma^2(k))};$$

其中, β 为幅值加权参数, $\beta \in [0,1]$, $W_{x1}(k)$ 为所述第一声道频域信号对应的维纳增益因子; $W_{x2}(k)$ 为所述第二声道频域信号对应的维纳增益因子; $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数, $\Gamma^2(k)$ 为所述

5 当前帧第 k 个频点的相干平方值, $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$, k 为频点索引值,

$k=0,1,\dots,N_{DFT}-1$, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

19. 根据权利要求 15 至 16 任一项所述的方法, 其特征在于, 所述第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式:

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

10 其中, β 为幅值加权参数, $\beta \in [0,1]$, $W_{x1}(k)$ 为所述第一声道频域信号对应的维纳增益因子; $W_{x2}(k)$ 为所述第二声道频域信号对应的维纳增益因子; $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数, $\Gamma^2(k)$ 为所述当前帧第 k 个频点的相干平方值, $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$, k 为频点索引值, $k=0,1,\dots,N_{DFT}-1$, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

15 20. 根据权利要求 15 至 19 任一项所述的方法, 其特征在于, 所述第一声道频域信号对应的维纳增益因子为所述第一声道频域信号的第一初始维纳增益因子, 所述第二声道频域信号对应的维纳增益因子为所述第二声道频域信号的第二初始维纳增益因子;

所述获得立体声音频信号中的当前帧之后, 所述方法还包括:

20 根据所述第一声道频域信号, 获得第一声道噪声功率谱的估计值; 根据所述第一声道噪声功率谱的估计值, 确定所述第一初始维纳增益因子;

根据所述第二声道频域信号, 获得第二声道噪声功率谱的估计值; 根据所述第二声道噪声功率谱的估计值, 确定所述第二初始维纳增益因子。

21. 根据权利要求 20 所述的方法, 其特征在于, 所述第一初始维纳增益因子 $W_{x1}^A(k)$ 满足以下公式:

$$25 \quad W_{x1}^A(k) = \frac{|X_1(k)|^2 - |\hat{N}_1(k)|^2}{|X_1(k)|^2};$$

所述第二初始维纳增益因子 $W_{x2}^A(k)$ 满足以下公式:

$$W_{x2}^A(k) = \frac{|X_2(k)|^2 - |\hat{N}_2(k)|^2}{|X_2(k)|^2};$$

30 其中, $|\hat{N}_1(k)|^2$ 为所述第一声道噪声功率谱的估计值, $|\hat{N}_2(k)|^2$ 为所述第二声道噪声功率谱的估计值; $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, k 为频点索引值, $k=0,1,\dots,N_{DFT}-1$, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

22. 根据权利要求 15 至 19 任一项所述的方法, 其特征在于, 所述第一声道频域信号对应的维纳增益因子为所述第一声道频域信号的第一改进维纳增益因子, 所述第二声道频域信号对应的维纳增益因子为所述第二声道频域信号的第二改进维纳增益因子;

在所述获得立体声音频信号中的当前帧之后, 所述方法还包括:

35 获得所述第一声道频域信号的第一初始维纳增益因子和所述第二声道频域信号的第二初始维纳增益因子;

对所述第一初始维纳增益因子构建二值掩蔽函数, 获得所述第一改进维纳增益因子;

对所述第二初始维纳增益因子构建二值掩蔽函数, 获得所述第二改进维纳增益因子。

23. 根据权利要求 22 所述的方法, 其特征在于, 所述第一改进维纳增益因子 $W_{x1}^B(k)$ 满足以下公式:

$$5 \quad W_{x1}^B(k) = \begin{cases} 1 & \text{if } W_{x1}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x1}^A(k) < \mu_0; \end{cases}$$

所述第二改进维纳增益因子 $W_{x2}^B(k)$ 满足以下公式:

$$W_{x2}^B(k) = \begin{cases} 1 & \text{if } W_{x2}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x2}^A(k) < \mu_0; \end{cases}$$

其中, μ_0 为维纳增益因子的二值掩蔽门限, $W_{x1}^A(k)$ 为所述第一初始维纳增益因子; $W_{x2}^A(k)$ 为所述第二初始维纳增益因子。

10 24. 根据权利要求 15 至 23 任一项所述的方法, 其特征在于, 所述第二加权函数 $\Phi_{new_2}(k)$ 满足以下公式:

$$\Phi_{new_2}(k) = \frac{1}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

其中, β 为幅值加权参数, $\beta \in [0, 1]$, $W_{x1}(k)$ 为所述第一声道维纳增益因子; $W_{x2}(k)$ 为所述第二声道维纳增益因子; $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数, $\Gamma^2(k)$ 为所述当前帧第 k 个频点的相干平方值, $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$, k 为频点索引值, $k=0, 1, \dots, N_{DFT}-1$, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

25. 一种立体声音频信号时延估计装置, 其特征在于, 包括:

20 第一获得模块, 用于获得立体声音频信号的当前帧, 所述当前帧包括第一声道音频信号和第二声道音频信号;

25 第一声道间时间差估计模块, 用于如果所述当前帧所包含的噪声信号的信号类型为相关性噪声信号类型, 则采用第一算法估计所述第一声道音频信号和所述第二声道音频信号的声道间时间差; 如果所述当前帧所包含的噪声信号的信号类型为弥散性噪声信号类型, 则采用第二算法估计所述第一声道音频信号和所述第二声道音频信号的声道间时间差;

其中, 所述第一算法包括采用第一加权函数对所述当前帧的频域互功率谱加权, 所述第二算法包括采用第二加权函数对所述当前帧的频域互功率谱加权, 所述第一加权函数与所述第二加权函数的构造因子不同。

30 26. 根据权利要求 25 所述的装置, 其特征在于, 所述装置还包括: 噪声相干值计算模块, 用于在所述第一获得模块获得所述当前帧之后, 获得所述当前帧的噪声相干值; 如果所述噪声相干值大于或者等于预设阈值, 则确定所述当前帧所包含的噪声信号的信号类型为所述相关性噪声信号类型; 或者, 如果所述噪声相干值小于所述预设阈值, 则确定所述当前帧所包含的噪声信号的信号类型为所述弥散性噪声信号类型。

35 27. 根据权利要求 26 所述的装置, 其特征在于, 所述装置还包括: 语音端点检测模块, 用于对所述当前帧进行语音端点检测; 所述噪声相干值计算模块, 具体用于如果检测结果显示所述当前帧的信号类型为噪声信号类型, 则计算所述当前帧的噪声相干值; 或者, 如果检测结果显示所述当前帧的信号类型为语音信号类型, 则将所述立

体声音频信号中的所述当前帧的前一帧的噪声相干值确定为所述当前帧的噪声相干值。

28.根据权利要求25至27任一项所述的装置，其特征在于，所述第一声道音频信号为第一声道时域信号，所述第二声道音频信号为第二声道时域信号；所述第一声道间时间差估计模块，用于对所述第一声道时域信号和所述第二声道时域信号进行时频变换，以获得第一声道频域信号和第二声道频域信号；根据所述第一声道频域信号和所述第二声道频域信号，计算所述当前帧的频域互功率谱；采用所述第一加权函数对所述频域互功率谱进行加权；根据加权后的频域互功率谱，获得所述声道间时间差的估计值；其中，所述第一加权函数的构造因子包括：所述第一声道频域信号对应的维纳增益因子、所述第二声道频域信号对应的维纳增益因子、幅值加权参数和当前帧的相干平方值。

29.根据权利要求25至27任一项所述的装置，其特征在于，所述第一声道音频信号为第一声道频域信号，所述第二声道音频信号为第二声道频域信号；所述第一声道间时间差估计模块，用于根据所述第一声道频域信号和所述第二声道频域信号，计算所述当前帧的频域互功率谱；采用所述第一加权函数对所述频域互功率谱进行加权；根据加权后的频域互功率谱，获得所述声道间时间差的估计值；其中，所述第一加权函数的构造因子包括：所述第一声道频域信号对应的维纳增益因子、所述第二声道频域信号对应的维纳增益因子、幅值加权参数和当前帧的相干平方值。

30.根据权利要求28或29所述的装置，其特征在于，所述第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式：

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \frac{\Gamma^2(k)}{(1.0-\Gamma^2(k))};$$

其中， β 为幅值加权参数， $\beta \in [0,1]$ ， $W_{x1}(k)$ 为所述第一声道频域信号对应的维纳增益因子； $W_{x2}(k)$ 为所述第二声道频域信号对应的维纳增益因子； $X_1(k)$ 为所述第一声道频域信号， $X_2(k)$ 为所述第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， $\Gamma^2(k)$ 为所述当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

31.根据权利要求28或29所述的装置，其特征在于，所述第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式：

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

其中， β 为幅值加权参数， $\beta \in [0,1]$ ， $W_{x1}(k)$ 为所述第一声道频域信号对应的维纳增益因子； $W_{x2}(k)$ 为所述第二声道频域信号对应的维纳增益因子； $X_1(k)$ 为所述第一声道频域信号， $X_2(k)$ 为所述第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， $\Gamma^2(k)$ 为所述当前帧第 k 个频点的相干平方值， $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

32.根据权利要求28至31任一项所述的装置，其特征在于，所述第一声道频域信号对应的维纳增益因子为所述第一声道频域信号的所述第一初始维纳增益因子，所述第二声道频域信号对应的维纳增益因子为所述第二声道频域信号的所述第二初始维纳增益因子；

所述第一声道间时间差估计模块,具体用于在所述第一获得模块获得所述当前帧之后,根据所述第一声道频域信号,获得所述第一声道噪声功率谱的估计值;根据所述第一声道噪声功率谱的估计值,确定所述第一初始维纳增益因子;根据所述第二声道频域信号,获得所述第二声道噪声功率谱的估计值;根据所述第二声道噪声功率谱的估计值,

5 确定所述第二初始维纳增益因子。

33.根据权利要求32所述的装置,其特征在于,所述第一初始维纳增益因子 $W_{x1}^A(k)$ 满足以下公式:

$$W_{x1}^A(k) = \frac{|X_1(k)|^2 - |\hat{N}_1(k)|^2}{|X_1(k)|^2};$$

所述第二初始维纳增益因子 $W_{x2}^A(k)$ 满足以下公式:

10
$$W_{x2}^A(k) = \frac{|X_2(k)|^2 - |\hat{N}_2(k)|^2}{|X_2(k)|^2};$$

其中, $|\hat{N}_1(k)|^2$ 为所述第一声道噪声功率谱的估计值, $|\hat{N}_2(k)|^2$ 为所述第二声道噪声功率谱的估计值; $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, k 为频点索引值, $k=0,1,\dots,N_{\text{DFT}}-1$, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

34.根据权利要求28至31任一项所述的装置,其特征在于,所述第一声道频域信号对应的维纳增益因子为所述第一声道频域信号的第一改进维纳增益因子,所述第二声道频域信号对应的维纳增益因子为所述第二声道频域信号的第二改进维纳增益因子;

15

所述第一声道间时间差估计模块,具体用于在所述第一获得模块获得所述当前帧之后,获得所述第一声道频域信号的第一初始维纳增益因子和所述第二声道频域信号的第二初始维纳增益因子;对所述第一初始维纳增益因子构建二值掩蔽函数,获得所述第一改进维纳增益因子;对所述第二初始维纳增益因子构建二值掩蔽函数,获得所述第二改进维纳增益因子。

20

35.根据权利要求34所述的装置,其特征在于,所述第一改进维纳增益因子 $W_{x1}^B(k)$ 满足以下公式:

25
$$W_{x1}^B(k) = \begin{cases} 1 & \text{if } W_{x1}^A(k) \geq \mu_0 \\ 0 & \text{if } W_{x1}^A(k) < \mu_0 \end{cases};$$

所述第二改进维纳增益因子 $W_{x2}^B(k)$ 满足以下公式:

$$W_{x2}^B(k) = \begin{cases} 1 & \text{if } W_{x2}^A(k) \geq \mu_0 \\ 0 & \text{if } W_{x2}^A(k) < \mu_0 \end{cases};$$

其中, μ_0 为维纳增益因子的二值掩蔽门限, $W_{x1}^A(k)$ 为所述第一初始维纳增益因子; $W_{x2}^A(k)$ 为所述第二初始维纳增益因子。

36.根据权利要求25至35任一项所述装置,其特征在于,所述第一声道音频信号为第一声道时域信号,所述第二声道音频信号为第二声道时域信号;所述第一声道间时间差估计模块,具体用于对所述第一声道时域信号和所述第二声道时域信号进行时频变换,以获得第一声道频域信号和第二声道频域信号;根据所述第一声道频域信号和所述第二声道频域信号,计算所述当前帧的频域互功率谱;采用所述第二加权函数对所述频域互功率谱进行加权,获得所述声道间时间差的估计值;其中,所述第二加权函数的构造因子包括:幅值加权参数和所述当前帧的相干平方值。

30

35

37.根据权利要求25至35任一项所述装置,其特征在于,所述第一声道音频信号为

第一声道频域信号, 所述第二声道音频信号为第二声道频域信号; 所述第一声道间时间差估计模块, 具体用于根据所述第一声道频域信号和所述第二声道频域信号, 计算所述当前帧的频域互功率谱; 采用所述第二加权函数对所述频域互功率谱进行加权; 根据加权后的频域互功率谱, 获得所述声道间时间差的估计值; 其中, 所述第二加权函数的构造因子包括: 幅值加权参数和所述当前帧的相干平方值。

38. 根据权利要求 37 所述的装置, 其特征在于, 所述第二加权函数 $\Phi_{new_2}(k)$ 满足以下公式:

$$\Phi_{new_2}(k) = \frac{1}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

其中, β 为幅度加权参数, $\beta \in [0,1]$, $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数, $\Gamma^2(k)$ 为所述当前帧第 k 个频点的相干平方值, $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$, k 为频点索引值, $k=0,1,\dots,N_{DFT}-1$, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

39. 一种立体声音频信号时延估计装置, 其特征在于, 包括:

第二获得模块, 用于获得立体声音频信号中的当前帧, 所述当前帧包括第一声道音频信号和第二声道音频信号;

第二声道间时间差估计模块, 用于根据所述第一声道音频信号和所述第二声道音频信号, 计算所述当前帧的频域互功率谱; 采用预设加权函数对所述频域互功率谱进行加权, 所述预设加权函数为第一加权函数或者第二加权函数; 根据加权后的频域互功率谱, 获得所述第一声道频域信号和所述第二声道频域信号的声道间时间差的估计值;

其中, 所述第一加权函数的构造因子包括: 所述第一声道频域信号对应的维纳增益因子、所述第二声道频域信号对应的维纳增益、幅值加权参数和所述当前帧的相干平方值; 所述第二加权函数的构造因子包括: 幅值加权参数和所述当前帧的相干平方值; 所述第一加权函数与所述第二加权函数的构造因子不同。

40. 根据权利要求 39 所述的装置, 其特征在于, 所述第一声道音频信号为第一声道时域信号, 所述第二声道音频信号为第二声道时域信号; 所述第二声道间时间差估计模块, 用于对所述第一声道时域信号和所述第二声道时域信号进行时频变换, 以获得第一声道频域信号和第二声道频域信号; 根据所述第一声道频域信号和所述第二声道频域信号, 计算所述当前帧的频域互功率谱。

41. 根据权利要求 39 所述的装置, 其特征在于, 所述第一声道音频信号为第一声道频域信号, 所述第二声道音频信号为第二声道频域信号。

42. 根据权利要求 39 至 41 任一项所述的装置, 其特征在于, 所述第一加权函数 $\Phi_{new_1}(k)$ 满足以下公式:

$$\Phi_{new_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \frac{\Gamma^2(k)}{(1.0-\Gamma^2(k))};$$

其中, β 为幅值加权参数, $\beta \in [0,1]$, $W_{x1}(k)$ 为所述第一声道频域信号对应的维纳增益因子; $W_{x2}(k)$ 为所述第二声道频域信号对应的维纳增益因子; $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数, $\Gamma^2(k)$ 为所述当前帧第 k 个频点的相干平方值, $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$, k 为频点索引值, $k=0,1,\dots,N_{DFT}-1$

1, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

43. 根据权利要求 39 至 41 任一项所述的装置, 其特征在于, 所述第一加权函数 $\Phi_{\text{new}_1}(k)$ 满足以下公式:

$$\Phi_{\text{new}_1}(k) = \frac{W_{x1}(k)W_{x2}(k)}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

5 其中, β 为幅值加权参数, $\beta \in [0,1]$, $W_{x1}(k)$ 为所述第一声道频域信号对应的维纳增益因子; $W_{x2}(k)$ 为所述第二声道频域信号对应的维纳增益因子; $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数, $\Gamma^2(k)$ 为所述当前帧第 k 个频点的相干平方值, $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$, k 为频点索引值, $k=0,1,\dots,N_{\text{DFT}}-1$, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

10 44. 根据权利要求 39 至 43 任一项所述的装置, 其特征在于, 所述第一声道频域信号对应的维纳增益因子为所述第一声道频域信号的第一初始维纳增益因子, 所述第二声道频域信号对应的维纳增益因子为所述第二声道频域信号的第二初始维纳增益因子;

所述第二声道间时间差估计模块, 具体用于在所述第二获得模块获得所述当前帧之后, 根据所述第一声道频域信号, 获得第一声道噪声功率谱的估计值; 根据所述第一声道噪声功率谱的估计值, 确定所述第一初始维纳增益因子; 根据所述第二声道频域信号, 获得第二声道噪声功率谱的估计值; 根据所述第二声道噪声功率谱的估计值, 确定所述第二初始维纳增益因子。

45. 根据权利要求 44 所述的装置, 其特征在于, 所述第一初始维纳增益因子 $W_{x1}^A(k)$ 满足以下公式:

20
$$W_{x1}^A(k) = \frac{|X_1(k)|^2 - |\hat{N}_1(k)|^2}{|X_1(k)|^2};$$

所述第二初始维纳增益因子 $W_{x2}^A(k)$ 满足以下公式:

$$W_{x2}^A(k) = \frac{|X_2(k)|^2 - |\hat{N}_2(k)|^2}{|X_2(k)|^2};$$

25 其中, $|\hat{N}_1(k)|^2$ 为所述第一声道噪声功率谱的估计值, $|\hat{N}_2(k)|^2$ 为所述第二声道噪声功率谱的估计值; $X_1(k)$ 为所述第一声道频域信号, $X_2(k)$ 为所述第二声道频域信号, k 为频点索引值, $k=0,1,\dots,N_{\text{DFT}}-1$, N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

46. 根据权利要求 39 至 43 任一项所述的装置, 其特征在于, 所述第一声道频域信号对应的维纳增益因子为所述第一声道频域信号的第一改进维纳增益因子, 所述第二声道频域信号对应的维纳增益因子为所述第二声道频域信号的第二改进维纳增益因子;

30 所述第二声道间时间差估计模块, 具体用于所述第二获得模块获得所述当前帧之后, 获得所述第一声道频域信号的第一初始维纳增益因子和所述第二声道频域信号的第二初始维纳增益因子; 对所述第一初始维纳增益因子构建二值掩蔽函数, 获得所述第一改进维纳增益因子; 对所述第二初始维纳增益因子构建二值掩蔽函数, 获得所述第二改进维纳增益因子。

35 47. 根据权利要求 46 所述的装置, 其特征在于, 所述第一改进维纳增益因子 $W_{x1}^B(k)$ 满足以下公式:

$$W_{x1}^B(k) = \begin{cases} 1 & \text{if } W_{x1}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x1}^A(k) < \mu_0; \end{cases}$$

所述第二改进维纳增益因子 $W_{x2}^B(k)$ 满足以下公式:

$$W_{x_2}^B(k) = \begin{cases} 1 & \text{if } W_{x_2}^A(k) \geq \mu_0; \\ 0 & \text{if } W_{x_2}^A(k) < \mu_0; \end{cases}$$

其中， μ_0 为维纳增益因子的二值掩蔽门限， $W_{x_1}^A$ 为所述第一初始维纳增益因子； $W_{x_2}^A(k)$ 为所述第二初始维纳增益因子。

- 48.根据权利要求 39 至 47 任一项所述的装置，其特征在于，所述第二加权函数
5 $\Phi_{new_2}(k)$ 满足以下公式：

$$\Phi_{new_2}(k) = \frac{1}{|X_1(k)X_2^*(k)|^\beta} \times \Gamma^2(k);$$

- 其中， $\beta \in [0,1]$ ， $X_1(k)$ 为所述第一声道频域信号， $X_2(k)$ 为所述第二声道频域信号， $X_2^*(k)$ 为 $X_2(k)$ 的共轭函数， $\Gamma^2(k)$ 为所述当前帧第 k 个频点的相干平方值，
10 $\Gamma^2(k) = \frac{|X_1(k)X_2^*(k)|^2}{|X_1(k)|^2|X_2(k)|^2}$ ， k 为频点索引值， $k=0,1,\dots,N_{DFT}-1$ ， N_{DFT} 为所述当前帧在进行时频变换后的频点总数。

49.一种音频编码装置，其特征在于，包括：相互耦合的非易失性存储器和处理器，所述处理器调用存储在所述存储器中的程序代码以执行如权利要求1至24任一项所描述的立体声音频信号时延估计方法。

- 50.一种计算机存储介质，其特征在于，包括计算机程序，所述计算机程序在计算机
15 上被执行时，使得所述计算机执行如权利要求1至24任一项所描述的立体声音频信号时延估计方法。

51.一种计算机可读存储介质，其特征在于，包括编码码流，所述编码码流包括根据如权利要求 1 至 24 任一项所描述的立体声音频信号时延估计方法获得的立体声音频信号的声道间时间差。