Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3068655 C 2022/06/14

(11)(21) 3 068 655

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2018/06/11
(87) Date publication PCT/PCT Publication Date: 2019/01/03
(45) Date de délivrance/Issue Date: 2022/06/14
(85) Entrée phase nationale/National Entry: 2019/12/30
(86) N° demande PCT/PCT Application No.: CN 2018/090631
(87) N° publication PCT/PCT Publication No.: 2019/001252
(30) Priorité/Priority: 2017/06/29 (CN201710515887.1)

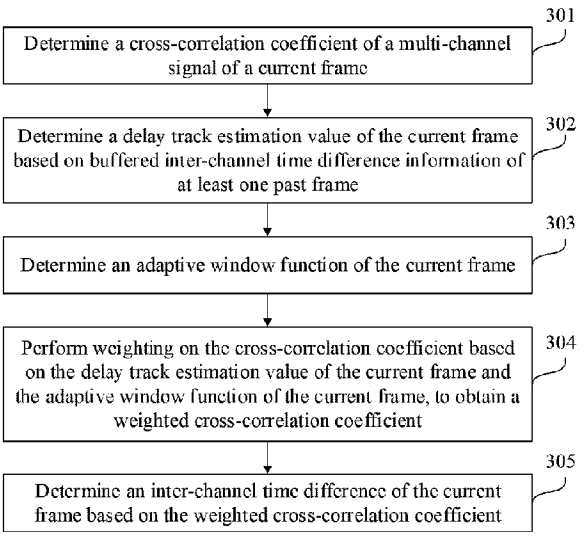
(51) Cl.Int./Int.Cl. *G10L 19/008* (2013.01)

(72) Inventeurs/Inventors:
SHLOMOT, EYAL, US;
LI, HAITING, CN;
MIAO, LEI, CN

(73) Propriétaire/Owner:
HUAWEI TECHNOLOGIES CO., LTD., CN

(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : PROCEDE ET DISPOSITIF D'ESTIMATION DE RETARD TEMPOREL
(54) Title: DELAY ESTIMATION METHOD AND APPARATUS



(57) Abrégé/Abstract:
This application discloses a delay estimation method and apparatus, and belongs to the audio processing field. The method includes: determining a cross- correlation coefficient of a multi-channel signal of a current frame; determining a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame; determining an adaptive window function of the current frame; performing weighting on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame, to obtain a weighted cross-correlation coefficient; and determining an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient, so as to resolve a problem that the cross-correlation coefficient is excessively smoothed or insufficiently smoothed, thereby improving accuracy of estimating an inter-channel time difference.

ABSTRACT

This application discloses a delay estimation method and apparatus, and belongs to the audio processing field. The method includes: determining a cross-correlation coefficient of a multi-channel signal of a current frame; determining a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame; determining an adaptive window function of the current frame; performing weighting on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame, to obtain a weighted cross-correlation coefficient; and determining an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient, so as to resolve a problem that the cross-correlation coefficient is excessively smoothed or insufficiently smoothed, thereby improving accuracy of estimating an inter-channel time difference.

Date Reçue/Date Received 2020-06-23

DELAY ESTIMATION METHOD AND APPARATUS

TECHNICAL FIELD

[0001] This application relates to the audio processing field, and in particular, to a delay estimation method and apparatus.

BACKGROUND

[0002] Compared with a mono signal, thanks to directionality and spaciousness, a multi-channel signal (such as a stereo signal) is favored by people. The multi-channel signal includes at least two mono signals. For example, the stereo signal includes two mono signals, namely, a left channel signal and a right channel signal. Encoding the stereo signal may be performing time-domain downmixing processing on the left channel signal and the right channel signal of the stereo signal to obtain two signals, and then encoding the obtained two signals. The two signals are a primary channel signal and a secondary channel signal. The primary channel signal is used to represent information about correlation between the two mono signals of the stereo signal. The secondary channel signal is used to represent information about a difference between the two mono signals of the stereo signal.

[0003] A smaller delay between the two mono signals indicates a stronger primary channel signal, higher coding efficiency of the stereo signal, and better encoding and decoding quality. On the contrary, a greater delay between the two mono signals indicates a stronger secondary channel signal, lower coding efficiency of the stereo signal, and worse encoding and decoding quality. To ensure a better effect of a stereo signal obtained through encoding and decoding, the delay between the two mono signals of the stereo signal, namely, an inter-channel time difference (ITD, Inter-channel Time Difference), needs to be estimated. The two mono signals are aligned by

Date Reçue/Date Received 2020-06-23

performing delay alignment processing is performed based on the estimated inter-channel time difference, and this enhances the primary channel signal.

[0004] A typical time-domain delay estimation method includes: performing smoothing processing on a cross-correlation coefficient of a stereo signal of a current frame based on a cross-correlation coefficient of at least one past frame, to obtain a smoothed cross-correlation coefficient, searching the smoothed cross-correlation coefficient for a maximum value, and determining an index value corresponding to the maximum value as an inter-channel time difference of the current frame. A smoothing factor of the current frame is a value obtained through adaptive adjustment based on energy of an input signal or another feature. The cross-correlation coefficient is used to indicate a degree of cross correlation between two mono signals after delays corresponding to different inter-channel time differences are adjusted. The cross-correlation coefficient may also be referred to as a cross-correlation function.

[0005] A uniform standard (the smoothing factor of the current frame) is used for an audio coding device, to smooth all cross-correlation values of the current frame. This may cause some cross-correlation values to be excessively smoothed, and/or cause other cross-correlation values to be insufficiently smoothed.

SUMMARY

[0006] To resolve a problem that an inter-channel time difference estimated by an audio coding device is inaccurate due to excessive smoothing or insufficient smoothing performed on a cross-correlation value of a cross-correlation coefficient of a current frame by the audio coding device, embodiments of this application provide a delay estimation method and apparatus.

[0007] According to a first aspect, a delay estimation method is provided. The method includes: determining a cross-correlation coefficient of a multi-channel signal of a current frame; determining a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame; determining an adaptive window function of the current frame; performing weighting

Date Reçue/Date Received 2020-06-23

on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame, to obtain a weighted cross-correlation coefficient; and determining an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient.

[0008] The inter-channel time difference of the current frame is predicted by calculating the delay track estimation value of the current frame, and weighting is performed on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame. The adaptive window function is a raised cosine-like window, and has a function of relatively enlarging a middle part and suppressing an edge part. Therefore, when weighting is performed on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame, if an index value is closer to the delay track estimation value, a weighting coefficient is greater, avoiding a problem that a first cross-correlation coefficient is excessively smoothed, and if the index value is farther from the delay track estimation value, the weighting coefficient is smaller, avoiding a problem that a second cross-correlation coefficient is insufficiently smoothed. In this way, the adaptive window function adaptively suppresses a cross-correlation value corresponding to the index value, away from the delay track estimation value, in the cross-correlation coefficient, thereby improving accuracy of determining the inter-channel time difference in the weighted cross-correlation coefficient. The first cross-correlation coefficient is a cross-correlation value corresponding to an index value, near the delay track estimation value, in the cross-correlation coefficient, and the second cross-correlation coefficient is a cross-correlation value corresponding to an index value, away from the delay track estimation value, in the cross-correlation coefficient.

[0009] With reference to the first aspect, in a first implementation of the first aspect, the determining an adaptive window function of the current frame includes: determining the adaptive window function of the current frame based on a smoothed inter-channel time difference estimation deviation of an $(n-k)^{th}$ frame, where $0 < k < n$, and the current frame is an n^{th} frame.

Date Reçue/Date Received 2020-06-23

[0010] The adaptive window function of the current frame is determined by using the smoothed inter-channel time difference estimation deviation of the $(n - k)^{th}$ frame, so that a shape of the adaptive window function is adjusted based on the smoothed inter-channel time difference estimation deviation, thereby avoiding a problem that a generated adaptive window function is inaccurate due to an error of the delay track estimation of the current frame, and improving accuracy of generating an adaptive window function.

[0011] With reference to the first aspect or the first implementation of the first aspect, in a second implementation of the first aspect, the determining an adaptive window function of the current frame includes: calculating a first raised cosine width parameter based on a smoothed inter-channel time difference estimation deviation of a previous frame of the current frame; calculating a first raised cosine height bias based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame; and determining the adaptive window function of the current frame based on the first raised cosine width parameter and the first raised cosine height bias.

[0012] A multi-channel signal of the previous frame of the current frame has a strong correlation with the multi-channel signal of the current frame. Therefore, the adaptive window function of the current frame is determined based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, thereby improving accuracy of calculating the adaptive window function of the current frame.

[0013] With reference to the second implementation of the first aspect, in a third implementation of the first aspect, a formula for calculating the first raised cosine width parameter is as follows:

$$win\_width1 = TRUNC(width\_par1 * (A * L\_NCSHIFT\_DS + 1)), \text{ and}$$

$$width\_par1 = a\_width1 * smooth\_dist\_reg + b\_width1; \text{ where}$$

$$a\_width1 = (xh\_width1 - xl\_width1)/(yh\_dist1 - yl\_dist1),$$

$$b\_width1 = xh\_width1 - a\_width1 * yh\_dist1,$$

[0014] win_width1 is the first raised cosine width parameter, TRUNC indicates

Date Reçue/Date Received 2020-06-23

rounding a value, L_NCSHIFT_DS is a maximum value of an absolute value of an inter-channel time difference, A is a preset constant, A is greater than or equal to 4, xh_width1 is an upper limit value of the first raised cosine width parameter, xl_width1 is a lower limit value of the first raised cosine width parameter, yh_dist1 is a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first raised cosine width parameter, yl_dist1 is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first raised cosine width parameter, smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, and xh_width1, xl_width1, yh_dist1, and yl_dist1 are all positive numbers.

[0015] With reference to the third implementation of the first aspect, in a fourth implementation of the first aspect,

$$width\_par1 = \min(width\_par1, xh\_width1); \text{ and}$$

$$width\_par1 = \max(width\_par1, xl\_width1), \text{ where}$$

min represents taking of a minimum value, and max represents taking of a maximum value.

[0016] When width_par1 is greater than the upper limit value of the first raised cosine width parameter, width_par1 is limited to be the upper limit value of the first raised cosine width parameter; or when width_par1 is less than the lower limit value of the first raised cosine width parameter, width_par1 is limited to the lower limit value of the first raised cosine width parameter, so as to ensure that a value of width_par1 does not exceed a normal value range of the raised cosine width parameter, thereby ensuring accuracy of a calculated adaptive window function.

[0017] With reference to any one of the second implementation to the fourth implementation of the first aspect, in a fifth implementation of the first aspect, a formula for calculating the first raised cosine height bias is as follows:

$$win\_bias1 = a\_bias1 * smooth\_dist\_reg + b\_bias1, \text{ where}$$

$$a\_bias1 = (xh\_bias1 - xl\_bias1)/(yh\_dist2 - yl\_dist2), \text{ and}$$

$$b\_bias1 = xh\_bias1 - a\_bias1 * yh\_dist2.$$

[0018] win_bias1 is the first raised cosine height bias, xh_bias1 is an upper limit

Date Reçue/Date Received 2020-06-23

value of the first raised cosine height bias, xl_bias1 is a lower limit value of the first raised cosine height bias, yh_dist2 is a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first raised cosine height bias, yl_dist2 is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first raised cosine height bias, smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, and yh_dist2, yl_dist2, xh_bias1, and xl_bias1 are all positive numbers.

[0019] With reference to the fifth implementation of the first aspect, in a sixth implementation of the first aspect,

$$win\_bias1 = \min(win\_bias1, xh\_bias1); \text{ and}$$

$$win\_bias1 = \max(win\_bias1, xl\_bias1), \text{ where}$$

min represents taking of a minimum value, and max represents taking of a maximum value.

[0020] When win_bias1 is greater than the upper limit value of the first raised cosine height bias, win_bias1 is limited to be the upper limit value of the first raised cosine height bias; or when win_bias1 is less than the lower limit value of the first raised cosine height bias, win_bias1 is limited to the lower limit value of the first raised cosine height bias, so as to ensure that a value of win_bias1 does not exceed a normal value range of the raised cosine height bias, thereby ensuring accuracy of a calculated adaptive window function.

[0021] With reference to any one of the second implementation to the fifth implementation of the first aspect, in a seventh implementation of the first aspect,

$$yh\_dist2 = yh\_dist1; \text{ and } yl\_dist2 = yl\_dist1.$$

[0022] With reference to any one of the first aspect, and the first implementation to the seventh implementation of the first aspect, in an eighth implementation of the first aspect,

$$\text{when } 0 \le k \le TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width1 - 1,$$

$$loc\_weight\_win(k) = win\_bias1;$$

$$\text{when } TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width1 \le k \le TRUNC(A$$

Date Reçue/Date Received 2020-06-23

* L_NCSHIFT_DS/2) + 2 * win_width1 – 1,

loc_weight_win(k) = 0.5 * (1 + win_bias1) + 0.5 * (1 – win_bias1) * cos(π * (k – TRUNC(A * L_NCSHIFT_DS/2))/(2 * win_width1)); and

when TRUNC(A * L_NCSHIFT_DS/2) + 2 * win_width1 ≤ k ≤ A * L_NCSHIFT_DS,

loc_weight_win(k) = win_bias1.

[0023] loc_weight_win(k) is used to represent the adaptive window function, where k = 0, 1, ..., A * L_NCSHIFT_DS; A is the preset constant and is greater than or equal to 4; L_NCSHIFT_DS is the maximum value of the absolute value of the inter-channel time difference; win_width1 is the first raised cosine width parameter; and win_bias1 is the first raised cosine height bias.

[0024] With reference to any one of the first implementation to the eighth implementation of the first aspect, in a ninth implementation of the first aspect, after the determining an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient, the method further includes: calculating a smoothed inter-channel time difference estimation deviation of the current frame based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, the delay track estimation value of the current frame, and the inter-channel time difference of the current frame.

[0025] After the inter-channel time difference of the current frame is determined, the smoothed inter-channel time difference estimation deviation of the current frame is calculated. When an inter-channel time difference of a next frame is to be determined, the smoothed inter-channel time difference estimation deviation of the current frame can be used, so as to ensure accuracy of determining the inter-channel time difference of the next frame.

[0026] With reference to the ninth implementation of the first aspect, in a tenth implementation of the first aspect, the smoothed inter-channel time difference estimation deviation of the current frame is obtained through calculation by using the following calculation formulas:

smooth_dist_reg_update = (1 – γ) * smooth_dist_reg + γ * dist_reg', and

Date Reçue/Date Received 2020-06-23

$$dist\_reg' = |reg\_prv\_corr - cur\_itd|.$$

[0027] smooth_dist_reg_update is the smoothed inter-channel time difference estimation deviation of the current frame; γ is a first smoothing factor, and $0 < \gamma < 1$; smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame; reg_prv_corr is the delay track estimation value of the current frame; and cur_itd is the inter-channel time difference of the current frame.

[0028] With reference to the first aspect, in an eleventh implementation of the first aspect, an initial value of the inter-channel time difference of the current frame is determined based on the cross-correlation coefficient; the inter-channel time difference estimation deviation of the current frame is calculated based on the delay track estimation value of the current frame and the initial value of the inter-channel time difference of the current frame; and the adaptive window function of the current frame is determined based on the inter-channel time difference estimation deviation of the current frame.

[0029] The adaptive window function of the current frame is determined based on the initial value of the inter-channel time difference of the current frame, so that the adaptive window function of the current frame can be obtained without a need of buffering a smoothed inter-channel time difference estimation deviation of an n^{th} past frame, thereby saving a storage resource.

[0030] With reference to the eleventh implementation of the first aspect, in a twelfth implementation of the first aspect, the inter-channel time difference estimation deviation of the current frame is obtained through calculation by using the following calculation formula:

$$dist\_reg = |reg\_prv\_corr - cur\_itd\_init|.$$

[0031] dist_reg is the inter-channel time difference estimation deviation of the current frame, reg_prv_corr is the delay track estimation value of the current frame, and cur_itd_init is the initial value of the inter-channel time difference of the current frame.

[0032] With reference to the eleventh implementation or the twelfth implementation of the first aspect, in a thirteenth implementation of the first aspect, a second raised cosine width parameter is calculated based on the inter-channel time

Date Reçue/Date Received 2020-06-23

difference estimation deviation of the current frame; a second raised cosine height bias is calculated based on the inter-channel time difference estimation deviation of the current frame; and the adaptive window function of the current frame is determined based on the second raised cosine width parameter and the second raised cosine height bias.

[0033] Optionally, formulas for calculating the second raised cosine width parameter are as follows:

$$win\_width2 = TRUNC(width\_par2 * (A * L\_NCSHIFT\_DS + 1)), \text{ and}$$

$$width\_par2 = a\_width2 * dist\_reg + b\_width2, \text{ where}$$

$$a\_width2 = (xh\_width2 - xl\_width2)/(yh\_dist3 - yl\_dist3), \text{ and}$$

$$b\_width2 = xh\_width2 - a\_width2 * yh\_dist3.$$

[0034] win_width2 is the second raised cosine width parameter, TRUNC indicates rounding a value, L_NCSHIFT_DS is a maximum value of an absolute value of an inter-channel time difference, A is a preset constant, A is greater than or equal to 4, A * L_NCSHIFT_DS + 1 is a positive integer greater than zero, xh_width2 is an upper limit value of the second raised cosine width parameter, xl_width2 is a lower limit value of the second raised cosine width parameter, yh_dist3 is an inter-channel time difference estimation deviation corresponding to the upper limit value of the second raised cosine width parameter, yl_dist3 is an inter-channel time difference estimation deviation corresponding to the lower limit value of the second raised cosine width parameter, dist_reg is the inter-channel time difference estimation deviation, xh_width2, xl_width2, yh_dist3, and yl_dist3 are all positive numbers.

[0035] Optionally, the second raised cosine width parameter meets:

$$width\_par2 = \min(width\_par2, xh\_width2), \text{ and}$$

$$width\_par2 = \max(width\_par2, xl\_width2), \text{ where}$$

min represents taking of a minimum value, and max represents taking of a maximum value.

[0036] When width_par2 is greater than the upper limit value of the second raised cosine width parameter, width_par2 is limited to be the upper limit value of the second raised cosine width parameter; or when width_par2 is less than the lower limit value of

Date Reçue/Date Received 2020-06-23

the second raised cosine width parameter, width_par2 is limited to the lower limit value of the second raised cosine width parameter, so as to ensure that a value of width_par2 does not exceed a normal value range of the raised cosine width parameter, thereby ensuring accuracy of a calculated adaptive window function.

[0037] Optionally, a formula for calculating the second raised cosine height bias is as follows:

$$win\_bias2 = a\_bias2 * dist\_reg + b\_bias2, \text{ where}$$

$$a\_bias2 = (xh\_bias2 - xl\_bias2)/(yh\_dist4 - yl\_dist4), \text{ and}$$

$$b\_bias2 = xh\_bias2 - a\_bias2 * yh\_dist4.$$

[0038] win_bias2 is the second raised cosine height bias, xh_bias2 is an upper limit value of the second raised cosine height bias, xl_bias2 is a lower limit value of the second raised cosine height bias, yh_dist4 is an inter-channel time difference estimation deviation corresponding to the upper limit value of the second raised cosine height bias, yl_dist4 is an inter-channel time difference estimation deviation corresponding to the lower limit value of the second raised cosine height bias, dist_reg is the inter-channel time difference estimation deviation, and yh_dist4, yl_dist4, xh_bias2, and xl_bias2 are all positive numbers.

[0039] Optionally, the second raised cosine height bias meets:

$$win\_bias2 = \min(win\_bias2, xh\_bias2), \text{ and}$$

$$win\_bias2 = \max(win\_bias2, xl\_bias2), \text{ where}$$

min represents taking of a minimum value, and max represents taking of a maximum value.

[0040] When win_bias2 is greater than the upper limit value of the second raised cosine height bias, win_bias2 is limited to be the upper limit value of the second raised cosine height bias; or when win_bias2 is less than the lower limit value of the second raised cosine height bias, win_bias2 is limited to the lower limit value of the second raised cosine height bias, so as to ensure that a value of win_bias2 does not exceed a normal value range of the raised cosine height bias, thereby ensuring accuracy of a calculated adaptive window function.

[0041] Optionally, yh_dist4 = yh_dist3, and yl_dist4 = yl_dist3.

Date Reçue/Date Received 2020-06-23

[0042] Optionally, the adaptive window function is represented by using the following formulas:

when $0 \le k \le TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width2 - 1$,

$$loc\_weight\_win(k) = win\_bias2;$$

when $TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width2 \le k \le TRUNC(A * L\_NCSHIFT\_DS/2) + 2 * win\_width2 - 1$,

$$loc\_weight\_win(k) = 0.5 * (1 + win\_bias2) + 0.5 * (1 - win\_bias2) * \cos(\pi * (k - TRUNC(A * L\_NCSHIFT\_DS/2))/(2 * win\_width2));$$ and

when $TRUNC(A * L\_NCSHIFT\_DS/2) + 2 * win\_width2 \le k \le A * L\_NCSHIFT\_DS$,

$$loc\_weight\_win(k) = win\_bias2.$$

[0043] loc_weight_win(k) is used to represent the adaptive window function, where k = 0, 1, ..., A * L_NCSHIFT_DS; A is the preset constant and is greater than or equal to 4; L_NCSHIFT_DS is the maximum value of the absolute value of the inter-channel time difference; win_width2 is the second raised cosine width parameter; and win_bias2 is the second raised cosine height bias.

[0044] With reference to any one of the first aspect, and the first implementation to the thirteenth implementation of the first aspect, in a fourteenth implementation of the first aspect, the weighted cross-correlation coefficient is represented by using the following formula:

$$c\_weight(x) = c(x) * loc\_weight\_win(x - TRUNC(reg\_prv\_corr) + TRUNC(A * L\_NCSHIFT\_DS/2) - L\_NCSHIFT\_DS).$$

[0045] c_weight(x) is the weighted cross-correlation coefficient; c(x) is the cross-correlation coefficient; loc_weight_win is the adaptive window function of the current frame; TRUNC indicates rounding a value; reg_prv_corr is the delay track estimation value of the current frame; x is an integer greater than or equal to zero and less than or equal to 2 * L_NCSHIFT_DS; and L_NCSHIFT_DS is the maximum value of the absolute value of the inter-channel time difference.

[0046] With reference to any one of the first aspect, and the first implementation to the fourteenth implementation of the first aspect, in a fifteenth implementation of the

Date Reçue/Date Received 2020-06-23

first aspect, before the determining an adaptive window function of the current frame, the method further includes: determining an adaptive parameter of the adaptive window function of the current frame based on a coding parameter of the previous frame of the current frame, where the coding parameter is used to indicate a type of a multi-channel signal of the previous frame of the current frame, or the coding parameter is used to indicate a type of a multi-channel signal of the previous frame of the current frame on which time-domain downmixing processing is performed; and the adaptive parameter is used to determine the adaptive window function of the current frame.

[0047] The adaptive window function of the current frame needs to change adaptively based on different types of multi-channel signals of the current frame, so as to ensure accuracy of an inter-channel time difference of the current frame obtained through calculation. It is of great probability that the type of the multi-channel signal of the current frame is the same as the type of the multi-channel signal of the previous frame of the current frame. Therefore, the adaptive parameter of the adaptive window function of the current frame is determined based on the coding parameter of the previous frame of the current frame, so that accuracy of a determined adaptive window function is improved without additional calculation complexity.

[0048] With reference to any one of the first aspect, and the first implementation to the fifteenth implementation of the first aspect, in a sixteenth implementation of the first aspect, the determining a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame includes: performing delay track estimation based on the buffered inter-channel time difference information of the at least one past frame by using a linear regression method, to determine the delay track estimation value of the current frame.

[0049] With reference to any one of the first aspect, and the first implementation to the fifteenth implementation of the first aspect, in a seventeenth implementation of the first aspect, the determining a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame includes: performing delay track estimation based on the buffered inter-channel time difference information of the at least one past frame by using a weighted linear regression method,

Date Reçue/Date Received 2020-06-23

to determine the delay track estimation value of the current frame.

[0050] With reference to any one of the first aspect, and the first implementation to the seventeenth implementation of the first aspect, in an eighteenth implementation of the first aspect, after the determining an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient, the method further includes: updating the buffered inter-channel time difference information of the at least one past frame, where the inter-channel time difference information of the at least one past frame is an inter-channel time difference smoothed value of the at least one past frame or an inter-channel time difference of the at least one past frame.

[0051] The buffered inter-channel time difference information of the at least one past frame is updated, and when the inter-channel time difference of the next frame is calculated, a delay track estimation value of the next frame can be calculated based on updated delay difference information, thereby improving accuracy of calculating the inter-channel time difference of the next frame.

[0052] With reference to the eighteenth implementation of the first aspect, in a nineteenth implementation of the first aspect, the buffered inter-channel time difference information of the at least one past frame is the inter-channel time difference smoothed value of the at least one past frame, and the updating the buffered inter-channel time difference information of the at least one past frame includes: determining an inter-channel time difference smoothed value of the current frame based on the delay track estimation value of the current frame and the inter-channel time difference of the current frame; and updating a buffered inter-channel time difference smoothed value of the at least one past frame based on the inter-channel time difference smoothed value of the current frame.

[0053] With reference to the nineteenth implementation of the first aspect, in a twentieth implementation of the first aspect, the inter-channel time difference smoothed value of the current frame is obtained by using the following calculation formula:

$$cur\_itd\_smooth = \varphi * reg\_prv\_corr + (1 - \varphi) * cur\_itd.$$

[0054] cur_itd_smooth is the inter-channel time difference smoothed value of the current frame, φ is a second smoothing factor, reg_prv_corr is the delay track estimation

Date Reçue/Date Received 2020-06-23

value of the current frame, cur_itd is the inter-channel time difference of the current frame, and φ is a constant greater than or equal to 0 and less than or equal to 1.

[0055] With reference to any one of the eighteenth implementation to the twentieth implementation of the first aspect, in a twenty-first implementation of the first aspect, the updating the buffered inter-channel time difference information of the at least one past frame includes: when a voice activation detection result of the previous frame of the current frame is an active frame or a voice activation detection result of the current frame is an active frame, updating the buffered inter-channel time difference information of the at least one past frame.

[0056] When the voice activation detection result of the previous frame of the current frame is the active frame or the voice activation detection result of the current frame is the active frame, it indicates that it is of great possibility that the multi-channel signal of the current frame is the active frame. When the multi-channel signal of the current frame is the active frame, validity of inter-channel time difference information of the current frame is relatively high. Therefore, it is determined, based on the voice activation detection result of the previous frame of the current frame or the voice activation detection result of the current frame, whether to update the buffered inter-channel time difference information of the at least one past frame, thereby improving validity of the buffered inter-channel time difference information of the at least one past frame.

[0057] With reference to at least one of the seventeenth implementation to the twenty-first implementation of the first aspect, in a twenty-second implementation of the first aspect, after the determining an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient, the method further includes: updating a buffered weighting coefficient of the at least one past frame, where the weighting coefficient of the at least one past frame is a coefficient in the weighted linear regression method, and the weighted linear regression method is used to determine the delay track estimation value of the current frame.

[0058] When the delay track estimation value of the current frame is determined by using the weighted linear regression method, the buffered weighting coefficient of the

Date Reçue/Date Received 2020-06-23

at least one past frame is updated, so that the delay track estimation value of the next frame can be calculated based on an updated weighting coefficient, thereby improving accuracy of calculating the delay track estimation value of the next frame.

[0059] With reference to the twenty-second implementation of the first aspect, in a twenty-third implementation of the first aspect, when the adaptive window function of the current frame is determined based on a smoothed inter-channel time difference of the previous frame of the current frame, the updating a buffered weighting coefficient of the at least one past frame includes: calculating a first weighting coefficient of the current frame based on the smoothed inter-channel time difference estimation deviation of the current frame; and updating a buffered first weighting coefficient of the at least one past frame based on the first weighting coefficient of the current frame.

[0060] With reference to the twenty-third implementation of the first aspect, in a twenty-fourth implementation of the first aspect, the first weighting coefficient of the current frame is obtained through calculation by using the following calculation formulas:

$$wgt\_par1 = a\_wgt1 * smooth\_dist\_reg\_update + b\_wgt1,$$

$$a\_wgt1 = (xl\_wgt1 - xh\_wgt1)/(yh\_dist1' - yl\_dist1'), \text{ and}$$

$$b\_wgt1 = xl\_wgt1 - a\_wgt1 * yh\_dist1'.$$

[0061] wgt_par1 is the first weighting coefficient of the current frame, smooth_dist_reg_update is the smoothed inter-channel time difference estimation deviation of the current frame, xh_wgt is an upper limit value of the first weighting coefficient, xl_wgt is a lower limit value of the first weighting coefficient, yh_dist1' is a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first weighting coefficient, yl_dist1' is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first weighting coefficient, and yh_dist1', yl_dist1', xh_wgt1, and xl_wgt1 are all positive numbers.

[0062] With reference to the twenty-fourth implementation of the first aspect, in a twenty-fifth implementation of the first aspect,

$$wgt\_par1 = \min(wgt\_par1, xh\_wgt1), \text{ and}$$

Date Reçue/Date Received 2020-06-23

wgt_par1 = max(wgt_par1, xl_wgt1), where

min represents taking of a minimum value, and max represents taking of a maximum value.

[0063] When wgt_par1 is greater than the upper limit value of the first weighting coefficient, wgt_par1 is limited to be the upper limit value of the first weighting coefficient; or when wgt_par1 is less than the lower limit value of the first weighting coefficient, wgt_par1 is limited to the lower limit value of the first weighting coefficient, so as to ensure that a value of wgt_par1 does not exceed a normal value range of the first weighting coefficient, thereby ensuring accuracy of the calculated delay track estimation value of the current frame.

[0064] With reference to the twenty-second implementation of the first aspect, in a twenty-sixth implementation of the first aspect, when the adaptive window function of the current frame is determined based on the inter-channel time difference estimation deviation of the current frame, the updating a buffered weighting coefficient of the at least one past frame includes: calculating a second weighting coefficient of the current frame based on the inter-channel time difference estimation deviation of the current frame; and updating a buffered second weighting coefficient of the at least one past frame based on the second weighting coefficient of the current frame.

[0065] Optionally, the second weighting coefficient of the current frame is obtained through calculation by using the following calculation formulas:

$$wgt\_par2 = a\_wgt2 * dist\_reg + b\_wgt2,$$

$$a\_wgt2 = (xl\_wgt2 - xh\_wgt2)/(yh\_dist2' - yl\_dist2'), \text{ and}$$

$$b\_wgt2 = xl\_wgt2 - a\_wgt2 * yh\_dist2'.$$

[0066] wgt_par2 is the second weighting coefficient of the current frame, dist_reg is the inter-channel time difference estimation deviation of the current frame, xh_wgt2 is an upper limit value of the second weighting coefficient, xl_wgt2 is a lower limit value of the second weighting coefficient, yh_dist2' is an inter-channel time difference estimation deviation corresponding to the upper limit value of the second weighting coefficient, yl_dist2' is an inter-channel time difference estimation deviation corresponding to the lower limit value of the second weighting coefficient, and

Date Reçue/Date Received 2020-06-23

yh_dist2', yl_dist2', xh_wgt2, and xl_wgt2 are all positive numbers.

[0067] Optionally, wgt_par2 = min(wgt_par2, xh_wgt2), and wgt_par2 = max(wgt_par2, xl_wgt2).

[0068] With reference to any one of the twenty-third implementation to the twenty-sixth implementation of the first aspect, in a twenty-seventh implementation of the first aspect, the updating a buffered weighting coefficient of the at least one past frame includes: when a voice activation detection result of the previous frame of the current frame is an active frame or a voice activation detection result of the current frame is an active frame, updating the buffered weighting coefficient of the at least one past frame.

[0069] When the voice activation detection result of the previous frame of the current frame is the active frame or the voice activation detection result of the current frame is the active frame, it indicates that it is of great possibility that the multi-channel signal of the current frame is the active frame. When the multi-channel signal of the current frame is the active frame, validity of a weighting coefficient of the current frame is relatively high. Therefore, it is determined, based on the voice activation detection result of the previous frame of the current frame or the voice activation detection result of the current frame, whether to update the buffered weighting coefficient of the at least one past frame, thereby improving validity of the buffered weighting coefficient of the at least one past frame.

[0070] According to a second aspect, a delay estimation apparatus is provided. The apparatus includes at least one unit, and the at least one unit is configured to implement the delay estimation method provided in any one of the first aspect or the implementations of the first aspect.

[0071] According to a third aspect, an audio coding device is provided. The audio coding device includes a processor and a memory connected to the processor.

[0072] The memory is configured to be controlled by the processor, and the processor is configured to implement the delay estimation method provided in any one of the first aspect or the implementations of the first aspect.

[0073] According to a fourth aspect, a computer readable storage medium is provided. The computer readable storage medium stores an instruction, and when the

Date Reçue/Date Received 2020-06-23

instruction is run on an audio coding device, the audio coding device is enabled to perform the delay estimation method provided in any one of the first aspect or the implementations of the first aspect.

BRIEF DESCRIPTION OF DRAWINGS

[0074] FIG. 1 is a schematic structural diagram of a stereo signal encoding and decoding system according to an example embodiment of this application;

[0075] FIG. 2 is a schematic structural diagram of a stereo signal encoding and decoding system according to another example embodiment of this application;

[0076] FIG. 3 is a schematic structural diagram of a stereo signal encoding and decoding system according to another example embodiment of this application;

[0077] FIG. 4 is a schematic diagram of an inter-channel time difference according to an example embodiment of this application;

[0078] FIG. 5 is a flowchart of a delay estimation method according to an example embodiment of this application;

[0079] FIG. 6 is a schematic diagram of an adaptive window function according to an example embodiment of this application;

[0080] FIG. 7 is a schematic diagram of a relationship between a raised cosine width parameter and inter-channel time difference estimation deviation information according to an example embodiment of this application;

[0081] FIG. 8 is a schematic diagram of a relationship between a raised cosine height bias and inter-channel time difference estimation deviation information according to an example embodiment of this application;

[0082] FIG. 9 is a schematic diagram of a buffer according to an example embodiment of this application;

[0083] FIG. 10 is a schematic diagram of buffer updating according to an example embodiment of this application;

[0084] FIG. 11 is a schematic structural diagram of an audio coding device according to an example embodiment of this application; and

Date Reçue/Date Received 2020-06-23

[0085] FIG. 12 is a block diagram of a delay estimation apparatus according to an embodiment of this application.

DESCRIPTION OF EMBODIMENTS

[0086] The words "first", "second" and similar words mentioned in this specification do not mean any order, quantity or importance, but are used to distinguish between different components. Likewise, "one", "a/an", or the like is not intended to indicate a quantity limitation either, but is intended to indicate existing at least one. "Connection", "link" or the like is not limited to a physical or mechanical connection, but may include an electrical connection, regardless of a direct connection or an indirect connection.

[0087] In this specification, "a plurality of" refers to two or more than two. The term "and/or" describes an association relationship for describing associated objects and represents that three relationships may exist. For example, A and/or B may represent the following three cases: Only A exists, both A and B exist, and only B exists. The character "/" generally indicates an "or" relationship between the associated objects.

[0088] FIG. 1 is a schematic structural diagram of a stereo encoding and decoding system in time domain according to an example embodiment of this application. The stereo encoding and decoding system includes an encoding component 110 and a decoding component 120.

[0089] The encoding component 110 is configured to encode a stereo signal in time domain. Optionally, the encoding component 110 may be implemented by using software, may be implemented by using hardware, or may be implemented in a form of a combination of software and hardware. This is not limited in this embodiment.

[0090] The encoding a stereo signal in time domain by the encoding component 110 includes the following steps:

[0091] (1) Perform time-domain preprocessing on an obtained stereo signal to obtain a preprocessed left channel signal and a preprocessed right channel signal.

[0092] The stereo signal is collected by a collection component and sent to the

Date Reçue/Date Received 2020-06-23

encoding component 110. Optionally, the collection component and the encoding component 110 may be disposed in a same device or in different devices.

[0093] The preprocessed left channel signal and the preprocessed right channel signal are two signals of the preprocessed stereo signal.

[0094] Optionally, the preprocessing includes at least one of high-pass filtering processing, pre-emphasis processing, sampling rate conversion, and channel conversion. This is not limited in this embodiment.

[0095] (2) Perform delay estimation based on the preprocessed left channel signal and the preprocessed right channel signal to obtain an inter-channel time difference between the preprocessed left channel signal and the preprocessed right channel signal.

[0096] (3) Perform delay alignment processing on the preprocessed left channel signal and the preprocessed right channel signal based on the inter-channel time difference, to obtain a left channel signal obtained after delay alignment processing and a right channel signal obtained after delay alignment processing.

[0097] (4) Encode the inter-channel time difference to obtain an encoding index of the inter-channel time difference.

[0098] (5) Calculate a stereo parameter used for time-domain downmixing processing, and encode the stereo parameter used for time-domain downmixing processing to obtain an encoding index of the stereo parameter used for time-domain downmixing processing.

[0099] The stereo parameter used for time-domain downmixing processing is used to perform time-domain downmixing processing on the left channel signal obtained after delay alignment processing and the right channel signal obtained after delay alignment processing.

[0100] (6) Perform, based on the stereo parameter used for time-domain downmixing processing, time-domain downmixing processing on the left channel signal and the right channel signal that are obtained after delay alignment processing, to obtain a primary channel signal and a secondary channel signal.

[0101] Time-domain downmixing processing is used to obtain the primary channel signal and the secondary channel signal.

Date Reçue/Date Received 2020-06-23

[0102] After the left channel signal and the right channel signal that are obtained after delay alignment processing are processed by using a time-domain downmixing technology, the primary channel signal (Primary channel, or referred to as a middle channel (Mid channel) signal), and the secondary channel (Secondary channel, or referred to as a side channel (Side channel) signal) are obtained.

[0103] The primary channel signal is used to represent information about correlation between channels, and the secondary channel signal is used to represent information about a difference between channels. When the left channel signal and the right channel signal that are obtained after delay alignment processing are aligned in time domain, the secondary channel signal is the weakest, and in this case, the stereo signal has a best effect.

[0104] Reference is made to a preprocessed left channel signal L and a preprocessed right channel signal R in an n^{th} frame shown in FIG. 4. The preprocessed left channel signal L is located before the preprocessed right channel signal R. In other words, compared with the preprocessed right channel signal R, the preprocessed left channel signal L has a delay, and there is an inter-channel time difference 21 between the preprocessed left channel signal L and the preprocessed right channel signal R. In this case, the secondary channel signal is enhanced, the primary channel signal is weakened, and the stereo signal has a relatively poor effect.

[0105] (7) Separately encode the primary channel signal and the secondary channel signal to obtain a first mono encoded bitstream corresponding to the primary channel signal and a second mono encoded bitstream corresponding to the secondary channel signal.

[0106] (8) Write the encoding index of the inter-channel time difference, the encoding index of the stereo parameter, the first mono encoded bitstream, and the second mono encoded bitstream into a stereo encoded bitstream.

[0107] The decoding component 120 is configured to decode the stereo encoded bitstream generated by the encoding component 110 to obtain the stereo signal.

[0108] Optionally, the encoding component 110 is connected to the decoding component 120 wiredly or wirelessly, and the decoding component 120 obtains,

Date Reçue/Date Received 2020-06-23

through the connection, the stereo encoded bitstream generated by the encoding component 110. Alternatively, the encoding component 110 stores the generated stereo encoded bitstream into a memory, and the decoding component 120 reads the stereo encoded bitstream in the memory.

[0109] Optionally, the decoding component 120 may be implemented by using software, may be implemented by using hardware, or may be implemented in a form of a combination of software and hardware. This is not limited in this embodiment.

[0110] The decoding the stereo encoded bitstream to obtain the stereo signal by the decoding component 120 includes the following several steps:

[0111] (1) Decode the first mono encoded bitstream and the second mono encoded bitstream in the stereo encoded bitstream to obtain the primary channel signal and the secondary channel signal.

[0112] (2) Obtain, based on the stereo encoded bitstream, an encoding index of a stereo parameter used for time-domain upmixing processing, and perform time-domain upmixing processing on the primary channel signal and the secondary channel signal to obtain a left channel signal obtained after time-domain upmixing processing and a right channel signal obtained after time-domain upmixing processing.

[0113] (3) Obtain the encoding index of the inter-channel time difference based on the stereo encoded bitstream, and perform delay adjustment on the left channel signal obtained after time-domain upmixing processing and the right channel signal obtained after time-domain upmixing processing to obtain the stereo signal.

[0114] Optionally, the encoding component 110 and the decoding component 120 may be disposed in a same device, or may be disposed in different devices. The device may be a mobile terminal that has an audio signal processing function, such as a mobile phone, a tablet computer, a laptop portable computer, a desktop computer, a Bluetooth speaker, a pen recorder, or a wearable device; or may be a network element that has an audio signal processing capability in a core network or a radio network. This is not limited in this embodiment.

[0115] For example, referring to FIG. 2, an example in which the encoding component 110 is disposed in a mobile terminal 130, and the decoding component 120

Date Reçue/Date Received 2020-06-23

is disposed in a mobile terminal 140. The mobile terminal 130 and the mobile terminal 140 are independent electronic devices with an audio signal processing capability, and the mobile terminal 130 and the mobile terminal 140 are connected to each other by using a wireless or wired network is used in this embodiment for description.

[0116] Optionally, the mobile terminal 130 includes a collection component 131, the encoding component 110, and a channel encoding component 132. The collection component 131 is connected to the encoding component 110, and the encoding component 110 is connected to the channel encoding component 132.

[0117] Optionally, the mobile terminal 140 includes an audio playing component 141, the decoding component 120, and a channel decoding component 142. The audio playing component 141 is connected to the decoding component 110, and the decoding component 110 is connected to the channel encoding component 132.

[0118] After collecting the stereo signal by using the collection component 131, the mobile terminal 130 encodes the stereo signal by using the encoding component 110 to obtain the stereo encoded bitstream. Then, the mobile terminal 130 encodes the stereo encoded bitstream by using the channel encoding component 132 to obtain a transmit signal.

[0119] The mobile terminal 130 sends the transmit signal to the mobile terminal 140 by using the wireless or wired network.

[0120] After receiving the transmit signal, the mobile terminal 140 decodes the transmit signal by using the channel decoding component 142 to obtain the stereo encoded bitstream, decodes the stereo encoded bitstream by using the decoding component 110 to obtain the stereo signal, and plays the stereo signal by using the audio playing component 141.

[0121] For example, referring to FIG. 3, this embodiment is described by using an example in which the encoding component 110 and the decoding component 120 are disposed in a same network element 150 that has an audio signal processing capability in a core network or a radio network.

[0122] Optionally, the network element 150 includes a channel decoding component 151, the decoding component 120, the encoding component 110, and a

Date Reçue/Date Received 2020-06-23

channel encoding component 152. The channel decoding component 151 is connected to the decoding component 120, the decoding component 120 is connected to the encoding component 110, and the encoding component 110 is connected to the channel encoding component 152.

[0123] After receiving a transmit signal sent by another device, the channel decoding component 151 decodes the transmit signal to obtain a first stereo encoded bitstream, decodes the stereo encoded bitstream by using the decoding component 120 to obtain a stereo signal, encodes the stereo signal by using the encoding component 110 to obtain a second stereo encoded bitstream, and encodes the second stereo encoded bitstream by using the channel encoding component 152 to obtain a transmit signal.

[0124] The another device may be a mobile terminal that has an audio signal processing capability, or may be another network element that has an audio signal processing capability. This is not limited in this embodiment.

[0125] Optionally, the encoding component 110 and the decoding component 120 in the network element may transcode a stereo encoded bitstream sent by the mobile terminal.

[0126] Optionally, in this embodiment, a device on which the encoding component 110 is installed is referred to as an audio coding device. In actual implementation, the audio coding device may also have an audio decoding function. This is not limited in this embodiment.

[0127] Optionally, in this embodiment, only the stereo signal is used as an example for description. In this application, the audio coding device may further process a multi-channel signal, where the multi-channel signal includes at least two channel signals.

[0128] Several nouns in the embodiments of this application are described below.

[0129] A multi-channel signal of a current frame is a frame of multi-channel signals used to estimate a current inter-channel time difference. The multi-channel signal of the current frame includes at least two channel signals. Channel signals of different channels may be collected by using different audio collection components in the audio coding device, or channel signals of different channels may be collected by different audio collection components in another device. The channel signals of different

Date Reçue/Date Received 2020-06-23

channels are transmitted from a same sound source.

[0130] For example, the multi-channel signal of the current frame includes a left channel signal L and a right channel signal R. The left channel signal L is collected by using a left channel audio collection component, the right channel signal R is collected by using a right channel audio collection component, and the left channel signal L and the right channel signal R are from a same sound source.

[0131] Referring to FIG. 4, an audio coding device is estimating an inter-channel time difference of a multi-channel signal of an n^{th} frame, and the n^{th} frame is the current frame.

[0132] A previous frame of the current frame is a first frame that is located before the current frame, for example, if the current frame is the n^{th} frame, the previous frame of the current frame is an $(n-1)^{th}$ frame.

[0133] Optionally, the previous frame of the current frame may also be briefly referred to as the previous frame.

[0134] A past frame is located before the current frame in time domain, and the past frame includes the previous frame of the current frame, first two frames of the current frame, first three frames of the current frame, and the like. Referring to FIG. 4, if the current frame is the n^{th} frame, the past frame includes: the $(n-1)^{th}$ frame, the $(n-2)^{th}$ frame, ..., and the first frame.

[0135] Optionally, in this application, at least one past frame may be M frames located before the current frame, for example, eight frames located before the current frame.

[0136] A next frame is a first frame after the current frame. Referring to FIG. 4, if the current frame is the n^{th} frame, the next frame is an $(n+1)^{th}$ frame.

[0137] A frame length is duration of a frame of multi-channel signals. Optionally, the frame length is represented by a quantity of sampling points, for example, a frame length N = 320 sampling points.

[0138] A cross-correlation coefficient is used to represent a degree of cross correlation between channel signals of different channels in the multi-channel signal of the current frame under different inter-channel time differences. The degree of cross

Date Reçue/Date Received 2020-06-23

correlation is represented by using a cross-correlation value. For any two channel signals in the multi-channel signal of the current frame, under an inter-channel time difference, if two channel signals obtained after delay adjustment is performed based on the inter-channel time difference are more similar, the degree of cross correlation is stronger, and the cross-correlation value is greater, or if a difference between two channel signals obtained after delay adjustment is performed based on the inter-channel time difference is greater, the degree of cross correlation is weaker, and the cross-correlation value is smaller.

[0139] An index value of the cross-correlation coefficient corresponds to an inter-channel time difference, and a cross-correlation value corresponding to each index value of the cross-correlation coefficient represents a degree of cross correlation between two mono signals that are obtained after delay adjustment and that are corresponding to each inter-channel time difference.

[0140] Optionally, the cross-correlation coefficient (cross-correlation coefficients) may also be referred to as a group of cross-correlation values or referred to as a cross-correlation function. This is not limited in this application.

[0141] Referring to FIG. 4, when a cross-correlation coefficient of a channel signal of an a^{th} frame is calculated, cross-correlation values between the left channel signal L and the right channel signal R are separately calculated under different inter-channel time differences.

[0142] For example, when the index value of the cross-correlation coefficient is 0, the inter-channel time difference is –N/2 sampling points, and the inter-channel time difference is used to align the left channel signal L and the right channel signal R to obtain the cross-correlation value k0;

when the index value of the cross-correlation coefficient is 1, the inter-channel time difference is (–N/2 + 1) sampling points, and the inter-channel time difference is used to align the left channel signal L and the right channel signal R to obtain the cross-correlation value k1;

when the index value of the cross-correlation coefficient is 2, the inter-channel time difference is (–N/2 + 2) sampling points, and the inter-channel time

Date Reçue/Date Received 2020-06-23

difference is used to align the left channel signal L and the right channel signal R to obtain the cross-correlation value k2;

when the index value of the cross-correlation coefficient is 3, the inter-channel time difference is (–N/2 + 3) sampling points, and the inter-channel time difference is used to align the left channel signal L and the right channel signal R to obtain the cross-correlation value k3; …, and

when the index value of the cross-correlation coefficient is N, the inter-channel time difference is N/2 sampling points, and the inter-channel time difference is used to align the left channel signal L and the right channel signal R to obtain the cross-correlation value kN.

[0143] A maximum value in k0 to kN is searched, for example, k3 is maximum. In this case, it indicates that when the inter-channel time difference is (–N/2 + 3) sampling points, the left channel signal L and the right channel signal R are most similar, in other words, the inter-channel time difference is closest to a real inter-channel time difference.

[0144] It should be noted that this embodiment is only used to describe a principle that the audio coding device determines the inter-channel time difference by using the cross-correlation coefficient. In actual implementation, the inter-channel time difference may not be determined by using the foregoing method.

[0145] FIG. 5 is a flowchart of a delay estimation method according to an example embodiment of this application. The method includes the following several steps.

[0146] Step 301: Determine a cross-correlation coefficient of a multi-channel signal of a current frame.

[0147] Step 302: Determine a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame.

[0148] Optionally, the at least one past frame is consecutive in time, and a last frame in the at least one past frame and the current frame are consecutive in time. In other words, the last past frame in the at least one past frame is a previous frame of the current frame. Alternatively, the at least one past frame is spaced by a predetermined quantity of frames in time, and a last past frame in the at least one past frame is spaced by a predetermined quantity of frames from the current frame. Alternatively, the at least one

Date Reçue/Date Received 2020-06-23

past frame is inconsecutive in time, a quantity of frames spaced between the at least one past frame is not fixed, and a quantity of frames between a last past frame in the at least one past frame and the current frame is not fixed. A value of the predetermined quantity of frames is not limited in this embodiment, for example, two frames.

[0149] In this embodiment, a quantity of past frames is not limited. For example, the quantity of past frames is 8, 12, and 25.

[0150] The delay track estimation value is used to represent a predicted value of an inter-channel time difference of the current frame. In this embodiment, a delay track is simulated based on the inter-channel time difference information of the at least one past frame, and the delay track estimation value of the current frame is calculated based on the delay track.

[0151] Optionally, the inter-channel time difference information of the at least one past frame is an inter-channel time difference of the at least one past frame, or an inter-channel time difference smoothed value of the at least one past frame.

[0152] An inter-channel time difference smoothed value of each past frame is determined based on a delay track estimation value of the frame and an inter-channel time difference of the frame.

[0153] Step 303: Determine an adaptive window function of the current frame.

[0154] Optionally, the adaptive window function is a raised cosine-like window function. The adaptive window function has a function of relatively enlarging a middle part and suppressing an edge part.

[0155] Optionally, adaptive window functions corresponding to frames of channel signals are different.

[0156] The adaptive window function is represented by using the following formulas:

when $0 \le k \le TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width - 1$,

$$loc\_weight\_win(k) = win\_bias;$$

when $TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width \le k \le TRUNC(A * L\_NCSHIFT\_DS/2) + 2 * win\_width - 1$,

$$loc\_weight\_win(k) = 0.5 * (1 + win\_bias) + 0.5 * (1 - win\_bias) * \cos(\pi * (k -$$

Date Reçue/Date Received 2020-06-23

TRUNC(A * L_NCSHIFT_DS/2))/(2 * win_width)); and

when TRUNC(A * L_NCSHIFT_DS/2) + 2 * win_width ≤ k ≤ A * L_NCSHIFT_DS,

loc_weight_win(k) = win_bias.

[0157] loc_weight_win(k) is used to represent the adaptive window function, where k = 0, 1, ..., A * L_NCSHIFT_DS; A is a preset constant greater than or equal to 4, for example, A = 4; TRUNC indicates rounding a value, for example, rounding a value of A * L_NCSHIFT_DS/2 in the formula of the adaptive window function; L_NCSHIFT_DS is a maximum value of an absolute value of an inter-channel time difference; win_width is used to represent a raised cosine width parameter of the adaptive window function; and win_bias is used to represent a raised cosine height bias of the adaptive window function.

[0158] Optionally, the maximum value of the absolute value of the inter-channel time difference is a preset positive number, and is usually a positive integer greater than zero and less than or equal to a frame length, for example, 40, 60, or 80.

[0159] Optionally, a maximum value of the inter-channel time difference or a minimum value of the inter-channel time difference is a preset positive integer, and the maximum value of the absolute value of the inter-channel time difference is obtained by taking an absolute value of the maximum value of the inter-channel time difference, or the maximum value of the absolute value of the inter-channel time difference is obtained by taking an absolute value of the minimum value of the inter-channel time difference.

[0160] For example, the maximum value of the inter-channel time difference is 40, the minimum value of the inter-channel time difference is –40, and the maximum value of the absolute value of the inter-channel time difference is 40, which is obtained by taking an absolute value of the maximum value of the inter-channel time difference and is also obtained by taking an absolute value of the minimum value of the inter-channel time difference.

[0161] For another example, the maximum value of the inter-channel time difference is 40, the minimum value of the inter-channel time difference is –20, and the

Date Reçue/Date Received 2020-06-23

maximum value of the absolute value of the inter-channel time difference is 40, which is obtained by taking an absolute value of the maximum value of the inter-channel time difference.

[0162] For another example, the maximum value of the inter-channel time difference is 40, the minimum value of the inter-channel time difference is –60, and the maximum value of the absolute value of the inter-channel time difference is 60, which is obtained by taking an absolute value of the minimum value of the inter-channel time difference.

[0163] It can be learned from the formula of the adaptive window function that the adaptive window function is a raised cosine-like window with a fixed height on both sides and a convexity in the middle. The adaptive window function includes a constant-weight window and a raised cosine window with a height bias. A weight of the constant-weight window is determined based on the height bias. The adaptive window function is mainly determined by two parameters: the raised cosine width parameter and the raised cosine height bias.

[0164] Reference is made to a schematic diagram of an adaptive window function shown in FIG. 6. Compared with a wide window 402, a narrow window 401 means that a window width of a raised cosine window in the adaptive window function is relatively small, and a difference between a delay track estimation value corresponding to the narrow window 401 and an actual inter-channel time difference is relatively small. Compared with the narrow window 401, the wide window 402 means that the window width of the raised cosine window in the adaptive window function is relatively large, and a difference between a delay track estimation value corresponding to the wide window 402 and the actual inter-channel time difference is relatively large. In other words, the window width of the raised cosine window in the adaptive window function is positively correlated with the difference between the delay track estimation value and the actual inter-channel time difference.

[0165] The raised cosine width parameter and the raised cosine height bias of the adaptive window function are related to inter-channel time difference estimation deviation information of a multi-channel signal of each frame. The inter-channel time

Date Reçue/Date Received 2020-06-23

difference estimation deviation information is used to represent a deviation between a predicted value of an inter-channel time difference and an actual value.

[0166] Reference is made to a schematic diagram of a relationship between a raised cosine width parameter and inter-channel time difference estimation deviation information shown in FIG. 7. If an upper limit value of the raised cosine width parameter is 0.25, a value of the inter-channel time difference estimation deviation information corresponding to the upper limit value of the raised cosine width parameter is 3.0. In this case, the value of the inter-channel time difference estimation deviation information is relatively large, and a window width of a raised cosine window in an adaptive window function is relatively large (refer to the wide window 402 in FIG. 6). If a lower limit value of the raised cosine width parameter of the adaptive window function is 0.04, a value of the inter-channel time difference estimation deviation information corresponding to the lower limit value of the raised cosine width parameter is 1.0. In this case, the value of the inter-channel time difference estimation deviation information is relatively small, and the window width of the raised cosine window in the adaptive window function is relatively small (refer to the narrow window 401 in FIG. 6).

[0167] Reference is made to a schematic diagram of a relationship between a raised cosine height bias and inter-channel time difference estimation deviation information shown in FIG. 8. If an upper limit value of the raised cosine height bias is 0.7, a value of the inter-channel time difference estimation deviation information corresponding to the upper limit value of the raised cosine height bias is 3.0. In this case, the smoothed inter-channel time difference estimation deviation is relatively large, and a height bias of a raised cosine window in an adaptive window function is relatively large (refer to the wide window 402 in FIG. 6). If a lower limit value of the raised cosine height bias is 0.4, a value of the inter-channel time difference estimation deviation information corresponding to the lower limit value of the raised cosine height bias is 1.0. In this case, the value of the inter-channel time difference estimation deviation information is relatively small, and the height bias of the raised cosine window in the adaptive window function is relatively small (refer to the narrow window 401 in FIG. 6).

Date Reçue/Date Received 2020-06-23

[0168] Step 304: Perform weighting on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame, to obtain a weighted cross-correlation coefficient.

[0169] The weighted cross-correlation coefficient may be obtained through calculation by using the following calculation formula:

$$c\_weight(x) = c(x) * loc\_weight\_win(x - TRUNC(reg\_prv\_corr) + TRUNC(A * L\_NCSHIFT\_DS/2) - L\_NCSHIFT\_DS).$$

[0170] c_weight(x) is the weighted cross-correlation coefficient; c(x) is the cross-correlation coefficient; loc_weight_win is the adaptive window function of the current frame; TRUNC indicates rounding a value, for example, rounding reg_prv_corr in the formula of the weighted cross-correlation coefficient, and rounding a value of A * L_NCSHIFT_DS/2; reg_prv_corr is the delay track estimation value of the current frame; and x is an integer greater than or equal to zero and less than or equal to 2 * L_NCSHIFT_DS.

[0171] The adaptive window function is the raised cosine-like window, and has the function of relatively enlarging a middle part and suppressing an edge part. Therefore, when weighting is performed on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame, if an index value is closer to the delay track estimation value, a weighting coefficient of a corresponding cross-correlation value is greater, and if the index value is farther from the delay track estimation value, the weighting coefficient of the corresponding cross-correlation value is smaller. The raised cosine width parameter and the raised cosine height bias of the adaptive window function adaptively suppress the cross-correlation value corresponding to the index value, away from the delay track estimation value, in the cross-correlation coefficient.

[0172] Step 305: Determine an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient.

[0173] The determining an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient includes: searching for a maximum value of the cross-correlation value in the weighted cross-correlation coefficient; and

Date Reçue/Date Received 2020-06-23

determining the inter-channel time difference of the current frame based on an index value corresponding to the maximum value.

[0174] Optionally, the searching for a maximum value of the cross-correlation value in the weighted cross-correlation coefficient includes: comparing a second cross-correlation value with a first cross-correlation value in the cross-correlation coefficient to obtain a maximum value in the first cross-correlation value and the second cross-correlation value; comparing a third cross-correlation value with the maximum value to obtain a maximum value in the third cross-correlation value and the maximum value; and in a cyclic order, comparing an i^{th} cross-correlation value with a maximum value obtained through previous comparison to obtain a maximum value in the i^{th} cross-correlation value and the maximum value obtained through previous comparison. It is assumed that $i = i + 1$, and the step of comparing an i^{th} cross-correlation value with a maximum value obtained through previous comparison is continuously performed until all cross-correlation values are compared, to obtain a maximum value in the cross-correlation values, where i is an integer greater than 2.

[0175] Optionally, the determining the inter-channel time difference of the current frame based on an index value corresponding to the maximum value includes: using a sum of the index value corresponding to the maximum value and the minimum value of the inter-channel time difference as the inter-channel time difference of the current frame.

[0176] The cross-correlation coefficient can reflect a degree of cross correlation between two channel signals obtained after a delay is adjusted based on different inter-channel time differences, and there is a correspondence between an index value of the cross-correlation coefficient and an inter-channel time difference. Therefore, an audio coding device can determine the inter-channel time difference of the current frame based on an index value corresponding to a maximum value of the cross-correlation coefficient (with a highest degree of cross correlation).

[0177] In conclusion, according to the delay estimation method provided in this embodiment, the inter-channel time difference of the current frame is predicted based on the delay track estimation value of the current frame, and weighting is performed on

Date Reçue/Date Received 2020-06-23

the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame. The adaptive window function is the raised cosine-like window, and has the function of relatively enlarging the middle part and suppressing the edge part. Therefore, when weighting is performed on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame, if an index value is closer to the delay track estimation value, a weighting coefficient is greater, avoiding a problem that a first cross-correlation coefficient is excessively smoothed, and if the index value is farther from the delay track estimation value, the weighting coefficient is smaller, avoiding a problem that a second cross-correlation coefficient is insufficiently smoothed. In this way, the adaptive window function adaptively suppresses a cross-correlation value corresponding to the index value, away from the delay track estimation value, in the cross-correlation coefficient, thereby improving accuracy of determining the inter-channel time difference in the weighted cross-correlation coefficient. The first cross-correlation coefficient is a cross-correlation value corresponding to an index value, near the delay track estimation value, in the cross-correlation coefficient, and the second cross-correlation coefficient is a cross-correlation value corresponding to an index value, away from the delay track estimation value, in the cross-correlation coefficient.

[0178] Steps 301 to 303 in the embodiment shown in FIG. 5 are described in detail below.

[0179] First, that the cross-correlation coefficient of the multi-channel signal of the current frame is determined in step 301 is described.

[0180] (1) The audio coding device determines the cross-correlation coefficient based on a left channel time domain signal and a right channel time domain signal of the current frame.

[0181] A maximum value T_{max} of the inter-channel time difference and a minimum value T_{min} of the inter-channel time difference usually need to be preset, so as to determine a calculation range of the cross-correlation coefficient. Both the maximum value T_{max} of the inter-channel time difference and the minimum value T_{min} of the inter-

Date Reçue/Date Received 2020-06-23

channel time difference are real numbers, and $T_{max} > T_{min}$. Values of T_{max} and T_{min} are related to a frame length, or values of T_{max} and T_{min} are related to a current sampling frequency.

[0182] Optionally, a maximum value L_NCSHIFT_DS of an absolute value of the inter-channel time difference is preset, to determine the maximum value T_{max} of the inter-channel time difference and the minimum value T_{min} of the inter-channel time difference. For example, the maximum value T_{max} of the inter-channel time difference = L_NCSHIFT_DS, and the minimum value T_{min} of the inter-channel time difference = –L_NCSHIFT_DS.

[0183] The values of T_{max} and T_{min} are not limited in this application. For example, if the maximum value L_NCSHIFT_DS of the absolute value of the inter-channel time difference is 40, $T_{max} = 40$, and $T_{min} = -40$.

[0184] In an implementation, an index value of the cross-correlation coefficient is used to indicate a difference between the inter-channel time difference and the minimum value of the inter-channel time difference. In this case, determining the cross-correlation coefficient based on the left channel time domain signal and the right channel time domain signal of the current frame is represented by using the following formulas:

[0185] In a case of $T_{min} \leq 0$ and $0 < T_{max}$,

when $T_{min} \leq i \leq 0$,

$$c(k) = \frac{1}{N+i} \sum_{j=0}^{N-1+i} \tilde{x}_R(j) \cdot \tilde{x}_L(j-i)$$, where $k = i - T_{min}$; and

when $0 < i \leq T_{max}$,

$$c(k) = \frac{1}{N+i} \sum_{j=0}^{N-1-i} \tilde{x}_R(j) \cdot \tilde{x}_L(j+i)$$, where $k = i - T_{min}$.

[0186] In a case of $T_{min} \leq 0$ and $T_{max} \leq 0$,

when $T_{min} \leq i \leq T_{max}$,

$$c(k) = \frac{1}{N+i} \sum_{j=0}^{N-1+i} \tilde{x}_R(j) \cdot \tilde{x}_L(j-i)$$, where $k = i - T_{min}$.

[0187] In a case of $T_{min} \geq 0$ and $T_{max} \geq 0$,

Date Reçue/Date Received 2020-06-23

when $T_{min} \le i \le T_{max}$,

$$c(k) = \frac{1}{N+i} \sum_{j=0}^{N-1-i} \tilde{x}_R(j) \cdot \tilde{x}_L(j+i), \text{ where } k = i - T_{min}.$$

[0188] N is a frame length, $\tilde{X}_L(j)$ is the left channel time domain signal of the current frame, $\tilde{X}_R(j)$ is the right channel time domain signal of the current frame, c(k) is the cross-correlation coefficient of the current frame, k is the index value of the cross-correlation coefficient, k is an integer not less than 0, and a value range of k is $[0, T_{max} - T_{min}]$.

[0189] It is assumed that $T_{max} = 40$, and $T_{min} = -40$. In this case, the audio coding device determines the cross-correlation coefficient of the current frame by using the calculation manner corresponding to the case that $T_{min} \le 0$ and $0 < T_{max}$. In this case, the value range of k is [0, 80].

[0190] In another implementation, the index value of the cross-correlation coefficient is used to indicate the inter-channel time difference. In this case, determining, by the audio coding device, the cross-correlation coefficient based on the maximum value of the inter-channel time difference and the minimum value of the inter-channel time difference is represented by using the following formulas:

[0191] In a case of $T_{min} \le 0$ and $0 < T_{max}$,

when $T_{min} \le i \le 0$,

$$c(i) = \frac{1}{N+i} \sum_{j=0}^{N-1+i} \tilde{x}_R(j) \cdot \tilde{x}_L(j-i); \text{ and}$$

when $0 < i \le T_{max}$,

$$c(i) = \frac{1}{N+i} \sum_{j=0}^{N-1-i} \tilde{x}_R(j) \cdot \tilde{x}_L(j+i).$$

[0192] In a case of $T_{min} \le 0$ and $T_{max} \le 0$,

when $T_{min} \le i \le T_{max}$,

$$c(i) = \frac{1}{N+i} \sum_{j=0}^{N-1+i} \tilde{x}_R(j) \cdot \tilde{x}_L(j-i).$$

[0193] In a case of $T_{min} \ge 0$ and $T_{max} \ge 0$,

Date Reçue/Date Received 2020-06-23

when $T_{min} \leq i \leq T_{max}$,

$$c(i) = \frac{1}{N+i} \sum_{j=0}^{N-1-i} \tilde{x}_R(j) \cdot \tilde{x}_L(j+i) .$$

[0194] N is a frame length, $\tilde{x}_L(j)$ is the left channel time domain signal of the current frame, $\tilde{x}_R(j)$ is the right channel time domain signal of the current frame, c(i) is the cross-correlation coefficient of the current frame, i is the index value of the cross-correlation coefficient, and a value range of i is $[T_{min}, T_{max}]$.

[0195] It is assumed that $T_{max} = 40$, and $T_{min} = -40$. In this case, the audio coding device determines the cross-correlation coefficient of the current frame by using the calculation formula corresponding to $T_{min} \leq 0$ and $0 < T_{max}$. In this case, the value range of i is [–40, 40].

[0196] Second, the determining a delay track estimation value of the current frame in step 302 is described.

[0197] In a first implementation, delay track estimation is performed based on the buffered inter-channel time difference information of the at least one past frame by using a linear regression method, to determine the delay track estimation value of the current frame.

[0198] This implementation is implemented by using the following several steps:

[0199] (1) Generate M data pairs based on the inter-channel time difference information of the at least one past frame and a corresponding sequence number, where M is a positive integer.

[0200] A buffer stores inter-channel time difference information of M past frames.

[0201] Optionally, the inter-channel time difference information is an inter-channel time difference. Alternatively, the inter-channel time difference information is an inter-channel time difference smoothed value.

[0202] Optionally, inter-channel time differences that are of the M past frames and that are stored in the buffer follow a first in first out principle. To be specific, a buffer location of an inter-channel time difference that is buffered first and that is of a past frame is in the front, and a buffer location of an inter-channel time difference that is

Date Reçue/Date Received 2020-06-23

buffered later and that is of a past frame is in the back.

[0203] In addition, for the inter-channel time difference that is buffered later and that is of the past frame, the inter-channel time difference that is buffered first and that is of the past frame moves out of the buffer first.

[0204] Optionally, in this embodiment, each data pair is generated by using inter-channel time difference information of each past frame and a corresponding sequence number.

[0205] A sequence number is referred to as a location of each past frame in the buffer. For example, if eight past frames are stored in the buffer, sequence numbers are 0, 1, 2, 3, 4, 5, 6, and 7 respectively.

[0206] For example, the generated M data pairs are: $\{(x_0, y_0), (x_1, y_1), (x_2, y_2) \ldots (x_r, y_r), \ldots,$ and $(x_{M-1}, y_{M-1})\}$. (x_r, y_r) is an $(r + 1)^{th}$ data pair, and x_r is used to indicate a sequence number of the $(r + 1)^{th}$ data pair, that is, $x_r = r$; and y_r is used to indicate an inter-channel time difference that is of a past frame and that is corresponding to the $(r + 1)^{th}$ data pair, where $r = 0, 1, \ldots,$ and $(M - 1)$.

[0207] FIG. 9 is a schematic diagram of eight buffered past frames. A location corresponding to each sequence number buffers an inter-channel time difference of one past frame. In this case, eight data pairs are: $\{(x_0, y_0), (x_1, y_1), (x_2, y_2) \ldots (x_r, yr), \ldots,$ and $(x_7, y_7)\}$. In this case, $r = 0, 1, 2, 3, 4, 5, 6,$ and 7.

[0208] (2) Calculate a first linear regression parameter and a second linear regression parameter based on the M data pairs.

[0209] In this embodiment, it is assumed that y_r in the data pairs is a linear function that is about x_r and that has a measurement error of ε_r. The linear function is as follows:

$$y_r = \alpha + \beta * x_r + \varepsilon_r.$$

[0210] α is the first linear regression parameter, β is the second linear regression parameter, and ε_r is the measurement error.

[0211] The linear function needs to meet the following condition: A distance between the observed value y_r (inter-channel time difference information actually buffered) corresponding to the observation point x_r and an estimation value $\alpha + \beta * x_r$ calculated based on the linear function is the smallest, to be specific, minimization of a

Date Reçue/Date Received 2020-06-23

cost function Q (α, β) is met.

[0212] The cost function Q (α, β) is as follows:

$$Q(\alpha, \beta) = \sum_{r=0}^{M1} \varepsilon_r = \sum_{r=0}^{M1} (y_r - \alpha - \beta \cdot x_r).$$

[0213] To meet the foregoing condition, the first linear regression parameter and the second linear regression parameter in the linear function need to meet the following:

$$\beta = \frac{\hat{XY} - \hat{X} * \hat{Y}}{\hat{X^2} - (\hat{X})^2};$$

$$\alpha = (\hat{Y} - \beta * \hat{X}) / M;$$

$$\hat{X} = \sum_{r=0}^{M-1} x_r;$$

$$\hat{Y} = \sum_{r=0}^{M-1} y_r;$$

$$\hat{X^2} = \sum_{r=0}^{M-1} x_r^2; \text{and}$$

$$\hat{XY} = \sum_{r=0}^{M-1} x_r * y_r.$$

[0214] x_r is used to indicate the sequence number of the $(r + 1)^{th}$ data pair in the M data pairs, and y_r is inter-channel time difference information of the $(r + 1)^{th}$ data pair.

[0215] (3) Obtain the delay track estimation value of the current frame based on the first linear regression parameter and the second linear regression parameter.

[0216] An estimation value corresponding to a sequence number of an $(M + 1)^{th}$ data pair is calculated based on the first linear regression parameter and the second linear regression parameter, and the estimation value is determined as the delay track estimation value of the current frame. A formula is as follows:

Date Reçue/Date Received 2020-06-23

$$reg\_prv\_corr = \alpha + \beta * M, \text{ where}$$

reg_prv_corr represents the delay track estimation value of the current frame, M is the sequence number of the $(M + 1)^{th}$ data pair, and $\alpha + \beta * M$ is the estimation value of the $(M + 1)^{th}$ data pair.

[0217] For example, M = 8. After α and β are determined based on the eight generated data pairs, an inter-channel time difference in a ninth data pair is estimated based on α and β, and the inter-channel time difference in the ninth data pair is determined as the delay track estimation value of the current frame, that is, $reg\_prv\_corr = \alpha + \beta * 8$.

[0218] Optionally, in this embodiment, only a manner of generating a data pair by using a sequence number and an inter-channel time difference is used as an example for description. In actual implementation, the data pair may alternatively be generated in another manner. This is not limited in this embodiment.

[0219] In a second implementation, delay track estimation is performed based on the buffered inter-channel time difference information of the at least one past frame by using a weighted linear regression method, to determine the delay track estimation value of the current frame.

[0220] This implementation is implemented by using the following several steps:

[0221] (1) Generate M data pairs based on the inter-channel time difference information of the at least one past frame and a corresponding sequence number, where M is a positive integer.

[0222] This step is the same as the related description in step (1) in the first implementation, and details are not described herein in this embodiment.

[0223] (2) Calculate a first linear regression parameter and a second linear regression parameter based on the M data pairs and weighting coefficients of the M past frames.

[0224] Optionally, the buffer stores not only the inter-channel time difference information of the M past frames, but also stores the weighting coefficients of the M past frames. A weighting coefficient is used to calculate a delay track estimation value of a corresponding past frame.

Date Reçue/Date Received 2020-06-23

[0225] Optionally, a weighting coefficient of each past frame is obtained through calculation based on a smoothed inter-channel time difference estimation deviation of the past frame. Alternatively, a weighting coefficient of each past frame is obtained through calculation based on an inter-channel time difference estimation deviation of the past frame.

[0226] In this embodiment, it is assumed that y_r in the data pairs is a linear function that is about x_r and that has a measurement error of ε_r. The linear function is as follows:

$$y_r = \alpha + \beta * x_r + \varepsilon_r.$$

[0227] α is the first linear regression parameter, β is the second linear regression parameter, and ε_r is the measurement error.

[0228] The linear function needs to meet the following condition: A weighting distance between the observed value y_r (inter-channel time difference information actually buffered) corresponding to the observation point x_r and an estimation value $\alpha + \beta * x_r$ calculated based on the linear function is the smallest, to be specific, minimization of a cost function Q (α, β) is met.

[0229] The cost function Q (α, β) is as follows:

$$Q(\alpha,\beta)=\sum_{r=0}^{M-1} w_r \cdot \varepsilon_r = \sum_{r=0}^{M-1} w_r \cdot (y_r - \alpha - \beta \cdot x_r).$$

[0230] w_r is a weighting coefficient of a past frame corresponding to an r^{th} data pair.

[0231] To meet the foregoing condition, the first linear regression parameter and the second linear regression parameter in the linear function need to meet the following:

$$\beta = \frac{\hat{W} * \hat{XY} - \hat{X} * \hat{Y}}{\hat{W} * \hat{X^2} - (\hat{X})^2};$$

$$\alpha = \frac{\hat{Y} - \beta * \hat{X}}{\hat{W}};$$

$$\hat{X} = \sum_{r=0}^{M-1} w_r * x_r;$$

Date Reçue/Date Received 2020-06-23

$$\hat{Y} = \sum_{r=0}^{M-1} w_r * y_r;$$

$$\hat{W} = \sum_{r=0}^{M-1} w_r;$$

$$\hat{X^2} = \sum_{r=0}^{M-1} w_r * x_r^2; \text{ and}$$

$$\hat{XY} = \sum_{r=0}^{M-1} w_r * x_r * y_r.$$

[0232] x_r is used to indicate a sequence number of the $(r + 1)^{th}$ data pair in the M data pairs, y_r is inter-channel time difference information in the $(r + 1)^{th}$ data pair, w_r is a weighting coefficient corresponding to the inter-channel time difference information in the $(r + 1)^{th}$ data pair in at least one past frame.

[0233] (3) Obtain the delay track estimation value of the current frame based on the first linear regression parameter and the second linear regression parameter.

[0234] This step is the same as the related description in step (3) in the first implementation, and details are not described herein in this embodiment.

[0235] Optionally, in this embodiment, only a manner of generating a data pair by using a sequence number and an inter-channel time difference is used as an example for description. In actual implementation, the data pair may alternatively be generated in another manner. This is not limited in this embodiment.

[0236] It should be noted that in this embodiment, description is provided by using an example in which a delay track estimation value is calculated only by using the linear regression method or in the weighted linear regression manner. In actual implementation, the delay track estimation value may alternatively be calculated in another manner. This is not limited in this embodiment. For example, the delay track estimation value is calculated by using a B-spline (B-spline) method, or the delay track estimation value is calculated by using a cubic spline method, or the delay track estimation value is calculated by using a quadratic spline method.

Date Reçue/Date Received 2020-06-23

[0237] Third, the determining an adaptive window function of the current frame in step 303 is described.

[0238] In this embodiment, two manners of calculating the adaptive window function of the current frame are provided. In a first manner, the adaptive window function of the current frame is determined based on a smoothed inter-channel time difference estimation deviation of a previous frame. In this case, inter-channel time difference estimation deviation information is the smoothed inter-channel time difference estimation deviation, and the raised cosine width parameter and the raised cosine height bias of the adaptive window function are related to the smoothed inter-channel time difference estimation deviation. In a second manner, the adaptive window function of the current frame is determined based on the inter-channel time difference estimation deviation of the current frame. In this case, the inter-channel time difference estimation deviation information is the inter-channel time difference estimation deviation, and the raised cosine width parameter and the raised cosine height bias of the adaptive window function are related to the inter-channel time difference estimation deviation.

[0239] The two manners are separately described below.

[0240] This first manner is implemented by using the following several steps:

[0241] (1) Calculate a first raised cosine width parameter based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame.

[0242] Because accuracy of calculating the adaptive window function of the current frame by using a multi-channel signal near the current frame is relatively high, in this embodiment, description is provided by using an example in which the adaptive window function of the current frame is determined based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame.

[0243] Optionally, the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame is stored in the buffer.

[0244] This step is represented by using the following formulas:

$$win\_width1 = TRUNC(width\_par1 * (A * L\_NCSHIFT\_DS + 1)), \text{ and}$$

Date Reçue/Date Received 2020-06-23

width_par1 = a_width1 * smooth_dist_reg + b_width1, where

a_width1 = (xh_width1 – xl_width1)/(yh_dist1 – yl_dist1),

b_width1 = xh_width1 – a_width1 * yh_dist1,

win_width1 is the first raised cosine width parameter, TRUNC indicates rounding a value, L_NCSHIFT_DS is the maximum value of the absolute value of the inter-channel time difference, A is a preset constant, and A is greater than or equal to 4.

[0245] xh_width1 is an upper limit value of the first raised cosine width parameter, for example, 0.25 in FIG. 7; xl_width1 is a lower limit value of the first raised cosine width parameter, for example, 0.04 in FIG. 7; yh_dist1 is a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first raised cosine width parameter, for example, 3.0 corresponding to 0.25 in FIG. 7; yl_dist1 is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first raised cosine width parameter, for example, 1.0 corresponding to 0.04 in FIG. 7.

[0246] smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, and xh_width1, xl_width1, yh_dist1, and yl_dist1 are all positive numbers.

[0247] Optionally, in the foregoing formula, b_width1 = xh_width1 – a_width1 * yh_dist1 may be replaced with b_width1 = xl_width1 – a_width1 * yl_dist1.

[0248] Optionally, in this step, width_par1 = min(width_par1, xh_width1), and width_par1 = max(width_par1, xl_width1), where min represents taking of a minimum value, and max represents taking of a maximum value. To be specific, when width_par1 obtained through calculation is greater than xh_width1, width_par1 is set to xh_width1; or when width_par1 obtained through calculation is less than xl_width1, width_par1 is set to xl_width1.

[0249] In this embodiment, when width_par1 is greater than the upper limit value of the first raised cosine width parameter, width_par1 is limited to be the upper limit value of the first raised cosine width parameter; or when width_par1 is less than the lower limit value of the first raised cosine width parameter, width_par1 is limited to the lower limit value of the first raised cosine width parameter, so as to ensure that a value

Date Reçue/Date Received 2020-06-23

of width_par1 does not exceed a normal value range of the raised cosine width parameter, thereby ensuring accuracy of a calculated adaptive window function.

[0250] (2) Calculate a first raised cosine height bias based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame.

[0251] This step is represented by using the following formula:

$$win\_bias1 = a\_bias1 * smooth\_dist\_reg + b\_bias1, \text{ where}$$

$$a\_bias1 = (xh\_bias1 - xl\_bias1)/(yh\_dist2 - yl\_dist2), \text{ and}$$

$$b\_bias1 = xh\_bias1 - a\_bias1 * yh\_dist2.$$

[0252] win_bias1 is the first raised cosine height bias; xh_bias1 is an upper limit value of the first raised cosine height bias, for example, 0.7 in FIG. 8; xl_bias1 is a lower limit value of the first raised cosine height bias, for example, 0.4 in FIG. 8; yh_dist2 is a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first raised cosine height bias, for example, 3.0 corresponding to 0.7 in FIG. 8; yl_dist2 is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first raised cosine height bias, for example, 1.0 corresponding to 0.4 in FIG. 8; smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame; and yh_dist2, yl_dist2, xh_bias1, and xl_bias1 are all positive numbers.

[0253] Optionally, in the foregoing formula, b_bias1 = xh_bias1 – a_bias1 * yh_dist2 may be replaced with b_bias1 = xl_bias1 – a_bias1 * yl_dist2.

[0254] Optionally, in this embodiment, win_bias1 = min(win_bias1, xh_bias1), and win_bias1 = max(win_bias1, xl_bias1). To be specific, when win_bias1 obtained through calculation is greater than xh_bias1, win_bias1 is set to xh_bias1; or when win_bias1 obtained through calculation is less than xl_bias1, win_bias1 is set to xl_bias1.

[0255] Optionally, yh_dist2 = yh_dist1, and yl_dist2 = yl_dist1.

[0256] (3) Determine the adaptive window function of the current frame based on the first raised cosine width parameter and the first raised cosine height bias.

[0257] The first raised cosine width parameter and the first raised cosine height bias

Date Reçue/Date Received 2020-06-23

are brought into the adaptive window function in step 303 to obtain the following calculation formulas:

when $0 \le k \le TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width1 - 1$,

$$loc\_weight\_win(k) = win\_bias1;$$

when $TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width1 \le k \le TRUNC(A * L\_NCSHIFT\_DS/2) + 2 * win\_width1 - 1$,

$$loc\_weight\_win(k) = 0.5 * (1 + win\_bias1) + 0.5 * (1 - win\_bias1) * \cos(\pi * (k - TRUNC(A * L\_NCSHIFT\_DS/2))/(2 * win\_width1));$$ and

when $TRUNC(A * L\_NCSHIFT\_DS/2) + 2 * win\_width1 \le k \le A * L\_NCSHIFT\_DS$,

$$loc\_weight\_win(k) = win\_bias1.$$

[0258] loc_weight_win(k) is used to represent the adaptive window function, where k = 0, 1, ..., A * L_NCSHIFT_DS; A is the preset constant greater than or equal to 4, for example, A = 4, L_NCSHIFT_DS is the maximum value of the absolute value of the inter-channel time difference; win_width1 is the first raised cosine width parameter; and win_bias1 is the first raised cosine height bias.

[0259] In this embodiment, the adaptive window function of the current frame is calculated by using the smoothed inter-channel time difference estimation deviation of the previous frame, so that a shape of the adaptive window function is adjusted based on the smoothed inter-channel time difference estimation deviation, thereby avoiding a problem that a generated adaptive window function is inaccurate due to an error of the delay track estimation of the current frame, and improving accuracy of generating an adaptive window function.

[0260] Optionally, after the inter-channel time difference of the current frame is determined based on the adaptive window function determined in the first manner, the smoothed inter-channel time difference estimation deviation of the current frame may be further determined based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, the delay track estimation value of the current frame, and the inter-channel time difference of the current frame.

[0261] Optionally, the smoothed inter-channel time difference estimation deviation

Date Reçue/Date Received 2020-06-23

of the previous frame of the current frame in the buffer is updated based on the smoothed inter-channel time difference estimation deviation of the current frame.

[0262] Optionally, after the inter-channel time difference of the current frame is determined each time, the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame in the buffer is updated based on the smoothed inter-channel time difference estimation deviation of the current frame.

[0263] Optionally, updating the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame in the buffer based on the smoothed inter-channel time difference estimation deviation of the current frame includes: replacing the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame in the buffer with the smoothed inter-channel time difference estimation deviation of the current frame.

[0264] The smoothed inter-channel time difference estimation deviation of the current frame is obtained through calculation by using the following calculation formulas:

$$smooth\_dist\_reg\_update = (1 - \gamma) * smooth\_dist\_reg + \gamma * dist\_reg', \text{ and}$$

$$dist\_reg' = |reg\_prv\_corr - cur\_itd|.$$

[0265] smooth_dist_reg_update is the smoothed inter-channel time difference estimation deviation of the current frame; γ is a first smoothing factor, and $0 < \gamma < 1$, for example, $\gamma = 0.02$; smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame; reg_prv_corr is the delay track estimation value of the current frame; and cur_itd is the inter-channel time difference of the current frame.

[0266] In this embodiment, after the inter-channel time difference of the current frame is determined, the smoothed inter-channel time difference estimation deviation of the current frame is calculated. When an inter-channel time difference of a next frame is to be determined, an adaptive window function of the next frame can be determined by using the smoothed inter-channel time difference estimation deviation of the current frame, thereby ensuring accuracy of determining the inter-channel time difference of

Date Reçue/Date Received 2020-06-23

the next frame.

[0267] Optionally, after the inter-channel time difference of the current frame is determined based on the adaptive window function determined in the foregoing first manner, the buffered inter-channel time difference information of the at least one past frame may be further updated.

[0268] In an update manner, the buffered inter-channel time difference information of the at least one past frame is updated based on the inter-channel time difference of the current frame.

[0269] In another update manner, the buffered inter-channel time difference information of the at least one past frame is updated based on an inter-channel time difference smoothed value of the current frame.

[0270] Optionally, the inter-channel time difference smoothed value of the current frame is determined based on the delay track estimation value of the current frame and the inter-channel time difference of the current frame.

[0271] For example, based on the delay track estimation value of the current frame and the inter-channel time difference of the current frame, the inter-channel time difference smoothed value of the current frame may be determined by using the following formula:

$$cur\_itd\_smooth = \varphi * reg\_prv\_corr + (1 - \varphi) * cur\_itd.$$

[0272] cur_itd_smooth is the inter-channel time difference smoothed value of the current frame, φ is a second smoothing factor, reg_prv_corr is the delay track estimation value of the current frame, and cur_itd is the inter-channel time difference of the current frame. φ is a constant greater than or equal to 0 and less than or equal to 1.

[0273] The updating the buffered inter-channel time difference information of the at least one past frame includes: adding the inter-channel time difference of the current frame or the inter-channel time difference smoothed value of the current frame to the buffer.

[0274] Optionally, for example, the inter-channel time difference smoothed value in the buffer is updated. The buffer stores inter-channel time difference smoothed values corresponding to a fixed quantity of past frames, for example, the buffer stores inter-

Date Reçue/Date Received 2020-06-23

channel time difference smoothed values of eight past frames. If the inter-channel time difference smoothed value of the current frame is added to the buffer, an inter-channel time difference smoothed value of a past frame that is originally located in a first bit (a head of a queue) in the buffer is deleted. Correspondingly, an inter-channel time difference smoothed value of a past frame that is originally located in a second bit is updated to the first bit. By analogy, the inter-channel time difference smoothed value of the current frame is located in a last bit (a tail of the queue) in the buffer.

[0275] Reference is made to a buffer updating process shown in FIG. 10. It is assumed that the buffer stores inter-channel time difference smoothed values of eight past frames. Before an inter-channel time difference smoothed value 601 of the current frame is added to the buffer (that is, the eight past frames corresponding to the current frame), an inter-channel time difference smoothed value of an $(i-8)^{th}$ frame is buffered in a first bit, and an inter-channel time difference smoothed value of an $(i-7)^{th}$ frame is buffered in a second bit, …, and an inter-channel time difference smoothed value of an $(i-1)^{th}$ frame is buffered in an eighth bit.

[0276] If the inter-channel time difference smoothed value 601 of the current frame is added to the buffer, the first bit (which is represented by a dashed box in the figure) is deleted, a sequence number of the second bit becomes a sequence number of the first bit, a sequence number of the third bit becomes the sequence number of the second bit, ..., and a sequence number of the eighth bit becomes a sequence number of a seventh bit. The inter-channel time difference smoothed value 601 of the current frame (an i^{th} frame) is located in the eighth bit, to obtain eight past frames corresponding to a next frame.

[0277] Optionally, after the inter-channel time difference smoothed value of the current frame is added to the buffer, the inter-channel time difference smoothed value buffered in the first bit may not be deleted, instead, inter-channel time difference smoothed values in the second bit to a ninth bit are directly used to calculate an inter-channel time difference of a next frame. Alternatively, inter-channel time difference smoothed values in the first bit to a ninth bit are used to calculate an inter-channel time difference of a next frame. In this case, a quantity of past frames corresponding to each

Date Reçue/Date Received 2020-06-23

current frame is variable. A buffer update manner is not limited in this embodiment.

[0278] In this embodiment, after the inter-channel time difference of the current frame is determined, the inter-channel time difference smoothed value of the current frame is calculated. When a delay track estimation value of the next frame is to be determined, the delay track estimation value of the next frame can be determined by using the inter-channel time difference smoothed value of the current frame. This ensures accuracy of determining the delay track estimation value of the next frame.

[0279] Optionally, if the delay track estimation value of the current frame is determined based on the foregoing second implementation of determining the delay track estimation value of the current frame, after the buffered inter-channel time difference smoothed value of the at least one past frame is updated, a buffered weighting coefficient of the at least one past frame may be further updated. The weighting coefficient of the at least one past frame is a weighting coefficient in the weighted linear regression method.

[0280] In the first manner of determining the adaptive window function, the updating the buffered weighting coefficient of the at least one past frame includes: calculating a first weighting coefficient of the current frame based on the smoothed inter-channel time difference estimation deviation of the current frame; and updating a buffered first weighting coefficient of the at least one past frame based on the first weighting coefficient of the current frame.

[0281] In this embodiment, for related descriptions of buffer updating, refer to FIG. 10. Details are not described again herein in this embodiment.

[0282] The first weighting coefficient of the current frame is obtained through calculation by using the following calculation formulas:

$$wgt\_par1 = a\_wgt1 * smooth\_dist\_reg\_update + b\_wgt1,$$

$$a\_wgt1 = (xl\_wgt1 - xh\_wgt1)/(yh\_dist1' - yl\_dist1'), \text{ and}$$

$$b\_wgt1 = xl\_wgt1 - a\_wgt1 * yh\_dist1'.$$

[0283] wgt_par1 is the first weighting coefficient of the current frame, smooth_dist_reg_update is the smoothed inter-channel time difference estimation deviation of the current frame, xh_wgt is an upper limit value of the first weighting

Date Reçue/Date Received 2020-06-23

coefficient, xl_wgt is a lower limit value of the first weighting coefficient, yh_dist1' is a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first weighting coefficient, yl_dist1' is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first weighting coefficient, and yh_dist1', yl_dist1', xh_wgt1, and xl_wgt1 are all positive numbers.

[0284] Optionally, wgt_par1 = min(wgt_par1, xh_wgt1), and wgt_par1 = max(wgt_par1, xl_wgt1).

[0285] Optionally, in this embodiment, values of yh_dist1', yl_dist1', xh_wgt1, and xl_wgt1 are not limited. For example, xl_wgt1 = 0.05, xh_wgt1 = 1.0, yl_dist1' = 2.0, and yh_dist1' = 1.0.

[0286] Optionally, in the foregoing formula, b_wgt1 = xl_wgt1 – a_wgt1 * yh_dist1' may be replaced with b_wgt1 = xh_wgt1 – a_wgt1 * yl_dist1'.

[0287] In this embodiment, xh_wgt1 > xl_wgt1, and yh_dist1' < yl_dist1'.

[0288] In this embodiment, when wgt_par1 is greater than the upper limit value of the first weighting coefficient, wgt_par1 is limited to be the upper limit value of the first weighting coefficient; or when wgt_par1 is less than the lower limit value of the first weighting coefficient, wgt_par1 is limited to the lower limit value of the first weighting coefficient, so as to ensure that a value of wgt_par1 does not exceed a normal value range of the first weighting coefficient, thereby ensuring accuracy of the calculated delay track estimation value of the current frame.

[0289] In addition, after the inter-channel time difference of the current frame is determined, the first weighting coefficient of the current frame is calculated. When the delay track estimation value of the next frame is to be determined, the delay track estimation value of the next frame can be determined by using the first weighting coefficient of the current frame, thereby ensuring accuracy of determining the delay track estimation value of the next frame.

[0290] In the second manner, an initial value of the inter-channel time difference of the current frame is determined based on the cross-correlation coefficient; the inter-channel time difference estimation deviation of the current frame is calculated based on

Date Reçue/Date Received 2020-06-23

the delay track estimation value of the current frame and the initial value of the inter-channel time difference of the current frame; and the adaptive window function of the current frame is determined based on the inter-channel time difference estimation deviation of the current frame.

[0291] Optionally, the initial value of the inter-channel time difference of the current frame is a maximum value that is of a cross-correlation value in the cross-correlation coefficient and that is determined based on the cross-correlation coefficient of the current frame, and an inter-channel time difference determined based on an index value corresponding to the maximum value.

[0292] Optionally, determining the inter-channel time difference estimation deviation of the current frame based on the delay track estimation value of the current frame and the initial value of the inter-channel time difference of the current frame is represented by using the following formula:

$$dist\_reg = |reg\_prv\_corr - cur\_itd\_init|.$$

[0293] dist_reg is the inter-channel time difference estimation deviation of the current frame, reg_prv_corr is the delay track estimation value of the current frame, and cur_itd_init is the initial value of the inter-channel time difference of the current frame.

[0294] Based on the inter-channel time difference estimation deviation of the current frame, determining the adaptive window function of the current frame is implemented by using the following steps.

[0295] (1) Calculate a second raised cosine width parameter based on the inter-channel time difference estimation deviation of the current frame.

[0296] This step may be represented by using the following formulas:

$$win\_width2 = TRUNC(width\_par2 * (A * L\_NCSHIFT\_DS + 1)), \text{ and}$$

$$width\_par2 = a\_width2 * dist\_reg + b\_width2, \text{ where}$$

$$a\_width2 = (xh\_width2 - xl\_width2)/(yh\_dist3 - yl\_dist3), \text{ and}$$

$$b\_width2 = xh\_width2 - a\_width2 * yh\_dist3.$$

[0297] win_width2 is the second raised cosine width parameter, TRUNC indicates rounding a value, L_NCSHIFT_DS is a maximum value of an absolute value of an inter-channel time difference, A is a preset constant, A is greater than or equal to 4, A *

Date Reçue/Date Received 2020-06-23

L_NCSHIFT_DS + 1 is a positive integer greater than zero, xh_width2 is an upper limit value of the second raised cosine width parameter, xl_width2 is a lower limit value of the second raised cosine width parameter, yh_dist3 is an inter-channel time difference estimation deviation corresponding to the upper limit value of the second raised cosine width parameter, yl_dist3 is an inter-channel time difference estimation deviation corresponding to the lower limit value of the second raised cosine width parameter, dist_reg is the inter-channel time difference estimation deviation, xh_width2, xl_width2, yh_dist3, and yl_dist3 are all positive numbers.

[0298] Optionally, in this step, b_width2 = xh_width2 – a_width2 * yh_dist3 may be replaced with b_width2 = xl_width2 – a_width2 * yl_dist3.

[0299] Optionally, in this step, width_par2 = min(width_par2, xh_width2), and width_par2 = max(width_par2, xl_width2), where min represents taking of a minimum value, and max represents taking of a maximum value. To be specific, when width_par2 obtained through calculation is greater than xh_width2, width_par2 is set to xh_width2; or when width_par2 obtained through calculation is less than xl_width2, width_par2 is set to xl_width2.

[0300] In this embodiment, when width_par2 is greater than the upper limit value of the second raised cosine width parameter, width_par2 is limited to be the upper limit value of the second raised cosine width parameter; or when width_par2 is less than the lower limit value of the second raised cosine width parameter, width_par2 is limited to the lower limit value of the second raised cosine width parameter, so as to ensure that a value of width_par2 does not exceed a normal value range of the raised cosine width parameter, thereby ensuring accuracy of a calculated adaptive window function.

[0301] (2) Calculate a second raised cosine height bias based on the inter-channel time difference estimation deviation of the current frame.

[0302] This step may be represented by using the following formula:

$$win\_bias2 = a\_bias2 * dist\_reg + b\_bias2, \text{ where}$$

$$a\_bias2 = (xh\_bias2 - xl\_bias2)/(yh\_dist4 - yl\_dist4), \text{ and}$$

$$b\_bias2 = xh\_bias2 - a\_bias2 * yh\_dist4.$$

[0303] win_bias2 is the second raised cosine height bias, xh_bias2 is an upper limit

Date Reçue/Date Received 2020-06-23

value of the second raised cosine height bias, xl_bias2 is a lower limit value of the second raised cosine height bias, yh_dist4 is an inter-channel time difference estimation deviation corresponding to the upper limit value of the second raised cosine height bias, yl_dist4 is an inter-channel time difference estimation deviation corresponding to the lower limit value of the second raised cosine height bias, dist_reg is the inter-channel time difference estimation deviation, and yh_dist4, yl_dist4, xh_bias2, and xl_bias2 are all positive numbers.

[0304] Optionally, in this step, b_bias2 = xh_bias2 – a_bias2 * yh_dist4 may be replaced with b_bias2 = xl_bias2 – a_bias2 * yl_dist4.

[0305] Optionally, in this embodiment, win_bias2 = min(win_bias2, xh_bias2), and win_bias2 = max(win_bias2, xl_bias2). To be specific, when win_bias2 obtained through calculation is greater than xh_bias2, win_bias2 is set to xh_bias2; or when win_bias2 obtained through calculation is less than xl_bias2, win_bias2 is set to xl_bias2.

[0306] Optionally, yh_dist4 = yh_dist3, and yl_dist4 = yl_dist3.

[0307] (3) The audio coding device determines the adaptive window function of the current frame based on the second raised cosine width parameter and the second raised cosine height bias.

[0308] The audio coding device brings the second raised cosine width parameter and the second raised cosine height bias into the adaptive window function in step 303 to obtain the following calculation formulas:

when $0 \le k \le TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width2 - 1$,

$$loc\_weight\_win(k) = win\_bias2;$$

when $TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width2 \le k \le TRUNC(A * L\_NCSHIFT\_DS/2) + 2 * win\_width2 - 1$,

$$loc\_weight\_win(k) = 0.5 * (1 + win\_bias2) + 0.5 * (1 - win\_bias2) * \cos(\pi * (k - TRUNC(A * L\_NCSHIFT\_DS/2))/(2 * win\_width2));$$ and

when $TRUNC(A * L\_NCSHIFT\_DS/2) + 2 * win\_width2 \le k \le A * L\_NCSHIFT\_DS$,

$$loc\_weight\_win(k) = win\_bias2.$$

Date Reçue/Date Received 2020-06-23

[0309] loc_weight_win(k) is used to represent the adaptive window function, where k = 0, 1, ..., A * L_NCSHIFT_DS; A is the preset constant greater than or equal to 4, for example, A = 4, L_NCSHIFT_DS is the maximum value of the absolute value of the inter-channel time difference; win_width2 is the second raised cosine width parameter; and win_bias2 is the second raised cosine height bias.

[0310] In this embodiment, the adaptive window function of the current frame is determined based on the inter-channel time difference estimation deviation of the current frame, and when the smoothed inter-channel time difference estimation deviation of the previous frame does not need to be buffered, the adaptive window function of the current frame can be determined, thereby saving a storage resource.

[0311] Optionally, after the inter-channel time difference of the current frame is determined based on the adaptive window function determined in the foregoing second manner, the buffered inter-channel time difference information of the at least one past frame may be further updated. For related descriptions, refer to the first manner of determining the adaptive window function. Details are not described again herein in this embodiment.

[0312] Optionally, if the delay track estimation value of the current frame is determined based on the second implementation of determining the delay track estimation value of the current frame, after the buffered inter-channel time difference smoothed value of the at least one past frame is updated, a buffered weighting coefficient of the at least one past frame may be further updated.

[0313] In the second manner of determining the adaptive window function, the weighting coefficient of the at least one past frame is a second weighting coefficient of the at least one past frame.

[0314] Updating the buffered weighting coefficient of the at least one past frame includes: calculating a second weighting coefficient of the current frame based on the inter-channel time difference estimation deviation of the current frame; and updating a buffered second weighting coefficient of the at least one past frame based on the second weighting coefficient of the current frame.

[0315] Calculating the second weighting coefficient of the current frame based on

Date Reçue/Date Received 2020-06-23

the inter-channel time difference estimation deviation of the current frame is represented by using the following formulas:

$$wgt\_par2 = a\_wgt2 * dist\_reg + b\_wgt2,$$

$$a\_wgt2 = (xl\_wgt2 - xh\_wgt2)/(yh\_dist2' - yl\_dist2'), \text{ and}$$

$$b\_wgt2 = xl\_wgt2 - a\_wgt2 * yh\_dist2'.$$

[0316] wgt_par2 is the second weighting coefficient of the current frame, dist_reg is the inter-channel time difference estimation deviation of the current frame, xh_wgt2 is an upper limit value of the second weighting coefficient, xl_wgt2 is a lower limit value of the second weighting coefficient, yh_dist2' is an inter-channel time difference estimation deviation corresponding to the upper limit value of the second weighting coefficient, yl_dist2' is an inter-channel time difference estimation deviation corresponding to the lower limit value of the second weighting coefficient, and yh_dist2', yl_dist2', xh_wgt2, and xl_wgt2 are all positive numbers.

[0317] Optionally, wgt_par2 = min(wgt_par2, xh_wgt2), and wgt_par2 = max(wgt_par2, xl_wgt2).

[0318] Optionally, in this embodiment, values of yh_dist2', yl_dist2', xh_wgt2, and xl_wgt2 are not limited. For example, xl_wgt2 = 0.05, xh_wgt2 = 1.0, yl_dist2' = 2.0, and yh_dist2' = 1.0.

[0319] Optionally, in the foregoing formula, b_wgt2 = xl_wgt2 – a_wgt2 * yh_dist2' may be replaced with b_wgt2 = xh_wgt2 – a_wgt2 * yl_dist2'.

[0320] In this embodiment, xh_wgt2 > x2_wgt1, and yh_dist2' < yl_dist2'.

[0321] In this embodiment, when wgt_par2 is greater than the upper limit value of the second weighting coefficient, wgt_par2 is limited to be the upper limit value of the second weighting coefficient; or when wgt_par2 is less than the lower limit value of the second weighting coefficient, wgt_par2 is limited to the lower limit value of the second weighting coefficient, so as to ensure that a value of wgt_par2 does not exceed a normal value range of the second weighting coefficient, thereby ensuring accuracy of the calculated delay track estimation value of the current frame.

[0322] In addition, after the inter-channel time difference of the current frame is determined, the second weighting coefficient of the current frame is calculated. When

Date Reçue/Date Received 2020-06-23

the delay track estimation value of the next frame is to be determined, the delay track estimation value of the next frame can be determined by using the second weighting coefficient of the current frame, thereby ensuring accuracy of determining the delay track estimation value of the next frame.

[0323] Optionally, in the foregoing embodiments, the buffer is updated regardless of whether the multi-channel signal of the current frame is a valid signal. For example, the inter-channel time difference information of the at least one past frame and/or the weighting coefficient of the at least one past frame in the buffer are/is updated.

[0324] Optionally, the buffer is updated only when the multi-channel signal of the current frame is a valid signal. In this way, validity of data in the buffer is improved.

[0325] The valid signal is a signal whose energy is higher than preset energy, and/or belongs to preset type, for example, the valid signal is a speech signal, or the valid signal is a periodic signal.

[0326] In this embodiment, a voice activity detection (Voice Activity Detection, VAD) algorithm is used to detect whether the multi-channel signal of the current frame is an active frame. If the multi-channel signal of the current frame is an active frame, it indicates that the multi-channel signal of the current frame is the valid signal. If the multi-channel signal of the current frame is not an active frame, it indicates that the multi-channel signal of the current frame is not the valid signal.

[0327] In a manner, it is determined, based on a voice activation detection result of the previous frame of the current frame, whether to update the buffer.

[0328] When the voice activation detection result of the previous frame of the current frame is the active frame, it indicates that it is of great possibility that the current frame is the active frame. In this case, the buffer is updated. When the voice activation detection result of the previous frame of the current frame is not the active frame, it indicates that it is of great possibility that the current frame is not the active frame. In this case, the buffer is not updated.

[0329] Optionally, the voice activation detection result of the previous frame of the current frame is determined based on a voice activation detection result of a primary channel signal of the previous frame of the current frame and a voice activation

Date Reçue/Date Received 2020-06-23

detection result of a secondary channel signal of the previous frame of the current frame.

[0330] If both the voice activation detection result of the primary channel signal of the previous frame of the current frame and the voice activation detection result of the secondary channel signal of the previous frame of the current frame are active frames, the voice activation detection result of the previous frame of the current frame is the active frame. If the voice activation detection result of the primary channel signal of the previous frame of the current frame and/or the voice activation detection result of the secondary channel signal of the previous frame of the current frame are/is not active frames/an active frame, the voice activation detection result of the previous frame of the current frame is not the active frame.

[0331] In another manner, it is determined, based on a voice activation detection result of the current frame, whether to update the buffer.

[0332] When the voice activation detection result of the current frame is an active frame, it indicates that it is of great possibility that the current frame is the active frame. In this case, the audio coding device updates the buffer. When the voice activation detection result of the current frame is not an active frame, it indicates that it is of great possibility that the current frame is not the active frame. In this case, the audio coding device does not update the buffer.

[0333] Optionally, the voice activation detection result of the current frame is determined based on voice activation detection results of a plurality of channel signals of the current frame.

[0334] If the voice activation detection results of the plurality of channel signals of the current frame are all active frames, the voice activation detection result of the current frame is the active frame. If a voice activation detection result of at least one channel of channel signal of the plurality of channel signals of the current frame is not the active frame, the voice activation detection result of the current frame is not the active frame.

[0335] It should be noted that, in this embodiment, description is provided by using an example in which the buffer is updated by using only a criterion about whether the current frame is the active frame. In actual implementation, the buffer may alternatively

Date Reçue/Date Received 2020-06-23

be updated based on at least one of unvoicing or voicing, period or aperiodic, transient or non-transient, and speech or non-speech of the current frame.

[0336] For example, if both the primary channel signal and the secondary channel signal of the previous frame of the current frame are voiced, it indicates that there is a great probability that the current frame is voiced. In this case, the buffer is updated. If at least one of the primary channel signal and the secondary channel signal of the previous frame of the current frame is unvoiced, there is a great probability that the current frame is not voiced. In this case, the buffer is not updated.

[0337] Optionally, based on the foregoing embodiments, an adaptive parameter of a preset window function model may be further determined based on a coding parameter of the previous frame of the current frame. In this way, the adaptive parameter in the preset window function model of the current frame is adaptively adjusted, and accuracy of determining the adaptive window function is improved.

[0338] The coding parameter is used to indicate a type of a multi-channel signal of the previous frame of the current frame, or the coding parameter is used to indicate a type of a multi-channel signal of the previous frame of the current frame in which time-domain downmixing processing is performed, for example, an active frame or an inactive frame, unvoicing or voicing, periodic or aperiodic, transient or non-transient, or speech or music.

[0339] The adaptive parameter includes at least one of an upper limit value of a raised cosine width parameter, a lower limit value of the raised cosine width parameter, an upper limit value of a raised cosine height bias, a lower limit value of the raised cosine height bias, a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine width parameter, a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine width parameter, a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine height bias, and a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine height bias.

[0340] Optionally, when the audio coding device determines the adaptive window

Date Reçue/Date Received 2020-06-23

function in the first manner of determining the adaptive window function, the upper limit value of the raised cosine width parameter is the upper limit value of the first raised cosine width parameter, the lower limit value of the raised cosine width parameter is the lower limit value of the first raised cosine width parameter, the upper limit value of the raised cosine height bias is the upper limit value of the first raised cosine height bias, and the lower limit value of the raised cosine height bias is the lower limit value of the first raised cosine height bias. Correspondingly, the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine width parameter is the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first raised cosine width parameter, the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine width parameter is the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first raised cosine width parameter, the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine height bias is the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first raised cosine height bias, and the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine height bias is the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first raised cosine height bias.

[0341] Optionally, when the audio coding device determines the adaptive window function in the second manner of determining the adaptive window function, the upper limit value of the raised cosine width parameter is the upper limit value of the second raised cosine width parameter, the lower limit value of the raised cosine width parameter is the lower limit value of the second raised cosine width parameter, the upper limit value of the raised cosine height bias is the upper limit value of the second raised cosine height bias, and the lower limit value of the raised cosine height bias is the lower limit value of the second raised cosine height bias. Correspondingly, the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of

Date Reçue/Date Received 2020-06-23

the raised cosine width parameter is the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the second raised cosine width parameter, the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine width parameter is the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the second raised cosine width parameter, the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine height bias is the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the second raised cosine height bias, and the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine height bias is the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the second raised cosine height bias.

[0342] Optionally, in this embodiment, description is provided by using an example in which the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine width parameter is equal to the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine height bias, and the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine width parameter is equal to the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine height bias.

[0343] Optionally, in this embodiment, description is provided by using an example in which the coding parameter of the previous frame of the current frame is used to indicate unvoicing or voicing of the primary channel signal of the previous frame of the current frame and unvoicing or voicing of the secondary channel signal of the previous frame of the current frame.

[0344] (1) Determine the upper limit value of the raised cosine width parameter and the lower limit value of the raised cosine width parameter in the adaptive parameter based on the coding parameter of the previous frame of the current frame.

Date Reçue/Date Received 2020-06-23

[0345] Unvoicing or voicing of the primary channel signal of the previous frame of the current frame and unvoicing or voicing of the secondary channel signal of the previous frame of the current frame are determined based on the coding parameter. If both the primary channel signal and the secondary channel signal are unvoiced, the upper limit value of the raised cosine width parameter is set to a first unvoicing parameter, and the lower limit value of the raised cosine width parameter is set to a second unvoicing parameter, that is, xh_width = xh_width_uv, and xl_width = xl_width_uv.

[0346] If both the primary channel signal and the secondary channel signal are voiced, the upper limit value of the raised cosine width parameter is set to a first voicing parameter, and the lower limit value of the raised cosine width parameter is set to a second voicing parameter, that is, xh_width = xh_width_v, and xl_width = xl_width_v.

[0347] If the primary channel signal is voiced, and the secondary channel signal is unvoiced, the upper limit value of the raised cosine width parameter is set to a third voicing parameter, and the lower limit value of the raised cosine width parameter is set to a fourth voicing parameter, that is, xh_width = xh_width_v2, and xl_width = xl_width_v2.

[0348] If the primary channel signal is unvoiced, and the secondary channel signal is voiced, the upper limit value of the raised cosine width parameter is set to a third unvoicing parameter, and the lower limit value of the raised cosine width parameter is set to a fourth unvoicing parameter, that is, xh_width = xh_width_uv2, and xl_width = xl_width_uv2.

[0349] The first unvoicing parameter xh_width_uv, the second unvoicing parameter xl_width_uv, the third unvoicing parameter xh_width_uv2, the fourth unvoicing parameter xl_width_uv2, the first voicing parameter xh_width_v, the second voicing parameter xl_width_v, the third voicing parameter xh_width_v2, and the fourth voicing parameter xl_width_v2 are all positive numbers, where xh_width_v < xh_width_v2 < xh_width_uv2 < xh_width_uv, and xl_width_uv < xl_width_uv2 < xl_width_v2 < xl_width_v.

[0350] Values of xh_width_v, xh_width_v2, xh_width_uv2, xh_width_uv,

Date Reçue/Date Received 2020-06-23

xl_width_uv, xl_width_uv2, xl_width_v2, and xl_width_v are not limited in this embodiment. For example, xh_width_v = 0.2, xh_width_v2 = 0.25, xh_width_uv2 = 0.35, xh_width_uv = 0.3, xl_width_uv = 0.03, xl_width_uv2 = 0.02, xl_width_v2 = 0.04, and xl_width_v = 0.05.

[0351] Optionally, at least one parameter of the first unvoicing parameter, the second unvoicing parameter, the third unvoicing parameter, the fourth unvoicing parameter, the first voicing parameter, the second voicing parameter, the third voicing parameter, and the fourth voicing parameter is adjusted by using the coding parameter of the previous frame of the current frame.

[0352] For example, that the audio coding device adjusts at least one parameter of the first unvoicing parameter, the second unvoicing parameter, the third unvoicing parameter, the fourth unvoicing parameter, the first voicing parameter, the second voicing parameter, the third voicing parameter, and the fourth voicing parameter based on the coding parameter of a channel signal of the previous frame of the current frame is represented by using the following formulas:

xh_width_uv = fach_uv * xh_width_init; xl_width_uv = facl_uv * xl_width_init;

xh_width_v= fach_v * xh_width_init; xl_width_v = facl_v * xl_width_init;

xh_width_v2 = fach_v2 * xh_width_init; xl_width_v2 = facl_v2 * xl_width_init; and

xh_width_uv2 = fach_uv2 * xh_width_init; and xl_width_uv2 = facl_uv2 * xl_width_init.

[0353] fach_uv, fach_v, fach_v2, fach_uv2, xh_width_init, and xl_width_init are positive numbers determined based on the coding parameter.

[0354] In this embodiment, values of fach_uv, fach_v, fach_v2, fach_uv2, xh_width_init, and xl_width_init are not limited. For example, fach_uv = 1.4, fach_v = 0.8, fach_v2 = 1.0, fach_uv2 = 1.2, xh_width_init = 0.25, and xl_width_init = 0.04.

[0355] (2) Determine the upper limit value of the raised cosine height bias and the lower limit value of the raised cosine height bias in the adaptive parameter based on the coding parameter of the previous frame of the current frame.

Date Reçue/Date Received 2020-06-23

[0356] Unvoicing or voicing of the primary channel signal of the previous frame of the current frame and unvoicing or voicing of the secondary channel signal of the previous frame of the current frame are determined based on the coding parameter. If both the primary channel signal and the secondary channel signal are the unvoiced, the upper limit value of the raised cosine height bias is set to a fifth unvoicing parameter, and the lower limit value of the raised cosine height bias is set to a sixth unvoicing parameter, that is, xh_bias = xh_bias_uv, and xl_bias = xl_bias_uv.

[0357] If both the primary channel signal and the secondary channel signal are voiced, the upper limit value of the raised cosine height bias is set to a fifth voicing parameter, and the lower limit value of the raised cosine height bias is set to a sixth voicing parameter, that is, xh_bias = xh_bias_v, and xl_bias = xl_bias_v.

[0358] If the primary channel signal is voiced, and the secondary channel signal is unvoiced, the upper limit value of the raised cosine height bias is set to a seventh voicing parameter, and the lower limit value of the raised cosine height bias is set to an eighth voicing parameter, that is, xh_bias = xh_bias_v2, and xl_bias = xl_bias_v2.

[0359] If the primary channel signal is unvoiced, and the secondary channel signal is voiced, the upper limit value of the raised cosine height bias is set to a seventh unvoicing parameter, and the lower limit value of the raised cosine height bias is set to an eighth unvoicing parameter, that is, xh_bias = xh_bias_uv2, and xl_bias = xl_bias_uv2.

[0360] The fifth unvoicing parameter xh_bias_uv, the sixth unvoicing parameter xl_bias_uv, the seventh unvoicing parameter xh_bias_uv2, the eighth unvoicing parameter xl_bias_uv2, the fifth voicing parameter xh_bias_v, the sixth voicing parameter xl_bias_v, the seventh voicing parameter xh_bias_v2, and the eighth voicing parameter xl_bias_v2 are all positive numbers, where xh_bias_v < xh_bias_v2 < xh_bias_uv2 < xh_bias_uv, xl_bias_v < xl_bias_v2 < xl_bias_uv2 < xl_bias_uv, xh_bias is the upper limit value of the raised cosine height bias, and xl_bias is the lower limit value of the raised cosine height bias.

[0361] In this embodiment, values of xh_bias_v, xh_bias_v2, xh_bias_uv2, xh_bias_uv, xl_bias_v, xl_bias_v2, xl_bias_uv2, and xl_bias_uv are not limited. For

Date Reçue/Date Received 2020-06-23

example, xh_bias_v = 0.8, xl_bias_v = 0.5, xh_bias_v2 = 0.7, xl_bias_v2 = 0.4, xh_bias_uv = 0.6, xl_bias_uv = 0.3, xh_bias_uv2 = 0.5, and xl_bias_uv2 = 0.2

[0362] Optionally, at least one of the fifth unvoicing parameter, the sixth unvoicing parameter, the seventh unvoicing parameter, the eighth unvoicing parameter, the fifth voicing parameter, the sixth voicing parameter, the seventh voicing parameter, and the eighth voicing parameter is adjusted based on the coding parameter of a channel signal of the previous frame of the current frame.

[0363] For example, the following formula is used for representation:

xh_bias_uv = fach_uv' * xh_bias_init; xl_bias_uv = facl_uv' * xl_bias_init;

xh_bias_v = fach_v' * xh_bias_init; xl_bias_v = facl_v' * xl_bias_init;

xh_bias_v2 = fach_v2' * xh_bias_init; xl_bias_v2 = facl_v2' * xl_bias_init;

xh_bias_uv2 = fach_uv2' * xh_bias_init; and xl_bias_uv2 = facl_uv2' * xl_bias_init.

[0364] fach_uv', fach_v', fach_v2', fach_uv2', xh_bias_init, and xl_bias_init are positive numbers determined based on the coding parameter.

[0365] In this embodiment, values of fach_uv', fach_v', fach_v2', fach_uv2', xh_bias_init, and xl_bias_init are not limited. For example, fach_v' = 1.15, fach_v2' = 1.0, fach_uv2' = 0.85, fach_uv' = 0.7, xh_bias_init = 0.7, and xl_bias_init = 0.4.

[0366] (3) Determine, based on the coding parameter of the previous frame of the current frame, the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine width parameter, and the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine width parameter in the adaptive parameter.

[0367] The unvoiced and voiced primary channel signals of the previous frame of the current frame and the unvoiced and voiced secondary channel signals of the previous frame of the current frame are determined based on the coding parameter. If both the primary channel signal and the secondary channel signal are unvoiced, the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine width parameter is set to a ninth unvoicing parameter, and the smoothed inter-channel time difference estimation deviation corresponding to

Date Reçue/Date Received 2020-06-23

the lower limit value of the raised cosine width parameter is set to a tenth unvoicing parameter, that is, yh_dist = yh_dist_uv, and yl_dist = yl_dist_uv.

[0368] If both the primary channel signal and the secondary channel signal are voiced, the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine width parameter is set to a ninth voicing parameter, and the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine width parameter is set to a tenth voicing parameter, that is, yh_dist = yh_dist_v, and yl_dist = yl_dist_v.

[0369] If the primary channel signal is voiced, and the secondary channel signal is unvoiced, the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine width parameter is set to an eleventh voicing parameter, and the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine width parameter is set to a twelfth voicing parameter, that is, yh_dist = yh_dist_v2, and yl_dist = yl_dist_v2.

[0370] If the primary channel signal is unvoiced, and the secondary channel signal is voiced, the smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the raised cosine width parameter is set to an eleventh unvoicing parameter, and the smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the raised cosine width parameter is set to a twelfth unvoicing parameter, that is, yh_dist = yh_dist_uv2, and yl_dist = yl_dist_uv2.

[0371] The ninth unvoicing parameter yh_dist_uv, the tenth unvoicing parameter yl_dist_uv, the eleventh unvoicing parameter yh_dist_uv2, the twelfth unvoicing parameter yl_dist_uv2, the ninth voicing parameter yh_dist_v, the tenth voicing parameter yl_dist_v, the eleventh voicing parameter yh_dist_v2, and the twelfth voicing parameter yl_dist_v2 are all positive numbers, where yh_dist_v < yh_dist_v2 < yh_dist_uv2 < yh_dist_uv, and yl_dist_uv < yl_dist_uv2 < yl_dist_v2 < yl_dist_v.

[0372] In this embodiment, values of yh_dist_v, yh_dist_v2, yh_dist_uv2, yh_dist_uv, yl_dist_uv, yl_dist_uv2, yl_dist_v2, and yl_dist_v are not limited.

Date Reçue/Date Received 2020-06-23

[0373] Optionally, at least one parameter of the ninth unvoicing parameter, the tenth unvoicing parameter, the eleventh unvoicing parameter, the twelfth unvoicing parameter, the ninth voicing parameter, the tenth voicing parameter, the eleventh voicing parameter, and the twelfth voicing parameter is adjusted by using the coding parameter of the previous frame of the current frame.

[0374] For example, the following formula is used for representation:

yh_dist_uv = fach_uv" * yh_dist_init; yl_dist_uv = facl_uv" * yl_dist_init;

yh_dist_v = fach_v" * yh_dist_init; yl_dist_v = facl_v" * yl_dist_init;

yh_dist_v2 = fach_v2" * yh_dist_init; yl_dist_v2 = facl_v2" * yl_dist_init;

yh_dist_uv2 = fach_uv2" * yh_dist_init; and yl_dist_uv2 = facl_uv2" * yl_dist_init.

[0375] fach_uv", fach_v", fach_v2", fach_uv2", yh_dist_init, and yl_dist_init are positive numbers determined based on the coding parameter, and values of the parameters are not limited in this embodiment.

[0376] In this embodiment, the adaptive parameter in the preset window function model is adjusted based on the coding parameter of the previous frame of the current frame, so that an appropriate adaptive window function is determined adaptively based on the coding parameter of the previous frame of the current frame, thereby improving accuracy of generating an adaptive window function, and improving accuracy of estimating an inter-channel time difference.

[0377] Optionally, based on the foregoing embodiments, before step 301, time-domain preprocessing is performed on the multi-channel signal.

[0378] Optionally, the multi-channel signal of the current frame in this embodiment of this application is a multi-channel signal input to the audio coding device, or a multi-channel signal obtained through preprocessing after the multi-channel signal is input to the audio coding device.

[0379] Optionally, the multi-channel signal input to the audio coding device may be collected by a collection component in the audio coding device, or may be collected by a collection device independent of the audio coding device, and is sent to the audio coding device.

Date Reçue/Date Received 2020-06-23

[0380] Optionally, the multi-channel signal input to the audio coding device is a multi-channel signal obtained after through analog-to-digital (Analog to Digital, A/D) conversion. Optionally, the multi-channel signal is a pulse code modulation (Pulse Code Modulation, PCM) signal.

[0381] A sampling frequency of the multi-channel signal may be 8 kHz, 16 kHz, 32 kHz, 44.1 kHz, 48 kHz, or the like. This is not limited in this embodiment.

[0382] For example, the sampling frequency of the multi-channel signal is 16 kHz. In this case, duration of a frame of multi-channel signals is 20 ms, and a frame length is denoted as N, where N = 320, in other words, the frame length is 320 sampling points. The multi-channel signal of the current frame includes a left channel signal and a right channel signal, the left channel signal is denoted as $x_L(n)$, and the right channel signal is denoted as $x_R(n)$, where n is a sampling point sequence number, and n = 0, 1, 2, ..., and (N – 1).

[0383] Optionally, if high-pass filtering processing is performed on the current frame, a processed left channel signal is denoted as $x_{L\_HP}(n)$, and a processed right channel signal is denoted as $x_{R\_HP}(n)$, where n is a sampling point sequence number, and n = 0, 1, 2, ..., and (N – 1).

[0384] FIG. 11 is a schematic structural diagram of an audio coding device according to an example embodiment of this application. In this embodiment of this application, the audio coding device may be an electronic device that has an audio collection and audio signal processing function, such as a mobile phone, a tablet computer, a laptop portable computer, a desktop computer, a Bluetooth speaker, a pen recorder, and a wearable device, or may be a network element that has an audio signal processing capability in a core network and a radio network. This is not limited in this embodiment.

[0385] The audio coding device includes a processor 701, a memory 702, and a bus 703.

[0386] The processor 701 includes one or more processing cores, and the processor 701 runs a software program and a module, to perform various function applications and process information.

Date Reçue/Date Received 2020-06-23

[0387] The memory 702 is connected to the processor 701 by using the bus 703. The memory 702 stores an instruction necessary for the audio coding device.

[0388] The processor 701 is configured to execute the instruction in the memory 702 to implement the delay estimation method provided in the method embodiments of this application.

[0389] In addition, the memory 702 may be implemented by any type of volatile or non-volatile storage device or a combination thereof, such as a static random access memory (SRAM), an electrically erasable programmable read-only memory (EEPROM), an erasable programmable read-only memory (EPROM), a programmable read-only memory (PROM), a read-only memory (ROM), a magnetic memory, a flash memory, a magnetic disk, or an optic disc.

[0390] The memory 702 is further configured to buffer inter-channel time difference information of at least one past frame and/or a weighting coefficient of the at least one past frame.

[0391] Optionally, the audio coding device includes a collection component, and the collection component is configured to collect a multi-channel signal.

[0392] Optionally, the collection component includes at least one microphone. Each microphone is configured to collect one channel of channel signal.

[0393] Optionally, the audio coding device includes a receiving component, and the receiving component is configured to receive a multi-channel signal sent by another device.

[0394] Optionally, the audio coding device further has a decoding function.

[0395] It may be understood that FIG. 11 shows merely a simplified design of the audio coding device. In another embodiment, the audio coding device may include any quantity of transmitters, receivers, processors, controllers, memories, communications units, display units, play units, and the like. This is not limited in this embodiment.

[0396] Optionally, this application provides a computer readable storage medium. The computer readable storage medium stores an instruction. When the instruction is run on the audio coding device, the audio coding device is enabled to perform the delay estimation method provided in the foregoing embodiments.

Date Reçue/Date Received 2020-06-23

[0397] FIG. 12 is a block diagram of a delay estimation apparatus according to an embodiment of this application. The delay estimation apparatus may be implemented as all or a part of the audio coding device shown in FIG. 11 by using software, hardware, or a combination thereof. The delay estimation apparatus may include a cross-correlation coefficient determining unit 810, a delay track estimation unit 820, an adaptive function determining unit 830, a weighting unit 840, and an inter-channel time difference determining unit 850.

[0398] The cross-correlation coefficient determining unit 810 is configured to determine a cross-correlation coefficient of a multi-channel signal of a current frame.

[0399] The delay track estimation unit 820 is configured to determine a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame.

[0400] The adaptive function determining unit 830 is configured to determine an adaptive window function of the current frame.

[0401] The weighting unit 840 is configured to perform weighting on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame, to obtain a weighted cross-correlation coefficient.

[0402] The inter-channel time difference determining unit 850 is configured to determine an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient.

[0403] Optionally, the adaptive function determining unit 830 is further configured to:

calculate a first raised cosine width parameter based on a smoothed inter-channel time difference estimation deviation of a previous frame of the current frame;

calculate a first raised cosine height bias based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame; and

determine the adaptive window function of the current frame based on the first raised cosine width parameter and the first raised cosine height bias.

Date Reçue/Date Received 2020-06-23

[0404] Optionally, the apparatus further includes: a smoothed inter-channel time difference estimation deviation determining unit 860.

[0405] The smoothed inter-channel time difference estimation deviation determining unit 860 is configured to calculate a smoothed inter-channel time difference estimation deviation of the current frame based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, the delay track estimation value of the current frame, and the inter-channel time difference of the current frame.

[0406] Optionally, the adaptive function determining unit 830 is further configured to:

determine an initial value of the inter-channel time difference of the current frame based on the cross-correlation coefficient;

calculate an inter-channel time difference estimation deviation of the current frame based on the delay track estimation value of the current frame and the initial value of the inter-channel time difference of the current frame; and

determine the adaptive window function of the current frame based on the inter-channel time difference estimation deviation of the current frame.

[0407] Optionally, the adaptive function determining unit 830 is further configured to:

calculate a second raised cosine width parameter based on the inter-channel time difference estimation deviation of the current frame;

calculate a second raised cosine height bias based on the inter-channel time difference estimation deviation of the current frame; and

determine the adaptive window function of the current frame based on the second raised cosine width parameter and the second raised cosine height bias.

[0408] Optionally, the apparatus further includes an adaptive parameter determining unit 870.

[0409] The adaptive parameter determining unit 870 is configured to determine an adaptive parameter of the adaptive window function of the current frame based on a coding parameter of the previous frame of the current frame.

Date Reçue/Date Received 2020-06-23

[0410] Optionally, the delay track estimation unit 820 is further configured to:

perform delay track estimation based on the buffered inter-channel time difference information of the at least one past frame by using a linear regression method, to determine the delay track estimation value of the current frame.

[0411] Optionally, the delay track estimation unit 820 is further configured to:

perform delay track estimation based on the buffered inter-channel time difference information of the at least one past frame by using a weighted linear regression method, to determine the delay track estimation value of the current frame.

[0412] Optionally, the apparatus further includes an update unit 880.

[0413] The update unit 880 is configured to update the buffered inter-channel time difference information of the at least one past frame.

[0414] Optionally, the buffered inter-channel time difference information of the at least one past frame is an inter-channel time difference smoothed value of the at least one past frame, and the update unit 880 is configured to:

determine an inter-channel time difference smoothed value of the current frame based on the delay track estimation value of the current frame and the inter-channel time difference of the current frame; and

update a buffered inter-channel time difference smoothed value of the at least one past frame based on the inter-channel time difference smoothed value of the current frame.

[0415] Optionally, the update unit 880 is further configured to:

determine, based on a voice activation detection result of the previous frame of the current frame or a voice activation detection result of the current frame, whether to update the buffered inter-channel time difference information of the at least one past frame.

[0416] Optionally, the update unit 880 is further configured to:

update a buffered weighting coefficient of the at least one past frame, where the weighting coefficient of the at least one past frame is a coefficient in the weighted linear regression method.

[0417] Optionally, when the adaptive window function of the current frame is

Date Reçue/Date Received 2020-06-23

determined based on a smoothed inter-channel time difference of the previous frame of the current frame, the update unit 880 is further configured to:

calculate a first weighting coefficient of the current frame based on the smoothed inter-channel time difference estimation deviation of the current frame; and

update a buffered first weighting coefficient of the at least one past frame based on the first weighting coefficient of the current frame.

[0418] Optionally, when the adaptive window function of the current frame is determined based on the smoothed inter-channel time difference estimation deviation of the current frame, the update unit 880 is further configured to:

calculate a second weighting coefficient of the current frame based on the inter-channel time difference estimation deviation of the current frame; and

update a buffered second weighting coefficient of the at least one past frame based on the second weighting coefficient of the current frame.

[0419] Optionally, the update unit 880 is further configured to:

when the voice activation detection result of the previous frame of the current frame is an active frame or the voice activation detection result of the current frame is an active frame, update the buffered weighting coefficient of the at least one past frame.

[0420] For related details, refer to the foregoing method embodiments.

[0421] Optionally, the foregoing units may be implemented by a processor in the audio coding device by executing an instruction in a memory.

[0422] It may be clearly understood by a person of ordinary skill in the art that, for ease and brief description, for a detailed working process of the foregoing apparatus and units, refer to a corresponding process in the foregoing method embodiments, and details are not described herein again.

[0423] In the embodiments provided in the present application, it should be understood that the disclosed apparatus and method may be implemented in other manners. For example, the described apparatus embodiments are merely examples. For example, the unit division may merely be logical function division and may be other division in actual implementation. For example, a plurality of units or components may

Date Reçue/Date Received 2020-06-23

be combined or integrated into another system, or some features may be ignored or not performed.

[0424] The foregoing descriptions are merely optional implementations of this application, but are not intended to limit the protection scope of this application. Any variation or replacement readily figured out by a person skilled in the art within the technical scope disclosed in this application shall fall within the protection scope of this application. Therefore, the protection scope of this application shall be subject to the protection scope of the claims.

Date Reçue/Date Received 2020-06-23

CLAIMS

What is claimed is:

1. A delay estimation method performed by an audio signal coding device, wherein the method comprises:

determining a cross-correlation coefficient of a multi-channel audio signal of a current frame;

determining a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame;

determining an adaptive window function of the current frame;

performing weighting on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame, to obtain a weighted cross-correlation coefficient;

determining an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient; and

encoding the inter-channel time difference of the current frame.

2. The method according to claim 1, wherein the determining an adaptive window function of the current frame comprises:

calculating a first raised cosine width parameter based on a smoothed inter-channel time difference estimation deviation of a previous frame of the current frame;

calculating a first raised cosine height bias based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame; and

determining the adaptive window function of the current frame based on the first raised cosine width parameter and the first raised cosine height bias.

3. The method according to claim 2, wherein the first raised cosine width parameter is obtained through calculation by using the following calculation formulas:

$$win\_width1 = TRUNC(width\_par1 * (A * L\_NCSHIFT\_DS + 1)), \text{ and}$$

$$width\_par1 = a\_width1 * smooth\_dist\_reg + b\_width1; \text{ wherein}$$

$$a\_width1 = (xh\_width1 - xl\_width1)/(yh\_dist1 - yl\_dist1),$$

Date Reçue/Date Received 2021-06-10

$$b\_width1 = xh\_width1 - a\_width1 * yh\_dist1,$$

wherein win_width1 is the first raised cosine width parameter, TRUNC indicates rounding a value, L_NCSHIFT_DS is a maximum value of an absolute value of an inter-channel time difference, A is a preset constant, A is greater than or equal to 4, xh_width1 is an upper limit value of the first raised cosine width parameter, xl_width1 is a lower limit value of the first raised cosine width parameter, yh_dist1 is a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first raised cosine width parameter, yl_dist1 is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first raised cosine width parameter, smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, and xh_width1, xl_width1, yh_dist1, and yl_dist1 are all positive numbers.

4. The method according to claim 3, wherein

$$width\_par1 = \min(width\_par1, xh\_width1), \text{ and}$$

$$width\_par1 = \max(width\_par1, xl\_width1),$$

wherein min represents taking of a minimum value, and max represents taking of a maximum value.

5. The method according to claim 3 or 4, wherein the first raised cosine height bias is obtained through calculation by using the following calculation formula:

$$win\_bias1 = a\_bias1 * smooth\_dist\_reg + b\_bias1, \text{ wherein}$$

$$a\_bias1 = (xh\_bias1 - xl\_bias1)/(yh\_dist2 - yl\_dist2),$$

$$b\_bias1 = xh\_bias1 - a\_bias1 * yh\_dist2,$$

wherein win_bias1 is the first raised cosine height bias, xh_bias1 is an upper limit value of the first raised cosine height bias, xl_bias1 is a lower limit value of the first raised cosine height bias, yh_dist2 is a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first raised cosine height bias, yl_dist2 is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first raised cosine height bias, smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, and yh_dist2, yl_dist2, xh_bias1, and xl_bias1

Date Reçue/Date Received 2021-06-10

are all positive numbers.

6. The method according to claim 5, wherein

$$win\_bias1 = \min(win\_bias1, xh\_bias1), \text{ and}$$

$$win\_bias1 = \max(win\_bias1, xl\_bias1),$$

wherein min represents taking of a minimum value, and max represents taking of a maximum value.

7. The method according to claim 5 or 6, wherein yh_dist2 = yh_dist1, and yl_dist2 = yl_dist1.

8. The method according to any one of claims 1 to 7, wherein the adaptive window function is represented by using the following formulas:

when $0 \le k \le TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width1 - 1$,

$$loc\_weight\_win(k) = win\_bias1;$$

when $TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width1 \le k \le TRUNC(A * L\_NCSHIFT\_DS/2) + 2 * win\_width1 - 1$,

$$loc\_weight\_win(k) = 0.5 * (1 + win\_bias1) + 0.5 * (1 - win\_bias1) * \cos(\pi * (k - TRUNC(A * L\_NCSHIFT\_DS/2))/(2 * win\_width1));$$ and

when $TRUNC(A * L\_NCSHIFT\_DS/2) + 2 * win\_width1 \le k \le A * L\_NCSHIFT\_DS$,

$$loc\_weight\_win(k) = win\_bias1;$$ wherein

wherein loc_weight_win(k) is used to represent the adaptive window function, wherein k = 0, 1, ..., A * L_NCSHIFT_DS; A is the preset constant and is greater than or equal to 4; L_NCSHIFT_DS is the maximum value of the absolute value of the inter-channel time difference; win_width1 is the first raised cosine width parameter; and win_bias1 is the first raised cosine height bias.

9. The method according to any one of claims 2 to 8, after the determining an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient, further comprising:

calculating a smoothed inter-channel time difference estimation deviation of the current frame based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, the delay track estimation value of the

Date Reçue/Date Received 2021-06-10

current frame, and the inter-channel time difference of the current frame; and

the smoothed inter-channel time difference estimation deviation of the current frame is obtained through calculation by using the following calculation formulas:

$$smooth\_dist\_reg\_update = (1 - \gamma) * smooth\_dist\_reg + \gamma * dist\_reg', \text{ and}$$

$$dist\_reg' = |reg\_prv\_corr - cur\_itd|,$$

wherein smooth_dist_reg_update is the smoothed inter-channel time difference estimation deviation of the current frame; γ is a first smoothing factor, and $0 < \gamma < 1$; smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame; reg_prv_corr is the delay track estimation value of the current frame; and cur_itd is the inter-channel time difference of the current frame.

10. The method according to claim 1, wherein the determining an adaptive window function of the current frame comprises:

determining an initial value of the inter-channel time difference of the current frame based on the cross-correlation coefficient;

calculating an inter-channel time difference estimation deviation of the current frame based on the delay track estimation value of the current frame and the initial value of the inter-channel time difference of the current frame; and

determining the adaptive window function of the current frame based on the inter-channel time difference estimation deviation of the current frame; and

the inter-channel time difference estimation deviation of the current frame is obtained through calculation by using the following calculation formula:

$$dist\_reg = |reg\_prv\_corr - cur\_itd\_init|,$$

wherein dist_reg is the inter-channel time difference estimation deviation of the current frame, reg_prv_corr is the delay track estimation value of the current frame, and cur_itd_init is the initial value of the inter-channel time difference of the current frame.

11. The method according to claim 10, wherein the determining the adaptive window function of the current frame based on the inter-channel time difference estimation deviation of the current frame comprises:

calculating a second raised cosine width parameter based on the inter-channel time difference estimation deviation of the current frame;

Date Reçue/Date Received 2021-06-10

calculating a second raised cosine height bias based on the inter-channel time difference estimation deviation of the current frame; and

determining the adaptive window function of the current frame based on the second raised cosine width parameter and the second raised cosine height bias.

12. The method according to any one of claims 1 to 11, wherein the weighted cross-correlation coefficient is obtained through calculation by using the following calculation formula:

$$c\_weight(x) = c(x) * loc\_weight\_win(x - TRUNC(reg\_prv\_corr) + TRUNC(A * L\_NCSHIFT\_DS/2) - L\_NCSHIFT\_DS),$$

wherein c_weight(x) is the weighted cross-correlation coefficient; c(x) is the cross-correlation coefficient; loc_weight_win is the adaptive window function of the current frame; TRUNC indicates rounding a value; reg_prv_corr is the delay track estimation value of the current frame; x is an integer greater than or equal to zero and less than or equal to 2 * L_NCSHIFT_DS; and L_NCSHIFT_DS is the maximum value of the absolute value of the inter-channel time difference.

13. The method according to any one of claims 1 to 12, before the determining an adaptive window function of the current frame, further comprising:

determining an adaptive parameter of the adaptive window function of the current frame based on a coding parameter of the previous frame of the current frame, wherein the coding parameter is used to indicate a type of a multi-channel signal of the previous frame of the current frame, or the coding parameter is used to indicate a type of a multi-channel signal of the previous frame of the current frame on which time-domain downmixing processing is performed; and the adaptive parameter is used to determine the adaptive window function of the current frame.

14. The method according to any one of claims 1 to 13, wherein the determining a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame comprises:

performing delay track estimation based on the buffered inter-channel time difference information of the at least one past frame by using a linear regression method, to determine the delay track estimation value of the current frame.

Date Reçue/Date Received 2021-06-10

15. The method according to any one of claims 1 to 13, wherein the determining a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame comprises:

performing delay track estimation based on the buffered inter-channel time difference information of the at least one past frame by using a weighted linear regression method, to determine the delay track estimation value of the current frame.

16. The method according to any one of claims 1 to 15, after the determining an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient, further comprising:

updating the buffered inter-channel time difference information of the at least one past frame, wherein the inter-channel time difference information of the at least one past frame is an inter-channel time difference smoothed value of the at least one past frame or an inter-channel time difference of the at least one past frame.

17. The method according to claim 16, wherein the inter-channel time difference information of the at least one past frame is the inter-channel time difference smoothed value of the at least one past frame, and the updating the buffered inter-channel time difference information of the at least one past frame comprises:

determining an inter-channel time difference smoothed value of the current frame based on the delay track estimation value of the current frame and the inter-channel time difference of the current frame; and

updating a buffered inter-channel time difference smoothed value of the at least one past frame based on the inter-channel time difference smoothed value of the current frame; wherein

the inter-channel time difference smoothed value of the current frame is obtained by using the following calculation formula:

$$cur\_itd\_smooth = \varphi * reg\_prv\_corr + (1 - \varphi) * cur\_itd,$$ wherein

cur_itd_smooth is the inter-channel time difference smoothed value of the current frame, φ is a second smoothing factor and is a constant greater than or equal to 0 and less than or equal to 1, reg_prv_corr is the delay track estimation value of the current frame, and cur_itd is the inter-channel time difference of the current frame.

Date Reçue/Date Received 2021-06-10

18. The method according to claim 16 or 17, wherein the updating the buffered inter-channel time difference information of the at least one past frame comprises:

when a voice activation detection result of the previous frame of the current frame is an active frame or a voice activation detection result of the current frame is an active frame, updating the buffered inter-channel time difference information of the at least one past frame.

19. The method according to any one of claims 15 to 18, after the determining an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient, further comprising:

updating a buffered weighting coefficient of the at least one past frame, wherein the weighting coefficient of the at least one past frame is a weighting coefficient in the weighted linear regression method.

20. The method according to claim 19, wherein when the adaptive window function of the current frame is determined based on a smoothed inter-channel time difference of the previous frame of the current frame, the updating a buffered weighting coefficient of the at least one past frame comprises:

calculating a first weighting coefficient of the current frame based on the smoothed inter-channel time difference estimation deviation of the current frame; and

updating a buffered first weighting coefficient of the at least one past frame based on the first weighting coefficient of the current frame, wherein

the first weighting coefficient of the current frame is obtained through calculation by using the following calculation formulas:

$$wgt\_par1 = a\_wgt1 * smooth\_dist\_reg\_update + b\_wgt1,$$

$$a\_wgt1 = (xl\_wgt1 - xh\_wgt1)/(yh\_dist1' - yl\_dist1'), \text{ and}$$

$$b\_wgt1 = xl\_wgt1 - a\_wgt1 * yh\_dist1',$$

wherein wgt_par1 is the first weighting coefficient of the current frame, smooth_dist_reg_update is the smoothed inter-channel time difference estimation deviation of the current frame, xh_wgt is an upper limit value of the first weighting coefficient, xl_wgt is a lower limit value of the first weighting coefficient, yh_dist1' is a smoothed inter-channel time difference estimation deviation corresponding to the

Date Reçue/Date Received 2021-06-10

upper limit value of the first weighting coefficient, yl_dist1' is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first weighting coefficient, and yh_dist1', yl_dist1', xh_wgt1, and xl_wgt1 are all positive numbers.

21. The method according to claim 20, wherein

$$wgt\_par1 = \min(wgt\_par1, xh\_wgt1), \text{ and}$$

$$wgt\_par1 = \max(wgt\_par1, xl\_wgt1),$$

wherein min represents taking of a minimum value, and max represents taking of a maximum value.

22. The method according to claim 19, wherein when the adaptive window function of the current frame is determined based on the inter-channel time difference estimation deviation of the current frame, the updating a buffered weighting coefficient of the at least one past frame comprises:

calculating a second weighting coefficient of the current frame based on the inter-channel time difference estimation deviation of the current frame; and

updating a buffered second weighting coefficient of the at least one past frame based on the second weighting coefficient of the current frame.

23. The method according to any one of claims 19 to 22, wherein the updating a buffered weighting coefficient of the at least one past frame comprises:

when a voice activation detection result of the previous frame of the current frame is an active frame or a voice activation detection result of the current frame is an active frame, updating the buffered weighting coefficient of the at least one past frame.

24. A delay estimation apparatus, wherein the apparatus comprises:

a cross-correlation coefficient determining unit, configured to determine a cross-correlation coefficient of a multi-channel audio signal of a current frame;

a delay track estimation unit, configured to determine a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame;

an adaptive function determining unit, configured to determine an adaptive window function of the current frame;

Date Reçue/Date Received 2021-06-10

a weighting unit, configured to perform weighting on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame, to obtain a weighted cross-correlation coefficient;

an inter-channel time difference determining unit, configured to determine an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient; and

a unit configured to encode the inter-channel time difference of the current frame.

25. The apparatus according to claim 24, wherein the adaptive function determining unit is configured to:

calculate a first raised cosine width parameter based on a smoothed inter-channel time difference estimation deviation of a previous frame of the current frame;

calculate a first raised cosine height bias based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame; and

determine the adaptive window function of the current frame based on the first raised cosine width parameter and the first raised cosine height bias.

26. The apparatus according to claim 25, wherein the first raised cosine width parameter is obtained through calculation by using the following calculation formulas:

$$win\_width1 = TRUNC(width\_par1 * (A * L\_NCSHIFT\_DS + 1)), \text{ and}$$

$$width\_par1 = a\_width1 * smooth\_dist\_reg + b\_width1; \text{ wherein}$$

$$a\_width1 = (xh\_width1 - xl\_width1)/(yh\_dist1 - yl\_dist1),$$

$$b\_width1 = xh\_width1 - a\_width1 * yh\_dist1,$$

win_width1 is the first raised cosine width parameter, TRUNC indicates rounding a value, L_NCSHIFT_DS is a maximum value of an absolute value of an inter-channel time difference, A is a preset constant, A is greater than or equal to 4, xh_width1 is an upper limit value of the first raised cosine width parameter, xl_width1 is a lower limit value of the first raised cosine width parameter, yh_dist1 is a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first raised cosine width parameter, yl_dist1 is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first raised cosine

Date Reçue/Date Received 2021-06-10

width parameter, smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, and xh_width1, xl_width1, yh_dist1, and yl_dist1 are all positive numbers.

27. The apparatus according to claim 26, wherein

$$width\_par1 = \min(width\_par1, xh\_width1), \text{ and}$$

$$width\_par1 = \max(width\_par1, xl\_width1), \text{ wherein}$$

min represents taking of a minimum value, and max represents taking of a maximum value.

28. The apparatus according to claim 26 or 27, wherein the first raised cosine height bias is obtained through calculation by using the following calculation formula:

$$win\_bias1 = a\_bias1 * smooth\_dist\_reg + b\_bias1, \text{ wherein}$$

$$a\_bias1 = (xh\_bias1 - xl\_bias1)/(yh\_dist2 - yl\_dist2),$$

$$b\_bias1 = xh\_bias1 - a\_bias1 * yh\_dist2,$$

win_bias1 is the first raised cosine height bias, xh_bias1 is an upper limit value of the first raised cosine height bias, xl_bias1 is a lower limit value of the first raised cosine height bias, yh_dist2 is a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first raised cosine height bias, yl_dist2 is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first raised cosine height bias, smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, and yh_dist2, yl_dist2, xh_bias1, and xl_bias1 are all positive numbers.

29. The apparatus according to claim 28, wherein

$$win\_bias1 = \min(win\_bias1, xh\_bias1), \text{ and}$$

$$win\_bias1 = \max(win\_bias1, xl\_bias1), \text{ wherein}$$

min represents taking of a minimum value, and max represents taking of a maximum value.

30. The apparatus according to claim 28 or 29, wherein yh_dist2 = yh_dist1, and yl_dist2 = yl_dist1.

31. The apparatus according to any one of claims 24 to 30, wherein the adaptive window function is represented by using the following formulas:

Date Reçue/Date Received 2021-06-10

when $0 \leq k \leq TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width1 - 1$,

$$loc\_weight\_win(k) = win\_bias1;$$

when $TRUNC(A * L\_NCSHIFT\_DS/2) - 2 * win\_width1 \leq k \leq TRUNC(A * L\_NCSHIFT\_DS/2) + 2 * win\_width1 - 1$,

$$loc\_weight\_win(k) = 0.5 * (1 + win\_bias1) + 0.5 * (1 - win\_bias1) * \cos(\pi * (k - TRUNC(A * L\_NCSHIFT\_DS/2))/(2 * win\_width1));$$ and

when $TRUNC(A * L\_NCSHIFT\_DS/2) + 2 * win\_width1 \leq k \leq A * L\_NCSHIFT\_DS$,

$$loc\_weight\_win(k) = win\_bias1;$$ wherein

loc_weight_win(k) is used to represent the adaptive window function, wherein k = 0, 1, ..., A * L_NCSHIFT_DS; A is the preset constant and is greater than or equal to 4; L_NCSHIFT_DS is the maximum value of the absolute value of the inter-channel time difference; win_width1 is the first raised cosine width parameter; and win_bias1 is the first raised cosine height bias.

32. The apparatus according to any one of claims 25 to 31, wherein the apparatus further comprises:

a smoothed inter-channel time difference estimation deviation determining unit, configured to calculate a smoothed inter-channel time difference estimation deviation of the current frame based on the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame, the delay track estimation value of the current frame, and the inter-channel time difference of the current frame; and

the smoothed inter-channel time difference estimation deviation of the current frame is obtained through calculation by using the following calculation formulas:

$$smooth\_dist\_reg\_update = (1 - \gamma) * smooth\_dist\_reg + \gamma * dist\_reg',$$ and

$$dist\_reg' = |reg\_prv\_corr - cur\_itd|,$$ wherein

smooth_dist_reg_update is the smoothed inter-channel time difference estimation deviation of the current frame; γ is a first smoothing factor, and $0 < \gamma < 1$; smooth_dist_reg is the smoothed inter-channel time difference estimation deviation of the previous frame of the current frame; reg_prv_corr is the delay track estimation value of the current frame; and cur_itd is the inter-channel time difference of the current frame.

Date Reçue/Date Received 2021-06-10

33. The apparatus according to any one of claims 24 to 32, wherein the weighted cross-correlation coefficient is obtained through calculation by using the following calculation formula:

c_weight(x) = c(x) * loc_weight_win(x – TRUNC(reg_prv_corr) + TRUNC(A * L_NCSHIFT_DS/2) – L_NCSHIFT_DS), wherein

c_weight(x) is the weighted cross-correlation coefficient; c(x) is the cross-correlation coefficient; loc_weight_win is the adaptive window function of the current frame; TRUNC indicates rounding a value; reg_prv_corr is the delay track estimation value of the current frame; x is an integer greater than or equal to zero and less than or equal to 2 * L_NCSHIFT_DS; and L_NCSHIFT_DS is the maximum value of the absolute value of the inter-channel time difference.

34. The apparatus according to any one of claims 24 to 33, wherein the delay track estimation unit is configured to:

perform delay track estimation based on the buffered inter-channel time difference information of the at least one past frame by using a linear regression method, to determine the delay track estimation value of the current frame.

35. The apparatus according to any one of claims 24 to 33, wherein the delay track estimation unit is configured to:

perform delay track estimation based on the buffered inter-channel time difference information of the at least one past frame by using a weighted linear regression method, to determine the delay track estimation value of the current frame.

36. The apparatus according to any one of claims 24 to 35, wherein the apparatus further comprises:

an update unit, configured to update the buffered inter-channel time difference information of the at least one past frame, wherein the inter-channel time difference information of the at least one past frame is an inter-channel time difference smoothed value of the at least one past frame or an inter-channel time difference of the at least one past frame.

37. The apparatus according to claim 36, wherein the inter-channel time difference information of the at least one past frame is the inter-channel time difference smoothed

Date Reçue/Date Received 2021-06-10

value of the at least one past frame, and the update unit is configured to:

determine an inter-channel time difference smoothed value of the current frame based on the delay track estimation value of the current frame and the inter-channel time difference of the current frame; and

update a buffered inter-channel time difference smoothed value of the at least one past frame based on the inter-channel time difference smoothed value of the current frame; wherein

the inter-channel time difference smoothed value of the current frame is obtained by using the following calculation formula:

$$cur\_itd\_smooth = \varphi * reg\_prv\_corr + (1 - \varphi) * cur\_itd,$$ wherein

cur_itd_smooth is the inter-channel time difference smoothed value of the current frame, φ is a second smoothing factor and is a constant greater than or equal to 0 and less than or equal to 1, reg_prv_corr is the delay track estimation value of the current frame, and cur_itd is the inter-channel time difference of the current frame.

38. The apparatus according to any one of claims 35 to 37, wherein the update unit is further configured to:

update a buffered weighting coefficient of the at least one past frame, wherein the weighting coefficient of the at least one past frame is a weighting coefficient in the weighted linear regression method.

39. The apparatus according to claim 38, wherein when the adaptive window function of the current frame is determined based on a smoothed inter-channel time difference of the previous frame of the current frame, the update unit is configured to:

calculate a first weighting coefficient of the current frame based on the smoothed inter-channel time difference estimation deviation of the current frame; and

update a buffered first weighting coefficient of the at least one past frame based on the first weighting coefficient of the current frame, wherein

the first weighting coefficient of the current frame is obtained through calculation by using the following calculation formulas:

$$wgt\_par1 = a\_wgt1 * smooth\_dist\_reg\_update + b\_wgt1,$$

$$a\_wgt1 = (xl\_wgt1 - xh\_wgt1)/(yh\_dist1' - yl\_dist1'),$$ and

Date Reçue/Date Received 2021-06-10

b_wgt1 = xl_wgt1 – a_wgt1 * yh_dist1', wherein

wgt_par1 is the first weighting coefficient of the current frame, smooth_dist_reg_update is the smoothed inter-channel time difference estimation deviation of the current frame, xh_wgt is an upper limit value of the first weighting coefficient, xl_wgt is a lower limit value of the first weighting coefficient, yh_dist1' is a smoothed inter-channel time difference estimation deviation corresponding to the upper limit value of the first weighting coefficient, yl_dist1' is a smoothed inter-channel time difference estimation deviation corresponding to the lower limit value of the first weighting coefficient, and yh_dist1', yl_dist1', xh_wgt1, and xl_wgt1 are all positive numbers.

40. The apparatus according to claim 39, wherein

wgt_par1 = min(wgt_par1, xh_wgt1), and

wgt_par1 = max(wgt_par1, xl_wgt1), wherein

min represents taking of a minimum value, and max represents taking of a maximum value.

41. An audio signal coding device, wherein the audio coding device comprises a processor, and a memory connected to the processor; and

the memory is configured to be controlled by the processor, and the processor is configured to implement the delay estimation method according to any one of claims 1 to 23.

Date Reçue/Date Received 2021-06-10

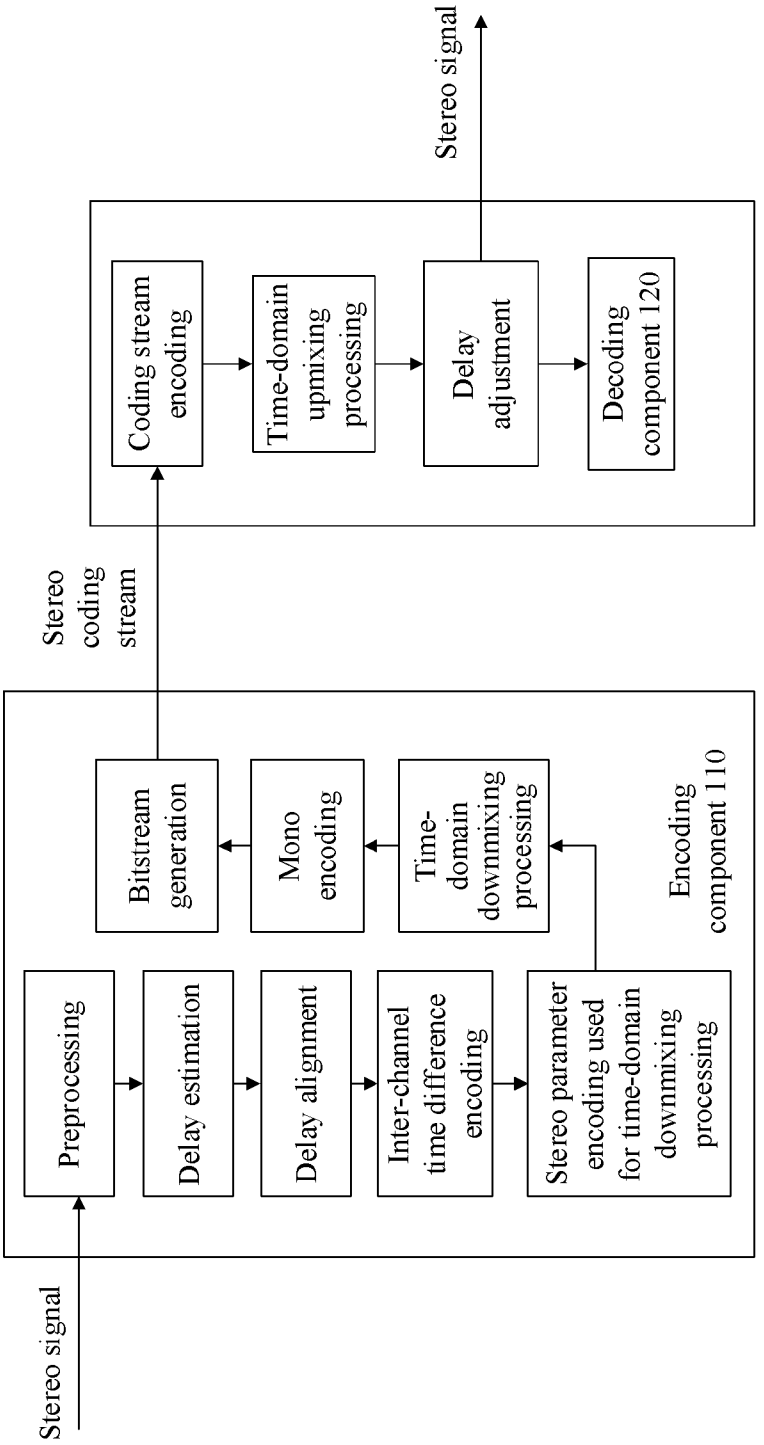


FIG. 1

Date Reçue/Date Received 2020-06-23

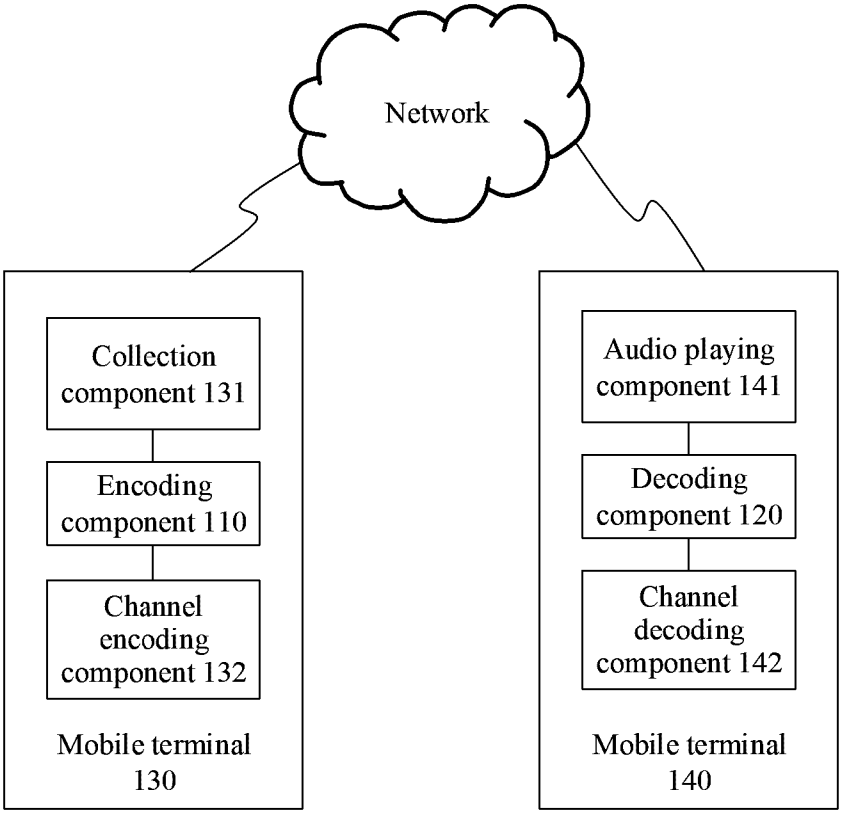


FIG. 2

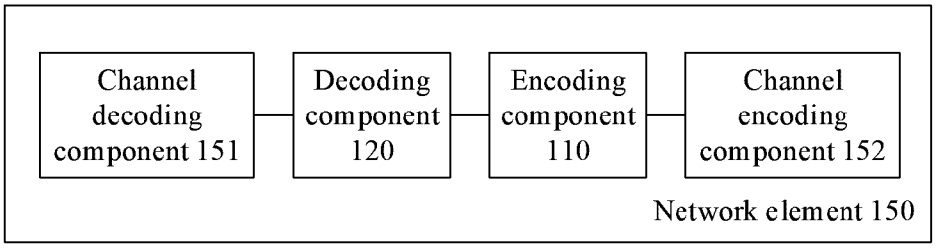


FIG. 3

Date Reçue/Date Received 2020-06-23

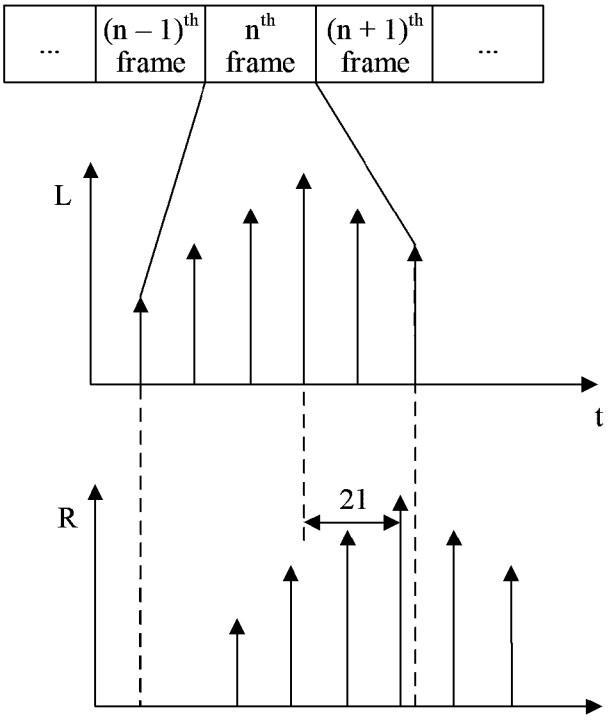


FIG. 4

Date Reçue/Date Received 2020-06-23

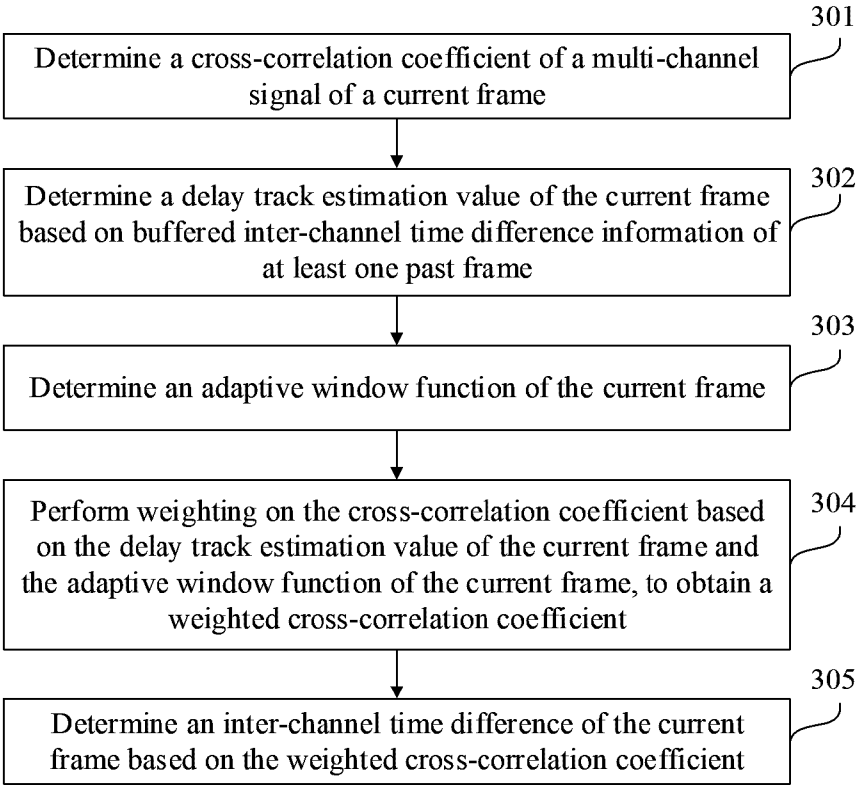


FIG. 5

Date Reçue/Date Received 2020-06-23

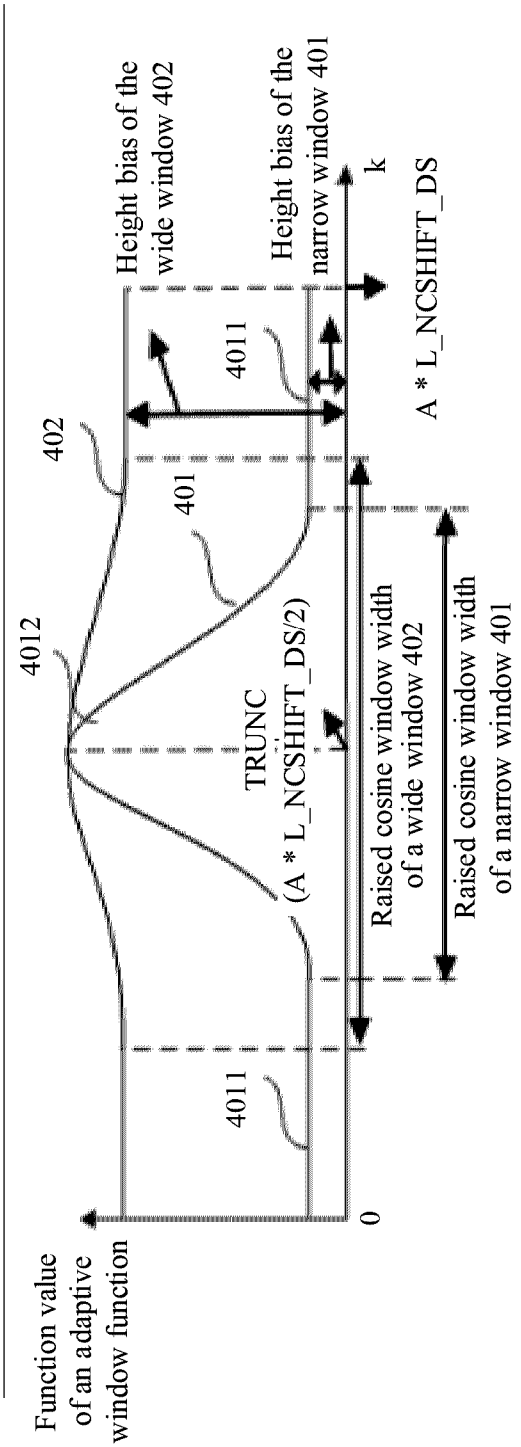


FIG. 6

Date Reçue/Date Received 2020-06-23

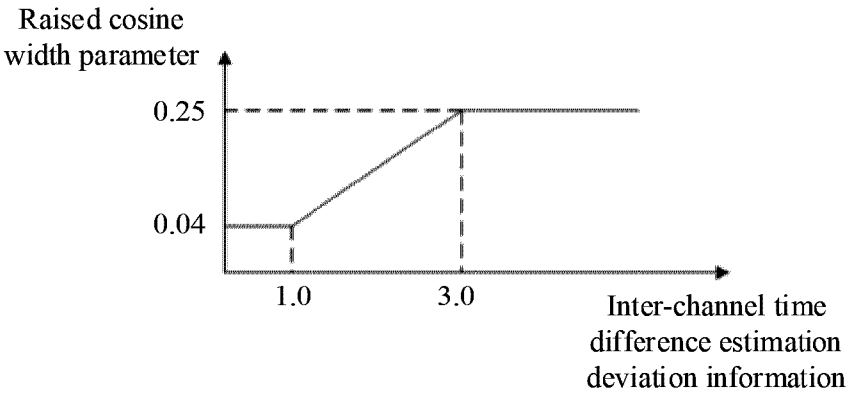


FIG. 7

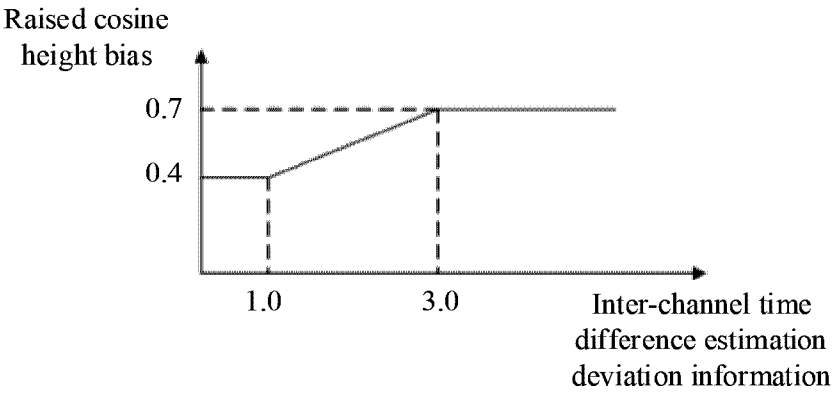


FIG. 8

0	1	2	3	4	5	6	7

FIG. 9

Date Reçue/Date Received 2020-06-23

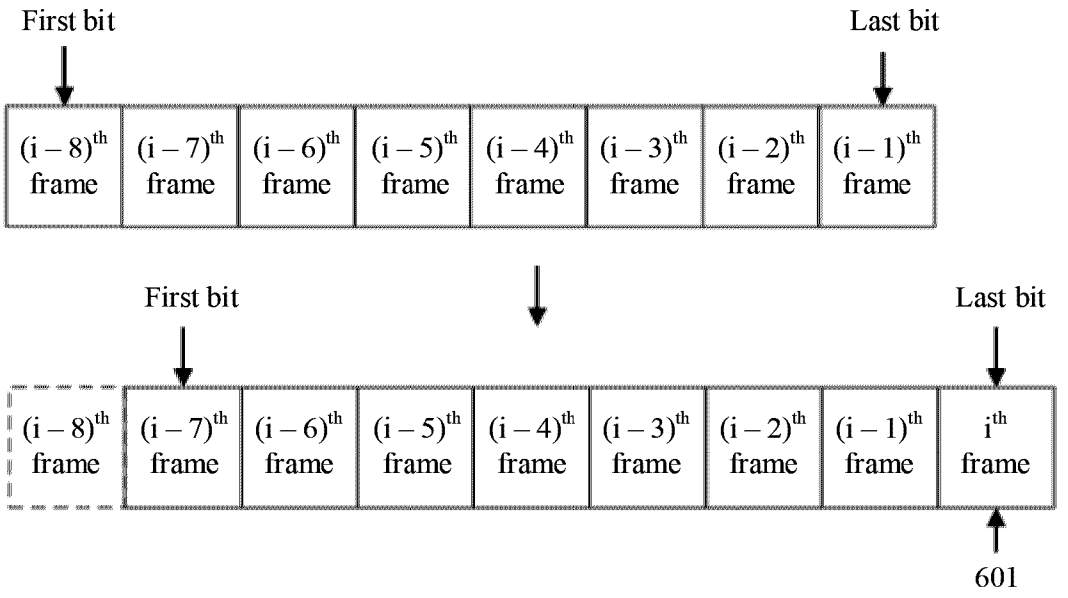


FIG. 10

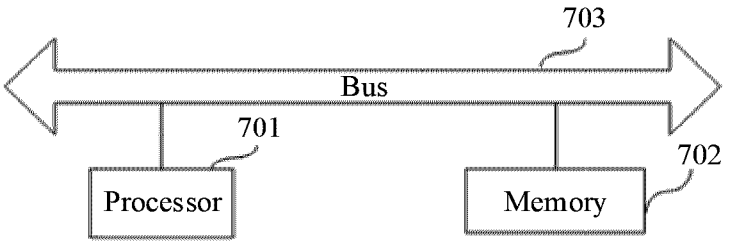


FIG. 11

Date Reçue/Date Received 2020-06-23

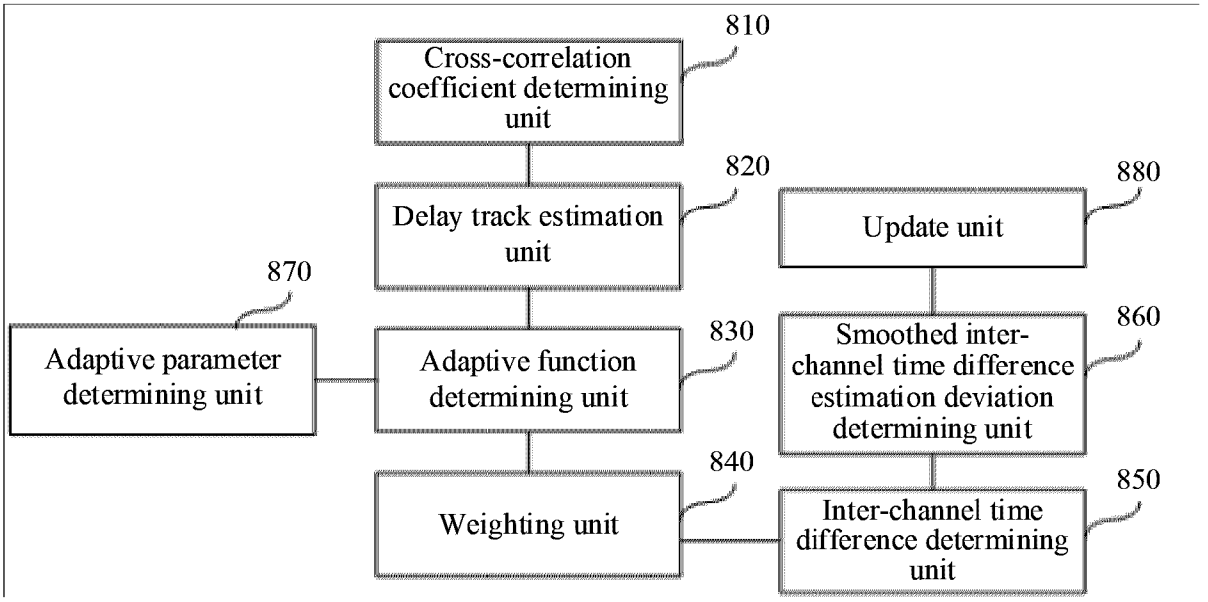
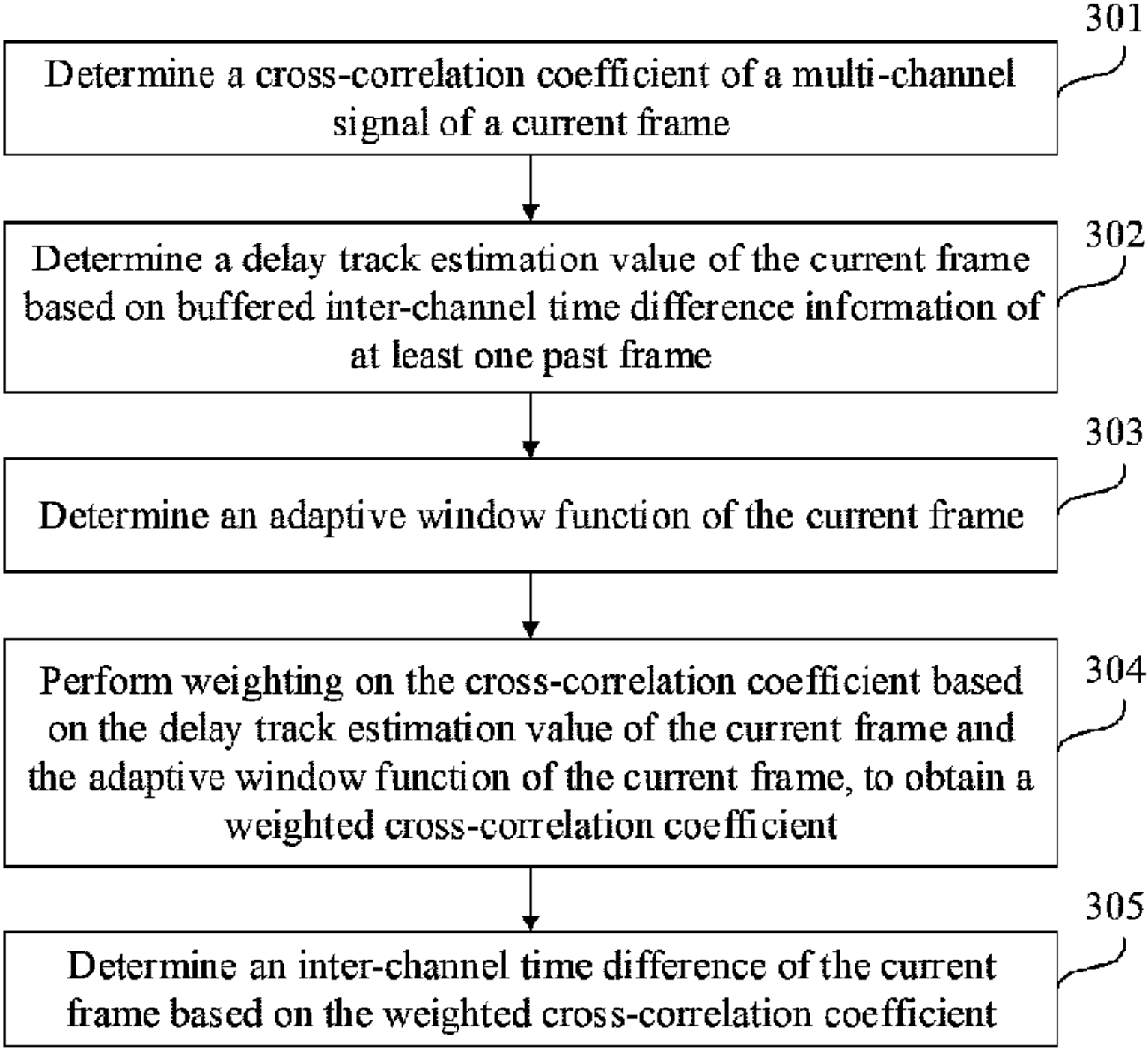


FIG. 12

Date Reçue/Date Received 2020-06-23


301
Determine a cross-correlation coefficient of a multi-channel signal of a current frame
302
Determine a delay track estimation value of the current frame based on buffered inter-channel time difference information of at least one past frame
303
Determine an adaptive window function of the current frame
304
Perform weighting on the cross-correlation coefficient based on the delay track estimation value of the current frame and the adaptive window function of the current frame, to obtain a weighted cross-correlation coefficient
305
Determine an inter-channel time difference of the current frame based on the weighted cross-correlation coefficient