



(19)中華民國智慧財產局

(12)發明說明書公告本

(11)證書號數：TW I844036 B

(45)公告日：中華民國 113 (2024) 年 06 月 01 日

(21)申請案號：111121698

(22)申請日：中華民國 111 (2022) 年 06 月 10 日

(51)Int. Cl. : **G10L19/008 (2013.01)****G10L19/20 (2013.01)****H04S3/00 (2006.01)**

(30)優先權：2021/06/18

中國大陸

202110680341.8

(71)申請人：大陸商華為技術有限公司(中國大陸) HUAWEI TECHNOLOGIES CO., LTD. (CN)  
中國大陸(72)發明人：高原 GAO, YUAN (CN)；劉帥 LIU, SHUAI (CN)；夏丙寅 XIA, BINGYIN (CN)；  
王賓 WANG, BIN (CN)；王喆 WANG, ZHE (CN)

(74)代理人：陳寧樺

(56)參考文獻：

CN 105940447B

CN 109804645A

US 2014/0358564A1

審查人員：蔡季霖

申請專利範圍項數：23 項 圖式數：13 共 81 頁

(54)名稱

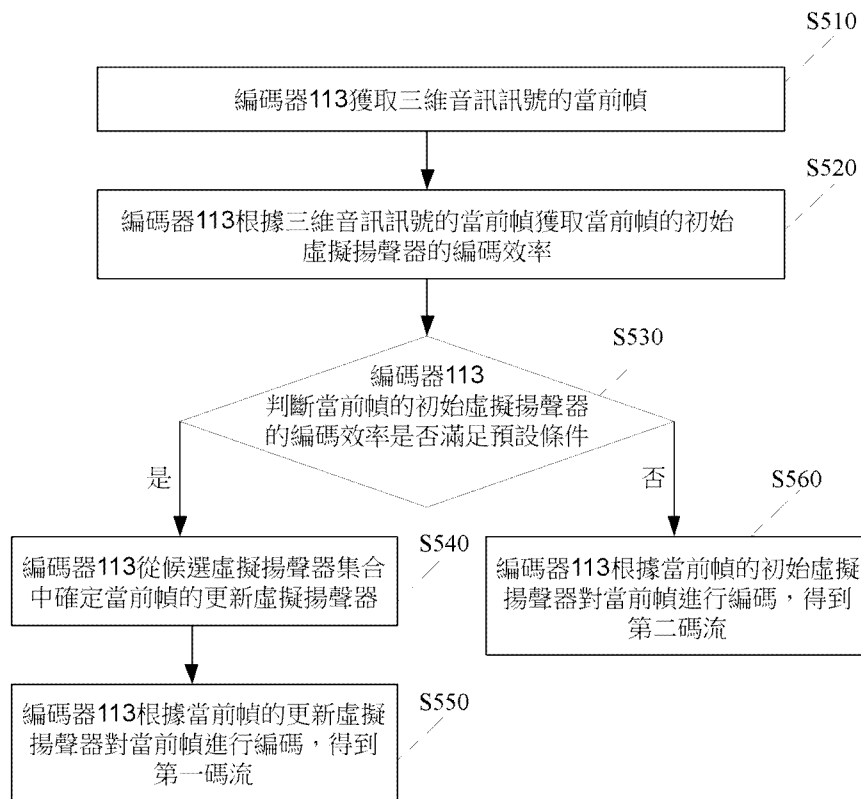
三維音訊訊號編碼方法、裝置、編碼器、系統、電腦程式和電腦可讀儲存介質

(57)摘要

本申請公開了一種三維音訊訊號編碼方法、裝置、編碼器、系統、電腦程式和電腦可讀儲存介質，涉及多媒體領域。該方法包括：編碼器獲取到三維音訊訊號的當前幀後，獲取當前幀的初始虛擬揚聲器的編碼效率，若當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器，以及根據當前幀的更新虛擬揚聲器對當前幀進行編碼，得到第一碼流，從而通過重選虛擬揚聲器，降低三維音訊訊號的不同幀之間進行編碼所使用的虛擬揚聲器的波動性，提高解碼端重建後三維音訊訊號的品質，以及解碼端播放的聲音的音質。若當前幀的初始虛擬揚聲器的編碼效率不滿足預設條件，根據當前幀的初始虛擬揚聲器對當前幀進行編碼，得到第二碼流。

The present application discloses a three-dimensional audio signal encoding method, apparatus, encoder, system, computer program and computer readable storage medium, which relates to the multimedia field. The method includes: after a current frame of a three-dimensional audio signal is obtained, obtaining, by an encoder, coding efficiency of an initial virtual speaker of the current frame; determining an updated virtual speaker of the current frame from a candidate virtual speaker set if the coding efficiency of the initial virtual speaker of the current frame meets a preset condition; and encoding the current frame according to the updated virtual speaker of the current frame to obtain a first bitstream, thereby reducing fluctuation of the virtual speaker used for encoding between different frames of the three-dimensional audio signal by reselecting the virtual speaker, improving quality of a reconstructed three-dimensional audio signal at the decoding end and the sound quality of the sound played at the decoding end; encoding the current frame according to the initial virtual speaker of the current frame to obtain the second bitstream if the encoding efficiency of the initial virtual speaker of the current frame does not meet the preset condition.

指定代表圖：



符號簡單說明：

S510、S520、S530、  
S540、S550、S560:步  
驟

【圖5】



I844036

## 【發明摘要】

【中文發明名稱】三維音訊訊號編碼方法、裝置、編碼器、系統、電腦程式和電腦可讀儲存介質

【英文發明名稱】THREE-DIMENSIONAL AUDIO SIGNAL ENCODING METHOD, APPARATUS, ENCODER, SYSTEM, COMPUTER PROGRAM AND COMPUTER READABLE STORAGE MEDIUM

【中文】本申請公開了一種三維音訊訊號編碼方法、裝置、編碼器、系統、電腦程式和電腦可讀儲存介質，涉及多媒體領域。該方法包括：編碼器獲取到三維音訊訊號的當前幀後，獲取當前幀的初始虛擬揚聲器的編碼效率，若當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器，以及根據當前幀的更新虛擬揚聲器對當前幀進行編碼，得到第一碼流，從而通過重選虛擬揚聲器，降低三維音訊訊號的不同幀之間進行編碼所使用的虛擬揚聲器的波動性，提高解碼端重建後三維音訊訊號的品質，以及解碼端播放的聲音的音質。若當前幀的初始虛擬揚聲器的編碼效率不滿足預設條件，根據當前幀的初始虛擬揚聲器對當前幀進行編碼，得到第二碼流。

【英文】The present application discloses a three-dimensional audio signal encoding method, apparatus, encoder, system, computer program and computer readable storage medium, which relates to the

multimedia field. The method includes: after a current frame of a three-dimensional audio signal is obtained, obtaining, by an encoder, coding efficiency of an initial virtual speaker of the current frame; determining an updated virtual speaker of the current frame from a candidate virtual speaker set if the coding efficiency of the initial virtual speaker of the current frame meets a preset condition; and encoding the current frame according to the updated virtual speaker of the current frame to obtain a first bitstream, thereby reducing fluctuation of the virtual speaker used for encoding between different frames of the three-dimensional audio signal by reselecting the virtual speaker, improving quality of a reconstructed three-dimensional audio signal at the decoding end and the sound quality of the sound played at the decoding end; encoding the current frame according to the initial virtual speaker of the current frame to obtain the second bitstream if the encoding efficiency of the initial virtual speaker of the current frame does not meet the preset condition.

【指定代表圖】圖5。

【代表圖之符號簡單說明】

S510、S520、S530、S540、S550、S560:步驟

【特徵化學式】

無

## 【發明說明書】

【中文發明名稱】三維音訊訊號編碼方法、裝置、編碼器、系統、電腦程式和電腦可讀儲存介質

【英文發明名稱】THREE-DIMENSIONAL AUDIO SIGNAL ENCODING METHOD, APPARATUS, ENCODER, SYSTEM, COMPUTER PROGRAM AND COMPUTER READABLE STORAGE MEDIUM

### 【技術領域】

【0001】本申請涉及多媒體領域，尤其涉及一種三維音訊訊號編碼方法、裝置、編碼器、系統、電腦程式和電腦可讀儲存介質。

### 【先前技術】

【0002】隨著高性能電腦和訊號處理技術的飛速發展，收聽者對語音、音訊體驗提出了越來越高的要求，沉浸式音訊能夠滿足人們在這方面的需求。例如，三維音訊技術在無線通訊（例如 4G/5G 等等）語音、虛擬實境/增強現實和媒體音訊等方面得到了廣泛應用。三維音訊技術是對真實世界中的聲音和三維聲場資訊進行獲取、處理、傳輸和渲染重播的音訊技術，使聲音具有強烈的空間感、包圍感及沉浸感，給收聽者以“身臨其境”的非凡聽覺體驗。

【0003】通常，採集設備（如：麥克風）採集大量的資料記錄三維聲場資訊，向重播設備（例如揚聲器，耳機等）傳輸三維音訊

訊號，以便於重播設備播放三維音訊。由於三維聲場資訊的資料量較大，導致需要大量的儲存空間儲存資料，以及傳輸三維音訊訊號的頻寬需求較高。為了解決上述問題，可以對三維音訊訊號進行壓縮，儲存或傳輸壓縮資料。目前，編碼器利用虛擬揚聲器對三維音訊訊號進行壓縮。但是，若編碼器對三維音訊訊號的不同幀進行編碼所使用的虛擬揚聲器波動性較大，導致重建後三維音訊訊號的品質較低，音質較差。因此，如何提高重建後三維音訊訊號的品質是一個亟待解決的問題。

#### 【發明內容】

【0004】 本申請提供了三維音訊訊號編碼方法、裝置、編碼器、系統、電腦程式和電腦可讀儲存介質，由此可以提高重建後三維音訊訊號的品質。

【0005】 第一方面，本申請提供了一種三維音訊訊號編碼方法，該方法由編碼器執行，具體包括如下步驟：編碼器獲取到三維音訊訊號的當前幀後，根據三維音訊訊號的當前幀獲取當前幀的初始虛擬揚聲器的編碼效率，編碼效率表示當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力。若當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，表示當前幀的初始虛擬揚聲器不能充分表達三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較弱，則編碼器從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器，以及根據當前

幀的更新虛擬揚聲器對當前幀進行編碼，得到第一碼流。若當前幀的初始虛擬揚聲器的編碼效率不滿足預設條件，表示當前幀的初始虛擬揚聲器充分表達了三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較強，則編碼器根據當前幀的初始虛擬揚聲器對當前幀進行編碼，得到第二碼流。其中，當前幀的初始虛擬揚聲器和當前幀的更新虛擬揚聲器均屬於候選虛擬揚聲器集合。

**【0006】** 如此，編碼器獲取到當前幀的初始虛擬揚聲器後，確定初始虛擬揚聲器的編碼效率，依據編碼效率表示的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力，確定是否重新選擇當前幀的虛擬揚聲器。在當前幀的初始虛擬揚聲器的編碼效率滿足預設條件時，也即是當前幀的初始虛擬揚聲器無法充分表示重建三維音訊訊號所屬聲場的場景下，重新選擇當前幀的虛擬揚聲器，將當前幀的更新虛擬揚聲器作為對當前幀進行編碼的虛擬揚聲器。從而，通過重選虛擬揚聲器，降低三維音訊訊號的不同幀之間進行編碼所使用的虛擬揚聲器的波動性，提高解碼端重建後三維音訊訊號的品質，以及解碼端播放的聲音的音質。

**【0007】** 具體地，編碼器可以根據以下四種方式中任一種獲取當前幀的初始虛擬揚聲器的編碼效率。

**【0008】** 方式一，編碼器根據三維音訊訊號的當前幀獲取當前幀的初始虛擬揚聲器的編碼效率包括：編碼器根據當前幀的初始虛擬揚聲器獲取重建後三維音訊訊號的重建當前幀後，根據重建當

前幀的能量與當前幀的能量確定當前幀的初始虛擬揚聲器的編碼效率。由於重建後三維音訊訊號的重建當前幀是由表達三維音訊訊號的聲場資訊的當前幀的初始虛擬揚聲器確定的，則編碼器依據重建當前幀的能量佔據當前幀的能量的比例關係能夠直觀準確地確定初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力，從而確保編碼器確定當前幀的初始虛擬揚聲器的編碼效率的準確度。例如，若重建當前幀的能量小於當前幀的能量的一半，表示當前幀的初始虛擬揚聲器不能充分表達三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較弱。

**【0009】** 方式二，編碼器根據三維音訊訊號的當前幀獲取當前幀的初始虛擬揚聲器的編碼效率包括：編碼器根據當前幀的初始虛擬揚聲器確定重建後三維音訊訊號的重建當前幀，以及根據當前幀和重建當前幀獲取當前幀的殘差訊號後，編碼器根據當前幀的虛擬揚聲器訊號的能量與當前幀的虛擬揚聲器訊號的能量和殘差訊號的能量之和的比值確定當前幀的初始虛擬揚聲器的編碼效率。需要說明的是，當前幀的虛擬揚聲器訊號的能量和殘差訊號的能量之和可以是編碼端待傳輸的訊號。從而，編碼器可以通過當前幀的虛擬揚聲器訊號的能量與待傳輸的訊號的能量的比值關係間接地確定初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力，避免編碼器確定重建當前幀，降低了編碼器確定當前幀的初始虛擬揚聲器的編碼效率的複雜度。例如，若當前幀的虛擬揚

聲器訊號的能量小於待傳輸的訊號的能量的一半，表示當前幀的初始虛擬揚聲器不能充分表達三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較弱。

【0010】 其中，編碼器根據當前幀的初始虛擬揚聲器獲取重建後三維音訊訊號的重建當前幀包括：根據當前幀的初始虛擬揚聲器確定當前幀的虛擬揚聲器訊號；根據當前幀的虛擬揚聲器訊號確定重建當前幀。示例地，重建當前幀的能量是根據重建當前幀的係數確定的，當前幀的能量是根據當前幀的係數確定的。

【0011】 方式三，編碼器根據三維音訊訊號的當前幀獲取當前幀的初始虛擬揚聲器的編碼效率包括：編碼器根據三維音訊訊號的當前幀確定聲源數量；根據當前幀的初始虛擬揚聲器的數量與聲源數量的比值確定當前幀的初始虛擬揚聲器的編碼效率。

【0012】 方式四，編碼器根據三維音訊訊號的當前幀獲取當前幀的初始虛擬揚聲器的編碼效率包括：編碼器根據三維音訊訊號的當前幀確定聲源數量，根據當前幀的初始虛擬揚聲器確定當前幀的虛擬揚聲器訊號，根據當前幀的虛擬揚聲器訊號的數量與聲源數量的比值確定當前幀的初始虛擬揚聲器的編碼效率。

【0013】 由於當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場，則當前幀的初始虛擬揚聲器可以表示三維音訊訊號所屬聲場的資訊，編碼器利用當前幀的初始虛擬揚聲器的數量與三維音訊訊號的聲源數量的關係確定當前幀的初始虛擬揚聲器的編碼效率，或者編碼器利用當前幀的虛擬揚聲器訊號的數量與三維音

訊訊號的聲源數量的關係確定當前幀的初始虛擬揚聲器的編碼效率，可以既確保編碼器確定當前幀的初始虛擬揚聲器的編碼效率的準確度，又降低了編碼器確定當前幀的初始虛擬揚聲器的編碼效率的複雜度。

**【0014】** 在編碼器根據上述方式一至方式四中任一方式確定當前幀的初始虛擬揚聲器的編碼效率小於第一閾值，即當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，編碼器可以根據下述可能的實現方式確定當前幀的更新虛擬揚聲器。可理解的，預設條件包括當前幀的初始虛擬揚聲器的編碼效率小於第一閾值。第一閾值的取值範圍可以是 0 至 1，或 0.5 至 1。例如，第一閾值可以是 0.35、0.65、0.75 或 0.85 等等。

**【0015】** 在一種可能的實現方式中，編碼器從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器包括：若當前幀的初始虛擬揚聲器的編碼效率小於第二閾值，將候選虛擬揚聲器集合中的預設虛擬揚聲器作為當前幀的更新虛擬揚聲器，第二閾值小於第一閾值。

**【0016】** 如此，在當前幀的初始虛擬揚聲器無法充分表示重建三維音訊訊號所屬聲場，導致解碼端重建後三維音訊訊號的品質較差的場景下，編碼器經過二次判斷當前幀的初始虛擬揚聲器的編碼效率，進一步提高了編碼器確定初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力的準確度。而且，編碼器通過定向選取當前幀的更新虛擬揚聲器，降低三維音訊訊號的不同幀之間進行

編碼所使用的虛擬揚聲器的波動性，提高解碼端重建後三維音訊訊號的品質，以及解碼端播放的聲音的音質。

【0017】 在另一種可能的實現方式中，編碼器從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器包括：若當前幀的初始虛擬揚聲器的編碼效率小於第一閾值，且大於第二閾值，將在先幀的虛擬揚聲器作為當前幀的更新虛擬揚聲器，在先幀的虛擬揚聲器為對三維音訊訊號的在先幀進行編碼所使用的虛擬揚聲器。由於編碼器將在先幀的虛擬揚聲器作為對當前幀進行編碼的虛擬揚聲器，從而降低了三維音訊訊號的不同幀之間進行編碼所使用的虛擬揚聲器的波動性，提高解碼端重建後三維音訊訊號的品質，以及解碼端播放的聲音的音質。

【0018】 可選地，該方法還包括：編碼器根據當前幀的初始虛擬揚聲器的編碼效率和在先幀的虛擬揚聲器的編碼效率確定當前幀的初始虛擬揚聲器的調整後編碼效率；若當前幀的初始虛擬揚聲器的編碼效率大於當前幀的初始虛擬揚聲器的調整後編碼效率，表明當前幀的初始虛擬揚聲器具有表示重建三維音訊訊號所屬聲場的能力，將當前幀的初始虛擬揚聲器作為當前幀的後續幀的虛擬揚聲器。從而，降低了三維音訊訊號的不同幀之間進行編碼所使用的虛擬揚聲器的波動性，提高解碼端重建後三維音訊訊號的品質，以及解碼端播放的聲音的音質。

【0019】 另外，三維音訊訊號可以為高階立體混響（higher order ambisonics，HOA）訊號。

【0020】 第二方面，本申請提供了一種三維音訊訊號編碼裝置，所述裝置包括用於執行第一方面或第一方面任一種可能設計中的三維音訊訊號編碼方法的各個模組。例如，三維音訊訊號編碼裝置包括通信模組、編碼效率獲取模組、虛擬揚聲器重選模組和編碼模組。所述通信模組，用於獲取三維音訊訊號的當前幀。所述編碼效率獲取模組，用於根據三維音訊訊號的當前幀獲取當前幀的初始虛擬揚聲器的編碼效率，當前幀的初始虛擬揚聲器屬於候選虛擬揚聲器集合。所述虛擬揚聲器重選模組，用於若當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器。所述編碼模組，用於根據當前幀的更新虛擬揚聲器對當前幀進行編碼，得到第一碼流。所述編碼模組，還用於若當前幀的初始虛擬揚聲器的編碼效率不滿足預設條件，根據當前幀的初始虛擬揚聲器對當前幀進行編碼，得到第二碼流。這些模組可以執行上述第一方面方法示例中的相應功能，具體參見方法示例中的詳細描述，此處不做贅述。

【0021】 第三方面，本申請提供一種編碼器，該編碼器包括至少一個處理器和記憶體，其中，所述記憶體用於儲存一組電腦指令；當處理器執行所述一組電腦指令時，執行第一方面或第一方面任一種可能實現方式中的三維音訊訊號編碼方法的操作步驟。

【0022】 第四方面，本申請提供一種系統，系統包括如第三方面所述的編碼器，以及解碼器，所述編碼器用於執行第一方面或第一方面任一種可能實現方式中的三維音訊訊號編碼方法的操作步

驟，所述解碼器用於解碼所述編碼器生成的碼流。

【0023】 第五方面，本申請提供一種電腦可讀儲存介質，包括：電腦軟體指令；當電腦軟體指令在編碼器中運行時，使得編碼器執行如第一方面或第一方面任意一種可能的實現方式中所述方法的操作步驟。

【0024】 第六方面，本申請提供一種電腦程式產品，當電腦程式產品在編碼器上運行時，使得編碼器執行如第一方面或第一方面任意一種可能的實現方式中所述方法的操作步驟。

【0025】 第七方面，本申請提供一種電腦可讀儲存介質，包括如第一方面或第一方面任意一種可能的實現方式中所述方法所獲得的碼流。

【0026】 本申請在上述各方面提供的實現方式的基礎上，還可以進行進一步組合以提供更多實現方式。

### 【圖式簡單說明】

#### 【0027】

圖 1 為本申請實施例提供的一種音訊編解碼系統的結構示意圖；

圖 2 為本申請實施例提供的一種音訊編解碼系統的場景示意圖；

圖 3 為本申請實施例提供的一種編碼器的結構示意圖；

圖 4 為本申請實施例提供的一種三維音訊訊號編解碼方法的

流程示意圖；

圖 5 為本申請實施例提供的一種三維音訊訊號編碼方法的流程示意圖；

圖 6 為本申請實施例提供的另一種編碼器的結構示意圖；

圖 7 為本申請實施例提供的另一種編碼器的結構示意圖；

圖 8 為本申請實施例提供的另一種編碼器的結構示意圖；

圖 9 為本申請實施例提供的另一種編碼器的結構示意圖；

圖 10 為本申請實施例提供的另一種三維音訊訊號編碼方法的流程示意圖；

圖 11 為本申請實施例提供的一種選擇虛擬揚聲器方法的流程示意圖；

圖 12 為本申請提供的一種三維音訊訊號編碼裝置的結構示意圖；

圖 13 為本申請提供的一種編碼器的結構示意圖。

### 【實施方式】

【0028】 為了下述各實施例的描述清楚簡潔，首先給出相關技術的簡要介紹。

【0029】 聲音 ( sound)是由物體振動產生的一種連續的波。產生振動而發出聲波的物體稱為聲源。聲波通過介質 (如：空氣、固體或液體) 傳播的過程中，人或動物的聽覺器官能感知到聲音。

【0030】 聲波的特徵包括音調、音強和音色。音調表示聲音的高

低。音強表示聲音的大小。音強也可以稱為響度或音量。音強的單位是分貝（decibel，dB）。音色又稱為音品。

【0031】聲波的頻率決定了音調的高低。頻率越高音調越高。物體在一秒鐘之內振動的次數稱為頻率，頻率單位是赫茲（hertz，Hz）。人耳能識別的聲音的頻率在 20 Hz~20000 Hz 之間。

【0032】聲波的幅度決定了音強的強弱。幅度越大音強越大。距離聲源越近，音強越大。

【0033】聲波的波形決定了音色。聲波的波形包括方波、鋸齒波、正弦波和脈衝波等。

【0034】根據聲波的特徵，聲音可以分為規則聲音和無規則聲音。無規則聲音是指聲源無規則地振動發出的聲音。無規則聲音例如是影響人們工作、學習和休息等的雜訊。規則聲音是指聲源規則地振動發出的聲音。規則聲音包括語音和樂音。聲音用電表示時，規則聲音是一種在時頻域上連續變化的類比訊號。該類比訊號可以稱為音訊訊號。音訊訊號是一種攜帶語音、音樂和音效的資訊載體。

【0035】由於人的聽覺具有辨別空間中聲源的位置分佈的能力，則聽音者聽到空間中的聲音時，除了能感受到聲音的音調、音強和音色外，還能感受到聲音的方位。

【0036】隨著人們對聽覺系統體驗的關注和品質要求與日俱增，為了增強聲音的縱深感、臨場感和空間感，則三維音訊技術應運而生。從而聽音者不僅感受到來自前、後、左和右的聲源發出的

聲音，而且感受到自己所處空間被這些聲源產生的空間聲場（簡稱“聲場”（sound field））所包圍的感覺，以及聲音向四周擴散的感覺，營造出一種使聽音者置身於影院或音樂廳等場所的“身臨其境”的音響效果。

【0037】三維音訊技術是指將人耳以外的空間假設為一個系統，耳膜處接收到的訊號為聲源發出的聲音經過耳朵以外系統濾波輸出的三維音訊訊號。例如，人耳以外的系統可以定義為系統衝擊回應  $h(n)$ ，任意一個聲源可以定義為  $x(n)$ ，耳膜處接收到的訊號為  $x(n)$  和  $h(n)$  的卷積結果。本申請實施例所述的三維音訊訊號可以是指高階立體混響（higher order ambisonics, HOA）訊號。三維音訊也可以稱為三維音效、空間音訊、三維聲場重建、虛擬 3D 音訊或雙耳音訊等。

【0038】眾所周知，聲波在理想介質中傳播，波數為  $k = w/c$ ，角頻率為  $w = 2\pi f$ ，其中， $f$  為聲波頻率， $c$  為聲速。聲壓  $p$  滿足公式(1)， $\nabla^2$  為拉普拉斯運算元。

【0039】  $\nabla^2 p + k^2 p = 0$       公式(1)

【0040】假設人耳以外的空間系統是一個球形，聽音者處於球的中心，從球外傳來的聲音在球面上有一個投影，過濾掉球面以外的聲音，假設聲源分佈在這個球面上，用球面上的聲源產生的聲場來擬合原始聲源產生的聲場，即三維音訊技術就是一個擬合聲場的方法。具體地，在球坐標系下求解公式(1)等式方程，在無源球形區域內，該公式(1)方程解為如下公式(2)。

$$\text{【0041】 } p(r, \theta, \varphi, k) = s \sum_{m=0}^{\infty} (2m+1) j^m j_m^{kr}(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} Y_{m,n}^{\sigma}(\theta_s, \varphi_s) Y_{m,n}^{\sigma}(\theta, \varphi)$$

公式(2)

【0042】 其中， $r$  表示球半徑， $\theta$  表示水平角， $\varphi$  表示俯仰角， $k$  表示波數， $s$  表示理想平面波的幅度， $m$  表示三維音訊訊號的階數序號（或稱為 HOA 訊號的階數序號）。 $j^m j_m^{kr}(kr)$  表示球貝塞爾函數，球貝塞爾函數又稱為徑向基函數，其中，第一個  $j$  表示虛數單位， $(2m+1)j^m j_m^{kr}(kr)$  不隨角度變化。 $Y_{m,n}^{\sigma}(\theta, \varphi)$  表示  $\theta, \varphi$  方向的球諧函數， $Y_{m,n}^{\sigma}(\theta_s, \varphi_s)$  表示聲源方向的球諧函數。三維音訊訊號係數滿足公式(3)。

$$\text{【0043】 } B_{m,n}^{\sigma} = s \cdot Y_{m,n}^{\sigma}(\theta_s, \varphi_s) \quad \text{公式(3)}$$

【0044】 將公式(3)代入公式(2)，公式(2)可以變形為公式(4)。

$$\text{【0045】 } p(r, \theta, \varphi, k) = \sum_{m=0}^{\infty} j^m j_m^{kr}(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{m,n}^{\sigma} Y_{m,n}^{\sigma}(\theta, \varphi) \quad \text{公式(4)}$$

【0046】 其中， $B_{m,n}^{\sigma}$  表示  $N$  階的三維音訊訊號係數，用於近似描述聲場。聲場是指介質中有聲波存在的區域。 $N$  為大於或等於 1 的整數。比如， $N$  的取值範圍為 2 至 6 的整數。本申請的實施例所述的三維音訊訊號的係數可以是指 HOA 係數或環境身歷聲 (ambisonic) 係數。

【0047】 三維音訊訊號是一種攜帶聲場中聲源的空間位置資訊的資訊載體，描述了空間中聽音者的聲場。公式(4)表明聲場可以在球面上按球諧函數展開，即聲場可以分解為多個平面波的疊加。因此，可以將三維音訊訊號描述的聲場使用多個平面波的疊加來表達，並通過三維音訊訊號係數重建聲場。

【0048】 相對 5.1 聲道的音訊訊號或 7.1 聲道的音訊訊號，由於  $N$  階的 HOA 訊號有  $(N+1)^2$  個聲道，則 HOA 訊號包括用於描述聲場的空間資訊的資料量較多。若採集設備（比如：麥克風）將該三維音訊訊號傳輸到重播設備（比如：揚聲器），需要消耗較大的頻寬。目前，編碼器可以利用空間壓縮環繞音訊編碼（spatial squeezed surround audio coding, S3AC）或定向音訊編碼（directional audio coding, DirAC）對三維音訊訊號進行壓縮編碼得到碼流，向重播設備傳輸碼流。重播設備對碼流進行解碼，並重建三維音訊訊號，播放重建後三維音訊訊號。從而降低向重播設備傳輸三維音訊訊號的資料量，以及頻寬的佔用。但是，編碼器對三維音訊訊號進行壓縮編碼的計算複雜度較高，佔用編碼器過多的計算資源。因此，如何降低對三維音訊訊號進行壓縮編碼的計算複雜度是一個亟待解決的問題。

【0049】 本申請實施例提供一種音訊編解碼技術，尤其是提供一種面向三維音訊訊號的三維音訊編解碼技術，具體提供一種採用較少的聲道表示三維音訊訊號的編解碼技術，以改進傳統的音訊編解碼系統。音訊編碼（或通常稱為編碼）包括音訊編碼和音訊解碼兩部分。音訊編碼在源側執行，通常包括處理（例如，壓縮）原始音訊以減少表示該原始音訊所需的資料量，從而更高效地儲存和/或傳輸。音訊解碼在目的側執行，通常包括相對於編碼器作逆處理，以重建原始音訊。編碼部分和解碼部分也合稱為編解碼。下面將結合附圖對本申請實施例的實施方式進行詳細描述。

【0050】 圖 1 為本申請實施例提供的一種音訊編解碼系統的結構示意圖。音訊編解碼系統 100 包括源設備 110 和目的設備 120。源設備 110 用於對三維音訊訊號進行壓縮編碼得到碼流，向目的設備 120 傳輸碼流。目的設備 120 對碼流進行解碼，並重建三維音訊訊號，播放重建後三維音訊訊號。

【0051】 具體地，源設備 110 包括音訊獲取器 111、預處理器 112、編碼器 113 和通信介面 114。

【0052】 音訊獲取器 111 用於獲取原始音訊。音訊獲取器 111 可以是任意類型的用於捕獲現實世界聲音的音訊採集設備，和/或任意類型的音訊生成設備。音訊獲取器 111 例如是用於生成電腦音訊的電腦音訊處理器。音訊獲取器 111 也可以為儲存音訊的任意類型的記憶體或記憶體。音訊包括現實世界聲音、虛擬場景（如：虛擬實境（virtual reality，VR）或增強現實（augmented reality，AR））聲音和/或其任意組合。

【0053】 預處理器 112 用於接收音訊獲取器 111 採集的原始音訊，並對原始音訊進行預處理，得到三維音訊訊號。例如，預處理器 112 執行的預處理包括聲道轉換、音訊格式轉換或去雜訊等。

【0054】 編碼器 113 用於接收預處理器 112 生成的三維音訊訊號，對三維音訊訊號進行壓縮編碼得到碼流。示例地，編碼器 113 可以包括空間編碼器 1131 和核心編碼器 1132。空間編碼器 1131 用於根據三維音訊訊號從候選虛擬揚聲器集合選取（或稱為搜索）虛擬揚聲器，根據三維音訊訊號和虛擬揚聲器生成虛擬揚聲器訊

號。虛擬揚聲器訊號也可以稱為重播訊號。核心編碼器 1132 用於對虛擬揚聲器訊號進行編碼，得到碼流。

【0055】 通信介面 114 用於接收編碼器 113 生成的碼流，通過通信通道 130 向目的設備 120 發送碼流，以便於目的設備 120 根據碼流重建三維音訊訊號。

【0056】 目的設備 120 包括播放器 121、後處理器 122、解碼器 123 和通信介面 124。

【0057】 通信介面 124 用於接收通信介面 114 發送的碼流，並將碼流傳輸給解碼器 123。以便於解碼器 123 根據碼流重建三維音訊訊號。

【0058】 通信介面 114 和通信介面 124 可用於通過源設備 110 與目的設備 120 之間的直連通信鏈路，例如直接有線或無線連接等，或者通過任意類型的網路，例如有線網路、無線網路或其任意組合、任意類型的私網和公網或其任意類型的組合，發送或接收原始音訊的相關資料。

【0059】 通信介面 114 和通信介面 124 均可配置為如圖 1 中從源設備 110 指向目的設備 120 的對應通信通道 130 的箭頭所指示的單向通信介面，或雙向通信介面，並且可用於發送和接收消息等，以建立連接，確認並交換與通信鏈路和/或例如編碼後的碼流傳輸等資料傳輸相關的任何其它資訊，等等。

【0060】 解碼器 123 用於對碼流進行解碼，並重建三維音訊訊號。示例地，解碼器 123 包括核心解碼器 1231 和空間解碼器 1232。核

心解碼器 1231 用於對碼流進行解碼，得到解碼後虛擬揚聲器訊號。空間解碼器 1232 用於根據候選虛擬揚聲器集合和解碼後虛擬揚聲器訊號重建三維音訊訊號，得到重建後三維音訊訊號。

【0061】 後處理器 122 用於接收解碼器 123 生成的重建後三維音訊訊號，對重建後三維音訊訊號進行後處理。例如，後處理器 122 執行的後處理包括音訊渲染、響度歸一化、用戶交互、音訊格式轉換或去雜訊等。

【0062】 播放器 121 用於根據重建後三維音訊訊號播放重建的聲音。

【0063】 需要說明的是，音訊獲取器 111 和編碼器 113 可以集成在一個物理設備上，也可以設置在不同的物理設備上，不予限定。示例地，如圖 1 所示的源設備 110 包括音訊獲取器 111 和編碼器 113，表示音訊獲取器 111 和編碼器 113 集成在一個物理設備上，則源設備 110 也可稱為採集設備。源設備 110 例如是無線接入網的媒體閘道、核心網的媒體閘道、轉碼設備、媒體資原始伺服器、AR 設備、VR 設備、麥克風或者其他採集音訊設備。若源設備 110 不包括音訊獲取器 111，表示音訊獲取器 111 和編碼器 113 是兩個不同的物理設備，源設備 110 可以從其他設備（如：採集音訊設備或儲存音訊設備）獲取原始音訊。

【0064】 另外，播放器 121 和解碼器 123 可以集成在一個物理設備上，也可以設置在不同的物理設備上，不予限定。示例地，如圖 1 所示的目的設備 120 包括播放器 121 和解碼器 123，表示播放

器 121 和解碼器 123 集成在一個物理設備上，則目的設備 120 也可稱為重播設備，目的設備 120 具有解碼和播放重建音訊的功能。目的設備 120 例如是揚聲器、耳機或其他播放音訊的設備。若目的設備 120 不包括播放器 121，表示播放器 121 和解碼器 123 是兩個不同的物理設備，目的設備 120 對碼流解碼重建三維音訊訊號後，將重建後三維音訊訊號傳輸給其他播放設備（如：揚聲器或耳機），由其他播放設備重播重建後三維音訊訊號。

**【0065】** 此外，圖 1 示出了源設備 110 和目的設備 120 可以集成在一個物理設備上，也可以設置在不同的物理設備上，不予限定。

**【0066】** 示例地，如圖 2 中的(a)所示，源設備 110 可以是錄音棚中的麥克風，目的設備 120 可以是揚聲器。源設備 110 可以採集各種樂器的原始音訊，將原始音訊廣播至編解碼設備，編解碼設備對原始音訊進行編解碼處理，得到重建後三維音訊訊號，由目的設備 120 重播重建後三維音訊訊號。又示例地，源設備 110 可以是終端設備中的麥克風，目的設備 120 可以是耳機。源設備 110 可以採集外界的聲音或終端設備合成的音訊。

**【0067】** 又示例地，如圖 2 中的(b)所示，源設備 110 和目的設備 120 集成在 VR 設備、AR 設備、混合現實（Mixed Reality，MR）設備或擴展現實（Extended Reality，ER）設備中，則 VR/AR/MR/ER 設備具備採集原始音訊、重播音訊和編解碼的功能。源設備 110 可以採集使用者發出的聲音和用戶所處的虛擬環境中虛擬物體發出的聲音。

【0068】 在這些實施例中，源設備 110 或其對應功能和目的設備 120 或其對應功能可以使用相同硬體和/或軟體或通過單獨的硬體和/或軟體或其任意組合來實現。根據描述，圖 1 所示的源設備 110 和/或目的設備 120 中的不同單元或功能的存在和劃分可能根據實際設備和應用而有所不同，這對技術人員來說是顯而易見的。

【0069】 上述音訊編解碼系統的結構只是示意性說明，在一些可能的實現方式中，音訊編解碼系統還可以包括其他設備，例如，音訊編解碼系統還可以包括端側設備或雲側設備。源設備 110 採集到原始音訊後，對原始音訊進行預處理，得到三維音訊訊號；並將三維音訊廣播至端側設備或雲側設備，由端側設備或雲側設備實現對三維音訊訊號進行編解碼的功能。

【0070】 本申請實施例提供的音訊訊號編解碼方法主要應用於編碼端。結合圖 3 對編碼器（如編碼器 300）的結構進行詳細說明。如圖 3 所示，編碼器 300 包括虛擬揚聲器配置單元 310、虛擬揚聲器集合生成單元 320、編碼分析單元 330、虛擬揚聲器選擇單元 340、虛擬揚聲器訊號生成單元 350 和編碼單元 360。

【0071】 虛擬揚聲器配置單元 310 用於根據編碼器配置資訊生成虛擬揚聲器配置參數，以便得到多個虛擬揚聲器。編碼器配置資訊包括但不限於：三維音訊訊號的階數（或通常稱為 HOA 階數），編碼位元速率，使用者自訂資訊，等。虛擬揚聲器配置參數包括但不限於：虛擬揚聲器的數量，虛擬揚聲器的階數，虛擬揚聲器的位置座標，等。虛擬揚聲器的數量例如是 2048、1669、1343、

1024、530、512、256、128 或 64 等。虛擬揚聲器的階數可以是 2 階至 6 階中任一個。虛擬揚聲器的位置座標包括水平角和俯仰角。

【0072】 虛擬揚聲器配置單元 310 輸出的虛擬揚聲器配置參數作為虛擬揚聲器集合生成單元 320 的輸入。

【0073】 虛擬揚聲器集合生成單元 320 用於根據虛擬揚聲器配置參數生成候選虛擬揚聲器集合，候選虛擬揚聲器集合包括多個虛擬揚聲器。具體地，虛擬揚聲器集合生成單元 320 根據虛擬揚聲器的數量確定了候選虛擬揚聲器集合包括的多個虛擬揚聲器，以及根據虛擬揚聲器的位置資訊（如：座標）和虛擬揚聲器的階數確定虛擬揚聲器的係數。示例地，虛擬揚聲器的座標確定方法包括但不限於：按等距規則產生多個虛擬揚聲器，或者根據聽覺感知原理生成非均勻分佈的多個虛擬揚聲器；然後，根據虛擬揚聲器的數量生成虛擬揚聲器的座標。

【0074】 根據上述三維音訊訊號的生成原理也可以生成虛擬揚聲器的係數。將公式(3)中的  $\theta_s$  和  $\varphi_s$  分別設置為虛擬揚聲器的位置座標， $B_{m,n}^\sigma$  表示 N 階的虛擬揚聲器的係數。虛擬揚聲器的係數也可以稱作 ambisonics 係數。

【0075】 編碼分析單元 330 用於對三維音訊訊號進行編碼分析，例如分析三維音訊訊號的聲場分佈特徵，即三維音訊訊號的聲源數量、聲源的方向性和聲源的彌散度等特徵。

【0076】 虛擬揚聲器集合生成單元 320 輸出的候選虛擬揚聲器集合包括的多個虛擬揚聲器的係數作為虛擬揚聲器選擇單元 340 的

輸入。

【0077】 編碼分析單元 330 輸出的三維音訊訊號的聲場分佈特徵作為虛擬揚聲器選擇單元 340 的輸入。

【0078】 虛擬揚聲器選擇單元 340 用於根據待編碼的三維音訊訊號、三維音訊訊號的聲場分佈特徵和多個虛擬揚聲器的係數確定與三維音訊訊號匹配的代表虛擬揚聲器。

【0079】 不限定的是，本申請實施例的編碼器 300 還可以不包括編碼分析單元 330，即編碼器 300 可以不對輸入訊號進行分析，虛擬揚聲器選擇單元 340 採用一種預設配置確定代表虛擬揚聲器。例如，虛擬揚聲器選擇單元 340 僅根據三維音訊訊號和多個虛擬揚聲器的係數確定與三維音訊訊號匹配的代表虛擬揚聲器。

【0080】 其中，編碼器 300 可以將從採集設備獲取的三維音訊訊號或採用人工音訊物件合成的三維音訊訊號作為編碼器 300 的輸入。另外，編碼器 300 輸入的三維音訊訊號可以是時域三維音訊訊號也可以是頻域三維音訊訊號，不予限定。

【0081】 虛擬揚聲器選擇單元 340 輸出的代表虛擬揚聲器的位置資訊和代表虛擬揚聲器的係數作為虛擬揚聲器訊號生成單元 350 和編碼單元 360 的輸入。

【0082】 虛擬揚聲器訊號生成單元 350 用於根據三維音訊訊號和代表虛擬揚聲器的屬性資訊生成虛擬揚聲器訊號。代表虛擬揚聲器的屬性資訊包括代表虛擬揚聲器的位置資訊、代表虛擬揚聲器的係數和三維音訊訊號的係數中至少一個。若屬性資訊為代表虛

擬揚聲器的位置資訊，根據代表虛擬揚聲器的位置資訊確定代表虛擬揚聲器的係數；若屬性資訊包括三維音訊訊號的係數，根據三維音訊訊號的係數獲取代表虛擬揚聲器的係數。具體地，虛擬揚聲器訊號生成單元 350 根據三維音訊訊號的係數和代表虛擬揚聲器的係數計算虛擬揚聲器訊號。

【0083】 示例地，假設矩陣  $A$  表示虛擬揚聲器的係數，矩陣  $X$  表示 HOA 訊號的係數。矩陣  $X$  為矩陣  $A$  的逆矩陣。採用最小二乘法求得理論的最優解  $w$ ， $w$  表示虛擬揚聲器訊號。虛擬揚聲器訊號滿足公式(5)。

【0084】  $w = A^{-1}X$       公式(5)

【0085】 其中， $A^{-1}$  表示矩陣  $A$  的逆矩陣。矩陣  $A$  的大小為  $(M \times C)$ ， $C$  表示代表虛擬揚聲器的數量， $M$  表示  $N$  階 HOA 訊號的聲道的數量， $a$  表示代表虛擬揚聲器的係數，矩陣  $X$  的大小為  $(M \times L)$ ， $L$  表示 HOA 訊號的係數的數量， $x$  表示 HOA 訊號的係數。代表虛擬揚聲器的係數可以是指代表虛擬揚聲器的 HOA 係數或代表虛擬揚

聲器的 ambisonics 係數。例如，

$$X = \begin{bmatrix} x_{11} & \cdot & \cdot & \cdot & x_{1L} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{M1} & \cdot & \cdot & \cdot & x_{ML} \end{bmatrix}。$$

$$A = \begin{bmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1C} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{M1} & \cdot & \cdot & \cdot & a_{MC} \end{bmatrix}，$$

【0086】 虛擬揚聲器訊號生成單元 350 輸出的虛擬揚聲器訊號作

為編碼單元 360 的輸入。

【0087】 可選地，為了提高解碼端重建三維音訊訊號的品質，編碼器 300 還可以預先估計重建後三維音訊訊號，利用預先估計的重建後三維音訊訊號生成殘差訊號，利用殘差訊號對虛擬揚聲器訊號進行補償，從而，提高編碼端的虛擬揚聲器訊號表示三維音訊訊號的聲源的聲場資訊的準確性。示例地，編碼器 300 還可以包括訊號重建單元 370 和殘差訊號生成單元 380。

【0088】 訊號重建單元 370 用於根據虛擬揚聲器選擇單元 340 輸出的代表虛擬揚聲器的位置資訊和代表虛擬揚聲器的係數，以及虛擬揚聲器訊號生成單元 350 輸出的虛擬揚聲器訊號預先估計重建後三維音訊訊號，得到重建後三維音訊訊號。訊號重建單元 370 輸出的重建後三維音訊訊號作為殘差訊號生成單元 380 的輸入。

【0089】 殘差訊號生成單元 380 用於根據重建後三維音訊訊號和待編碼的三維音訊訊號生成殘差訊號。殘差訊號可以表示由虛擬揚聲器訊號得到的重建後三維音訊訊號後與原始的三維音訊訊號相比的差值。殘差訊號生成單元 380 輸出的殘差訊號作為殘差訊號選擇單元 390 和訊號補償單元 3100 的輸入。

【0090】 編碼單元 360 可以對虛擬揚聲器訊號和殘差訊號進行編碼得到碼流。為了提高編碼器 300 的編碼效率，可以從殘差訊號中選取部分殘差訊號供編碼單元 360 進行編碼。可選地，編碼器 300 還可以包括殘差訊號選擇單元 390 和訊號補償單元 3100。

【0091】 殘差訊號選擇單元 390 用於根據虛擬揚聲器訊號和殘差

訊號確定待編碼的殘差訊號。示例地，殘差訊號包含 $(N+1)2$  個係數，殘差訊號選擇單元 390 可以從 $(N+1)2$  個係數中選取小於 $(N+1)2$  個的係數作為待編碼的殘差訊號。殘差訊號選擇單元 390 輸出的待編碼的殘差訊號作為編碼單元 360 和訊號補償單元 3100 的輸入。

【0092】 由於殘差訊號選擇單元 390 選擇小於  $N$  階 ambisonic 係數的係數個數作為待傳輸的殘差訊號，與  $N$  階 ambisonic 係數的殘差訊號相比會有資訊丟失，因此訊號補償單元 3100 對不傳輸的殘差訊號進行資訊補償。訊號補償單元 3100 用於根據待編碼的三維音訊訊號、殘差訊號和待編碼的殘差訊號確定補償資訊，補償資訊用於指示待編碼的殘差訊號和不傳輸的殘差訊號的相關資訊，例如補償資訊用於指示待編碼的殘差訊號和不傳輸的殘差訊號的差值，以便於解碼端提供解碼的準確率。

【0093】 編碼單元 360 用於對虛擬揚聲器訊號、待編碼的殘差訊號和補償資訊進行核心編碼處理，得到碼流。核心編碼處理包括但不限於：變換、量化、心理聲學模型、雜訊整形、頻寬擴展、下混、算數編碼和碼流產生等。

【0094】 值得注意的是，空間編碼器 1131 可以包括虛擬揚聲器配置單元 310、虛擬揚聲器集合生成單元 320、編碼分析單元 330、虛擬揚聲器選擇單元 340 和虛擬揚聲器訊號生成單元 350，即虛擬揚聲器配置單元 310、虛擬揚聲器集合生成單元 320、編碼分析單元 330、虛擬揚聲器選擇單元 340、虛擬揚聲器訊號生成單元 350、

訊號重建單元 370、殘差訊號生成單元 380、殘差訊號選擇單元 390 和訊號補償單元 3100 實現了空間編碼器 1131 的功能。核心編碼器 1132 可以包括編碼單元 360，即編碼單元 360 實現了核心編碼器 1132 的功能。

【0095】 圖 3 所示的編碼器可以生成一個虛擬揚聲器訊號，也可以生成多個虛擬揚聲器訊號。多個虛擬揚聲器訊號可以由圖 3 所示的編碼器多次執行得到，也可以由圖 3 所示的編碼器一次執行得到。

【0096】 接下來，結合附圖對三維音訊訊號的編解碼過程進行說明。圖 4 為本申請實施例提供的一種三維音訊訊號編解碼方法的流程示意圖。在這裡由圖 1 中源設備 110 和目的設備 120 執行三維音訊訊號編解碼過程為例進行說明。如圖 4 所示，該方法包括以下步驟。

【0097】 S410、源設備 110 獲取三維音訊訊號的當前幀。

【0098】 如上述實施例所述，若源設備 110 攜帶音訊獲取器 111，源設備 110 可以通過音訊獲取器 111 採集原始音訊。可選地，源設備 110 也可以接收其他設備採集的原始音訊；或者從源設備 110 中的記憶體或其他記憶體獲取原始音訊。原始音訊可以包括即時採集的現實世界聲音、設備儲存的音訊和由多個音訊合成的音訊中至少一種。本實施例對原始音訊的獲取方式以及原始音訊的類型不予限定。

【0099】 源設備 110 獲取到原始音訊後，根據三維音訊技術和原

始音訊生成三維音訊訊號，以便於目的設備 120 重播重建後三維音訊訊號，也即是目的設備 120 重播由重建後三維音訊訊號生成的聲音時，為聽音者提供“身臨其境”的音響效果。生成三維音訊訊號的具體方法可以參考上述實施例中預處理器 112 的闡述和現有技術的闡述。

**【0100】** 另外，音訊訊號是一個連續的類比訊號。在音訊訊號處理過程中，可以先對音訊訊號進行採樣，生成幀序列的數位訊號。幀可以包括多個採樣點。幀也可以指採樣得到的採樣點。幀也可以包括對幀劃分得到的子幀。幀也可以指對幀劃分得到的子幀。例如一幀長度為  $L$  個採樣點，劃分為  $N$  個子幀，那麼每個子幀對應  $L/N$  個採樣點。音訊編解碼通常是指處理包含多個採樣點的音訊幀序列。

**【0101】** 音訊幀可以包括當前幀或在先幀。本申請的各個實施例所述的當前幀或在先幀可以是指幀或是子幀。當前幀是指在當前時刻進行編解碼處理的幀。在先幀是指在當前時刻之前時刻已進行編解碼處理的幀。在先幀可以是當前時刻的前一時刻或者前多個時刻的幀。本申請的實施例中，三維音訊訊號的當前幀是指在當前時刻進行編解碼處理的一幀三維音訊訊號。在先幀是指在當前時刻之前時刻已進行編解碼處理的一幀三維音訊訊號。三維音訊訊號的當前幀可以是指三維音訊訊號的待編碼當前幀。三維音訊訊號的當前幀可以簡稱為當前幀。三維音訊訊號的在先幀可以簡稱為在先幀。

【0102】 S420、源設備 110 確定候選虛擬揚聲器集合。

【0103】 在一種情形下，源設備 110 的記憶體中預先配置有候選虛擬揚聲器集合。源設備 110 可以從記憶體中讀取候選虛擬揚聲器集合。候選虛擬揚聲器集合包括多個虛擬揚聲器。虛擬揚聲器表示空間聲場中虛擬存在的揚聲器。虛擬揚聲器用於根據三維音訊訊號計算虛擬揚聲器訊號，以便於目的設備 120 重播重建後三維音訊訊號，也即是以便於目的設備 120 重播由重建後三維音訊訊號生成的聲音。

【0104】 在另一種情形下，源設備 110 的記憶體中預先配置有虛擬揚聲器配置參數。源設備 110 根據虛擬揚聲器配置參數生成候選虛擬揚聲器集合。可選地，源設備 110 根據自身的計算資源（如：處理器）能力和當前幀的特徵（如：通道和資料量）即時生成候選虛擬揚聲器集合。

【0105】 生成候選虛擬揚聲器集合的具體方法可以參考現有技術，以及上述實施例中虛擬揚聲器配置單元 310 和虛擬揚聲器集合生成單元 320 的闡述。

【0106】 S430、源設備 110 根據三維音訊訊號的當前幀，從候選虛擬揚聲器集合中選取當前幀的代表虛擬揚聲器。

【0107】 源設備 110 可以根據匹配投影法（match-projection，MP）從候選虛擬揚聲器集合中選取當前幀的代表虛擬揚聲器。

【0108】 源設備 110 還可以根據當前幀的係數與虛擬揚聲器的係數對虛擬揚聲器進行投票，根據虛擬揚聲器的投票值從候選虛擬

揚聲器集合中選擇當前幀的代表虛擬揚聲器。從候選虛擬揚聲器集合中搜索有限數量的當前幀的代表虛擬揚聲器，作為待編碼的當前幀的最佳匹配虛擬揚聲器，從而實現對待編碼的三維音訊訊號進行資料壓縮的目的。

**【0109】** 需要說明的是，當前幀的代表虛擬揚聲器屬於候選虛擬揚聲器集合。當前幀的代表虛擬揚聲器的數量小於或等於候選虛擬揚聲器集合包含的虛擬揚聲器的數量。

**【0110】** S440、源設備 110 根據三維音訊訊號的當前幀和當前幀的代表虛擬揚聲器生成虛擬揚聲器訊號。

**【0111】** 源設備 110 根據當前幀的係數和當前幀的代表虛擬揚聲器的係數生成虛擬揚聲器訊號。生成虛擬揚聲器訊號的具體方法可以參考現有技術，以及上述實施例中虛擬揚聲器訊號生成單元 350 的闡述。

**【0112】** S450、源設備 110 根據當前幀的代表虛擬揚聲器和虛擬揚聲器訊號生成重建後三維音訊訊號。

**【0113】** 源設備 110 根據當前幀的代表虛擬揚聲器的係數和虛擬揚聲器訊號的係數生成重建後三維音訊訊號。生成重建後三維音訊訊號的具體方法可以參考現有技術，以及上述實施例中訊號重建單元 370 的闡述。

**【0114】** S460、源設備 110 根據三維音訊訊號的當前幀和重建後三維音訊訊號生成殘差訊號。

**【0115】** S470、源設備 110 根據三維音訊訊號的當前幀和殘差訊

號生成補償資訊。

【0116】 生成殘差訊號和補償資訊的具體方法可以參考現有技術，以及上述實施例中殘差訊號生成單元 380 和訊號補償單元 3100 的闡述。

【0117】 S480、源設備 110 對虛擬揚聲器訊號、殘差訊號和補償資訊進行編碼得到碼流。

【0118】 源設備 110 可以對虛擬揚聲器訊號、殘差訊號和補償資訊進行變換或量化等編碼操作，生成碼流，從而實現對待編碼的三維音訊訊號進行資料壓縮的目的。生成碼流的具體方法可以參考現有技術，以及上述實施例中編碼單元 360 的闡述。

【0119】 S490、源設備 110 向目的設備 120 發送碼流。

【0120】 源設備 110 可以對原始音訊全部編碼完成後，向目的設備 120 發送原始音訊的碼流。或者，源設備 110 也可以以幀為單位，即時對三維音訊訊號進行編碼處理，對一幀編碼完成後發送一幀的碼流。發送碼流的具體方法可以參考現有技術，以及上述實施例中通信介面 114 和通信介面 124 的闡述。

【0121】 S4100、目的設備 120 對源設備 110 發送的碼流進行解碼，重建三維音訊訊號，得到重建後三維音訊訊號。

【0122】 目的設備 120 接收到碼流後，對碼流進行解碼得到虛擬揚聲器訊號，再根據候選虛擬揚聲器集合和虛擬揚聲器訊號重建三維音訊訊號，得到重建後三維音訊訊號。目的設備 120 重播重建後三維音訊訊號，也即是目的設備 120 重播由重建後三維音訊

訊號生成的聲音。或者，目的設備 120 將重建後三維音訊訊號傳輸給其他播放設備，由其他播放設備播放重建後三維音訊訊號，也即是由其他播放設備播放由重建後三維音訊訊號生成的聲音，使得聽音者置身於影院、音樂廳或虛擬場景等場所的“身臨其境”的音響效果更加逼真。

【0123】目前，在虛擬揚聲器搜索過程中，編碼器依據待編碼的三維音訊訊號和虛擬揚聲器之間的相關計算的結果作為虛擬揚聲器的選擇衡量指標。若編碼器對每一個係數傳輸一個虛擬揚聲器，則無法達到資料壓縮的目的，且會對編碼器造成沉重的計算負擔。但是，若編碼器對三維音訊訊號的不同幀進行編碼所使用的虛擬揚聲器波動性較大，導致重建後三維音訊訊號的品質較低，解碼端播放的聲音的音質較差。因此，本申請實施例提供一種選擇虛擬揚聲器的方法，編碼器獲取到當前幀的初始虛擬揚聲器後，確定初始虛擬揚聲器的編碼效率，依據編碼效率表示的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力，確定是否重新選擇當前幀的虛擬揚聲器。在當前幀的初始虛擬揚聲器的編碼效率滿足預設條件時，也即是當前幀的初始虛擬揚聲器無法充分表示重建三維音訊訊號所屬聲場的場景下，重新選擇當前幀的虛擬揚聲器，將當前幀的更新虛擬揚聲器作為對當前幀進行編碼的虛擬揚聲器。從而，通過重選虛擬揚聲器，降低三維音訊訊號的不同幀之間進行編碼所使用的虛擬揚聲器的波動性，提高解碼端重建後三維音訊訊號的品質，以及解碼端播放的聲音的音質。

【0124】 在本申請實施例中，編碼效率也可以稱為重建聲場效率、重建三維音訊訊號效率或虛擬揚聲器選擇效率。

【0125】 接下來，結合附圖對選擇虛擬揚聲器的過程進行詳細說明。圖 5 為本申請實施例提供的一種三維音訊訊號編碼方法的流程示意圖。在這裡由圖 1 中源設備 110 中編碼器 113 執行選擇虛擬揚聲器過程為例進行說明。如圖 5 所示，該方法包括以下步驟。

【0126】 S510、編碼器 113 獲取三維音訊訊號的當前幀。

【0127】 編碼器 113 可以獲取由音訊獲取器 111 採集的原始音訊經過預處理 112 處理後的三維音訊訊號的當前幀。關於三維音訊訊號的當前幀相關解釋可以參考上述 S410 的闡述。

【0128】 S520、編碼器 113 根據三維音訊訊號的當前幀獲取當前幀的初始虛擬揚聲器的編碼效率。

【0129】 編碼器 113 根據三維音訊訊號的當前幀，從候選虛擬揚聲器集合中選取當前幀的初始虛擬揚聲器。當前幀的初始虛擬揚聲器屬於候選虛擬揚聲器集合。當前幀的初始虛擬揚聲器的數量小於或等於候選虛擬揚聲器集合包含的虛擬揚聲器的數量。關於獲取初始虛擬揚聲器的具體方法可以參考上述 S420 和 S430，以及下述圖 11 中獲取代表虛擬揚聲器的闡述。

【0130】 當前幀的初始虛擬揚聲器的編碼效率表示當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力。可理解的，若當前幀的初始虛擬揚聲器充分表達了三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的

能力較強。若當前幀的初始虛擬揚聲器不能充分表達三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較弱。

【0131】 下面對編碼器 113 獲取當前幀的初始虛擬揚聲器的編碼效率的方法進行說明。

【0132】 在第一種可能的實現方式中，編碼器 113 根據重建當前幀的能量與當前幀的能量確定當前幀的初始虛擬揚聲器的編碼效率後，執行 S530。其中，編碼器 113 先根據三維音訊訊號的當前幀和當前幀的初始虛擬揚聲器確定當前幀的虛擬揚聲器訊號，以及，根據當前幀的初始虛擬揚聲器和虛擬揚聲器訊號確定重建後三維音訊訊號的重建當前幀。需要說明的是，這裡的重建後三維音訊訊號的重建當前幀是編碼端預先估計的重建後三維音訊訊號，並非解碼端進行重建的重建後三維音訊訊號。具體地，關於生成當前幀的虛擬揚聲器訊號和重建後三維音訊訊號的重建當前幀的具體方法可以參考上述 S440 和 S450 中的闡述。當前幀的初始虛擬揚聲器的編碼效率可以滿足如下公式(6)。

【0133】 
$$R' = \frac{NRG_1}{NRG_2} \quad \text{公式(6)}$$

【0134】 其中， $R'$ 表示當前幀的初始虛擬揚聲器的編碼效率。 $NRG_1$ 表示重建當前幀的能量。 $NRG_2$ 表示當前幀的能量。

【0135】 在一些實施例中，重建當前幀的能量是根據重建當前幀的係數確定的。當前幀的能量是根據當前幀的係數確定的。例如，

編碼器 113 可以計算重建當前幀的每個通道的能量的表徵值  $R_1$ 、 $R_2$  至  $R_t$ ， $R_t = \text{norm}(SR_t)$ 。 $\text{norm}()$ 表示求取二範數運算， $SR_t$  表示重建當前幀的第  $t$  個通道包含的修正的離散余弦變換 (Modified Discrete Cosine Transform, MDCT) 係數。若三維音訊訊號為 HOA 訊號， $t$  的取值範圍為 1 至 (HOA 訊號的階數+1) 的平方。

【0136】 編碼器 113 可計算當前幀的能量的表徵值  $N_1$ 、 $N_2$  至  $N_t$ ， $N_t = \text{norm}(SN_t)$ 。 $SN_t$  表示當前幀的第  $t$  個通道包含的 MDCT 係數。

【0137】 因此，當前幀的初始虛擬揚聲器的編碼效率  $R' = \text{sum}(R) / \text{sum}(N)$ 。其中， $\text{sum}(R)$ 表示  $R_1$  至  $R_t$  之和， $NRG_1$ 等於  $\text{sum}(R)$ 。 $\text{sum}(N)$  表示  $N_1$  至  $N_t$  之和。 $NRG_2$ 等於  $\text{sum}(N)$ 。

【0138】 在第二種可能的實現方式中，編碼器 113 根據當前幀的虛擬揚聲器訊號的能量與當前幀的虛擬揚聲器訊號的能量和殘差訊號的能量之和的比值確定當前幀的初始虛擬揚聲器的編碼效率後，執行 S530。其中，當前幀的虛擬揚聲器訊號的能量和殘差訊號的能量之和可以表示傳輸訊號的能量。編碼器 113 先根據三維音訊訊號的當前幀和當前幀的初始虛擬揚聲器確定當前幀的虛擬揚聲器訊號，以及，根據當前幀的初始虛擬揚聲器和虛擬揚聲器訊號確定重建後三維音訊訊號的重建當前幀，根據當前幀和重建當前幀獲取當前幀的殘差訊號。具體地，關於生成殘差訊號的具體方法可以參考上述 S460 中的闡述。當前幀的初始虛擬揚聲器的編碼效率可以滿足如下公式(7)。

$$\text{【0139】 } R' = \frac{NRG_3}{NRG_3 + NRG_4} \quad \text{公式(7)}$$

【0140】 其中， $R'$ 表示當前幀的初始虛擬揚聲器的編碼效率。 $NRG_3$ 表示當前幀的虛擬揚聲器訊號的能量。 $NRG_4$ 表示殘差訊號的能量。

【0141】 在第三種可能的實現方式中，編碼器 113 根據當前幀的初始虛擬揚聲器的數量與聲源數量的比值確定當前幀的初始虛擬揚聲器的編碼效率後，執行 S530。其中，編碼器 113 可以根據三維音訊訊號的當前幀確定聲源數量。具體地，關於確定三維音訊訊號的聲源數量的具體方法可以參考上述編碼分析單元 330 中的闡述。當前幀的初始虛擬揚聲器的編碼效率可以滿足如下公式(8)。

$$\text{【0142】 } R' = \frac{N_1}{N_2} \quad \text{公式(8)}$$

【0143】 其中， $R'$ 表示當前幀的初始虛擬揚聲器的編碼效率。 $N_1$ 表示當前幀的初始虛擬揚聲器的數量。 $N_2$ 表示三維音訊訊號的聲源數量。聲源數量例如可以是根據實際場景預先佈置的。聲源數量可以是大於等於 1 的整數。

【0144】 在第四種可能的實現方式中，編碼器 113 根據當前幀的虛擬揚聲器訊號的數量與三維音訊訊號的聲源數量的比值確定當前幀的初始虛擬揚聲器的編碼效率後，執行 S530。當前幀的初始虛擬揚聲器的編碼效率可以滿足如下公式(9)。

$$\text{【0145】 } R' = \frac{N_3}{N_2} \quad \text{公式(9)}$$

【0146】 其中， $R'$ 表示當前幀的初始虛擬揚聲器的編碼效率。 $N_3$ 表

示當前幀的虛擬揚聲器訊號的數量。 $N_2$ 表示三維音訊訊號的聲源數量。

【0147】 S530、編碼器 113 判斷當前幀的初始虛擬揚聲器的編碼效率是否滿足預設條件。

【0148】 若當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，表示當前幀的初始虛擬揚聲器不能充分表達三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較弱，編碼器 113 執行 S540 和 S550。

【0149】 若當前幀的初始虛擬揚聲器的編碼效率不滿足預設條件，表示當前幀的初始虛擬揚聲器充分表達了三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較強，編碼器 113 執行 S560。

【0150】 示例地，預設條件包括當前幀的初始虛擬揚聲器的編碼效率小於第一閾值。編碼器 113 可以判斷當前幀的初始虛擬揚聲器的編碼效率是否小於第一閾值。

【0151】 需要說明的是，針對上述四種不同的可能的實現方式，第一閾值的取值範圍可能不同。

【0152】 例如，在第一種可能的實現方式中，第一閾值的取值範圍可以為 0.5 至 1。可理解的，若編碼效率小於 0.5，表示重建當前幀的能量小於當前幀的能量的一半，表示當前幀的初始虛擬揚聲器不能充分表達三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較弱。

【0153】 又如，在第二種可能的實現方式中，第一閾值的取值範圍可以為 0.5 至 1。可理解的，若編碼效率小於 0.5，表示當前幀的虛擬揚聲器訊號的能量小於傳輸訊號的能量的一半，表示當前幀的初始虛擬揚聲器不能充分表達三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較弱。

【0154】 又如，在第三種可能的實現方式中，第一閾值的取值範圍可以為 0 至 1。可理解的，若編碼效率小於 1，表示當前幀的初始虛擬揚聲器的數量小於三維音訊訊號的聲源數量，表示當前幀的初始虛擬揚聲器不能充分表達三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較弱。例如，當前幀的初始虛擬揚聲器的數量可以是 2，三維音訊訊號的聲源數量可以是 4。當前幀的初始虛擬揚聲器的數量是聲源數量的一半，表示當前幀的初始虛擬揚聲器不能充分表達三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較弱。

【0155】 又如，在第四種可能的實現方式中，第一閾值的取值範圍可以為 0 至 1。可理解的，若編碼效率小於 1，表示當前幀的虛擬揚聲器訊號的數量小於三維音訊訊號的聲源數量，表示當前幀的初始虛擬揚聲器不能充分表達三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較弱。例如，當前幀的虛擬揚聲器訊號的數量可以是 2，三維音訊訊

號的聲源數量可以是 4。當前幀的虛擬揚聲器訊號的數量是聲源數量的一半，表示當前幀的初始虛擬揚聲器不能充分表達三維音訊訊號的聲場資訊，當前幀的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力較弱。

**【0156】** 在一些實施例中，第一閾值也可以是一個具體的值。例如，第一閾值為 0.65。

**【0157】** 可理解的，第一閾值越大，預設條件越嚴格，則編碼器 113 進行重新選擇虛擬揚聲器的幾率越大且選擇當前幀的虛擬揚聲器的複雜度越高，三維音訊訊號的不同幀之間進行編碼所使用的虛擬揚聲器的波動性越小；反之，第一閾值越小，預設條件越寬鬆，則編碼器 113 進行重新選擇虛擬揚聲器的幾率越小且選擇當前幀的虛擬揚聲器的複雜度越低，三維音訊訊號的不同幀之間進行編碼所使用的虛擬揚聲器的波動性越大。第一閾值可以根據實際的應用場景進行設置，本實施例對第一閾值的具體取值不予限定。

**【0158】** S540、編碼器 113 從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器。

**【0159】** 在一種可能的示例中，如圖 6 所示，圖 6 與圖 3 的區別在於，編碼器 300 還包含後處理單元 3200。後處理單元 3200 分別與虛擬揚聲器訊號生成單元 350 和訊號重建單元 370 連接。後處理單元 3200 可以從訊號重建單元 370 獲取重建後三維音訊訊號的重建當前幀後，根據重建當前幀的能量與當前幀的能量確定當前

幀的初始虛擬揚聲器的編碼效率。若後處理單元 3200 確定當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器。進而，後處理單元 3200 將當前幀的更新虛擬揚聲器回饋給訊號重建單元 370、虛擬揚聲器訊號生成單元 350 和編碼單元 360，虛擬揚聲器訊號生成單元 350 根據當前幀的更新虛擬揚聲器和當前幀生成虛擬揚聲器訊號，訊號重建單元 370 根據當前幀的更新虛擬揚聲器和更新虛擬揚聲器訊號生成重建後三維音訊訊號。使得殘差訊號生成單元 380、殘差訊號選擇單元 390、訊號補償單元 3100 和編碼單元 360 中每個單元的輸入和輸出均是與當前幀的更新虛擬揚聲器相關的資訊（如：重建後三維音訊訊號和虛擬揚聲器訊號），與依據當前幀的初始虛擬揚聲器生成的資訊不同。可理解地，在後處理單元 3200 獲取到當前幀的更新虛擬揚聲器後，編碼器 113 根據更新虛擬揚聲器執行 S440 至 S480 的步驟。

**【0160】** 如圖 7 所示，圖 7 與圖 6 的區別在於，編碼器 300 還包含後處理單元 3200。後處理單元 3200 分別與虛擬揚聲器訊號生成單元 350 和殘差訊號生成單元 380 連接。後處理單元 3200 可以從虛擬揚聲器訊號生成單元 350 獲取當前幀的虛擬揚聲器訊號，以及從殘差訊號生成單元 380 獲取殘差訊號後，根據當前幀的虛擬揚聲器訊號的能量與當前幀的虛擬揚聲器訊號的能量和殘差訊號的能量之和的比值確定當前幀的初始虛擬揚聲器的編碼效率。若後處理單元 3200 確定當前幀的初始虛擬揚聲器的編碼效率滿足預

設條件，從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器。

【0161】 如圖 8 所示，圖 8 與圖 6 的區別在於，編碼器 300 還包含後處理單元 3200。後處理單元 3200 分別與編碼分析單元 330 和虛擬揚聲器選擇單元 340 連接。後處理單元 3200 可以從編碼分析單元 330 獲取三維音訊訊號的聲源數量，以及從虛擬揚聲器選擇單元 340 獲取當前幀的初始虛擬揚聲器的數量後，根據當前幀的初始虛擬揚聲器的數量與三維音訊訊號的聲源數量的比值確定當前幀的初始虛擬揚聲器的編碼效率。若後處理單元 3200 確定當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器。當前幀的初始虛擬揚聲器的數量可以是預先設置或者是虛擬揚聲器選擇單元 340 分析得到的。

【0162】 如圖 9 所示，圖 9 與圖 8 的區別在於，編碼器 300 還包含後處理單元 3200。後處理單元 3200 分別與編碼分析單元 330 和虛擬揚聲器訊號生成單元 350 連接。後處理單元 3200 可以從編碼分析單元 330 獲取三維音訊訊號的聲源數量，以及從虛擬揚聲器訊號生成單元 350 獲取當前幀的虛擬揚聲器訊號的數量後，根據當前幀的虛擬揚聲器訊號的數量與三維音訊訊號的聲源數量的比值確定當前幀的初始虛擬揚聲器的編碼效率。若後處理單元 3200 確定當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器。當前幀的虛擬揚聲器訊號的數量可以是預先設置或者是虛擬揚聲器選擇單

元 340 分析得到的。

【0163】 若當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，編碼器 113 可以進一步根據小於第一閾值的第二閾值判斷編碼效率，以便於編碼器 113 重選當前幀的虛擬揚聲器的準確性。

【0164】 示例地，如圖 10 所示，圖 10 所述的方法流程是對圖 5 中 S540 所包括的具體操作過程的闡述。

【0165】 S541、編碼器 113 判斷當前幀的初始虛擬揚聲器的編碼效率是否小於第二閾值。

【0166】 若當前幀的初始虛擬揚聲器的編碼效率小於或等於第二閾值，執行 S542；若當前幀的初始虛擬揚聲器的編碼效率大於第二閾值，且編碼效率小於第一閾值，執行 S543。

【0167】 S542、編碼器 113 將候選虛擬揚聲器集合中的預設虛擬揚聲器作為當前幀的更新虛擬揚聲器。

【0168】 預設虛擬揚聲器可以是指定的虛擬揚聲器。指定的虛擬揚聲器可以是虛擬揚聲器集合中任意一個虛擬揚聲器。例如，指定的虛擬揚聲器的水平角為 100 度，且俯仰角為 50 度。

【0169】 預設虛擬揚聲器可以是根據標準揚聲器佈局的虛擬揚聲器或非標準揚聲器佈局的虛擬揚聲器。標準揚聲器可以是指依據 22.2 聲道、7.1.4 聲道、5.1.4 聲道、7.1 聲道或 5.1 聲道等設置的揚聲器。非標準揚聲器可以是指根據實際場景預先佈置的揚聲器。

【0170】 預設虛擬揚聲器還可以是根據聲場中聲源位置確定的虛擬揚聲器。聲源位置可以是從上述編碼分析單元 330 獲得，或者

從待編碼的三維音訊訊號中獲得。

【0171】 S543、編碼器 113 將在先幀的虛擬揚聲器作為當前幀的更新虛擬揚聲器。

【0172】 在先幀的虛擬揚聲器為對三維音訊訊號的在先幀進行編碼所使用的虛擬揚聲器。

【0173】 需要說明的是，編碼器 113 將當前幀的更新虛擬揚聲器作為當前幀的代表虛擬揚聲器對當前幀進行編碼。

【0174】 可選地，若當前幀的初始虛擬揚聲器的編碼效率大於第二閾值，且編碼效率小於第一閾值，編碼器 113 還可以根據當前幀的初始虛擬揚聲器的編碼效率和在先幀的虛擬揚聲器的編碼效率確定當前幀的初始虛擬揚聲器的調整後編碼效率。示例地，編碼器 113 可以根據當前幀的初始虛擬揚聲器的編碼效率和在先幀的虛擬揚聲器的平均編碼效率生成當前幀的初始虛擬揚聲器的調整後編碼效率。調整後編碼效率滿足公式(10)。

【0175】 
$$MR' = \frac{(R' + MR)}{2} \quad \text{公式(10)}$$

【0176】 其中， $R'$  表示當前幀的初始虛擬揚聲器的編碼效率。 $MR'$  表示調整後編碼效率， $MR$  表示在先幀的虛擬揚聲器的平均編碼效率。在先幀可以是指當前幀之前的一個或多個幀。

【0177】 若當前幀的初始虛擬揚聲器的編碼效率大於當前幀的初始虛擬揚聲器的調整後編碼效率，表示當前幀的初始虛擬揚聲器相比在先幀的虛擬揚聲器可以充分地表達三維音訊訊號的聲場資

訊。因此，編碼器 113 將當前幀的初始虛擬揚聲器作為當前幀的後續幀的虛擬揚聲器。從而，進一步地降低三維音訊訊號的不同幀進行編碼所使用的虛擬揚聲器波動性，確保提高解碼端重建後三維音訊訊號的品質，以及解碼端播放的聲音的音質。

**【0178】** 若當前幀的初始虛擬揚聲器的編碼效率小於當前幀的初始虛擬揚聲器的調整後編碼效率，表示當前幀的初始虛擬揚聲器相比在先幀的虛擬揚聲器不能充分地表達三維音訊訊號的聲場資訊，可以將在先幀的虛擬揚聲器作為當前幀的後續幀的虛擬揚聲器。

**【0179】** 需要說明的是，第二閾值可以是一個具體的值。第二閾值小於第一閾值。例如，第二閾值為 0.55。第一閾值和第二閾值的具體取值本實施例不予限定。

**【0180】** 可選地，在當前幀的初始虛擬揚聲器的編碼效率滿足預設條件的場景下，編碼器 113 可以根據預設細微性調整第一閾值。例如，預設細微性可以為 0.1。示例地，第一閾值為 0.65，第二閾值為 0.55，第三閾值為 0.45。若當前幀的初始虛擬揚聲器的編碼效率小於或等於第二閾值，編碼器 113 可以判斷當前幀的初始虛擬揚聲器的編碼效率是否小於第三閾值。

**【0181】** S550、編碼器 113 根據當前幀的更新虛擬揚聲器對當前幀進行編碼，得到第一碼流。

**【0182】** 編碼器 113 根據當前幀的更新虛擬揚聲器和當前幀生成更新虛擬揚聲器訊號，根據當前幀的更新虛擬揚聲器和更新虛擬

揚聲器訊號生成更新重建後三維音訊訊號，根據更新重建當前幀和當前幀確定更新殘差訊號；根據當前幀和更新殘差訊號確定第一碼流。編碼器 113 可以根據上述 S430 至 S480 的闡述生成第一碼流，即編碼器 113 更新當前幀的初始虛擬揚聲器，利用當前幀的更新虛擬揚聲器、更新殘差訊號和更新補償資訊進行編碼得到第一碼流。

**【0183】** S560、編碼器 113 根據當前幀的初始虛擬揚聲器對當前幀進行編碼，得到第二碼流。

**【0184】** 編碼器 113 可以根據上述 S430 至 S480 的闡述生成第二碼流，即編碼器 113 無需更新當前幀的初始虛擬揚聲器，利用當前幀的初始虛擬揚聲器、殘差訊號和補償資訊進行編碼得到第二碼流。

**【0185】** 如此，在當前幀的初始虛擬揚聲器無法充分表示重建三維音訊訊號所屬聲場，導致解碼端重建後三維音訊訊號的品質較差的場景下，編碼器可以依據初始虛擬揚聲器的編碼效率指示的初始虛擬揚聲器用於重建三維音訊訊號所屬聲場的能力，確定重新選擇當前幀的虛擬揚聲器，則編碼器將當前幀的更新虛擬揚聲器作為對當前幀進行編碼的虛擬揚聲器。從而，編碼器通過重選虛擬揚聲器，降低三維音訊訊號的不同幀之間進行編碼所使用的虛擬揚聲器的波動性，提高解碼端重建後三維音訊訊號的品質，以及解碼端播放的聲音的音質。

**【0186】** 在一些實施例中，源設備 110 根據當前幀的係數與虛擬

揚聲器的係數對虛擬揚聲器進行投票，根據虛擬揚聲器的投票值從候選虛擬揚聲器集合中選擇當前幀的代表虛擬揚聲器，實現對待編碼的三維音訊訊號進行資料壓縮的目的。在本實施例中，當前幀的代表虛擬揚聲器可以作為上述各實施例的初始虛擬揚聲器。

【0187】 圖 11 為本申請實施例提供的一種選擇虛擬揚聲器方法的流程示意圖。圖 11 所述的方法流程是對圖 4 中 S430 所包括的具體操作過程的闡述。在這裡由圖 1 所示的源設備 110 中編碼器 113 執行選擇虛擬揚聲器過程為例進行說明。具體地實現虛擬揚聲器選擇單元 340 的功能。如圖 11 所示，該方法包括以下步驟。

【0188】 S1110、編碼器 113 獲取當前幀的代表係數。

【0189】 代表係數可以是指頻域代表係數或時域代表係數。頻域代表係數也可以稱為頻域代表頻點或頻譜代表係數。時域代表係數也可以稱為時域代表採樣點。

【0190】 示例地，編碼器 113 獲取到三維音訊訊號的當前幀的第四數量個係數，以及第四數量個係數的頻域特徵值後，根據第四數量個係數的頻域特徵值，從第四數量個係數中選取第三數量個代表係數，進而，根據第三數量個代表係數從候選虛擬揚聲器集合中選取第二數量個當前幀的代表虛擬揚聲器。其中，所述第四數量個係數包括第三數量個代表係數，第三數量小於第四數量，表示第三數量個代表係數是第四數量個係數中的部分係數。三維音訊訊號的當前幀為 HOA 訊號；係數的頻域特徵值是依據 HOA

訊號的係數確定的。

【0191】 如此，由於編碼器從當前幀的全部係數中選取部分係數作為代表係數，利用較少數量的代表係數代替當前幀的全部係數從候選虛擬揚聲器集合中選取代表虛擬揚聲器，因此有效地降低了編碼器搜索虛擬揚聲器的計算複雜度，從而降低了對三維音訊訊號進行壓縮編碼的計算複雜度以及減輕了編碼器的計算負擔。

【0192】 S1120、編碼器 113 根據當前幀的代表係數對候選虛擬揚聲器集合中虛擬揚聲器的投票值，從候選虛擬揚聲器集合中選取當前幀的代表虛擬揚聲器。

【0193】 編碼器 113 根據當前幀的代表係數與虛擬揚聲器的係數對候選虛擬揚聲器集合中的虛擬揚聲器進行投票，根據虛擬揚聲器的當前幀最終投票值從候選虛擬揚聲器集合中選擇（搜索）當前幀的代表虛擬揚聲器。

【0194】 示例地，編碼器 113 根據當前幀的第三數量個代表係數、候選虛擬揚聲器集合和投票輪數確定第一數量個虛擬揚聲器和第一數量個投票值，根據第一數量個投票值，從第一數量個虛擬揚聲器中選取第二數量個當前幀的代表虛擬揚聲器，第二數量小於第一數量，表示第二數量個當前幀的代表虛擬揚聲器是候選虛擬揚聲器集合中的部分虛擬揚聲器。可理解的，虛擬揚聲器與投票值一一對應。例如，第一數量個虛擬揚聲器包括第一虛擬揚聲器，第一數量個投票值包括第一虛擬揚聲器的投票值，第一虛擬揚聲器與第一虛擬揚聲器的投票值對應。第一虛擬揚聲器的投票值用

於表徵對當前幀進行編碼時使用第一虛擬揚聲器的優先順序。候選虛擬揚聲器集合包括第五數量個虛擬揚聲器，第五數量個虛擬揚聲器包括第一數量個虛擬揚聲器，第一數量小於或等於第五數量，投票輪數為大於或等於 1 的整數，且投票輪數小於或等於第五數量。

**【0195】** 目前，在虛擬揚聲器搜索過程中，編碼器依據待編碼的三維音訊訊號和虛擬揚聲器之間的相關計算的結果作為虛擬揚聲器的選擇衡量指標。而且，若編碼器對每一個係數傳輸一個虛擬揚聲器，則無法達到高效資料壓縮的目的，會對編碼器造成沉重的計算負擔。本申請實施例提供的選擇虛擬揚聲器的方法，編碼器利用較少數量的代表係數代替當前幀的全部係數對候選虛擬揚聲器集合中每個虛擬揚聲器進行投票，依據投票值選取當前幀的代表虛擬揚聲器。進而，編碼器利用當前幀的代表虛擬揚聲器對待編碼的三維音訊訊號進行壓縮編碼，不僅有效地提升了對三維音訊訊號進行壓縮編碼的壓縮率，而且降低了編碼器搜索虛擬揚聲器的計算複雜度，從而降低了對三維音訊訊號進行壓縮編碼的計算複雜度以及減輕了編碼器的計算負擔。

**【0196】** 第二數量用於表徵編碼器選取的當前幀的代表虛擬揚聲器的數量。第二數量越大表示當前幀的代表虛擬揚聲器的數量越大，三維音訊訊號的聲場資訊越多；第二數量越小表示當前幀的代表虛擬揚聲器的數量越小，三維音訊訊號的聲場資訊越少。因此，可通過設置第二數量控制編碼器選取的當前幀的代表虛擬揚

聲器的數量。例如，第二數量可以是預設的，又如，第二數量可以是根據當前幀確定的。示例地，第二數量的取值可以是 1、2、4 或 8。

**【0197】** 需要說明的是，編碼器先遍歷候選虛擬揚聲器集合包含的虛擬揚聲器，利用從候選虛擬揚聲器集合中選取的當前幀的代表虛擬揚聲器對當前幀進行壓縮。但是，若連續幀選取的虛擬揚聲器的結果差異較大，會導致重建後三維音訊訊號的聲像不穩定，降低重建後三維音訊訊號的音質。在本申請的實施例中，編碼器 113 可以依據在先幀的代表虛擬揚聲器的在先幀最終投票值對候選虛擬揚聲器集合包含的虛擬揚聲器的當前幀初始投票值進行更新處理，得到虛擬揚聲器的當前幀最終投票值，則根據虛擬揚聲器的當前幀最終投票值從候選虛擬揚聲器集合中選取當前幀的代表虛擬揚聲器。從而，通過參考在先幀的代表虛擬揚聲器來選取當前幀的代表虛擬揚聲器，使編碼器對當前幀選擇當前幀的代表虛擬揚聲器時傾向於選擇與在先幀的代表虛擬揚聲器相同的虛擬揚聲器，增加連續幀之間的方位的連續性，克服了連續幀選取的虛擬揚聲器的結果差異較大的問題。因此，本申請的實施例還可以包括 S1130。

**【0198】** S1130、編碼器 113 根據在先幀的代表虛擬揚聲器的在先幀最終投票值調整候選虛擬揚聲器集合中虛擬揚聲器的當前幀初始投票值，獲得虛擬揚聲器的當前幀最終投票值。

**【0199】** 編碼器 113 根據當前幀的代表係數與虛擬揚聲器的係數

對候選虛擬揚聲器集合中的虛擬揚聲器進行投票，得到虛擬揚聲器的當前幀初始投票值後，根據在先幀的代表虛擬揚聲器的在先幀最終投票值調整候選虛擬揚聲器集合中虛擬揚聲器的當前幀初始投票值，獲得虛擬揚聲器的當前幀最終投票值。在先幀的代表虛擬揚聲器為編碼器 113 對在先幀進行編碼時使用的虛擬揚聲器。

**【0200】** 編碼器 113 根據第一數量個投票值，以及第六數量個在先幀最終投票值，獲取第七數量個虛擬揚聲器與當前幀對應的第七數量個當前幀最終投票值，根據第七數量個當前幀最終投票值，從第七數量個虛擬揚聲器中選取第二數量個當前幀的代表虛擬揚聲器，第二數量小於第七數量，表示第二數量個當前幀的代表虛擬揚聲器是第七數量個虛擬揚聲器中的部分虛擬揚聲器。其中，第七數量個虛擬揚聲器包括第一數量個虛擬揚聲器，且第七數量個虛擬揚聲器包括第六數量個虛擬揚聲器，第六數量個虛擬揚聲器包含的虛擬揚聲器為對三維音訊訊號的在先幀進行編碼所使用的在先幀的代表虛擬揚聲器。在先幀的代表虛擬揚聲器集合包含的第六數量個虛擬揚聲器與所述第六數量個在先幀最終投票值一一對應。

**【0201】** 在虛擬揚聲器搜索過程中，由於真實聲源的位置與虛擬揚聲器的位置不一定重合，會導致虛擬揚聲器不一定能夠與真實聲源形成一一對應關係，且由於在實際的複雜場景下，可能出現有限數量的虛擬揚聲器集合無法表徵聲場中所有聲源的情況，此時，幀與幀之間搜索到的虛擬揚聲器可能會發生頻繁跳變，這種

跳變會明顯地影響聽音者的聽覺感受，導致解碼重建後三維音訊訊號中出現明顯的不連續和雜訊現象。本申請的實施例提供的選擇虛擬揚聲器的方法通過繼承在先幀的代表虛擬揚聲器，即對於相同編號的虛擬揚聲器，用在先幀最終投票值調整當前幀初始投票值，使得編碼器更傾向於選擇在先幀的代表虛擬揚聲器，從而降低幀與幀之間的虛擬揚聲器的頻繁跳變，增強了幀之間的訊號方位的連續性，提高了重建後三維音訊訊號的聲像的穩定性，確保重建後三維音訊訊號的音質。

**【0202】** 在一些實施例中，若當前幀是原始音訊中第一幀，編碼器 113 執行 S1110 至 S1120。若當前幀是原始音訊中第二幀以上的任意一幀，編碼器 113 可以先判斷是否複用在先幀的代表虛擬揚聲器對當前幀進行編碼或判斷是否進行虛擬揚聲器搜索，確保連續幀之間的方位的連續性，並降低編碼複雜度。本申請的實施例還可以包括 S1140。

**【0203】** S1140、編碼器 113 根據在先幀的代表虛擬揚聲器和當前幀判斷是否進行虛擬揚聲器搜索。

**【0204】** 若編碼器 113 確定進行虛擬揚聲器搜索，執行 S1110 至 S1130。可選地，編碼器 113 可以先執行 S1110，即編碼器 113 獲取當前幀的代表係數，編碼器 113 根據當前幀的代表係數和在先幀的代表虛擬揚聲器的係數判斷是否進行虛擬揚聲器搜索，若編碼器 113 確定進行虛擬揚聲器搜索，再執行 S1120 至 S1130。

**【0205】** 若編碼器 113 確定不進行虛擬揚聲器搜索，執行 S1150。

【0206】 S1150、編碼器 113 確定複用在先幀的代表虛擬揚聲器對當前幀進行編碼。

【0207】 編碼器 113 複用在先幀的代表虛擬揚聲器和當前幀生成虛擬揚聲器訊號，對虛擬揚聲器訊號進行編碼得到碼流，向目的設備 120 發送碼流。

【0208】 可選地，在本申請實施例提供的重新虛擬揚聲器的過程中，若當前幀的初始虛擬揚聲器是根據在先幀的代表虛擬揚聲器的投票值確定的，而當前幀的初始虛擬揚聲器的編碼效率小於第一閾值，編碼器 113 可以將在先幀的代表虛擬揚聲器的投票值清零，從而，避免編碼器 113 選擇不能充分表達三維音訊訊號的聲場資訊在先幀的代表虛擬揚聲器，導致重建後三維音訊訊號的品質較低，解碼端播放的聲音的音質較差。

【0209】 可以理解的是，為了實現上述實施例中的功能，編碼器包括了執行各個功能相應的硬體結構和/或軟體模組。本領域技術人員應該很容易意識到，結合本申請中所公開的實施例描述的各示例的單元及方法步驟，本申請能夠以硬體或硬體和電腦軟體相結合的形式來實現。某個功能究竟以硬體還是電腦軟體驅動硬體的方式來執行，取決於技術方案的特定應用場景和設計約束條件。

【0210】 上文中結合圖 1 至圖 11，詳細描述了根據本實施例所提供的三維音訊訊號編碼方法，下面將結合圖 12 和圖 13，描述根據本實施例所提供的三維音訊訊號編碼裝置和編碼器。

【0211】 圖 12 為本實施例提供的可能的三維音訊訊號編碼裝置的

結構示意圖。這些三維音訊訊號編碼裝置可以用於實現上述方法實施例中編碼三維音訊訊號的功能，因此也能實現上述方法實施例所具備的有益效果。在本實施例中，該三維音訊訊號編碼裝置可以是如圖 1 所示的編碼器 113，或者如圖 3 所示的編碼器 300，還可以是應用於終端設備或伺服器的模組（如晶片）。

【0212】 如圖 12 所示，三維音訊訊號編碼裝置 1200 包括通信模組 1210、編碼效率獲取模組 1220、虛擬揚聲器重選模組 1230、編碼模組 1240 和儲存模組 1250。三維音訊訊號編碼裝置 1200 用於實現上述圖 5 和圖 10 中所示的方法實施例中編碼器 113 的功能。

【0213】 通信模組 1210 用於獲取三維音訊訊號的當前幀。可選地，通信模組 1210 也可以接收其他設備獲取的三維音訊訊號的當前幀；或者從儲存模組 1250 獲取三維音訊訊號的當前幀。三維音訊訊號為 HOA 訊號；係數的頻域特徵值是依據二維向量確定的，二維向量包括 HOA 訊號的 HOA 係數。

【0214】 編碼效率獲取模組 1220，用於根據三維音訊訊號的當前幀獲取當前幀的初始虛擬揚聲器的編碼效率，當前幀的初始虛擬揚聲器屬於候選虛擬揚聲器集合。當三維音訊訊號編碼裝置 1200 用於實現圖 5 和圖 10 所示的方法實施例中編碼器 113 的功能時，編碼效率獲取模組 1220 用於實現 S520 的相關功能。

【0215】 虛擬揚聲器重選模組 1230，用於若當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，從候選虛擬揚聲器集合中確定當前幀的更新虛擬揚聲器。當三維音訊訊號編碼裝置 1200 用於實現

圖 5 所示的方法實施例中編碼器 113 的功能時，虛擬揚聲器重選模組 1230 用於實現 S530 和 S540 的相關功能。當三維音訊訊號編碼裝置 1200 用於實現圖 10 所示的方法實施例中編碼器 113 的功能時，虛擬揚聲器重選模組 1230 用於實現 S530、S541 至 S543 的相關功能。

【0216】 若所述當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，編碼模組 1240 用於根據所述當前幀的更新虛擬揚聲器對所述當前幀進行編碼，得到第一碼流。

【0217】 若所述當前幀的初始虛擬揚聲器的編碼效率不滿足所述預設條件，編碼模組 1240 用於根據所述當前幀的初始虛擬揚聲器對所述當前幀進行編碼，得到第二碼流。

【0218】 當三維音訊訊號編碼裝置 1200 用於實現圖 5 和圖 10 所示的方法實施例中編碼器 113 的功能時，編碼模組 1240 用於實現 S550 和 S560 的相關功能。

【0219】 儲存模組 1250 用於儲存與三維音訊訊號相關的係數，候選虛擬揚聲器集合，在先幀的代表虛擬揚聲器集合，碼流，以及選取的係數和虛擬揚聲器等，以便於編碼模組 1240 對當前幀進行編碼得到碼流，並將碼流傳輸至解碼器。

【0220】 應理解的是，本申請實施例的三維音訊訊號編碼裝置 1200 可以通過專用積體電路（application-specific integrated circuit, ASIC）實現，或可程式設計邏輯器件（programmable logic device, PLD）實現，上述 PLD 可以是複雜程式邏輯器件（complex

programmable logical device, CPLD), 現場可程式設計閘陣列 (field-programmable gate array, FPGA), 通用陣列邏輯 (generic array logic, GAL) 或其任意組合。也可以通過軟體實現圖 5 和圖 10 所示的三維音訊訊號編碼方法時, 三維音訊訊號編碼裝置 1200 及其各個模組也可以為軟體模組。

【0221】 有關上述通信模組 1210、編碼效率獲取模組 1220、虛擬揚聲器重選模組 1230、編碼模組 1240 和儲存模組 1250 更詳細的描述可以參考圖 5 和圖 10 所示的方法實施例中相關描述直接得到, 這裡不加贅述。

【0222】 圖 13 為本實施例提供的一種編碼器 1300 的結構示意圖。如圖所示, 編碼器 1300 包括處理器 1310、匯流排 1320、記憶體 1330 和通信介面 1340。

【0223】 應理解, 在本實施例中, 處理器 1310 可以是中央處理器 (central processing unit, CPU), 該處理器 1310 還可以是其他通用處理器、數位訊號處理器 (digital signal processing, DSP)、ASIC、FPGA 或者其他可程式設計邏輯器件、分立門或者電晶體邏輯器件、分立硬體元件等。通用處理器可以是微處理器或者是任何常規的處理器等。

【0224】 處理器還可以是圖形處理器 (graphics processing unit, GPU)、神經網路處理器 (neural network processing unit, NPU)、微處理器或一個或多個用於控制本申請方案程式執行的積體電路。

【0225】 通信介面 1340 用於實現編碼器 1300 與外部設備或器件的通信。在本實施例中，通信介面 1340 用於接收三維音訊訊號。

【0226】 匯流排 1320 可以包括一通路，用於在上述元件（如處理器 1310 和記憶體 1330）之間傳送信息。匯流排 1320 除包括資料匯流排之外，還可以包括電源匯流排、控制匯流排和狀態訊號匯流排等。但是為了清楚說明起見，在圖中將各種匯流排都標為匯流排 1320。

【0227】 作為一個示例，編碼器 1300 可以包括多個處理器。處理器可以是一個多核（multi-CPU）處理器。這裡的處理器可以指一個或多個設備、電路、和/或用於處理資料（例如電腦程式指令）的計算單元。處理器 1310 可以調用記憶體 1330 儲存的與三維音訊訊號相關的係數，候選虛擬揚聲器集合，在先幀的代表虛擬揚聲器集合，以及選取的係數和虛擬揚聲器等。

【0228】 值得說明的是，圖 13 中僅以編碼器 1300 包括 1 個處理器 1310 和 1 個記憶體 1330 為例，此處，處理器 1310 和記憶體 1330 分別用於指示一類器件或設備，具體實施例中，可以根據業務需求確定每種類型的器件或設備的數量。

【0229】 記憶體 1330 可以對應上述方法實施例中用於儲存與三維音訊訊號相關的係數，候選虛擬揚聲器集合，在先幀的代表虛擬揚聲器集合，以及選取的係數和虛擬揚聲器等資訊的儲存介質，例如，磁片，如機械硬碟或固態硬碟。

【0230】 上述編碼器 1300 可以是一個通用設備或者是一個專用設

備。例如，編碼器 1300 可以是基於 X86、ARM 的伺服器，也可以為其他的專用伺服器，如策略控制和計費（policy control and charging，PCC）伺服器等。本申請實施例不限定編碼器 1300 的類型。

【0231】 應理解，根據本實施例的編碼器 1300 可對應於本實施例中的三維音訊訊號編碼裝置 1200，並可以對應於執行根據圖 5 和圖 10 中任一方法中的相應主體，並且三維音訊訊號編碼裝置 1200 中的各個模組的上述和其它操作和/或功能分別為了實現圖 5 和圖 10 中的各個方法的相應流程，為了簡潔，在此不再贅述。

【0232】 本申請實施例還提供一種系統，該系統包括解碼器和如圖 13 所示的編碼器，編碼器和解碼器用於實現上述圖 5 和圖 10 所示的方法步驟，為了簡潔，在此不再贅述。

【0233】 本實施例中的方法步驟可以通過硬體的方式來實現，也可以由處理器執行軟體指令的方式來實現。軟體指令可以由相應的軟體模組組成，軟體模組可以被存放於隨機存取記憶體（random access memory，RAM）、快閃記憶體、唯讀記憶體（read-only memory，ROM）、可程式設計唯讀記憶體（programmable ROM，PROM）、可擦除可程式設計唯讀記憶體（erasable PROM，EPROM）、電可擦除可程式設計唯讀記憶體（electrically EPROM，EEPROM）、寄存器、硬碟、移動硬碟、CD-ROM 或者本領域熟知的任何其它形式的儲存介質中。一種示例性的儲存介質耦合至處理器，從而使處理器能夠從該儲存介質讀取資訊，且可向該儲存

介質寫入資訊。當然，儲存介質也可以是處理器的組成部分。處理器和儲存介質可以位於 ASIC 中。另外，該 ASIC 可以位於網路設備或終端設備中。當然，處理器和儲存介質也可以作為分立元件存在於網路設備或終端設備中。

**【0234】** 在上述實施例中，可以全部或部分地通過軟體、硬體、固件或者其任意組合來實現。當使用軟體實現時，可以全部或部分地以電腦程式產品的形式實現。所述電腦程式產品包括一個或多個電腦程式或指令。在電腦上載入和執行所述電腦程式或指令時，全部或部分地執行本申請實施例所述的流程或功能。所述電腦可以是通用電腦、專用電腦、電腦網路、網路設備、使用者設備或者其它可程式設計裝置。所述電腦程式或指令可以儲存在電腦可讀儲存介質中，或者從一個電腦可讀儲存介質向另一個電腦可讀儲存介質傳輸，例如，所述電腦程式或指令可以從一個網站、電腦、伺服器或資料中心通過有線或無線方式向另一個網站、電腦、伺服器或資料中心進行傳輸。所述電腦可讀儲存介質可以是電腦能夠存取的任何可用介質或者是集成一個或多個可用介質的伺服器、資料中心等資料存放裝置。所述可用介質可以是磁性介質，例如，軟碟、硬碟、磁帶；也可以是光介質，例如，數位視訊光碟（digital video disc，DVD）；還可以是半導體介質，例如，固態硬碟（solid state drive，SSD）。

**【0235】** 以上所述，僅為本申請的具體實施方式，但本申請的保護範圍並不局限於此，任何熟悉本技術領域的技術人員在本申請

揭露的技術範圍內，可輕易想到各種等效的修改或替換，這些修改或替換都應涵蓋在本申請的保護範圍之內。因此，本申請的保護範圍應以請求項的保護範圍為準。

**【符號說明】**

**【0236】**

100:音訊編解碼系統

110:源設備

111:音訊獲取器

112:預處理器

113:編碼器

114:通信介面

1131:空間編碼器

1132:核心編碼器

130:通信通道

120:目的設備

121:播放器

122:後處理器

123:解碼器

124:通信介面

1231:核心解碼器

1232 空間解碼器

- 300:編碼器
- 310:虛擬揚聲器配置單元
- 320:虛擬揚聲器集合生成單元
- 330:編碼分析單元
- 340:虛擬揚聲器選擇單元
- 350:虛擬揚聲器訊號生成單元
- 360:編碼單元
- 370:訊號重建單元
- 380:殘差訊號生成單元
- 390:殘差訊號選擇單元
- 3100:訊號補償單元
- 3200:後處理單元
- 1200:三維音訊訊號編碼裝置
- 1210:通信模組
- 1220:編碼效率獲取模組
- 1230:虛擬揚聲器重選模組
- 1240:編碼模組
- 1250:儲存模組
- 1300:編碼器
- 1310:處理器
- 1320:匯流排
- 1330:記憶體

1340:通信介面

S410、S420、S430、S440、S450、S460、S470、S480、S490、  
S4100、S510、S520、S530、S540、S550、S560、S541、S542、  
S543、S1110、S1120、S1130、S1140、S1150:步驟

## 【發明申請專利範圍】

【請求項1】 一種三維音訊訊號編碼方法，其中，包括：

獲取三維音訊訊號的當前幀；

根據所述三維音訊訊號的當前幀獲取所述當前幀的初始虛擬揚聲器的編碼效率，所述當前幀的初始虛擬揚聲器屬於候選虛擬揚聲器集合；

若所述當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，從所述候選虛擬揚聲器集合中確定所述當前幀的更新虛擬揚聲器，以及根據所述當前幀的更新虛擬揚聲器對所述當前幀進行編碼，得到第一碼流；

若所述當前幀的初始虛擬揚聲器的編碼效率不滿足所述預設條件，根據所述當前幀的初始虛擬揚聲器對所述當前幀進行編碼，得到第二碼流；

其中，所述預設條件包括所述當前幀的初始虛擬揚聲器的編碼效率小於第一閾值；

其中，所述從所述候選虛擬揚聲器集合中確定所述當前幀的更新虛擬揚聲器包括：

若所述當前幀的初始虛擬揚聲器的編碼效率小於第二閾值，將所述候選虛擬揚聲器集合中的預設虛擬揚聲器作為所述當前幀的更新虛擬揚聲器，所述第二閾值小於所述第一閾值；

或，若所述當前幀的初始虛擬揚聲器的編碼效率小於第一閾值，且大於第二閾值，將在先幀的虛擬揚聲器作為所述當前幀的

更新虛擬揚聲器，所述在先幀的虛擬揚聲器為對所述三維音訊訊號的在先幀進行編碼所使用的虛擬揚聲器。

【請求項2】 如請求項1所述的方法，其中，所述根據所述三維音訊訊號的當前幀獲取所述當前幀的初始虛擬揚聲器的編碼效率包括：

根據所述當前幀的初始虛擬揚聲器獲取重建後三維音訊訊號的重建當前幀；

根據所述重建當前幀的能量與所述當前幀的能量確定所述當前幀的初始虛擬揚聲器的編碼效率。

【請求項3】 如請求項2所述的方法，其中，所述重建當前幀的能量是根據所述重建當前幀的係數確定的，所述當前幀的能量是根據所述當前幀的係數確定的。

【請求項4】 如請求項1所述的方法，其中，所述根據所述三維音訊訊號的當前幀獲取所述當前幀的初始虛擬揚聲器的編碼效率包括：

根據所述當前幀的初始虛擬揚聲器獲取重建後三維音訊訊號的重建當前幀；

根據所述三維音訊訊號的當前幀和所述重建後三維音訊訊號的重建當前幀獲取所述當前幀的殘差訊號；

獲取所述當前幀的虛擬揚聲器訊號和所述殘差訊號的能量和；

根據所述當前幀的虛擬揚聲器訊號的能量與所述能量和的比值確定所述當前幀的初始虛擬揚聲器的編碼效率。

【請求項5】 如請求項2或4所述的方法，其中，所述根據所述當前幀的初始虛擬揚聲器獲取重建後三維音訊訊號的重建當前幀包括：

根據所述當前幀的初始虛擬揚聲器確定所述當前幀的虛擬揚聲器訊號；

根據所述當前幀的虛擬揚聲器訊號確定所述重建當前幀。

【請求項6】 如請求項1所述的方法，其中，所述根據所述三維音訊訊號的當前幀獲取所述當前幀的初始虛擬揚聲器的編碼效率包括：

根據所述三維音訊訊號的當前幀確定聲源數量；

根據所述當前幀的初始虛擬揚聲器的數量與所述聲源數量確定所述當前幀的初始虛擬揚聲器的編碼效率。

【請求項7】 如請求項1所述的方法，其中，所述根據所述三維音訊訊號的當前幀獲取所述當前幀的初始虛擬揚聲器的編碼效率包括：

根據所述三維音訊訊號的當前幀確定聲源數量；

根據所述當前幀的初始虛擬揚聲器確定所述當前幀的虛擬揚聲器訊號；

根據所述當前幀的虛擬揚聲器訊號的數量與所述三維音訊訊號的聲源數量確定所述當前幀的初始虛擬揚聲器的編碼效率。

【請求項8】 如請求項1所述的方法，其中，所述方法還包括：

根據所述當前幀的初始虛擬揚聲器的編碼效率和所述在先幀的虛擬揚聲器的編碼效率確定所述當前幀的初始虛擬揚聲器的調整後編碼效率；

若所述當前幀的初始虛擬揚聲器的編碼效率大於所述當前幀的初始虛擬揚聲器的調整後編碼效率，將所述當前幀的初始虛擬揚聲器作為所述當前幀的後續幀的虛擬揚聲器。

【請求項9】 如請求項1至4中任一項所述的方法，其中，所述三維音訊訊號為高階立體混響HOA訊號。

【請求項10】 一種三維音訊訊號編碼裝置，其中，包括：

通信模組，用於獲取三維音訊訊號的當前幀；

編碼效率獲取模組，用於根據所述三維音訊訊號的當前幀獲取所述當前幀的初始虛擬揚聲器的編碼效率，所述當前幀的初始虛擬揚聲器屬於候選虛擬揚聲器集合；

虛擬揚聲器重選模組，用於若所述當前幀的初始虛擬揚聲器的編碼效率滿足預設條件，從所述候選虛擬揚聲器集合中確定所述當前幀的更新虛擬揚聲器；

編碼模組，用於根據所述當前幀的更新虛擬揚聲器對所述當前幀進行編碼，得到第一碼流；

所述編碼模組，還用於若所述當前幀的初始虛擬揚聲器的編碼效率不滿足所述預設條件，根據所述當前幀的初始虛擬揚聲器對所述當前幀進行編碼，得到第二碼流；

其中，所述預設條件包括所述當前幀的初始虛擬揚聲器的編碼效率小於第一閾值；

其中，所述虛擬揚聲器重選模組從所述候選虛擬揚聲器集合中確定所述當前幀的更新虛擬揚聲器時，具體用於：

若所述當前幀的初始虛擬揚聲器的編碼效率小於第二閾值，將所述候選虛擬揚聲器集合中的預設虛擬揚聲器作為所述當前幀的更新虛擬揚聲器，所述第二閾值小於所述第一閾值；

或，若所述當前幀的初始虛擬揚聲器的編碼效率小於第一閾值，且大於第二閾值，將在先幀的虛擬揚聲器作為所述當前幀的更新虛擬揚聲器，所述在先幀的虛擬揚聲器為對所述三維音訊訊號的在先幀進行編碼所使用的虛擬揚聲器。

**【請求項11】** 如請求項10所述的裝置，其中，所述編碼效率獲取模組根據所述三維音訊訊號的當前幀獲取所述當前幀的初始虛擬揚聲器的編碼效率時，具體用於：

根據所述當前幀的初始虛擬揚聲器獲取重建後三維音訊訊號的重建當前幀；

根據所述重建當前幀的能量與所述當前幀的能量確定所述當前幀的初始虛擬揚聲器的編碼效率。

**【請求項12】** 如請求項11所述的裝置，其中，所述重建當前幀的能量是根據所述重建當前幀的係數確定的，所述當前幀的能量是根據所述當前幀的係數確定的。

**【請求項13】** 如請求項10所述的裝置，其中，所述編碼效率獲取模組根據所述三維音訊訊號的當前幀獲取所述當前幀的初始虛擬揚聲器的編碼效率時，具體用於：

根據所述當前幀的初始虛擬揚聲器獲取重建後三維音訊訊號的重建當前幀；

根據所述三維音訊訊號的當前幀和所述重建後三維音訊訊號的重建當前幀獲取所述當前幀的殘差訊號；

獲取所述當前幀的虛擬揚聲器訊號和所述殘差訊號的能量和；

根據所述當前幀的虛擬揚聲器訊號的能量與所述能量和的比值確定所述當前幀的初始虛擬揚聲器的編碼效率。

**【請求項14】** 如請求項11或13所述的裝置，其中，所述編碼效率獲取模組根據所述當前幀的初始虛擬揚聲器獲取重建後三維音訊訊號的重建當前幀時，具體用於：

根據所述當前幀的初始虛擬揚聲器確定所述當前幀的虛擬揚聲器訊號；

根據所述當前幀的虛擬揚聲器訊號確定所述重建當前幀。

**【請求項15】** 如請求項10所述的裝置，其中，所述編碼效率獲取模組根據所述三維音訊訊號的當前幀獲取所述當前幀的初始虛擬揚聲器的編碼效率時，具體用於：

根據所述三維音訊訊號的當前幀確定聲源數量；

根據所述當前幀的初始虛擬揚聲器的數量與所述聲源數量確定所述當前幀的初始虛擬揚聲器的編碼效率。

**【請求項16】** 如請求項10所述的裝置，其中，所述編碼效率獲取模組根據所述三維音訊訊號的當前幀獲取所述當前幀的初始虛擬揚聲器的編碼效率時，具體用於：

根據所述三維音訊訊號的當前幀確定聲源數量；

根據所述當前幀的初始虛擬揚聲器確定所述當前幀的虛擬揚聲器訊號；

根據所述當前幀的虛擬揚聲器訊號的數量與所述三維音訊訊號的聲源數量確定所述當前幀的初始虛擬揚聲器的編碼效率。

**【請求項17】** 如請求項10所述的裝置，其中，所述虛擬揚聲器重選模組還用於：

根據所述當前幀的初始虛擬揚聲器的編碼效率和所述在先幀的虛擬揚聲器的編碼效率確定所述當前幀的初始虛擬揚聲器的調整後編碼效率；

若所述當前幀的初始虛擬揚聲器的編碼效率大於所述當前幀的初始虛擬揚聲器的調整後編碼效率，將所述當前幀的初始虛擬揚聲器作為所述當前幀的後續幀的虛擬揚聲器。

**【請求項18】** 如請求項10至13中任一項所述的裝置，其中，所述三維音訊訊號為高階立體混響HOA訊號。

**【請求項19】** 一種編碼器，其中，所述編碼器包括至少一個處理器和記憶體，其中，所述記憶體用於儲存電腦程式，使得所述電腦程式被所述至少一個處理器執行時實現如請求項1至9中任一項所述的三維音訊訊號編碼方法。

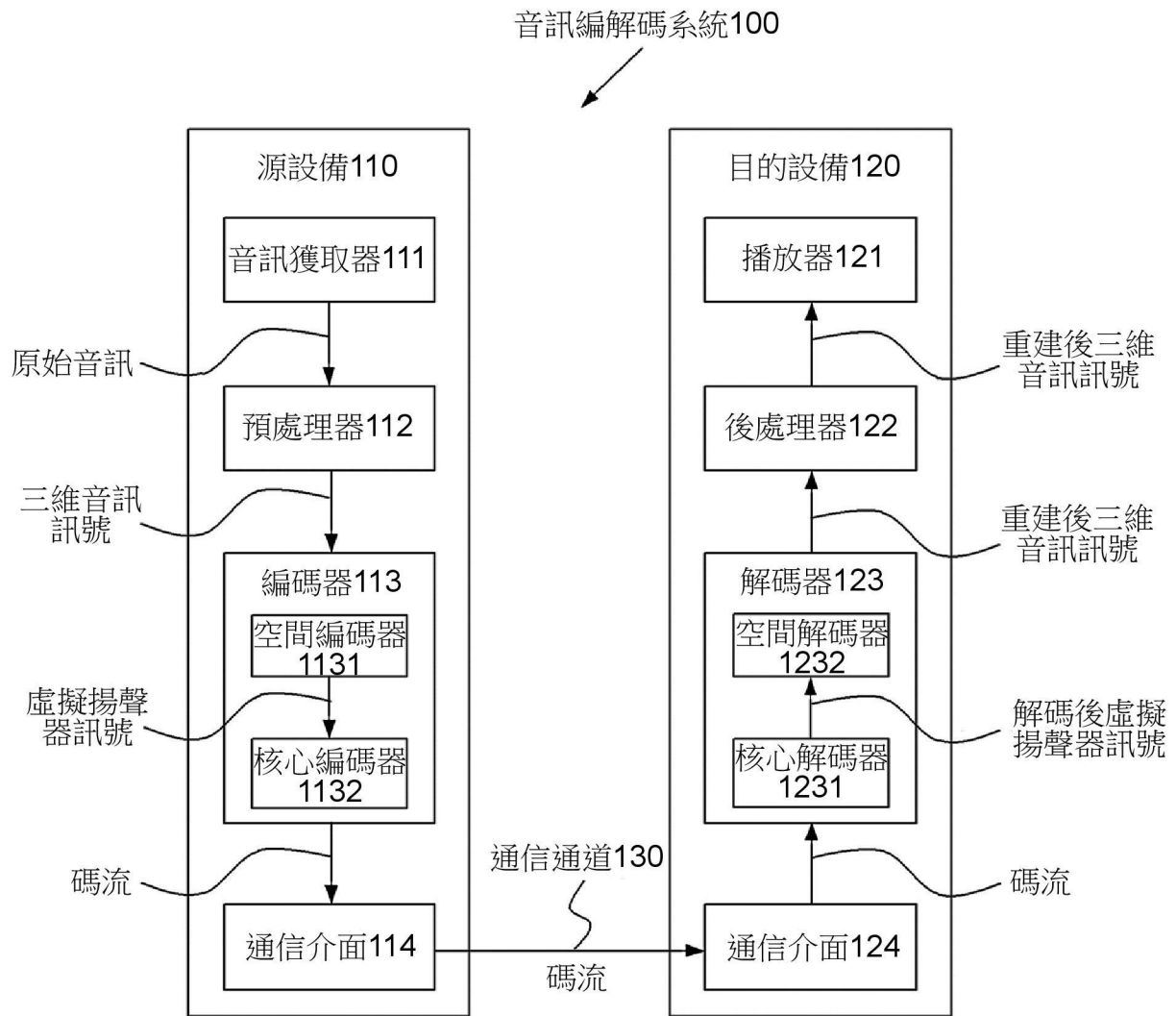
【請求項20】 一種三維音訊訊號編解碼系統，其中，所述系統包括如請求項19所述的編碼器，以及解碼器，所述編碼器用於執行上述請求項1至9中任一項所述的方法的操作步驟，所述解碼器用於解碼所述編碼器生成的碼流。

【請求項21】 一種電腦程式，其中，所述電腦程式被執行時實現如請求項1至9中任一項所述的三維音訊訊號編碼方法。

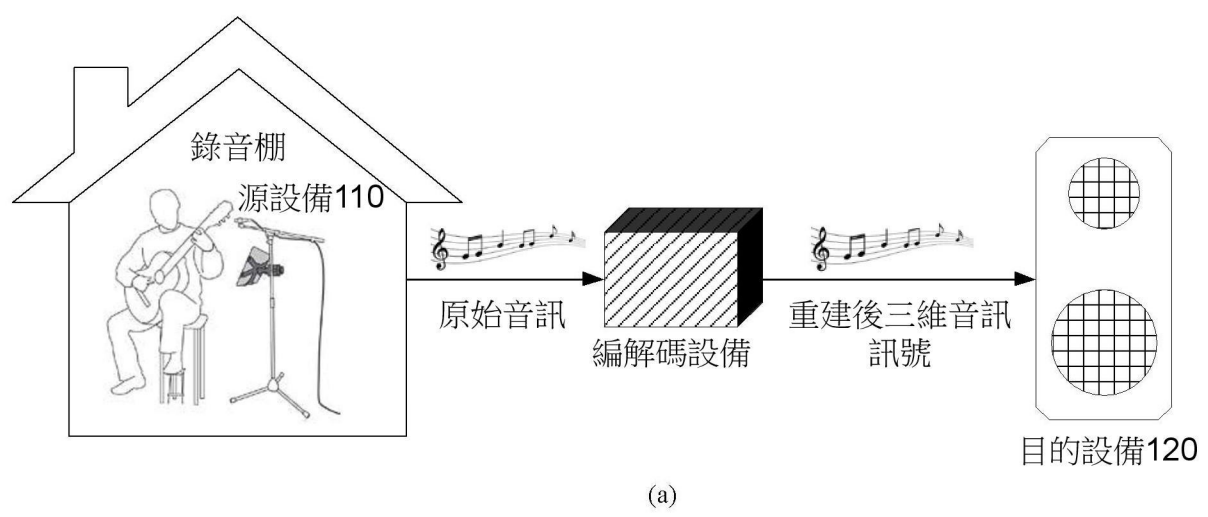
【請求項22】 一種電腦可讀儲存介質，其中，包括電腦軟體指令；當電腦軟體指令在編碼器中運行時，使得所述編碼器執行如請求項1至9中任一項所述的三維音訊訊號編碼方法。

【請求項23】 一種電腦可讀儲存介質，其中，包括如請求項1至9中任一項所述的三維音訊訊號編碼方法所獲得的碼流。

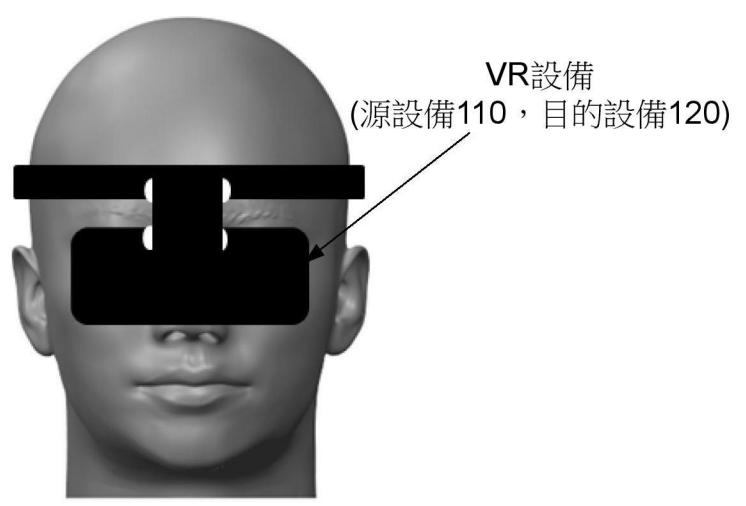
【發明圖式】



【圖1】

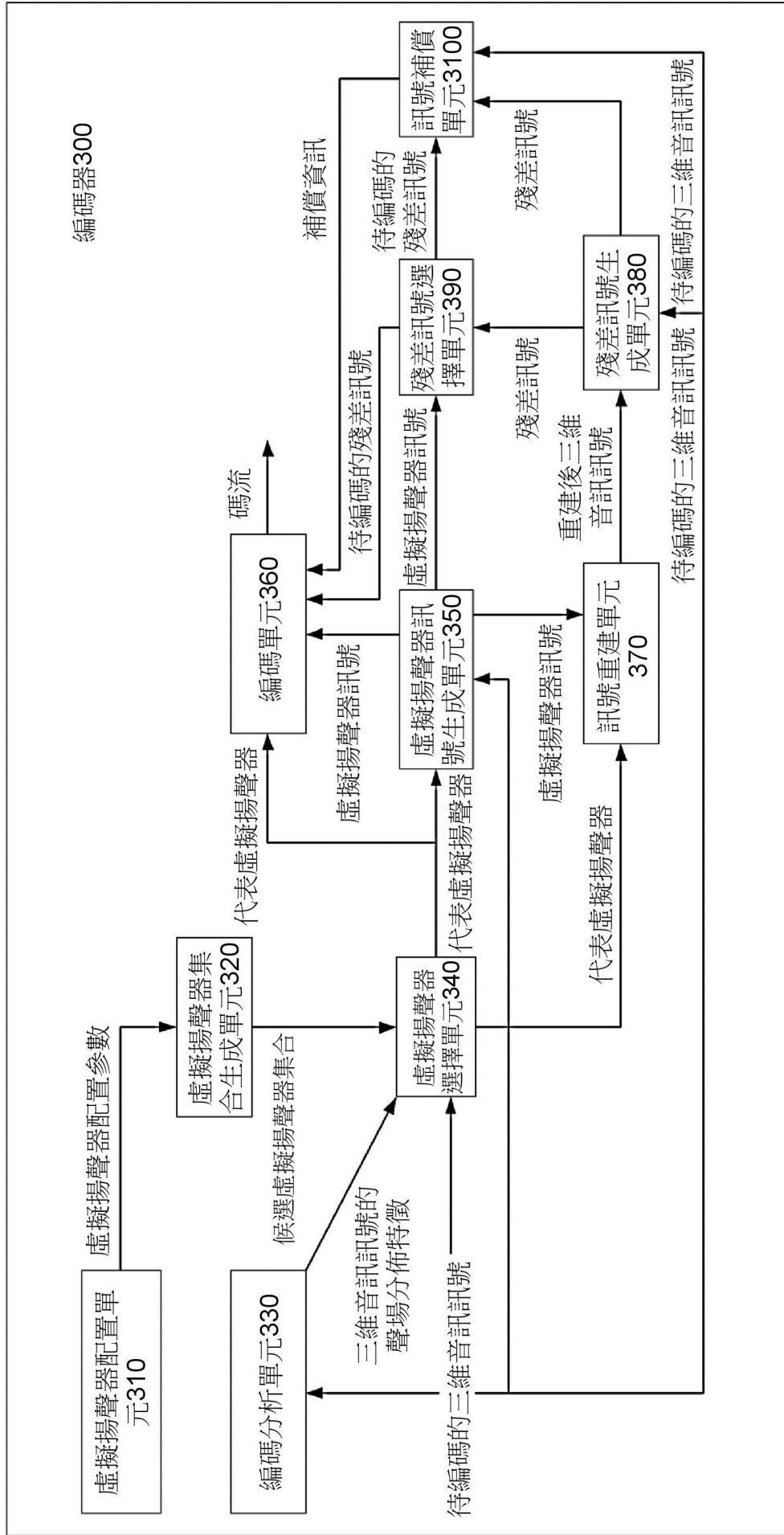


(a)

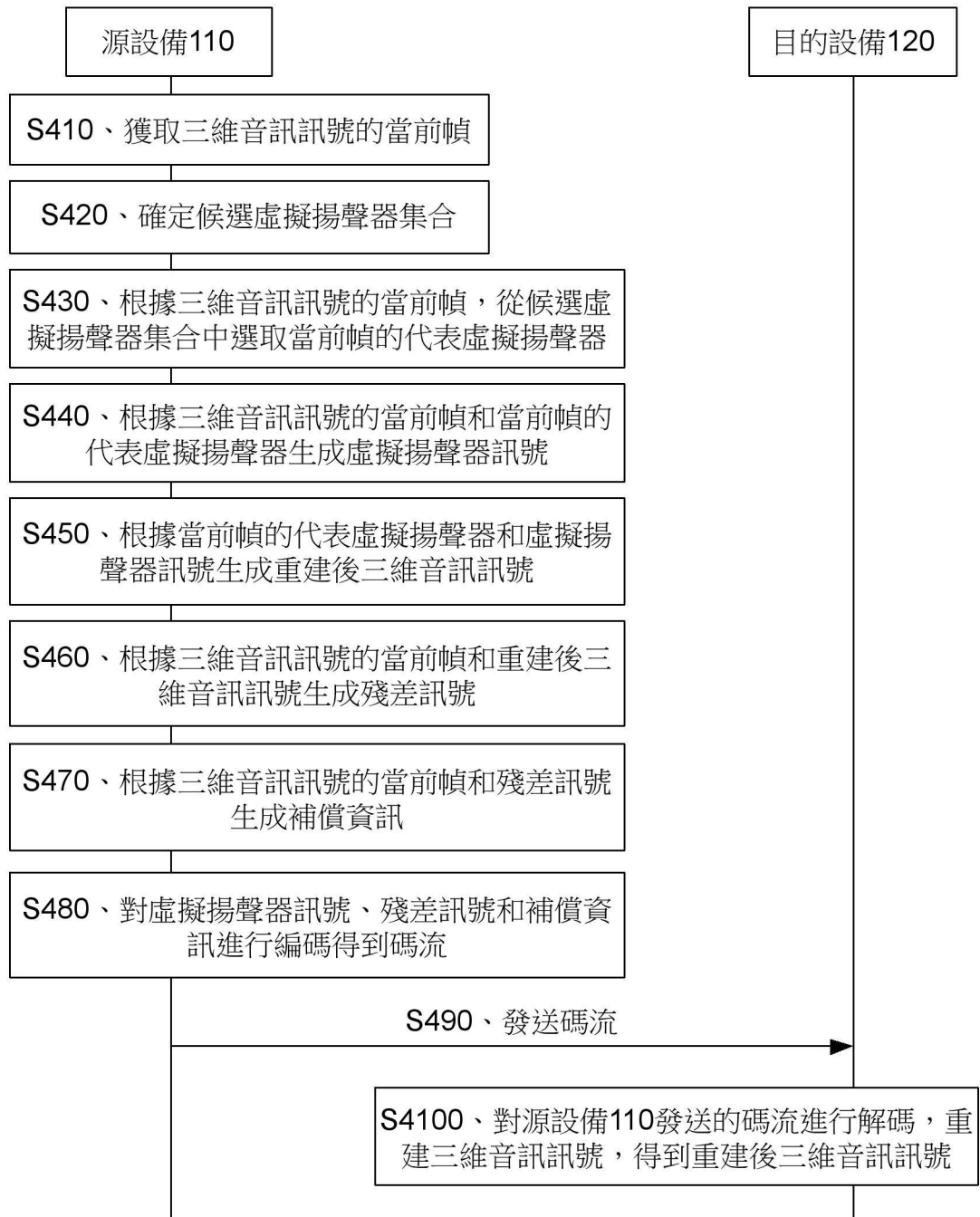


(b)

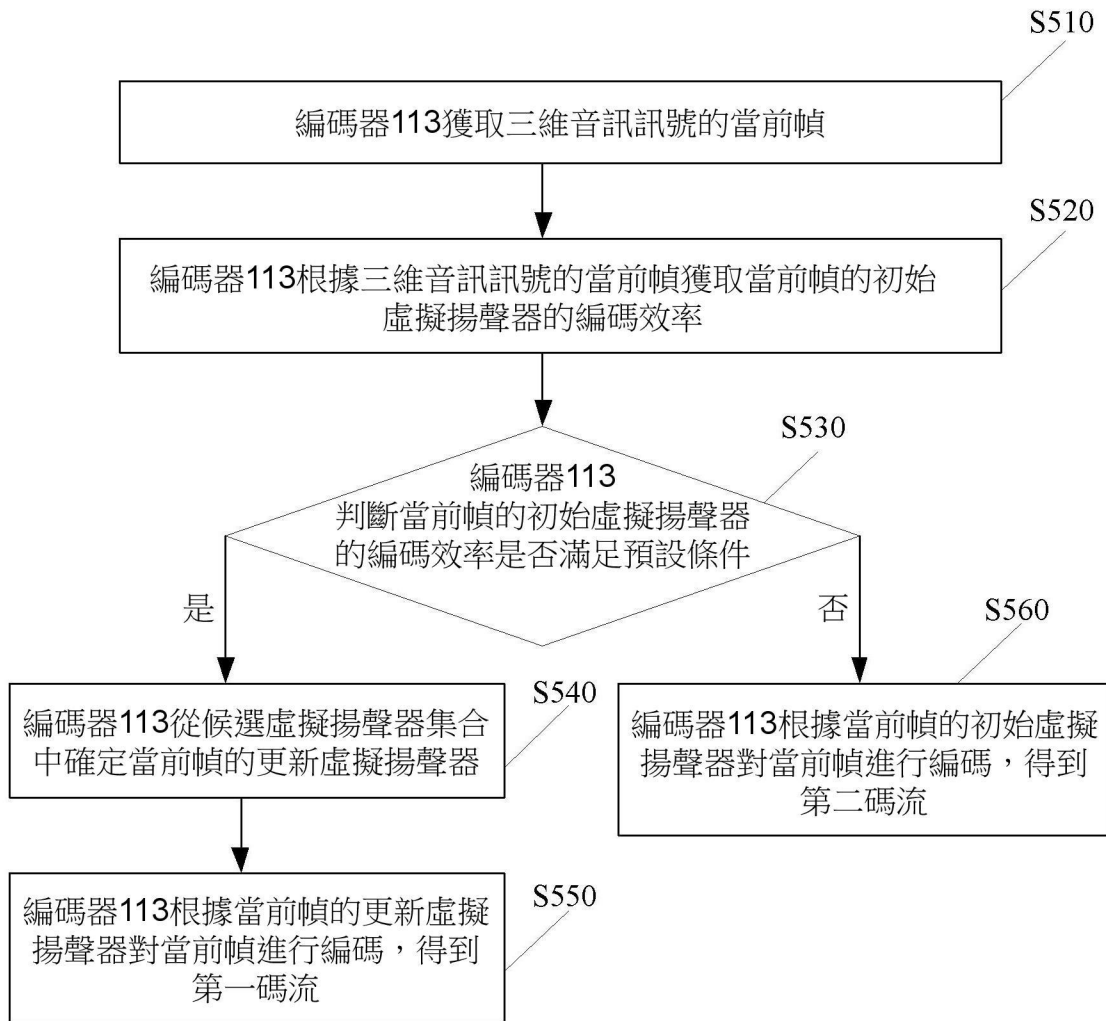
【圖2】



【圖3】

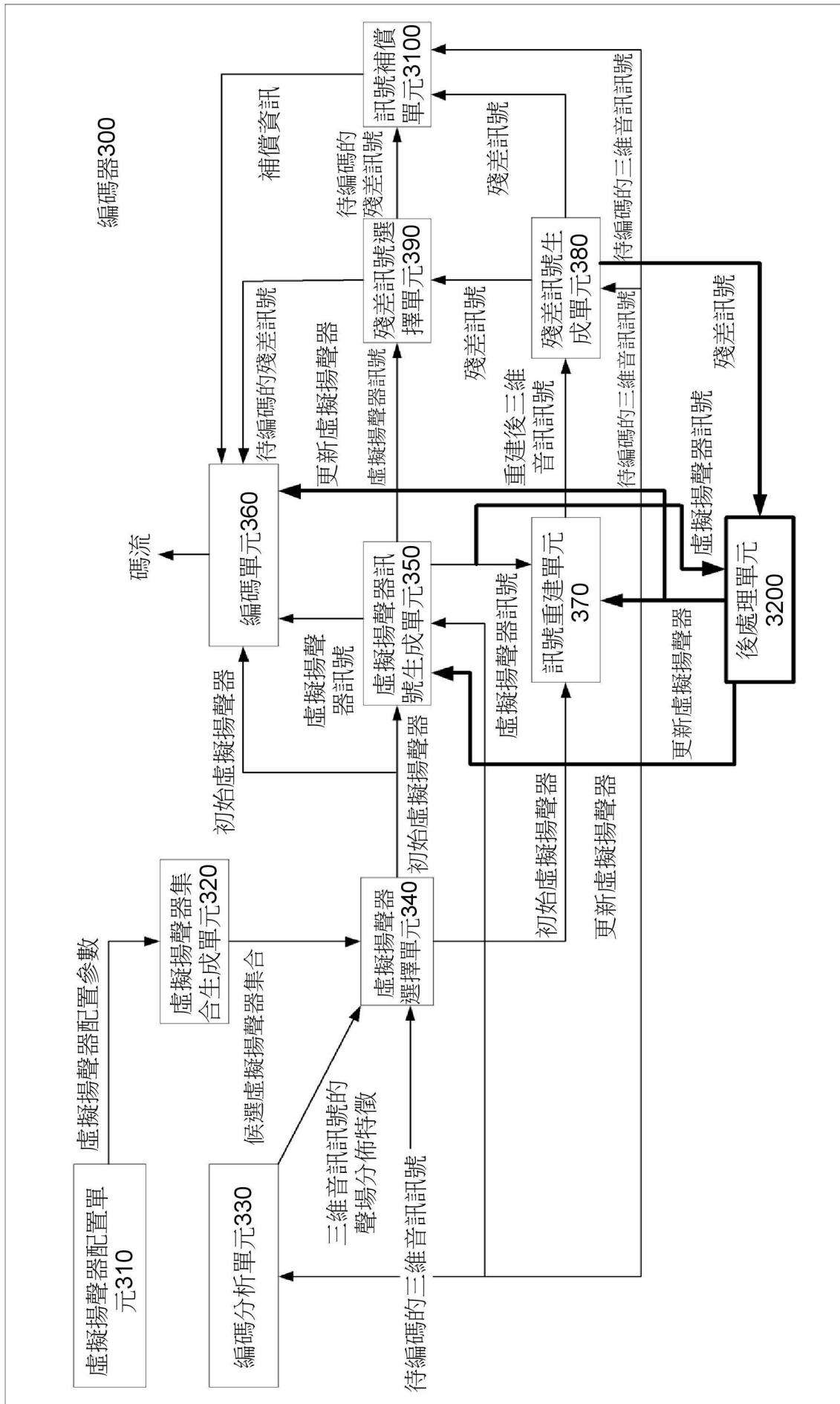


【圖4】



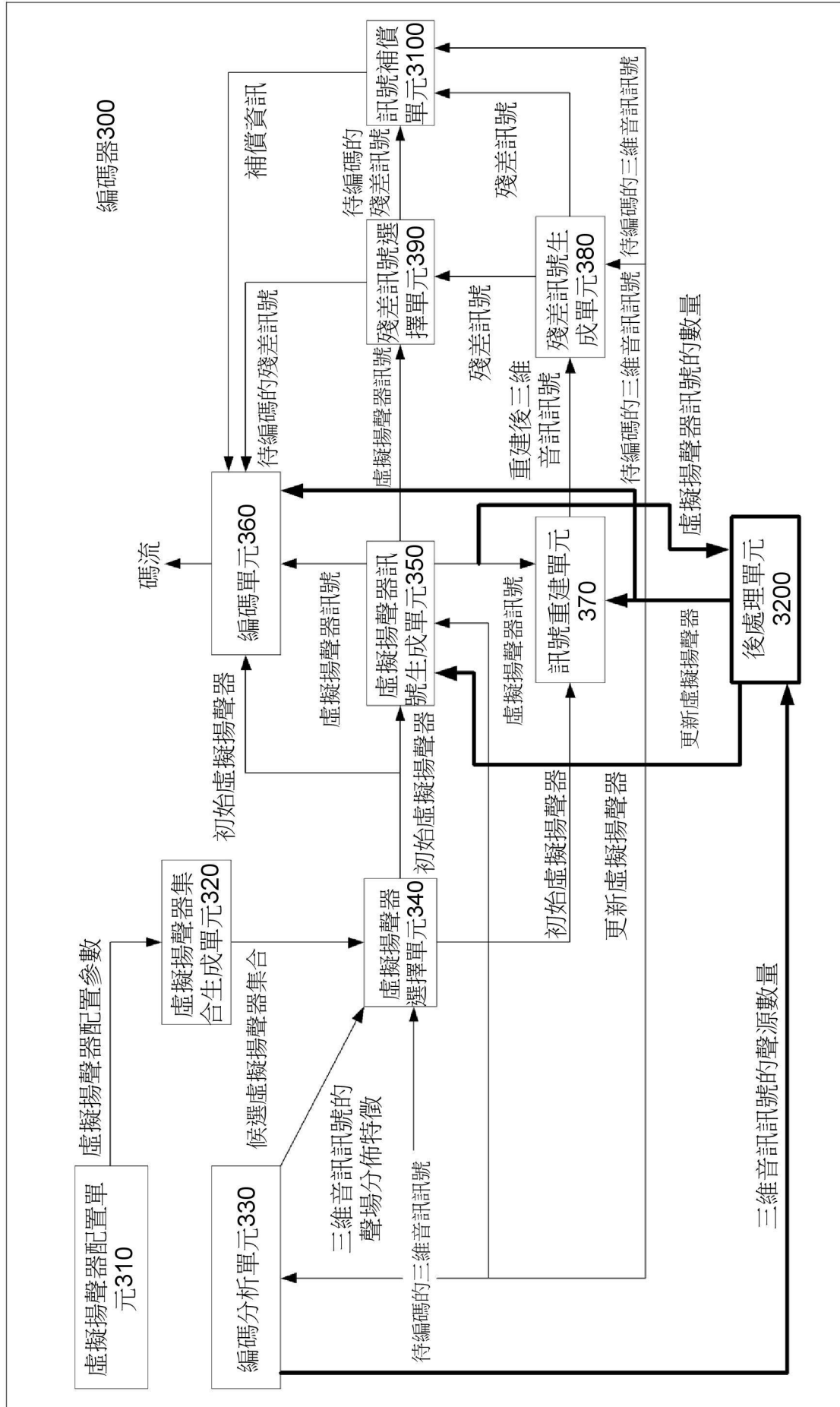
【圖5】



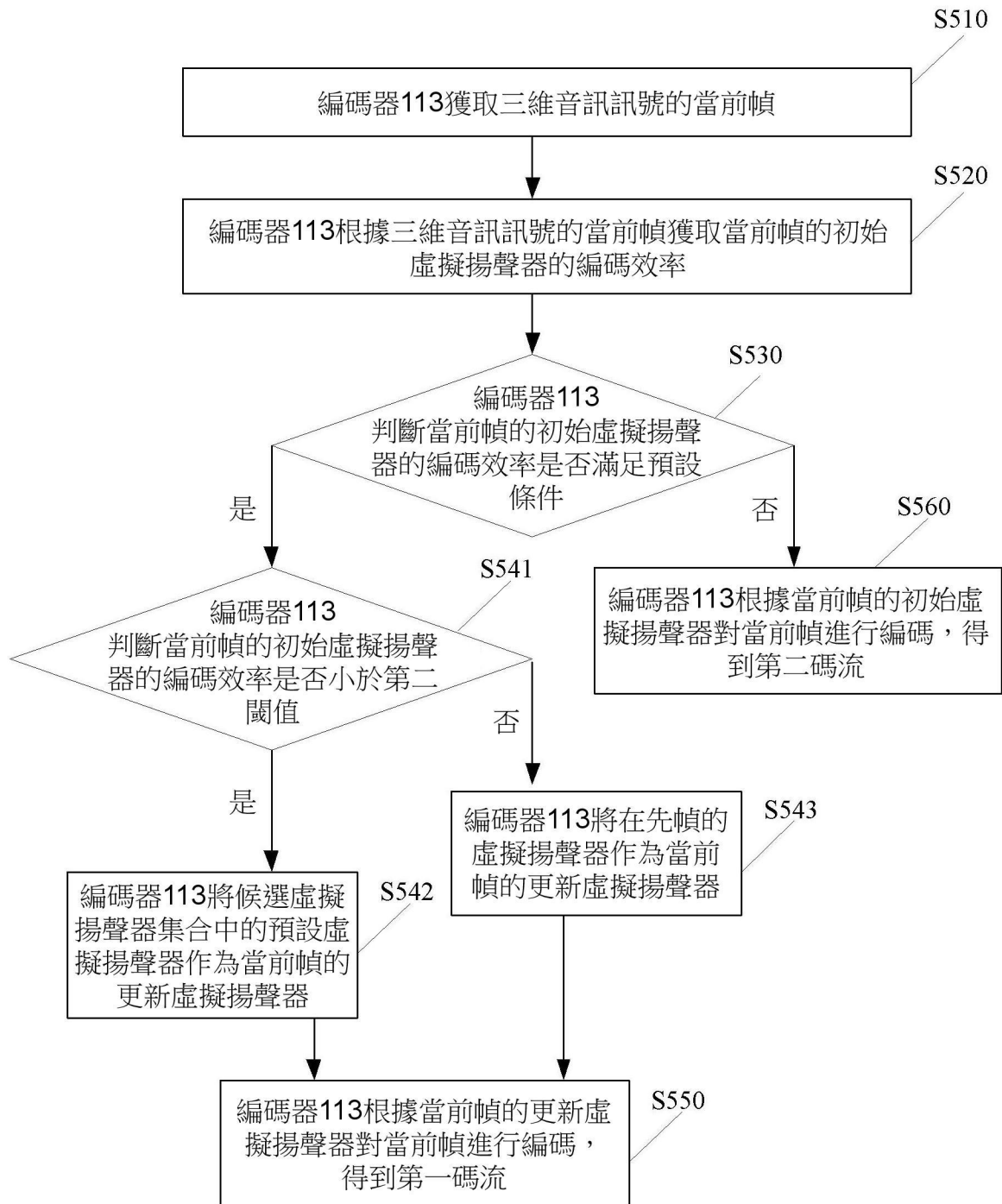


【圖7】

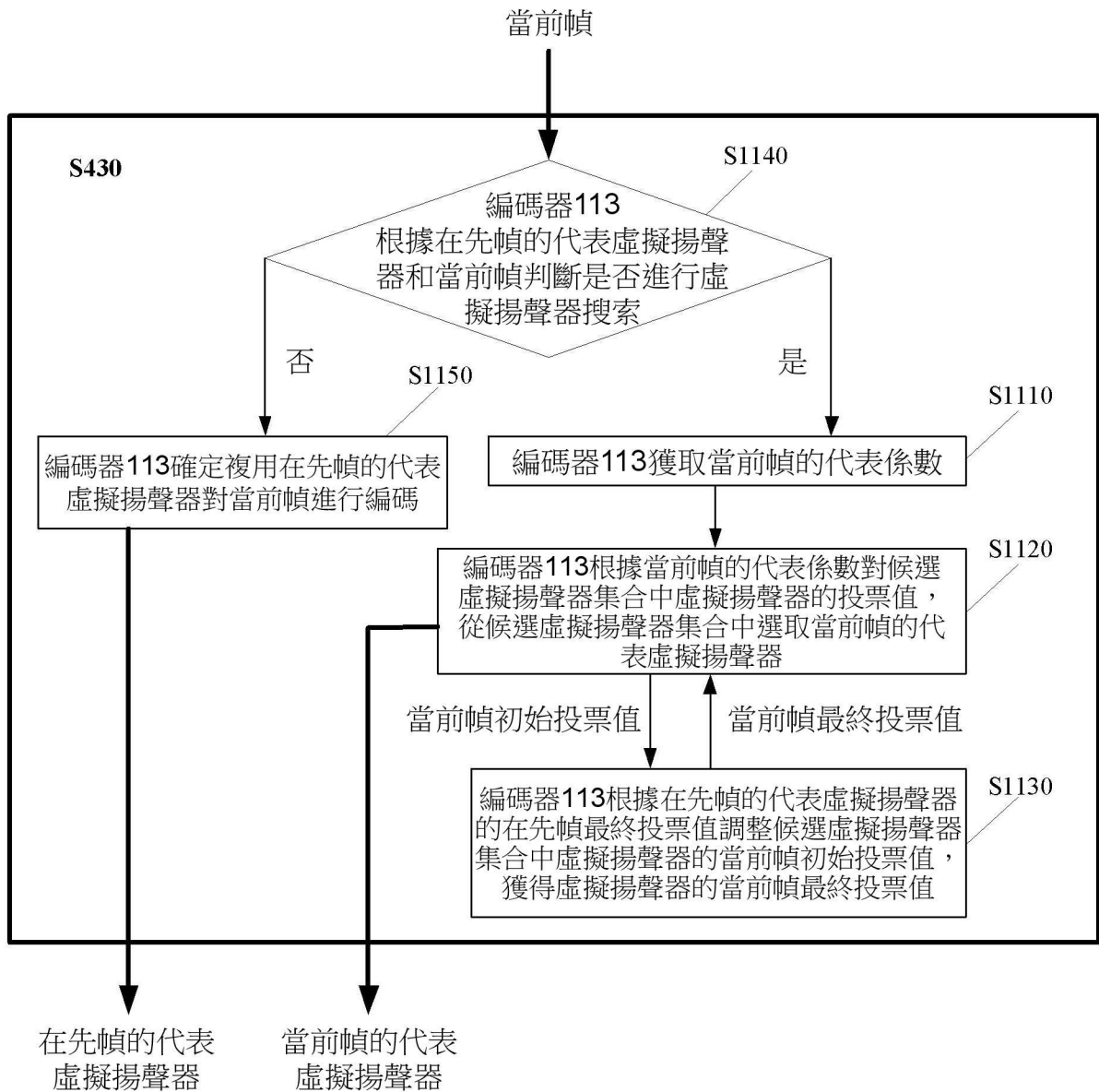




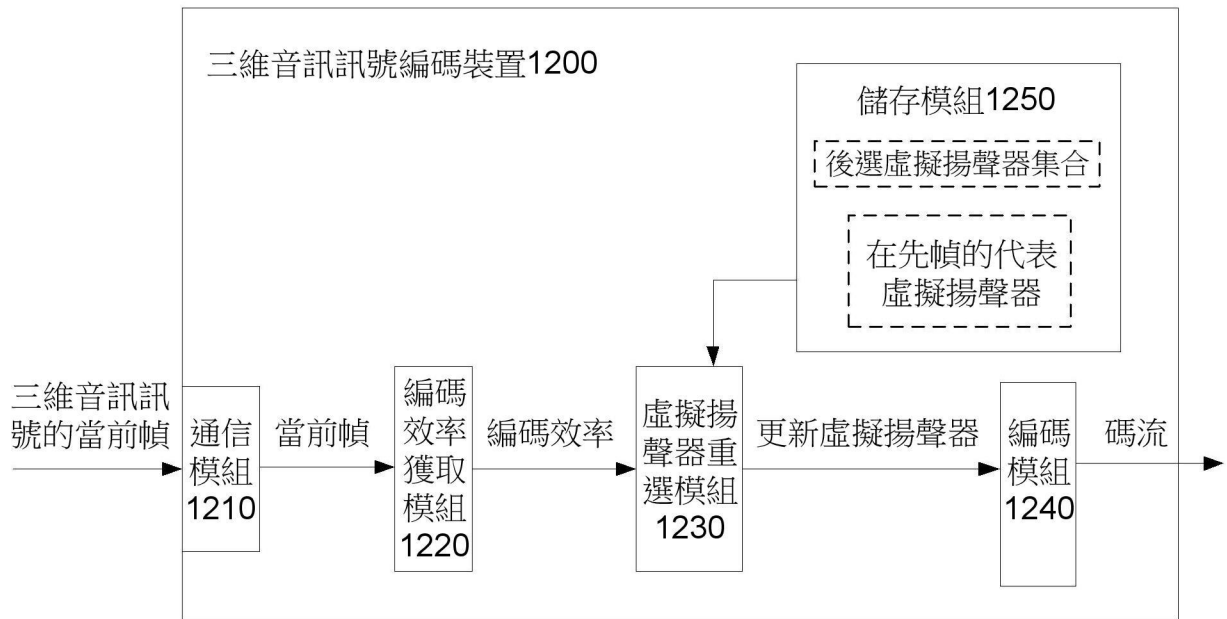
【圖9】



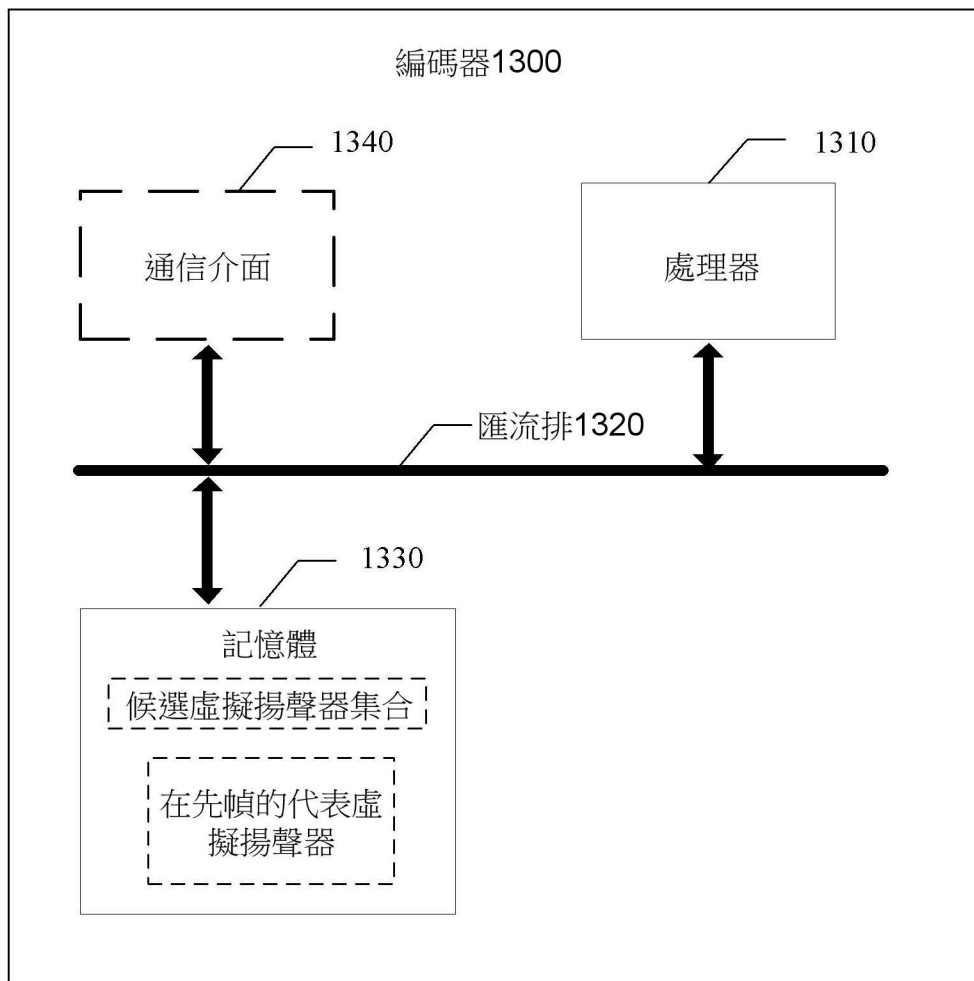
【圖10】



【圖11】



【圖12】



【圖13】