



(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 I729997 B

(45)公告日：中華民國 110 (2021) 年 06 月 11 日

(21)申請案號：105127327

(22)申請日：中華民國 105 (2016) 年 08 月 25 日

(51)Int. Cl. : H04N21/233 (2011.01)

H04N21/2368(2011.01)

H04N21/854 (2011.01)

(30)優先權：2015/08/25 美國 62/209,779

2016/08/24 美國 15/246,370

2015/08/25 美國 62/209,764

(71)申請人：美商高通公司 (美國) QUALCOMM INCORPORATED (US)

92121 美國加州聖地牙哥莫爾豪斯 5775

(72)發明人：史塔克漢莫 湯瑪士 STOCKHAMMER, THOMAS (DE) ; 森 迪潘強 SEN,

DIPANJAN (US) ; 彼得斯 尼爾斯 古恩瑟 PETERS, NILS GUNTHER (DE) ; 金

墨永 KIM, MOO YOUNG (KR)

(74)代理人：林怡芳

(56)參考文獻：

TW 201511583A

US 2007/0291837A1

US 2014/0019587A1

審查人員：陳逸

申請專利範圍項數：35 項 圖式數：19 共 108 頁

(54)名稱

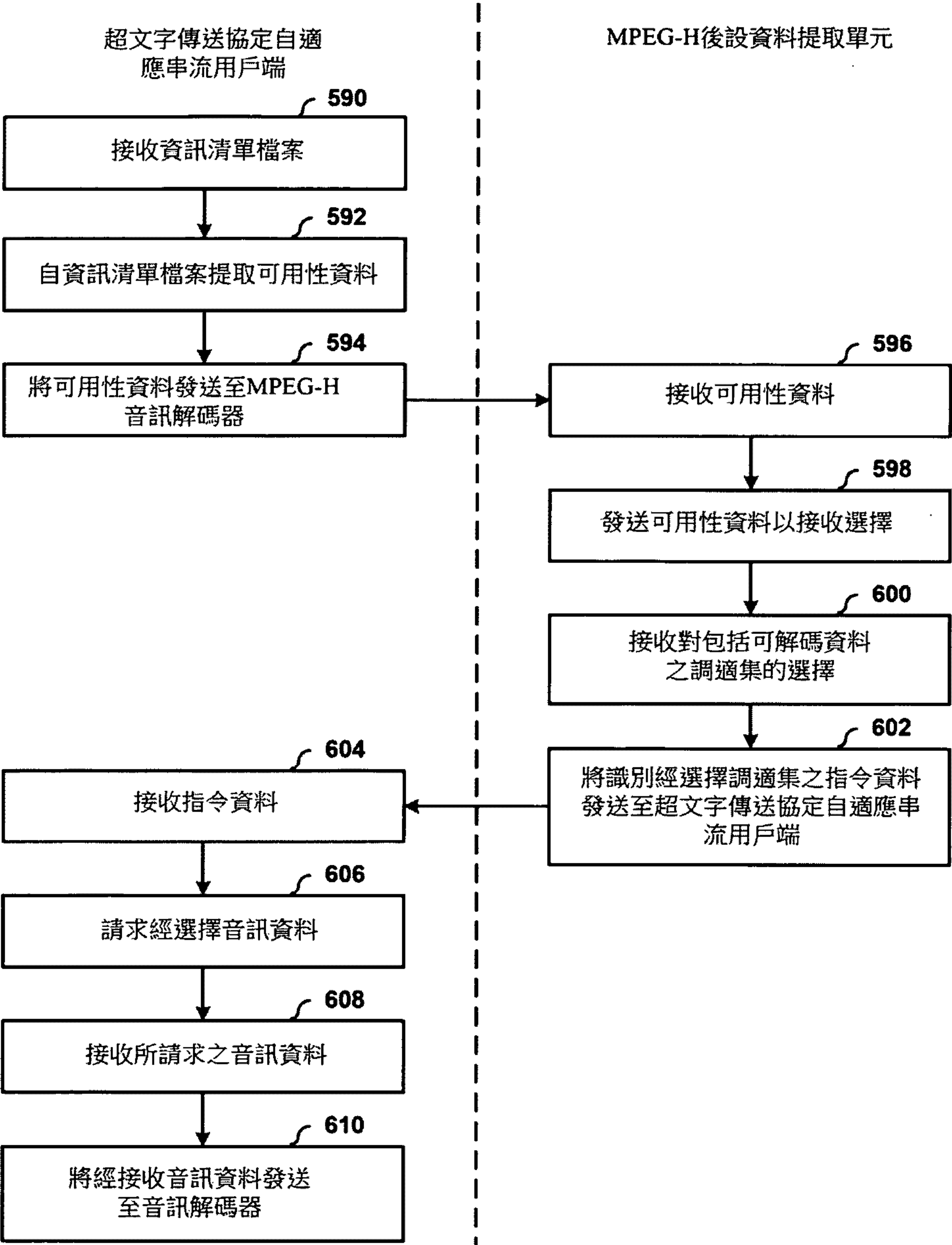
傳輸經寫碼音訊資料

(57)摘要

在一個實例中，一種用於擷取音訊資料之器件包括一或多個處理器，該一或多個處理器經組態以進行以下操作：接收表示複數個可用調適集之可用性資料，該等可用調適集包括一基於場景之音訊調適集及一或多個基於物件之音訊調適集；接收選擇資料，其識別該基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料；及一記憶體，其經組態以儲存該等音訊調適集之所擷取資料。

In one example, a device for retrieving audio data includes one or more processors configured to receive availability data representative of a plurality of available adaptation sets, the available adaptation sets including a scene-based audio adaptation set and one or more object-based audio adaptation sets, receive selection data identifying which of the scene-based audio adaptation set and the one or more object-based audio adaptation sets are to be retrieved, and provide instruction data to a streaming client to cause the streaming client to retrieve data for each of the adaptation sets identified by the selection data, and a memory configured to store the retrieved data for the audio adaptation sets.

指定代表圖：



【圖19】

符號簡單說明：

- 590 . . . 動作
- 592 . . . 動作
- 594 . . . 動作
- 596 . . . 動作
- 598 . . . 動作
- 600 . . . 動作
- 602 . . . 動作
- 604 . . . 動作
- 606 . . . 動作
- 608 . . . 動作
- 610 . . . 動作



I729997

【發明摘要】

公告本

【中文發明名稱】

傳輸經寫碼音訊資料

【英文發明名稱】

TRANSPORTING CODED AUDIO DATA

【中文】

在一個實例中，一種用於擷取音訊資料之器件包括一或多個處理器，該一或多個處理器經組態以進行以下操作：接收表示複數個可用調適集之可用性資料，該等可用調適集包括一基於場景之音訊調適集及一或多個基於物件之音訊調適集；接收選擇資料，其識別該基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料；及一記憶體，其經組態以儲存該等音訊調適集之所擷取資料。

【英文】

In one example, a device for retrieving audio data includes one or more processors configured to receive availability data representative of a plurality of available adaptation sets, the available adaptation sets including a scene-based audio adaptation set and one or more object-based audio adaptation sets, receive selection data identifying which of the scene-based audio adaptation set and the one or more object-based audio adaptation sets are to be retrieved, and provide instruction data to a streaming client to cause the streaming client to retrieve data for each of the adaptation sets identified by the selection data, and a memory configured to store the retrieved data for the audio adaptation sets.

【指定代表圖】

圖19

【代表圖之符號簡單說明】

590	動作
592	動作
594	動作
596	動作
598	動作
600	動作
602	動作
604	動作
606	動作
608	動作
610	動作

【發明說明書】

【中文發明名稱】

傳輸經寫碼音訊資料

【英文發明名稱】

TRANSPORTING CODED AUDIO DATA

【技術領域】

本發明係關於經編碼媒體資料之儲存及輸送。

【先前技術】

數位視訊能力可併入至廣泛範圍之器件中，包括數位電視、數位直播系統、無線廣播系統、個人數位助理(PDA)、膝上型或桌上型電腦、數位攝影機、數位記錄器件、數位媒體播放器、視訊遊戲器件、視訊遊戲控制台、蜂巢式或衛星無線電電話、視訊電話會議器件及其類似者。數位視訊器件實施視訊壓縮技術，諸如由MPEG-2、MPEG-4、ITU-T H.263或ITU-T H.264/MPEG-4，第10部分，進階視訊寫碼(AVC)、ITU-T H.265/高效視訊寫碼(HEVC)界定之標準及此等標準之擴展中所描述的彼等技術，以更有效地傳輸及接收數位視訊資訊。

高階立體混響(HOA)信號(其常常由複數個球諧係數(SHC)或其他階層式元素表示)為音場之三維表示。HOA或SHC表示可按獨立於用以播放自SHC信號轉譯之多聲道音訊信號的本端揚聲器幾何佈置之方式來表示音場。

在已編碼諸如音訊或視訊資料之媒體資料之後，媒體資料可經封包化以用於傳輸或儲存。可將媒體資料組譯成符合多種標準中之任一者的媒體檔案，該等標準諸如國際標準化組織(ISO)基本媒體檔案格式及其擴展。

【發明內容】

一般而言，本發明描述用於使用諸如HTTP動態自適應串流(DASH)之串流媒體輸送技術來輸送三維(3D)音訊資料的技術。舉例而言，3D音訊資料可包括一或多個HOA信號及/或球諧係數(SHC)之一或多個集合。詳言之，根據本發明之技術，各種類型之音訊資料可提供於相異調適集(例如，根據DASH)中。舉例而言，第一調適集可包括場景音訊資料，調適集之第一集合可包括聲道音訊資料，且調適集之第二集合可包括物件音訊資料。場景音訊資料可大體對應於背景雜訊。聲道音訊資料可大體對應於專用於特定聲道(例如，對於特定的對應揚聲器)之音訊資料。物件音訊資料可對應於自在三維空間中產生聲音之物件記錄的音訊資料。舉例而言，一物件可對應於一樂器、正說話之一個人或其他產生聲音之真實世界物件。

可用性資料可用以指示包括該等類型之音訊資料中之每一者的調適集，其中可根據(例如) MPEG-H 3D音訊資料格式而格式化該可用性資料。因此，諸如MPEG-H 3D音訊解碼器之專用處理單元可用以解碼可用性資料。選擇資料(例如，使用者輸入或預組態之資料)可用以選擇該等類型之音訊資料中之哪些係待擷取。接著，可指示串流用戶端(諸如DASH用戶端)擷取經選定調適集之資料。

在一個實例中，一種擷取音訊資料之方法包括：接收表示複數個可用調適集之可用性資料，該等可用調適集包括一基於場景之音訊調適集及一或多個基於物件之音訊調適集；接收選擇資料，其識別該基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該

等調適集中之每一者的資料。

在另一實例中，一種用於擷取音訊資料之器件包括一或多個處理器，該一或多個處理器經組態以進行以下操作：接收表示複數個可用調適集之可用性資料，該等可用調適集包括一基於場景之音訊調適集及一或多個基於物件之音訊調適集；接收選擇資料，其識別該基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料；及一記憶體，其經組態以儲存該等音訊調適集之所擷取資料。

在另一實例中，一種用於擷取音訊資料之器件包括：用於接收表示複數個可用調適集之可用性資料的構件，該等可用調適集包括一基於場景之音訊調適集及一或多個基於物件之音訊調適集；用於接收選擇資料的構件，該選擇資料識別該基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及用於將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料的構件。

在另一實例中，一種電腦可讀儲存媒體上儲存有指令，該等指令在執行時造成一處理器進行以下操作：接收表示複數個可用調適集之可用性資料，該等可用調適集包括一基於場景之音訊調適集及一或多個基於物件之音訊調適集；接收選擇資料，其識別該基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料。

在隨附圖式及以下描述中闡述一或多個實例之細節。其他特徵、目標及優點將自該描述及圖式以及申請專利範圍顯而易見。

【圖式簡單說明】

圖1為說明實施用於經由網路而串流傳輸媒體資料之技術之實例系統的方塊圖。

圖2為更詳細地說明擷取單元之組件之實例集合的方塊圖。

圖3A及圖3B為說明實例多媒體內容之元素的概念圖。

圖4為說明實例媒體檔案之元素的方塊圖，實例媒體檔案可對應於表示之區段。

圖5A及圖5B為說明用於輸送諸如經編碼3D音訊資料之經編碼媒體資料之實例系統的方塊圖。

圖6A及圖6B為說明單獨地串流傳輸來自基於物件之內容的各種類型之資料之另一實例的方塊圖。

圖7A至圖7C為說明根據本發明之技術之另一實例系統的方塊圖。

圖8為說明根據本發明之技術之另外實例系統的方塊圖。

圖9為根據本發明之技術的另一實例系統。

圖10為說明可使用本發明之技術之另一實例系統的概念圖。

圖11為說明可實施本發明之技術之另一實例系統的概念圖。

圖12為說明用於ATSC 3.0之實例概念性協定模型的概念圖。

圖13A及圖13B為表示多層音訊資料之實例的概念圖。

圖14A及圖14B為說明多層音訊資料之額外實例的概念圖。

圖15為說明根據本發明之技術之傳送可調式HOA資料之另一實例系統的方塊圖。

圖16為說明根據本發明之技術之實例架構的概念圖。

圖17為說明根據本發明之技術之實例用戶端器件的方塊圖。

圖18為說明用於執行本發明之技術之實例方法的流程圖。

圖19為說明用於執行本發明之技術之另一實例方法的流程圖。

【實施方式】

本申請案主張2015年8月25日申請之美國臨時申請案第62/209,779號及2015年8月25日申請之美國臨時申請案第62/209,764號的權益，該等美國臨時申請案中之每一者的整體內容特此以引用之方式併入本文中。

一般而言，本發明描述用於輸送諸如經編碼三維(3D)音訊資料之經編碼媒體資料的技術。環繞聲之演進已提供許多用於娛樂的輸出格式。此類消費型環繞聲格式之實例大部分為「聲道」式的，此係因為其隱含地指定對處於某些幾何座標的擴音器之饋入。消費型環繞聲格式包括風行的5.1格式(其包括以下六個聲道：左前(FL)、右前(FR)、中心或前中心、左後或左環繞、右後或右環繞，及低頻效應(LFE))、發展中的7.1格式、包括高度揚聲器之各種格式，諸如7.1.4格式及22.2格式(例如，用於供超高清晰度電視標準使用)。非消費型格式可涵括常常被稱為「環繞陣列」的任何數目個揚聲器(成對稱及非對稱幾何佈置)。此類陣列之一個實例包括定位於截頂二十面體(truncated icosohedron)之拐角上的座標處之32個擴音器。

至未來MPEG-H編碼器之輸入視情況為以下三種可能格式中之一者：
(i)傳統的基於聲道之音訊(如上文所論述)，其必須經由處於預先指定之位置處的擴音器播放；(ii)基於物件之音訊，其涉及用於單音訊物件的離散脈碼調變(PCM)資料，相關聯的後設資料含有單音訊物件的位置座標(以

及其他資訊)；及(iii)基於場景之音訊，其涉及使用球諧基底函數之係數(其亦被稱作「球諧係數」或SHC、「高階立體混響」或HOA及「HOA係數」)來表示音場。在「MPEG-H 3D Audio—The New Standard for Coding of Immersive Spatial Audio, Jürgen Herre, Senior Member, IEEE, Johannes Hilpert, Achim Kuntz 及 Jan Plogsties, IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, 第9卷, 第5號, 2015年8月」中更詳細地描述實例MPEG-H編碼器。

新MPEG-H 3D音訊標準針對基於聲道、物件及SCE之音訊串流中之每一者提供標準化音訊位元串流，且提供後續的可調適且與揚聲器幾何佈置(及揚聲器之數目)以及播放位置處之聲學條件無關的解碼(涉及轉譯器)。

如IEEE論文(第771頁)所指出，HOA提供較多係數信號，且因此，提供增大的空間選擇性，其允許在較小串音的情況下轉譯擴音器信號，從而產生縮減的音響度偽訊。與物件相反，未以明確的幾何後設資料形式傳達HOA之空間資訊，而以係數信號自身的形式傳達空間資訊。因此，立體混響/HOA並不十分適於允許對聲音場景中之個別物件進行存取。然而，對於內容建立者，使用階層式元素集合來表示音場存在較多靈活性。階層式元素集合可指其中元素經排序而使得低階元素之基本集合提供經模型化音場之完整表示的元素集合。在該集合經擴展以包括高階元素時，表示變得更詳細，從而增加解析度。

階層式元素集合之一個實例為球諧係數(SHC)之集合。以下陳述式示範使用SHC進行的對音場之描述或表示：

$$p_i(t, r_r, \theta_r, \varphi_r) = \sum_{\omega=0}^{\infty} \left[4\pi \sum_{n=0}^{\infty} j_n(kr_r) \sum_{m=-n}^n A_n^m(k) Y_n^m(\theta_r, \varphi_r) \right] e^{j\omega t},$$

該陳述式展示在時間 t 處，音場之任何點 $\{r, \theta, \varphi\}$ 處的壓力 P_i 可由SHC， $A_n^m(k)$ 唯一地表示。此處， $k = \frac{\omega}{c}$ ， c 為聲速(約343 m/s)， $\{r, \theta, \varphi\}$ 為參考點(或觀測點)， $j_n(\cdot)$ 為具有階數 n 之球面貝塞耳函數(spherical Bessel function)，且 $Y_n^m(\theta, \varphi)$ 為具有階數 n 及子階數 m 之球諧基底函數。可認識到，方括弧中之項為信號之頻域表示(亦即， $S(\omega, r, \theta, \varphi)$)，其可藉由各種時間-頻率變換(諸如，離散傅立葉變換(DFT)、離散餘弦變換(DCT)或小波變換)來近似。階層式集合之其他實例包括小波變換係數之集合及多解析度基底函數之係數的其他集合。

本發明之技術可用以輸送如上文所論述使用諸如HTTP動態自適應串流(DASH)之串流協定編碼的音訊資料。DASH之各種態樣在(例如)「資訊技術—HTTP動態自適應串流(DASH)—第1部分：媒體呈現描述及區段格式(Information Technology—Dynamic Adaptive Streaming over HTTP (DASH)—Part 1: Media Presentation Description and Segment Formats)」，ISO/IEC 23089-1，2012年4月1日；及第三代合作夥伴計劃(3rd Generation Partnership Project)；技術規範群組服務及系統態樣(Technical Specification Group Services and System Aspects)；透明端對端封包交換串流服務(PSS) (Transparent end-to-end Packet-switched Streaming Service (PSS))；漸進下載及HTTP動態自適應串流(Progressive Download and Dynamic Adaptive Streaming over HTTP)(3GP-DASH)(版本12) 3GPP TS 26.247, V12.1.0，2013年12月中描述。

在HTTP串流中，頻繁使用之操作包括HEAD、GET及部分GET。HEAD操作擷取與給定的統一資源定位符(URL)或統一資源名稱(URN)相

關聯之檔案的標頭，但不擷取與URL或URN相關聯之有效負載。**GET**操作擷取與給定URL或URN相關聯之整個檔案。部分**GET**操作接收位元組範圍作為輸入參數且擷取檔案之連續數目個位元組，其中位元組之數目對應於接收到的位元組範圍。因此，**HTTP**串流可提供電影片段，此係因為部分**GET**操作可得到一或多個個別的電影片段。在電影片段中，可能存在不同播放軌之若干播放軌片段。在**HTTP**串流中，媒體呈現可為用戶端可存取之資料之結構化集合。用戶端可請求且下載媒體資料資訊以向使用者提供串流服務。

在使用**HTTP**串流來串流**3GPP**資料之實例中，可能存在多媒體內容之視訊及/或音訊資料的多個表示。

如下文所解釋，不同表示可對應於針對**HoA**，亦即，基於場景之音訊之可調式寫碼的不同形式。

可在媒體呈現描述(**MPD**)資料結構中定義此等表示之資訊清單。媒體呈現可對應於**HTTP**串流用戶端器件可存取之資料之結構化集合。**HTTP**串流用戶端器件可請求且下載媒體資料資訊以向用戶端器件之使用者呈現串流服務。媒體呈現可在**MPD**資料結構中描述，**MPD**資料結構可包括**MPD**之更新。

媒體呈現可含有一或多個週期之序列。週期可由**MPD**中之*Period*元素來定義。每一週期可具有**MPD**中之屬性*start*。對於每一週期，**MPD**可包括*start*屬性及*availableStartTime*屬性。對於實況服務，週期之*start*屬性與**MPD**屬性*availableStartTime*之總和可按UTC格式指定週期之可用性時間，詳言之，對應週期中之每一表示的第一媒體區段。對於點播服務，第一週期之*start*屬性可為0。對於任何其他週期，*start*屬性可指定對應週

期之開始時間相對於第一週期之開始時間的時間偏移。每一週期可延長，直至下一週期開始為止，或在最後一個週期的狀況下，直至媒體呈現結束為止。週期開始時間可為精確的。週期開始時間可反映由播放所有先前週期之媒體產生的實際時序。

每一週期可含有同一媒體內容之一或多個表示。表示可為音訊或視訊資料之數個替代的經編碼版本中之一者。表示可因編碼類型而異(例如，對於視訊資料，因位元速率、解析度及/或編碼解碼器而異，及對於音訊資料，因位元速率、語言及/或編碼解碼器而異)。術語表示可用以指經編碼音訊或視訊資料的對應於多媒體內容之特定週期且以特定方式編碼之部分。

特定週期之表示可指派至由MPD中之屬性(其指示表示所屬之調適集)指示之群組。同一調適集中之表示通常被視為彼此之替代例，此係因為用戶端器件可在此等表示之間動態地且順暢地切換，例如以執行頻寬調適。舉例而言，特定週期之視訊資料之每一表示可指派至同一調適集，以使得可選擇該等表示中之任一者進行解碼以呈現對應週期之多媒體內容的媒體資料(諸如視訊資料或音訊資料)。作為另一實例，音訊調適集之表示可包括同一類型之音訊資料，其以不同位元速率經編碼以支援頻寬調適。在一些實例中，一個週期內之媒體內容可由來自群組0 (若存在)之一個表示或由來自每一非零群組的至多一個表示之組合來表示。週期之每一表示之時序資料可相對於該週期之開始時間來表達。

一表示可包括一或多個區段。每一表示可包括初始化區段，或表示之每一區段可自初始化。當存在時，初始化區段可含有用於存取表示之初始化資訊。一般而言，初始化區段不含有媒體資料。區段可由識別符唯一地

參考，諸如統一資源定位符(URL)、統一資源名稱(URN)或統一資源識別符(URI)。MPD可提供每一區段之識別符。在一些實例中，MPD亦可提供呈*range*屬性之形式的位元組範圍，該屬性可對應於可藉由URL、URN或URI存取之檔案內之區段的資料。

可選擇不同表示以用於大體上同時擷取不同類型之媒體資料。舉例而言，用戶端器件可選擇音訊表示、視訊表示及計時文字表示，自該等表示擷取區段。在一些實例中，用戶端器件可選擇特定調適集以用於執行頻寬調適。即，用戶端器件可選擇包括視訊表示之視訊調適集、包括音訊表示之調適集及/或包括計時文字之調適集。

本發明之技術可用以將媒體(例如，3D音訊)資料多工至(例如)MPEG-2系統中，MPEG-2系統在「資訊技術—移動圖像及相關聯音訊資訊之通用寫碼—第1部分：系統(Information technology—Generic coding of moving pictures and associated audio information—Part 1: Systems)」，ISO/IEC 13818-1:2013 (亦為ISO/IEC 13818-1:2015)中描述。該系統規範藉由存取單元描述串流/播放軌，每一存取單元具有時間戳記。對存取單元進行多工，且通常存在可如何執行此多工之某一靈活性。MPEG-H音訊准許將所有物件之樣本置放於一個串流中，例如，可將具有相同時間碼之所有樣本映射至一個存取單元中。在系統層級處，有可能產生一個主串流及多個補充串流，其允許將物件分離至不同系統串流中。系統串流產生靈活性：其允許不同傳遞路徑、混合傳遞、根本不傳遞及其類似者。

包括媒體資料(例如，音訊及/或視訊資料)之檔案可根據在(例如)「資訊技術—音訊-視覺物件之寫碼—第12部分：ISO基本媒體檔案格式(Information technology-Coding of audio-visual objects—Part 12: ISO

base media file format)」，ISO/IEC 14496-12:2012中描述的ISO基本媒體檔案格式(BMFF)形成。在ISO BMFF中，串流為播放軌一存取單元含於電影資料(mdat)方塊中。每一播放軌得到電影標頭中之樣本條目，且可實體上找到描述樣本之樣本表。藉由使用電影片段，分佈式儲存亦係可能的。

在MPEG-2輸送串流(TS)中，串流為基本串流。MPEG-2 TS中存在較小靈活性，但一般而言，該等技術相似於ISO BMFF。儘管含有媒體資料(例如，經編碼3D音訊資料)之檔案可根據上文所論述之各種技術中之任一者形成，但本發明描述關於ISO BMFF/檔案格式之技術。因此，3D音訊資料(例如，場景音訊資料、物件音訊資料及/或聲道音訊資料)可根據MPEG-H 3D音訊編碼且根據(例如) ISO BMFF囊封。相似地，可用性資料可根據MPEG-H 3D音訊編碼。因此，與DASH用戶端分離之單元或器件(諸如MPEG-H 3D音訊解碼器)可解碼可用性資料且判定該等調適集中之哪些係待擷取，接著將指令資料發送至DASH用戶端以造成DASH用戶端擷取經選定調適集之資料。

一般而言，檔案可含有經編碼媒體資料，諸如經編碼3D音訊資料。在DASH中，此等檔案可被稱作表示之「區段」，如上文所論述。此外，內容提供者可使用各種調適集來提供媒體內容，如上文所提及。關於3D音訊資料，場景音訊資料可提供於一個調適集中。此調適集可包括場景音訊資料之多種可切換(即，替代)表示(例如，在位元速率上彼此相異，但在其他方面大體上相同)。相似地，音訊物件可各自提供於各別調適集中。替代地，調適集可包括多個音訊物件，及/或一或多個音訊物件可提供於多個調適集中。

根據本發明之技術，用戶端器件(例如，使用者設備，「UE」)可包括MPEG-H音訊解碼器或經組態以解碼及剖析音訊後設資料(其可根據MPEG-H 3D音訊標準而格式化)之其他單元。音訊後設資料可包括對可用調適集(包括一或多個場景調適集及一或多個音訊物件調適集)之描述。更特定而言，音訊後設資料可包括場景及/或物件音訊資料與包括該場景/物件音訊資料之調適集之間的映射。此類後設資料可在本文中被稱作可用性資料。

音訊解碼器(或其他單元)可進一步自使用者介面接收選擇資料。使用者可選擇需要輸出場景物件及/或音訊物件中之哪些。替代地，使用者可選擇音訊設定檔(例如，「電影」、「音樂會」、「視訊遊戲」等等)，且使用者介面(或其他單元)可經組態以判定場景物件及音訊物件中之哪些對應於經選定音訊設定檔。

音訊解碼器(或其他單元)可基於選擇資料及可用性資料而判定該等調適集中之哪些係待擷取。音訊解碼器可接著將指令資料提供至(例如)用戶端器件之DASH用戶端。指令資料可指示該等調適集中之哪些係待擷取，或更特定而言，從調適集中之哪些擷取資料。DASH用戶端可接著選擇經選定調適集之表示，且因此自經選定表示擷取區段(例如，使用HTTP GET或部分GET請求)。

以此方式，DASH用戶端可接收可用性資料及音訊資料兩者。然而，可根據與音訊資料不同之格式(例如，呈MPEG-H 3D音訊格式而非ISO BMFF)而格式化可用性資料。亦可以與其他後設資料(諸如媒體呈現描述(MPD)或可包括可用性資料之其他資訊清單檔案的資料)不同之方式而格式化可用性資料。因此，DASH用戶端可能夠正確地剖析及解譯可用性

資料。因此，MPEG-H 3D音訊解碼器(或與DASH用戶端分離之其他單元或器件)可解碼可用性資料，且將指示將從調適集中之哪些者擷取音訊資料的指令資料提供至DASH用戶端。當然，DASH用戶端亦可自視訊調適集擷取視訊資料，及/或擷取諸如計時文字資料之其他媒體資料。藉由自分離單元或器件接收此指令資料，DASH用戶端能夠選擇適當的調適集及自經選定適當的調適集擷取媒體資料。

圖1為說明實施用於經由網路而串流傳輸媒體資料之技術之實例系統10的方塊圖。在此實例中，系統10包括內容準備器件20、伺服器器件60及用戶端器件40。用戶端器件40與伺服器器件60係藉由網路74以通信方式耦接，該網路74可包含網際網路。在一些實例中，內容準備器件20與伺服器器件60亦可藉由網路74或另一網路耦接，或可直接以通信方式耦接。在一些實例中，內容準備器件20與伺服器器件60可包含相同器件。

在圖1之實例中，內容準備器件20包含音訊源22及視訊源24。音訊源22可包含(例如)麥克風，其產生表示待藉由音訊編碼器26編碼之所俘獲音訊資料的電信號。替代地，音訊源22可包含儲存媒體(其儲存先前記錄之音訊資料)、音訊資料產生器(諸如電腦化之合成器)或任何其他音訊資料源。視訊源24可包含：視訊攝影機，其產生待藉由視訊編碼器28編碼之視訊資料；儲存媒體，其編碼有先前記錄之視訊資料；視訊資料產生單元，諸如電腦圖形源；或任何其他視訊資料源。內容準備器件20未必在所有實例中均以通信方式耦接至伺服器器件60，而可將多媒體內容儲存至由伺服器器件60讀取之分離媒體。

原始音訊及視訊資料可包含類比或數位資料。類比資料在藉由音訊編碼器26及/或視訊編碼器28編碼之前可被數位化。音訊源22可在說話參與

者正在說話時自說話參與者獲得音訊資料，且視訊源24可同時獲得說話參與者之視訊資料。在其他實例中，音訊源22可包含含有所儲存之音訊資料的電腦可讀儲存媒體，且視訊源24可包含含有所儲存之視訊資料的電腦可讀儲存媒體。以此方式，本發明中所描述之技術可應用於實況、串流、即時音訊及視訊資料或所存檔的、預先記錄的音訊及視訊資料。

對應於視訊訊框之音訊訊框通常為含有藉由音訊源22俘獲(或產生)(伴隨含於視訊訊框內的藉由視訊源24俘獲(或產生)之視訊資料)之音訊資料的音訊訊框。舉例而言，當說話參與者通常藉由說話而產生音訊資料時，音訊源22俘獲音訊資料，且視訊源24同時(即，在音訊源22正俘獲音訊資料的同時)俘獲說話參與者之視訊資料。因此，音訊訊框在時間上可對應於一或多個特定視訊訊框。因此，對應於視訊訊框之音訊訊框通常對應於一種情形，其中音訊資料及視訊資料同時被俘獲(或將以其他方式同時呈現)，且對於該情形，音訊訊框及視訊訊框分別包含已同時俘獲之音訊資料及視訊資料。另外，可單獨地產生音訊資料，其將與視訊及其他音訊資料(例如，旁白)同時呈現。

在一些實例中，音訊編碼器26可對每一經編碼音訊訊框中表示記錄經編碼音訊訊框的音訊資料之時間的時戳進行編碼，且類似地，視訊編碼器28可對每一經編碼視訊訊框中表示記錄經編碼視訊訊框的視訊資料之時間的時戳進行編碼。在此等實例中，對應於視訊訊框之音訊訊框可包含：包含時戳之音訊訊框及包含相同時戳之視訊訊框。內容準備器件20可包括內部時脈，音訊編碼器26及/或視訊編碼器28可根據該內部時脈產生時戳，或音訊源22及視訊源24可使用該內部時脈以分別使音訊資料及視訊資料與時戳相關聯。

在一些實例中，音訊源22可向音訊編碼器26發送對應於記錄音訊資料之時間的資料，且視訊源24可向視訊編碼器28發送對應於記錄視訊資料之時間的資料。在一些實例中，音訊編碼器26可對經編碼音訊資料中之序列識別符進行編碼以指示經編碼音訊資料之相對時間排序，但未必指示記錄音訊資料之絕對時間，且相似地，視訊編碼器28亦可使用序列識別符來指示經編碼視訊資料之相對時間排序。相似地，在一些實例中，序列識別符可映射或以其它方式與時戳相關聯。

音訊編碼器26通常產生經編碼音訊資料之串流，而視訊編碼器28產生經編碼視訊資料之串流。每一個別資料串流(不論音訊或視訊)可被稱作基本串流。基本串流為表示之單個經數位寫碼(可能經壓縮)之分量。舉例而言，表示之經寫碼視訊或音訊部分可為基本串流。基本串流可在被囊封於視訊檔案內之前被轉換成封包化基本串流(PES)。在相同表示內，可使用串流ID來區分屬於一個基本串流之PES封包與其他PES封包。基本串流之資料之基本單元為封包化基本串流(PES)封包。因此，經寫碼視訊資料大體對應於基本視訊串流。相似地，音訊資料對應於一或多個各別基本串流。

在圖1之實例中，內容準備器件20之囊封單元30自視訊編碼器28接收包含經寫碼視訊資料之基本串流，且自音訊編碼器26接收包含經寫碼音訊資料之基本串流。在一些實例中，視訊編碼器28及音訊編碼器26可各自包括用於自經編碼資料形成PES封包的封包化器。在其他實例中，視訊編碼器28及音訊編碼器26可各自與用於自經編碼資料形成PES封包之各別封包化器介接。在另外其他實例中，囊封單元30可包括用於自經編碼音訊及視訊資料形成PES封包之封包化器。

視訊編碼器28可以多種方式對多媒體內容之視訊資料進行編碼，以便產生多媒體內容之各種位元速率且具有各種特性之不同表示，該等特性諸如像素解析度、訊框速率、對各種寫碼標準之符合性、對各種寫碼標準之各種設定檔及/或設定檔層級之符合性、具有一或多個視圖之表示(例如，對於二維或三維播放)或其他此等特性。相似地，音訊編碼器26可以具有各種特性的多種不同方式編碼音訊資料。如下文更詳細地論述，舉例而言，音訊編碼器26可形成音訊調適集，其各自包括基於場景之音訊資料、基於聲道之音訊資料及/或基於物件之音訊資料中之一或多者。另外或在替代例中，音訊編碼器26可形成包括可調式音訊資料之調適集。舉例而言，音訊編碼器26可針對基層、左側/右側資訊及高度資訊形成調適集，如下文更詳細地論述。

如本發明中所使用，表示可包含音訊資料、視訊資料、文字資料(例如，用於隱藏式輔助字幕)或其他此資料中之一者。表示可包括諸如音訊基本串流或視訊基本串流之基本串流。每一PES封包可包括stream_id，其識別PES封包所屬之基本串流。囊封單元30負責將基本串流組譯成各種表示之視訊檔案(例如，區段)。囊封單元30自音訊編碼器26及視訊編碼器28接收表示之基本串流的PES封包，且自該等PES封包形成對應的網路抽象層(NAL)單元。

囊封單元30可向輸出介面32提供多媒體內容之一或多個表示的資料以及資訊清單檔案(例如，MPD)。輸出介面32可包含網路介面或用於對儲存媒體進行寫入之介面，諸如通用串列匯流排(USB)介面、CD或DVD寫入器或燒錄器、至磁性或快閃儲存媒體之介面，或用於儲存或傳輸媒體資料之其他介面。囊封單元30可向輸出介面32提供多媒體內容之表示中之

每一者的資料，該輸出介面可經由網路傳輸或儲存媒體向伺服器器件60發送該資料。在圖1之實例中，伺服器器件60包括儲存各種多媒體內容64之儲存媒體62，每一多媒體內容64包括各別資訊清單檔案66及一或多個表示68A至68N (表示68)。在一些實例中，輸出介面32亦可將資料直接發送至網路74。

在一些實例中，表示68可分成若干調適集。即，表示68之各種子集可包括各別共同特性集合，諸如編碼解碼器、設定檔及層級、解析度、視圖數目、區段之檔案格式、可識別待與待解碼及呈現之表示及/或音訊資料(例如，由揚聲器發出)一起顯示的文字之語言或其他特性的文字類型資訊、可描述調適集中之表示之場景的攝影機角度或真實世界攝影機視角的攝影機角度資訊、描述對於特定觀眾之內容適合性的分級資訊，或其類似資訊。

資訊清單檔案66可包括指示對應於特定調適集之表示68之子集以及該等調適集之共同特性的資料。資訊清單檔案66亦可包括表示調適集之個別表示的個別特性(諸如位元速率)之資料。以此方式，調適集可提供簡化的網路頻寬調適。調適集中之表示可使用資訊清單檔案66之調適集元素的子代元素來指示。

伺服器器件60包括請求處理單元70及網路介面72。在一些實例中，伺服器器件60可包括複數個網路介面。此外，伺服器器件60之特徵中之任一者或全部可在內容傳遞網路之其他器件(諸如路由器、橋接器、代理器件、交換器或其他器件)上實施。在一些實例中，內容傳遞網路之中間器件可快取多媒體內容64之資料，且包括大體上符合伺服器器件60之組件的組件。一般而言，網路介面72經組態以經由網路74來發送及接收資

料。

請求處理單元70經組態以自用戶端器件(諸如用戶端器件40)接收對儲存媒體62之資料的網路請求。舉例而言，請求處理單元70可實施超文字傳送協定(HTTP)版本1.1，如RFC 2616，R.費爾丁(R.Fielding)等人(網路工作組，IETF，1999年6月)的「超文字傳送協定-HTTP/1.1 (Hypertext Transfer Protocol—HTTP/1.1)」中所描述。即，請求處理單元70可經組態以接收HTTP GET或部分GET請求，且回應於該等請求而提供多媒體內容64之資料。請求可指定表示68中之一者的區段，例如使用區段之URL。在一些實例中，該等請求亦可指定區段之一或多個位元組範圍，因此包含部分GET請求。請求處理單元70可經進一步組態以服務於HTTP HEAD請求以提供表示68中之一者之區段的標頭資料。在任何狀況下，請求處理單元70可經組態以處理該等請求以將所請求之資料提供至請求器件(諸如用戶端器件40)。

另外或替代地，請求處理單元70可經組態以經由諸如eMBMS之廣播或多播協定而傳遞媒體資料。內容準備器件20可用與所描述大體上相同的方式產生DASH區段及/或子區段，但伺服器器件60可使用eMBMS或另一廣播或多播網路輸送協定來傳遞此等區段或子區段。舉例而言，請求處理單元70可經組態以自用戶端器件40接收多播群組加入請求。即，伺服器器件60可向與特定媒體內容(例如，實況事件之廣播)相關聯的用戶端器件(包括用戶端器件40)通告與多播群組相關聯之網際網路協定(IP)位址。用戶端器件40又可提交加入多播群組之請求。此請求可遍及網路74 (例如，組成網路74之路由器)傳播，以使得致使該等路由器將去往與多播群組相關聯之IP位址的訊務導向至訂用的用戶端器件(諸如用戶端器件40)。

如圖1之實例中所說明，多媒體內容64包括資訊清單檔案66，其可對應於媒體呈現描述(MPD)。資訊清單檔案66可含有不同替代表示68（例如，具有不同品質之視訊服務)的描述，且該描述可包括(例如)編碼解碼器資訊、設定檔值、層級值、位元速率及表示68之其他描述性特性。用戶端器件40可擷取媒體呈現之MPD以判定如何存取表示68之區段。

詳言之，擷取單元52可擷取用戶端器件40之組態資料(未展示)以判定視訊解碼器48之解碼能力及視訊輸出端44之轉譯能力。組態資料亦可包括由用戶端器件40之使用者選擇的語言偏好中之任一者或全部、對應於由用戶端器件40之使用者設定的深度偏好之一或多個攝影機視角及/或由用戶端器件40之使用者選擇的分級偏好。舉例而言，擷取單元52可包含網頁瀏覽器或媒體用戶端，其經組態以提交HTTP GET及部分GET請求。擷取單元52可對應於由用戶端器件40之一或多個處理器或處理單元(未展示)執行的軟體指令。在一些實例中，關於擷取單元52所描述的功能性之全部或部分可在硬體或硬體、軟體及/或韌體之組合中實施，其中可提供必需的硬體以執行軟體或韌體之指令。

擷取單元52可將用戶端器件40之解碼及轉譯能力與由資訊清單檔案66之資訊指示之表示68的特性進行比較。擷取單元52可首先擷取資訊清單檔案66之至少一部分以判定表示68之特性。舉例而言，擷取單元52可請求描述一或多個調適集之特性的資訊清單檔案66之一部分。擷取單元52可選擇具有用戶端器件40之寫碼及轉譯能力可滿足之特性的表示68之子集(例如，一調適集)。擷取單元52可接著(例如)判定用於調適集中之表示的位元速率，判定網路頻寬之一當前可用量，且自具有網路頻寬可滿足之位元速率的表示中之一者擷取區段。

一般而言，較高位元速率表示可產生較高品質之播放，而較低位元速率表示可在可用網路頻寬減少時提供足夠品質之播放。因此，當可用網路頻寬相對高時，擷取單元52可自相對高位元速率之表示擷取資料，而當可用網路頻寬較低時，擷取單元52可自相對低位元速率之表示擷取資料。以此方式，用戶端器件40可經由網路74串流傳輸多媒體資料，同時亦適應網路74之改變的網路頻寬可用性。

另外或替代地，擷取單元52可經組態以根據諸如eMBMS或IP多播之廣播或多播網路協定而接收資料。在此等實例中，擷取單元52可提交一請求以加入與特定媒體內容相關聯之多播網路群組。在加入該多播群組之後，擷取單元52可接收該多播群組之資料而無需發佈另外請求至伺服器器件60或內容準備器件20。當不再需要該多播群組之資料時，擷取單元52可提交一請求以離開該多播群組，例如用來停止播放或將聲道改變至不同多播群組。

網路介面54可接收經選定表示之區段的資料且將該資料提供至擷取單元52，擷取單元52又可將該等區段提供至解囊封單元50。解囊封單元50可將一視訊檔案之元素解囊封成組成性PES串流，解封包化該等PES串流以擷取經編碼資料，且取決於經編碼資料為一音訊串流抑或視訊串流之部分(例如，如由串流之PES封包標頭所指示)而將經編碼資料發送至音訊解碼器46或視訊解碼器48。音訊解碼器46解碼經編碼音訊資料，且將經解碼音訊資料發送至音訊輸出端42，而視訊解碼器48解碼經編碼視訊資料，且將經解碼視訊資料發送至視訊輸出端44，經解碼視訊資料可包括串流之複數個視圖。音訊輸出端42可包含一或多個揚聲器，而視訊輸出端44可包括一或多個顯示器。儘管未在圖1中展示，但用戶端器件40亦可包

括一或多個使用者介面，諸如鍵盤、滑鼠、指標、觸控式螢幕器件、遠端控制介面(例如，藍芽或紅外線遙控器)或其類似者。

視訊編碼器28、視訊解碼器48、音訊編碼器26、音訊解碼器46、囊封單元30、擷取單元52及解囊封單元50各自可在適用的情況下實施為多種合適處理電路系統中之任一者，合適處理電路系統諸如一或多個微處理器、數位信號處理器(DSP)、特定應用積體電路(ASIC)、場可程式化閘陣列(FPGA)、離散邏輯電路系統、軟體、硬體、韌體或其任何組合。視訊編碼器28及視訊解碼器48中之每一者可包括於一或多個編碼器或解碼器中，編碼器或解碼器中之任一者可經整合為組合式視訊編碼器/解碼器(編碼解碼器(CODEC))之部分。同樣地，音訊編碼器26及音訊解碼器46中之每一者可包括於一或多個編碼器或解碼器中，編碼器或解碼器中之任一者可經整合為組合式CODEC之部分。包括視訊編碼器28、視訊解碼器48、音訊編碼器26、音訊解碼器46、囊封單元30、擷取單元52及/或解囊封單元50之裝置可包含積體電路、微處理器及/或無線通信器件，諸如蜂巢式電話。

用戶端器件40、伺服器器件60及/或內容準備器件20可經組態以根據本發明之技術操作。出於實例之目的，本發明關於用戶端器件40及伺服器器件60描述此等技術。然而，應理解，代替伺服器器件60 (或除了伺服器器件60以外)，內容準備器件20可經組態以執行此等技術。

囊封單元30可形成NAL單元，該等NAL單元包含識別NAL所屬之程式的標頭，以及有效負載，例如音訊資料、視訊資料或描述NAL單元對應於的輸送或程式串流的資料。舉例而言，在H.264/AVC中，NAL單元包括1位元組標頭及變化大小之有效負載。在有效負載中包括視訊資料之

NAL單元可包含各種粒度層級之視訊資料。舉例而言，NAL單元可包含視訊資料區塊、複數個區塊、視訊資料之圖塊或視訊資料之整個圖像。囊封單元30可自視訊編碼器28接收呈基本串流之PES封包之形式的經編碼視訊資料。囊封單元30可使每一基本串流與對應程式相關聯。

囊封單元30亦可組譯來自複數個NAL單元之存取單元。一般而言，存取單元可包含用於表示視訊資料之訊框以及對應於該訊框之音訊資料(當此音訊資料可用時)的一或多個NAL單元。存取單元通常包括針對一個輸出時間執行個體之所有NAL單元，例如針對一個時間執行個體之所有音訊及視訊資料。舉例而言，若每一視圖具有20訊框每秒(fps)之訊框速率，則每一時間執行個體可對應於0.05秒之時間間隔。在此時間間隔期間，可同時轉譯相同存取單元(相同時間執行個體)之所有視圖的特定訊框。在一個實例中，存取單元可包含一個時間執行個體中之經寫碼圖像，其可呈現為初級經寫碼圖像。

因此，存取單元可包含共同時間執行個體之所有音訊訊框及視訊訊框，例如對應於時間X之所有視圖。本發明亦將特定視圖之經編碼圖像稱作「視圖分量」。即，視圖分量可包含在特定時間針對特定視圖之經編碼圖像(或訊框)。因此，存取單元可被定義為包含共同時間執行個體之所有視圖分量。存取單元之解碼次序未必與輸出或顯示次序相同。

媒體呈現可包括媒體呈現描述(MPD)，其可含有不同替代表示(例如，具有不同品質之視訊服務)之描述，且該描述可包括(例如)編碼解碼器資訊、設定檔值及層級值。MPD為資訊清單檔案(諸如資訊清單檔案66)之一個實例。用戶端器件40可擷取媒體呈現之MPD以判定如何存取各種呈現之電影片段。電影片段可位於視訊檔案之電影片段方塊(moof方塊)

中。

資訊清單檔案66 (其可包含(例如)MPD)可通告表示68之區段的可用性。即，MPD可包括指示表示68中之一者之第一片段變得可用時之掛鐘時間的資訊，以及指示表示68內之區段之持續時間的資訊。以此方式，用戶端器件40之擷取單元可基於開始時間以及在特定區段之前的區段之持續時間而判定何時每一區段可用。

在囊封單元30已基於所接收之資料將NAL單元及/或存取單元組譯成視訊檔案之後，囊封單元30將視訊檔案傳遞至輸出介面32以用於輸出。在一些實例中，囊封單元30可將視訊檔案儲存在本端，或經由輸出介面32而將視訊檔案發送至遠端伺服器，而非將視訊檔案直接發送至用戶端器件40。輸出介面32可包含(例如)傳輸器、收發器、用於將資料寫入至電腦可讀媒體之器件(諸如光碟機、磁性媒體驅動(例如，軟碟機))、通用串列匯流排(USB)埠、網路介面或其他輸出介面。輸出介面32將視訊檔案輸出至電腦可讀媒體，諸如傳輸信號、磁性媒體、光學媒體、記憶體、隨身碟或其他電腦可讀媒體。

網路介面54可經由網路74而接收NAL單元或存取單元，且經由擷取單元52而將NAL單元或存取單元提供至解囊封單元50。解囊封單元50可將視訊檔案之元素解囊封成組成性PES串流，解封包化該等PES串流以擷取經編碼資料，且取決於經編碼資料為音訊串流抑或視訊串流之部分(例如，如由串流之PES封包標頭所指示)而將經編碼資料發送至音訊解碼器46或視訊解碼器48。音訊解碼器46解碼經編碼音訊資料，且將經解碼音訊資料發送至音訊輸出端42，而視訊解碼器48解碼經編碼視訊資料，且將經解碼視訊資料發送至視訊輸出端44，經解碼視訊資料可包括串流之複

數個視圖。

如圖2中所展示且關於圖2更詳細地論述，擷取單元52可包括(例如) DASH用戶端。DASH用戶端可經組態以與音訊解碼器46互動，音訊解碼器46可表示MPEG-H 3D音訊解碼器。儘管未在圖1中展示，但音訊解碼器46可經進一步組態以自使用者介面(例如，如圖5至圖9中所展示)接收使用者輸入。因此，DASH用戶端可將可用性資料發送至音訊解碼器46，可用性資料可判定哪些調適集對應於哪些類型之音訊資料(例如，場景、物件及/或聲道音訊資料)。音訊解碼器46可進一步經由使用者介面而(例如)自使用者或自預先組態之選擇接收選擇資料。音訊解碼器46可接著將指令資料發送至擷取單元52 (待發送至DASH用戶端)以造成DASH用戶端擷取經選定調適集(其對應於經選定類型之音訊資料，例如，場景、聲道及/或物件資料)之音訊資料。

圖2為更詳細地說明圖1之擷取單元52之組件之實例集合的方塊圖。應理解，圖2之擷取單元52僅為一個實例；在其他實例中，擷取單元52可僅對應於DASH用戶端。在此實例中，擷取單元52包括eMBMS中間軟體單元100、DASH用戶端110及媒體應用程式112。圖2亦展示圖1之音訊解碼器46，DASH用戶端110可與音訊解碼器46互動，如下文所論述。

在此實例中，eMBMS中間軟體單元100進一步包括eMBMS接收單元106、快取記憶體104及伺服器單元102。在此實例中，eMBMS接收單元106經組態以經由eMBMS接收資料，例如，根據T.Paila等人在「FLUTE-經由單向輸送之檔案傳遞(FLUTE-File Delivery over Unidirectional Transport)」(網路工作群組，RFC 6726，2012年11月)(可於<http://tools.ietf.org/html/rfc6726>獲得)中所描述的單向輸送檔案傳遞(File

Delivery over Unidirectional Transport，FLUTE)。即，eMBMS接收單元106可經由廣播而自(例如)伺服器器件60 (其可充當BM-SC)接收檔案。

當eMBMS中間軟體單元100接收檔案之資料時，eMBMS中間軟體單元可將所接收之資料儲存於快取記憶體104中。快取記憶體104可包含電腦可讀儲存媒體，諸如快閃記憶體、硬碟、RAM或任何其他合適的儲存媒體。

代理伺服器102可充當DASH用戶端110之伺服器。舉例而言，代理伺服器102可將MPD檔案或其他資訊清單檔案提供至DASH用戶端110。代理伺服器102可通告MPD檔案中之區段的可用性時間，以及可擷取該等區段之超連結。此等超連結可包括對應於用戶端器件40之本端主機位址首碼(例如，IPv4之127.0.0.1)。以此方式，DASH用戶端110可使用HTTP GET或部分GET請求而自代理伺服器102請求區段。舉例而言，對於可自連結http://127.0.0.1/rep1/seg3獲得之區段，DASH用戶端110可建構包括針對http://127.0.0.1/rep1/seg3之請求的HTTP GET請求，且將該請求提交至代理伺服器102。代理伺服器102可自快取記憶體104擷取所請求之資料且回應於此等請求而將資料提供至DASH用戶端110。

儘管在圖2之實例中，擷取單元52包括eMBMS中間軟體單元100，但應理解，在其他實例中，可提供其他類型之中間軟體。舉例而言，代替eMBMS中間軟體100，可提供諸如高級電視系統委員會(ATSC)或國家電視系統委員會(NTSC)中間軟體之廣播中間軟體以分別接收ATSC或NTSC廣播信號。代替eMBMS接收單元106，此類ATSC或NTSC中間軟體將包括ATSC或NTSC接收單元，但仍包括如圖2之實例中所展示的代理伺服器及快取記憶體。接收單元可接收及快取所有接收到的廣播資料，且代理伺

服器可簡單地僅將所請求之媒體資料(例如，所請求之音訊資料)發送至 DASH用戶端110。

此外，DASH用戶端110可與如上文關於圖1所論述之音訊解碼器46互動。即，DASH用戶端110可接收資訊清單檔案或包括可用性資料之其他資料集。可根據(例如) MPEG-H 3D音訊而格式化可用性資料。此外，可用性資料可描述哪一/哪些調適集包括各種類型之音訊資料，諸如場景音訊資料、聲道音訊資料、物件音訊資料及/或可調式音訊資料。DASH用戶端110可自音訊解碼器46接收選擇資料，其中選擇資料可(例如)基於使用者之選擇而指示待擷取音訊資料之調適集。

圖3A為說明實例多媒體內容120之元素的概念圖。多媒體內容120可對應於多媒體內容64 (圖1)，或儲存於儲存媒體62中之另一多媒體內容。在圖3A之實例中，多媒體內容120包括媒體呈現描述(MPD) 122及複數個表示124A至124N (表示124)。表示124A包括視情況選用之標頭資料126及區段128A至128N (區段128)，而表示124N包括視情況選用之標頭資料130及區段132A至132N (區段132)。為了方便起見，使用字母N來指定表示124中之每一者中的最後一個電影片段。在一些實例中，表示124之間可存在不同數目個電影片段。

MPD 122可包含與表示124分離之資料結構。MPD 122可對應於圖1之資訊清單檔案66。同樣地，表示124可對應於圖2之表示68。一般而言，MPD 122可包括大體上描述表示124之特性的資料，諸如寫碼及轉譯特性、調適集、MPD 122所對應之設定檔、文字類型資訊、攝影機角度資訊、分級資訊、特技模式資訊(例如，指示包括時間子序列之表示的資訊)及/或用於擷取遠端週期(例如，用於在播放期間將針對性廣告插入至媒體

內容中)之資訊。

標頭資料126 (當存在時)可描述區段128之特性，例如隨機存取點(RAP，其亦被稱作串流存取點(SAP))之時間位置、區段128中之哪一者包括隨機存取點、與區段128內之隨機存取點之位元組偏移、區段128之統一資源定位符(URL)，或區段128之其他態樣。標頭資料130 (當存在時)可描述區段132之相似特性。另外或替代地，此等特性可完全包括於MPD 122內。

區段128、132包括一或多個經寫碼媒體樣本。區段128之經寫碼媒體樣本中之每一者可具有相似特性，例如，語言(若包括語音的話)、位置、CODEC及頻寬要求。此等特性可藉由MPD 122之資料來描述，不過此資料在圖3A之實例中未說明。MPD 122可包括如3GPP規範所描述之特性，同時添加了本發明中所描述的發信資訊中之任一者或全部。

區段128、132中之每一者可與唯一的統一資源定位符(URL)相關聯。因此，區段128、132中之每一者可使用串流網路協定(諸如DASH)來獨立地擷取。以此方式，諸如用戶端器件40之目的地器件可使用HTTP GET請求來擷取區段128或132。在一些實例中，用戶端器件40可使用HTTP部分GET請求來擷取區段128或132之特定位元組範圍。

圖3B為說明表示124BA至124BD (表示124B)之另一實例集合的概念圖。在此實例中，假定各種表示124B各自對應於不同的各別調適集。

可調式基於場景之音訊可包括關於再現佈局之資訊。可存在不同類型之基於場景之音訊編碼解碼器。遍及本發明描述不同實例。舉例而言，基於場景之音訊可調式編碼解碼器類型0可包括：層0包括音訊左側及音訊右側聲道，層1包括水平HOA分量，且層2包括與擴音器之高度相關的1階

HOA之高度資訊(此為圖13A及圖13B中之情境)。

在第二實例中，基於場景之音訊可調式編碼解碼器類型1可包括：層0包括音訊左側及音訊右側聲道，層1包括水平HOA分量，且層2包括與擴音器之高度相關的1階HOA之高度資訊(例如，如圖14A及圖14B中所展示)。

在第三實例中，基於場景之音訊可調式編碼解碼器類型2可包括：層0包括單聲道，層1包括音訊左側及音訊右側聲道，層2包括音訊前部及音訊背部聲道，且層3包括1階HOA之高度資訊。

在第四實例中，基於場景之音訊可調式編碼解碼器類型3可包括：層0包括呈W、X及Y信號形式的1階僅水平HOA資訊。層1包括音訊左側及音訊右側聲道，層2包括音訊前部及音訊背部聲道。

在第五實例中，可使用第一實例至第四實例，且額外層可包括擴音器之不同陣列的高度資訊，例如，在於先前實例中之揚聲器可位於的水平平面下方或上方的高度處。

因此，表示124各自對應於包括各種類型之基於場景之可調式音訊資料的不同調適集。儘管展示四個實例表示124，但應理解，可提供任何數目個調適集(及彼等調適集內之任何數目個表示)。

在圖3B之實例中，表示124BA包括類型0可調式基於場景之音訊資料，表示124BB包括類型1可調式基於場景之音訊資料，表示124BC包括類型2可調式基於場景之音訊資料，且表示124BD包括類型3可調式基於場景之音訊資料。表示124B包括對應類型之各別區段。即，表示124BA包括標頭資料類型0 126BA及類型0區段128BA至128BN，表示124BB包括標頭資料類型1 126BB及類型1區段128CA至128CN，表示124BC包括標

頭資料類型2 126BC及類型2區段128DA至128DN，且表示124BD包括標頭資料類型3 126BD及類型3區段128EA至128EN。在MPD 122B中描述各種調適集(詳言之，包括於調適集中之可調式音訊層以及表示124B中之哪些對應於哪些調適集)。

圖4為說明實例媒體檔案150之元素的方塊圖，實例媒體檔案可對應於表示之區段，諸如圖3之區段114、124中之一者。區段128、132中之每一者可包括大體上符合圖4之實例中所說明之資料之配置的資料。媒體檔案150可稱為囊封區段。如上文所描述，根據ISO基本媒體檔案格式及其擴展的媒體檔案將資料儲存於一系列物件(其被稱為「方塊」)中。在圖4之實例中，媒體檔案150包括檔案類型(FTYP)方塊152、電影(MOOV)方塊154、區段索引(sidx)方塊162、電影片段(MOOF)方塊164及電影片段隨機存取(MFRA)方塊166。儘管圖4表示視訊檔案之實例，但應理解，根據ISO基本媒體檔案格式及其擴展，其他媒體檔案可包括其他類型之媒體資料(例如，音訊資料、計時文字資料或其類似者)，其在結構上相似於媒體檔案150之資料。

檔案類型(FTYP)方塊152通常描述媒體檔案150之檔案類型。檔案類型方塊152可包括識別描述媒體檔案150之最佳用途之規範的資料。檔案類型方塊152可替代地置放在MOOV方塊154、電影片段方塊164及/或MFRA方塊166之前。

在圖4之實例中，MOOV方塊154包括電影標頭(MVHD)方塊156、播放軌(TRAK)方塊158及一或多個電影延伸(MVEX)方塊160。一般而言，MVHD方塊156可描述媒體檔案150之一般特性。舉例而言，MVHD方塊156可包括描述媒體檔案150何時最初建立、媒體檔案150何時經最後修

改、媒體檔案150之時間標度、媒體檔案150之播放持續時間的資料，或大體上描述媒體檔案150之其他資料。

TRAK方塊158可包括媒體檔案150之播放軌的資料。TRAK方塊158可包括播放軌標頭(TKHD)方塊，其描述對應於TRAK方塊158之播放軌的特性。在一些實例中，TRAK方塊158可包括經寫碼視訊圖像，而在其他實例中，播放軌之經寫碼視訊圖像可包括於電影片段164中，其可藉由TRAK方塊158及/或sidx方塊162之資料參考。

在一些實例中，媒體檔案150可包括多於一個播放軌。因此，MOOV方塊154可包括數個TRAK方塊，其等於媒體檔案150中之播放軌的數目。TRAK方塊158可描述媒體檔案150之對應播放軌的特性。舉例而言，TRAK方塊158可描述對應播放軌之時間及/或空間資訊。當囊封單元30(圖3)在視訊檔案(諸如媒體檔案150)中包括參數集播放軌時，相似於MOOV方塊154之TRAK方塊158的TRAK方塊可描述參數集播放軌的特性。囊封單元30可在描述參數集播放軌之TRAK方塊內發信序列層級SEI訊息存在於參數集播放軌中。

MVEX方塊160可描述對應電影片段164之特性，例如，發信媒體檔案150除了包括於MOOV方塊154(若存在)內之視訊資料以外亦包括電影片段164。在串流傳輸視訊資料的內容背景中，經寫碼視訊圖像可包括於電影片段164中，而非包括於MOOV方塊154中。因此，所有經寫碼視訊樣本可包括於電影片段164中，而非包括於MOOV方塊154中。

MOOV方塊154可包括數個MVEX方塊160，其等於媒體檔案150中之電影片段164的數目。MVEX方塊160中之每一者可描述電影片段164中之對應電影片段的特性。舉例而言，每一MVEX方塊可包括電影延伸標頭方

塊(MEHD)方塊，其描述電影片段164中之對應電影片段的持續時間。

如上文所提及，囊封單元30 (圖1)可儲存視訊樣本中之序列資料集，其並不包括實際經寫碼視訊資料。視訊樣本可大體上對應於存取單元，其為特定時間執行個體下之經寫碼圖像的表示。在AVC的內容背景中，經寫碼圖像包括一或多個VCL NAL單元，其含有用以建構存取單元之所有像素的資訊，及其他相關聯的非VCL NAL單元(諸如SEI訊息)。因此，囊封單元30可在電影片段164中之一者中包括序列資料集，其可包括序列層級SEI訊息。囊封單元30可另外在對應於電影片段164中之一者的MVEX方塊160之一者內發信序列資料集及/或序列層級SEI訊息存在於電影片段164中之一者中。

SIDX方塊162為媒體檔案150之視情況選用之元素。即，符合3GPP檔案格式或其他此等檔案格式之視訊檔案未必包括SIDX方塊162。根據3GPP檔案格式之實例，SIDX方塊可用以識別區段(例如，含於媒體檔案150內之區段)之子區段。3GPP檔案格式將子區段定義為「具有對應媒體資料方塊的一或多個連續電影片段方塊之自含式集合，且含有藉由電影片段方塊引用之資料的媒體資料方塊必須跟在彼電影片段方塊之後且在含有關於同一播放軌之資訊的下一個電影片段方塊之前」。3GPP檔案格式亦指示SIDX方塊「含有對由方塊記錄之(子)區段之子區段參考的序列。所參考子區段在呈現時間上鄰接。相似地，由區段索引方塊參考之位元組始終在區段內鄰接。所參考大小給出所參考材料中之位元組數目的計數」。

SIDX方塊162大體上提供表示包括於媒體檔案150中之區段之一或多個子區段的資訊。舉例而言，此資訊可包括子區段開始及/或結束之播放時間、子區段之位元組偏移、子區段是否包括(例如，開始於)串流存取點

(SAP)、SAP之類型(例如，SAP為瞬時解碼器再新(IDR)圖像、清潔隨機存取(CRA)圖像、斷鏈存取(BLA)圖像或其類似者)、SAP在子區段中的位置(就播放時間及/或位元組偏移而言)及其類似者。

電影片段164可包括一或多個經寫碼視訊圖像。在一些實例中，電影片段164可包括一或多個圖像群組(GOP)，其中之每一者可包括數個經寫碼視訊圖像，例如訊框或圖像。另外，如上文所描述，在一些實例中，電影片段164可包括序列資料集。電影片段164中之每一者可包括電影片段標頭方塊(MFHD，圖4中未展示)。MFHD方塊可描述對應電影片段之特性，諸如電影片段之序號。電影片段164可按序號次序包括於媒體檔案150中。

MFRA方塊166可描述媒體檔案150之電影片段164內的隨機存取點。此可輔助執行特技模式，諸如執行對由媒體檔案150囊封之區段內之特定時間位置(即，播放時間)的尋找。在一些實例中，MFRA方塊166通常係視情況選用的且無需包括於視訊檔案中。同樣地，用戶端器件(諸如用戶端器件40)未必需要參考MFRA方塊166來正確地解碼及顯示媒體檔案150之視訊資料。MFRA方塊166可包括數個播放軌片段隨機存取(TFRA)方塊(未展示)，其數目等於媒體檔案150之播放軌的數目，或在一些實例中等於媒體檔案150之媒體播放軌(例如，非提示播放軌)的數目。

在一些實例中，電影片段164可包括一或多個串流存取點(SAP)。同樣地，MFRA方塊166可提供對SAP在媒體檔案150內之位置的指示。因此，媒體檔案150之時間子序列可由媒體檔案150之SAP形成。時間子序列亦可包括其他圖像，諸如依賴於SAP之P訊框及/或B訊框。時間子序列之訊框及/或圖塊可配置於區段內，以使得時間子序列的取決於子序列之其

他訊框/圖塊之訊框/圖塊可被恰當地解碼。舉例而言，在資料之階層式配置中，用於其他資料之預測的資料亦可包括於時間子序列中。

圖5A為說明用於輸送諸如經編碼3D音訊資料之經編碼媒體資料之實例系統200的方塊圖。系統200包括基於物件之內容202，其自身包括後設資料204、場景資料206、聲道資料208之各種集合及物件資料210之各種集合。圖5B大體上相似於圖5A，其例外之處在於圖5B包括基於音訊之內容202'而非圖5A之基於物件之內容202。將基於物件之內容202提供至MPEG-H音訊編碼器212，MPEG-H音訊編碼器212包括音訊編碼器214及多工器216。MPEG-H音訊編碼器212可大體上對應於音訊編碼器26（圖1）。多工器216可形成囊封單元30（圖1）之部分或與囊封單元30互動。儘管圖5A中未展示，但應理解，亦可提供視訊編碼單元及多工單元，如圖1中所展示。

在此實例中，MPEG-H音訊編碼器212接收基於物件之內容202且造成音訊編碼器214編碼基於物件之內容202。將經編碼及經多工音訊資料218輸送至MPEG-H音訊解碼器220，MPEG-H音訊解碼器220包括後設資料提取單元222、場景資料提取單元224及物件資料提取單元226。提供使用者介面228以允許使用者經由應用程式設計介面(API) 230而存取所提取後設資料之版本，以使得使用者可選擇場景資料206、聲道資料208及/或物件資料210中之一或多者以在播放期間進行轉譯。根據選定的場景、聲道及/或物件，場景資料提取單元224及物件資料提取單元226可提取所請求之場景、聲道及/或物件資料，MPEG-H音訊解碼器220在播放期間解碼該等資料且將資料提供至音訊轉譯單元232。

在圖5A之實例中，將基於物件之內容202的所有資料提供於單一串流

中，其由經編碼及經多工音訊資料218表示。然而，多個串流可用以分離地提供基於物件之內容202的不同元素。舉例而言，圖6A及圖6B為說明來自基於物件之內容202（或基於音訊之內容202'）的各種類型之資料被分離地串流傳輸之其他實例的方塊圖。詳言之，在圖6A及圖6B之實例中，將場景資料206之經編碼版本提供於串流240中，串流240亦可包括聲道資料208之經編碼版本。

在圖6A及圖6B之實例中，以串流242A至242N（串流242）的形式提供物件資料210之經編碼版本。物件資料210與串流242之間的映射可以任何方式形成。舉例而言，物件資料210之集合及串流242之間可存在一對一映射，可將物件資料210之多個集合提供於串流242中之單一串流中，及/或串流242中之一或多者可包括物件資料210之一個集合的資料。可使用空中信號（諸如高級電視系統委員會(ATSC)或國家電視系統委員會(NTSC)信號）、基於電腦網路之廣播或多播（諸如eMBMS）或基於電腦網路之單播（諸如HTTP）來傳輸串流218、240、242。以此方式，當並不需要物件資料210之某些集合時，MPEG-H音訊解碼器220可避免接收串流242中之對應串流的資料。

根據本發明之一些實例，每一場景可具有組態資訊（例如，在電影標頭中，諸如圖4之MOOV方塊154）。組態資訊可含有關於物件及物件表示何者之資訊。組態資訊亦可含有可由互動性引擎使用之某種資訊。習知地，此組態資訊已為靜態的，且幾乎不可改變。然而，可使用MPEG-2 TS之技術在頻帶內修改此資訊。組態資訊亦描述物件至不同串流的映射，如圖6A及圖6B中所展示。

諸如圖6A之串流240的主串流可包括組態資訊以及在何處尋找全部物

件(例如，物件資料210)。舉例而言，串流240可包括指示串流242中之哪些含有物件資料210中之哪些的資料。串流242可被稱作「補充串流」，此係因為其僅可攜載物件資料210中的所含有之物件資料的存取單元。一般而言，每一物件可攜載於補充串流242中之個別補充串流中，不過如上文所論述，補充串流可攜載多個物件之資料，及/或一物件可攜載於多個補充串流中。

使用者介面228與後設資料提取單元222之間存在API 230。API 230可允許與包括於主串流中之後設資料之組態記錄互動。因此，API 230可允許使用者或其他實體選擇物件資料210之一或多個物件及界定其轉譯。舉例而言，使用者可選擇物件資料210之哪些物件係需要的，以及播放所要物件中之每一者的音量。

在以下論述中，假定將物件資料210之每一物件提供於分離補充串流中(例如，物件資料210與串流242之間存在一對一及至…上(onto)關係)。然而，應理解，物件資料210可經多工及經映射為傳遞最佳化。根據DASH，可將每一補充串流映射至一或多個表示中。

圖7A至圖7C為說明根據本發明之技術之另一實例系統250的方塊圖。系統250大體上包括相似於圖5A、圖5B、圖6A及圖6B之系統200之元素的元素，該等元素在圖7A及圖7B中以相同方式編號。然而，系統250另外包括媒體伺服器252，其在圖5A、圖5B、圖6A及圖6B中未展示。圖7C大體上相似於圖7A，其例外之處在於圖7C包括基於音訊之內容202'而非圖7A之基於物件之內容202。

根據本發明之技術，媒體伺服器252提供經編碼後設資料254、場景及聲道調適集256及多種物件調適集260A至260N (物件調適集260)。如圖

7B中所展示，場景及聲道調適集256包括表示258A至258M（表示258），物件調適集260A包括表示262A至262P（表示262），且物件調適集260N包括表示264A至264Q（表示264）。儘管在此實例中，場景及聲道調適集256經展示為單一調適集，但在其他實例中，可提供場景資料及聲道資料之分離調適集。即，在一些實例中，第一調適集可包括場景資料且第二調適集可包括聲道資料。

在圖7A及圖7B之實例中，根據以下映射提供內容。存在一個主物件，其為入口點且攜載組態資訊。每一物件被作為一個調適集(其為可選擇的)提供。在每一調適集內，提供多個表示(其為可切換的)。即，給定調適集之每一表示可具有不同位元速率以支援頻寬調適。提供指向物件之後設資料(物件與調適集之間可單獨地存在映射，例如，在MPEG-H後設資料中)。在此實例中，所有表示經時間對準以准許同步化及切換。

在接收器(其包括MPEG-H音訊解碼器220)處，最初假定所有物件係可用的。所含資料之標記可被視為「不透明的」，此係因為傳遞機制無需判定何種資料係由給定串流攜載。實情為，可使用抽象標記。表示之選擇通常為DASH用戶端操作之部分，但可由API 230支援。圖8中展示DASH用戶端之實例，如下文所論述。

圖8為說明根據本發明之技術之另外實例系統的方塊圖。詳言之，在圖8中，內容傳遞網路(其由雲端表示)提供經編碼後設資料254、場景及聲道調適集256及物件調適集260，以及媒體呈現描述(MPD) 270。儘管圖8中未展示，但媒體伺服器252可形成內容傳遞網路之部分。

另外，圖8說明DASH用戶端280。在此實例中，DASH用戶端280包括選擇單元282以及下載及切換單元284。選擇單元282通常負責選擇調適

集及對來自調適集之表示進行初始選擇，例如，根據自後設資料提取單元222接收到的選擇(其基於經由API 230而自使用者介面228接收到的選擇)。

以下內容為根據本發明之技術之基本操作序列的一個實例，出於實例及解釋之目的而參考圖8之元素。首先，DASH用戶端280下載MPD 270 (272)及音訊資料之主集合，音訊資料含有音訊後設資料及每一可用音訊物件之一個表示(即，每一可用的音訊調適集)。組態資訊可用於MPEG-H音訊解碼器220之後設資料提取單元222 (其經由API 230而與使用者介面228介接)以便手動選擇/取消選擇物件或使用者代理選擇/取消選擇(即，自動化選擇/取消選擇)。同樣地，DASH用戶端280之選擇單元282接收選擇資訊。即，MPEG-H音訊解碼器220告知DASH用戶端280關於哪一調適集(其由描述符或其他資料元素標記)待被選擇或取消選擇。此交換由圖8之元素274表示。

選擇單元282接著向下載及切換單元284提供指令以擷取經選定調適集之資料及停止下載經取消選擇之調適集的資料。因此，下載及切換單元284自內容傳遞網路擷取經選定(而非經取消選擇)調適集之資料(276)。舉例而言，下載及切換單元284可向內容傳遞網路提交HTTP GET或部分GET請求以擷取經選定調適集之經選定表示的區段。

在一些實例中，因為取消選擇某些調適集，所以下載及切換單元284可將先前已被分配至經取消選擇之調適集的頻寬分配至保持選定的其他調適集。因此，下載及切換單元284可選擇經選定調適集中之一或多者的較高位元速率(且因此，較高品質)的表示。在一些實例中，DASH用戶端280與MPEG-H音訊解碼器220交換關於某些調適集之品質預期的資訊。

舉例而言，MPEG-H音訊解碼器220可接收經選定調適集中之每一者的相對音量，且判定相較於具有較低相對音量之調適集，對於具有較高相對音量之調適集應擷取較高品質表示。

在一些實例中，代替停止經取消選擇之調適集的擷取，DASH用戶端280可簡單地擷取調適集之最低位元速率表示的資料，其可藉由未由MPEG-H音訊解碼器220解碼而緩衝。以此方式，若在未來的某一時刻，再次選定經取消選擇之調適集中之一者，則可立即解碼彼調適集之經緩衝資料。若需要且若頻寬可用，則下載及切換單元284可在重新選擇之後切換至此調適集之較高位元速率表示。

在擷取經選定調適集之資料之後，下載及切換單元284將資料提供至MPEG-H音訊解碼器220 (278)。因此，在由場景資料提取單元224及物件資料提取單元226中之對應資料提取單元進行提取之後，MPEG-H音訊解碼器220解碼所接收之資料，及將經解碼資料提供至音訊轉譯單元232以用於轉譯，及最終呈現。

亦可提供超出API 230之各種額外API。舉例而言，可提供API以用於發信MPD 270中之資料。可明確地發信MPD 270之後設資料以作為待下載以用於MPEG-H音訊之一個物件。MPD 270亦可發信需要下載之所有音訊調適集。此外，MPD 270可發信待用於選擇之每一調適集之標籤。

同樣地，可針對MPEG-H音訊解碼器220與DASH用戶端280之間的選擇及偏好邏輯來定義API。DASH用戶端280可使用此API將組態資訊提供至MPEG-H音訊解碼器220。MPEG-H音訊解碼器220可將標籤提供至DASH用戶端280，該標籤指示出於資料擷取之目的而被選定的調適集。MPEG-H音訊解碼器220亦可提供表示各種調適集之相對重要性的某一加

權，其由DASH用戶端280使用以選擇經選定調適集之適當表示。

此外，可針對將經多工媒體資料自DASH用戶端280提供至MPEG-H音訊解碼器220而定義API。DASH用戶端280通常下載指派至調適集之資料塊。DASH用戶端280以多工及附註方式提供資料，且亦實施切換邏輯以用於在調適集之表示之間進行切換。

以此方式，圖8表示一種用於擷取音訊資料之器件的一實例，該器件包括：一或多個處理器，該一或多個處理器經組態以進行以下操作：接收表示複數個可用調適集之可用性資料，該等可用調適集包括一基於場景之音訊調適集及一或多個基於物件之音訊調適集；接收選擇資料，其識別該基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料；及一記憶體，其經組態以儲存該等音訊調適集之該所擷取資料。

圖9為根據本發明之技術的另一實例系統。一般而言，圖9大體上相似於圖8之實例。圖8與圖9之間的區別在於：在圖9中，後設資料提取單元222'提供在MPEG-H音訊解碼器220'外部。因此，在圖9中，互動274'在選擇單元282與後設資料提取單元222'之間發生以用於提供表示可用調適集之後設資料及用於選擇(及/或取消選擇)可用調適集。否則，圖9之實例可以與圖8之實例大體上相一致的方式操作。然而，應強調，使用者介面無需與MPEG-H音訊解碼器220'直接地互動以執行本發明之技術。

以此方式，圖9表示一種用於擷取音訊資料之器件的一實例，該器件包括：一或多個處理器，該一或多個處理器經組態以進行以下操作：接收表示複數個可用調適集之可用性資料，該等可用調適集包括一基於場景之

音訊調適集及一或多個基於物件之音訊調適集；接收選擇資料，其識別該基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料；及一記憶體，其經組態以儲存該等音訊調適集之該所擷取資料。

圖10為說明可使用本發明之技術之另一實例系統350的概念圖。在圖10之實例中，系統350包括媒體伺服器352，媒體伺服器352準備媒體內容及將媒體內容提供至廣播伺服器354及HTTP內容傳遞網路(CDN) 358。舉例而言，廣播伺服器354可為一廣播多媒體服務中心(BMSC)。廣播伺服器354經由廣播傳輸器356而廣播媒體信號。各種使用者設備(UE)用戶端器件364A至364N (用戶端器件364) (諸如電視、個人電腦或諸如蜂巢式電話、平板電腦或其類似者之行動器件)可接收所廣播之信號。廣播傳輸器356可根據諸如ATSC或NTSC之空中標準操作。

HTTP CDN 358可經由基於電腦之網路提供媒體內容，基於電腦之網路可使用基於HTTP之串流，例如，DASH。另外或替代地，CDN 358可使用基於網路之廣播或多播協定(諸如eMBMS)經由基於電腦之網路來廣播或多播媒體內容。CDN 358包括複數個伺服器器件360A至360N (伺服器器件360)，伺服器器件360A至360N經由單播、廣播及/或多播協定而傳輸資料。在一些實例中，CDN 358根據長期演進(LTE)經由一eNode-B(諸如eNode-B 362)而在一無線電存取網路(RAN)上傳遞內容。

各種使用案例可在圖10之系統中發生。舉例而言，一些媒體分量可經由廣播(例如，藉由廣播伺服器354)傳遞，而其他媒體分量僅可經由單播獲得以作為一或多個伴隨串流。舉例而言，基於場景之音訊內容可由廣播

伺服器經由廣播傳輸器廣播，而物件音訊資料僅可自HTTP CDN 358獲得。在另一實例中，可經由單播傳遞資料以縮減聲道切換時間。

圖11為說明可實施本發明之技術之另一實例系統370的概念圖。圖11之實例在概念上相似於關於圖3所描述之實例。即，在圖11之實例系統370中，廣播DASH伺服器376將媒體資料提供至廣播檔案輸送封裝器378(例如，用於廣播傳遞檔案)。舉例而言，廣播檔案輸送封裝器378及廣播檔案輸送接收器380可根據Paila等人在「FLUTE-單向輸送檔案傳遞(FLUTE-File Delivery over Unidirectional Transport)」(網際網路工程任務小組，RFC 6726，2012年11月，可於tools.ietf.org/html/rfc6726獲得)中所描述的單向輸送檔案傳遞(FLUTE)操作。替代地，廣播檔案輸送封裝器378及廣播檔案輸送接收器380可根據單向輸送即時物件傳遞(ROUTE)協定操作。

在又一實例中，廣播檔案輸送封裝器378及廣播檔案輸送接收器380可根據諸如ATSC或NTSC之空中廣播協定操作。舉例而言，MBMS服務層可與DASH層組合以用於ATSC 3.0。此組合可按以IP為中心之方式提供分層清潔(layering-clean)MBMS服務層實施。亦可存在跨越多個傳遞路徑及方法之統一同步化。此系統亦可經由廣播而向DASH提供清潔的、經最佳化支援，該支援可提供許多益處。增強型AL FEC支援可向所有服務組件提供恆定的服務品質(QoS)。此外，此實例系統可支援各種使用案例且產生各種益處，諸如快速聲道改變及/或低潛時。

在圖11之實例中，廣播DASH伺服器376使用統一時間碼(UTC)源372來判定時序資訊，從而判定將何時傳輸媒體資料。DASH播放器384最終使用由本端UTC源374提供之時序資訊而自廣播檔案輸送接收器380接收

MPD及媒體資料382。替代地，DASH播放器384可自CDN 386擷取MPD及媒體資料382'。DASH播放器384可提取經時間對準之經壓縮媒體資料390及將經時間對準之經壓縮媒體資料390傳遞至CODEC 388（其可表示圖1之音訊解碼器46及視訊解碼器48）。CODEC 388可接著解碼經編碼媒體資料以產生經時間對準之媒體樣本及像素392，經時間對準之媒體樣本及像素392可被呈現(例如，經由圖1之音訊輸出端42及視訊輸出端44)。

圖12為說明用於ATSC 3.0之實例概念性協定模型400的概念圖。在模型400中，線性及基於應用之服務412包括線性TV、互動式服務、伴隨螢幕、個性化服務、緊急警示及使用情況報告，且可包括使用(例如) HTML 5及/或JavaScript實施之其他應用。

模型400之編碼、格式化及服務管理資料410包括各種編碼解碼器(例如，用於音訊及視訊資料)、ISO BMFF檔案、使用經加密媒體擴展(EME)之加密及/或共同加密(CENC)、媒體處理單元(MPU)、NRT檔案、發信物件及各種類型之發信資料。

在此實例中，在模型400之傳遞層408處，存在MPEG媒體輸送協定(MMTP)資料、ROUTE資料、應用層前向錯誤校正(AL FEC)資料(其可為視情況選用的)、統一資料報協定(UDP)資料及傳輸控制協定(TCP)資料406、超文字傳送協定(HTTP)資料以及網際網路協定(IP)資料404。可使用經由實體層402的廣播及/或寬頻帶傳輸來輸送此資料。

圖13A為表示多層音訊資料700之概念圖。雖然此實例描繪具有三個子層之第一層，但在其他實例中，三個子層可為三個分離層。

在圖13A之實例中，高階立體混響音訊資料之兩個或大於兩個層的第一層(其包括基底子層702、第一增強子層704及第二增強子層706)可包含

高階立體混響係數，其對應於具有等於或小於一階之一或多個球面基底函數。在一些實例中，第二層(亦即，第三增強層)包含基於向量之主要音訊資料。在一些實例中，基於向量之主要音訊至少包含主要音訊資料及經編碼V向量，其中經由應用線性可逆變換而自高階立體混響音訊資料分解得到經編碼V向量。Herre等人在2015年4月10日申請之美國臨時申請案62/145,960，「MPEG-H 3D音訊—沉浸式空間音訊之寫碼的新標準(MPEG-H 3D Audio—The New Standard for Coding of Immersive Spatial Audio)」(IEEE 9，Journal of Selected Topics in Signal Processing 5，2015年8月)包括關於V向量之額外資訊。在其他實例中，基於向量之主要音訊資料至少包含額外的高階立體混響聲道。在另外其他實例中，基於向量之主要音訊資料至少包含自動增益校正旁頻帶。在其他實例中，基於向量之主要音訊資料至少包含主要音訊資料、經編碼V向量、額外高階立體混響聲道及自動增益校正旁頻帶，其中經由應用線性可逆變換而自高階立體混響音訊資料分解得到經編碼V向量。

在圖13A之實例中，第一層702可包含至少三個子層。在一些實例中，至少三個子層之第一子層(亦即，基底層702)至少包含與左側音訊聲道相關聯之高階立體混響音訊資料。在其他實例中，至少三個子層之第一子層(亦即，基底層702)至少包含與右側音訊聲道相關聯之高階立體混響音訊資料。在另外其他實例中，至少三個子層之第一子層(亦即，基底層702)至少包含用於進行自動增益校正之旁頻帶。在其他實例中，至少三個子層之第一子層(亦即，基底層702)至少包含與左側音訊聲道及右側音訊聲道相關聯之高階立體混響音訊資料及用於進行自動增益校正之旁頻帶。

在一些實例中，圖13A之至少三個子層的第二子層(亦即，第一增強

層704)至少包含與定域化聲道相關聯之高階立體混響音訊資料。在其他實例中，至少三個子層之第二子層(亦即，第一增強層704)至少包含用於進行自動增益校正之旁頻帶。在另外其他實例中，至少三個子層之第二子層(亦即，第一增強層704)至少包含與定域化聲道相關聯之高階立體混響音訊資料及用於進行自動增益校正之旁頻帶。

在一些實例中，至少三個子層之第三子層(亦即，第二增強層706)至少包含與高度聲道相關聯之高階立體混響音訊資料。在其他實例中，至少三個子層之第三子層(亦即，第二增強層706)至少包含用於進行自動增益校正之旁頻帶。在另外其他實例中，至少三個子層之第三子層(亦即，第二增強層706)至少包含與高度聲道相關聯之高階立體混響音訊資料及用於進行自動增益校正之旁頻帶。

在圖13A之實例(其中存在四個分離層(亦即，基底層702、第一增強層704、第二增強層706及第三增強層))中，音訊寫碼器件可執行錯誤檢查程序。在一些實例中，音訊寫碼器件可對第一層(亦即，基底層702)執行錯誤檢查程序。在另一實例中，音訊寫碼器件可對第一層(亦即，基底層702)執行錯誤檢查程序及抑制對第二層、第三層及第四層執行錯誤檢查程序。在又一實例中，音訊寫碼器件可對第一層執行錯誤檢查程序，響應於判定第一層(亦即，基底層702)無誤，音訊寫碼器件可對第二層(亦即，第一增強層704)執行錯誤檢查程序，且音訊寫碼器件可抑制對第三層及第四層執行錯誤檢查程序。在又一實例中，音訊寫碼器件可對第一層執行錯誤檢查程序；響應於判定第一層(亦即，基底層702)無誤，音訊寫碼器件可對第二層執行錯誤檢查程序；響應於判定第二層(亦即，第一增強層704)無誤，音訊寫碼器件可對第三層(亦即，第二增強層)執行錯誤檢查程序，

且音訊寫碼器件可抑制對第四層執行錯誤檢查程序。在又一實例中，音訊寫碼器件可對第一層執行錯誤檢查程序；響應於判定第一層(亦即，基底層702)無誤，音訊寫碼器件可對第二層執行錯誤檢查程序；響應於判定第二層(亦即，第一增強層704)無誤，音訊寫碼器件可對第三層執行錯誤檢查程序；且響應於判定第三層(亦即，第二增強層706)無誤，音訊寫碼器件可對第四層(亦即，第三增強層)執行錯誤檢查程序。在以上實例(其中音訊寫碼器件對第一層(亦即，基底層702)執行錯誤檢查程序)中之任一者中，第一層可被視為相對於錯誤強固的強固層。

根據本發明之技術，在一個實例中，可將來自上文所描述之各種層(例如，基底層702、第二層704、第三層706及第四層)中之每一者的資料提供於各別調適集內。即，基底層調適集可包括包含對應於基底層702之資料的一或多個表示，第二層調適集可包括包含對應於第二層704之資料的一或多個表示，第三層調適集可包括包含對應於第三層706之資料的一或多個表示，且第四層調適集可包括包含對應於第四層之資料的一或多個表示。

圖13B為表示多層音訊資料之另一實例的概念圖。圖13B之實例大體上相似於圖13A之實例。然而，在此實例中，並未執行UHJ解相關。

圖14A為說明多層音訊資料710之另一實例的概念圖。雖然此實例描繪具有三個子層之第一層，但在其他實例中，三個子層可為三個分離層。

在圖14A之實例中，高階立體混響音訊資料之兩個或大於兩個層的第一層(其包括基底子層712、第一增強子層及第二增強子層)可包含高階立體混響係數，其對應於具有等於或小於一階之一或多個球面基底函數。在一些實例中，第二層(亦即，第三增強層)包含基於向量之主要音訊資料；

在一些實例中，基於向量之主要音訊至少包含主要音訊資料及經編碼V向量，其中經由應用線性可逆變換而自高階立體混響音訊資料分解得到經編碼V向量。在其他實例中，基於向量之主要音訊資料至少包含額外的高階立體混響聲道。在另外其他實例中，基於向量之主要音訊資料至少包含自動增益校正旁頻帶。在其他實例中，基於向量之主要音訊資料至少包含主要音訊資料、經編碼V向量、額外高階立體混響聲道及自動增益校正旁頻帶，其中經由應用線性可逆變換而自高階立體混響音訊資料分解得到經編碼V向量。

在圖14A之實例中，第一層可包含至少三個子層。在一些實例中，至少三個子層之第一子層(亦即，基底層712)至少包含與0階立體混響相關聯之高階立體混響音訊資料。在其他實例中，至少三個子層之第一子層(亦即，基底層712)至少包含用於進行自動增益校正之旁頻帶。在另外其他實例中，至少三個子層之第一子層(亦即，基底層712)至少包含與0階立體混響相關聯之高階立體混響音訊資料及用於進行自動增益校正之旁頻帶。

在一些實例中，至少三個子層之第二子層(亦即，第一增強層714)至少包含與X分量相關聯之高階立體混響音訊資料。在其他實例中，至少三個子層之第二子層(亦即，第一增強層714)至少包含與Y分量相關聯之高階立體混響音訊資料。在其他實例中，至少三個子層之第二子層(亦即，第一增強層714)至少包含用於進行自動增益校正之旁頻帶。在另外其他實例中，至少三個子層之第二子層(亦即，第一增強層714)至少包含與X分量及Y分量相關聯之高階立體混響音訊資料，及用於進行自動增益校正之旁頻帶。

在一些實例中，至少三個子層之第三子層(亦即，第二增強層716)至

少包含與Z分量相關聯之高階立體混響音訊資料。在其他實例中，至少三個子層之第三子層(亦即，第二增強層716)至少包含用於進行自動增益校正之旁頻帶。在另外其他實例中，至少三個子層之第三子層(亦即，第二增強層716)至少包含與Z分量相關聯之高階立體混響音訊資料及用於進行自動增益校正之旁頻帶。

在圖14A之實例(其中存在四個分離層(亦即，基底層712、第一增強層714、第二增強層716及第三增強層))中，音訊寫碼器件可執行錯誤檢查程序。在一些實例中，音訊寫碼器件可對第一層(亦即，基底層712)執行錯誤檢查程序。在另一實例中，音訊寫碼器件可對第一層(亦即，基底層712)執行錯誤檢查程序及抑制對第二層、第三層及第四層執行錯誤檢查程序。在又一實例中，音訊寫碼器件可對第一層執行錯誤檢查程序，響應於判定第一層(亦即，基底層712)無誤，音訊寫碼器件可對第二層(亦即，第一增強層714)執行錯誤檢查程序，且音訊寫碼器件可抑制對第三層及第四層執行錯誤檢查程序。在又一實例中，音訊寫碼器件可對第一層執行錯誤檢查程序；響應於判定第一層(亦即，基底層712)無誤，音訊寫碼器件可對第二層執行錯誤檢查程序；響應於判定第二層(亦即，第一增強層714)無誤，音訊寫碼器件可對第三層(亦即，第二增強層716)執行錯誤檢查程序，且音訊寫碼器件可抑制對第四層執行錯誤檢查程序。在又一實例中，音訊寫碼器件可對第一層執行錯誤檢查程序；響應於判定第一層(亦即，基底層712)無誤，音訊寫碼器件可對第二層執行錯誤檢查程序；響應於判定第二層(亦即，第一增強層714)無誤，音訊寫碼器件可對第三層執行錯誤檢查程序；且響應於判定第三層(亦即，第二增強層716)無誤，音訊寫碼器件可對第四層(亦即，第三增強層)執行錯誤檢查程序。在以上實例

(其中音訊寫碼器件對第一層(亦即，基底層712)執行錯誤檢查程序)中之任一者中，第一層可被視為相對於錯誤強固的強固層。

根據本發明之技術，在一個實例中，可將來自上文所描述之各種層(例如，基底層712、第二層、第三層及第四層)中之每一者的資料提供於各別調適集內。即，基底層712調適集可包括包含對應於基底層712之資料的一或多個表示，第二層調適集可包括包含對應於第二層714之資料的一或多個表示，第三層調適集可包括包含對應於第三層716之資料的一或多個表示，且第四層調適集可包括包含對應於第四層之資料的一或多個表示。

圖14B為表示多層音訊資料之另一實例的概念圖。圖14B之實例大體上相似於圖14A之實例。然而，在此實例中，並未執行模式矩陣解相關。

圖15為說明根據本發明之技術之傳送可調式HOA資料之另一實例系統的方塊圖。一般而言，圖15之元素大體上相似於圖8及圖9之元素。即，圖15說明包括MPEG-H音訊解碼器440之系統，MPEG-H音訊解碼器440與DASH用戶端430互動以自內容傳遞網路擷取音訊資料。圖15之與圖8及圖9之元素以相似方式命名的元素通常與如上文所論述之彼等元素以相同方式組態。然而，在此實例中，提供多個調適集，其各自對應於基於場景之音訊資料之層(或子層)，例如，如上文關於圖13A、圖13B、圖14A及圖14B所論述。

詳言之，在此實例中，CDN 420提供基於場景之可調式音訊內容422，基於場景之可調式音訊內容422包括媒體內容之經編碼後設資料424，其包括基於場景之音訊的基底層(呈基於場景之音訊、基底層調適集426的形式)及複數個增強層(呈基於場景之音訊、增強層調適集428A至

428N (調適集428)的形式)。舉例而言，基底層可包括單聲道音訊資料，第一增強層可提供左側/右側資訊，第二增強層可提供前部/背部資訊，且第三增強層可提供高度資訊。媒體內容係由MPD 421描述。

因此，使用者可經由使用者介面448指示需要哪些類型之資訊。使用者介面448可包括多種輸入及/或輸出介面中之任一者，諸如顯示器、鍵盤、滑鼠、觸控板、觸控式螢幕、軌跡墊、遙控器、麥克風、按鈕、撥號盤、滑件、開關或其類似者。舉例而言，若僅僅單一揚聲器可用，則DASH用戶端430可僅自基於場景之音訊、基底層調適集426擷取資料。然而，若多個揚聲器可用，則取決於揚聲器之配置，DASH用戶端430可自基於場景之音訊、增強層調適集428中之對應者擷取左側/右側資訊、前部/背部資訊及/或高度資訊中之任一者或全部。

下文描述DASH中之音訊資料之可調性的兩個實例類型。第一實例為靜態器件可調性。在此實例中，基底層與增強層表示不同的源信號。舉例而言，基底層可表示1080p 30fps SDR，且增強層可表示4K 60fps HDR。此情形之主要原因為支援對較低品質的存取以用於器件調適，例如，基底層係由一種器件類別選定，且增強層係由第二器件類別選定。在靜態器件可調性之實例中，將基底層及增強層提供於不同調適集中。即，器件可選擇該等調適集中之一或多者(例如，藉由自不同調適集中之互補表示獲取資料)。

第二實例係關於動態存取頻寬可調性。在此實例中，產生一個基底層及一或多個增強層。然而，所有層皆呈現相同的源信號(例如，1080p 60fps)。舉例而言，根據DASH之技術，此可支援自適應串流。即，基於所估計之頻寬的可用量，可下載/存取較多或較少之增強層。在此實例

中，基底層及增強係提供於一個調適集中且係可順暢地切換的。相較於廣播/多播傳遞，此實例可更有關於單播傳遞。

第三實例可包括靜態器件可調性技術與動態存取頻寬可調性技術之組合。

可使用DASH來支援此等實例中之每一者。

在圖15之實例中，DASH用戶端430首先接收MPD 421 (460)。選擇單元432判定可用調適集及調適集內之表示。接著選擇單元432將表示可用調適集(詳言之，可用可調式音訊層)之資料提供至MPEG-H音訊解碼器440之後設資料提取單元442(462)。在此實例中，使用者或其他實體經由API 450而將所要音訊層之選擇提供至MPEG-H音訊解碼器440。接著將此等選擇傳遞至選擇單元432。選擇單元432告知下載及切換單元434關於所要調適集以及初始表示選擇(例如，基於可用的網路頻寬)。

下載及切換單元434接著自所要調適集中之每一者的一個表示擷取資料(464)，例如，藉由將HTTP GET或部分GET請求提交至CDN 420之伺服器。在接收所請求之資料之後，下載及切換單元434將所擷取之資料提供至MPEG-H音訊解碼器440 (466)。場景資料提取單元444提取相關場景資料，且可調式音訊層解碼單元446解碼各種層中之每一者的音訊資料。最終，MPEG-H音訊解碼器440將經解碼音訊層提供至音訊轉譯單元452，音訊轉譯單元452轉譯音訊資料以供音訊輸出端454播放。音訊輸出端454可大體對應於圖1之音訊輸出端42。舉例而言，音訊輸出端454可包括呈多種配置之一或多個揚聲器。舉例而言，音訊輸出端454可包括單一揚聲器、左側及右側立體聲揚聲器、5.1配置揚聲器、7.1配置揚聲器或用以提供3D音訊之在各種高度處之揚聲器。

一般而言，上文關於圖8及圖9所論述之各種技術亦可由圖15之系統執行。

圖16為說明根據本發明之技術之實例架構的概念圖。圖16之實例包括發送器470及兩個接收器，接收器482及接收器494。

發送器470包括視訊編碼器472及音訊編碼器474。視訊編碼器472編碼視訊資料506，而音訊編碼器474編碼音訊資料508。在此實例中，發送器470可準備複數個表示，例如，三個音訊表示，表示1、表示2及表示3。因此，經編碼音訊資料508可包括表示1、表示2及表示3中之每一者的音訊資料。檔案格式囊封器476接收經編碼視訊資料506及經編碼音訊資料508且形成經囊封資料510。DASH分段器478形成區段512，區段512中之每一者包括經囊封經編碼之音訊或視訊資料的分離集合。ROUTE發送器480在各種對應位元串流中發送區段。在此實例中，位元串流514包括所有音訊資料(例如，表示1、2及3中之每一者)，而位元串流514'包括表示1及3但省略表示2。

接收器482包括視訊解碼器484，場景、物件及聲道音訊解碼器486，檔案格式剖析器488，DASH用戶端490及ROUTE接收器492；而接收器494包括視訊解碼器496，場景及聲道音訊解碼器498，檔案格式剖析器500，DASH用戶端502及ROUTE接收器504。

最終，在此實例中，接收器482接收位元串流514，其包括表示1、表示2及表示3中之每一者的資料。然而，接收器494接收位元串流514'，其包括表示1及表示3之資料。此可因為發送器與接收器494之間的網路條件並未提供用以擷取全部三種可用表示之資料的足夠量之頻寬，或因為耦接至接收器494之轉譯器件不能夠使用來自表示2之資料。舉例而言，若表

示2包括音訊資料之高度資訊，而接收器494與左側/右側立體聲系統相關聯，則來自表示2之資料對於轉譯經由接收器494接收到之音訊資料可為不必要的。

在此實例中，ROUTE接收器492接收位元串流514且在本端快取經接收區段直至DASH用戶端490請求區段為止。當區段可用性資訊(例如)基於經通告掛鐘時間指示區段可用(或應可用)時，DASH用戶端490可請求該等區段。DASH用戶端490可接著自ROUTE接收器492請求區段。DASH用戶端490可將區段510發送至檔案格式剖析器488。檔案格式剖析器488可解囊封區段且判定經解囊封資料對應於經編碼音訊資料508抑或經編碼視訊資料506。檔案格式剖析器488將經編碼音訊資料508傳遞至場景、物件及聲道音訊解碼器486且將經編碼視訊資料506傳遞至視訊解碼器484。

在此實例中，ROUTE接收器504接收位元串流514'且在本端快取經接收區段直至DASH用戶端502請求區段為止。當區段可用性資訊(例如)基於經通告掛鐘時間指示區段可用(或應可用)時，DASH用戶端502可請求該等區段。DASH用戶端502可接著自ROUTE接收器504請求區段。DASH用戶端502可將區段510'發送至檔案格式剖析器5070。檔案格式剖析器500可解囊封區段且判定經解囊封資料對應於經編碼音訊資料508' (其省略表示2，如上文所論述)抑或經編碼視訊資料506。檔案格式剖析器500將經編碼音訊資料508'傳遞至場景及聲道音訊解碼器498且將經編碼視訊資料506傳遞至視訊解碼器496。

本發明之技術可應用於多種使用案例中。舉例而言，本發明之技術可用以向兩個或大於兩個不同的接收器提供器件可調性。作為另一實例，不

同輸送工作階段(transport session)可攜載物件流及/或不同的可調式音訊層流。作為又一實例，該等技術可支援回溯相容性，此係因為傳統接收器可僅僅擷取基底層，而進階接收器可存取基底層及一或多個增強層。此外，如上文所論述，可組合媒體資料之寬頻帶、廣播/多播及/或單播接收以支援增強品質(其可被描述為混合式可調性)。此外，此等技術可支援未來技術，諸如8K信號及HDR擴展層、可調式音訊及/或即時基底層與NRT增強層之組合的技術。歸因於貫穿堆疊之功能性分離，此等使用案例中之每一者可由DASH/ROUTE支援。

以此方式，圖16表示用於擷取音訊資料之器件(接收器482、494)的實例，該等器件包括：一或多個處理器，該一或多個處理器經組態以進行以下操作：接收表示複數個可用調適集之可用性資料，該等可用調適集包括一基於場景之音訊調適集及一或多個基於物件之音訊調適集；接收選擇資料，其識別該基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料；及一記憶體，其經組態以儲存該等音訊調適集之該所擷取資料。

圖17為說明根據本發明之技術之實例用戶端器件520的方塊圖。用戶端器件520包括網路介面522，網路介面522通常提供至基於電腦之網路(諸如網際網路)的連接性。舉例而言，網路介面522可包含一或多個網路介面卡(NIC)，網路介面卡可根據諸如乙太網路及/或一或多個無線網路標準(諸如IEEE 802.11a、b、g、n或其類似者)之多種網路協定操作。

用戶端器件520亦包括DASH用戶端524。DASH用戶端524通常實施DASH技術。儘管在此實例中，用戶端器件520包括DASH用戶端524，但

在其他實例中，除了DASH用戶端524以外，用戶端器件520亦可包括中間軟體單元，例如，如上文關於圖2所論述。一般而言，DASH用戶端524自媒體內容之一或多個調適集選擇適當表示，例如，如由音訊控制器530及視訊控制器420所引導，如下文所論述。

用戶端器件520包括用於分別控制音訊及視訊資料之選擇的音訊控制器530及視訊控制器420。音訊控制器530通常根據本發明之技術操作，如上文所論述。舉例而言，音訊控制器530可經組態以接收表示可用音訊資料之後設資料(例如，自MPD或其他資料結構，諸如自MPEG-H後設資料)可用音訊資料可包括基於場景之音訊、基於聲道之音訊、基於物件之音訊或其任何組合。此外，如上文所論述，基於場景之音訊可為可調式的，亦即，具有多個層(其可提供於分離的各別調適集中)。一般而言，音訊控制器530之音訊後設資料處理單元532判定哪些類型之音訊資料係可用的。

音訊後設資料處理單元532與API 536互動，API 536提供位於使用者介面550中之一或多者與音訊後設資料處理單元532之間的介面。舉例而言，使用者介面550可包括顯示器、一或多個揚聲器、鍵盤、滑鼠、指標、軌跡板、觸控式螢幕、遙控器、麥克風、開關、撥號盤、滑件或其類似者中之一或多者，以用於自使用者接收輸入及將音訊及/或視訊輸出提供至使用者。因此，使用者可經由使用者介面550選擇所要音訊及視訊資料。

舉例而言，使用者可以多種組態中之任一者的形式將一或多個揚聲器連接至用戶端器件520。此等組態可包括單一揚聲器、立體聲揚聲器、3.1環繞、5.1環繞、7.1環繞或用於3D音訊之在多個高度及位置處之揚聲器。因此，使用者可經由使用者介面550而將揚聲器配置之指示提供至用戶端

器件520。相似地，使用者可提供對視訊組態(例如，二維視訊、三維視訊或多維視訊(例如，具有多個視角之三維視訊))之選擇。使用者介面550可經由API 426而與視訊控制器420互動，API 426以大體上相似於API 536之方式提供至視訊後設資料處理單元422的介面。

因此，音訊後設資料處理單元532可選擇適當調適集，將自該等調適集擷取音訊資料，而視訊後設資料處理單元422可選擇適當調適集，將自該等調適集擷取視訊資料。音訊後設資料處理單元532及視訊後設資料處理單元422可將待自其擷取音訊及視訊資料之調適集的指示提供至DASH用戶端524。DASH用戶端524又選擇調適集之表示且自經選擇表示擷取媒體資料(分別擷取音訊或視訊資料)。DASH用戶端524可基於(例如)可用網路頻寬、調適集之優先級或其類似者選擇該等表示。DASH用戶端524可經由網路介面522提交針對來自經選擇表示之資料的HTTP GET或部分GET請求，且回應於該等請求，經由網路介面522接收所請求之資料。DASH用戶端524可接著將所接收之資料提供至音訊控制器530或視訊控制器420。

音訊解碼器534解碼自DASH用戶端524接收之音訊資料，且視訊解碼器424解碼自DASH用戶端524接收之視訊資料。音訊解碼器534將經解碼音訊資料提供至音訊轉譯器538，而視訊解碼器424將經解碼視訊資料提供至視訊轉譯器428。音訊轉譯器538轉譯經解碼音訊資料，且視訊轉譯器428轉譯經解碼視訊資料。音訊轉譯器538將經轉譯音訊資料提供至使用者介面550以供呈現，而視訊轉譯器428將經轉譯視訊資料提供至使用者介面550以供呈現。

以此方式，圖17表示一種用於擷取音訊資料之器件的一實例，該器件

包括：一或多個處理器，該一或多個處理器經組態以進行以下操作：接收表示複數個可用調適集之可用性資料，該等可用調適集包括一基於場景之音訊調適集及一或多個基於物件之音訊調適集；接收選擇資料，其識別該基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料；及一記憶體，其經組態以儲存該等音訊調適集之該所擷取資料。

圖18為說明用於執行本發明之技術之實例方法的流程圖。在此實例中，關於伺服器器件及用戶端器件解釋方法。出於實例及解釋之目的，關於伺服器器件60 (圖1)論述伺服器器件之動作，且關於用戶端器件40 (圖1)論述用戶端器件之動作。然而，應理解，其他伺服器及用戶端器件可經組態以執行所論述之功能性。

首先，伺服器器件60編碼音訊資料(560)。舉例而言，音訊編碼器26 (圖1)、MPEG-H音訊編碼器212 (圖5至圖7)或音訊編碼器474 (圖16)編碼音訊資料，諸如場景音訊資料、聲道音訊資料、可調式音訊資料及/或物件音訊資料。伺服器器件60亦將音訊資料囊封成(例如)諸如ISO BMFF之檔案格式以用於串流傳輸音訊資料(562)。詳言之，囊封單元30 (圖1)、多工器216 (圖5、圖6)、廣播檔案輸送封裝器378 (圖11)或檔案格式囊封器476 (圖16)將經編碼音訊資料囊封成諸如根據(例如) ISO BMFF格式化之區段的可輸送檔案。伺服器器件60亦編碼可用性資料(564)。可用性資料可包括於諸如DASH之MPD的資訊清單檔案中。可用性資料自身可根據諸如MPEG-H 3D音訊之音訊編碼格式格式化。因此，伺服器器件60可將資訊清單檔案中之可用性資料發送至用戶端器件40 (566)。

用戶端器件40可接收資訊清單檔案及因此可用性資料(568)。如下文更詳細地論述，用戶端器件40之DASH用戶端可接收資訊清單檔案及提取可用性資料。然而，因為可用性資料可根據諸如MPEG-H 3D音訊之音訊編碼格式格式化，所以DASH用戶端可將可用性資料發送至MPEG-H 3D音訊解碼器(諸如圖1之音訊解碼器46)。用戶端器件40可接著自可用性資料判定待擷取之音訊資料(570)。舉例而言，如下文所論述，DASH用戶端可自(例如) MPEG-H 3D音訊解碼器(諸如圖1之音訊解碼器46)接收指示待自其擷取媒體資料之調適集的指令資料。用戶端器件40可接著根據指令資料請求所判定之音訊資料(572)。

在一個實例中，用戶端器件40可自所有可用音訊調適集請求音訊資料，但僅自未經選擇調適集(即，未由自(例如)MPEG-H 3D音訊解碼器接收之指令資料的選擇資料識別的調適集)之最低位元速率表示請求音訊資料。在此實例中，用戶端器件40可針對經選擇調適集執行頻寬調適。以此方式，若使用者選擇改變，則用戶端器件40可立即對至少一些音訊資料進行存取，且可開始針對新選擇之調適集執行頻寬調適(例如，自新選擇之調適集的較高位元速率表示擷取音訊資料)。

在另一實例中，用戶端器件40可簡單地僅自經選擇調適集請求音訊資料，及避免請求未經選擇調適集之任何音訊資料。

在任何狀況下，伺服器器件60可接收對音訊資料之請求(574)。伺服器器件60可接著將所請求之音訊資料發送至用戶端器件40 (576)。替代地，在另一實例中，伺服器器件60可經由網路廣播或多播或空中廣播而將音訊資料傳輸至用戶端器件40，且用戶端器件40可自中間軟體單元(例如，圖2之eMBMS中間軟體單元100)請求經選擇調適集資料。

用戶端器件40可接收音訊資料(578)。舉例而言，DASH用戶端可接收所請求之音訊資料。用戶端器件40亦可解碼及呈現音訊資料(580)。解碼可由音訊解碼器46 (圖1)、MPEG-H音訊解碼器220 (圖5至圖8)、MPEG-H音訊解碼器220' (圖9)、CODEC 388 (圖11)、MPEG-H音訊解碼器440 (圖15)、場景、物件及聲道音訊解碼器486 (圖16)、場景及聲道音訊解碼器498 (圖16)或音訊解碼器534 (圖17)執行，而呈現可由音訊輸出端42 (圖1)、音訊轉譯單元232 (圖5至圖9)、音訊輸出端454 (圖15)或使用者介面550 (圖17)執行。

圖19為說明用於執行本發明之技術之另一實例方法的流程圖。在此實例中，方法被描述為由DASH用戶端及MPEG-H後設資料提取單元執行。出於實例之目的，關於DASH用戶端280 (圖8)及後設資料提取單元222 (圖8)論述圖19之實例方法。然而，應理解，可執行其他實例。舉例而言，後設資料提取單元可與MPEG-H音訊解碼器分離，如圖9之實例中所展示。

首先，在此實例中，DASH用戶端280接收資訊清單檔案(590)。舉例而言，資訊清單檔案可包含DASH之MPD檔案。DASH用戶端280可接著自資訊清單檔案提取可用性資料(592)。可根據MPEG-H 3D音訊而格式化可用性資料。因此，DASH用戶端280可將可用性資料發送至後設資料提取單元222 (594)。

後設資料提取單元222可接收可用性資料(596)。後設資料提取單元可提取可用性資料(其可指示何種類型之音訊資料可用(例如，場景、聲道、物件及/或可調式音訊資料))，且將用於呈現之此等可用資料集合的指示發送至使用者以接收指示選擇擷取哪些音訊資料集合的選擇資料(598)。回

應於選擇資料，後設資料提取單元222可接收對包括待擷取之可解碼資料之調適集的選擇(600)。詳言之，後設資料提取單元222可接收對待擷取之音訊資料之類型的選擇，且判定(使用可用性資料)經選擇類型之音訊資料與對應調適集之間的映射。後設資料提取單元222可接著將指示將自其擷取音訊資料之調適集的指令資料發送至DASH用戶端280 (602)。

因此，DASH用戶端280可接收指令資料(604)。DASH用戶端280可接著請求經選擇音訊資料(606)。舉例而言，DASH用戶端280可擷取經選擇音訊調適集之音訊資料的相對高品質集合(例如，使用頻寬調適技術)及未經選擇音訊調適集之相對低品質或最低可用位元速率表示。替代地，DASH用戶端280可僅僅擷取經選擇音訊調適集之音訊資料，而不擷取未經選擇音訊調適集之任何音訊資料。

在一些實例中，DASH用戶端280可接收經選擇音訊調適集之相對品質等級的指示。舉例而言，比較一個調適集與另一調適集之相對品質的相對品質等級。在此實例中，若如選擇資料所指示，一個調適集相較於另一調適集具有較高相對品質值，則DASH用戶端280可優先化自具有較高相對品質值之調適集的相對較高位元速率表示擷取音訊資料。

在任何狀況下，DASH用戶端280可接著接收所請求之音訊資料(608)。舉例而言，DASH用戶端280可自外部伺服器器件接收所請求之音訊資料(例如，在請求為發送至外部伺服器器件之單播請求的情況下)，或自中間軟體單元接收所請求之音訊資料(例如，在中間軟體單元首先接收音訊資料及快取經接收音訊資料以供DASH用戶端280進行後續擷取的情況下)。DASH用戶端280可接著將經接收音訊資料發送至MPEG-H音訊解碼器(610)。MPEG-H音訊解碼器可包括後設資料提取單元222 (如圖8之

實例中所展示)或與後設資料提取單元222' (如圖9之實例中所展示)分離。

以此方式，圖19之方法表示一種擷取音訊資料之方法的實例，該方法包括：接收表示複數個可用調適集之可用性資料，該等可用調適集包括一基於場景之音訊調適集及一或多個基於物件之音訊調適集；接收選擇資料，其識別該基於場景之音訊調適集與該一或多個基於物件之音訊調適集中之哪些係待擷取；及將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料。

在一或多個實例中，所描述之功能可以硬體、軟體、韌體或其任何組合實施。若以軟體實施，則該等功能可作為一或多個指令或程式碼而儲存於電腦可讀媒體上或經由電腦可讀媒體進行傳輸，且藉由基於硬體之處理單元執行。電腦可讀媒體可包括電腦可讀儲存媒體(其對應於諸如資料儲存媒體之有形媒體)或通信媒體(其包括(例如)根據通信協定促進電腦程式自一處傳送至另一處的任何媒體)。以此方式，電腦可讀媒體通常可對應於(1)為非暫時性的有形電腦可讀儲存媒體，或(2)通信媒體，諸如信號或載波。資料儲存媒體可為可由一或多個電腦或一或多個處理器存取以擷取用於實施本發明中所描述之技術之指令、程式碼及/或資料結構的任何可用媒體。電腦程式產品可包括電腦可讀媒體。

作為實例而非限制，此類電腦可讀儲存媒體可包含RAM、ROM、EEPROM、CD-ROM或其他光碟儲存器、磁碟儲存器或其他磁性儲存器件、快閃記憶體或可用以儲存呈指令或資料結構形式之所要程式碼且可由電腦存取的任何其他媒體。又，將任何連接恰當地稱為電腦可讀媒體。舉例而言，若使用同軸電纜、光纜、雙絞線、數位用戶線(DSL)或無線技術(諸如紅外線、無線電及微波)自網站、伺服器或其他遠端源傳輸指令，則

同軸電纜、光纜、雙絞線、DSL或無線技術(諸如紅外線、無線電及微波)包括於媒體之定義中。然而，應理解，電腦可讀儲存媒體及資料儲存媒體不包括連接、載波、信號或其他暫時性媒體，而實情為關於非暫時性有形儲存媒體。如本文中所使用，磁碟及光碟包括緊密光碟(CD)、雷射光碟、光學光碟、數位影音光碟(DVD)、軟性磁碟和藍光光碟，其中磁碟通常以磁性方式再現資料，而光碟使用雷射以光學方式再現資料。以上各者之組合亦應包括於電腦可讀媒體之範疇內。

可由諸如一或多個數位信號處理器(DSP)、通用微處理器、特殊應用積體電路(ASIC)、場可程式化邏輯陣列(FPGA)或其他等效整合式或離散邏輯電路之一或多個處理器來執行指令。因此，如本文中所使用之術語「處理器」可指上述結構或適合於實施本文中所描述之技術的任何其他結構中之任一者。另外，在一些態樣中，本文中所描述之功能性可提供於經組態用於編碼及解碼之專用硬體及/或軟體模組內，或併入組合式編碼解碼器中。又，該等技術可完全實施於一或多個電路或邏輯元件中。

本發明之技術可在廣泛多種器件或裝置中實施，該等器件或裝置包括無線手持機、積體電路(IC)或IC集合(例如，晶片組)。在本發明中描述各種組件、模組或單元以強調經組態以執行所揭示技術之器件的功能態樣，但未必要求由不同硬體單元來實現。實情為，如上文所描述，各種單元可結合合適的軟體及/或韌體而組合於編碼解碼器硬體單元中或由互操作性硬體單元之集合提供，該等硬體單元單元包括如上文所描述之一或多個處理器。

各種實例已於本文中予以描述。此等及其他實例係在以下申請專利範圍之範疇內。

【符號說明】

10	系統
20	內容準備器件
22	音訊源
24	視訊源
26	音訊編碼器
28	視訊編碼器
30	囊封單元
32	輸出介面
40	用戶端器件
42	音訊輸出端
44	視訊輸出端
46	音訊解碼器
48	視訊解碼器
50	解囊封單元
52	擷取單元
54	網路介面
60	伺服器器件
62	儲存媒體
64	多媒體內容
66	資訊清單檔案
68A	表示
68N	表示

70	請求處理單元
72	網路介面
74	網路
100	eMBMS 中間軟體單元/eMBMS 中間軟體
102	伺服器單元/代理伺服器
104	快取記憶體
106	eMBMS 接收單元
110	超文字傳送協定動態自適應串流(DASH)用戶端
112	媒體應用程式
120	多媒體內容
122	媒體呈現描述(MPD)
122B	媒體呈現描述(MPD)
124A	表示
124BA	表示
124BB	表示
124BC	表示
124BD	表示
124N	表示
126	標頭資料
126BA	標頭資料類型0
126BB	標頭資料類型1
126BC	標頭資料類型2
126BD	標頭資料類型3

128A	區段
128B	區段
128BA	類型0區段
128BB	類型0區段
128BN	類型0區段
128CA	類型1區段
128CB	類型1區段
128CN	類型1區段
128DA	類型2區段
128DB	類型2區段
128DN	類型2區段
128EA	類型3區段
128EB	類型3區段
128EN	類型3區段
128N	區段
130	標頭資料
132A	區段
132B	區段
132N	區段
150	媒體檔案
152	檔案類型方塊
154	電影(MOOV)方塊
156	電影標頭(MVHD)方塊

158	播放軌(TRAK)方塊
160	電影延伸(MVEX)方塊
162	區段索引(sidx/SIDX)方塊
164	電影片段方塊/電影片段
166	電影片段隨機存取(MFRA)方塊
200	系統
202	基於物件之內容
202'	基於音訊之內容
204	後設資料
206	場景資料
208	聲道資料
210	物件資料
212	移動圖像專家組(MPEG-H)音訊編碼器
214	音訊編碼器
216	多工器
218	音訊資料/串流
220	移動圖像專家組(MPEG-H)音訊解碼器
220'	移動圖像專家組(MPEG-H)音訊解碼器
222	後設資料提取單元
224	場景資料提取單元
226	物件資料提取單元
228	使用者介面
230	應用程式設計介面(API)

232	音訊轉譯單元
240	串流
242A	串流
242N	串流
250	系統
252	媒體伺服器
254	經編碼後設資料
256	場景及聲道調適集
258A	表示
258M	表示
260A	物件調適集
260N	物件調適集
262A	表示
262P	表示
264A	表示
264Q	表示
270	媒體呈現描述(MPD)
272	媒體呈現描述(MPD)
274	元素
274'	互動
276	內容傳遞網路
278	移動圖像專家組(MPEG-H)音訊解碼器
280	超文字傳送協定動態自適應串流(DASH)用戶端

282	選擇單元
284	下載及切換單元
350	系統
352	媒體伺服器
354	廣播伺服器
356	廣播傳輸器
358	超文字傳送協定內容傳遞網路(HTTP CDN)
360A	伺服器器件
360N	伺服器器件
362	eNode-B
364A	用戶端器件
364B	用戶端器件
364C	用戶端器件
364N	用戶端器件
370	系統
372	統一時間碼(UTC)源
374	本端統一時間碼(UTC)源
376	廣播超文字傳送協定動態自適應串流(DASH)伺服器
378	廣播檔案輸送封裝器
380	廣播檔案輸送接收器
382	媒體呈現描述(MPD)及媒體資料
382'	媒體呈現描述(MPD)及/或媒體資料
384	超文字傳送協定動態自適應串流(DASH)播放器

386	內容傳遞網路(CDN)
388	編碼解碼器(CODEC)
390	經時間對準之經壓縮媒體資料
392	經時間對準之媒體樣本及像素
400	協定模型
402	實體層
404	網際網路協定資料
406	統一資料報協定資料及傳輸控制協定資料
408	傳遞層
410	服務管理資料
420	內容傳遞網路(CDN)
421	媒體呈現描述(MPD)
422	基於場景之可調式音訊內容/視訊後設資料處理單元
424	經編碼後設資料/視訊解碼器
426	基於場景之音訊、基底層調適集/應用程式設計介面(API)
428A	基於場景之音訊、增強層調適集
428N	基於場景之音訊、增強層調適集
430	超文字傳送協定動態自適應串流(DASH)用戶端
432	選擇單元
434	下載及切換單元
440	移動圖像專家組(MPEG-H)音訊解碼器
442	後設資料提取單元
444	場景資料提取單元

446	可調式音訊層解碼單元
448	使用者介面
450	應用程式設計介面(API)
452	音訊轉譯單元
454	音訊輸出端
460	媒體呈現描述(MPD)
462	移動圖像專家組(MPEG-H)音訊解碼器
464	調適集
466	移動圖像專家組(MPEG-H)音訊解碼器
470	發送器
472	視訊編碼器
474	音訊編碼器
476	檔案格式囊封器
478	超文字傳送協定動態自適應串流(DASH)分段器
480	單向輸送即時物件傳遞(ROUTE)發送器
482	接收器
484	視訊解碼器
486	場景、物件及聲道音訊解碼器
488	檔案格式剖析器
490	超文字傳送協定動態自適應串流(DASH)用戶端
492	單向輸送即時物件傳遞(ROUTE)接收器
494	接收器
496	視訊解碼器

498	場景及聲道音訊解碼器
500	檔案格式剖析器
502	超文字傳送協定動態自適應串流(DASH)用戶端
504	單向輸送即時物件傳遞(ROUTE)接收器
506	視訊資料
508	音訊資料
508'	音訊資料
510	經囊封資料/區段
510'	區段
512	區段
514	位元串流
514'	位元串流
520	用戶端器件
522	網路介面
524	超文字傳送協定動態自適應串流(DASH)用戶端
530	音訊控制器
532	音訊後設資料處理單元
534	音訊解碼器
536	應用程式設計介面(API)
538	音訊轉譯器
550	使用者介面
560	動作
562	動作

564	動作
566	動作
568	動作
570	動作
572	動作
574	動作
576	動作
578	動作
580	動作
590	動作
592	動作
594	動作
596	動作
598	動作
600	動作
602	動作
604	動作
606	動作
608	動作
610	動作
700	多層音訊資料
702	基底子層/第一層/基底層
704	第一增強子層/第一增強層/第二層

706	第二增強子層/第二增強層/第三層
710	多層音訊資料
712	基底子層/基底層
714	第一增強層/第二層
716	第二增強層/第三層

【發明申請專利範圍】

【第1項】

一種擷取音訊資料之方法，該方法包含：

接收表示複數個可用調適集之可用性資料，該等可用調適集包括一或多個基於場景之音訊調適集及一或多個基於物件之音訊調適集，該一或多個基於物件之音訊調適集包括音訊物件之音訊資料及表示該等音訊物件之位置座標之後設資料，且該一或多個基於場景之音訊調適集包括使用球諧係數表示一音場之音訊資料且包含一或多個可用可調式音訊調適集，該一或多個可用可調式音訊調適集之每一者對應於可調式音訊資料之各別層；

接收選擇資料，其識別該一或多個基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及

將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料。

【第2項】

如請求項1之方法，其中該串流用戶端經組態以根據界定一第一資料格式之一串流協定而擷取由該選擇資料識別的該等調適集中之每一者的該資料，且其中根據不同於由該串流協定界定之該第一資料格式的一第二資料格式而格式化該可用性資料。

【第3項】

如請求項2之方法，其中該串流協定包含HTTP動態自適應串流(DASH)，其中該第一資料格式包含ISO基本媒體檔案格式(ISO BMFF)、ISO BMFF之擴展或MPEG-2輸送串流(MPEG-2 TS)中之一者，且其中該第二資料格式包含MPEG-H 3D音訊。

【第4項】

如請求項1之方法，其進一步包含接收表示待擷取之該等調適集中之至少一者的一相對品質之品質資料，其中提供該指令資料包含形成該指令資料以造成該串流用戶端擷取該等調適集中之對應於該相對品質的該至少一者的一表示。

【第5項】

如請求項4之方法，其中接收該品質資料包含接收該等調適集中之該至少一者的一相對音量，該方法進一步包含：回應於判定該等調適集中之該至少一者的該相對音量高於由該選擇資料識別之其他調適集的相對音量，形成該指令資料以造成該串流用戶端擷取該至少一個調適集之一表示，該表示相較於由該選擇資料識別之該等其他調適集之表示的位元速率具有一相對較高位元速率。

【第6項】

如請求項1之方法，其進一步包含在接收該指令資料之前藉由該串流用戶端擷取該等可用調適集中之每一者的資料。

【第7項】

如請求項第6項之方法，其進一步包含回應於該指令資料而藉由該串流用戶端進行以下操作：

判定該等可用調適集中之不應擷取之至少一者；

判定在接收該指令資料之前已分配至該等可用調適集中之該至少一者的頻寬之一量；

根據該指令資料而將該所判定量之頻寬分配至待擷取之該等調適集中之一或多者；及

基於該經分配頻寬量而調整待擷取之該等調適集中之一或多者的表示選擇。

【第8項】

如請求項1之方法，其進一步包含藉由該串流用戶端接收包括該可用性資料之一資訊清單檔案。

【第9項】

如請求項8之方法，其中該資訊清單檔案包含一媒體呈現描述(MPD)。

【第10項】

如請求項1之方法，其中接收該可用性資料包含自該串流用戶端接收該可用性資料。

【第11項】

如請求項1之方法，其中接收該選擇資料包含自一使用者介面接收該選擇資料。

【第12項】

如請求項1之方法，其中提供該指令資料包含提供該指令資料以造成該串流用戶端僅擷取由該選擇資料識別之該等調適集的資料及阻止擷取未由該選擇資料識別之調適集的任何資料。

【第13項】

如請求項1之方法，其中提供該指令資料包含提供該指令資料以造成該串流用戶端針對由該選擇資料識別之該等調適集執行頻寬調適，及自未由該選擇資料識別之該等可用調適集之具有最低可用位元速率的表示擷取資料。

【第14項】

如請求項1之方法，其進一步包含形成該指令資料以識別待針對由該選擇資料識別的該等調適集中之每一者擷取的表示。

【第15項】

如請求項1之方法，其進一步包含根據該指令資料藉由該串流用戶端發送各別HTTP GET或部分GET請求以擷取資料。

【第16項】

如請求項1之方法，其中該串流用戶端包含一HTTP動態自適應串流(DASH)用戶端。

【第17項】

如請求項1之方法，其中該串流用戶端包括於一中間軟體單元中，該中間軟體單元進一步包括一廣播或多播接收單元及一代理伺服器，該方法進一步包含藉由該串流用戶端根據該指令資料經由單播自該代理伺服器擷取快速获取媒體資料。

【第18項】

一種用於擷取音訊資料之器件，該器件包含：

一或多個處理器，該一或多個處理器經組態以：

接收表示複數個可用調適集之可用性資料，該等可用調適集包括一或多個基於場景之音訊調適集及一或多個基於物件之音訊調適集，該一或多個基於物件之音訊調適集包括音訊物件之音訊資料及表示該等音訊物件之位置座標之後設資料，且該一或多個基於場景之音訊調適集包括使用球諧係數表示一音場之音訊資料且包含一或多個可用可調式音訊調適集，該一或多個可用可調式音訊調適集之每一者對應於可調式音訊資料之各別

層；

接收選擇資料，其識別該一或多個基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及

將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料；及

一記憶體，其經組態以儲存該等音訊調適集之該所擷取資料。

【第19項】

如請求項18之器件，其中該串流用戶端包含一HTTP動態自適應串流(DASH)用戶端，該DASH用戶端經組態以將針對包括標頭資料之媒體資料的請求發送至一伺服器器件。

【第20項】

如請求項19之器件，其中該DASH用戶端經組態以自該伺服器器件接收包括該標頭資料之該媒體資料。

【第21項】

如請求項18之器件，其中該串流用戶端經組態以根據界定一第一資料格式之一串流協定而擷取由該選擇資料識別的該等調適集中之每一者的該資料，且其中根據不同於由該串流協定界定之該第一資料格式的一第二資料格式而格式化該可用性資料。

【第22項】

如請求項18之器件，其進一步包含一使用者介面，該使用者介面經組態以基於一使用者定義之動作而接收使用者提供之選擇資料及將該選擇資料提供至該一或多個處理器。

【第23項】

如請求項18之器件，其進一步包含一移動圖像專家組(MPEG-H)音訊解碼器，該MPEG-H音訊解碼器經組態以解碼基於場景之音訊資料、基於聲道之音訊資料或基於物件之音訊資料中之至少一者。

【第24項】

如請求項18之器件，其中該一或多個處理器包含一MPEG-H音訊解碼器。

【第25項】

如請求項18之器件，其中該一或多個處理器包含一MPEG-H音訊解碼器之一後設資料處理單元。

【第26項】

如請求項18之器件，其進一步包含一MPEG-H音訊解碼器，其中該一或多個處理器包含與該MPEG-H音訊解碼器分離之一後設資料處理單元。

【第27項】

一種用於擷取音訊資料之器件，該器件包含：

用於接收表示複數個可用調適集之可用性資料的構件，該等可用調適集包括一或多個基於場景之音訊調適集及一或多個基於物件之音訊調適集，該一或多個基於物件之音訊調適集包括音訊物件之音訊資料及表示該等音訊物件之位置座標之後設資料，且該一或多個基於場景之音訊調適集包括使用球譜係數表示一音場之音訊資料且包含一或多個可用可調式音訊調適集，該一或多個可用可調式音訊調適集之每一者對應於可調式音訊資料之各別層；

用於接收選擇資料的構件，該選擇資料識別該一或多個基於場景之

音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及

用於將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料的構件。

【第28項】

如請求項27之器件，其中該串流用戶端經組態以根據界定一第一資料格式之一串流協定而擷取由該選擇資料識別的該等調適集中之每一者的該資料，且其中根據不同於由該串流協定界定之該第一資料格式的一第二資料格式而格式化該可用性資料。

【第29項】

如請求項27之器件，其中用於接收該選擇資料的該構件包含用於自一使用者介面接收該選擇資料的構件。

【第30項】

如請求項27之器件，其中用於提供該指令資料的該構件包含用於提供該指令資料以造成該串流用戶端僅擷取由該選擇資料識別的該等調適集之資料並阻止擷取未由該選擇資料識別的調適集之任何資料的構件。

【第31項】

如請求項27之器件，其中該串流用戶端包含一HTTP動態自適應串流(DASH)用戶端，該DASH用戶端經組態以將針對包括標頭資料之媒體資料的請求發送至一伺服器器件。

【第32項】

一種儲存有指令於其上之非暫時性電腦可讀儲存媒體，該等指令在執行時造成一處理器進行以下操作：

接收表示複數個可用調適集之可用性資料，該等可用調適集包括一

或多個基於場景之音訊調適集及一或多個基於物件之音訊調適集，該一或多個基於物件之音訊調適集包括音訊物件之音訊資料及表示該等音訊物件之位置座標之後設資料，且該一或多個基於場景之音訊調適集包括使用球諧係數表示一音場之音訊資料且包含一或多個可用可調式音訊調適集，該一或多個可用可調式音訊調適集之每一者對應於可調式音訊資料之各別層；

接收選擇資料，其識別該一或多個基於場景之音訊調適集及該一或多個基於物件之音訊調適集中之哪些係待擷取；及

將指令資料提供至一串流用戶端以造成該串流用戶端擷取由該選擇資料識別的該等調適集中之每一者的資料。

【第33項】

如請求項32之非暫時性電腦可讀儲存媒體，其中該串流用戶端經組態以根據界定一第一資料格式之一串流協定而擷取由該選擇資料識別的該等調適集中之每一者的該資料，且其中根據不同於由該串流協定界定之該第一資料格式的一第二資料格式而格式化該可用性資料。

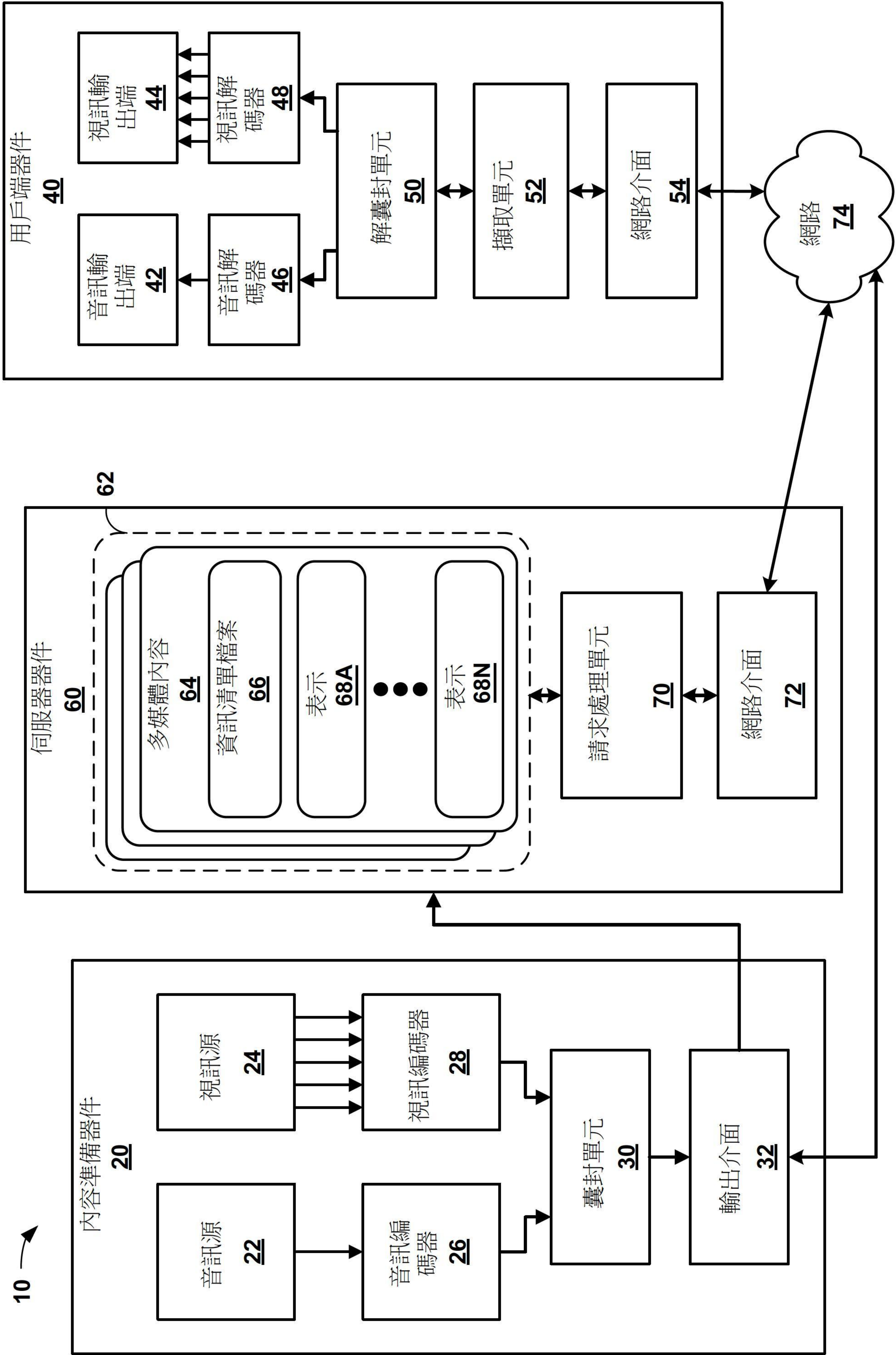
【第34項】

如請求項32之非暫時性電腦可讀儲存媒體，其中造成該處理器接收該選擇資料之該等指令包含造成該處理器自一使用者介面接收該選擇資料之指令。

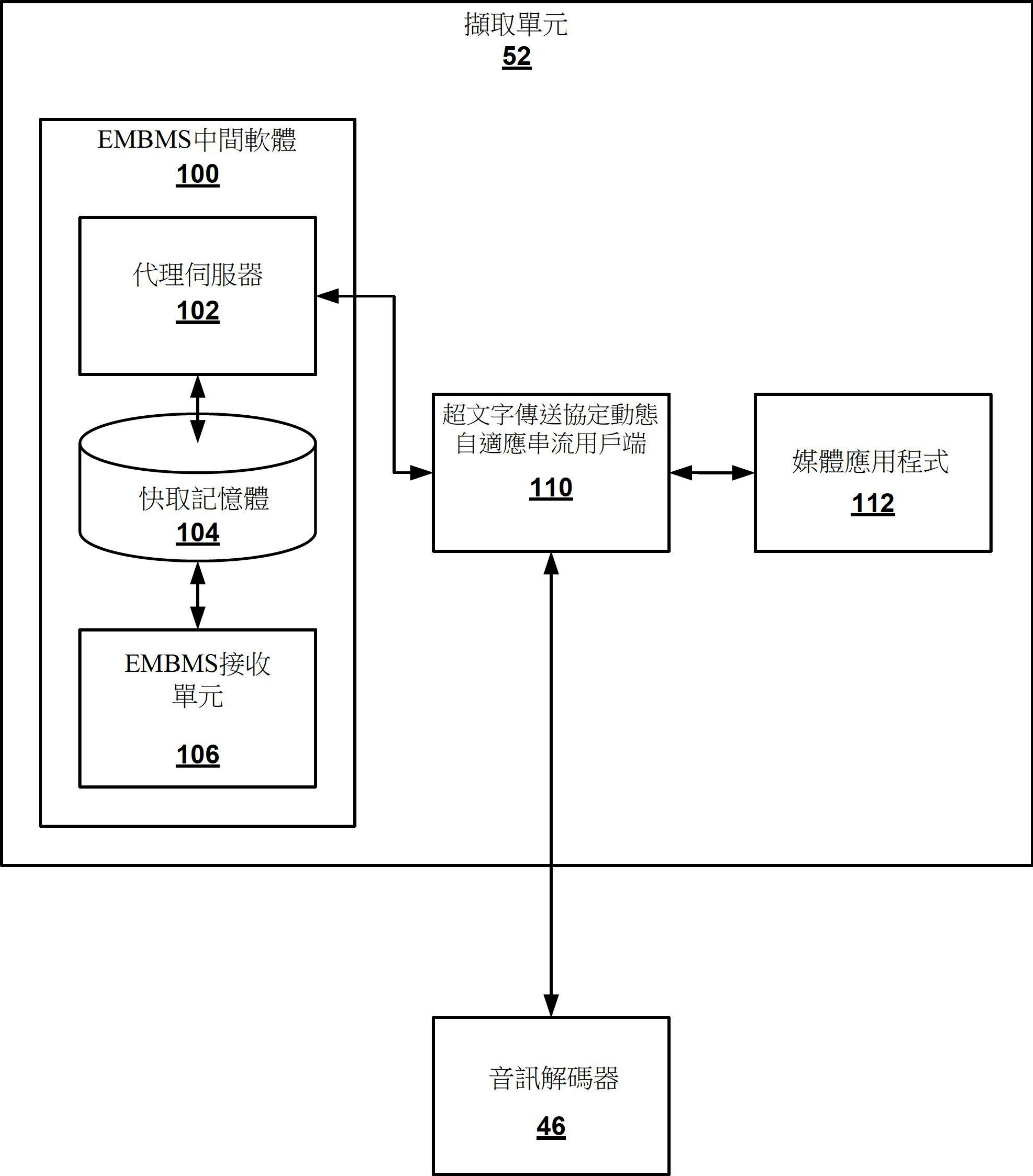
【第35項】

如請求項32之非暫時性電腦可讀儲存媒體，其中該串流用戶端包含一HTTP動態自適應串流(DASH)用戶端，該DASH用戶端經組態以將針對包括標頭資料之媒體資料的請求發送至一伺服器器件。

【發明圖式】

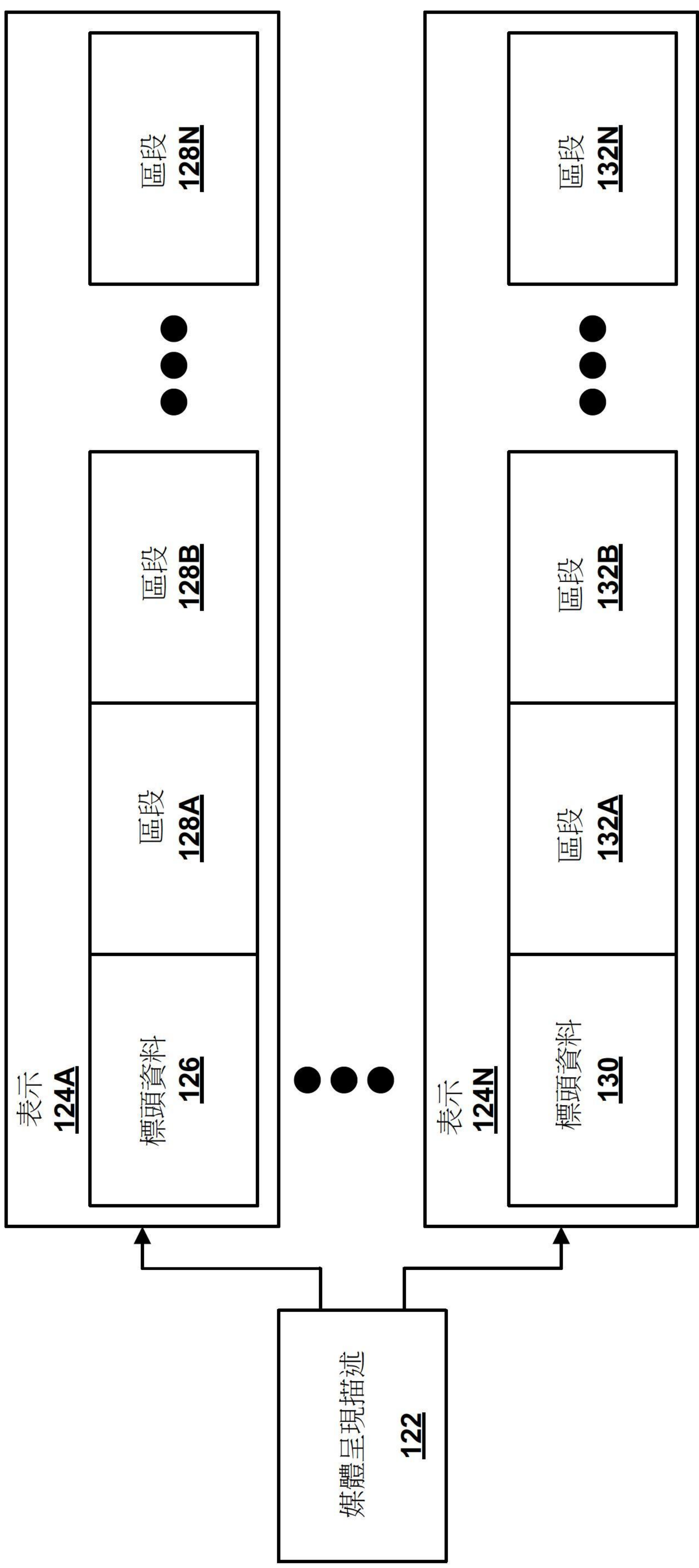


【圖1】

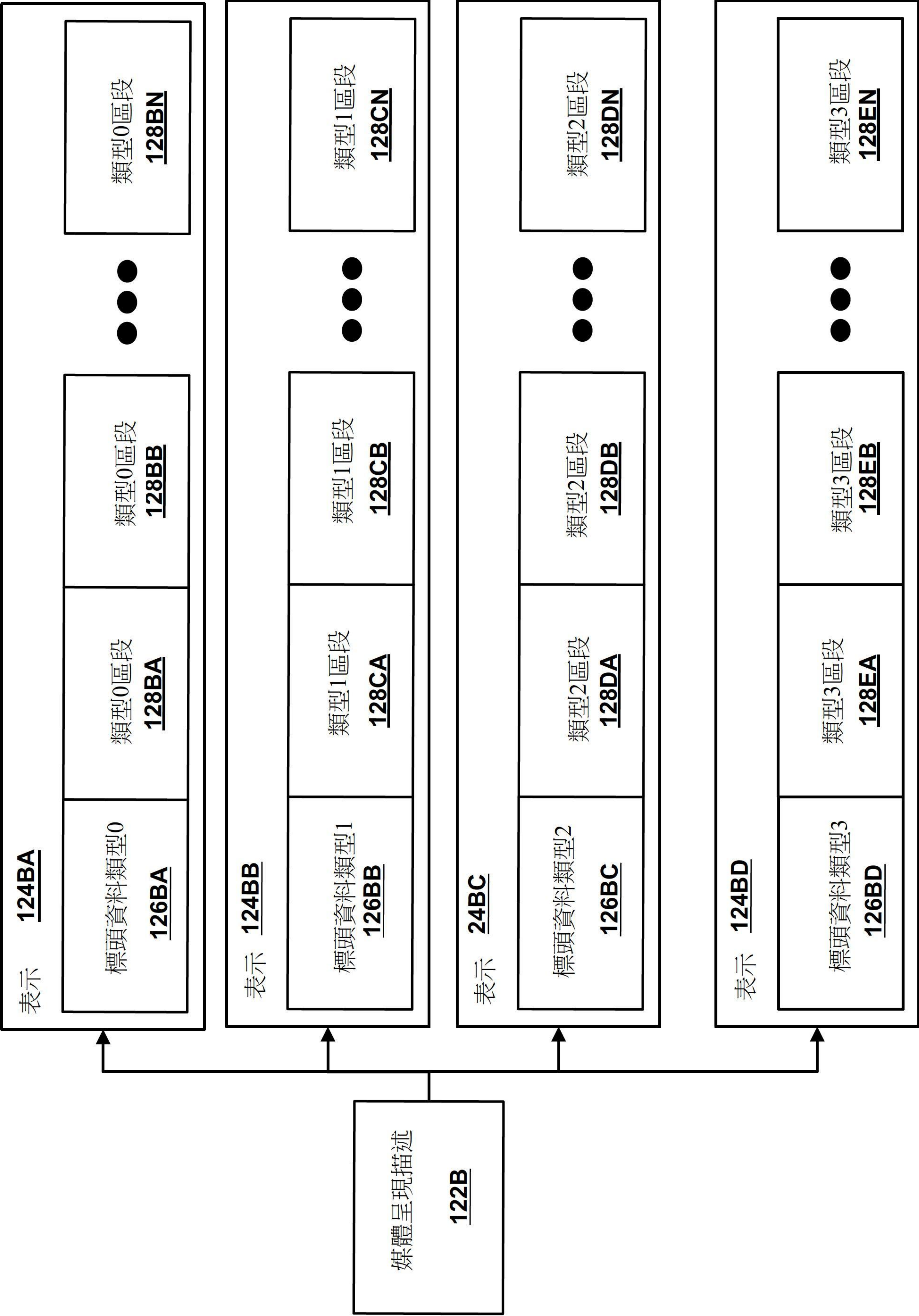


【圖2】

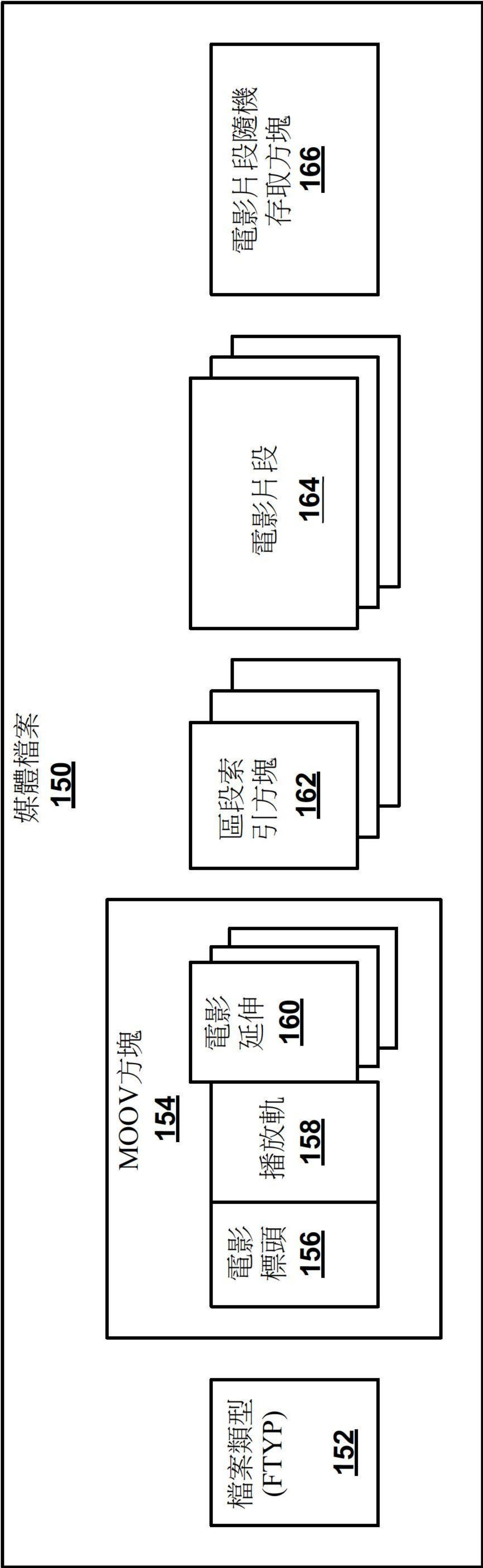
120



【圖3A】

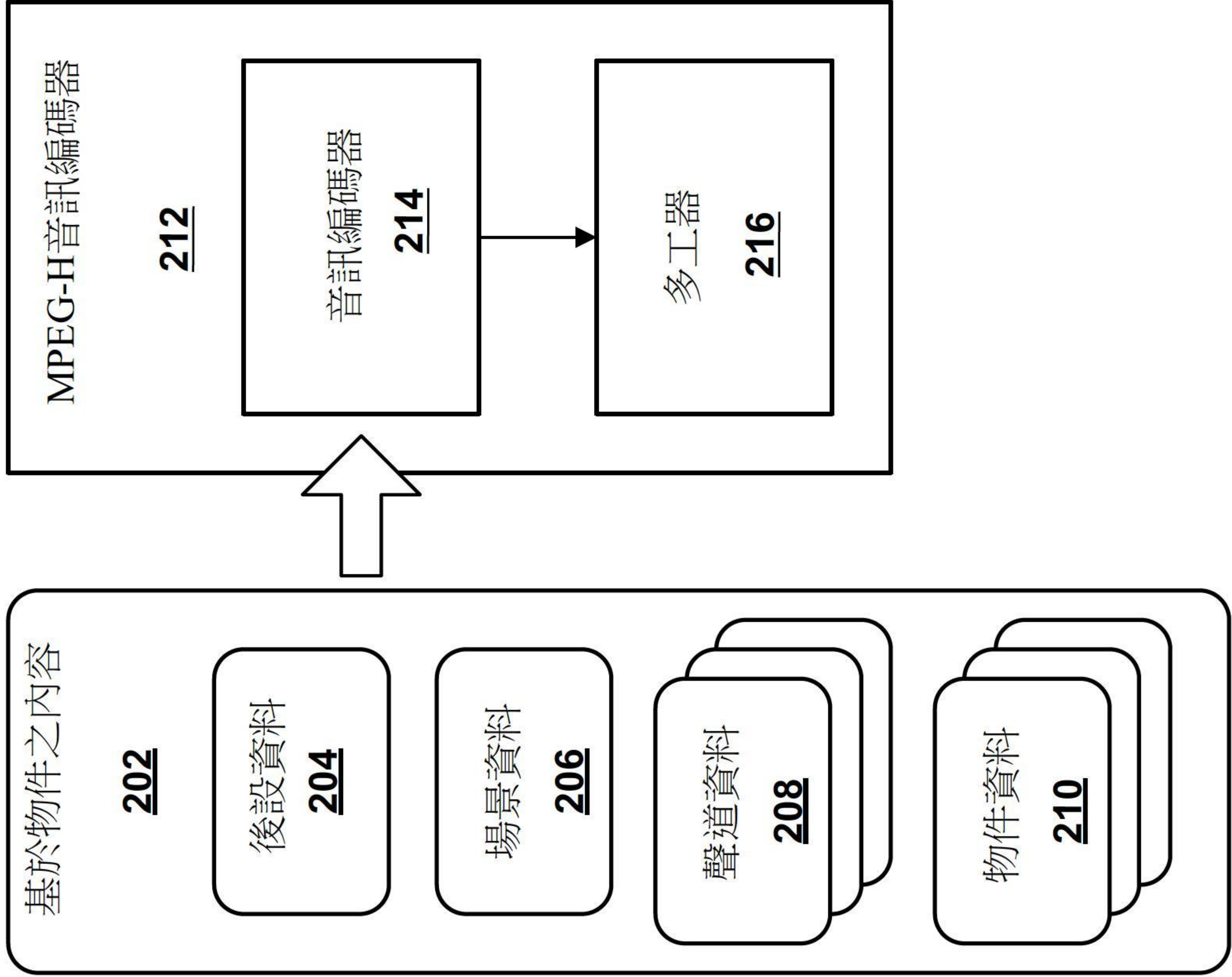


【圖3B】

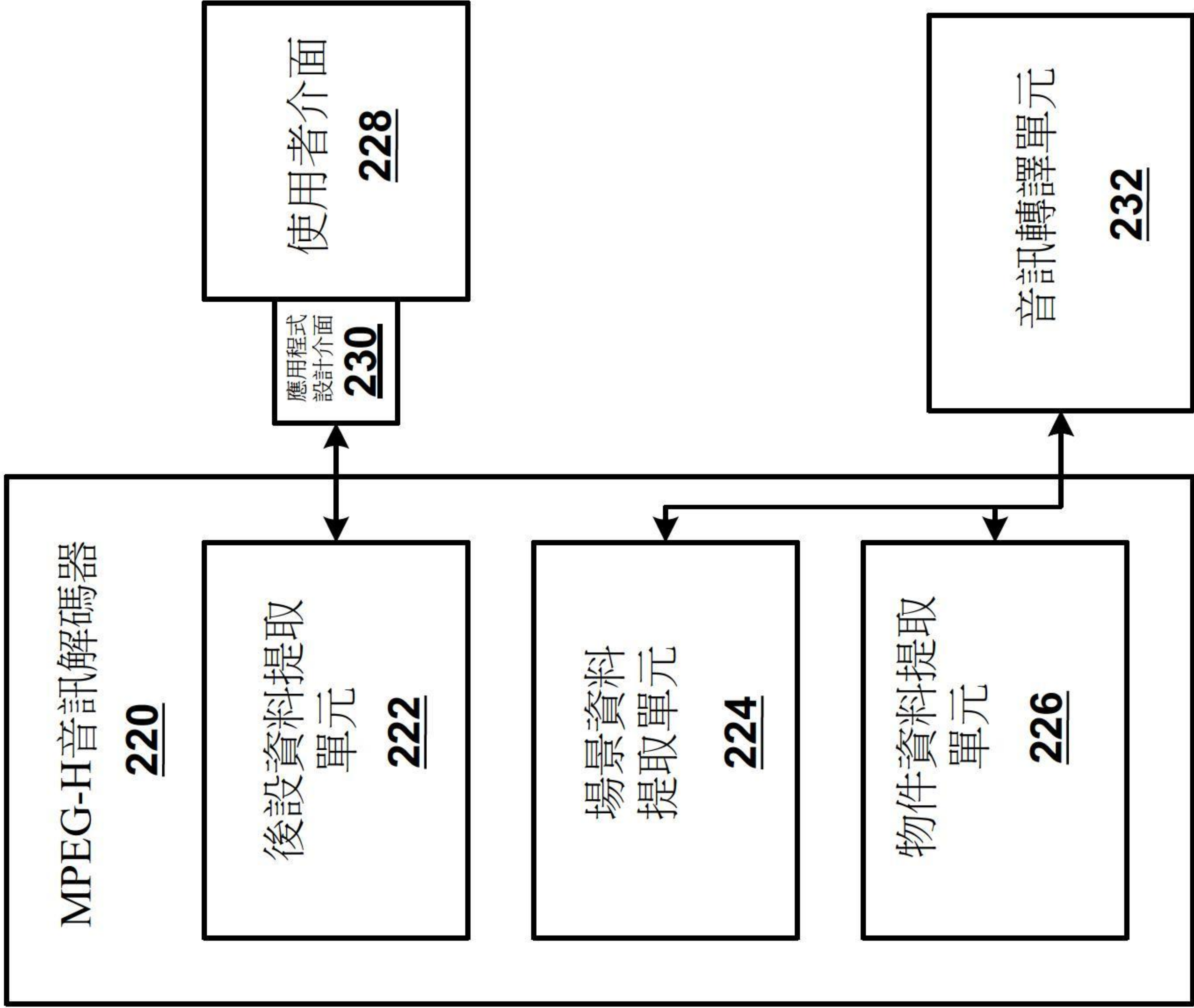


【圖4】

200 →

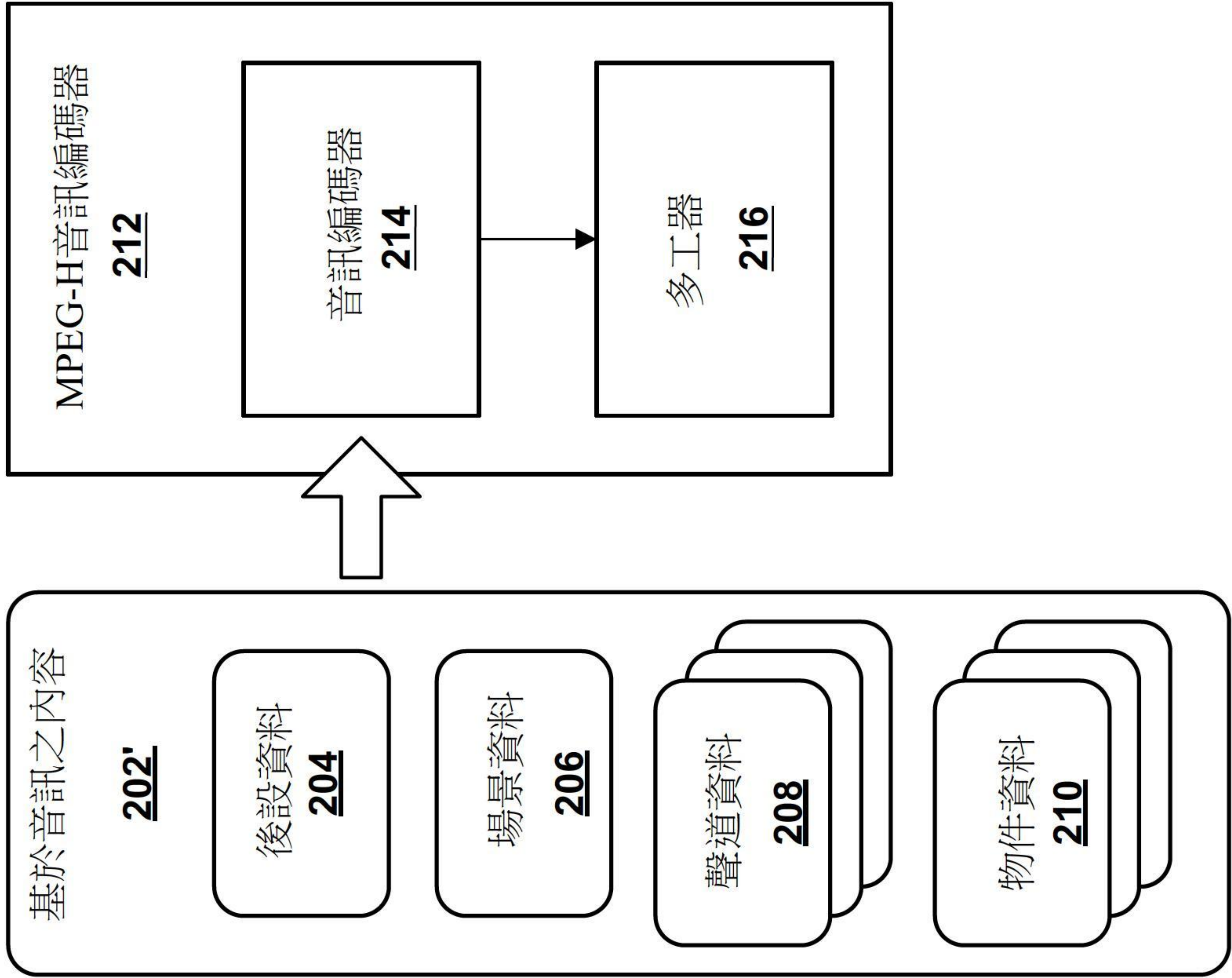


← 218

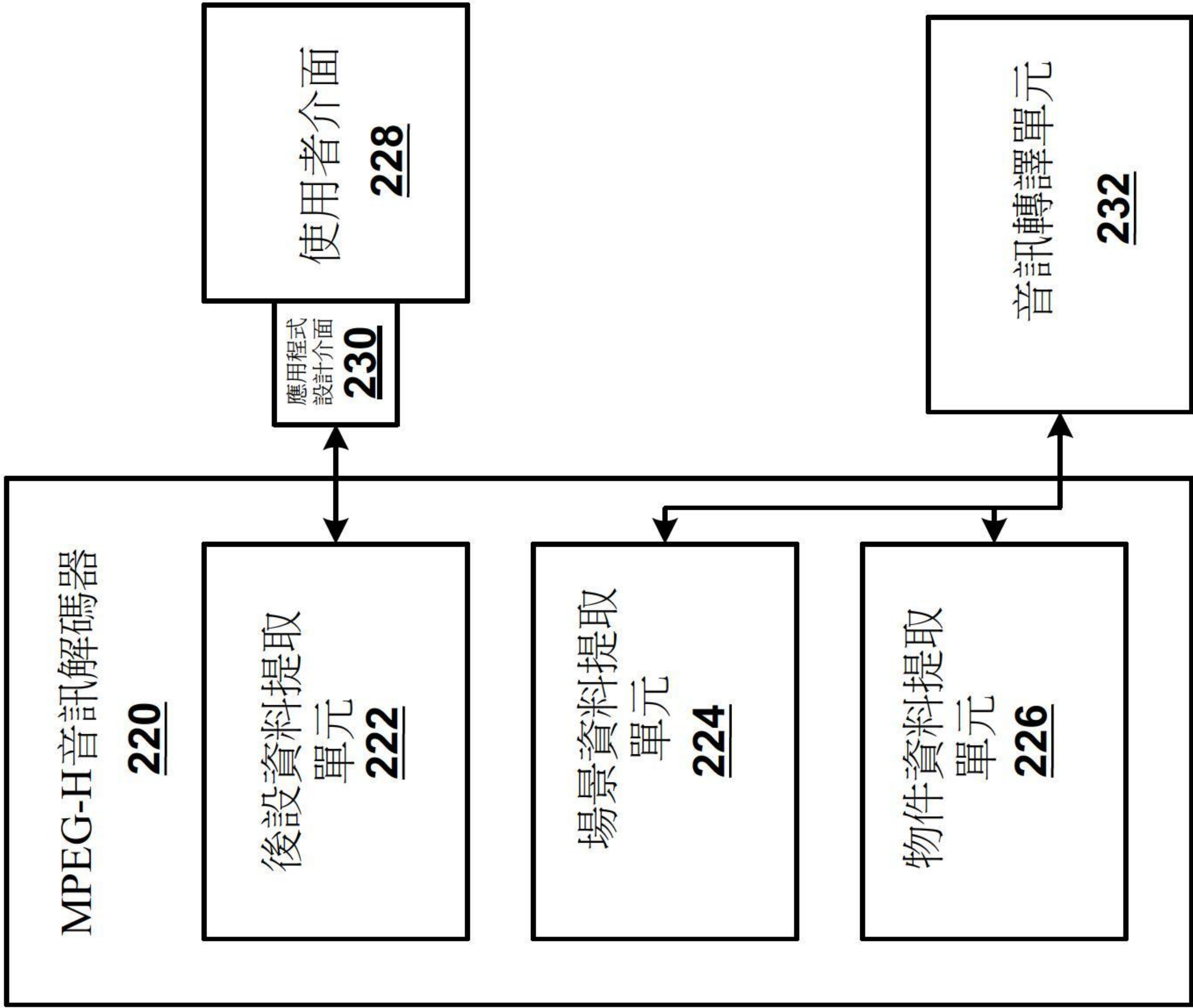


【圖 5A】

200 →

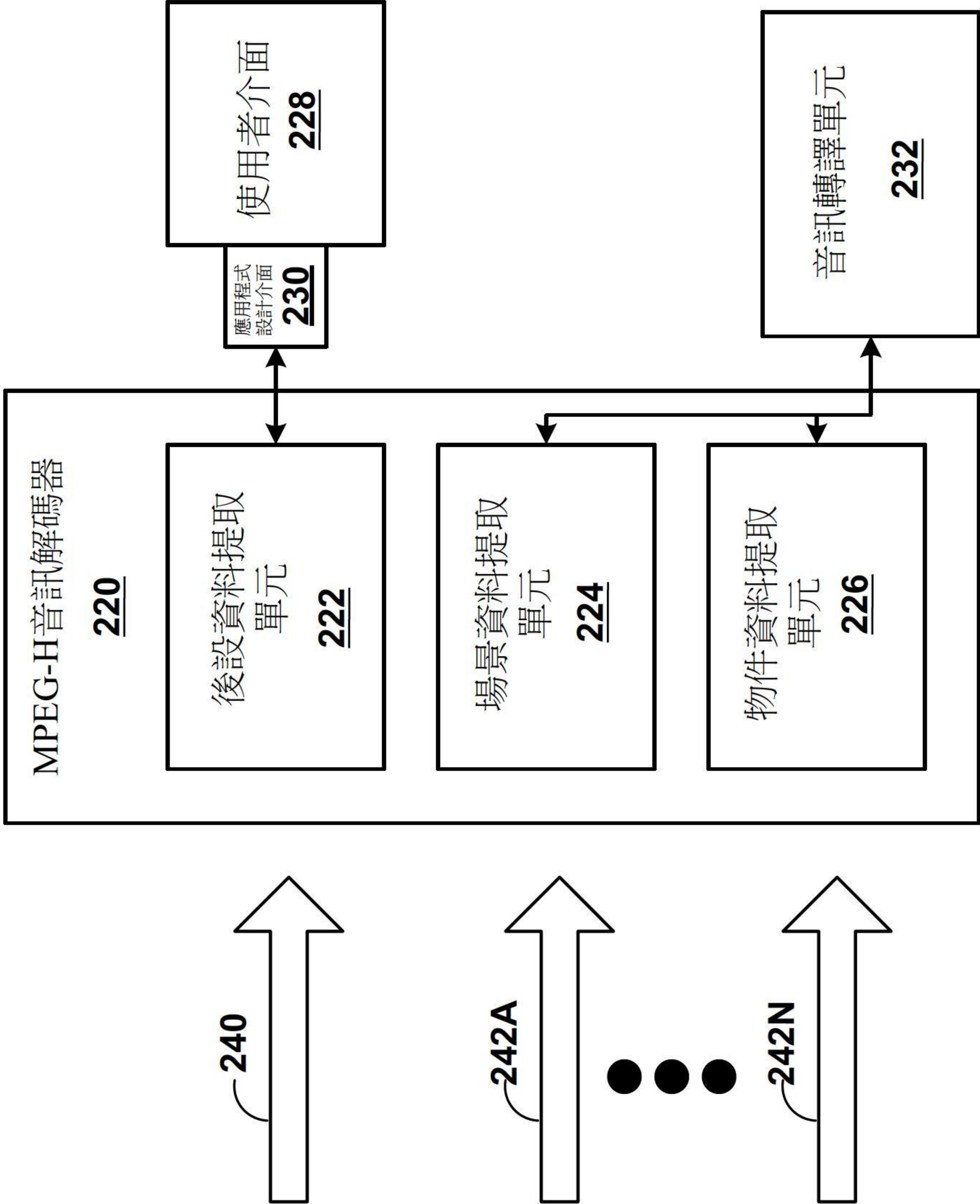
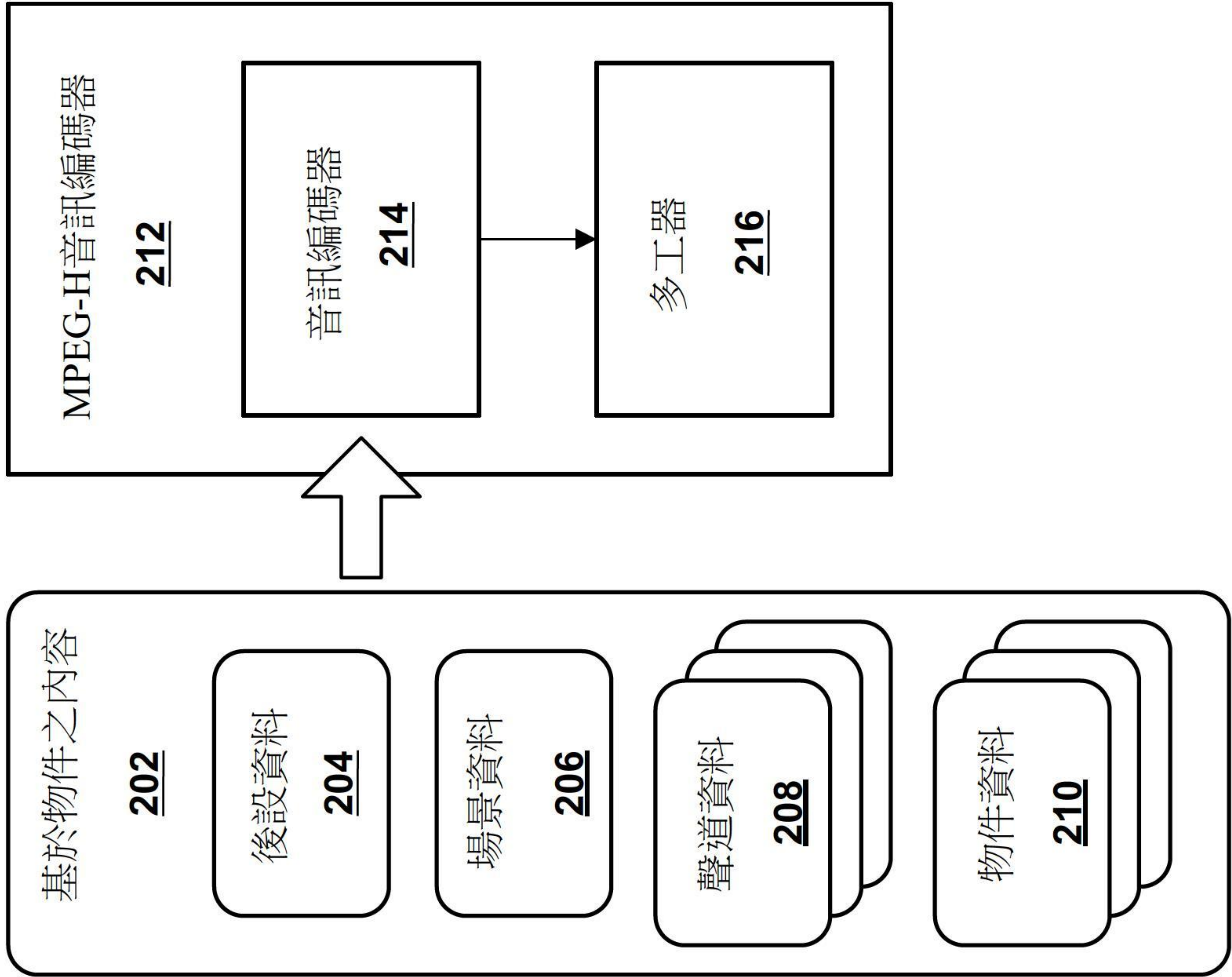


218

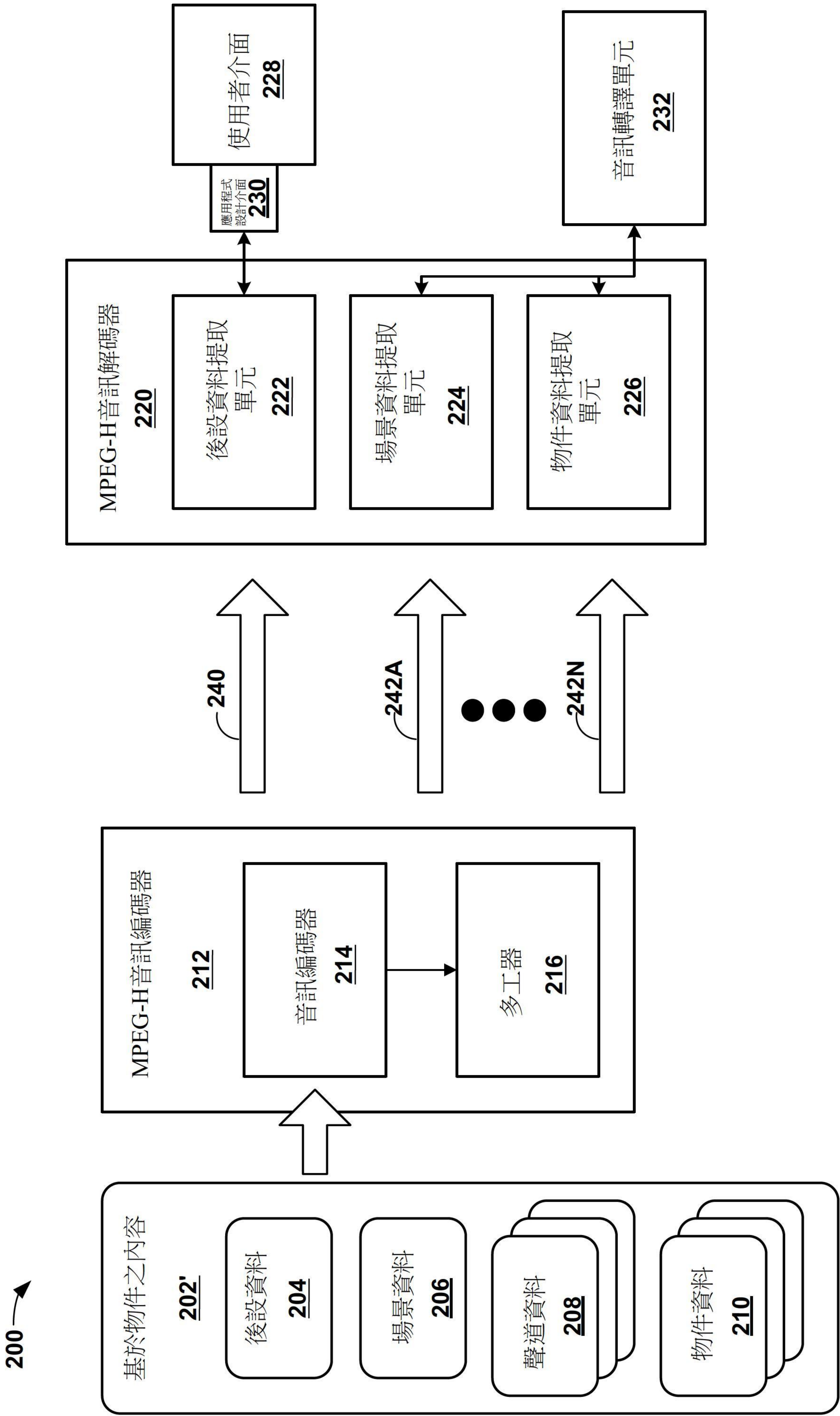


【圖5B】

200

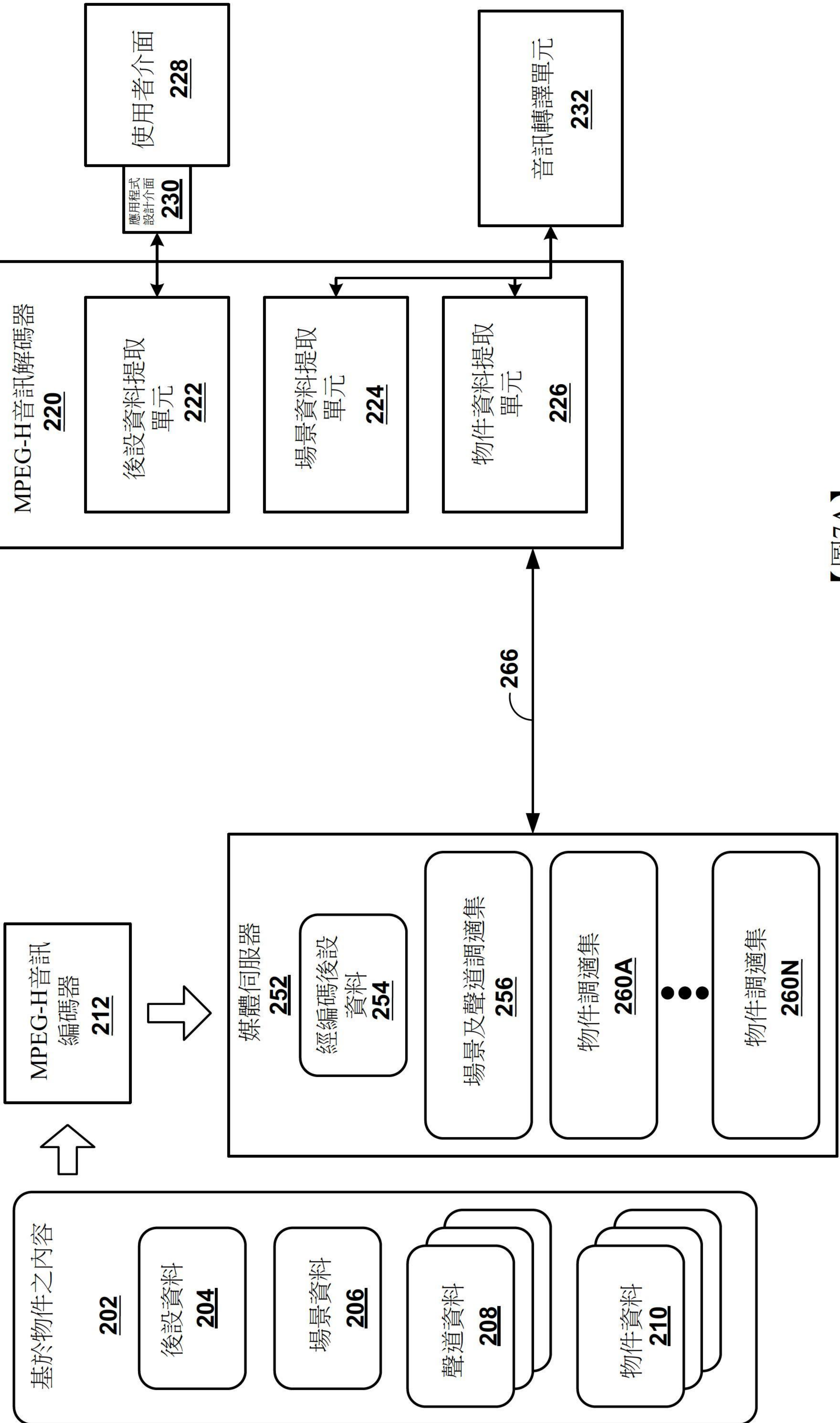


【圖6A】

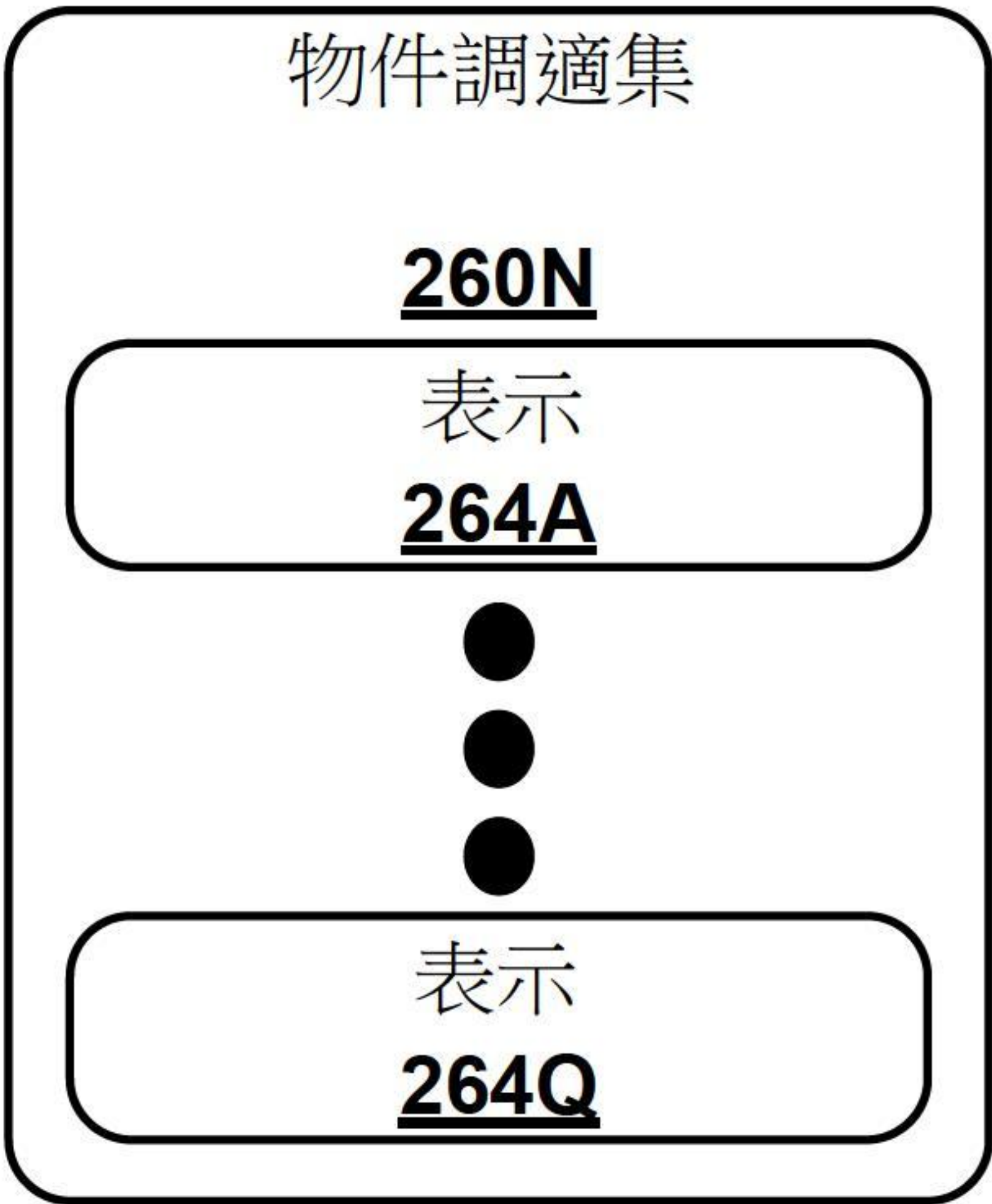
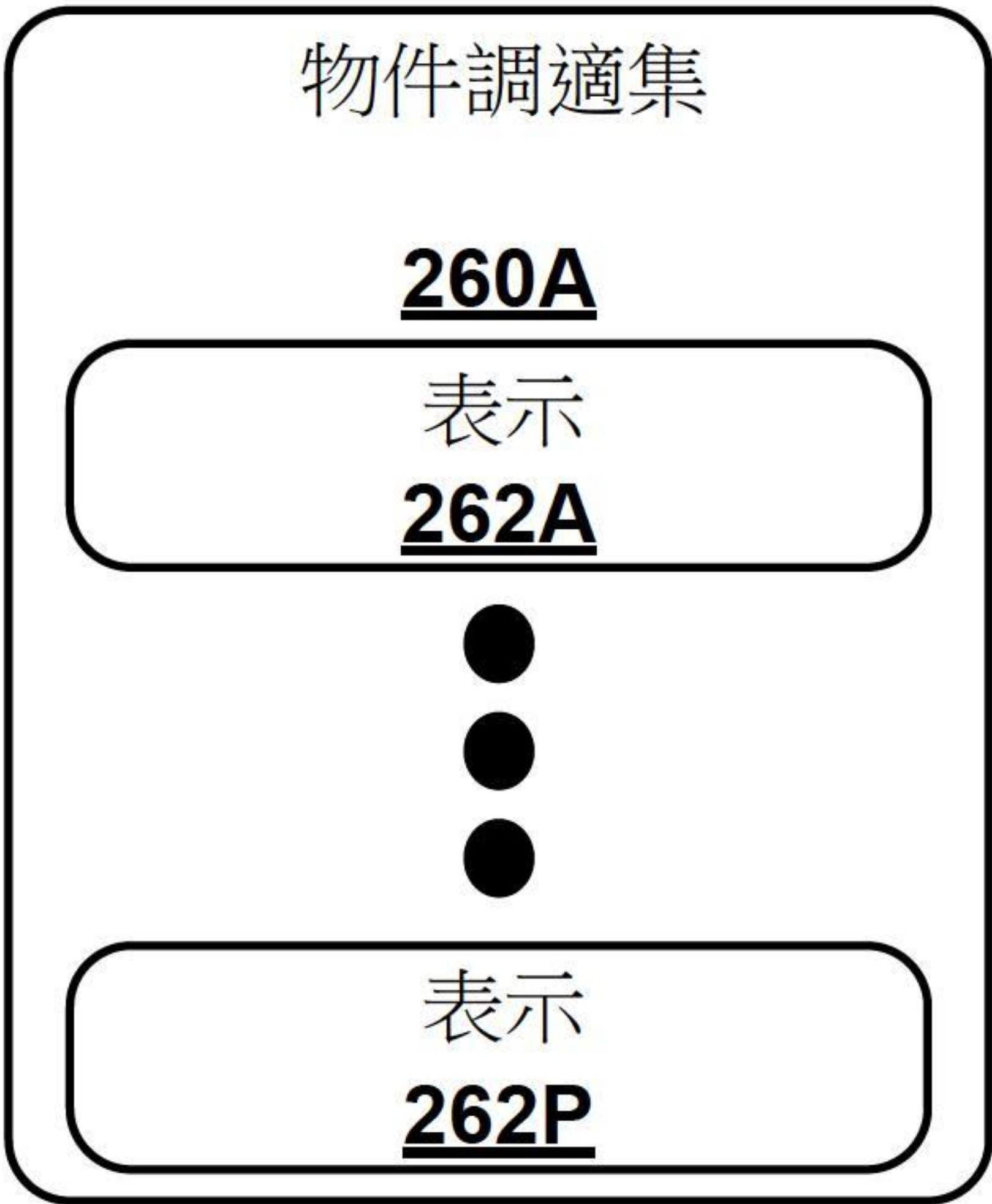
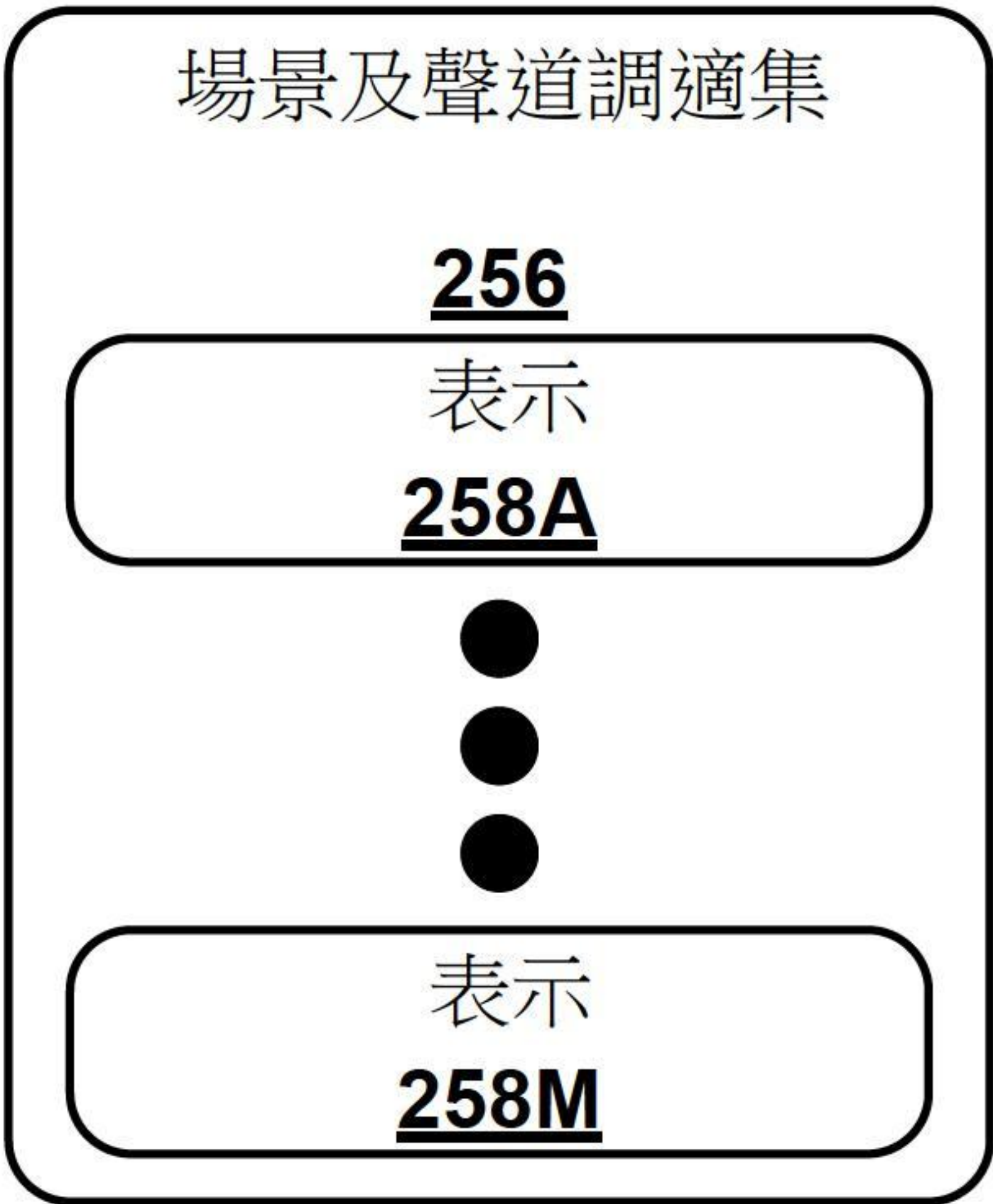


【圖6B】

250 →

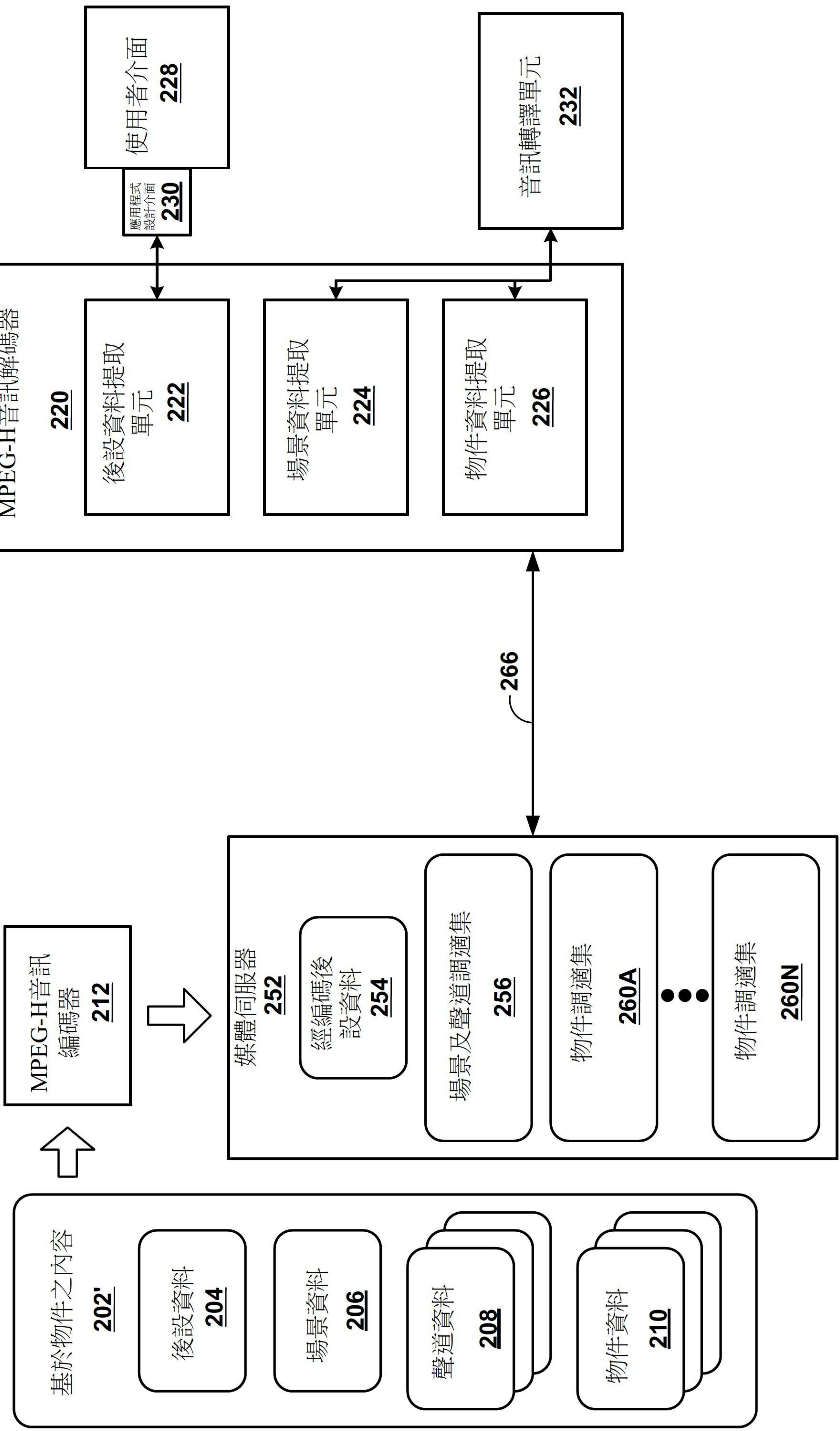


【圖7A】

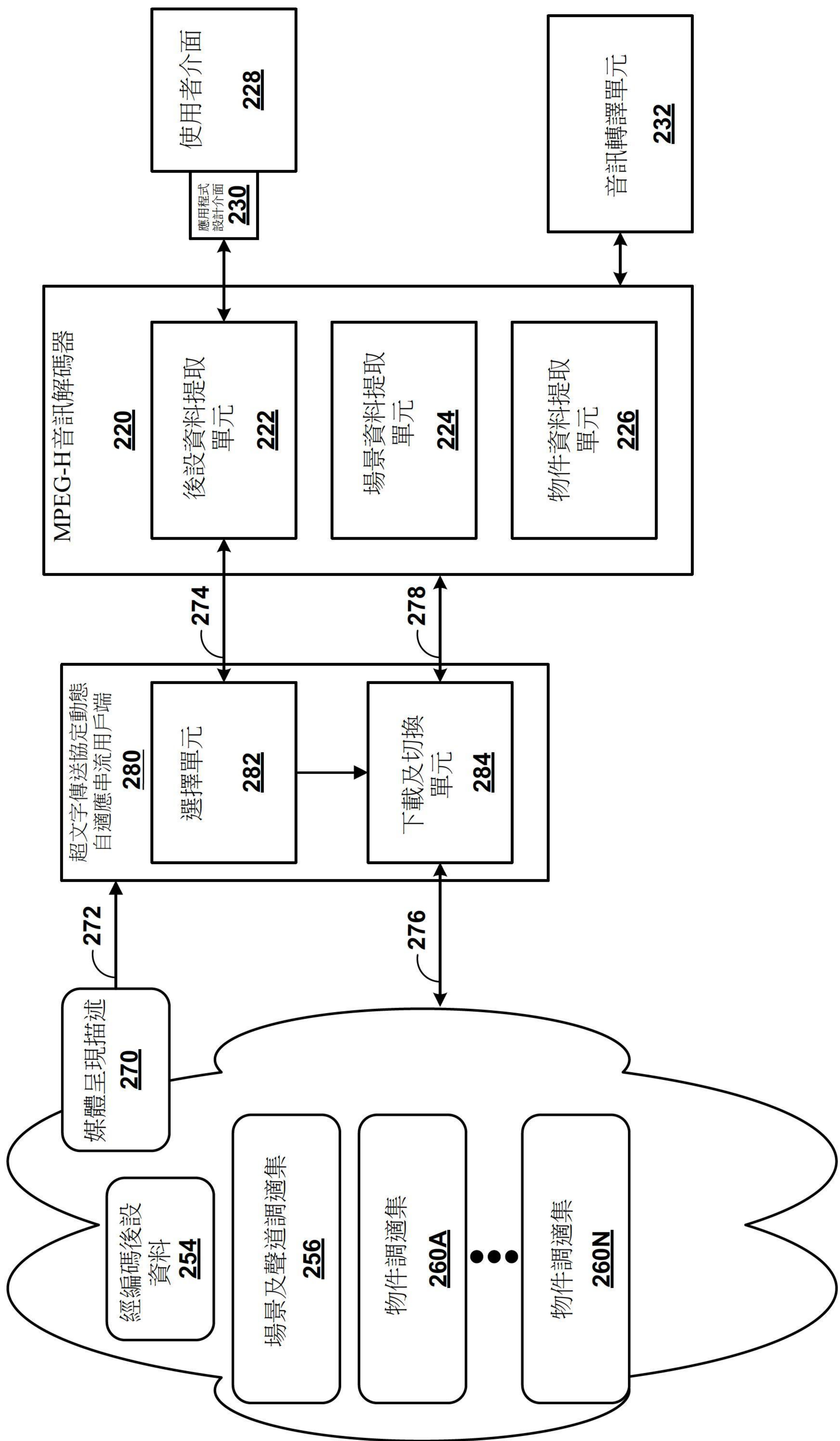


【圖7B】

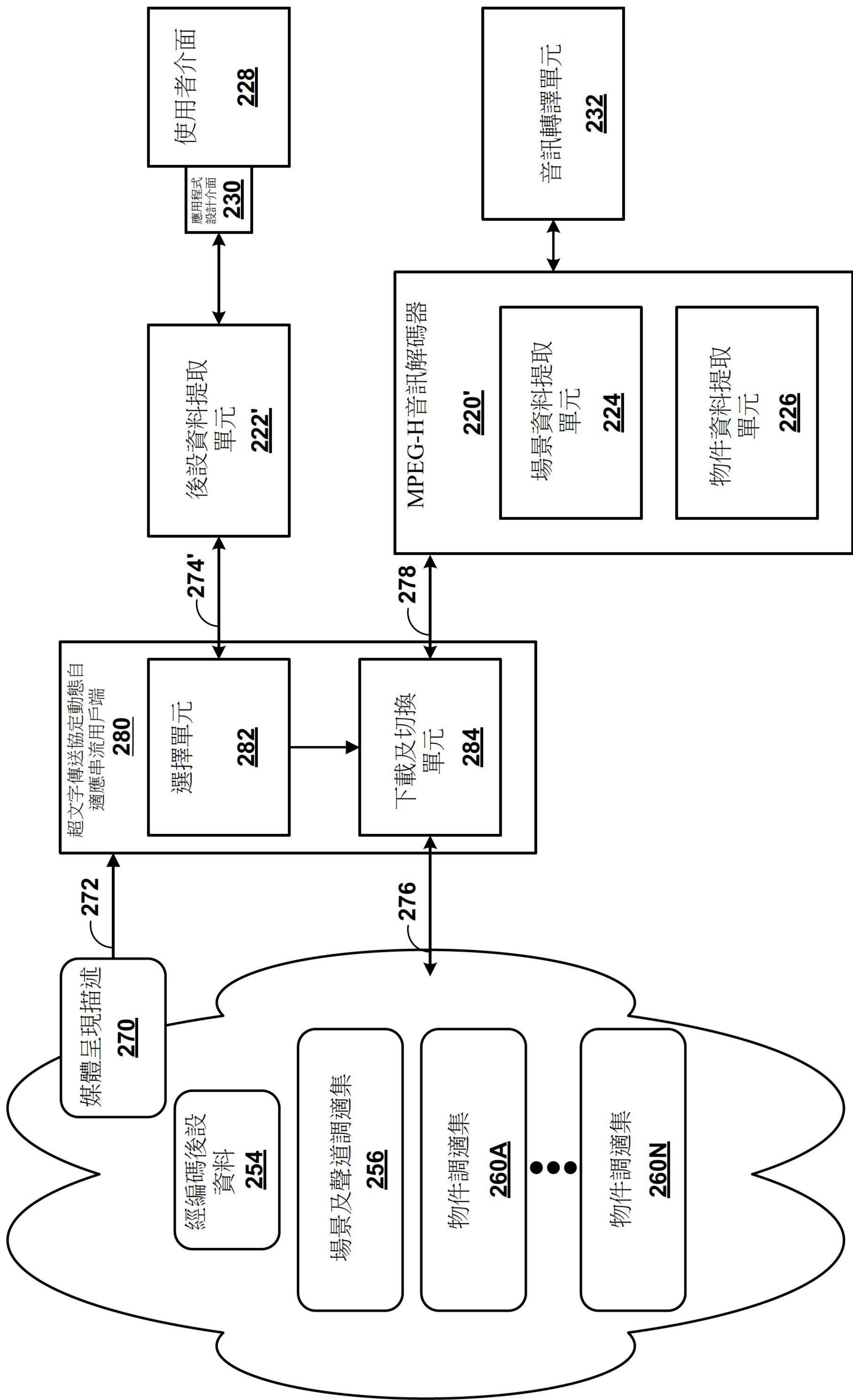
250 →



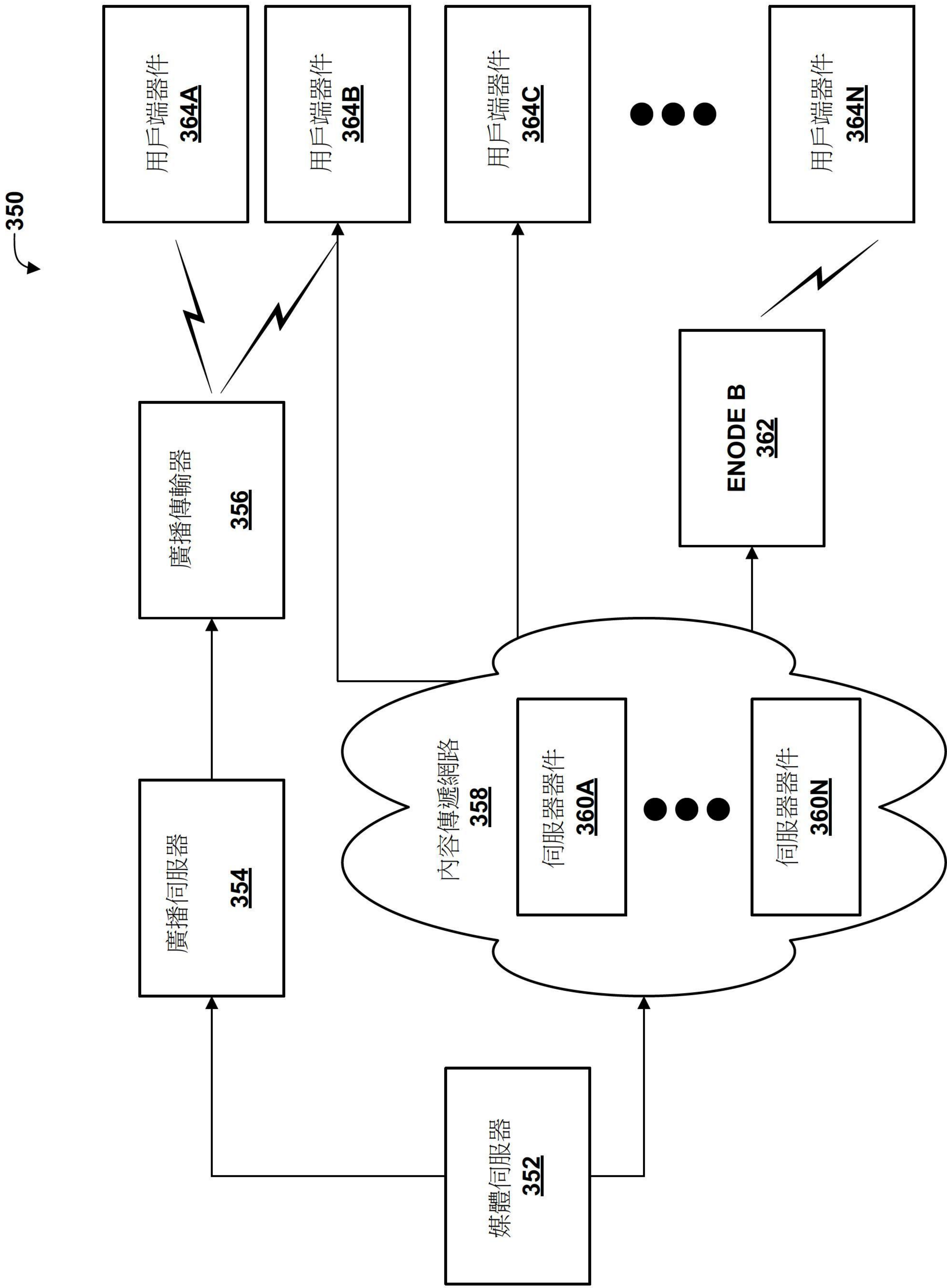
【圖7C】



【圖8】

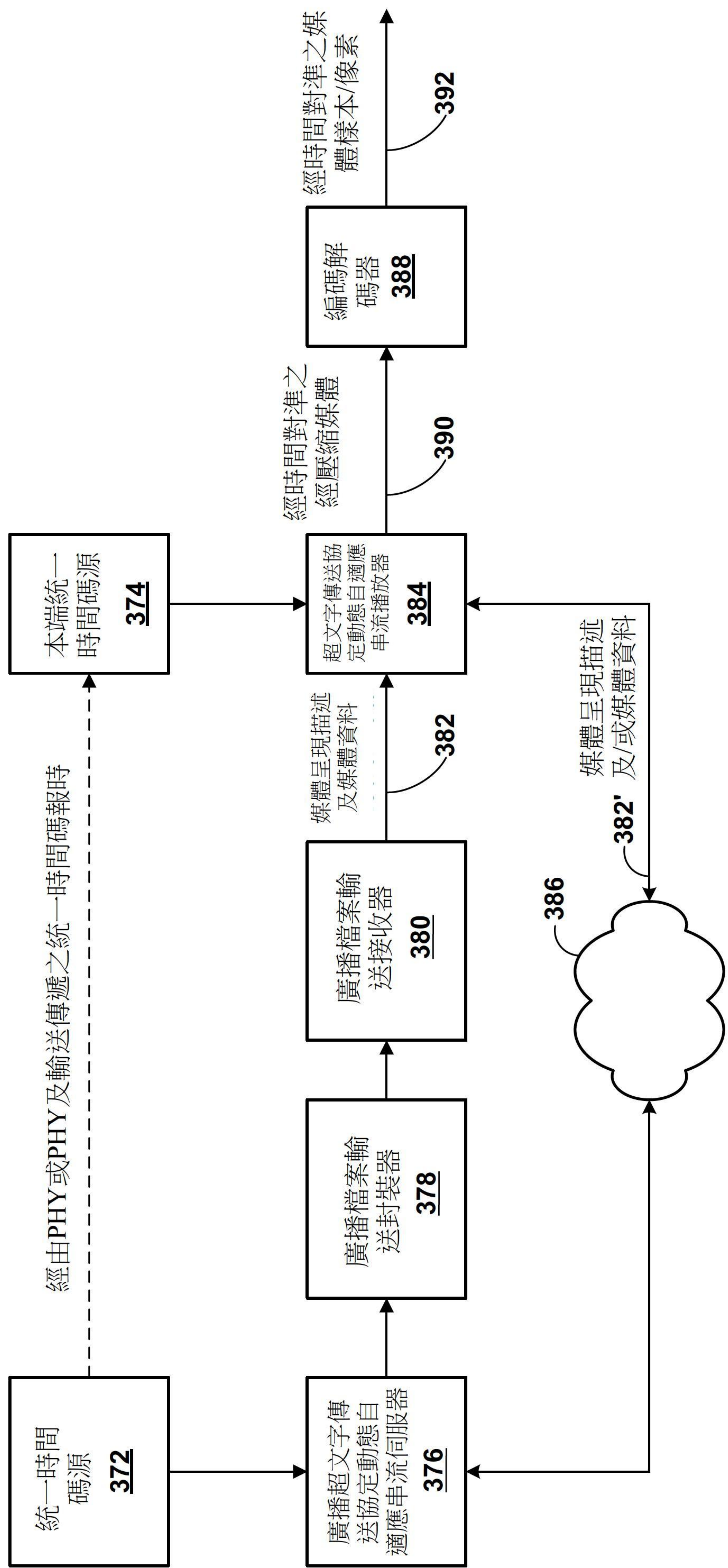


【圖9】



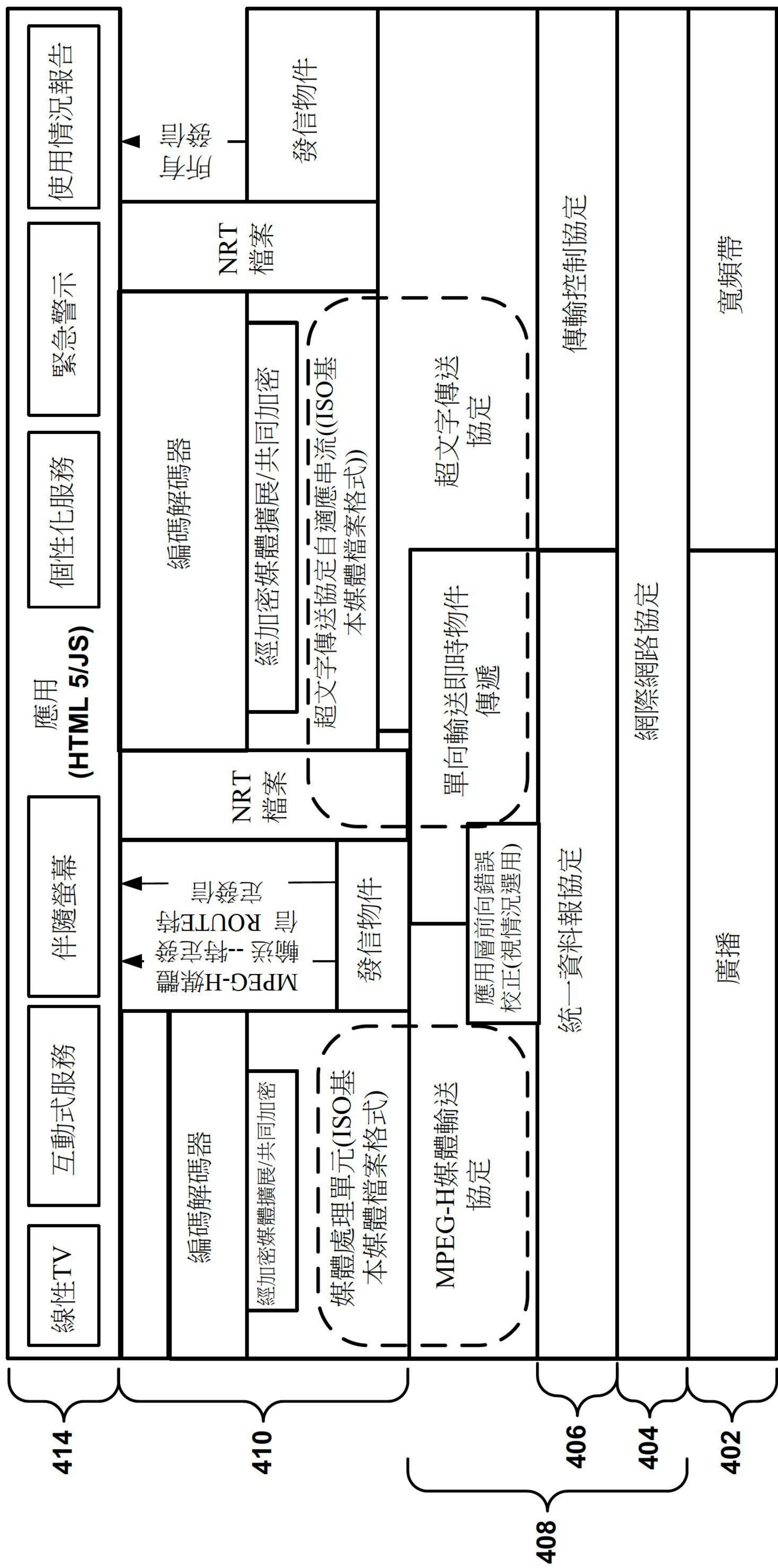
【圖10】

370

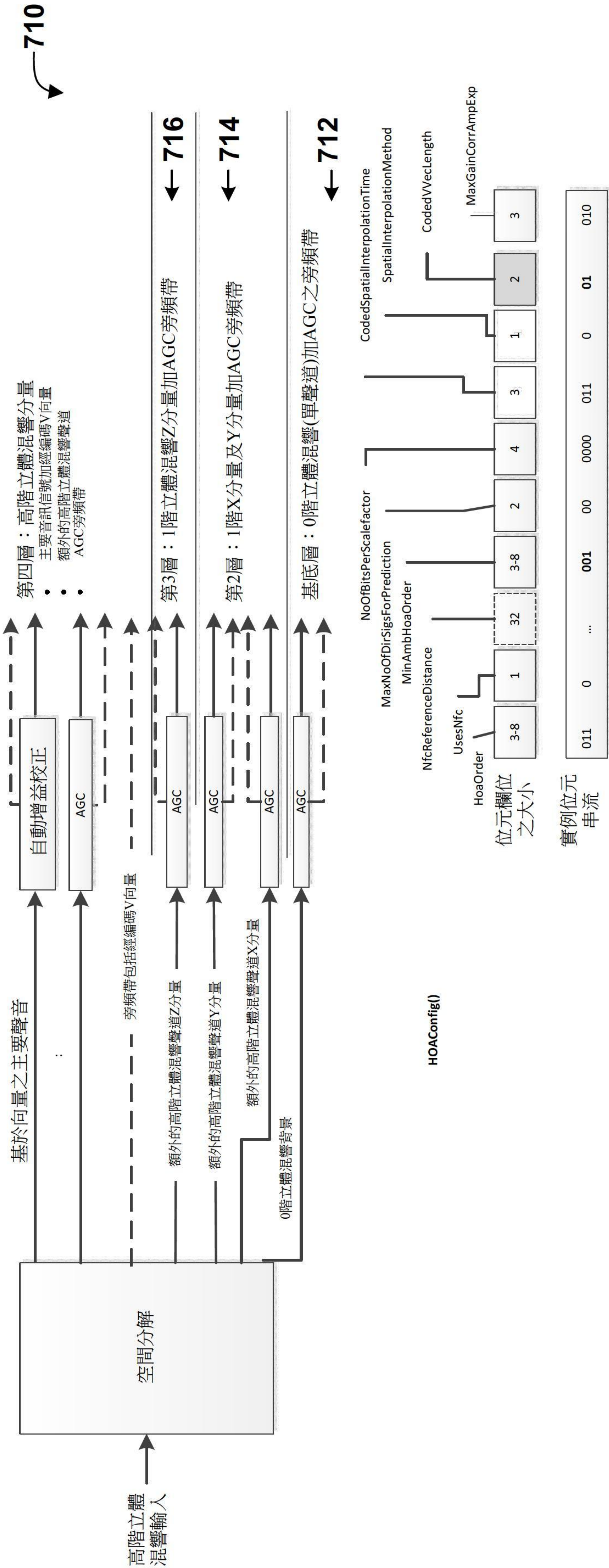


【圖11】

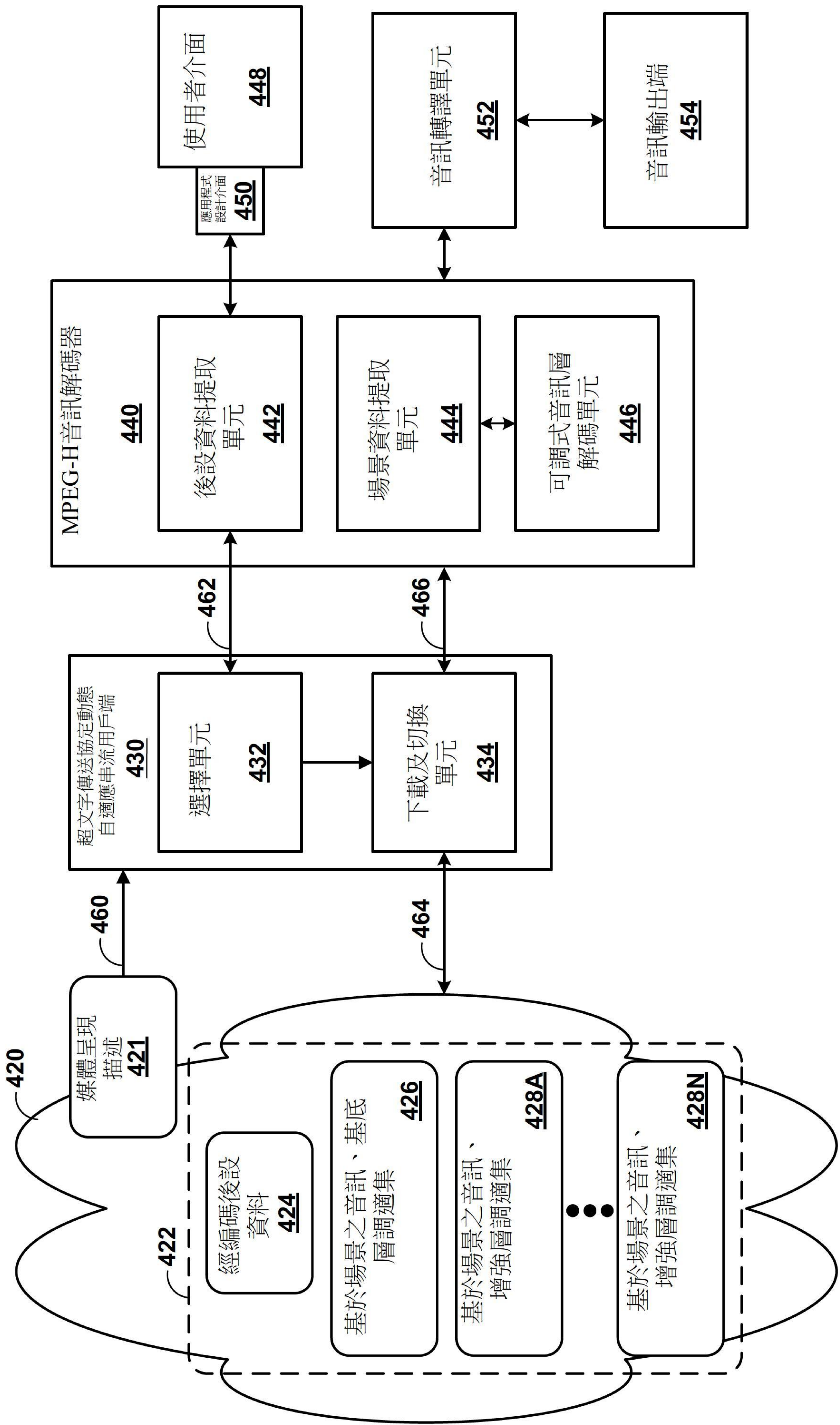
400



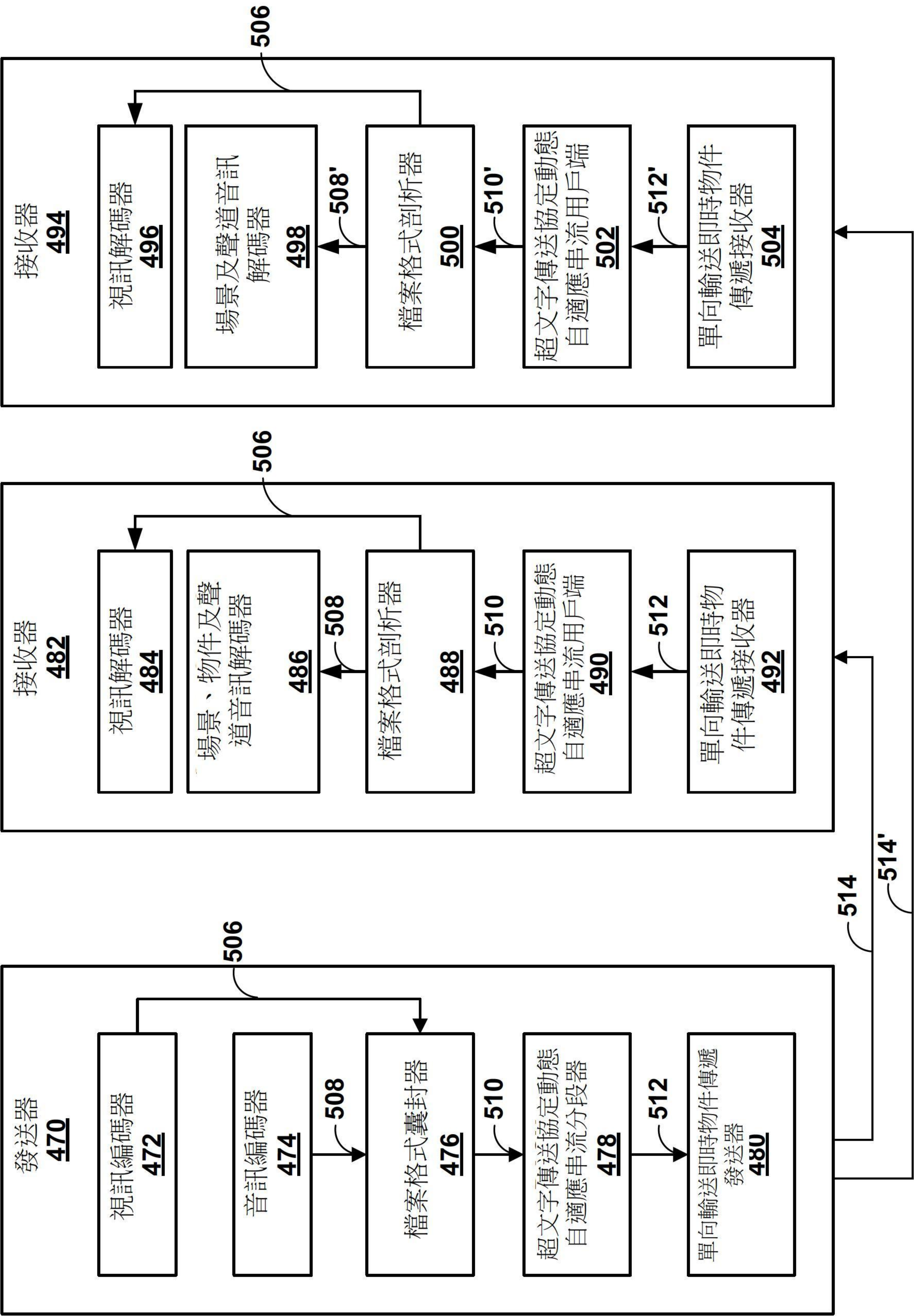
【圖12】



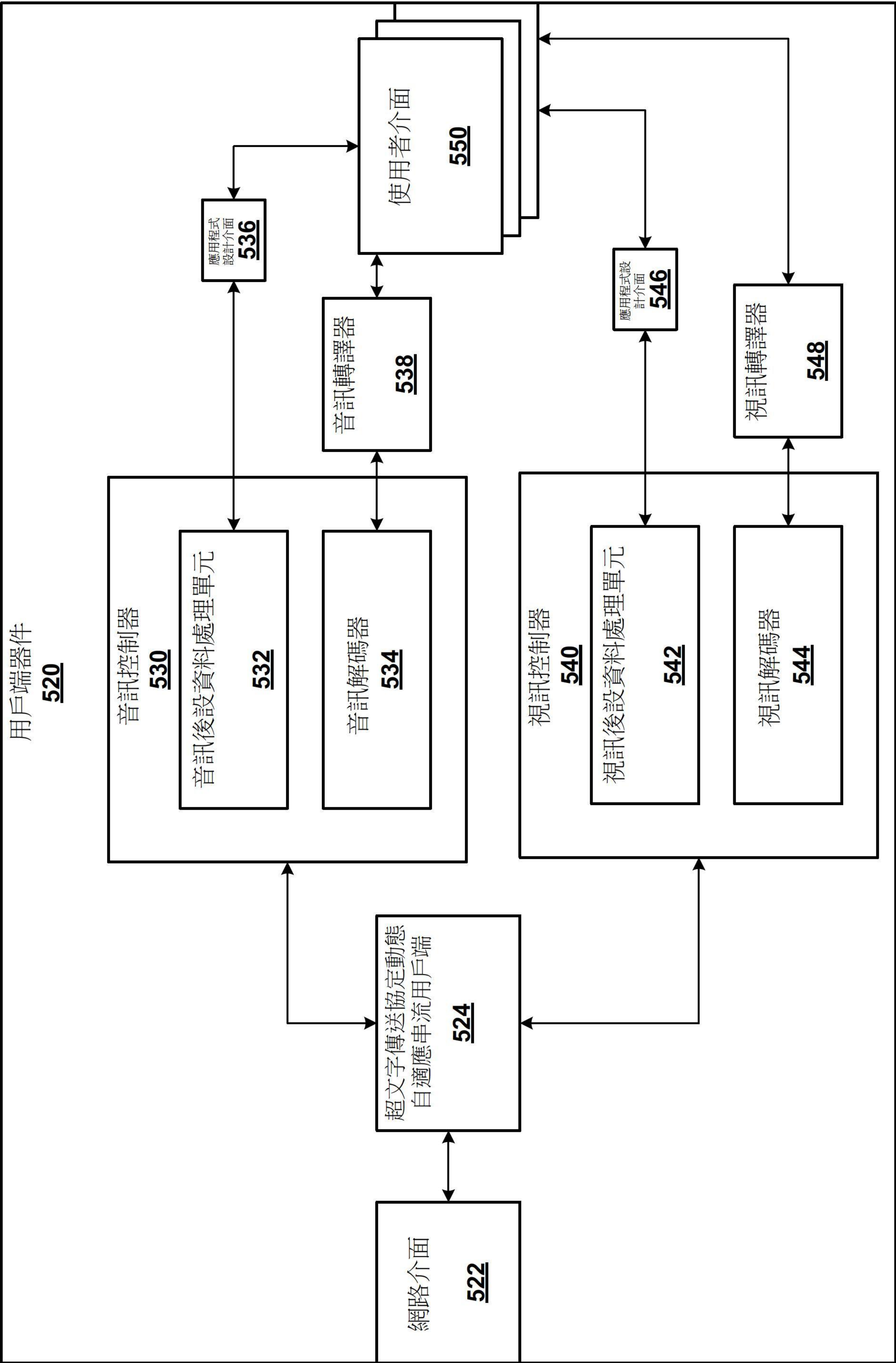
【圖14B】



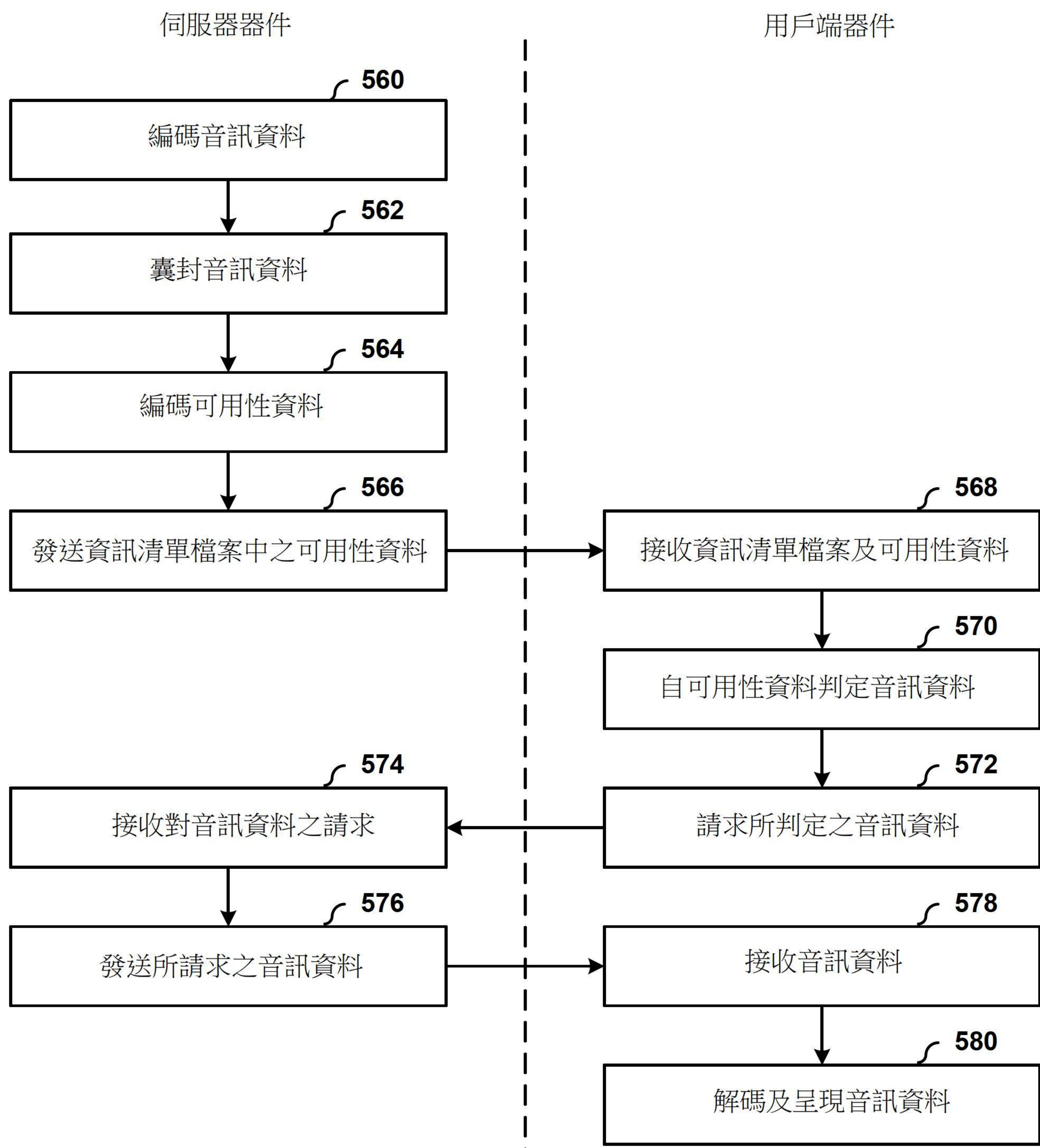
【圖15】



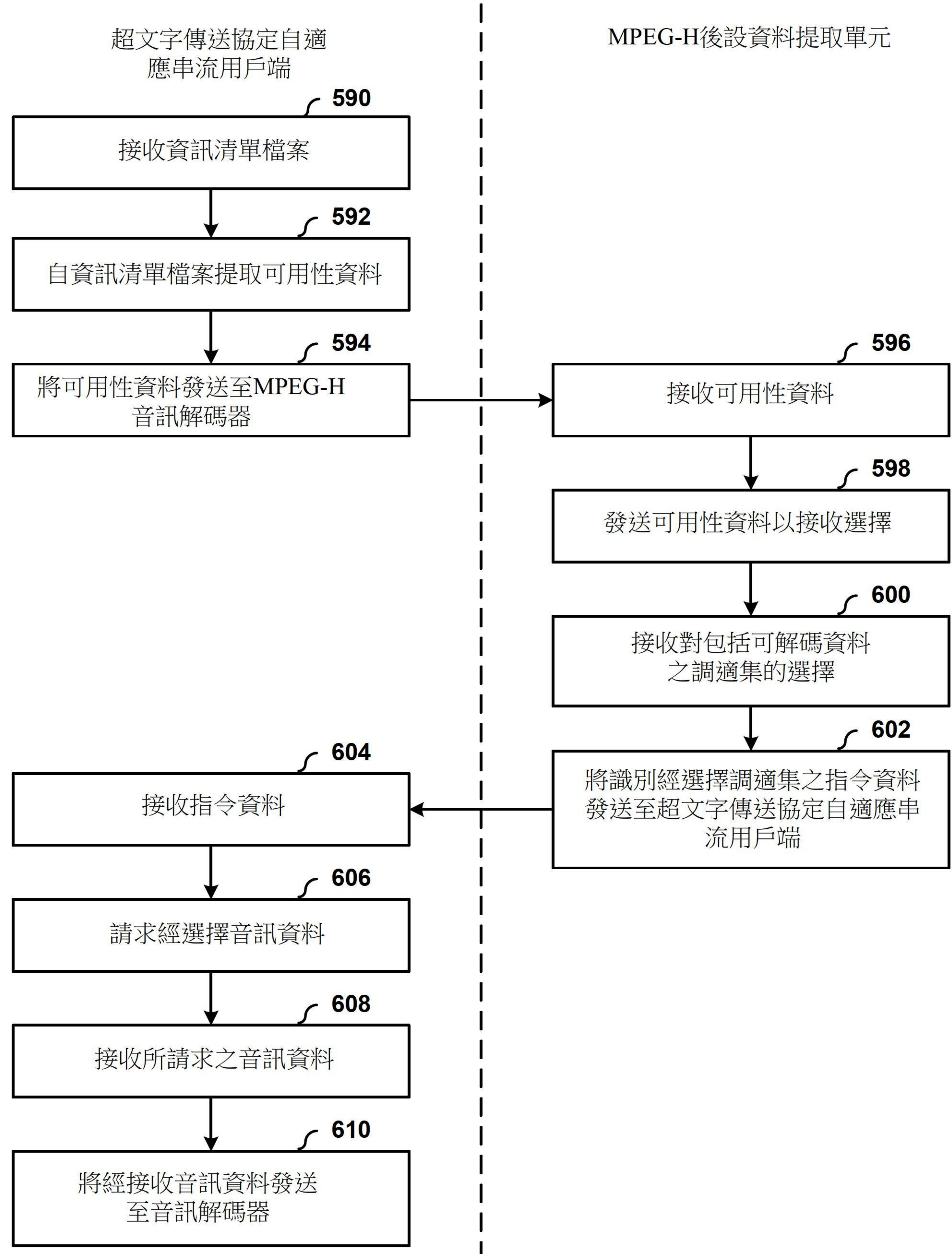
【圖16】



【圖17】



【圖18】



【圖19】