I441166

(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 I441166 B**

**(45)公告日： 中華民國 103 (2014) 年 06 月 11 日**

(21)申請案號：098121852 (22)申請日： 中華民國 98 (2009) 年 06 月 29 日

(51)Int. Cl.： ***G10L19/02 (2013.01)***

(30)優先權：2008/07/11 美國 61/079,875

2009/06/16 世界智慧財產權組織 PCT/EP2009/004339

(71)申請人：弗勞恩霍夫爾協會 (德國) FRAUNHOFER-GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG E. V. (DE)
德國

(72)發明人：福契斯 古拉米 FUCHS, GUILLAUME (FR)；拜爾 史蒂芬 BAYER, STEFAN (AT)；赫史菲爾德 珍斯 HIRSCHFELD, JENS (DE)；希瑞 喬根 HERRE, JUERGEN (DE)；列康提 傑瑞米 LECOMTE, JEREMIE (FR)；納吉爾 費德瑞克 NAGEL, FREDERIK (DE)；瑞德貝曲 尼可拉斯 RETTELBACH, NIKOLAUS (DE)；維比涅克 史蒂芬 WABNIK, STEFAN (DE)；橫谷義一 YOKOTANI, YOSHIKAZU (DE)

(74)代理人：惲軼群；陳文郎

(56)參考文獻：

| | | | |
|---|---|---|---|
| TW | 469422 | TW | I222623 |
| TW | I280560 | TW | 200739517A |
| US | 2003/0231775A1 | US | 2008/0162121A1 |

審查人員：李昭俊

申請專利範圍項數：17 項 圖式數：7 共 0 頁

(54)名稱

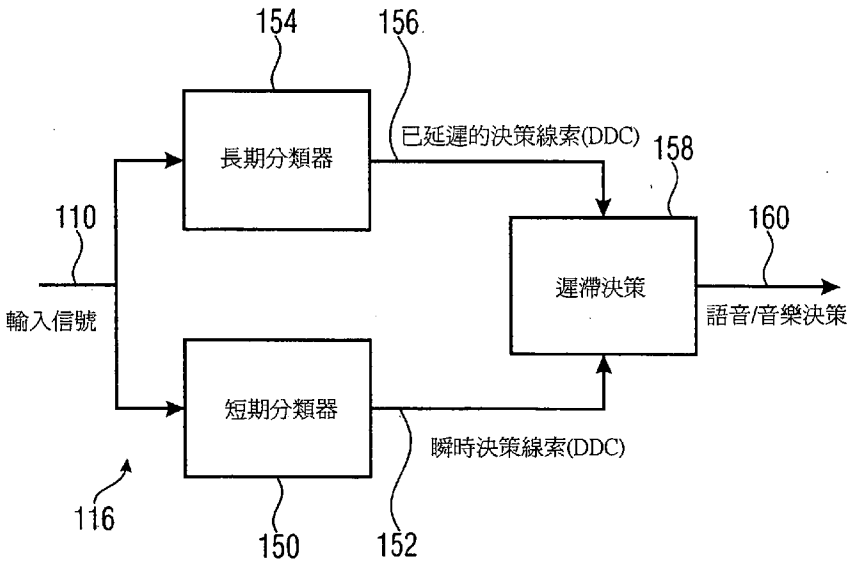
用以將信號之不同區段分類之方法與鑑別器

METHOD AND DISCRIMINATOR FOR CLASSIFYING DIFFERENT SEGMENTS OF A SIGNAL

(57)摘要

為了將包含至少一第一類型及一第二類型區段例如音訊區段及語音區段之一信號的不同區段分類，該信號係基於擷取自該信號之至少一個短期特徵做短期分類及遞送一短期分類結果。該信號也係基於擷取自該信號之至少一個短期特徵及至少一個長期特徵做長期分類及遞送一長期分類結果。該短期分類結果及長期分類結果組合而提供一輸出信號，指示該信號之一區段是否屬於第一類型或屬於第二類型。

For classifying different segments of a signal which comprises segments of at least a first type and second type, e.g. audio and speech segments, the signal is short-term classified (150) on the basis of the at least one short-term feature extracted from the signal and a short-term classification result (152) is delivered. The signal is also long-term classified (154) on the basis of the at least one short-term feature and at least one long-term feature extracted from the signal and a long-term classification result (156) is delivered. The short-term classification result (152) and the long-term classification result (156) are combined (158) to provide an output signal (160) indicating whether a segment of the signal is of the first type or of the second type.




第1圖

110・・・輸入線

116・・・語音/音樂鑑別器

150・・・短期分類器

152・・・輸出線

154・・・長期分類器

156・・・輸出線

158・・・遲滯決策電路

160・・・輸出信號、輸出線、語音/音樂決策信號

# 發明專利說明書

公告本

（本說明書格式、順序，請勿任意更動，※記號部分請勿填寫）

※申請案號：98121852

※申請日：98.6.29 ※IPC分類：G10L 19/02 ~~(2006.01)~~ (2013.01)

## 一、發明名稱：（中文/英文）

用以將信號之不同區段分類之方法與鑑別器

METHOD AND DISCRIMINATOR FOR CLASSIFYING DIFFERENT SEGMENTS OF A SIGNAL

## 二、中文發明摘要：

為了將包含至少一第一類型及一第二類型區段例如音訊區段及語音區段之一信號的不同區段分類，該信號係基於擷取自該信號之至少一個短期特徵做短期分類及遞送一短期分類結果。該信號也係基於擷取自該信號之至少一個短期特徵及至少一個長期特徵做長期分類及遞送一長期分類結果。該短期分類結果及長期分類結果組合而提供一輸出信號，指示該信號之一區段是否屬於第一類型或屬於第二類型。

## 三、英文發明摘要：

For classifying different segments of a signal which comprises segments of at least a first type and second type, e.g. audio and speech segments, the signal is short-term classified (150) on the basis of the at least one short-term feature extracted from the signal and a short-term classification result (152) is delivered. The signal is also long-term classified (154) on the basis of the at least one short-term feature and at least one long-term feature extracted from the signal and a long-term classification result (156) is delivered. The short-term classification result (152) and the long-term classification result (156) are combined (158) to provide an output signal (160) indicating whether a segment of the signal is of the first type or of the second type.

## 四、指定代表圖：

(一)本案指定代表圖為：第（ 1 ）圖。

(二)本代表圖之元件符號簡單說明：

110...輸入線
116...語音/音樂鑑別器
150...短期分類器
152...輸出線
154...長期分類器
156...輸出線
158...遲滯決策電路
160...輸出信號、輸出線、語音/音樂決策信號

## 五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

## 六、發明說明：

### 【發明所屬之技術領域】

本發明係關於一種用於將包含至少一第一類型及一第二類型之區段之一信號的不同區段分類之辨法。本發明之實施例係關於音訊編碼領域，特別係關於當編碼一音訊信號時語音/音樂之鑑別。

### 【先前技術】

發明背景

技藝界已知頻域編碼方案諸如MP3或AAC。此等頻域編碼器係基於時域/頻域變換；隨後之量化階段，其中該量化誤差係使用得自心理聲學模組之資訊控制；及一編碼階段，其中該等已量化之頻譜係數及相對應之旁資訊係使用碼表進行熵編碼。

另一方面，具有極為適合用於語音處理之編碼器諸如AMR-WB+，如述於3GPP TS 26.290。此種語音編碼方案進行時域信號之線性預測濾波。此種LP濾波係由所輸入的時域信號之線性預測分析導算出。然後所得LP濾波係數經編碼且傳輸作為旁資訊。該方法稱作為線性預測編碼(LPC)。於濾波器之輸出端，預測殘餘信號或預測誤差信號也稱作為激勵信號係使用ACELP編碼器之藉合成分析階段編碼；或另外，使用變換編碼器編碼，該變換編碼器係使用有重疊之傅立葉變換。ACELP與變換編碼激勵編碼(也稱作為TCX編碼)間之決策係使用閉環演繹法則或開環演繹法則判定。

頻域音訊編碼方案諸如高效率AAC編碼方案，組合AAC編碼方案及頻譜帶寬複製技術也可組合至於「MPEG環繞」一詞下方為已知之聯合立體聲或多頻道編碼工具。頻域編碼方案之優點在於其對音樂信號於低位元率顯示高品質。但問題在於語音信號於低位元率的品質。

另一方面，語音編碼器諸如AMR-WB+也具有高頻加強階段及立體聲功能。語音編碼方案即使於低位元率也對語音信號顯示高品質，但對於低位元率之音樂信號顯示品質不佳。

有鑑於前述可利用之編碼方案，其中若干較為適合編碼語音，而其它方案較為適合用於編碼音樂。域編碼之音訊信號的自動分段及分類為多種多媒體應用上的重要工具，可用來選擇出現於音訊信號之各不同類適當方法。應用之總體效能強力仰賴該音訊信號分類之可信度。確實，錯誤分類可能產生不適當的選擇以及隨後方法不適當的微調。

第6圖顯示取決於音訊信號之鑑別，用於適當編碼語音及音樂之習知編碼器設計。編碼器設計包含一語音編碼分支100包括一適當語音編碼器102，例如AMR-WB+語音編碼器，如「已擴展的自適應多速率-寬帶(AMR-WB+)編碼解碼器」，3GPP TS 26.290 V6.3.0，2005-06，技術規格所述。此外，編碼器設計包含一音樂編碼分支104其包含一音樂編碼器106例如AAC音樂編碼器，例如說明於動畫及關聯的音訊之大致編碼：進階音訊編碼。國際標準13818-7, ISO/IEC

JTC1/SC29/WG11動畫專家群1997年。

編碼器102及106之輸出端係連結至多工器108之輸入端。編碼器102及106之輸入端係可選擇性地連結至攜帶一輸入音訊信號之一輸入線110。輸入音訊信號係利用第6圖示意顯示且可藉開關控制器114控制之一開關112而選擇性施用至語音編碼器102或音樂編碼器106。此外，編碼器設計包含一語音/音樂鑑別器116，於其輸入端也接收輸入音訊信號，及輸出一控制信號至該開關控制器114。開關控制器114進一步於線118上輸出一模式指示器信號，該信號係輸入多工器108之第二輸入端，使得模式指示器信號可連同一已編碼信號發送。模式指示器信號可只有一個位元，指示與該模式指示器位元相關聯之一資料區塊為語音編碼或為音樂編碼，因此例如於解碼器無需再做鑑別。反而，基於連同已編碼資料遞送至該解碼器端之模式指示器位元，可產生一適當切換信號，基於該信號，該模式指示器可用於安排所接收的已編碼的資料路由至一適當語音解碼器或音樂解碼器。

第6圖為用來數位編碼施加至線110之語音信號及音樂信號之傳統編碼器設計。通常語音編碼器用於語音為較佳，而音訊編碼器用於音樂為較佳。經由使用多編碼器系統，其根據輸入信號的本質由一個編碼器切換至另一個編碼器，可設計出統一編碼方案。此處一項重要問題係設計出可驅動該切換元件之極為適當的輸入信號分類器。該分類器為第6圖所示語音/音樂鑑別器116。通常音訊信號之可

靠的分類導入高延遲；而另一方面，延遲為即時應用的一項重要因素。

通常期望藉語音/音樂鑑別器所導入的總演繹法則延遲夠低，因而允許將被切換的編碼器用於即時應用。

第7圖示例顯示如第6圖所示編碼器設計中所遭遇的延遲。假設施加於輸入線110之信號欲以16 kHz取樣率以1024樣本之訊框基準編碼，使得語音/音樂鑑別可遞送每個訊框一個決策，亦即每64毫秒一個決策。介於兩個編碼器間之變換例如係以WO 2008/071353 A2所述方式執行，語音/音樂鑑別器不會顯著增加被切換解碼器之演繹法則延遲，該解碼器共有1600個樣本，但未考慮語音/音樂鑑別器所需延遲。進一步，期望對同一個訊框提供語音/音樂決策，此處判定AAC區塊切換。該情況係顯示於第7圖，第7圖示例顯示具有2048個樣本長度之AAC長區塊120，亦即長區塊120包含兩個1024個樣本之訊框，一個1024個樣本訊框之AAC短區塊122，及一個1024個樣本訊框之AMR-WB+超訊框124。

於第7圖，分別基於涵蓋相等時間長度個別有1024個樣本之訊框126及128做AAC區塊切換決策及語音/音樂決策。兩項決策係於此特定位置進行來允許編碼可用於由一個模式適當變成另一個模式之一時間變遷窗使用。結果，藉兩項決策導入至少512+64樣本延遲。本延遲必需加至藉50%重疊所產生的1024個樣本之延遲，形成AAC MDCT，獲得至少1600個樣本之延遲。於習知AAC，只存在有區塊切換，延遲恰為1600個樣本。當於訊框126檢測得變遷時，

本延遲為由長區塊切換至短區塊時所需。此種變換長度的切換為避免前回波假信號所需。第7圖已解碼之訊框130表示第一個完整訊框，總而言之(長區塊或短區塊)其可於解碼器端重建。

於使用AAC作為音樂編碼器之已切換的編碼器中，來自於決策階段之切換決策須避免添加過多額外的延遲至原先的AAC延遲。額外延遲係來自於前瞻訊框132，該訊框為決策階段之信號分析所需。於例如16 kHz之取樣率，AAC延遲為100毫秒；習知語音/音樂鑑別器使用約500毫秒之前瞻，將導致具有600毫秒延遲之已切換的編碼結構。則總延遲將變成原先AAC延遲的6倍。

如前文說明之習知辨法之缺點為用於音訊信號可靠的分類，導入高度非期望的延遲，因而需要有新穎辦法可用於鑑別包括不同類型區段之一信號，其中藉該鑑別器所導入之額外演繹法則延遲夠低，因而已切換的編碼器也可用於即時應用。

J. Wang等人「具有階層式傾斜決策樹之即時語音/音樂分類」，ICASSP 2008，IEEE聲學國際會議，語音及信號處理2008年，2008年3月31日至2008年4月4日敘述一種使用由等數訊框所導算出之短期特徵及長期特徵進行語音/音樂分類之辦法。此等短期特徵及長期特徵係用來分類信號，但只探勘短期特徵有限的性質，例如並未探勘分類反應性，但該反應性對大部分音訊編碼應用上扮演重要角色。

## 【發明內容】

發明概要

本發明之一目的係提供一種於不同類型之信號區段做鑑別同時維持藉該鑑別所導入之任何延遲極低之改良式辦法。

本目的可藉如申請專利範圍第1項之方法及如申請專利範圍第14項之鑑別器達成。

本發明之一個實施例提供一種用於分類一信號之不同區段之方法，該信號包含至少一第一類型及一第二類型之區段，該方法包含：

基於擷取自該信號之至少一個短期特徵，短期分類該信號及遞送短期分類結果；

基於擷取自該信號之至少一個短期特徵及至少一個長期特徵對該信號做長期分類及遞送一長期分類結果；及

組合該短期分類結果與該長期分類結果而提供一輸出信號指示該信號之一區段是否為第一類型或為第二類型。

本發明之另一個實施例提供一種鑑別器，包含：

一短期分類器，配置來接收一信號且基於擷取自該信號之至少一個短期特徵，提供該信號之一短期分類結果，該信號包含至少一第一類型及一第二類型之區段；

一長期分類器，配置來接收該信號且基於擷取自該信號之至少一個短期特徵及至少一個長期特徵而提供該信號之一長期分類結果；

一決策電路，配置來將該短期分類結果與該長期分類結果組合而提供一輸出信號，該輸出信號指示該信號之一

區段是否為第一類型或為第二類型。

本發明之實施例提供基於該短期分析結果與該長期分析結果比較之該輸出信號。

本發明之實施例係有關一種辦法將一音訊信號之不同的非重疊的短時間區段分類為語音或非語音或其它類別。該辦法係基於特徵之擷取及對兩個不同分析視窗長度之統計學分析。第一個視窗為長視窗，主要係看過去。第一個視窗用來對該信號之分類獲得可靠但延遲的決策線索。第二視窗的長度短，主要係考慮目前時間所處理的區段或稱作為目前區段。第二視窗係用來獲得瞬時決策線索。兩個決策線索經最佳組合，較佳係使用遲滯決策獲得得自該延遲線索之記憶資訊及得自該瞬時線索之瞬時資訊。

本發明之實施例使用短期特徵用於短期分類器及長期分類器二者，讓兩種分類器探勘同一種特徵之不同統計學。短期分類器將只擷取瞬時資訊，原因在於其只能存取一個特徵集合。例如可探勘該等特徵的平均值。另一方面，長期分類器由於考慮數個訊框故可存取若干特徵集合。結果長期分類器經由探勘比短期分類器更多個訊框之統計學，因而可探勘該信號之更多特性。舉例言之，長期分類器可探勘隨著時間之經過該等特徵之變因或該等特徵之演化。如此，長期分類器比短期分類器可探勘更多個特徵，但導入延遲或拖延。但長期特徵儘管導入延遲或拖延，將造成長期分類更強勁更可靠。於若干實施例中，短期分類器及長期分類器考慮相同短期特徵，可一次運算且由兩個

分類器使用。如此，於此種實施例中，長期分類器可直接由短期分類器接收短期特徵。

藉此該新穎辦法允許獲得強勁但導入低度延遲之一種分類。除了習知辦法外，本發明之實施例限制由語音/音樂決策所導入的延遲，同時保有可靠的決策。於本發明之一個實施例中，前瞻係限於128個樣本，結果導致總延遲只有108毫秒。

圖式簡單說明

後文將參考附圖說明本發明之實施例，附圖中：

第1圖為根據本發明之實施例，一種語音/音樂鑑別器之方塊圖；

第2圖示例顯示由第1圖之鑑別器之長期分類器及短期分類器所使用之分析視窗；

第3圖示例顯示用於第1圖之鑑別器之遲滯決策；

第4圖為包含根據本發明之實施例之一鑑別器之一編碼方案實例之方塊圖；

第5圖為與該第4圖之編碼方案相對應之解碼方案之方塊圖；

第6圖顯示用於依據一音訊信號之鑑別而分開編碼語音及音樂之一種習知編碼器設計；及

第7圖示例顯示於第6圖所示編碼器設計中遭遇之延遲。

## 【實施方式】

較佳實施例之詳細說明

第1圖為根據本發明之實施例之一種語音/音樂鑑別器116之方塊圖。語音/音樂鑑別器116包含一短期分類器150，於其輸入端接收一輸入信號，例如包含語音區段及音樂區段之一音訊信號。短期分類器150於一輸出線152上輸出一短期分類結果亦即瞬時決策線索。鑑別器116進一步包含一長期分類器154，其接收該輸入信號且於輸出線156上輸出長期分類結果，亦即已延遲的決策線索。進一步，設置一遲滯決策電路158其係以容後詳述之方式組合得自該短期分類器150及長期分類器154之輸出信號，來產生一語音/音樂決策信號，該信號係於線160輸出，且可用來以前文參考第6圖所述方式控制一輸入信號之一區段的進一步處理，以及語音/音樂決策信號160可用來安排已經被分類之該輸入信號區段路由至一語音編碼器或音訊編碼器。

如此，根據本發明之實施例，兩個不同分類器150及154並列用於透過輸入線110而施加至個別分類器之該輸入信號。二分類器稱作為分類器154及短期分類器150，其中藉分析於分析視窗上運算之各項特徵的統計學而可區別兩個分類器。二分類器遞送輸出信號152及156，亦即瞬時決策線索(IDC)及已延遲的決策線索(DDC)。短期分類器150係基於針對捕捉有關該輸入信號本質之瞬時資訊之短期特徵產生IDC。IDC係有關可快速且隨時改變之該信號之短期屬性。結果，短期特徵預期為反應性而不會將長期延遲導入整個鑑別過程。例如由於語音被視為5毫秒至20毫秒時間之準穩態，於以16 kHz取樣之信號，可每16毫秒一個訊框運

算短期特徵。長期分類器154係基於由於長期觀察信號之特徵(長期特徵)產生DDC，因而允許達成更可靠的分類。

第2圖示例顯示由第1圖所示長期分類器154及短期分類器150所使用之分析視窗。假設於取樣率16 kHz，一個訊框1024個樣本，長期分類器視窗162之長度為4*1024+128樣本，亦即長期分類器視窗162跨四個音訊信號訊框，長期分類器154進行分析時需要額外128個樣本。此種額外延遲也稱作為「前瞻」係於第2圖中指示於元件符號164。第2圖也顯示短期分類器視窗166，共有1024+128樣本，亦即跨據一個音訊信號訊框且需額外延遲來分析一目前區段。該目前區段指示於128作為必需進行語音/音樂決策之區段。

第2圖指示之長期分類器視窗夠長而可獲得語音之4-Hz能量調變特性。4-Hz能量調變為傳統上例如由Scheirer E.及Slaney M.，「強勁多重特徵語音/音樂鑑別器之組成及評估」，ICASSP' 97，慕尼黑1997年所使用之強勁語音/音樂鑑別器傳統上所探勘之語音之相關的且具有鑑別能力的特性。4-Hz能量調變是一種經由觀察長時間區段之信號而可擷取之特徵。藉語音/音樂鑑別器所導入之額外延遲係等於128個樣本之前瞻164，該前瞻為分類器150及154各自所需來作個別分析，例如感官式線性預測分析，由H. Hermansky說明於「語音之感官式線性預測(plp)分析」，美國聲波學會期刊第87卷第4期1738-1752頁，1990年及H. Hermansky等人「語音之基於感官之線性預測分析」，ICASSP 5.509-512，1985。如此，當使用前述實施例之鑑別器於如第6圖所示之

編碼器設計，切換之編碼器102及106之總延遲將為1600+128樣本，等於108毫秒，該延遲夠低而可供即時應用。

現在參考第3圖，說明用於獲得一語音/音樂決策信號160之鑑別器116之分類器150及154之輸出信號152與156之組合。根據本發明之實施例，已延遲之決策線索DDC及瞬時決策線索IDC係使用一遲滯決策組合。遲滯決策廣用於後處理決策俾便穩定決策。第3圖示例顯示二態遲滯決策呈DDC及IDC之函數俾測定語音/音樂決策信號是否須指示該輸入信號之目前所處理的區段為一語音區段或一音樂區段。特性遲滯週期顯示於第3圖，IDC及DDC藉分類器150及154規度化因而其數值為-1至1，其中-1表示全然為音樂狀之似然度，及1表示全然為語音狀之似然度。

決策係基於函數F (IDC, DDC)之數值，其實例說明如下。第3圖中，F1 (DDC, IDC)指示F (IDC, DDC)由音樂態至語音態須交叉的一臨界值。F2 (DDC, IDC)指示F (IDC, DDC)由語音態至音樂態須交叉的一臨界值。然後基於如下虛擬碼計算具有指數n之一目前區段或目前訊框之最終決策D(n)：

```
%Hysteresis Decision Pseudo Code
If(D(n-1)==music)
     If(F(IDC,DDC)<F1(DDC,IDC))
          D(n)==music
     Else
          D(n)==speech
Else
     If(F(IDC,DDC)>F2(DDC,IDC))
          D(n)==speech
     Else
          D(n)==music
%End Hysteresis Decision Pseudo Code
```

根據本發明之實施例，函數F (IDC, DDC)及前述臨界值列舉如下：

```
F(IDC,DDC)=IDC
F1(IDC,DDC)=0.4-0.4*DDC
F2(IDC,DDC)=-0.4-0.4*DDC
```

另外，可做下列定義：

```
F(IDC,DDC)=(2*IDC+DDC)/3
F1(IDC,DDC)=-0.75*DDC
F2(IDC,DDC)=-0.75*DDC
```

當使用最末定義時，遲滯週期消失，只基於獨特的自適應臨界值做決策。

本發明並非限於前文說明之遲滯決策。後文將說明組合分析結果用來獲得該輸出信號之額外實施例。

經由探勘DDC及IDC二者特性製作臨界值，可使用單純的臨界值決策來替代遲滯決策。由於DDC係來自於該信號之長期觀察，故DDC被視為較為可靠的鑑別線索。但DDC部分係基於該信號之過去觀察運算。習知分類器只將DDC值與臨界值0比較，當DDC大於0時將一區段分類為語音狀，否則即分類為音樂狀，將有延遲決策。於本發明之一個實施例中，發明人經由探勘IDC來自適應臨界值決策，使得該決策更具有反應性。用於此項目的，臨界值可基於下述虛擬碼自適應：

```
% Pseudo code of adaptive thresholding
If(DDC>-0.5*IDC)
     D(n)==speech
Else
     D(n)==music
%End of adaptive thresholding
```

於另一個實施例中，DDC可比IDC更可靠地使用。IDC已知具有反應性，但不如DDC可靠。此外，觀察過去區段與目前區段間之DDC的演化，獲得另一項指示，指示第2圖之訊框166如何影響於區段162上計算的DDC。DDC(n)標示法用於DDC之目前值，而DDC(n-1)用於過去值。使用二數值DDC(n)及DDC(n-1)，經由使用決策樹可更可靠地做出IDC，說明如下：

```
% Pseudo code of decision tree
If(IDC>0 && DDC(n)>0)
     D(n)=speech
Else if (IDC<0 && DDC(n)<0)
     D(n)=music
Else if (IDC>0 && DDC(n)-DDC(n-1)>0)
     D(n)=speech
Else if (IDC<0 && DDC(n)-DDC(n-1)<0)
     D(n)=music
Else if (DDC>0)
     D(n)=speech
Else
     D(n)=music
%End of decision tree
```

於前述決策樹中，若二線索顯示相同的似然度，則直接做決策。若二線索獲得相矛盾的指示，則注意DDC的演化。若差值DDC(n)-DDC(n-1)為正，則假設目前區段為語音狀。否則可假設目前區段為音樂狀。若此新指示之方向與IDC相同，則取最終決策。若兩項嘗試皆未能獲得明白決策，則經由只考慮已延遲的線索DDC做決策，原因在於IDC之可靠度無法證實。

後文將說明根據本發明之實施例之個別分類器150及

154之進一步細節。

首先參考長期分類器154，發現用於由256樣本之每個子訊框擷取一特徵集合亦同。該第一特徵為感官式線性預測聲譜係數(PLPCC)，由H. Hermansky說明於「語音之感官式線性預測(plp)分析」，美國聲波學會期刊第87卷第4期1738-1752頁，1990年及H. Hermansky等人「語音之基於感官之線性預測分析」，ICASSP 5.509-512，1985。經由使用人類聽覺感官估算，PLPCC可有效用於揚聲器分類。此項特徵可用於鑑別語音及音樂，確實允許藉觀察隨著時間之經過的特徵變化來區別語音之特性共振峰及語音之音節4-Hz調變。

但為了獲得更為強勁，PLPCC組合另一項特徵，其可捕捉音高資訊，音高資訊乃語音的另一項重要特性，於編碼上具有關鍵重要性。確實，語音編碼仰賴假設輸入信號為一虛擬單一週期信號。語音編碼方案可有效用於此種信號。另一方面，語音之音高特性對音樂編碼器之多種編碼效率有害。語音的天然顫音產生平順音高延遲起伏波動，造成音樂編碼器的頻率表示型態無法大為緊壓能量，而該大為緊壓為獲得高度編碼效率所需。

可測定下列音高特徵：

聲門脈衝能比：

本特徵係運算聲門脈衝與LPC殘餘信號間之能量比。聲門脈衝係藉使用撿選波峰演繹法則而由LPC殘餘信號中擷取。通常有聲區段之LPC殘餘獲得來自於聲門振動的大

脈衝狀結構。於有聲區段期間該項特徵為高。

長期增益預測：

於長期預測期間通常於語音編碼器運算增益(例如「已擴展的自適應多速率-寬帶(AMR-WB+)編碼解碼器」，3GPP TS 26.290 V6.3.0，2005-06，技術規格)。本特徵測量該信號之週期性且係基於音高延遲估算。

音高延遲起伏波動：

本特徵判定比較最末子訊框時，本音高延遲估算之差。對有聲語音，本特徵須為低但不可為零且係順利演化。

一旦長期分類器已經擷取要求的特徵集合，對此等所擷取的特徵使用統計分類器。首先係於語音訓練集合及音樂訓練集合擷取各項特徵來訓練該分類器。所擷取的特徵對二訓練集合規度化至平均值為0及變因為1。對各個訓練集合，已擷取的且已規度化特徵被聚集於一長期分類器視窗，使用5高斯藉高斯混合模型(GMM)來模型化。於訓練序列結束時，獲得及儲存一規度化參數集合及二GMM參數集合。

對各個欲分類之訊框，首先擷取各項特徵及以規度化參數來規度化。使用語音類別之GMM及音樂類別之GMM，對所擷取的且已規度化之特徵運算語音之最大似然度(lld_speech)及音樂之最大似然度(lld_music)。已延遲決策線索DDC隨後計算如下：

DDC=(lld_speech–lld_music)/(abs(lld_music)+abs(lld_speech))

DDC係以-1至1為邊界，當語音之最大似然度高於音樂之最大似然度lld_speech>lld_music時，DDC為正。

短期分類器係用作為短期特徵PLPCC。除了於長期分類器之外，本特徵只於視窗128分析。本特徵之統計學係使用5高斯藉高斯混合模型(GMM)於此短時間探勘。訓練兩個模型，一個用於音樂，另一個用於語音。值得注意者為兩個模型係與對長期分類器所得之模型不同。對各個欲分類之訊框，首先擷取PLPCC，使用語音類別之GMM及音樂類別之GMM分別運算語音之最大似然度(lld_speech)及音樂之最大似然度(lld_music)。然後求出瞬時決策線索IDC如下：

IDC＝(lld_speech-lld_music)/(abs(lld_music)+abs(lld_speech))

IDC係以-1至1為界。

如此，基於特徵「感官式線性預測聲譜係數(PLPCC)」，短期分類器150產生該信號之短期分類結果；及基於相同特徵「感官式線性預測聲譜係數(PLPCC)」，及前述額外特徵例如音高特徵，長期分類器154產生該信號之長期分類結果。此外，長期分類器由於接取較長的觀察視窗，故可探勘共享特徵亦即PLPCC之不同特性。如此當組合短期結果與長期結果時，短期特徵被充分考慮用於分類，亦即短期特徵之性質經充分探勘。

以下將說明個別分類器150及154之又一個實例之進一步細節。

根據本實例藉短期分類器分析之短期特徵主要係與前述感官式線性預測聲譜係數(PLPCC)相對應。PLPCC廣用於語音及揚聲器辨識及MFCC(參見上文)。保有PLPCC，原因在於PLPCC與用於大部分近代語音編碼器且已經於已切換的音訊編碼器中實施之線性預測(LP)共享大部分功能。如同LP，PLPCC可擷取語音的共振峰結構，但經由做感官考量，PLPCC與揚聲器更為獨立無關，如此與語言資訊更具有相關。對16 kHz取樣的輸入信號使用16階級。

除了PLPCC外，調聲強度作為短期特徵。調聲強度本身並未考慮作為實際上用於鑑別，反而係與特徵維度方面之PLPCC相關有利。調聲強度允許於特徵維度中至少帶入兩個簇集分別係與語音的有聲發音及無聲發音相對應。係基於使用不同參數做考績計算，該等參數包含零交叉計數器(zc)、頻譜傾斜(tilt)、音高穩定性(ps)、及音高之規度化相關性(nc)。四項參數皆被規度化為0至1，0係與典型無聲信號相對應，1係與典型有聲信號相對應。於本實施例中，由VMR-WB語音編碼器所使用的語音分類標準激發調聲強度，該標準係述於Milan Jelinek及Redwan Salami，「於vmr-wb標準之寬帶語音編碼進階」，音訊、語音及語言處理之IEEE議事錄第15卷第4期1167-1179頁，2007年5月。係基於以自我相關性為基礎所演化的音高追蹤器。對訊框指數k，調聲強度u(k)具有下述形式：

$$v(k)=\frac{1}{5}(2*nc(k)+2*ps(k)+tilt(k)+zc(k))$$

短期特徵之鑑別能力係藉高斯混合模型(GMMS)作為分類器評估。應用兩個GMM，一個GMM用於語音類別，另一個GMM用於音樂類別。改變混合物數目來評估對表現效能的影響。表1顯示不同數目混合物之準確率。對四個連續訊框的每個區段運算判定。總延遲係等於64毫秒，其對已切換之音訊編碼為適合。可觀察到效能隨著混合物數目而增加。1-GMMs與5-GMMs間之間隙特別重要，可藉下述事實說明，語音之共振峰表現太複雜而無法只藉一個高斯來定義。

表1：短期特徵分類準確度，以%表示

| | 1-GMMs | 5-GMMs | 10-GMMs | 20-GMMs |
|---|---|---|---|---|
| 語音 | 95.33 | 96.52 | 97.02 | 97.60 |
| 音樂 | 92.17 | 91.97 | 91.61 | 91.77 |
| 平均 | 93.75 | 94.25 | 94.31 | 94.68 |

現在說明長期分類器154，發現許多研究工作者例如M.J. Carey等人，「語音及音樂鑑別之特徵比較」，聲學、語音及信號處理，第12期149至152頁，1999年三月，考慮統計特徵之變因比特徵本身更具有鑑別力。至於粗略之一般規則，考慮音樂更穩定具有更低變因。相反地，語音更容易藉其顯著4-Hz能量調變來區別，原因在於語音信號係週期性介於有聲區段與無聲區段間改變。此外連續不同的音素讓語音特徵較不恆定。於本實施例中，考慮兩個長期特徵，一個特徵係基於變因運算，而另一個特徵係基於語音之音高輪廓的先驗知識。長期特徵自適應於低延遲SMD(語

音/音樂鑑別)。

PLPCC的移動變因包含對涵蓋數個視窗之重疊分析視窗對各個PLPCC集合運算變因來強調最末的視窗。為了限制所導入的潛在延遲，分析視窗為非對稱性，只考慮目前視窗及過去史。於第一步驟中，PLPCC的移動平均$ma_m(k)$係對最末N個訊框運算說明如下：

$$ma_m(k)=\sum_{i=0}^{N-1}PLPC_m(k-i)\cdot w(i)$$

此處PLPm(k)為來自於第k個視窗共m個係數之第m個聲譜係數。移動變因$mv_m(k)$隨後定義為：

$$mv_m(k)=\sum_{i=0}^{N-1}(PLPC_m(k-i)-ma_m(k))^2\cdot w(i)$$

此處w為長度N之視窗，於本實施例中具有斜坡斜率定義如下：

$$w(i)=(N-i)/N\cdot(N+1)/2$$

該移動變因最終於聲譜維度上求平均：

$$mv(k)=\frac{1}{M}\sum_{m=0}^{M}mv_m(k)$$

語音之音高具有顯著性質，部分性質只在長期分析視窗上觀察得。確實語音之音高於有聲區段順利起伏波動，但罕見為恆定。相反地，音樂於整個音符期間表現出較常見恆定音高，而在變遷期間突然改變。經由觀察長期區段之音高輪廓，長期特徵涵蓋此項特性。音高輪廓參數pc(k)定義為：

$$pc(k)=\begin{cases}0 & 若 \quad |p(k)-p(k-1)|<1\\ 0,5 & 若 \quad 1\le|p(k)-p(k-1)|<2\\ 1 & 若 \quad 2\le|p(k)-p(k-1)|<20\\ 0,5 & 若 \quad 20\le|p(k)-p(k-1)|<25\\ 0 & 否則\end{cases}$$

此處p(k)為對於16 Hz取樣之LP殘餘信號於訊框指數k運算得之音高延遲。由音高輪廓參數，計算語音指標sm(k)，使得語音於有聲區段期間預期顯示出平順起伏的音高延遲，而於無聲區段期間顯示朝向高頻強力頻譜傾斜：

$$sm(k)=\begin{cases}nc(k)\cdot pc(k) & 若 \quad v(k)\ge 0.5\\ (1-nc(k))\cdot(1-tilt(k)) & 否則\end{cases}$$

此處nc(k)、tilt(k)及v(k)係定義如前(參考短期分類器)。然後語音指標藉如前述定義之視窗w加權且於最末N個訊框積分：

$$ams(k)=\sum_{i=0}^{N} m(k-1)w(i)$$

音高輪廓也是該信號是否適合用於語音編碼或音訊編碼的重要指標。確實語音編碼器主要係作用於時域，假設該信號為斜坡且於約5毫秒之短期區段為準穩態。藉此方式，可將語音的自然音高起伏波動有效模型化。相反地，同樣起伏波動對於探勘長期分析視窗之線性變換之一般音訊編碼器的效率有害。信號的主要能量展開於數個已變化的係數。

至於短期特徵，使用統計分類器也評估長期特徵，藉此獲得長期分類結果(DDC)。兩項特徵係使用N＝25個訊框

運算，亦即考慮該信號之400毫秒過去史。於縮小的一度空間使用3-GMM之前，先應用線性判別式分析(LDA)。表2顯示當對四個連續訊框區段分類時，對訓練集合及測試集合測量得之效能。

表2：長期特徵分類準確率，以%表示

| | 訓練集合 | 測試集合 |
|---|---|---|
| 語音 | 97.99 | 97.84 |
| 音樂 | 95.93 | 95.44 |
| 平均 | 96.96 | 96.64 |

根據本發明之實施例之組合型分類器系統適當組合短期特徵及長期特徵，使得兩種特徵對最終決策做出其特定貢獻。用於此項目的，可使用如前文說明之遲滯最終決策階段，此處記憶效應係藉DDC或長期鑑別線索(LTDC)驅動；而瞬時輸入係來自於IDC或短期鑑別線索(STDC)。兩個線索皆為長期分類器及短期分類器之輸出信號，如第1圖所示。基於IDC做決策，但藉DDC穩定化，DDC動態控制觸發狀態改變之臨界值。

長期分類器154先前使用LDA接著為3-GMM定義之長期特徵及短期特徵。DDC係等於最末4 X K訊框運算得之語音類別及音樂類別之長期分類器似然度之對數比。考慮的訊框數目可隨著參數K改變，俾便對最終決策加上更多或更少的記憶效應。相反地，短期分類器只使用有5-GMM之短期特徵，顯示效能與複雜度間之良好折衷。IDC係等於只對最末4個訊框運算得語音類別及音樂類別之短期分類器似

然度之對數比。

為了評估本發明辦法，特別對已切換之音訊編碼，評估三種不同效能。第一效能測量值為習知語音對音樂(SvM)效能。係對音樂項目及語音項目之一個大集合做評估。第二效能測量係對每3秒鐘語音區段與音樂區段交替之一個大型獨特項做測量。則該鑑別準確度稱作為音樂之後/之前語音(SabM)效能，主要係反映出系統的反應性。最後，經由對語音/音樂項目之一個大集合進行分類來評估決策的穩定性。語音與音樂間之混合對各項目係於不同位準進行。然後經由運算於訊框總數發生之類別切換數目比，獲得語音/音樂(SoM)效能。

長期分類器及短期分類器用作為評估習知單一分類器辦法之參考。短期分類器顯示良好反應性，同時具有低穩定性及較低總鑑別能力。另一方面，長期分類器，特別將訊框數目提高4ＸK，經由折衷決策反應性而可達成較高穩定性及鑑別能力。比較前述習知辦法，根據本發明之組合型分類器系統之效能有若干優點。其中一項優點為可保有良好純粹語音對音樂鑑別效能，同時保有系統之反應性。另一項優點為反應性與穩定性間做出良好折衷。

後文參考第4圖及第5圖，示例顯示編碼及解碼方案實例，其包括根據本發明之實施例運算之鑑別器或決策階段。

根據第4圖所示編碼方案實例，單聲信號、立體聲信號或多頻道信號係輸入一共通前處理階段200。

共通前處理階段200具有聯合立體聲功能、環繞功能及

/或帶寬擴展功能。於階段200之輸出端，有單聲頻道、立體聲頻道或多頻道其係輸入一個或多個開關202。當階段200有兩個或多個輸出端，例如當階段200輸出一立體聲信號或一多頻道信號時，可對階段200之各個輸出端設置開關202。舉例言之，立體聲信號之第一頻道可為語音頻道，立體聲信號之第二頻道可為音樂頻道。於此種情況下，決策階段204中之決策於同一個時間瞬間於兩個頻道間可不同。

開關202係藉決策階段204控制。決策階段包含根據本發明之實施例之一鑑別器，及接收輸入階段200之一信號或由階段200輸出之一信號作為輸出信號。另外，決策階段204也包括含括於該單聲信號、立體聲信號或多頻道信號之旁資訊，或該旁資訊至少係與此種信號關聯，此處該存在的資訊係於原先產生單聲信號、立體聲信號或多頻道信號時產生。

於一個實施例中，決策階段並未控制前處理階段200，階段204與階段200間之箭頭不存在。於又一個實施例中，階段200之處理係藉決策階段204控制至某個程度，俾便基於該決策選擇於階段200之一項或多項參數。但如此不影響階段200之一般演繹法則，使得階段200之主要功能為作用狀態而與階段204之決策無關。

決策階段204致動開關202俾便將共通前處理階段之輸出信號饋至第4圖上分支示例說明之一頻率編碼部206或第4圖下分支示例顯示之一LPC域編碼部208。

於一個實施例中，開關202是介於二編碼分支206、208

間切換。於額外實施例中，可有額外編碼分支，諸如第三編碼分支，或甚至第四編碼分支或甚至更多編碼分支。於有三個編碼分支之一個實施例中，第三編碼分支可類似第二編碼分支，但包括與第二分支208之激勵編碼器210不同之一激勵編碼器。於此種實施例中，第二分支包含LPC階段212及基於碼簿之激勵編碼器210諸如ACELP；及第三分支包含LPC階段及於LPC階段輸出信號之頻譜表示型態運算之一激勵編碼器。

頻域編碼分支包含一頻譜變換區塊214，其可操作來將該共通前處理階段輸出信號變換成頻譜域。頻譜變換區塊可包括MDCT演繹法則、QMF、FFT演繹法則、子波分析或濾波器組諸如具有某個數目之濾波器組頻道之臨界取樣的濾波器組，此處於本濾波器組之子頻帶信號可為實數值信號或複數值信號。頻譜變換區塊214之輸出係使用頻譜音訊編碼器216編碼，頻譜音訊編碼器216可包括如由AAC編碼方案已知之處理區塊。

下編碼分支208包含一來源模型分析器諸如LPC 212，其輸出兩種信號。一種信號為LPC資訊信號，其係用於控制LPC合成濾波器之濾波特性。此種LPC資訊傳輸至解碼器。另一個LPC階段212輸出信號為激勵信號或LPC域信號，其係輸入激勵編碼器210。激勵編碼器210可來自於任何來源濾波器模型編碼器諸如CELP編碼器、ACELP編碼器或任何其它處理LPC域信號之編碼器。

另一種激勵編碼器之實施例為該激勵信號之變換編

碼。於此種實施例中，激勵信號並未使用ACELP碼簿機制編碼，反而激勵信號被變換成頻譜表示型態，頻譜表示型態數值諸如濾波器組情況下之子頻帶信號或變換諸如FFT情況下之頻率係數係經編碼來獲得資料壓縮。此種激勵編碼器之實施例為由AMR-WB+已知之TCX編碼模式。

於決策階段204之決策可為信號自適應性，使得決策階段204進行音樂/語音鑑別，且控制開關202，使得音樂信號係輸入上分支206，而語音信號係輸入下分支208。於一個實施例中，決策階段204將其決策資訊饋入一輸出位元流，因此解碼器可使用本決策資訊來進行正確解碼運算。

此種解碼器示例說明於第5圖。於傳輸後，藉頻譜音訊編碼器216輸出之信號係輸入頻譜音訊解碼器218。頻譜音訊解碼器218之輸出係輸入時域變換器220。第4圖之激勵編碼器210之輸出信號係輸入激勵解碼器222，其輸出一LPC域信號。該LPC域信號係輸入LPC合成階段224，其接收由相對應的LPC分析階段212所產生之LPC資訊作為額外輸入信號。時域變換器220之輸出信號及/或LPC合成階段224之輸出信號係輸入一開關226。開關226係透過開關控制信號控制，該開關控制信號例如係由決策階段204所產生，或由外部提供諸如由原先單聲信號、立體聲信號或多頻道信號之形成器由外部提供。

開關226之輸出信號為完全單聲信號，其隨後輸入一共通後處理階段228，該階段執行聯合立體聲處理或帶寬擴展處理等。另外，開關之輸出信號也可為立體聲信號或多頻

道信號。當前處理包括頻道減至二頻道時，其為立體聲信號。當頻道減至三頻道或絲毫也無頻道減少而只進行頻譜帶複製時，其甚至可為多頻道信號。

依據該共通後處理階段之特定功能而定，單聲信號、立體聲信號或多頻道信號被輸出，當該共通後處理階段228執行帶寬擴展操作時，具有比輸入區塊228更大的帶寬。

於一個實施例中，開關226介於兩個解碼分支218、220與222、224間切換。於又一實施例中，可有額外解碼分支，諸如第三解碼分支或甚至第四解碼分支或甚至更多解碼分支。於有三個解碼分支之實施例中，第三解碼分支可類似第二解碼分支，但包括與於第二分支222、224之激勵解碼器222不同之激勵解碼器。於此種實施例中，第二分支包含LPC階段224及基於碼簿之激勵解碼器諸如於ACELP；而第三分支包含LPC階段及於LPC階段224輸出信號之頻譜表示型態上操作之一激勵解碼器。

於另一個實施例中，該共通前處理階段包含一環繞/立體聲區塊，其產生聯合立體聲參數及一單聲輸出信號作為輸出信號，該單聲輸出信號係經由將具有兩個或多個頻道之輸入信號降混所產生。通常，於該區塊輸出端之信號可為有更多頻道之信號，但因降混操作，於該區塊輸出端之頻道數目將小於輸入該區塊之頻道數目。於本實施例中，頻率編碼分支包含一頻譜變換階段及一隨後連結的量化/編碼階段。該量化/編碼階段可包括如由近代頻域編碼器諸如AAC編碼器已知之任一項功能。此外，於該量化/編碼階

段之量化操作可透過心理聲學模型控制，該心理聲學模型產生心理聲學資訊諸如對該頻率之心理聲學遮蔽臨界值，此處本資訊係輸入該階段。較佳係使用MDCT操作進行頻譜變換，又更佳係使用時間已翹曲的MDCT操作，此處強度或通常為翹曲強度可控制於零至高翹曲強度間。於零翹曲強度中，MDCT操作為技藝界已知之直通式MDCT操作。LPC域編碼器包括ACELP核心，計算音高增益、音高滯後及/或碼簿資訊諸如碼簿指數及碼增益。

雖然若干圖式示例說明裝置之方塊圖，但須注意此等圖式同時也示例說明一種方法，其中各個方塊之功能係對應於方法之步驟。

前文說明之本發明實施例係基於包含不同區段或不同訊框之一音訊輸入信號做說明，該等不同區段或訊框係與語音資訊或音樂資訊有關。本發明並非限於此等實施例，反而將包含至少第一型區段及第二型區段之一信號之不同區段分類之辦法也可應用於包含三個或更多個不同區段類型之音訊信號，各區段類型期望藉不同的編碼方案編碼。此等區段類型之實例為：

-穩態/非穩態區段可用於使用不同濾波器組、視窗或編碼自適應性。舉例言之，暫態應使用細緻時間解析度濾波器組編碼；純粹彎曲須藉細緻頻率解析度濾波器組編碼。

-有聲/無聲：有聲區段可藉語音編碼器諸如CELP良好處理；但用於無聲區段則浪費太多位元。參數編碼將較為有效。

-靜默/作用狀態：靜默可使用比作用狀態區段更少的位元編碼。

-諧波/非諧波：較佳係使用於頻域之線性預測用於諧波區段編碼。

此外，本發明並非囿限於音訊技術領域，反而所述分類信號之辦法也可應用至其它種信號，例如視訊信號或資料信號，其中個別信號包括不同類型之區段而要求不同的處理，例如：

本發明可自適應於全部需要一時間信號分段之即時應用。舉例言之，來自於監控視訊攝影機之面部檢測可基於一分類器，該分類器判定一訊框之各個像素(此處一訊框係對應於時間n)所拍相片(是否屬於一個人的臉部)。該分類(亦即臉部分段)係對該視訊串流之各個單一訊框進行。但使用本發明，目前訊框的分段可考慮過去連續的訊框，利用連續圖像有強力相關性之優點而獲得更佳分段準確度。則可應用兩個分類器。一個只考慮目前訊框，另一個分類器考慮包括目前訊框及過去訊框之一訊框集合。最後分類器積分訊框集合，判定臉部位置之機率區。該分類器之判定只對目前訊框進行判定，隨後與該機率區做比較。然後讓決策生效或經修改。

本發明之實施例使用開關用於介於二分支間切換，使得只有一個分支接收欲處理信號，而另一個分支並未接收信號。但於另一個實施例中，開關將配置於處理階段或處理分支例如音訊編碼器或語音編碼器之後，故二分支可並

列處理同一個信號。由其中一個分支輸出之信號被選用來輸出，例如被寫入一輸出位元流。

雖然本發明之實施例係基於數位信號做說明，但其區段係藉於特定取樣率獲得之預定樣本數目測定，本發明並未限於此等信號，反而本發明也可應用於類比信號，其中該區段係由類比信號之特定頻率範圍或時間週期決定。此外，本發明之實施例將組合包括鑑別器之編碼器做說明。基本上發現根據本發明之實施例用於分類信號之方法也可應用於接收一已編碼信號之解碼器，不同編碼方案可經分類來允許已編碼信號提供予一適當解碼器。

依據本發明方法之若干實施要求，本發明方法可於硬體或於軟體實施。實施可使用數位儲存媒體進行，特別為有可電子讀取控制信號儲存於其上之碟片、DVD或CD，其係與可規劃電腦系統協力合作因而可執行本發明方法。因此本發明為一種有程式碼儲存於一機器可讀取載體上之一種電腦程式產品，該程式碼於電腦程式產品於電腦上跑時可運算來執行本發明方法。換言之，本發明方法為一種具有程式碼之一電腦程式，用於當該電腦程式於電腦上跑時該程式碼可執行至少一種本發明之方法。

前述實施例僅供舉例說明本發明之原理。須了解此處所述配置及細節之修改及變化為熟諳技藝人士顯然易知。因此意圖僅受隨附之申請專利範圍之範圍所限而非受藉此處實施例之說明及解釋呈現之特定細節所限。

於前述實施例中，描述之信號包含多個訊框，其中評

估一目前訊框用於切換決策。注意評估用於切換決策之該信號之目前區段可為一個訊框，但本發明並非限於此種實施例。反而該信號之一區段也可包含多數亦即兩個或更多個訊框。

此外，於前述實施例中，短期分類器及長期分類器使用相同短期特徵。此種辦法由於不同理由而可使用，例如只需運算短期特徵一次，藉兩個分類器以不同方式探勘短期特徵，將減少系統的複雜度，原因在於該短期特徵將藉短期分類器或長期分類器中之一者計算而提供予另一個分類器。又，短期分類器結果與長期分類器結果間之比較將更具有相關性，原因在於兩個分類器共享共通特徵，經由比較長期分類結果與短期分類結果，更容易推定於長期分類結果中目前訊框之貢獻。

但本發明並未限於此種辦法，長期分類器並未限於使用與短期分類器相同的短期特徵，亦即短期分類器與長期分類器二者可計算彼此不同之其個別的短期特徵。

雖然前述實施例述及使用PLPCC作為短期特徵，但須注意也可考慮其它特徵，例如PLPCC之變化例。

## 【圖式簡單說明】

第1圖為根據本發明之實施例，一種語音/音樂鑑別器之方塊圖；

第2圖示例顯示由第1圖之鑑別器之長期分類器及短期分類器所使用之分析視窗；

第3圖示例顯示用於第1圖之鑑別器之遲滯決策；

第4圖為包含根據本發明之實施例之一鑑別器之一編碼方案實例之方塊圖；

第5圖為與該第4圖之編碼方案相對應之解碼方案之方塊圖；

第6圖顯示用於依據一音訊信號之鑑別而分開編碼語音及音樂之一種習知編碼器設計；及

第7圖示例顯示於第6圖所示編碼器設計中遭遇之延遲。

【主要元件符號說明】

100...語音編碼分支
102...語音編碼器
104...音樂編碼分支
106...音樂編碼器
108...多工器
110...輸入線
112...開關
114...切換控制器、開關控制器
116...語音/音樂鑑別器
118...線、模式指示器信號
120...AAC長區塊
122...AAC短區塊
124...AMR-WB+超訊框
126...訊框
128...訊框、目前區段
130...已解碼訊框
132...前瞻訊框
150...短期分類器
152...輸出線、輸出信號
154...長期分類器
156...輸出線、輸出信號
158...遲滯決策電路
160...輸出信號、輸出線、語音/音樂決策信號
162...長期分類器視窗
164...前瞻
166...短期分類器視窗
200...輸出階段
202...開關
204...決策階段

206...頻率編碼部、頻率編碼分支

208...LPC域編碼部、LPC域編碼分支

210...激勵編碼器

212...LPC階段、來源模型分析器

214...頻譜變換區塊

216...頻譜音訊編碼器

218...頻譜音訊解碼器

220...時域變換器

222...激勵解碼器

224...LPC合成階段

226...開關

228...共通後處理階段、區塊

## 七、申請專利範圍：

1. 一種用於分類一音訊信號之不同區段之方法，該音訊信號包含語音及音樂區段，該方法包含：

基於擷取自該音訊信號之至少一個短期特徵來判定該音訊信號之目前區段是否為一語音區段或一音樂區段，以短期分類該音訊信號，並遞送一短期分類結果以指示該音訊信號之該目前區段為語音區段或音樂區段；

基於擷取自該音訊信號之至少一個短期特徵及至少一個長期特徵來判定該音訊信號之目前區段是否為語音區段或音樂區段，以長期分類該音訊信號，及遞送一長期分類結果以指示該音訊信號之該目前區段為語音區段或音樂區段；及

組合該短期分類結果及該長期分類結果而提供一輸出信號來指示該音訊信號之該目前區段是否為語音區段或為音樂區段。

2. 如申請專利範圍第1項之方法，其中該組合步驟包含基於該短期分類結果與長期分類結果之比較而提供該輸出信號。

3. 如申請專利範圍第1項之方法，其中

該至少一個短期特徵係經由分析欲分類之該音訊信號之目前區段獲得；及

該至少一個長期特徵係經由分析該音訊信號之該目前區段及該音訊信號之一個或多個先前區段獲得。

4. 如申請專利範圍第1項之方法，其中

該至少一個短期特徵係經由分析一第一長度之分析視窗及一第一分析方法獲得；及

該至少一個長期特徵係經由分析一第二長度之分析視窗及第二分析方法獲得，該第一長度係比該第二長度短，及該第一分析方法與該第二分析方法不同。

5. 如申請專利範圍第4項之方法，其中該第一長度跨據該音訊信號之一目前區段，該第二長度跨據該音訊信號之該目前區段及該音訊信號之一個或多個先前區段，及該第一長度與該第二長度包含涵蓋一分析週期之一額外週期。

6. 如申請專利範圍第1項之方法，其中組合該短期分類結果與該長期分類結果包含基於一組合結果之遲滯決策，其中該組合結果包含該短期分類結果及該長期分類結果，各自藉一預定之加權因數加權。

7. 如申請專利範圍第1項之方法，其中該音訊信號為一數位信號及該音訊信號之一區段包含以特定取樣率獲得之預定數目樣本。

8. 如申請專利範圍第1項之方法，其中

該至少一個短期特徵包含PLPCC參數；及

該至少一個長期特徵包含音高特性資訊。

9. 如申請專利範圍第1項之方法，其中用於短期分類之該短期特徵與用於長期分類之該短期特徵為相同或相異。

10. 一種用於處理包含語音及音樂區段之一音訊信號之方

法，該方法包含：

根據申請專利範圍第1項之方法分類該信號之一目前區段；

依據由該分類步驟所提供之該輸出信號，來依照第一方法或第二方法處理該目前區段；及

輸出該已處理的區段。

11. 如申請專利範圍第10項之方法，其中

當該輸出信號指示該區段為一語音區段時，該區段係藉語音編碼器處理；及

當該輸出信號指示該區段為一音樂區段時，該區段係藉音樂編碼器處理。

12. 如申請專利範圍第11項之方法，進一步包含：

組合該已編碼區段與來自指出該區段類型之該輸出信號之資訊。

13. 一種電腦程式，當其於一電腦上執行時用以進行如申請專利範圍第1項之方法。

14. 一種鑑別器，包含：

一短期分類器，其配置來接收一音訊信號及判定該音訊信號之目前區段是否為一語音區段或一音樂區段，且提供基於擷取自該音訊信號之至少一個短期特徵之該音訊信號之一短期分類結果，該短期分類結果指示該音訊信號之該目前區段為語音區段或音樂區段，該音訊信號包含語音及音樂區段；

一長期分類器，其配置來接收一音訊信號及判定該

音訊信號之目前區段是否為語音區段或音樂區段，且提供基於擷取自該音訊信號之至少一個短期特徵及至少一個長期特徵之該音訊信號之一長期分類結果，該長期分類結果指示該音訊信號之該目前區段為語音區段或音樂區段；及

一決策電路，其配置來組合該短期分類結果及該長期分類結果而提供一輸出信號，該輸出信號指示該音訊信號之該目前區段是否為語音區段或為音樂區段。

15. 如申請專利範圍第14項之鑑別器，其中該決策電路係配置來基於該短期分類結果與長期分類結果之比較提供該輸出信號。

16. 一種音訊信號處理裝置，其包含：

一輸入裝置，其配置來接收一欲處理之音訊信號，其中該音訊信號包含語音及音樂區段；

一第一處理級裝置，其配置來處理語音區段；

一第二處理級裝置，其配置來處理音樂區段；

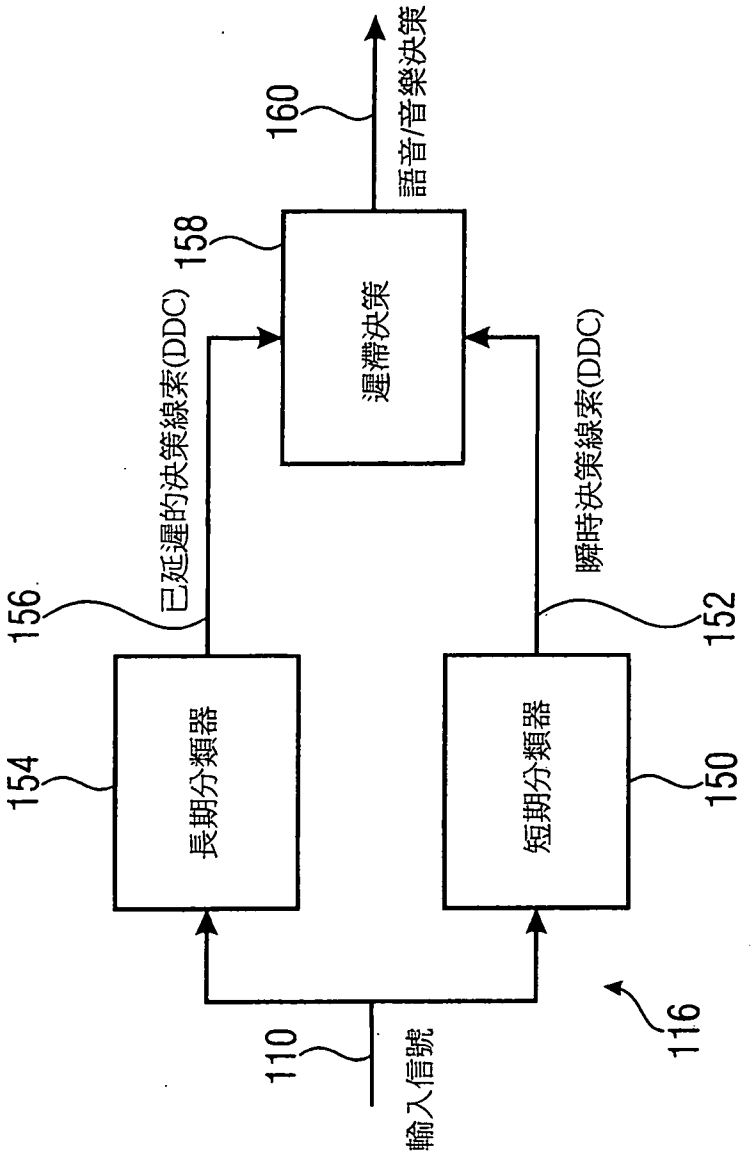
耦接至該輸入裝置之如申請專利範圍第14 項或第15項之一鑑別器；及

耦接於該輸入裝置與該第一處理級裝置及第二處理級裝置間之一切換裝置，其配置來依據得自該鑑別器之輸出信號而將得自該輸入裝置之音訊信號施加至該第一處理級裝置及第二處理級裝置中之一者。

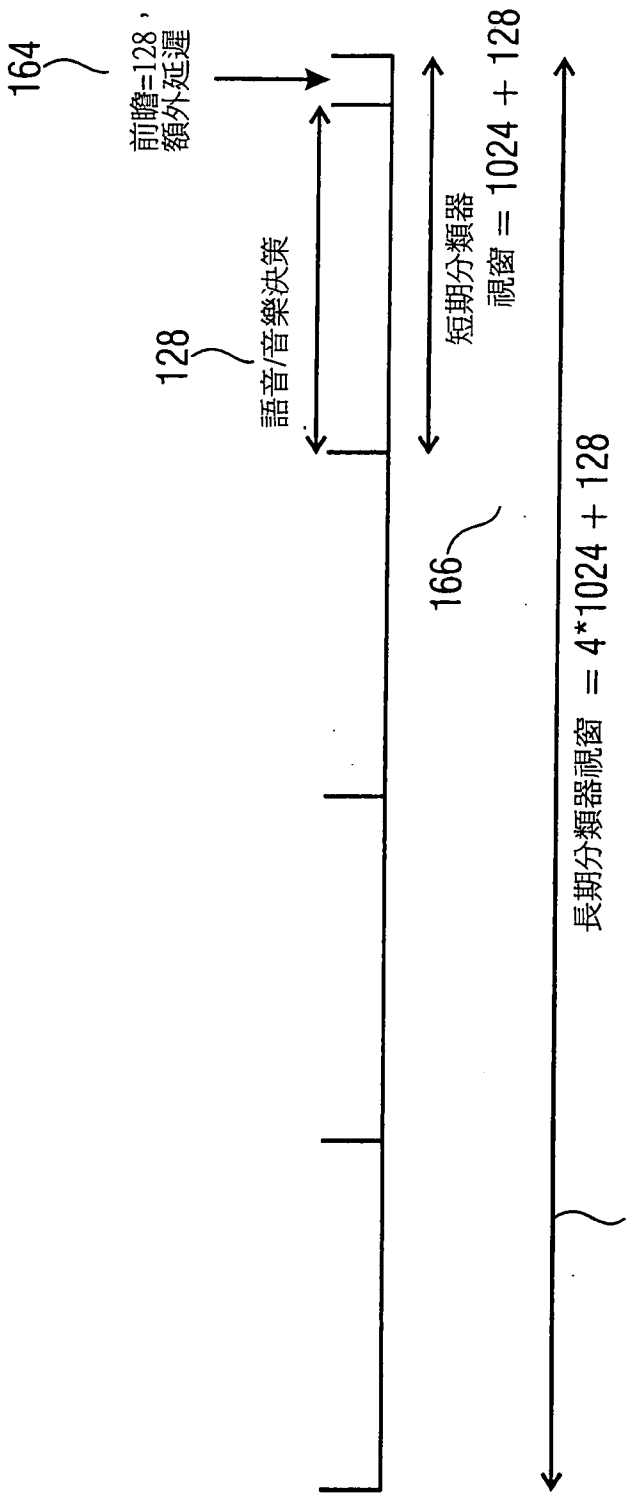
17. 一種音訊編碼器，其包含：

如申請專利範圍第16項之音訊信號處理裝置，

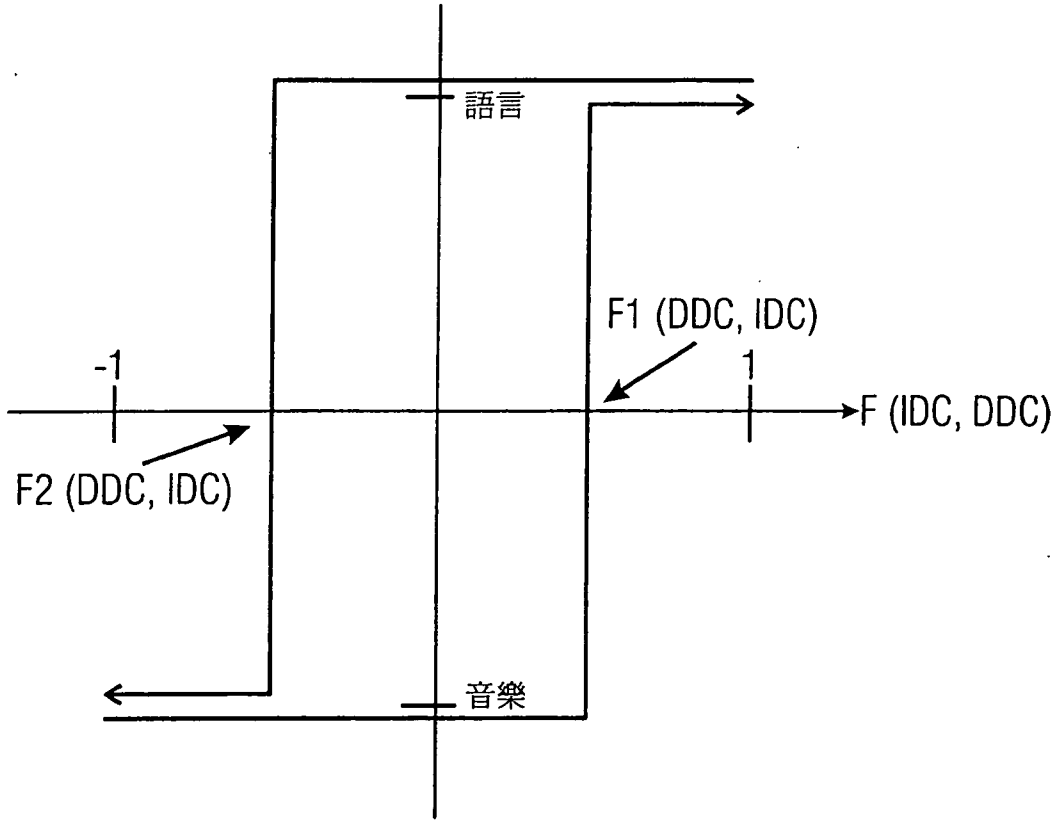
其中該第一處理級裝置包含一語音編碼器及該第二處理級裝置包含一音樂編碼器。


輸入信號
110
116
短期分類器
150
瞬時決策線索(DDC)
152
長期分類器
154
已延遲的決策線索(DDC)
156
遲滯決策
158
語音/音樂決策
160

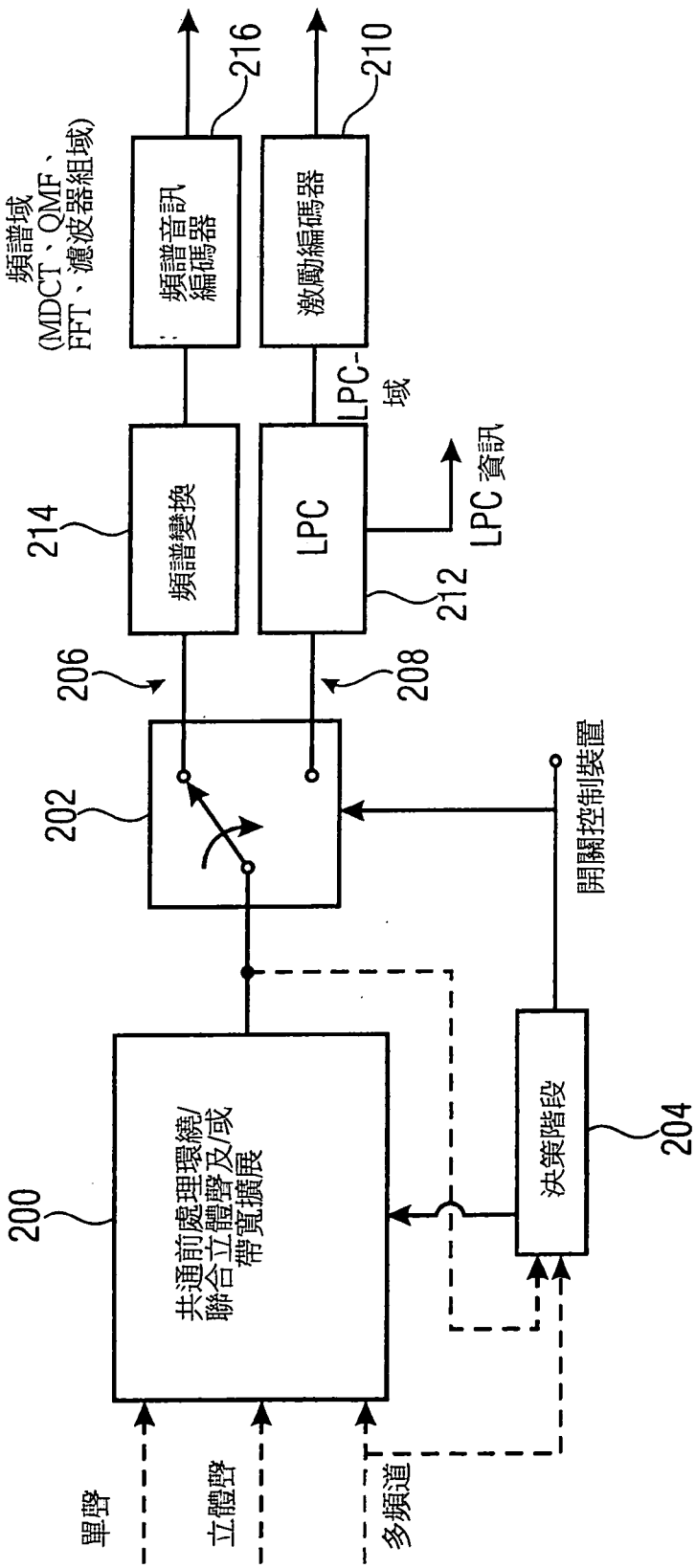
第1圖


164
前瞻=128，
額外延遲
128
語音/音樂決策
短期分類器
視窗 = 1024 + 128
166
長期分類器視窗 = 4*1024 + 128
162

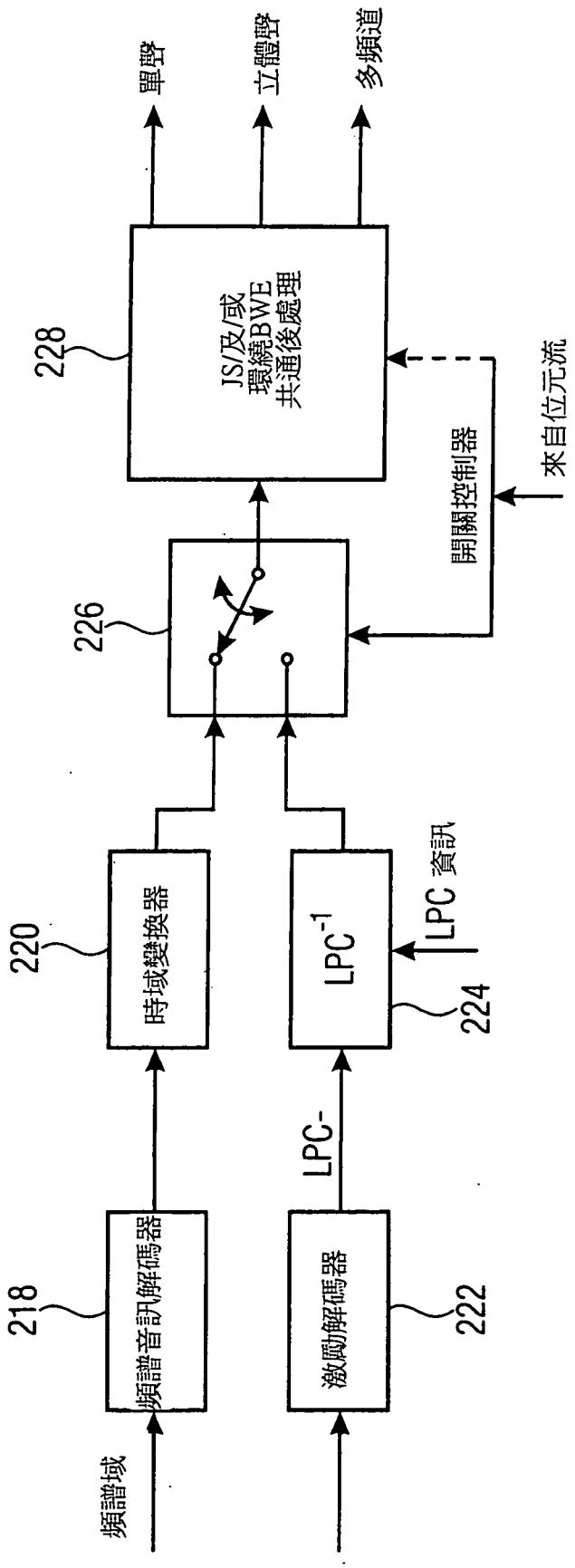
第2圖



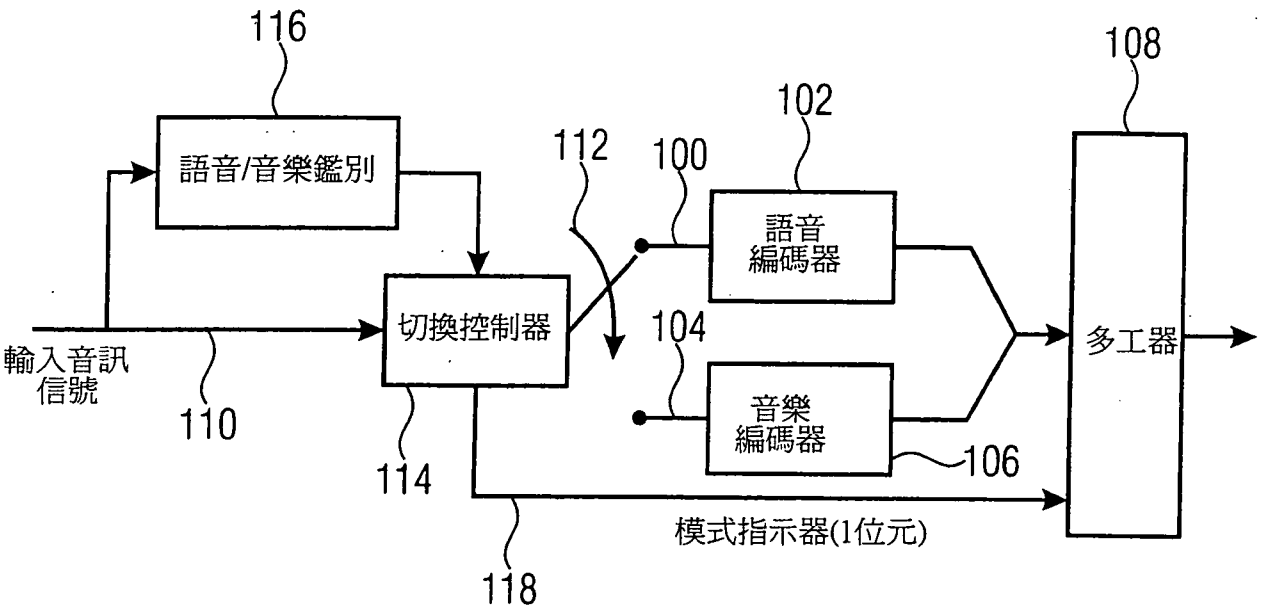

第3圖


單聲
立體聲
多頻道
共通前處理環繞/
聯合立體聲及/或
帶寬擴展
200
決策階段
204
開關控制裝置
202
206
208
頻譜變換
214
LPC
212
LPC 資訊
LPC-
域
激勵編碼器
210
頻譜音訊
編碼器
216
頻譜域
(MDCT、QMF、
FFT、濾波器組域)

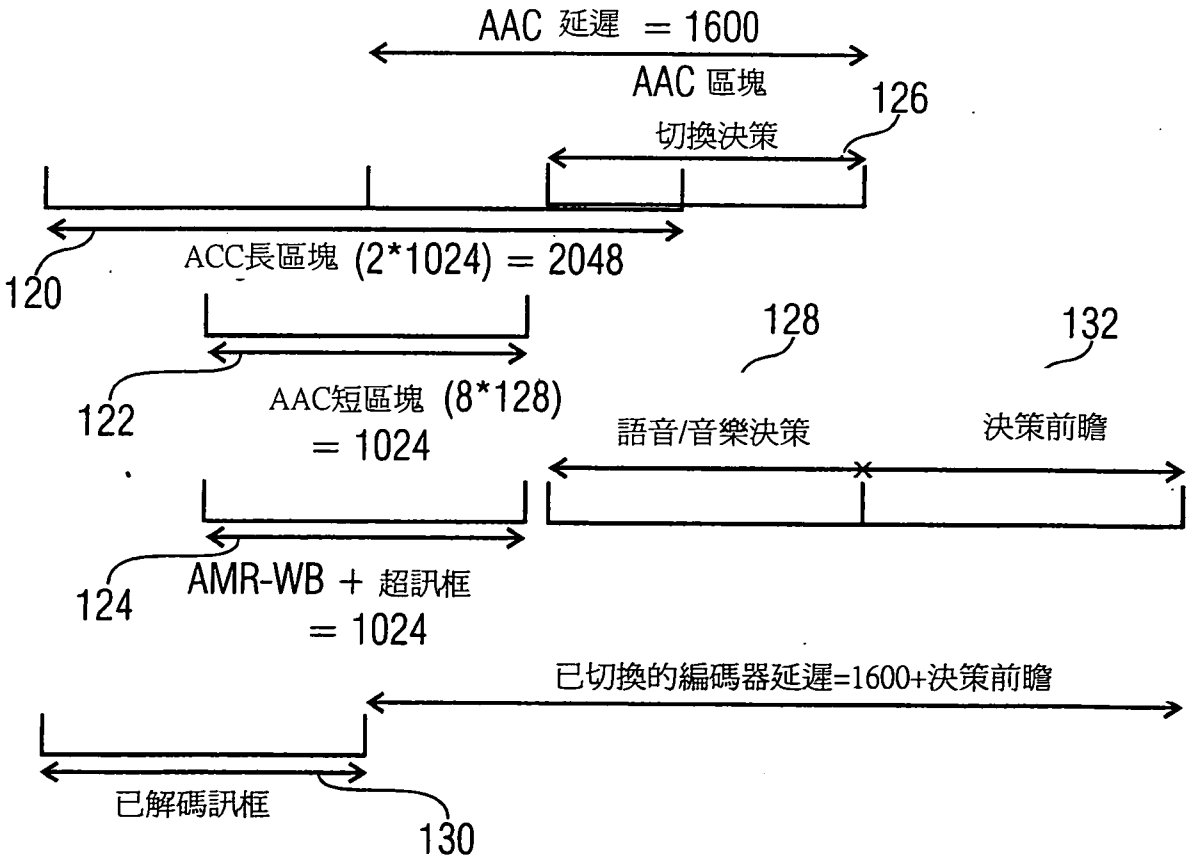
第4圖
(編碼器)


單聲
立體聲
多頻道
228
JS/及/或
環繞BWE
共通後處理
來自位元流
開關控制器
226
220
時域變換器
LPC$^{-1}$
LPC 資訊
224
LPC-
218
頻譜音訊解碼器
激勵解碼器
222
頻譜域

第5圖
(解碼器)


116
語音/音樂鑑別
108
102
112
100
語音
編碼器
切換控制器
104
多工器
輸入音訊
信號
110
音樂
編碼器
106
114
模式指示器(1位元)
118

第6圖


AAC 延遲 = 1600
AAC 區塊
126
切換決策
ACC長區塊 (2*1024) = 2048
120
128
132
AAC短區塊 (8*128)
= 1024
122
語音/音樂決策
決策前瞻
AMR-WB + 超訊框
= 1024
124
已切換的編碼器延遲=1600+決策前瞻
已解碼訊框
130

第7圖