M653962

(19)中華民國智慧財產局

(12)新型說明書公告本 (11)證書號數：TW M653962 U

(45)公告日： 中華民國 113 (2024) 年 04 月 11 日

(21)申請案號：112211484 (22)申請日： 中華民國 112 (2023) 年 10 月 24 日

(51)Int. Cl.： *G11B20/10 (2006.01)*

(71)申請人：華碩電腦股份有限公司(中華民國) ASUSTEK COMPUTER INC. (TW)
臺北市北投區立德路 15 號

(72)新型創作人：范植貿 FAN, CHI-MAO (TW)；廖書巧 LIAO, SHU-CHIAO (TW)；吳信夫 WU, SHIN-FU (TW)；沈鴻儒 SHEN, HONG-RU (TW)；林鋐儒 LIN, HUNG-JU (TW)；陳冠廷 CHEN, GUAN-TING (TW)；郭志榮 KUO, CHIH-RONG (TW)

(74)代理人：李文賢；盧建川；陳政大

申請專利範圍項數：15 項 圖式數：9 共 27 頁

(54)名稱

數位音訊系統

(57)摘要

本案揭露一種數位音訊系統，包含一處理裝置以及一資料庫。處理裝置內建有一音源分離模型及一音訊轉錄模型，處理裝置載入一輸入音訊檔，並透過音源分離模型對輸入音訊檔進行關鍵特徵擷取及訊號分離處理，以獲得複數分離音檔，音訊轉錄模型分別載入該些分離音檔，以透過音訊轉錄模型進行音訊轉錄而分別取得對應之複數轉換檔案，處理裝置將該些轉換檔案分別轉換為複數樂譜。資料庫電性連接處理裝置，用以儲存輸入音訊檔及該些樂譜。

指定代表圖：

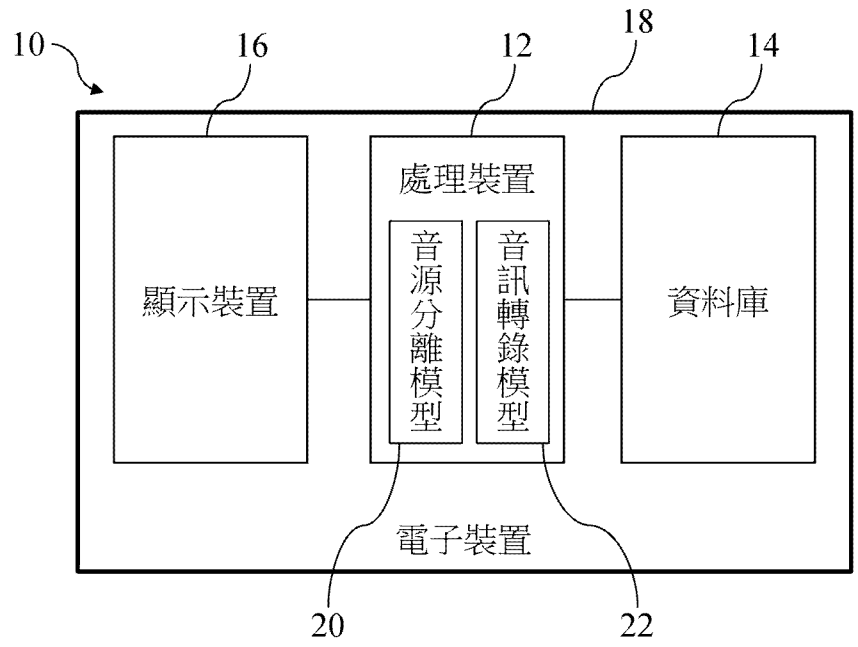


圖1

符號簡單說明：

10:數位音訊系統

12:處理裝置

14:資料庫

16:顯示裝置

18:電子裝置

20:音源分離模型

22:音訊轉錄模型

專利案號：112211484

公告本

M653962

【新型摘要】

【中文新型名稱】數位音訊系統

【中文】

本案揭露一種數位音訊系統，包含一處理裝置以及一資料庫。處理裝置內建有一音源分離模型及一音訊轉錄模型，處理裝置載入一輸入音訊檔，並透過音源分離模型對輸入音訊檔進行關鍵特徵擷取及訊號分離處理，以獲得複數分離音檔，音訊轉錄模型分別載入該些分離音檔，以透過音訊轉錄模型進行音訊轉錄而分別取得對應之複數轉換檔案，處理裝置將該些轉換檔案分別轉換為複數樂譜。資料庫電性連接處理裝置，用以儲存輸入音訊檔及該些樂譜。

【指定代表圖】　　圖1

【代表圖之符號簡單說明】

10:數位音訊系統

12:處理裝置

14:資料庫

16:顯示裝置

18:電子裝置

20:音源分離模型

22:音訊轉錄模型

【新型說明書】

【中文新型名稱】數位音訊系統

【技術領域】

【0001】 本案係有關一種結合人工智慧（AI）技術進行音訊轉譜之數位音訊系統。

【先前技術】

【0002】 目前若想彈奏尚未公開樂譜的音樂時，往往都是藉由具備樂理知識的專業人士，經由聽覺與音感慢慢的將整首歌的旋律記錄於樂譜上，分享給其他人。這個過程是曠日廢時的，且若遇到由許多樂器組合的複雜旋律，更是會花費大量人力與資源來記錄整理至樂譜上，且由於是人工記錄，也容易發生差錯。

【0003】 目前將輸入音訊轉錄成音樂數位介面（Musical Instrument Digital Interface，MIDI）的方法，雖然節省了人工針對特定樂器將音符一一對應至樂譜的過程，而大幅加速了整個流程，但因為採取傳統依據頻率轉換的數學公式，例如短時距傅立葉轉換（Short-Time Fourier Transform，STFT）、常數Q轉換（Constant Q Transform，CQT）等的方式，往往會有泛化（Generalization）能力不高、需花費較長時間進行推論等問題。更重要的是，傳統的曲譜作法，缺少了音源分離（Music Source Separation，MSS）技術，導致轉錄成MIDI檔案時，容易受到輸入音訊的背景環境音或其他副旋律樂器所影響，導致輸出的樂譜效果不彰。

【新型內容】

【0004】 本案提供一種數位音訊系統，包含一處理裝置以及一資料庫。在數位音訊系統中，處理裝置係內建有一音源分離模型及一音訊轉錄模型，處理裝置接收一輸入音訊檔，並透過音源分離模型對輸入音訊檔進行關鍵特徵擷取及訊號分離處理，以獲得複數分離音檔，音訊轉錄模型會分別載入該些分離音檔，以透過音訊轉錄模型分別取得對應之複數轉換檔案，處理裝置再將這些轉換檔案分別轉換為複數樂譜。資料庫電性連接處理裝置，用以儲存輸入音訊檔及樂譜。

【0005】 綜上所述，本案之數位音訊系統係為一種結合人工智慧（AI）技術的音訊轉譜流程的平台架構，在使用者匯入音樂後，經音源分離（Music Source Separation，MSS）與音訊轉錄（Music Transcription，MT）等相關技術，以取得特定樂器的樂譜輸出。由於本案採用人工智慧深度學習模型，依據龐大的數據庫支持，不會受限於單一樂器，除了擁有較高的曲譜正確性之外，其推論時間比傳統方法迅速、準確。並且，透過音源分離模型之設計，可用以分離複雜多樂器的旋律，此過程不僅可以讓後續的音訊轉錄更加準確，還可以取代傳統人工分離不同旋律的過程。

【圖式簡單說明】

【0006】

圖1為根據本案一實施例之數位音訊系統的方塊示意圖。

圖2為根據本案一實施例之數位音訊系統產生之樂譜顯示於顯示裝置上的結構示意圖。

圖3為根據本案一實施例之數位音訊系統使用之音源分離模型的架構示意圖。

圖4為根據本案一實施例之數位音訊系統使用之音訊轉錄模型的架構示意圖。

圖5為根據本案另一實施例之數位音訊系統使用之音訊轉錄模型的架構示意圖。

圖6為根據本案一實施例之音訊轉錄模型中之轉移視窗轉換單元的架構示意圖。

圖7為根據本案一實施例之數位音訊系統之應用程式介面示意圖。

圖8為根據本案一實施例於訓練音源分離模型或音訊轉錄模型之流程示意圖。

圖9為根據本案另一實施例之數位音訊系統的方塊示意圖。

【實施方式】

【0007】 以下將配合相關圖式來說明本案的實施例。此外，實施例中的圖式有省略部份元件或結構，以清楚顯示本案的技術特點。在這些圖式中，相同的標號表示相同或類似的元件或電路。

【0008】 請參閱圖 1 所示，一數位音訊系統 10 包含一處理裝置 12、一資料庫 14 以及一顯示裝置 16，處理裝置 12 電性連接至資料庫 14，且處理裝置 12 電性連接顯示裝置 16。在一實施例中，處理裝置 12、資料庫 14 及顯示裝置 16 可以內建在一電子裝置 18 中，例如，電子裝置 18 可為一個人電腦、一筆記型電腦或是一平板電腦等，但本案不以此為限。在另一實施例中，資料庫 14 更可獨立於電子裝置 18 之外，使

處理裝置 12 訊號連接至外部的資料庫 14，以進行資料的存取。

【0009】 如圖 1 所示，在數位音訊系統 10 中，處理裝置 12 係內建有一音源分離模型 20 及一音訊轉錄模型 22。處理裝置 12 自外部裝置（圖中未示）或資料庫 14 載入一輸入音訊檔，此輸入音訊檔可包含哼唱、交響樂、流行音樂等各種複雜多元的音頻訊號，且支援各式音訊格式的檔案類型（例如 mp3、wav、flac 等）。處理裝置 12 將輸入音訊檔載入至音源分離模型 20 中，以透過音源分離模型 20 對輸入音訊檔進行關鍵特徵擷取及訊號分離處理，以獲得複數分離音檔。然後，音訊轉錄模型 22 會分別載入這些分離音檔，以透過音訊轉錄模型 22 進行音訊轉錄而分別取得對應之複數轉換檔案，在一實施例中，此轉換檔案係為一樂器數位介面（MIDI）檔案。在取得各種樂器轉錄過後的轉換檔案之後，處理裝置 12 再將這該些轉換檔案分別轉換為複數樂譜 23，並可選擇性將樂譜 23 輸出並顯示在顯示裝置 16 上，如圖 2 所示。在取得各樂器的樂譜 23 之後，這些轉錄成功的樂譜 23 會以圖像檔案格式儲存在資料庫 14 中，且輸入音訊檔亦儲存在資料庫 14 中，以方便後續使用者透過顯示裝置 16 觀看或查閱。

【0010】 在一實施例中，由於轉換過後樂譜 23 有時會因為音符排列的隨機性，導致產生的結果難以閱讀，因此處理裝置 12 更可透過一數學佈局演算法來整理音符，以符合樂曲的格式與方格，達到簡化樂譜易閱讀之目的。

【0011】 在一實施例中，請同時參閱圖 1 及圖 3 所示，本案使用之音源分離模型 20 更進一步包含一共享編碼器 28 以及一多頭解碼器 30，

且多頭解碼器 30 包含第一解碼器 301、第二解碼器 302 至第 N 解碼器 303，使共享編碼器 28 同時電性連接第一解碼器 301、第二解碼器 302 至第 N 解碼器 303。當有輸入音訊檔 24 載入時，處理裝置 12 先利用一短時距傅立葉轉換（STFT）將輸入音訊檔 24 轉換至一原始時頻圖 26，再將原始時頻圖 26 載入至共享編碼器 28 中進行關鍵特徵的提取。在經過權重提取關鍵特徵之後，再經由多頭解碼器 30 分別取得複數分離音源時頻圖 32，再利用一反短時距傅立葉轉換（iSTFT）將該些分離音源時頻圖 32 分別轉換為複數分離音檔 34。其中，輸入音訊檔 24 及該些分離音檔 34 係具有相同檔案格式。本案利用多頭解碼器 30 的方式，除了能利用單一模型實現多種樂器之分離音檔 34 輸出之外，還具備辨認輸入音訊檔 24 含有何種樂器之能力。再者，採用音源分離模型 20 除了有過濾環境背景音的優勢之外，對於一般使用者可以選擇針對人聲主旋律或樂器伴奏副旋律，進行分離；針對專業編曲者，除主副旋律分離之外，音源分離模型 20 還可透過前述過程辨別這首歌可能包含何種樂器，提供給使用者分析，以利後續創作編曲。

【0012】 在一實施例中，請同時參閱圖 1 及圖 4 所示，在音訊轉錄模型 22 中，係先將每一分離音檔 34 經一常數 Q 轉換來轉換成一時頻圖 36，並利用一卷積神經網路層 38 對時頻圖 36 提取特徵資訊，以獲得複數語音後驗特徵圖（Phonetic Posterior Gram，PPG）40，在此係以三個語音後驗特徵圖 40 為例，包含一起始點（Onset）後驗特徵圖 401、一音符（Note）後驗特徵圖 402 以及一音高（Pitch）後驗特徵圖 403，起始點後驗特徵圖 401 用來記錄音符的開始時間，音符後驗特徵圖 402

用來記錄各時間段的音符，音高後驗特徵圖 403 用來紀錄各時間的音高。每一語音後驗特徵圖 40 上的橫軸為時間，縱軸為對應類別（起始點、音符、音高）的後驗機率，利用交叉熵（cross-entropy）計算與地面實況（ground-truth）的差異逐漸學習收斂至對應的結果，而由於語音後驗特徵圖 40 的正負樣本比例差距很大（正樣本：有起始點、音符、音高出現的地方；負樣本：沒任何類別變化的黑色部分），並利用類別平衡（class-balanced）解決正負樣本差距問題（正樣本類別權重=0.95，負樣本類別權重=0.05），以獲得如圖所示之語音後驗特徵圖 40。最後，再對該些語音後驗特徵圖 40 進行一後處理，以產生轉換檔案（樂器數位介面檔案）42。其中，每一卷積神經網路層 38 更進一步至少包含一卷積層 381、一批次標準化層 382 以及一線性整流層 383，以提取時頻圖 36 中的重要特徵資訊。

【0013】 以詳細的音訊轉錄過程而言，如圖 4 所示，在音訊轉錄模型 22 中，先利用第一卷積神經網路層 384 對時頻圖 36 提取特徵資訊，以獲得音高後驗特徵圖 403；取得音高後驗特徵圖 403 之後，第二卷積神經網路層 385 會對音高後驗特徵圖 403 進行特徵資訊的提取，以產生音符後驗特徵圖 402；然後再利用第三卷積神經網路層 386 對時頻圖 36 提取特徵資訊後與音符後驗特徵圖 402 串接在一起，並經過一卷積層 44 處理後，以產生起始點後驗特徵圖 401。在取得起始點後驗特徵圖 401、音符後驗特徵圖 402 以及音高後驗特徵圖 403 之後，一起經過後處理整合，即可產生對應之轉換檔案 42。

【0014】 在一實施例中，請同時參閱圖 1 及圖 5 所示，為了增加音

訊轉錄的準確率，除了在特定位置使用卷積層 44、50、52 之外，例如時頻圖 36 之後的卷積層 50、52 以及起始點後驗特徵圖 401 之前的卷積層 44，其餘的處理模組皆都改為使用提取全域資訊特徵能力較強的轉移視窗轉換器層（Shifted-Window (Swin) Transformer layer）46。在音訊轉錄模型 22 中，係先將一分離音檔 34 經一常數 Q 轉換來轉換成一時頻圖 36，並經過一卷積層 50 處理後，再利用一轉移視窗轉換器層 46 對時頻圖 36 提取全域資訊特徵，以獲得複數語音後驗特徵圖 40，在此係以三個語音後驗特徵圖 40 為例，包含一起始點後驗特徵圖 401、一音符後驗特徵圖 402 以及一音高後驗特徵圖 403。再對該些語音後驗特徵圖 40 進行一後處理，以產生轉換檔案（樂器數位介面檔案）42。

【0015】 其中，每一轉移視窗轉換器層 46 更進一步包含複數轉移視窗轉換單元 48，其數量通常為 4-8 個的偶數個轉移視窗轉換單元 48。請同時參閱圖 5 及圖 6 所示，每一轉移視窗轉換單元 48 更包含依序串接的一第一層標準化層 481、一視窗多頭自注意力（Window Multi-Head Self-Attention）層 482、一第二層標準化層 483、一第一多層感知（Multi-Layer Perceptron，MLP）層 484、一第三層標準化層 485、一轉移視窗多頭自注意力（Shifted Window MSA）層 486、一第四層標準化層 487 以及一第二多層感知層 488，以利用視窗多頭自注意力層 482 降低計算複雜度，並利用轉移視窗多頭自注意力層 486 避免計算到重複的特徵區域。

【0016】 以詳細的音訊轉錄過程而言，如圖 5 所示，在音訊轉錄模型 22 中，時頻圖 36 先經過卷積層 50 處理後，利用第一轉移視窗轉換

器層 461 對處理後之時頻圖 36 提取特徵資訊，以獲得音高後驗特徵圖 403；取得音高後驗特徵圖 403 之後，第二轉移視窗轉換器層 462 會對音高後驗特徵圖 403 進行特徵資訊的提取，以產生音符後驗特徵圖 402；然後，時頻圖 36 先經過卷積層 52 處理後，再利用第三轉移視窗轉換器層 463 對處理後之時頻圖 36 提取特徵資訊後與音符後驗特徵圖 402 串接在一起，並經過一卷積層 44 處理後，以產生起始點後驗特徵圖 401。在取得起始點後驗特徵圖 401、音符後驗特徵圖 402 以及音高後驗特徵圖 403 之後，三者一起經過後處理整合，即可產生對應之轉換檔案 42。

【0017】 在一實施例中，為了進一步加速推論運算時間與降低硬體設備規格需求，處理裝置 12 更可利用一邊緣計算技術處理訓練好的音源分離模型 20 及音訊轉錄模型 22，使音源分離模型 20 及音訊轉錄模型 22 之權重檔轉換為邊緣計算技術所接受的檔案格式，以更佳容易部署於硬體規格受限的電子裝置 18。在一實施例中，邊緣計算技術係為 OpenVINO 技術。

【0018】 在一實施例中，處理裝置 12 係為一中央處理單元（central processing unit，CPU），或是其他一般用途或特殊用途的微處理器（microprocessor）、微控制單元（micro control unit，MCU）、數位信號處理器（digital signal processor，DSP）、可程式化控制器、特殊應用積體電路（application specific integrated circuit，ASIC），或其他類似元件或上述元件的組合，本案不以此為限。

【0019】 在一實施例中，請同時參閱圖 1 及圖 7 所示，處理裝置

12 更可提供一應用程式介面 54 顯示於顯示裝置 16 上，以透過此應用程式介面 54 執行將音訊轉為樂譜的整個軟體流程，並可透過此應用程式介面 54 選擇待轉譜之樂器。此應用程式介面 54 包含一選項區 541、一主畫面顯示區 542、一功能執行選項 543 以及一類別資訊區 544，其中類別資訊區 544 包含有人聲選項 544a、貝斯選項 544b 以及鼓選項 544c，以供選擇欲分離的樂器或人聲，但本案不以此為限。

【0020】 在本案之數位音訊系統 10 所使用之音源分離模型 20 及音訊轉錄模型 22 中，模型的效能會根據訓練資料的豐富度、準確度影響到最終的轉譜效果，而本案在使用數位音訊系統 10 中之音源分離模型 20 及音訊轉錄模型 22 進行轉譜之前，會先訓練音源分離模型 20 及音訊轉錄模型 22，並在完成音源分離模型 20 及音訊轉錄模型 22 的訓練之後，將音源分離模型 20 及音訊轉錄模型 22 串接起來，並結合應用程式介面 54，使處理裝置 12 具有完整的音源分離模型 20 及音訊轉錄模型 22。

【0021】 由於音源分離模型 20 及音訊轉錄模型 22 是處理不同的任務，所以使用的訓練資料集是完不同且各自獨立訓練的。以下將分別敘述音源分離模型 20 及音訊轉錄模型 22 的建立流程。

【0022】 處理裝置 12 在建立音源分離模型 20 之主要步驟如圖 8 所示之步驟 S10 至步驟 S14 所示。請同時參閱圖 1 及圖 8 所示，如步驟 S10 所示，準備訓練資料，本案可以使用現成的分離資料集（例如，MUSDB18 等）作為訓練資料，若想讓音源分離模型 20 能辨別更多不同種類的樂器（除了人聲、鼓與貝斯之外），亦可藉由自行產生各樂器

的音軌，混合成多樂器之音訊，以作為訓練資料。如步驟 S12 所示，開始訓練模型，本案採用多任務學習（Multi-Task Learning，MTL）的方式訓練模型，將訓練資料輸入至原始人工智慧模型中持續進行訓練修正，直至最終預測輸出之分離音檔與訓練資料中之音檔相符。最後如步驟 S14 所示，進行模型權重處理，在模型訓練完成之後，即可取得最終的模型權重，並據此建立最終之音源分離模型 20。在本案使用之音源分離模型 20 中，係採用雙向長短期記憶（Bidirectional Long Short Term Memory，BiLSTM）作為模型中主要的提取特徵技術。

【0023】 處理裝置 12 在建立音訊轉錄模型 22 之主要步驟亦如圖 8 所示，但在訓練音訊轉錄模型 22 中，係使用常數 Q 轉換後的頻譜圖與對應之樂器數位介面（MIDI）檔案的反常數 Q 轉換譜圖做為訓練資料對，將音訊轉錄模型 22 視為一個映射（Mapping）模型，模型架構係採用深度卷積神經網路（Convolution Neural Network，CNN）來實現。請同時參閱圖 1 及圖 8 所示，如步驟 S10 所示，準備訓練資料，本案可以使用常數 Q 轉換後的頻譜圖與對應之樂器數位介面檔案的反常數 Q 轉換譜圖做為訓練資料。如步驟 S12 所示，開始訓練模型，將訓練資料輸入至一原始人工智慧模型中持續進行訓練修正，直至最終預測輸出之轉換檔案（樂器數位介面檔案）與訓練資料中之轉換檔案相符。最後如步驟 S14 所示，進行模型權重處理，在模型訓練完成之後，即可取得最終的模型權重，並據此建立最終之音訊轉錄模型 22。

【0024】 在一實施例中，如圖 9 所示，處理裝置 12 在訓練音源分離模型 20 及音訊轉錄模型 22 時，更可採用一圖形處理器（GPU）56

進行訓練的運算處理，此圖形處理器 56 電性連接處理裝置 12，使處理裝置 12 可搭配圖形處理器 56 來分別訓練音源分離模型 20 及音訊轉錄模型 22，以加速訓練模型的時間。

【0025】 本案採用音源分離模型及音訊轉錄模型等之人工智慧模型實現相關功能，除了擁有較高的曲譜正確率之外，為了避免使用者因深度學習推論所需較高硬體設備之限制，本案更配合邊緣運算技術之應用，不僅加速了整體流程的推理速度、降低硬體規格門檻，同時保持原有模型的效果，使一般人也能受惠於本案提出的應用軟體，並達到應用人工智慧技術協助發展音樂娛樂教育產業的願景。對於一般使用者而言，本案之數位音訊系統可以協助一些音樂演奏者，因樂曲時常伴有多種樂器而難以捕捉單一樂器旋律、新歌沒樂譜等多項限制，而無法練習的使用者；而對於專業編曲者而言，本案之數位音訊系統可定位為編曲輔助工具，先快速產生各種樂器的樂譜輸出，再經由編曲者確認提取需要部份，故可以大幅減少編曲、譜曲的時間。

【0026】 綜上所述，本案之數位音訊系統係為一種結合人工智慧（AI）技術的音訊轉譜流程的平台架構，在使用者匯入音樂後，經音源分離（Music Source Separation，MSS）與音訊轉錄（Music Transcription，MT）等相關技術，以取得特定樂器的樂譜輸出。由於本案採用人工智慧深度學習模型（音源分離模型及音訊轉錄模型），依據龐大的數據庫支持，不會受限於單一樂器，除了擁有較高的曲譜正確性之外，其推論時間比傳統方法迅速、準確。並且，透過音源分離模型之設計，可用以分離複雜多樂器的旋律，此過程不僅可以讓後續的音訊

轉錄更加準確，還可以取代傳統人工分離不同旋律的過程。

【0027】 以上所述之實施例僅係為說明本案之技術思想及特點，其目的在使熟悉此項技術者能夠瞭解本案之內容並據以實施，當不能以之限定本案之專利範圍，即大凡依本案所揭示之精神所作之均等變化或修飾，仍應涵蓋在本案之申請專利範圍內。

【符號說明】

【0028】

10:數位音訊系統

12:處理裝置

14:資料庫

16:顯示裝置

18:電子裝置

20:音源分離模型

22:音訊轉錄模型

23:樂譜

24:輸入音訊檔

26:原始時頻圖

28:共享編碼器

30:多頭解碼器

301:第一解碼器

302:第二解碼器

303:第N解碼器

32:分離音源時頻圖

34:分離音檔

36:時頻圖

38:卷積神經網路層

381:卷積層

382:批次標準化層

383:線性整流層

384:第一卷積神經網路層

385:第二卷積神經網路層

386:第三卷積神經網路層

40:語音後驗特徵圖

401:起始點後驗特徵圖

402:音符後驗特徵圖

403:音高後驗特徵圖

42:轉換檔案

44:卷積層

46:轉移視窗轉換器層

461:第一轉移視窗轉換器層

462:第二轉移視窗轉換器層

463:第三轉移視窗轉換器層

48:轉移視窗轉換單元

481:第一層標準化層

482:視窗多頭自注意力層

483:第二層標準化層

484:第一多層感知層

485:第三層標準化層

486:轉移視窗多頭自注意力層

487:第四層標準化層

488:第二多層感知層

50:卷積層

52:卷積層

54:應用程式介面

541:選項區

542:主畫面顯示區

543:功能執行選項

544:類別資訊區

544a:人聲選項

544b:貝斯選項

544c:鼓選項

56:圖形處理器

S10~S14:步驟

【新型申請專利範圍】

【請求項1】 一種數位音訊系統，包含：

一處理裝置，內建有一音源分離模型及一音訊轉錄模型，該處理裝置載入一輸入音訊檔，並透過該音源分離模型對該輸入音訊檔進行關鍵特徵擷取及訊號分離處理，以獲得複數分離音檔，該音訊轉錄模型會分別載入該些分離音檔，以透過該音訊轉錄模型分別取得對應之複數轉換檔案，該處理裝置將該些轉換檔案分別轉換為複數樂譜；以及

一資料庫，電性連接該處理裝置，用以儲存該輸入音訊檔及該些樂譜。

【請求項2】 如請求項1所述之數位音訊系統，其中該音源分離模型更包含一共享編碼器以及一多頭解碼器，該共享編碼器係權重提取該關鍵特徵後，再經由該多頭解碼器分別取得複數分離音源時頻圖，進而將該些分離音源時頻圖轉換為該些分離音檔。

【請求項3】 如請求項2所述之數位音訊系統，其中該處理裝置先利用一短時距傅立葉轉換將該輸入音訊檔轉換至一原始時頻圖，再將該原始時頻圖載入該共享編碼器中進行該關鍵特徵的提取；以及該處理裝置利用一反短時距傅立葉轉換將該些分離音源時頻圖轉換為該些分離音檔。

【請求項4】 如請求項1所述之數位音訊系統，其中該輸入音訊檔及該些分離音檔係具有相同檔案格式。

【請求項5】 如請求項1所述之數位音訊系統，其中該音訊轉錄模型係先將一該分離音檔轉換成一時頻圖，並利用一卷積神經網路層對該時

頻圖提取特徵資訊，以獲得複數語音後驗特徵圖，再對該些語音後驗特徵圖進行一後處理，以產生該些轉換檔案。

【請求項6】 如請求項5所述之數位音訊系統，其中該些語音後驗特徵圖係包含一起始點後驗特徵圖、一音符後驗特徵圖以及一音高後驗特徵圖。

【請求項7】 如請求項1所述之數位音訊系統，其中該音訊轉錄模型係先將一該分離音檔轉換成一時頻圖，並利用一轉移視窗轉換器層對該時頻圖提取全域資訊特徵，以獲得複數語音後驗特徵圖，再對該些語音後驗特徵圖進行一後處理，以產生該些轉換檔案。

【請求項8】 如請求項7所述之數位音訊系統，其中該些語音後驗特徵圖係包含一起始點後驗特徵圖、一音符後驗特徵圖以及一音高後驗特徵圖。

【請求項9】 如請求項1所述之數位音訊系統，其中該些轉換檔案係為一樂器數位介面（MIDI）檔案。

【請求項10】 如請求項1所述之數位音訊系統，其中該處理裝置更可透過一數學佈局演算法來整理音符，以簡化該些樂譜。

【請求項11】 如請求項1所述之數位音訊系統，其中該些樂譜係以圖像檔案格式儲存在該資料庫中。

【請求項12】 如請求項1所述之數位音訊系統，更包含一顯示裝置，電性連接該處理裝置，以顯示該些樂譜。

【請求項13】 如請求項12所述之數位音訊系統，其中該處理裝置更可提供一應用程式介面顯示於該顯示裝置上，以透過該應用程式介面選擇待轉譜之樂器。

【請求項14】 如請求項1所述之數位音訊系統，更包含一圖形處理器，該圖形處理器電性連接該處理裝置，使該處理裝置可搭配該圖形處理器來分別訓練該音源分離模型及該音訊轉錄模型。

【請求項15】 如請求項1所述之數位音訊系統，其中處理裝置更可利用一邊緣計算技術處理該音源分離模型及該音訊轉錄模型，使該音源分離模型及該音訊轉錄模型之權重檔轉換為該邊緣計算技術所接受的檔案格式。

【新型圖式】

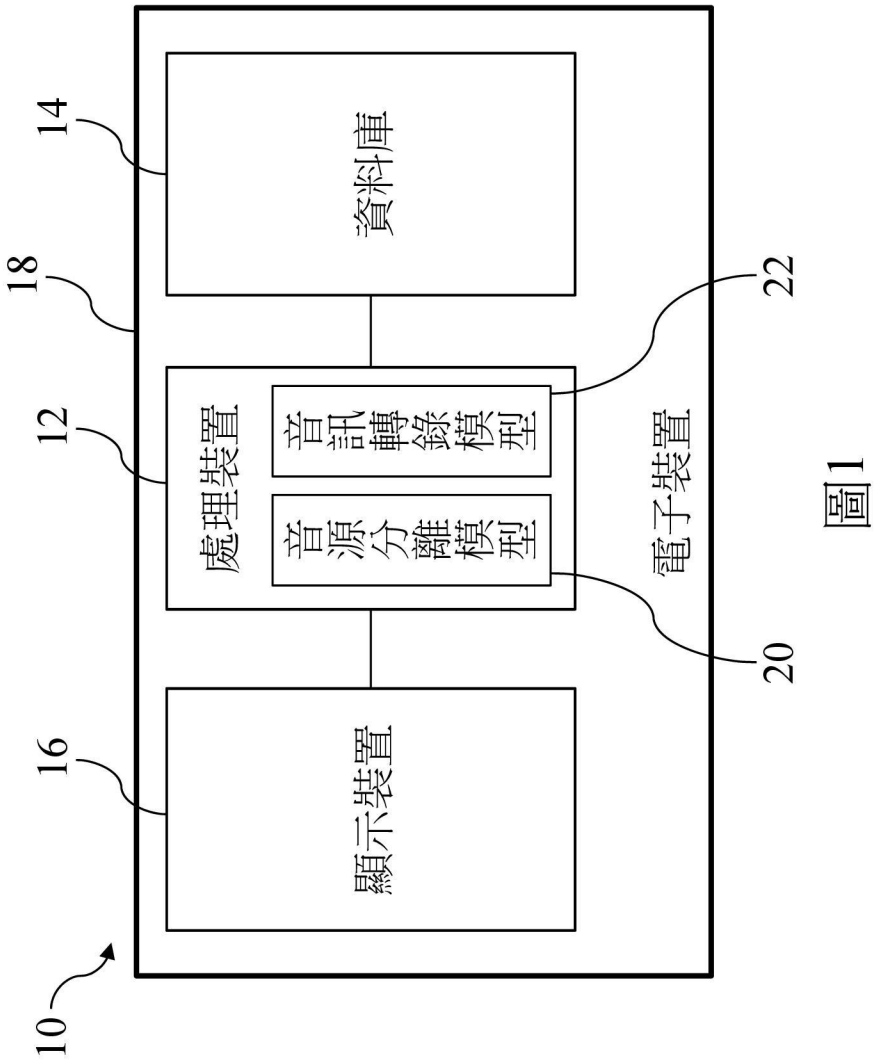


圖1

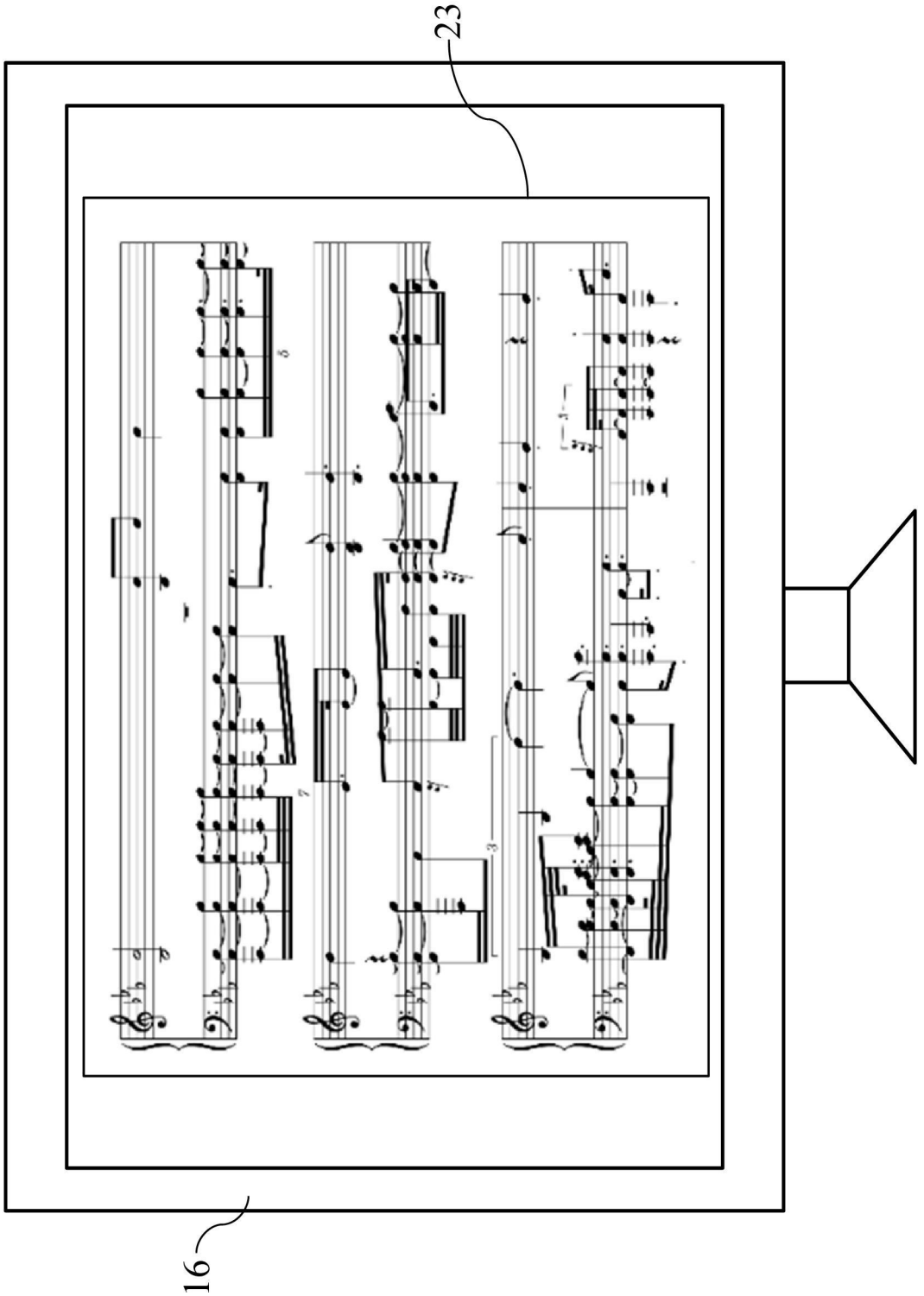


圖2

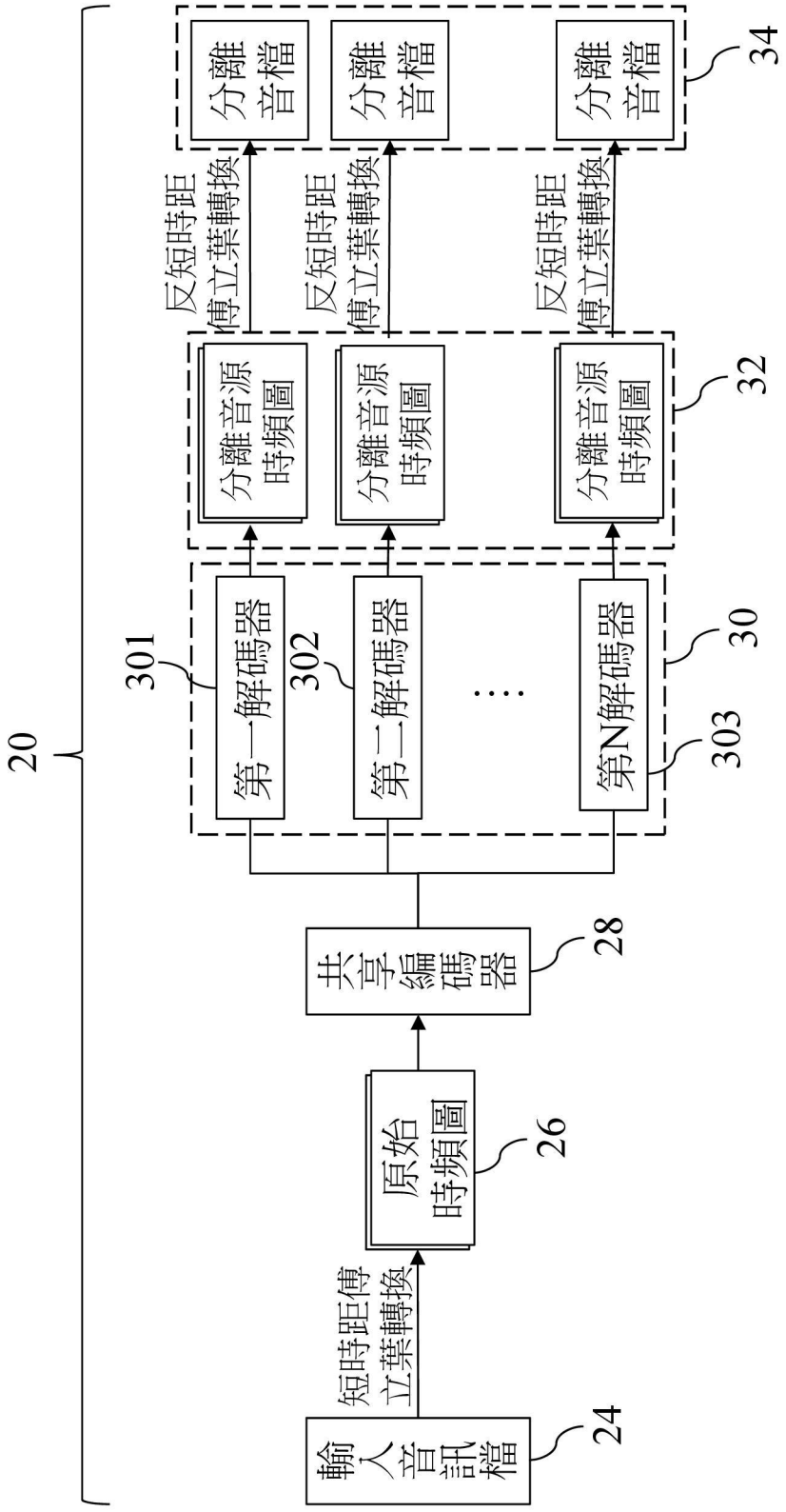

20
輸入音訊檔
24
短時距傅立葉轉換
原始時頻圖
26
共享編碼器
28
第一解碼器
301
第二解碼器
302
第N解碼器
303
30
分離音源時頻圖
32
反短時距傅立葉轉換
分離音檔
34

圖3

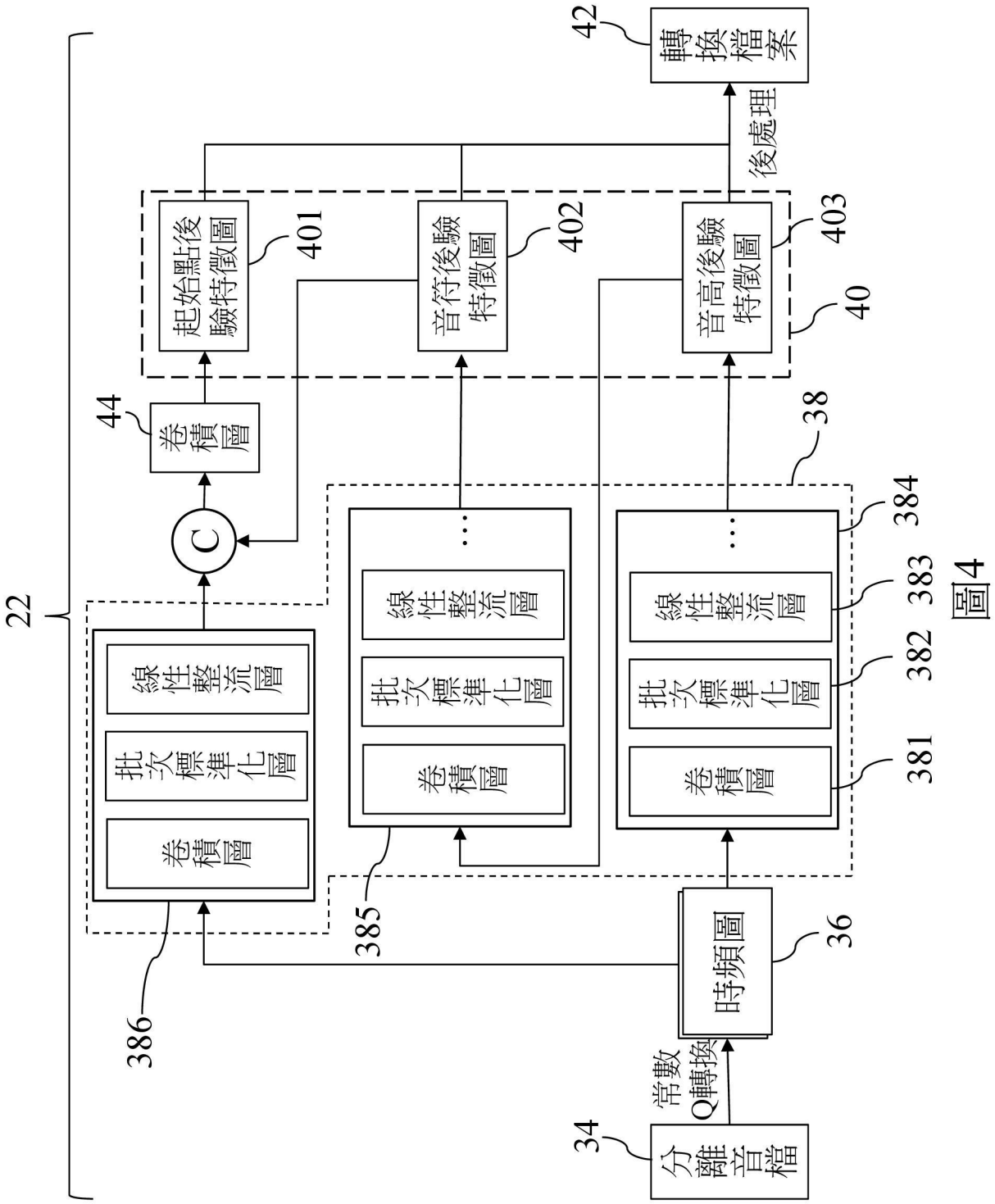

22
42
轉換檔案
後處理
起始點後驗特徵圖
401
音符後驗特徵圖
402
音高後驗特徵圖
403
40
44
卷積層
C
38
384
383
382
381
線性整流層
批次標準化層
卷積層
385
386
36
時頻圖
常數Q轉換
34
分離音檔

圖4

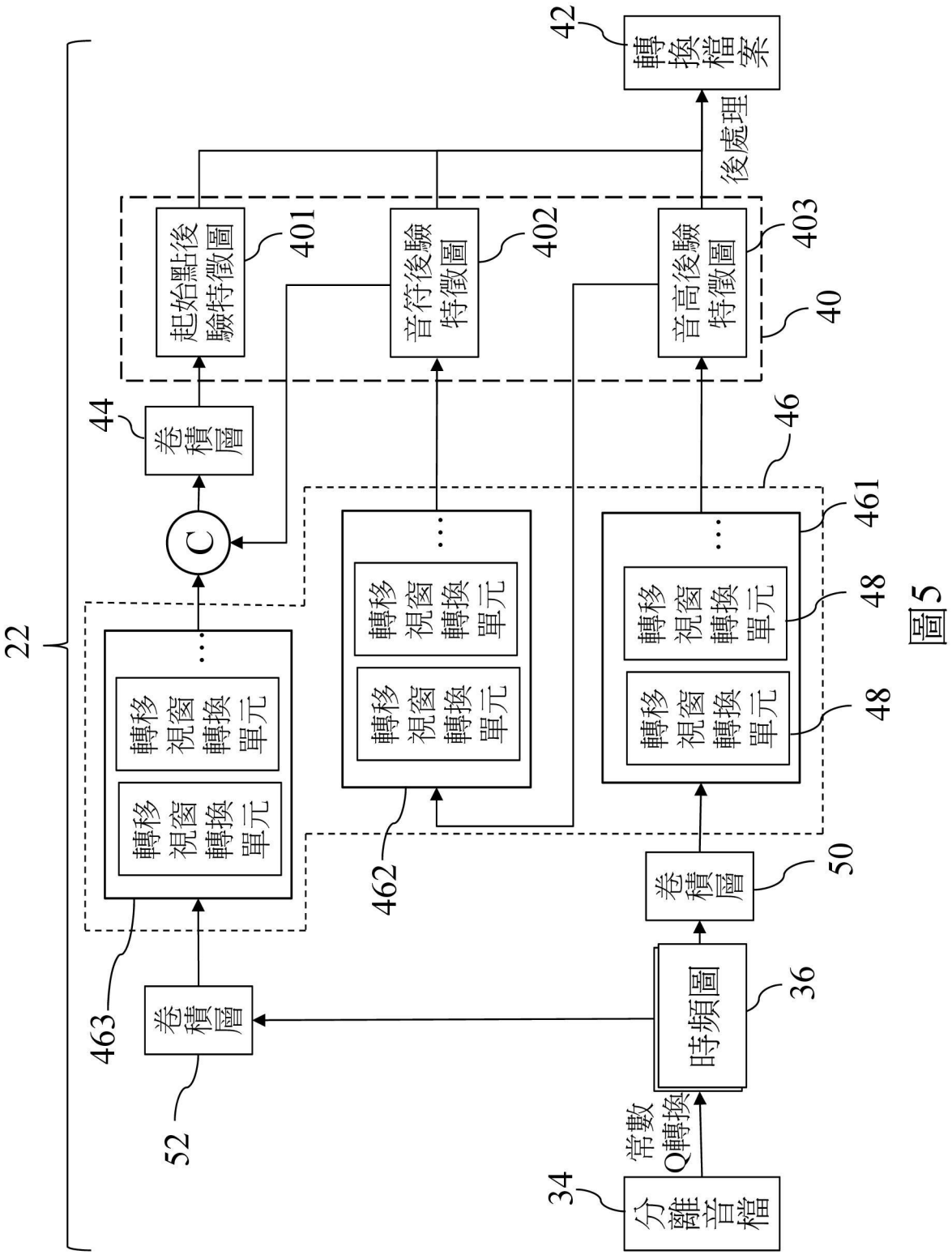

22
42
轉換檔案
後處理
起始點後驗特徵圖
401
音符後驗特徵圖
402
音高後驗特徵圖
403
40
44
卷積層
C
46
461
轉移視窗轉換單元
48
462
463
52
卷積層
時頻圖
36
50
卷積層
常數Q轉換
34
分離音檔

圖5

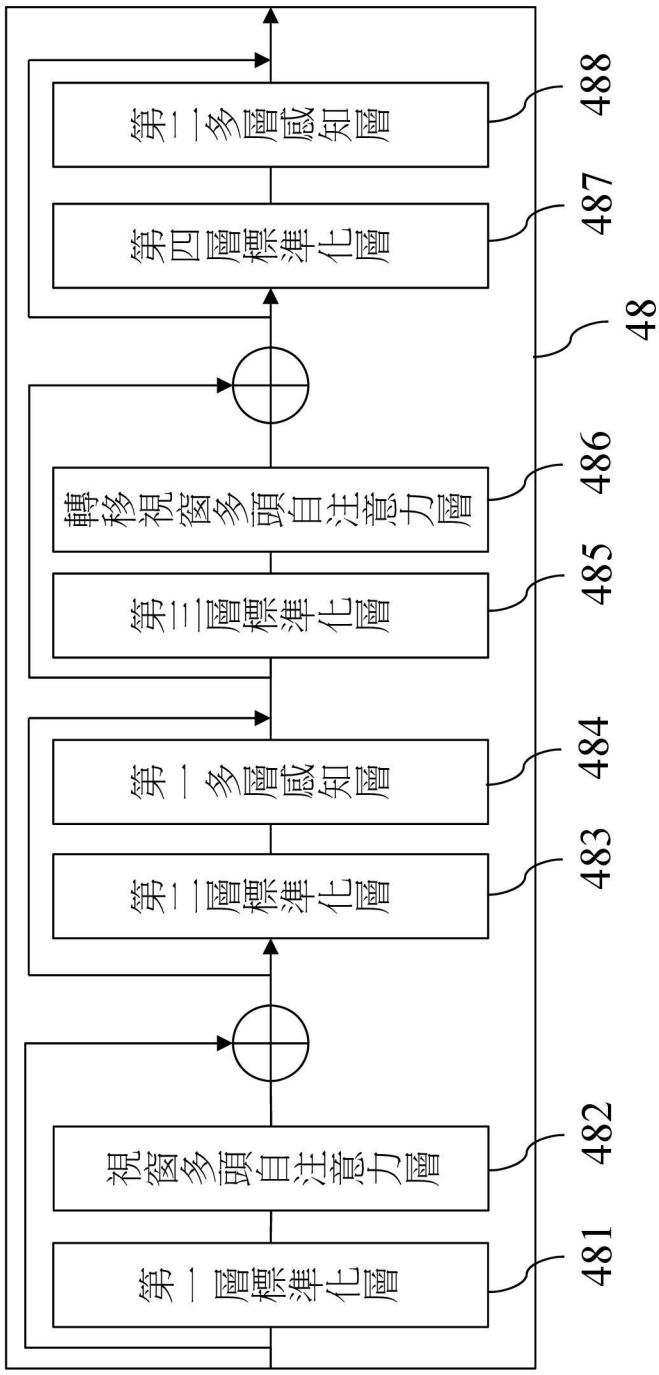

第一層標準化層
481
視窗多頭自注意力層
482
第二層標準化層
483
第一多層感知層
484
第三層標準化層
485
轉移視窗多頭自注意力層
486
第四層標準化層
487
第二多層感知層
488
48

圖6

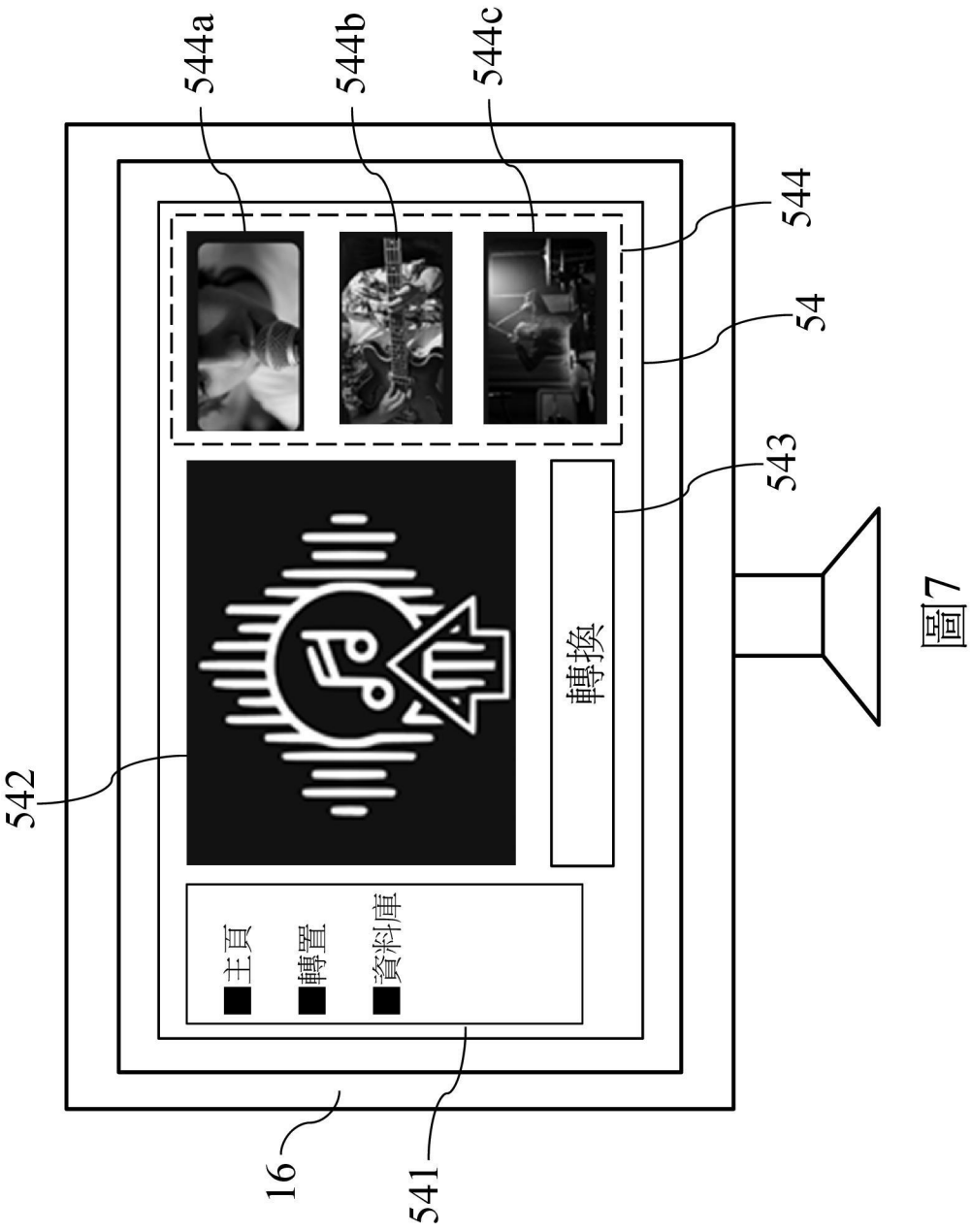


圖7

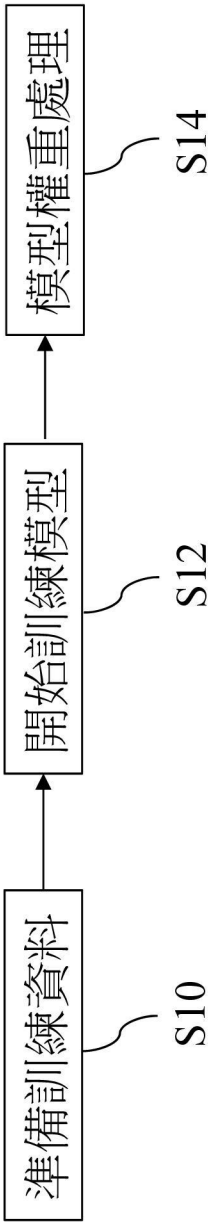


圖8

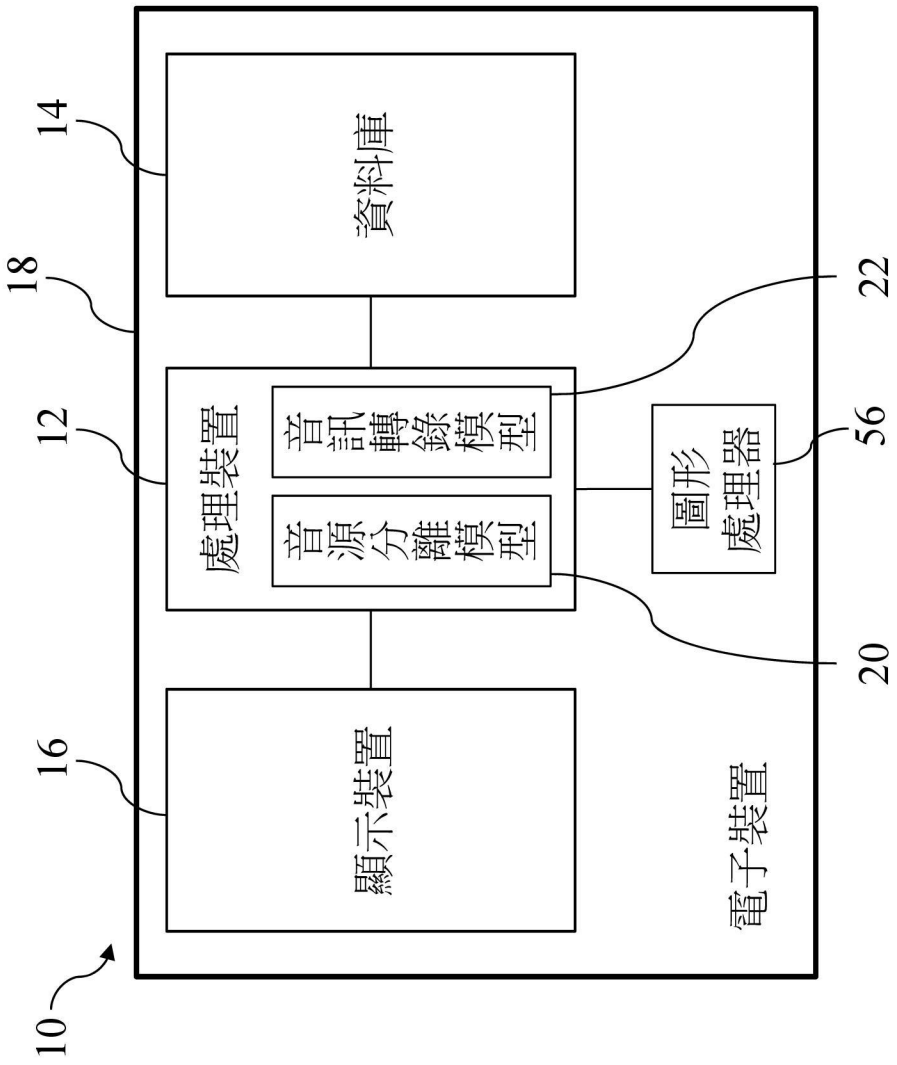


圖9