201820315

(19)中華民國智慧財產局

(12)發明說明書公開本 (11)公開編號：TW 201820315 A

(43)公開日： 中華民國 107 (2018) 年 06 月 01 日

(21)申請案號：106140244 (22)申請日： 中華民國 106 (2017) 年 11 月 21 日

(51)Int. Cl.： *G10L25/60 (2013.01)* *G10L17/22 (2013.01)*
G06F9/44 (2006.01) *H04R5/033 (2006.01)*

(30)優先權：2016/11/21 法國 1661324

(71)申請人：法國國立高等礦業電信學校聯盟 (法國) INSTITUT MINES TELECOM (FR)
法國
布盧埃 拉斐爾 (法國) BLOUET, RAPHAEL (FR)
法國

(72)發明人：埃西德 西林姆 ESSID, SLIM (FR)；布盧埃 拉斐爾 BLOUET, RAPHAEL (FR)

(74)代理人：許世正

申請實體審查：無 申請專利範圍項數：11 項 圖式數：4 共 26 頁

(54)名稱

改良型音訊耳機裝置及其聲音播放方法、電腦程式

IMPROVED AUDIO HEADSET DEVICE

(57)摘要

本發明涉及在頭戴式或耳塞式聲音播放裝置上播放聲音的資料處理。聲音播放裝置可由使用者在一環境下攜帶，且包含至少一喇叭、至少一麥克風和一處理電路的一接頭，處理電路包含用以接收來自至少麥克風之訊號的輸入介面、用以讀取要在喇叭播放的至少一聲音內容的處理單元以及用以至少配送喇叭要播放的聲音訊號的輸出介面。處理單元用以分析來自該麥克風的該些訊號，以識別來自環境、對應多個預設類別的目標聲音的聲音；根據用戶偏好準則，選擇至少一識別出的聲音；以及藉由聲音內容與選擇的聲音的選擇混合，建立喇叭要播放的聲音訊號。

The invention relates to data processing for sound playing on a sound playing device (DIS), headset or ear bud type, portable by a user in an environment (ENV). The device comprises at least one speaker (HP), at least one microphone (MIC), and a connection to a processing circuit comprising: an input interface (IN) for receiving signals coming at least from the microphone; a processing unit (PROC, MEM) for reading at least one audio content to play on the speaker; and an output interface (OUT) for delivering at least the audio signals to be played by the speaker. The processing unit is arranged for: a) Analyzing the signals coming from the microphone for identifying sounds coming from the environment and corresponding to predetermined classes of target sounds; b) Selecting at least one identified sound, according to a user preference criterion; and c) Building said audio signals to be played by the speaker, by a selected mixing of the audio content and the selected sound.

指定代表圖：

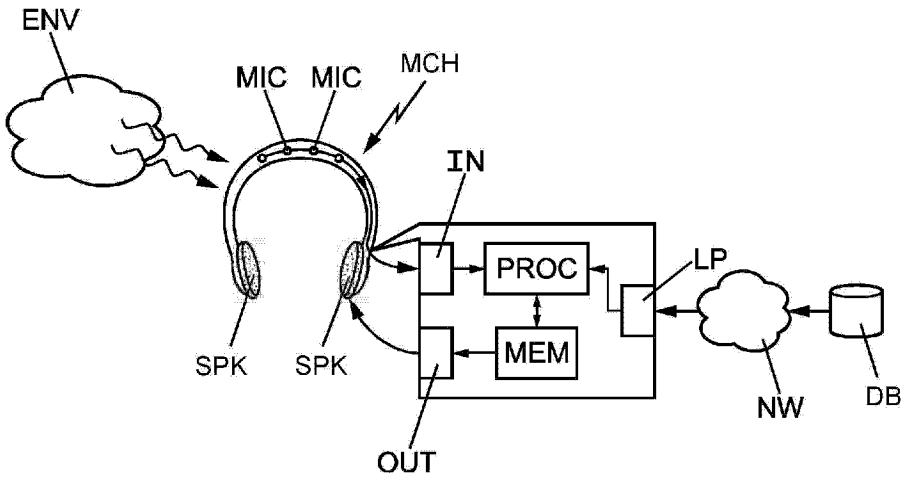

圖 1

符號簡單說明：

BD・・・資料庫

DIS・・・聲音播放裝置

ENV・・・環境

IN・・・輸入介面

HP・・・喇叭

LP・・・通訊模組

MEM・・・記憶體

MIC・・・麥克風

NW・・・網路

OUT・・・輸出介面

PROC・・・處理器

【發明說明書】

【中文發明名稱】　改良型音訊耳機裝置及其聲音播放方法、電腦程式

【英文發明名稱】　IMPROVED AUDIO HEADSET DEVICE

【技術領域】

【0001】　本發明係關於一種可攜式聲音收聽裝置，其可涉及一種具有左右聽筒或左右可攜式耳塞的音頻耳機。

【先前技術】

【0002】　已知的抗噪音頻耳機是基於利用麥克風陣列來擷取使用者的聲音環境。一般來說，這些裝置試圖即時地建立理想的過濾器，藉由過濾器以最大幅度地減少使用者感知到的聲音訊號中聲音環境所帶來的貢獻。近來已有人提出一種環境噪音過濾器，環境噪音過濾器可為使用者自己描述之環境類型的函數，然後使用者可選擇不同的雜訊消除模式（例如，辦公室、外面等）。在此情況下，所述“外面”模式可提供回注環境訊號（但比沒有使用過濾器的水平低得多，並且以允許使用者保持在能知曉其環境的方式進行環境訊號的回注）。

【0003】　並且，選擇性的聲音頭戴式耳機和耳塞式耳機是眾所周知，可以個人化地聆聽該環境。最近出現的這些產品可以在兩個軸向上改變對環境的感知：

增加感知（言語的可理解性）；以及

使聽覺系統免於受環境噪音的影響。

【0004】　可涉及可被智慧型手機應用程式配置的音頻耳機。由於說話聲通常是位於使用者前方，因此可在吵雜的環境中放大說話聲。

【0005】　也可涉及連接智慧型手機的音頻耳機，智慧型手機讓使用者可以配置對聲音環境的感知：調整音量、增加等化器或音效。

【0006】　因此，舉互動式的頭戴式耳機和耳塞式耳機為例，為了添加真實感，頭戴式耳機和耳塞式耳機可使聲音環境（遊戲、歷史重組）變豐

富，或者可引導使用者進行活動（虛擬教練）。

【0007】 至終，一些助聽器用來改善有聽力障礙之使用者的體驗的方法提供創新的方向，例如改善空間選擇性（例如，跟隨使用者眼睛的方向）。

【0008】 然而，這些現有的不同實施方式無法進行下列功能：

分析與解讀使用者的活動、使用者消費的內容以及使用者所沉浸的環境（尤其是音景（soundscape））；

根據這些分析結果，自動修改聲音渲染。

【0009】 一般，頭戴式抗噪耳機僅僅是基於以多聲道的方式擷取使用者的環境。不管環境的性質如何，頭戴式抗噪耳機試圖全面地減少對於使用者感知到之訊號的貢獻；即使環境包含潛在令人感興趣的資訊，頭戴式抗噪耳機仍這麼做。因此，這些裝置傾向於將使用者隔絕於其環境。

【0010】 頭戴式音頻耳機的選擇性原型讓使用者可以藉由利用等化過濾器或增加語音清晰度的方式配置其聲音環境。利用這些裝置，可改善使用者對環境的感知能力，但這些裝置並沒有真的根據使用者的狀態或出現在環境中的聲音類別來修正所產生的內容。在此配置下，正在以大聲量聽音樂的使用者總是與其環境隔絕，以及在此配置下，一直需要有一種能讓使用者從其環境中獲取有關資訊的裝置。

【0011】 當然，互動式的頭戴式耳機和耳塞式耳機可配備有感測器，用來載入、產生與位置（例如與觀光相關）或活動（遊戲、運動訓練）相關的內容。當某些裝置甚至有用來監控使用者活動的慣性或生理感測器，然後可以根據對來自感測器的訊號進行分析的結果來產生某些內容時，所產生的內容並不是從包含對使用者周遭音景進行分析的一自動產生的過程中產生，並且此內容也不允許自動從環境中選擇與使用者有關的成分。而且，運作模式是固定的，且不會自動隨著聲音環境的時間改變而改變，更不用說隨其他變數參數，例如使用者的生理狀態，而改變。

【0012】 本發明試圖改善這樣的情況。

【發明內容】

【0013】 為這目的，提出一種藉由計算機資料處理手段實現的方法，用以在頭戴式或耳塞式的一聲音播放裝置上播放聲音，該聲音播放裝置可由一環境中的使用者攜帶，且包含：

至少一喇叭；

至少一麥克風；

一處理電路的一接頭；

該處理電路，包含：

一輸入介面，用以接收至少來自該麥克風的訊號；

一處理單元，用以讀取要在該喇叭上播放的至少一聲音內容；以及

一輸出介面，用以至少配送該喇叭要播放的聲音訊號。

尤其是，該處理單元更配置來實現下列步驟：

a）分析來自麥克風的訊號，以識別來自環境且對應多個預設類別的目標聲音的聲音；

b）根據用戶偏好準則，選擇至少一識別出的聲音；以及

c）藉由選擇性混合該聲音內容與選擇的該聲音，來建立要被該喇叭播放的該些聲音訊號。

【0014】 在一可能的實施例中，該裝置包含多個麥克風，並且分析來自麥克風的訊號更包含處理來自麥克風的訊號，以從環境中區分出聲音來源。

【0015】 例如，在步驟 c）中被選擇的聲音可：

至少在頻率和持續時間上被分析；

在處理訊號、區分來源後，藉由過濾的方式被改善，並與該聲音內容混合。

【0016】 在裝置包含至少兩個喇叭且應用 3D 音效在喇叭上播放訊號

的一實施例中，可考量環境中偵測到釋出所選擇之聲音的聲源位置，進而以混合的方式對來源施加一聲音空間化效果。

【0017】 在一實施例中，該裝置可更包含一人機介面的一接頭，此人機介面讓使用者可以輸入偏好來選擇來自該環境的聲音（廣義來說，稍後會看到），然後再藉由從該使用者輸入且儲存在記憶體中的偏好歷程學習，以決定該用戶偏好準則。

【0018】 在一（另一或附加）實施例中，該裝置可更包含一用戶偏好資料庫的一接頭，該用戶偏好準接著藉由分析所述資料庫的內容來設定。

【0019】 該裝置可更包含針對該裝置使用者的一或多個狀態感測器的接頭，因此用戶偏好準則考量到使用者的目前狀態，然後以廣泛的含義來定義使用者的“環境”。

【0020】 在這樣的環境中，該裝置可包含該裝置的該使用者可使用的一行動終端的一接頭，此終端更有利的是包含針對該使用者狀態的一或多個感測器。

【0021】 處理單元可更設置來根據感測到的使用者狀態，從多個內容中選擇要讀取的內容。

【0022】 在一實施例中，預設目標聲音的類別可至少包含說話聲音，可為說話聲音預先錄製聲紋。

【0023】 而且，例如步驟 a）可選擇性地包含下列運作的至少其中之一：

建構並應用一動態過濾器，動態過濾器是用來消除來自麥克風之訊號中的噪音的；

對來自多個麥克風的訊號進行來源分離處理以及利用例如波束成形（beamforming）來識別（對於裝置的使用者而言）感興趣的來源，以對來自環境的音源進行區域化及離析；

選用這些感興趣的來源特有的的參數，為了後續用對由感興趣來源擷取到的聲音進行空間化混音的方式進行播放；

藉由已知聲音類別（語音、音樂、噪音等）的分類系統（例如，藉由深度神經網路），識別對應所述來源（在不同的空間方向上）的各種聲音類別；

以及透過其他用來分類音景（例如，辦公室、外面街道、大眾運輸等的聲音識別）的技術來進行可能的識別。

【0024】 再者，例如步驟 c）可選擇性地包含下列運作的至少其中之一：

時間、光譜和/或空間過濾（例如，Weiner 過濾和/或 Duet 演算法），以從多個麥克風擷取到的一或多個聲音串流中加強某一特定的聲音來源（根據前述來源分離模組取出的參數）；

3D 聲音渲染，例如使用頭部關聯傳遞函數（Head Related Transfer Function，HRTF）過濾技術。

【0025】 本發明也以電腦程式做為目標，此電腦程式包含多個指令，當此程式被一處理器執行時，這些指令會實現上述的方法。

【0026】 本發明也以頭戴式或耳塞式的聲音播放裝置作為目標，此聲音播放裝置可由一環境的使用者攜帶，並且包含：

至少一喇叭；

至少一麥克風；

一處理電路的一接頭；

該處理電路，包含：

一輸入介面，用以至少從該麥克風接收多個訊號；

一處理單元，用以讀取要在該喇叭上播放的至少一聲音內容；以及

一輸出介面，用以至少配送要被該喇叭播放的該些聲音訊號。

該處理單元更設置用來：

分析來自該麥克風的該些訊號，以識別來自該環境且對應多個預設類別的目標聲音的聲音；

根據一用戶偏好準則，選擇至少一識別出的聲音；以及

藉由選擇性混合該聲音內容與選擇的該聲音，來建立要被該喇叭播放的該些聲音訊號。

【0027】 因此，本發明提出一種包含智能音頻裝置的系統，此智能音頻裝置的系統例如包含一個由感測器構成的網路、至少一喇叭和一終端（例如，智慧型手機）。此系統的創舉是能夠即時自動地管理要給使用者的“最佳音軌”，“最佳音軌”意味著與使用者的環境和自身狀況最相襯的多媒體內容。

【0028】 使用者自身狀況可由下列定義：

i）偏好（音樂類型、感興趣的聲音類別等）的收集；

ii）使用者的活動（休息中、在辦公室、在運動訓練中等）；

iii）使用者的生理狀態（壓力、疲勞、努力等）和/或社會情感（人格、心情、情感等）。

【0029】 所產生的多媒體內容可包含一（將在耳機中產生的）主要聲音內容，也可能包含可經由智慧型手機型的終端播放的次要多媒體內容（文字、影像、視訊）。

【0030】 各種內容項目兼備有來自用戶內容庫的項目（儲存在終端或雲端的音樂、視訊等）、系統中的感測器網路擷取到的結果以及系統所產生的合成元素（通知、聲音或文字廣告短曲、舒適噪音等）。

【0031】 因此，所述系統可自動地分析使用者所在環境，以及預測使用者可能感興趣的成分，為了能藉由最佳地將可能感興趣的成分疊加在使

用者消費的內容（代表性地是指使用者聆聽的音樂）上，以增強和控制的方式來播放可能感興趣的成分。

【0032】 有效地播放內容要考量內容的性質以及從環境中選用之成分（在更精密的實施例中會一併考量使用者的自身狀況）。耳機中所產生的聲音串流不再來自兩個同時發生的來源：

一主要來源（音樂或無線電廣播或其他），以及

一擾亂的來源（周遭噪音），

但來自對資訊串流的採集，這些資訊串流的相對貢獻是根據其相關性來調整。

【0033】 因此，在播放火車站內的廣播訊息的同時，也會降低與使用者不相干的周遭噪音，使得即使使用者正在聽高音量的音樂，也可以清楚了解到此廣播訊息。增加一智慧型處理模組，尤其是搭載了來源分離（source separation）的演算法以及音景分類的演算法的智慧型處理模組，使得上述目的變成可能。直接應用的優點是：若偵測到某一類別的目標聲音時，將使用者重新連於其環境並提醒使用者；以及由於一推薦引擎會掌管前述的各種內容項目，因此可隨時自動地產生符合使用者期望的內容。

【0034】 適當的做法是，回想一下目前最新的技術的設備不允許自動識別出現在使用者之環境的每一聲音類別，以根據其在環境中的識別結果，將每一聲音類別聯於符合使用者期望的處理程序（例如，提升或降低聲量、產生警訊）。目前最新的技術並未採用音景分析技術，也未採用使用者的狀態或活動，來計算聲音渲染（rendering）。

【圖式簡單說明】

【0035】 在閱讀以下示範性實施例的詳細說明及檢視附圖後將可知本發明的其他優點和特徵，其中附圖包含：

【0036】 圖 1 表明根據本發明第一實施例的裝置；

【0037】 圖 2 表明根據本發明第二實施例的裝置，此裝置連接一行動

終端；

【0038】 圖 3 表明根據本發明一實施例的方法步驟；以及

【0039】 圖 4 根據一特定實施例具體說明圖 3 的方法步驟。

【實施方式】

【0040】 請參考圖 1，一種例如由環境 ENV 中的使用者所穿戴的（頭戴式或耳塞式）聲音播放裝置 DIS，此裝置至少包含：

一個（或範例中所示的兩個）喇叭 HP；

至少一感測器，例如麥克風 MIC（或範例所示的一麥克風陣列，用來捕捉來自環境中之聲音的方向）；以及

一處理電路的一接頭。

【0041】 處理電路可直接結合到耳機中而被喇叭外殼覆蓋（如圖 1 所示），或者處理電路可以圖 2 所示的不同方式實施於一用戶端 TER，例如智慧型手機型的行動終端，或者甚至將處理電路分散在多個用戶端（智慧型手機和連接的物件，此物件可包含其他感測器）。在此變化態樣中，是藉由 USB 或近程射頻（例如藍芽或其他）連接來實現耳機（或耳塞）與終端內的專用處理電路間的連線，並且耳機（或耳塞）相當於可與包含在終端 TER 內的 BT2 收發器進行通訊的 BT1 收發器。也可能採用一種將處理電路分散在耳機外殼與一終端的混合方案。

【0042】 在上述實施例中的一或另一中，處理電路包含：

一輸入介面 IN，用以接收來自至少該麥克風 MIC 的多個訊號；

一處理單元，典型地包含一處理器 PROC 和記憶體 MEM，用以對應環境 ENV 來解讀來自麥克風的訊號，進而學習（例如，藉由分類或甚至例如藉由指紋型匹配）；

一輸出介面 OUT，用以至少配送取決於該環境、要被該喇叭播放的聲音訊號。

【0043】 記憶體 MEM 可儲存本發明含義中的一電腦程式的多個指

令，且記憶體 MEM 除了可儲存長期數據以外，也可儲存臨時數據（計算結果或其他），長期數據例如後續將看到的用戶偏好或一致的模板定義資料或其他資料。

【0044】 在一精密的實施例中，輸入介面 IN 是連接一麥克風陣列，也連接一慣性感測器（裝備在耳機上或終端內）。

【0045】 用戶偏好資料可就地儲存在記憶體 MEM 中，如上所述。在另一做法中，資料可能與其他資料一起儲存於一遠端資料庫 DB 中，可藉由透過區域或廣域網路 NW 的通訊方式來存取。為此，與所述網路相配的一通訊模組 LP 可裝備於耳機或終端 TER 中。

【0046】 有利的是，人機介面可允許使用者定義、實施其偏好。在圖 2 裝置 DIS 與終端 TER 配對的實施例中，人機介面可輕易地對應例如智慧型手機 TER 的觸控螢幕。或者，此介面可直接裝備於耳機上。

【0047】 在圖 2 的實施例中，有利的是，利用終端 TER 中有附加感測器的優勢在一般認知上充實使用者環境的定義同樣是可行的。這些附加的感測器可以是使用者（腦波圖量測、心律量測、計步器等）專用的生理感測器，或者是其他用來改善對環境與當前使用者狀態成對的認知的感測器。此外，此定義可包含使用者直接回報其活動、其自身狀況和其環境。

【0048】 環境的定義可更進一步考量：

收集可存取內容以及收集諮詢內容（音樂、視訊、無線電廣播等）的歷程；

也可與關聯於該使用者的音樂庫的元資料（例如類型、分段收聽事件）相關聯；

另外，他們的智慧型手機的導航和應用程式歷程；

他們的串流（經由服務供應者）或本地內容消費的歷程；

在連上社群網路期間，使用者的偏好與活動。

【0049】 因此，廣而言之，輸入介面可連於感測器的收集結果，且也

包含多個連接模組（尤其是 LP 介面），用以描繪出使用者的環境以及他們的習性和偏好（內容消費、串流媒體活動和/或社群網路的歷程）。

【0050】 請參考圖 3，來說明前述處理單元所執行的處理程序：監控該環境，也可能監控該使用者狀態，以描繪可在輸出多媒體串流中播放的有關資訊。在一實施例中，此監控動作是以自動選用重要參數的方式實現，以經由訊號處理和人工智慧模組，尤其是機器學習，產生輸出多媒體串流（圖 3 的步驟 S7 所表示）。圖示中標註為 P1、P2 等的參數一般可為在喇叭上進行播放所應考量的環境參數。舉例來說，如果環境中擷取到的聲音被識別出是要播放的語音訊號：

一第一參數集合可為理想的過濾器（Weiner 型過濾器）的係數，藉由過濾器可加強語音訊號，以提升其清晰度；

一第二參數是環境中擷取到且要播放之聲音的方向性，聲音播放例如是採用立體音的渲染技術（利用 HRTF 型轉換函數的播放技術）；

等。

【0051】 因此，將理解的是，這些參數 P1、P2 等廣言之也可解讀成環境與使用者自身狀況的描述符，提供給一程式來產生“理想音軌（soundtrack）”給該使用者。此音軌是藉由編排其內容、來自環境的項目以及合成的項目來獲得。

【0052】 在第一步驟 S1 期間，處理單元呼叫用來從裝置 DIS 乘載的麥克風或麥克風陣列 MIC 收集訊號的輸入介面。自然地，步驟 S2 或步驟 S3 之終端 TER 內的其他感測器（慣性或其他）（相連的心律圖、腦波圖的感測器等）可傳遞其訊號給處理單元。此外，藉由記憶體 MEM 和/或處理單元的資料庫 BD 將除了擷取到的訊號（較佳的是來自步驟 S5 的使用者和/或步驟 S6 的內容與社群網路連線的消費歷程）以外的資料傳送至處理單元。

【0053】 在步驟 S4，收集環境和使用者狀態（以下統稱為“環境”）

特有的所有資料和訊號，並以步驟 S7 計算機模組的實施方式來解讀所述的環境，以藉由人工智慧解碼該環境。為達到這個目的，此解碼模組可採用一學習庫，以在步驟 S9 取出要用來普遍地將環境模型化的有關參數 P1、P2、P3 等。學習庫可例如為遠端的，且在步驟 S8 中可經由網路 NW（和通訊介面 LP）被呼叫。

【0054】 如稍後參照圖 4 詳細描述，要播放的音景特別是由步驟 S10 的參數所產生，並在步驟 S11 以聲音訊號的形式傳送至喇叭 HP。此音景可伴隨圖形資訊，例如在步驟 S12 要顯示在終端螢幕 TER 上的元資料。

【0055】 因此，藉由下列進行一環境訊號分析：

環境的一標識，以評估用來表徵使用者的環境及自身狀況的預測模型（該等模型會與一推薦引擎一併使用，這將在稍後參照圖 4 會看到）；以及

一細微聲學分析（fine acoustic analysis），用來產生操控要播放的聲音內容所需之更精確的參數（例如，特定音源的分離/強化、音效、混合、空間化或其他）。

【0056】 環境的標識用來藉由自動學習來表徵環境/使用者自身狀況的配對。其主要涉及：

偵測在一些預先記錄的類別中是否有些類別的目標聲音出現在使用者的環境中，以及適當地判斷其來源方向。一開始，使用者可藉由其終端或藉由預先定義的操作模式一個接一個地定義目標聲音的類別；

判斷使用者的活動：休息、待在辦公室、在健身房活動或其他；

判斷使用者的情緒狀態和生理狀態（例如，從計步器得知“處於良好的健康狀況”，或從使用者的腦波圖得知“感到有壓力”）；

藉由內容分析的技術手段（電腦聽覺和視覺技術以及自然語言處理）來描述使用者消費的內容。

【0057】 這些用來聲音播放（例如，3D 播放）的聲學參數可由細微聲學分析計算獲得。

【0058】 現在請參考圖 4，於步驟 S17，利用一推薦引擎從“環境”中接收描述符，尤其是所識別的聲音事件的類別（參數 P1、P2 等），並且於步驟 S19，推薦引擎在此基礎上提供一建議模型（或模型的組合）。為此，推薦引擎可利用用戶內容的特性描述以及用戶內容與外部內容間的相似處，也一併使用於步驟 S15 輸入進學習庫的使用者偏好和/或於步驟 S18 的其他使用者的標準偏好。在此步驟中，該使用者也可用其終端來操作，以在步驟 S24 輸入例如關於要播放的內容或內容清單的偏好。

【0059】 根據環境與使用者狀態，從收集的推薦中選擇一恰當的推薦模型（例如，在健身房裡的使用者做明顯運動之情境下的節奏音樂組中）。接著，於步驟 S20 中實現一編排引擎，編排引擎將參數 P1、P2 等合併道推薦模型中，以於步驟 S21 中調製一編排程式。此時，其涉及一慣常程序，此慣常程序例如建議：

在使用者的內容中尋找一特定型態的內容；

考量使用者的自身狀況（例如，其活動）以及來自環境中被參數 P1、P2 等識別出的某些類型的聲音；

根據一音量與編排引擎所定義的空間渲染（3D 聲音）來混合內容。

【0060】 嚴格來說，在步驟 S22 會牽涉到用於聲音訊號的合成引擎，合成引擎用於根據下列事項調製要在步驟 S11 和 S12 播放的訊號：

用戶內容（來自步驟 S25（作為步驟 S6 的子步驟），當然，在步驟 S21 中編排引擎會選擇一內容項目）；

在環境中擷取到的聲音訊號（S1，在合成來自要播放的環境的聲音的情況下可能為參數 P1、P2 等）；以及

用於通知（砰聲、鈴聲或其他）的其他聲音，該等聲音可能會被合成，通知可通報外部事件且可與要播放的內容（在步驟 S21 中選自步驟 S16）混合。

調製要在步驟 S11 和 S12 播放的訊號也可能根據步驟 S23 中所定義的 3D 渲染。

【0061】 因此，在一特定的實施例中，會根據三個主要步驟，使產生的串流與使用者的預期相稱並使串流根據串流產生的情境（context）被優化：

利用一推薦引擎即時過濾、選擇為了對多媒體串流（稱“受控的本體（reality）”）進行聲音播放（也可能視覺播放）而要混合的內容項目；

利用一媒體編排引擎規劃內容項目的時間、頻率和空間編排，也定義各別的音量；

利用一合成引擎根據編排引擎所建立的程序產生用來聲音渲染（也可能用來視覺渲染）的訊號，也可能產生用來聲音空間化的訊號。

【0062】 所產生的多媒體串流至少包含聲音訊號，但可能包含文字、觸覺或視覺提醒。聲音訊號包含下列的混合：

從使用者的內容庫選出的內容（音樂、視訊等）；

也可能包含選出的內容與下列的混合：

經由感測器陣列 MIC 擷取、從聲音環境（therefore filtered）中挑選出、被提升（例如，經由來源分離技術）且經處理過的聲音，其中此聲音具有可調整為適於引入所述混合中的頻率紋理、強度和空間定位；以及

在步驟 S16 中從一聲音資料庫檢索到的合成項目（例如，聲音或文字通知/廣告音樂（jingle）、舒適噪音（舒適噪音）等），

其中所選出的內容是步驟 S24 中由使用者按照其偏好所輸入，或者是由推薦引擎依據使用者的狀態及所在環境直接推薦。

【0063】 推薦引擎連帶地基於：

使用者的偏好，而使用者的偏好是明確地經由調查的做法來獲得，或者是間接地藉由利用對使用者自身狀況進行解碼的結果來獲得；

協同過濾和社交圖譜的技術，其一次採用多個使用者的模型（步驟 S18）；

來自使用者之內容的描述及這些內容的相似點的描述，以建立用來決定應該對使用者播放哪個內容項目的模型。

【0064】 隨著時間改變，會持續更新模型，以適應使用者的變化。

【0065】 編排引擎會規劃：

每個內容項目應該要播放的時間，尤其是要呈現使用者的內容的順序（例如，被挑選出的音樂在播放清單中的順序）以及播放外界聲音或通知的時間：即時或延遲（例如，在播放清單中兩個被選出者之間），而不會在不恰當的時間打擾到正在聆聽或活動的使用者；

每個內容項目的空間位置（用於 3D 渲染）；

必須應用在每個內容項目的各種音效（增益、過濾、等化、動態壓縮、迴音或迴響、時間減速/加速、移調等）。

【0066】 所述的規劃是基於從解碼使用者所在環境以及使用者自身狀況所構建成的模型與規則。例如，麥克風擷取到的聲音事件的空間位置以及關聯於聲音事件的增益程度是取決於圖 3 中步驟 S7 進行環境解碼的音源定位偵測結果。

【0067】 合成引擎分別仰賴訊號處理技術、自然語言和影像來進行聲音、文字和視覺資料（影像或視訊）輸出的合成，並且連帶地產生多媒體輸出，例如視訊。

【0068】 就合成聲音輸出來說，可採用時間、光譜和/或空間過濾技

術。舉例來說，首先在短時間視窗上進行局部性的合成，並且在利用加法復原法（addition-recovery）重新組織訊號之後，將訊號傳送給至少兩個喇叭（每個耳朵一個）。增益（功率）和各種音效應用於各種內容項目，例如編排引擎提供的增益和各種音效。

【0069】 在一特定的實施例中，利用視窗（window）的處理程序可包含過濾（例如，維納（Wiener）過濾），經由過濾從一或多個擷取到的聲音串流中加強一特定的聲音來源（例如編排引擎想要的）。

【0070】 在一特定的實施例中，處理程序可包含 3D 聲音渲染，3D聲音渲染可能採用HRTF過濾技術(HRTF轉換函數“頭部關聯轉換函數”)。

【0071】 在用來表明最小限度的實施方式的第一個範例中，

使用者所在環境的描述僅限於使用者的聲音環境；

使用者自身的狀況僅限於使用者的偏好：目標聲音的類別、使用者想接收的通知，而這些偏好是使用者利用其終端所定義；

裝置（可能搭配所述的終端）配備有慣性感測器（加速度計、陀螺儀和磁力計）；

當偵測出使用者所在環境中的目標聲音的類別時，會自動修改播放參數；

可記錄簡訊；

可傳送通知給使用者，以提醒使用者偵測到感興趣的事件。

【0072】 分析擷取到的訊號，以決定：

出現在使用者所在環境的聲音類別以及來自的方向，並且為了那目的：

藉由各別分析每一個方向的內容，來偵測最強聲音能量的方向；

整體判斷每個聲音類別的分布方向（例如，利用來源分離技術）；

描述使用者所在環境的模型參數以及提供給推薦引擎之參數。

【0073】 在用來說明更精密的實施方式的第二個範例中，包含一麥克風陣列、一視訊攝影機、計步器、慣性感測器（加速度計、陀螺儀、磁力計）以及生理感測器的一感測器組可擷取使用者的視覺環境和聲音環境(麥克風和相機）、表徵使用者運動的資料（慣性感測器、計步器）以及使用者的生理參數（腦波圖（EEG）、心電圖（ECG）、肌電圖（EMG）、膚電流（electrodermal）），也擷取使用者正在查閱的所有內容（音樂、無線電廣播、視訊、導航歷程以及使用者的智慧型手機應用程式）。接著，分析各種串流，以取出與使用者的活動、情緒、疲勞程度及環境（例如，在健身房用跑步機、好心情、低疲勞度）相關的資訊。可產生適合該環境及使用者自身狀況的音樂串流（例如，根據使用者的音樂品味、周遭環境和疲勞度所選擇的每一項目組成的播放清單）。然後，所有的聲音來源會在使用者的耳機中被刪除，並且當使用者附近的運動教練的聲音被識別出（之前預先記錄的聲紋）時，會將運動教練的聲音與串流混合並採用雙耳渲染（binaural rendering）技術進行空間播放（例如藉由頭部關聯傳遞函數）。

【符號說明】

【0074】

BD	資料庫
BT1、BT2	收發器
DIS	聲音播放裝置
ENV	環境
IN	輸入介面
HP	喇叭
LP	通訊模組
MEM	記憶體

MIC	麥克風
NW	網路
OUT	輸出介面
PROC	處理器
TER	用戶端

201820315
專利案號：106140244
申請日：
IPC 分類：

【發明摘要】

【中文發明名稱】 改良型音訊耳機裝置及其聲音播放方法、電腦程式

【英文發明名稱】 IMPROVED AUDIO HEADSET DEVICE

【中文】

本發明涉及在頭戴式或耳塞式聲音播放裝置上播放聲音的資料處理。聲音播放裝置可由使用者在一環境下攜帶，且包含至少一喇叭、至少一麥克風和一處理電路的一接頭，處理電路包含用以接收來自至少麥克風之訊號的輸入介面、用以讀取要在喇叭播放的至少一聲音內容的處理單元以及用以至少配送喇叭要播放的聲音訊號的輸出介面。處理單元用以分析來自該麥克風的該些訊號，以識別來自環境、對應多個預設類別的目標聲音的聲音；根據用戶偏好準則，選擇至少一識別出的聲音；以及藉由聲音內容與選擇的聲音的選擇混合，建立喇叭要播放的聲音訊號。

【英文】

The invention relates to data processing for sound playing on a sound playing device (DIS), headset or ear bud type, portable by a user in an environment (ENV). The device comprises at least one speaker (HP), at least one microphone (MIC), and a connection to a processing circuit comprising: an input interface (IN) for receiving signals coming at least from the microphone; a processing unit (PROC, MEM) for reading at least one audio content to play on the speaker; and an output interface (OUT) for delivering at least the audio signals to be played by the speaker. The processing unit is arranged for: a) Analyzing the signals coming from the microphone for identifying sounds coming from the environment and corresponding to predetermined classes of target sounds; b) Selecting at least one identified sound, according to a user preference criterion; and c)

Building said audio signals to be played by the speaker, by a selected mixing of the audio content and the selected sound.

【指定代表圖】　圖1。

【代表圖之符號簡單說明】

BD	資料庫
DIS	聲音播放裝置
ENV	環境
IN	輸入介面
HP	喇叭
LP	通訊模組
MEM	記憶體
MIC	麥克風
NW	網路
OUT	輸出介面
PROC	處理器

【特徵化學式】

無

【發明申請專利範圍】

【第1項】一種藉由計算機資料處理手段實現在一頭戴式或耳塞式的聲音播放裝置上播放聲音的方法，該聲音播放裝置可讓一使用者於一環境中攜帶，該裝置包含：

至少一喇叭；

至少一麥克風；

一處理電路的一接頭；

該處理電路，包含：

一輸入介面，用以至少從該麥克風接收多個訊號；

一處理單元，用以讀取要在該喇叭上播放的至少一聲音內容；以及

一輸出介面，用以至少配送要被該喇叭播放的該些聲音訊號，

其特徵在於該處理單元更設置來實現以下步驟：

a）分析來自該麥克風的該些訊號，以識別來自該環境且對應多個預設類別的目標聲音的聲音；

b）根據一用戶偏好準則，選擇至少一識別出的聲音；以及

c）藉由選擇性混合該聲音內容與選擇的該聲音，來建立要被該喇叭播放的該些聲音訊號，

其中該裝置包含多個麥克風，以及分析來自該些麥克風的該些訊號更包含處理來自該些麥克風的該些訊號，以從該環境中區分出聲音來源。

【第2項】 根據請求項 1 所述的方法，其特徵在於步驟 c）中所選擇的聲音：

至少在頻率和持續時間上被分析；以及

在處理訊號、區分來源後，藉由過濾的方式被改善，並與該聲音內容混合。

【第3項】 如請求項 1 所述的方法，其特徵在於該裝置包含至少兩個喇叭，該些喇叭應用 3D 音效、該環境中偵測到且釋出一選擇聲音的聲源位置來播放該些訊號，以對該混合中的該來源加入聲音空間化效果。

【第4項】 如請求項 1 所述的方法，其特徵在於該裝置包含一人機介面的一接頭，該人機介面允許一使用者輸入偏好，以選擇來自該環境的聲音，其中該用戶偏好準則是藉由從使用者輸入偏好的歷程中學習的方式來決定，並且該用戶偏好準則會被儲存於記憶體中。

【第5項】 如請求項 1 所述的方法，其特徵在於該裝置更包含一用戶偏好資料庫的一接頭，該用戶偏好準則是藉由分析該用戶偏好資料庫的內容來設定。

【第6項】 如請求項 1 所述的方法，其特徵在於該裝置更包含針對該裝置的一使用者的至少一狀態感測器的一接頭，其中該用戶偏好準則考量到該使用者的目前狀態。

【第7項】 如請求項 6 所述的方法，其特徵在於該裝置包含該裝置的該使用者可用的一行動終端的一接頭，該終端包含該使用者之狀態的一或多個感測器。

【第8項】 如請求項 6 所述的方法，其特徵在於該處理單元更設置來從多項內容中選擇要讀取的內容，該要讀取的內容取決於該使用者的該狀態。

【第9項】 如請求項 1 所述的方法，其特徵在於該些該些預設類別的目標聲音至少包含說話聲音、預先記錄的聲紋。

【第10項】 一電腦程式，其特徵在於該電腦程式包含多個指令，當該電腦程式被一處理器執行時，該些指令用來實現請求項 1 所述的方法的指令。

【第11項】 一種頭戴式或耳塞式的聲音播放裝置，該聲音播放裝置可由一環境中的一使用者攜帶，該裝置包含：

至少一喇叭；

至少一麥克風；

一處理電路的一接頭；

該處理電路包含：

一輸入介面，用以接收至少來自該麥克風的多筆訊號；

一處理單元，用以讀取要在該喇叭上播放的至少一聲音內容；以及

一輸出介面，用以配送該喇叭要播放的至少該些聲音訊號，

其特徵在於該處理單元更設置來：

分析來自該麥克風的該些訊號，以識別來自該環境且對應多個預設類別的目標聲音的聲音；

根據一用戶偏好準則，選擇至少一識別出的聲音；以及

選擇性混合該聲音內容與選擇的該聲音，以建立要讓該喇叭播放的該些聲音訊號，

其中該裝置包含多個麥克風，分析來自該些麥克風的該些訊號更包含處理來自該些麥克風的該些訊號，以從該環境中區分出聲音來源。