201828285

(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 201828285 A

(43)公開日： 中華民國 107 (2018) 年 08 月 01 日

(21)申請案號：106101958 (22)申請日： 中華民國 106 (2017) 年 01 月 19 日

(51)Int. Cl.： *G10L19/02 (2013.01)* *G10L15/08 (2006.01)*

(71)申請人：阿里巴巴集團服務有限公司 (香港地區) ALIBABA GROUP SERVICES LIMITED (HK)

香港

(72)發明人：杜志軍 (CN)

(74)代理人：林志剛

申請實體審查：有 申請專利範圍項數：18 項 圖式數：9 共 33 頁

(54)名稱

音頻識別方法和系統

(57)摘要

本申請實施例公開了一種音頻識別方法，包括：對待識別音頻檔案的語譜圖中的第一特徵點進行擴散處理，得到特徵點圖，所述第一特徵點的數量為多個；在目標音頻檔案的語譜圖中查找是否存在與所述特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點；若是，則確定所述待識別音頻檔案為所述目標音頻檔案的一部分。本申請還公佈了一種音頻識別系統實施例。利用本實施例可以在音頻識別中提高特徵點匹配成功率。

指定代表圖：

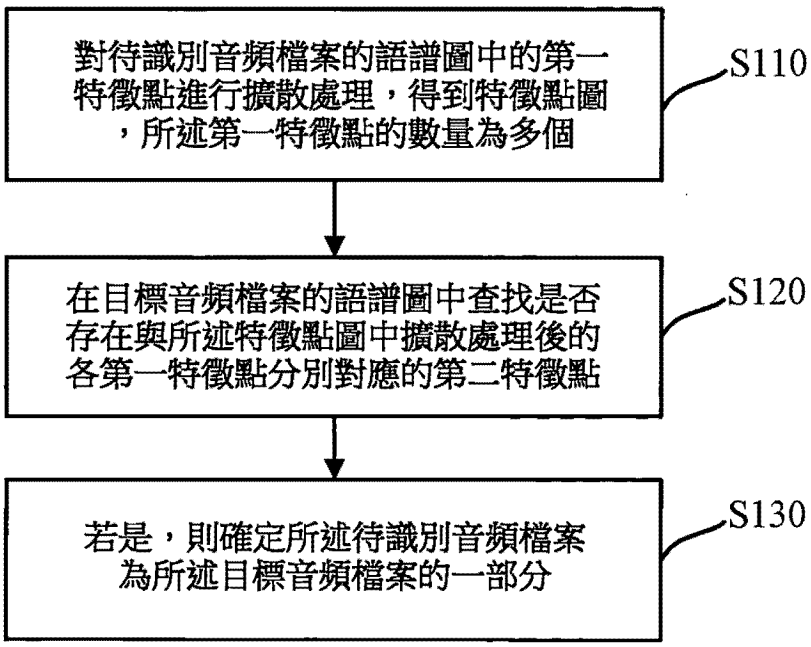

圖 2

發明專利說明書

（本說明書格式、順序，請勿任意更動）

【發明名稱】(中文/英文)

音頻識別方法和系統

【技術領域】

本申請關於互聯網技術領域，特別關於一種音頻識別方法及系統。

【先前技術】

隨著互聯網技術的不斷發展，互聯網已成為人們生活中必不可少的工具。利用互聯網設備實現未知音頻的識別，並基於音頻識別的互動，成為一種新的應用趨勢。

基於音頻識別的互動有多種應用，一種應用例如是：用戶聽到一首不知道歌名的歌曲，可以錄製該歌曲的一段音頻，然後利用音頻識別技術，可以識別出這首歌的歌名、歌手等資訊。

現有技術中，一般是提取待識別音頻的特徵點，利用特徵點對進行識別。如圖 1 所示，橫軸代表時間，縱軸代表頻率。提取的特徵點為圖中的“X”；兩個特徵點構成一個特徵點對，在目標區域內有 8 個特徵點對；採用特徵點對的方式在資料庫中進行識別，資料庫記憶體儲存有歌曲的特徵點及歌曲資訊如歌名、歌手等；如果在資料庫中能在相同的目標區域內匹配到一樣的特徵點對，則匹配成

功；進而可以得到對應的歌曲資訊。然而，由於錄製音頻時不可避免的受到噪音的影響，提取的特徵點不一定都在正常的位置出現，所以導致特徵點對匹配成功的機率較低。

綜上所述，現有技術中存在音頻識別中特徵點匹配成功率低的問題。

【發明內容】

本申請實施例的目的是提供一種音頻識別方法及系統，用以解決現有技術中音頻識別中特徵點匹配成功率低的問題。

為解決上述技術問題，本申請一實施例提供的音頻識別方法，包括：

對待識別音頻檔案的語譜圖中的第一特徵點進行擴散處理，得到特徵點圖，所述第一特徵點的數量為多個；

在目標音頻檔案的語譜圖中查找是否存在與所述特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點；

若是，則確定所述待識別音頻檔案為所述目標音頻檔案的一部分。

本申請一實施例提供的音頻識別系統，包括：

擴散單元，用於對待識別音頻檔案的語譜圖中的第一特徵點進行擴散處理，得到特徵點圖，所述第一特徵點的數量為多個；

查找單元，在目標音頻檔案的語譜圖中查找是否存在與所述特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點；

確定單元，用於在目標音頻檔案的語譜圖中查找到與所述特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點時，則確定所述待識別音頻檔案為所述目標音頻檔案的一部分。

由以上本申請實施例提供的技術方案可見，本申請實施例提供的一種音頻識別方法及系統，透過對待識別音頻檔案的語譜圖中的第一特徵點進行擴散處理，可以減少所述第一特徵點受噪音影響產生的偏差；從而提高擴散處理後的第一特徵點與目標音頻檔案的匹配率，即提高了特徵點匹配成功率。

【圖式簡單說明】

為了更清楚地說明本申請實施例或現有技術中的技術方案，下面將對實施例或現有技術描述中所需要使用的附圖作簡單地介紹，顯而易見地，下面描述中的附圖僅僅是本申請中記載的一些實施例，對於本領域普通技術人員來講，在不付出創造性勞動性的前提下，還可以根據這些附圖獲得其他的附圖。

圖 1 為現有技術中利用特徵點對進行識別的示意圖；

圖 2 為本申請一實施例中提供的音頻識別方法的流程圖；

圖 3 為待識別音頻檔案的語譜圖的示意圖；

圖 4a 為擴散處理前的第一特徵點的示意圖；

圖 4b 為擴散處理後的第一特徵點的示意圖；

圖 5 為圖 1 中 S120 步驟的方法流程圖；

圖 6 為在目標音頻檔案的語譜圖中查找與特徵點圖中擴散處理後第一特徵點分別對應的第二特徵點的示意圖；

圖 7 為本申請一實施例中提供的音頻識別方法的流程圖；

圖 8a 為在語譜圖中確定的第一特徵點的示意圖；

圖 8b 為圖 8a 的局部放大圖；

圖 9 為本申請一實施例中提供的音頻識別系統的模組示意圖。

【實施方式】

為了使本技術領域的人員更好地理解本申請中的技術方案，下面將結合本申請實施例中的附圖，對本申請實施例中的技術方案進行清楚、完整地描述，顯然，所描述的實施例僅僅是本申請一部分實施例，而不是全部的實施例。基於本申請中的實施例，本領域普通技術人員在沒有付出創造性勞動前提下所獲得的所有其他實施例，都應當屬於本申請保護的範圍。

圖 2 為本申請一實施例中提供的音頻識別方法的流程圖。本實施例中，所述音頻識別方法包括如下步驟：

S110：對待識別音頻檔案的語譜圖中的第一特徵點進

行擴散處理，得到特徵點圖，所述第一特徵點的數量為多個。

語譜圖也稱為語音頻譜圖，一般是透過處理接收的時域信號得到。一般地，語譜圖的橫坐標用來表示時間，縱坐標用來表示頻率，坐標點值表示語音資料的能量。通常可以採用二維平面來表達三維資訊，所以坐標點值所表示的語音資料的能量值，大小可以透過顏色來表示。例如透過彩色的方式表示，顏色越深的可以表示該坐標點的語音能量越強；反之，顏色越淺的可以表示該坐標點的語音能量越弱。還可以透過灰度的方式表示，顏色越接近於白色的可以表示該坐標點的語音能量越強；反之，顏色越接近於黑色的可以表示該坐標點的語音能量越弱。

這樣，語譜圖可以直觀的表示語音信號隨時間變化的頻譜特性。任一給定頻率成分在給定時刻的強弱用相應點的灰度或色調的濃淡來表示。

具體地，語譜圖可以透過如下步驟獲得：

A1：對待識別音頻檔案按照預設時間進行分幀處理。

所述預設時間可以是用戶根據過往經驗得出的經驗值。本實施例中所述預設時間包括 32 毫秒。即對待識別音頻檔案按照 32 毫秒進行分幀處理，得到每 32 毫秒為一幀，幀疊 16 毫秒的音頻片段。

A2：對分幀處理後的音頻片段進行短時頻譜分析，得到語譜圖。

所述短時頻譜分析包括快速傅立葉變化（Fast Fourier

Transformation，FFT）。FFT 是離散傅立葉變換的快速算法，利用 FFT 可以將音頻信號轉變為記錄了時間域與頻率域的聯合分佈資訊的語譜圖。

由於以 32 毫秒分幀處理，而 32 毫秒對應了 8000hz 採樣，使得 FFT 計算後可以得到 256 頻率點。

如圖 3 中橫軸可以代表幀數，即音頻檔案分幀處理後的幀數的個數，對應了語譜圖的寬度；縱軸可以代表頻率，共有 256 個頻率點，對應了語譜圖的高度；坐標點值表示第一特徵點的能量。

較佳地，在對分幀處理後的音頻片段進行短時頻譜分析之後，還可以包括：

A3：提取所述短時頻譜分析後 300-2khz 頻率段。

由於一般的歌曲主要的頻率是集中在 300-2khz 這個頻率段上的，所以本實施例透過提取 300-2khz 這個頻率段後，即可以消除其它頻率段噪音對所述頻率段的負面影響。

在本申請的另一實施例中，在 S110 步驟之前，還可以包括：

將待識別音頻檔案的語譜圖的第一特徵點的能量值歸一化為第一特徵點的灰度值。

本實施例中，由於經過 FFT 之後的第一特徵點的能量值範圍較大，有時可能達到 0-2^8，甚至 0-2^16（能量值範圍與音頻檔案的信號強度呈正比）；所以，這裡將所述能量值歸一化到 0-255 的範圍內；使得 0-255 可以對應為

灰度值，0代表黑色，255代表白色。

一般的歸一化方法包括：遍歷整個語譜圖中的第一特徵點的能量值，獲得最大值和最小值；

對所述第一特徵點進行歸一化：

$$v=255.\frac{(v-v_{min})}{(v_{max}-v_{min})} \qquad (2)$$

其中，V為第一特徵點的能量值；V_{min}為最小值；V_{max}為最大值。

本申請實施例可以是採用上述一般的歸一化方法。然而，這種歸一化方法，對於可能存在某些弱音時，獲得的V_{min}太小，例如可能趨近於 0，使得歸一化公式變為了$v=255.\frac{v}{v_{max}}$，這樣就與V_{min}無關了。因此這樣的V_{min}不具有代表性，影響了整體的歸一化處理結果。

本申請實施例中提供了一種新的歸一化方法，可以包括：

以第一預設長度為窗口逐幀遍歷語譜圖；

獲取所述窗口內第一特徵點的能量值中的局部最大值和局部最小值；

根據所述局部最大值和局部最小值將第一特徵點的能量值歸一化為第一特徵點的灰度值。

利用（2）所示的公式，其中，V為第一特徵點的能量值；V_{min}為局部最小值；V_{max}為局部最大值。

本實施例以分幀處理後說明，所述第一預設長度可以包括當前幀的前 T 幀到當前幀的後 T 幀。即所述第一預設

長度為 2T 幀，2T+1 幀大於 1 秒。

本實施例提供的歸一化方法，對於某些弱音，只能影響在其所在的第一預設長度內的歸一化結果，不能影響在第一預設長度之外的歸一化結果。所以這樣的歸一化方法可以減少弱音對整體歸一化結果的影響。

所述擴散處理，可以包括高斯函數（Gauss function）擴散處理，即利用高斯函數對第一特徵點進行擴散處理；還可以包括放大處理，即將第一特徵點放大若干倍，例如放大 10 倍。

以下以高斯函數擴散處理為例，利用如下公式：

$$f(x)=ae^{-(x-b)^2/c^2} \qquad (1)$$

其中 a、b 與 c 為常數，且 a>0。

即利用公式（1）對第一特徵點的半徑或直徑進行高斯函數擴散處理。

以下以將第一特徵點放大處理為例。將所述第一特徵點的半徑或直徑放大處理，例如將半徑或直徑放大 10 倍。當然，在某些實施例中，還可以將所述第一特徵點放大若干倍後變為圓形、菱形、矩形等中的至少一種。

如圖 4a 所示，在擴散處理前的白點（待識別音頻檔案的第一特徵點）與黑點（目標音頻檔案的特徵點）存在偏差，進而最後匹配得到的第二特徵點就少；如圖 4b 所示，在擴散處理後的白點從一個點擴散成了一個區域，並且所述區域與黑點都重合。

擴散處理可以使得第一特徵點由點擴散為區域，進而

可以對噪音有一定的抗干擾能力，例如由於噪音干擾的影響，錄製的音頻的第一特徵點可能與原始的音頻的第一特徵點位置有少許的偏差，而透過所述擴散處理後可以忽略這個偏差，增加匹配得到的第二特徵點的數量。

S120：在目標音頻檔案的語譜圖中查找是否存在與所述特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點。

如圖 5 所示，所述 S120 步驟，具體可以包括：

S121：以所述特徵點圖為窗口逐幀遍歷所述目標音頻檔案的語譜圖；

S122：每次遍歷過程中將所述窗口內所述目標音頻檔案的語譜圖中坐標位於所述窗口內擴散處理後第一特徵點的坐標範圍內的特徵點確定為第二特徵點；

S123：查找所述窗口內所述目標音頻檔案的語譜圖中是否存在與所述擴散處理後各第一特徵點分別對應的各第二特徵點。

如圖 6 所示，為在目標音頻檔案的語譜圖中查找與特徵點圖中擴散處理後第一特徵點分別對應的第二特徵點的示意圖。假設特徵點圖的幀數為 N；目標音頻檔案的語譜圖的幀數為 L，所述 L 大於或等於 N。首先在所述目標音頻檔案的語譜圖中幀數為[0，N]的區域內查找；之後在[1，N+1]的區域內查找；這樣逐幀查找，直到[L-N，L]的區域結束遍歷。在每次遍歷過程中每一幀的[t，t+N]的窗口內其中 t 為幀數，將目標音頻檔案的語譜圖中坐標位於

擴散處理後第一特徵點的坐標範圍內的特徵點確定為第二特徵點。在目標音頻檔案內查找與所述擴散處理後各第一特徵點分別對應的各第二特徵點。

在其它實施例中，還可以是遍歷資料庫中所有的音頻檔案。這樣，可以更精確的識別出待識別音頻檔案的音頻資訊。

S130：若是，則確定所述待識別音頻檔案為所述目標音頻檔案的一部分。

如果在目標音頻檔案的語譜圖中查找到與所述擴散處理後各第一特徵點分別對應的第二特徵點，則可以確定所述待識別音頻檔案為所述目標音頻檔案的一部分。

透過本實施例中，對待識別音頻檔案的語譜圖中的第一特徵點進行擴散處理，可以減少所述第一特徵點受噪音影響產生的偏差；從而提高擴散處理後的第一特徵點與目標音頻檔案的匹配率，即可以實現了提高音頻特徵點匹配成功率。

在本申請的一實施例中，所述 S122 步驟，具體可以包括：

確定所述窗口內所述目標音頻檔案的語譜圖中坐標位於所述窗口內擴散處理後第一特徵點的坐標範圍內的特徵點與第一特徵點的匹配度；

將所述匹配度大於第一閾值的特徵點確定為第二特徵點。

所述匹配度包括所述窗口內語譜圖中位於擴散處理後

第一特徵點的坐標範圍內的特徵點個數與第一特徵點個數的比值或所述窗口內語譜圖中位於擴散處理後第一特徵點的坐標範圍內的特徵點對應的第一特徵點的能量值或者灰度值之和。所述第一閾值可以是用戶根據綜合相關因素的一個統計結果。

以所述窗口內語譜圖中位於擴散處理後第一特徵點的坐標範圍內的特徵點個數與第一特徵點個數的比值為例，例如擴散後第一特徵點為 100 個，所述特徵點為 60 個；則所述第一特徵點與所述特徵點的匹配度為 60%。如果所述第一閾值為 80%，那麼將所述特徵點確定為第二特徵點。

以所述窗口內語譜圖中位於擴散處理後第一特徵點的坐標範圍內的特徵點對應的第一特徵點的能量值之和為例，例如特徵點有 10 個，那麼將這 10 個特徵點對應的 10 個第一特徵點的能量值相加，得到能量值之和。如果所述能量值之和大於所述第一閾值，那麼將所述特徵點確定為第二特徵點。

以所述窗口內語譜圖中位於擴散處理後第一特徵點的坐標範圍內的特徵點對應的第一特徵點的灰度值之和為例，例如特徵點有 10 個，那麼將這 10 個特徵點對應的 10 個第一特徵點的灰度值相加，得到灰度值之和。如果所述灰度值之和大於所述第一閾值，那麼將所述特徵點確定為第二特徵點。

在本申請的一實施例中，在 S110 步驟之前，還可以

包括 S101、S102，如圖 7 所示：

S101：將待識別音頻檔案的語譜圖中包含的能量值或者灰度值大於第二閾值的特徵點作為關鍵點；

所述第二閾值可以是用戶根據綜合相關因素的一個統計結果；第二閾值越小，可以提取的關鍵點就越多，進而可能造成後續匹配時間越久；第二閾值越大，可以提取的關鍵點就越少，進而可能造成後續匹配的成功機率過低。

S102：若所述關鍵點的能量值或者灰度值在預設區域內為最大值，則將該關鍵點確定為第一特徵點；

所述預設區域可以是以所述關鍵點為中心並根據預設半徑確定的圓形區域；或者以所述關鍵點為中心並根據預設長和寬確定的矩形區域。

所述預設區域可以是用戶根據綜合相關因素的一個統計結果；預設區域越小，可以確定的第一特徵點越多，進而可能造成後續匹配時間越久；預設區域越大，可以確定的第一特徵點越少，進而可能造成後續匹配的成功機率過低。

如圖 8a 所示，為確定的第一特徵點在語譜圖上的示意圖。圖中白點即第一特徵點。具體的，假設所述第二預設閾值為 30，所述預設區域為 15*15（以關鍵點為中心，橫坐標取 15 幀，縱坐標取長度 15），如圖 8b 所示，為圖 8a 的局部放大示意圖；圖中白點的能量值或者灰度值即大於第一預設閾值 30 並且在預設區域 15*15 內依然是最大值，提取出這樣的點作為第一特徵點。

本申請實施例與上一實施例不同之處在於，透過提取語譜圖中能量值或者灰度值大的特徵點作為第一特徵點，從而可以排除能量弱的特徵點對後續匹配的干擾，並且還可以大大的減少擴散處理的資料量，進而提高系統性能。

在本申請的一實施例中，所述目標音頻檔案可以攜帶有音頻資訊。本申請應用於歌曲識別場景中時，所述音頻資訊可以包括歌曲名。用戶錄製一段不知道歌曲名的待識別音頻檔案或待識別音頻檔案就是一首不知道歌曲名的歌曲，當確定待識別音頻檔案為目標音頻檔案的一部分時，就可以識別出所述待識別音頻檔案的歌曲名。

圖 9 為本申請一實施例中提供的音頻識別系統的模組示意圖。本實施例中，所述音頻識別系統包括：

擴散單元 210，用於對待識別音頻檔案的語譜圖中的第一特徵點進行擴散處理，得到特徵點圖，所述第一特徵點的數量為多個；

查找單元 220，在目標音頻檔案的語譜圖中查找是否存在與所述特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點；

確定單元 230，用於在目標音頻檔案的語譜圖中查找到與所述特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點的區域時，則確定所述待識別音頻檔案為所述目標音頻檔案的一部分。

較佳地，在所述擴散單元 210 之前，還可以包括：

歸一化單元，用於將待識別音頻檔案的語譜圖的第一

特徵點的能量值歸一化為第一特徵點的灰度值。

較佳地，所述擴散處理包括高斯函數擴散處理或者放大處理中的至少一種。

較佳地，所述歸一化單元，具體可以包括：

第一歸一化子單元，用於以第一預設長度為窗口逐幀遍歷語譜圖；

第二歸一化子單元，用於獲取所述窗口內第一特徵點的能量值中的局部最大值和局部最小值；

第三歸一化子單元，用於根據所述局部最大值和局部最小值將第一特徵點的能量值歸一化為第一特徵點的灰度值。

較佳地，所述查找單元 220，具體可以包括：

第一查找子單元，用於以所述特徵點圖為窗口逐幀遍歷所述目標音頻檔案的語譜圖；

第二查找子單元，用於每次遍歷過程中將所述窗口內所述目標音頻檔案的語譜圖中坐標位於所述窗口內擴散處理後第一特徵點的坐標範圍內的特徵點確定為第二特徵點；

第三查找子單元，用於查找所述窗口內所述目標音頻檔案的語譜圖中是否存在與所述擴散處理後各第一特徵點分別對應的各第二特徵點。

較佳地，所述第二查找子單元，具體可以包括：

第四查找子單元，用於確定所述窗口內所述目標音頻檔案的語譜圖中坐標位於所述窗口內擴散處理後第一特徵

點的坐標範圍內的特徵點與所述第一特徵點的匹配度；

第五查找子單元，用於將所述匹配度大於第一閾值的特徵點確定為第二特徵點。

較佳地，所述匹配度包括所述窗口內語譜圖中位於擴散處理後第一特徵點的坐標範圍內的特徵點個數與第一特徵點個數的比值或所述窗口內語譜圖中位於擴散處理後第一特徵點的坐標範圍內的特徵點對應的第一特徵點的能量值或者灰度值之和。

較佳地，在所述擴散處理之前，還可以包括：

第一處理單元，用於將待識別音頻檔案的語譜圖中包含的能量值或者灰度值大於第二閾值的特徵點作為關鍵點；

第二處理單元，用於在所述關鍵點的能量值或者灰度值在預設區域內為最大值時，將該關鍵點確定為第一特徵點。

較佳地，所述目標音頻檔案攜帶有音頻資訊，所述音頻資訊包括歌曲名。

在20世紀90年代，對於一個技術的改進可以很明顯地區分是硬體上的改進（例如，對二極體、電晶體、開關等電路結構的改進）還是軟體上的改進（對於方法流程的改進）。然而，隨著技術的發展，當今的很多方法流程的改進已經可以視為硬體電路結構的直接改進。設計人員幾乎都透過將改進的方法流程編程到硬體電路中來得到相應的硬體電路結構。因此，不能說一個方法流程的改進就不

能用硬體實體模組來實現。例如，可編程邏輯裝置（Programmable Logic Device, PLD）（例如現場可編程閘陣列（Field Programmable Gate Array，FPGA））就是這樣一種集成電路，其邏輯功能由用戶對裝置編程來確定。由設計人員自行編程來把一個數位系統“集成”在一片 PLD 上，而不需要請晶片製造廠商來設計和製作專用的集成電路晶片。而且，如今，取代手工地製作集成電路晶片，這種編程也多半改用“邏輯編譯器（logic compiler）”軟體來實現，它與程式開發撰寫時所用的軟體編譯器相類似，而要編譯之前的原始代碼也得用特定的編程語言來撰寫，此稱之為硬體描述語言（Hardware Description Language，HDL），而 HDL 也並非僅有一種，而是有許多種，如 ABEL（Advanced Boolean Expression Language）、AHDL（Altera Hardware Description Language）、Confluence、CUPL（Cornell University Programming Language）、HDCal、JHDL（Java Hardware Description Language）、Lava、Lola、MyHDL、PALASM、RHDL（Ruby Hardware Description Language）等，目前最普遍使用的是 VHDL（Very-High-Speed Integrated Circuit Hardware Description Language）與 Verilog。本領域技術人員也應該清楚，只需要將方法流程用上述幾種硬體描述語言稍作邏輯編程並編程到集成電路中，就可以很容易得到實現該邏輯方法流程的硬體電路。

控制器可以按任何適當的方式實現，例如，控制器可

以採取例如微處理器或處理器以及儲存可由該（微）處理器執行的電腦可讀程式代碼（例如軟體或韌體）的電腦可讀媒體、邏輯閘、開關、專用集成電路（Application Specific Integrated Circuit，ASIC）、可編程邏輯控制器和嵌入微控制器的形式，控制器的例子包括但不限於以下微控制器：ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20 以及 Silicone Labs C8051F320，記憶體控制器還可以被實現為記憶體的控制邏輯的一部分。本領域技術人員也知道，除了以純電腦可讀程式代碼方式實現控制器以外，完全可以透過將方法步驟進行邏輯編程來使得控制器以邏輯閘、開關、專用集成電路、可編程邏輯控制器和嵌入微控制器等的形式來實現相同功能。因此這種控制器可以被認為是一種硬體部件，而對其內包括的用於實現各種功能的裝置也可以視為硬體部件內的結構。或者甚至，可以將用於實現各種功能的裝置視為既可以是實現方法的軟體模組又可以是硬體部件內的結構。

上述實施例闡明的系統、裝置、模組或單元，具體可以由電腦晶片或實體實現，或者由具有某種功能的產品來實現。

為了描述的方便，描述以上裝置時以功能分為各種單元分別描述。當然，在實施本申請時可以把各單元的功能在同一個或多個軟體和／或硬體中實現。

本領域內的技術人員應明白，本發明的實施例可提供為方法、系統、或電腦程式產品。因此，本發明可採用完

全硬體實施例、完全軟體實施例、或結合軟體和硬體方面的實施例的形式。而且，本發明可採用在一個或多個其中包含有電腦可用程式代碼的電腦可用儲存媒體（包括但不限於磁盤記憶體、CD-ROM、光學記憶體等）上實施的電腦程式產品的形式。

本發明是參照根據本發明實施例的方法、設備（系統）、和電腦程式產品的流程圖和／或方框圖來描述的。應理解可由電腦程式指令實現流程圖和／或方框圖中的每一流程和／或方框、以及流程圖和／或方框圖中的流程和／或方框的結合。可提供這些電腦程式指令到通用電腦、專用電腦、嵌入式處理機或其他可編程資料處理設備的處理器以產生一個機器，使得透過電腦或其他可編程資料處理設備的處理器執行的指令產生用於實現在流程圖一個流程或多個流程和／或方框圖一個方框或多個方框中指定的功能的裝置。

這些電腦程式指令也可儲存在能引導電腦或其他可編程資料處理設備以特定方式工作的電腦可讀記憶體中，使得儲存在該電腦可讀記憶體中的指令產生包括指令裝置的製造品，該指令裝置實現在流程圖一個流程或多個流程和／或方框圖一個方框或多個方框中指定的功能。

這些電腦程式指令也可裝載到電腦或其他可編程資料處理設備上，使得在電腦或其他可編程設備上執行一系列操作步驟以產生電腦實現的處理，從而在電腦或其他可編程設備上執行的指令提供用於實現在流程圖一個流程或多

個流程和／或方框圖一個方框或多個方框中指定的功能的步驟。

在一個典型的配置中，計算設備包括一個或多個處理器(CPU)、輸入/輸出介面、網路介面和記憶體。

記憶體可能包括電腦可讀媒體中的非永久性記憶體，隨機存取記憶體(RAM)和/或非揮發性記憶體等形式，如唯讀記憶體(ROM)或快閃記憶體(flash RAM)。記憶體是電腦可讀媒體的示例。

電腦可讀媒體包括永久性和非永久性、可移動和非可移動媒體可以由任何方法或技術來實現資訊儲存。資訊可以是電腦可讀指令、資料結構、程式的模組或其他資料。電腦的儲存媒體的例子包括，但不限於相變記憶體(PRAM)、靜態隨機存取記憶體(SRAM)、動態隨機存取記憶體(DRAM)、其他類型的隨機存取記憶體(RAM)、唯讀記憶體(ROM)、電可擦除可編程唯讀記憶體(EEPROM)、快閃記憶體或其他記憶體技術、唯讀光碟唯讀記憶體(CD-ROM)、數位多功能光碟(DVD)或其他光學儲存、磁盒式磁帶，磁帶磁盤儲存或其他磁性儲存設備或任何其他非傳輸媒體，可用於儲存可以被計算設備訪問的資訊。按照本文中的界定，電腦可讀媒體不包括暫存電腦可讀媒體(transitory media)，如調製的資料信號和載波。

還需要說明的是，術語“包括”、“包含”或者其任何其他變體意在涵蓋非排他性的包含，從而使得包括一系列要素的過程、方法、商品或者設備不僅包括那些要素，而且

還包括沒有明確列出的其他要素，或者是還包括為這種過程、方法、商品或者設備所固有的要素。在沒有更多限制的情況下，由語句“包括一個……”限定的要素，並不排除在包括所述要素的過程、方法、商品或者設備中還存在另外的相同要素。

本領域技術人員應明白，本申請的實施例可提供為方法、系統或電腦程式產品。因此，本申請可採用完全硬體實施例、完全軟體實施例或結合軟體和硬體方面的實施例的形式。而且，本申請可採用在一個或多個其中包含有電腦可用程式代碼的電腦可用儲存媒體（包括但不限於磁碟記憶體、CD-ROM、光學記憶體等）上實施的電腦程式產品的形式。

本申請可以在由電腦執行的電腦可執行指令的一般上下文中描述，例如程式模組。一般地，程式模組包括執行特定任務或實現特定抽象資料類型的常式、程式、物件、元件、資料結構等等。也可以在分散式運算環境中實踐本申請，在這些分散式運算環境中，由透過通信網路而被連接的遠程處理設備來執行任務。在分散式運算環境中，程式模組可以位於包括儲存設備在內的本地和遠程電腦儲存媒體中。

本說明書中的各個實施例均採用遞進的方式描述，各個實施例之間相同相似的部分互相參見即可，每個實施例重點說明的都是與其他實施例的不同之處。尤其，對於系統實施例而言，由於其基本相似於方法實施例，所以描述

的比較簡單，相關之處參見方法實施例的部分說明即可。

以上所述僅為本申請的實施例而已，並不用於限制本申請。對於本領域技術人員來說，本申請可以有各種更改和變化。凡在本申請的精神和原理之內所作的任何修改、等同替換、改進等，均應包含在本申請的申請專利範圍的範圍之內。

【符號說明】

210：擴散單元

220：查找單元

230：確定單元

201828285

發明摘要

※申請案號：106101958

※申請日：106年01月19日　　**※IPC分類：** *G10L 19/02* (2013.01)
G10L 15/08 (2006.01)

【發明名稱】(中文/英文)

音頻識別方法和系統

【中文】

本申請實施例公開了一種音頻識別方法，包括：對待識別音頻檔案的語譜圖中的第一特徵點進行擴散處理，得到特徵點圖，所述第一特徵點的數量為多個；在目標音頻檔案的語譜圖中查找是否存在與所述特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點；若是，則確定所述待識別音頻檔案為所述目標音頻檔案的一部分。本申請還公佈了一種音頻識別系統實施例。利用本實施例可以在音頻識別中提高特徵點匹配成功率。

【英文】

【代表圖】

【本案指定代表圖】：第(２)圖。

【本代表圖之符號簡單說明】：無

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：無

申請專利範圍

1. 一種音頻識別方法，其特徵在於，包括：

對待識別音頻檔案的語譜圖中的第一特徵點進行擴散處理，得到特徵點圖，該第一特徵點的數量為多個；

在目標音頻檔案的語譜圖中查找是否存在與該特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點；

若是，則確定該待識別音頻檔案為該目標音頻檔案的一部分。

2. 如申請專利範圍第 1 項所述的方法，其中，在所述對待識別音頻檔案的語譜圖中的第一特徵點進行擴散處理之前，還包括：

將待識別音頻檔案的語譜圖中的第一特徵點的能量值歸一化為第一特徵點的灰度值。

3. 如申請專利範圍第 1 或 2 項所述的方法，其中，該擴散處理包括高斯函數擴散處理或者放大處理中的至少一種。

4. 如申請專利範圍第 2 項所述的方法，其中，將待識別音頻檔案的語譜圖中的第一特徵點的能量值歸一化為第一特徵點的灰度值，具體包括：

以第一預設長度為窗口逐幀遍歷語譜圖；

獲取該窗口內第一特徵點的能量值中的局部最大值和局部最小值；

根據該局部最大值和局部最小值將第一特徵點的能量值歸一化為第一特徵點的灰度值。

5. 如申請專利範圍第 1 或 2 項所述的方法，其中，所述在目標音頻檔案的語譜圖中查找是否存在與該特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點，具體包括：

以該特徵點圖為窗口逐幀遍歷該目標音頻檔案的語譜圖；

每次遍歷過程中將該窗口內該目標音頻檔案的語譜圖中坐標位於該窗口內擴散處理後第一特徵點的坐標範圍內的特徵點確定為第二特徵點；

查找該窗口內該目標音頻檔案的語譜圖中是否存在與該擴散處理後各第一特徵點分別對應的各第二特徵點。

6. 如申請專利範圍第 5 項所述的方法，其中，所述將該窗口內該目標音頻檔案的語譜圖中坐標位於該窗口內擴散處理後第一特徵點的坐標範圍內的特徵點確定為第二特徵點，包括：

確定該窗口內該目標音頻檔案的語譜圖中坐標位於該窗口內擴散處理後第一特徵點的坐標範圍內的特徵點與第一特徵點的匹配度；

將該匹配度大於第一閾值的特徵點確定為第二特徵點。

7. 如申請專利範圍第 6 項所述的方法，其中，該匹配度包括該窗口內語譜圖中位於擴散處理後第一特徵點的坐標範圍內的特徵點個數與第一特徵點個數的比值或該窗口內語譜圖中位於擴散處理後第一特徵點的坐標範圍內的

特徵點對應的第一特徵點的能量值或者灰度值之和。

8. 如申請專利範圍第 1 或 2 項所述的方法，其中，在所述對待識別音頻檔案的語譜圖的第一特徵點進行擴散處理之前，還包括：

將待識別音頻檔案的語譜圖中包含的能量值或者灰度值大於第二閾值的特徵點作為關鍵點；

若該關鍵點的能量值或者灰度值在預設區域內為最大值，則將該關鍵點確定為第一特徵點。

9. 如申請專利範圍第 1 項所述的方法，其中，該目標音頻檔案攜帶有音頻資訊，該音頻資訊包括歌曲名。

10. 一種音頻識別系統，其特徵在於，包括：

擴散單元，用於對待識別音頻檔案的語譜圖中的第一特徵點進行擴散處理，得到特徵點圖，該第一特徵點的數量為多個；

查找單元，在目標音頻檔案的語譜圖中查找是否存在與該特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點；

確定單元，用於在目標音頻檔案的語譜圖中查找到與該特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點時，則確定該待識別音頻檔案為該目標音頻檔案的一部分。

11. 如申請專利範圍第 10 項所述的系統，其中，在該擴散單元之前，還包括：

歸一化單元，用於將待識別音頻檔案的語譜圖中的第

一特徵點的能量值歸一化為第一特徵點的灰度值。

12. 如申請專利範圍第 10 或 11 項所述的系統，其中，該擴散處理包括高斯函數擴散處理或者放大處理中的至少一種。

13. 如申請專利範圍第 11 項所述的系統，其中，該歸一化單元，具體包括：

第一歸一化子單元，用於以第一預設長度為窗口逐幀遍歷語譜圖；

第二歸一化子單元，用於獲取該窗口內第一特徵點的能量值中的局部最大值和局部最小值；

第三歸一化子單元，用於根據該局部最大值和局部最小值將第一特徵點的能量值歸一化為第一特徵點的灰度值。

14. 如申請專利範圍第 10 或 11 項所述的系統，其中，該查找單元，具體包括：

第一查找子單元，用於以該特徵點圖為窗口逐幀遍歷該目標音頻檔案的語譜圖；

第二查找子單元，用於每次遍歷過程中將該窗口內該目標音頻檔案的語譜圖中坐標位於該窗口內擴散處理後第一特徵點的坐標範圍內的特徵點確定為第二特徵點；

第三查找子單元，用於查找該窗口內該目標音頻檔案的語譜圖中是否存在與該擴散處理後各第一特徵點分別對應的各第二特徵點。

15. 如申請專利範圍第 14 項所述的系統，其中，該

第二查找子單元，具體包括：

第四查找子單元，用於確定該窗口內該目標音頻檔案的語譜圖中坐標位於該窗口內擴散處理後第一特徵點的坐標範圍內的特徵點與第一特徵點的匹配度；

第五查找子單元，用於將該匹配度大於第一閾值的特徵點確定為第二特徵點。

16. 如申請專利範圍第 15 項所述的系統，其中，該匹配度包括該窗口內語譜圖中位於擴散處理後第一特徵點的坐標範圍內的特徵點個數與第一特徵點個數的比值或該窗口內語譜圖中位於擴散處理後第一特徵點的坐標範圍內的特徵點對應的第一特徵點的能量值或者灰度值之和。

17. 如申請專利範圍第 10 或 11 項所述的系統，其中，在該擴散處理之前，還包括：

第一處理單元，用於將待識別音頻檔案的語譜圖中包含的能量值或者灰度值大於第二閾值的特徵點作為關鍵點；

第二處理單元，用於在該關鍵點的能量值或者灰度值在預設區域內為最大值時，將該關鍵點確定為第一特徵點。

18. 如申請專利範圍第 10 項所述的系統，其中，該目標音頻檔案攜帶有音頻資訊，該音頻資訊包括歌曲名。

圖式

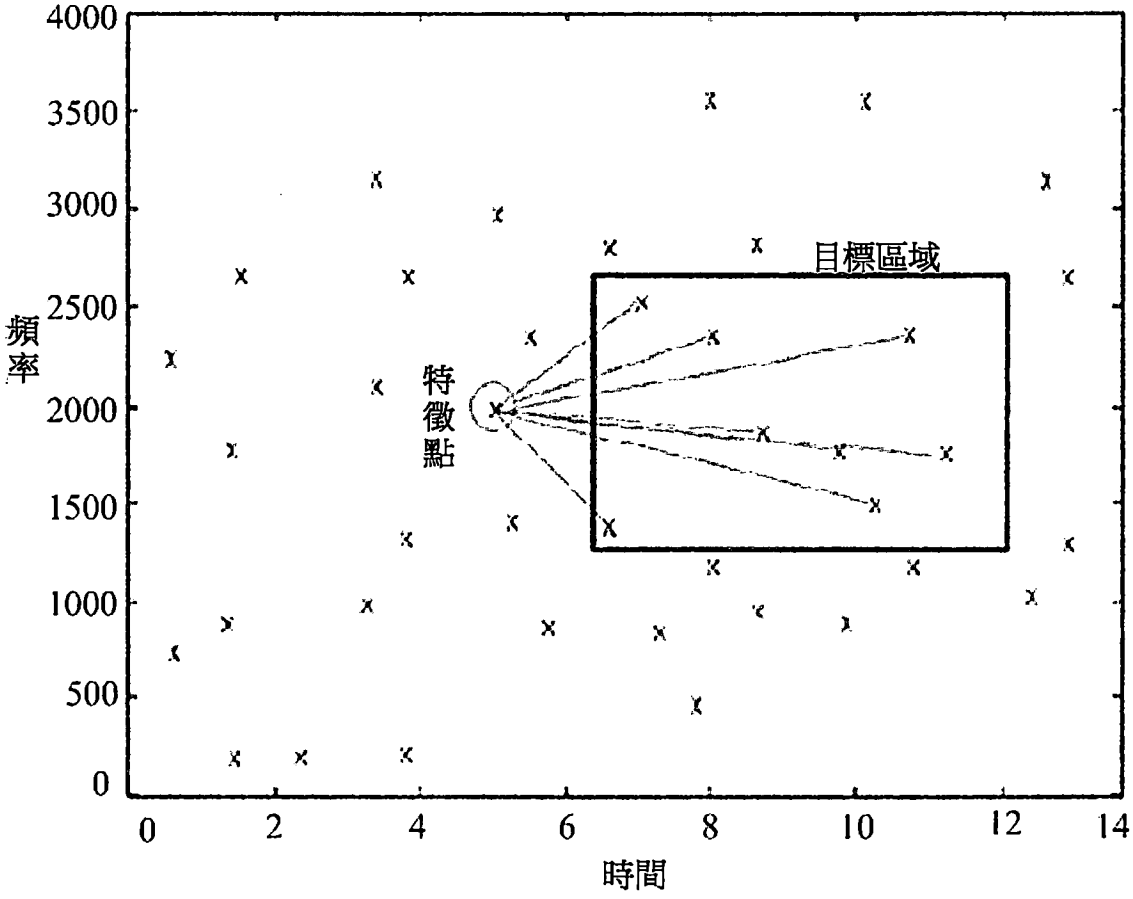
4000
3500
3000
2500
2000
1500
1000
500
0
頻率
0
2
4
6
8
10
12
14
時間
目標區域
特徵點

圖 1

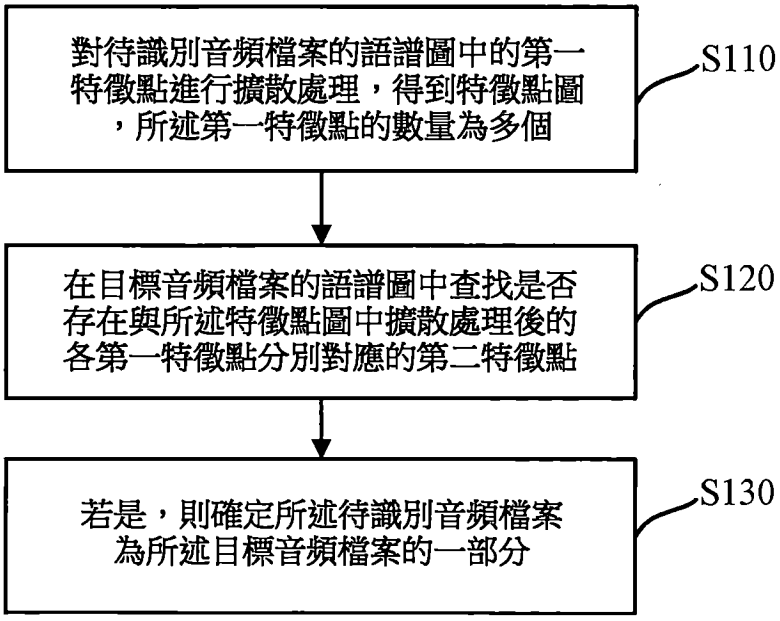
對待識別音頻檔案的語譜圖中的第一特徵點進行擴散處理，得到特徵點圖，所述第一特徵點的數量為多個
S110
在目標音頻檔案的語譜圖中查找是否存在與所述特徵點圖中擴散處理後的各第一特徵點分別對應的第二特徵點
S120
若是，則確定所述待識別音頻檔案為所述目標音頻檔案的一部分
S130

圖 2