



(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 201246183 A1

(43)公開日：中華民國 101 (2012) 年 11 月 16 日

(21)申請案號：101102127

(22)申請日：中華民國 101 (2012) 年 01 月 19 日

(51)Int. Cl. : **G10L11/00 (2006.01)**
G10L15/10 (2006.01)

G10L15/02 (2006.01)

(30)優先權：2011/02/10 美國 13/025,060

(71)申請人：雅虎股份有限公司 (美國) YAHOO! INC. (US)
美國

(72)發明人：畢羅拉夫 社吉 BILOBROV, SERGIY (CA)

(74)代理人：陳傳岳；郭兩嵐

申請實體審查：有 申請專利範圍項數：21 項 圖式數：8 共 43 頁

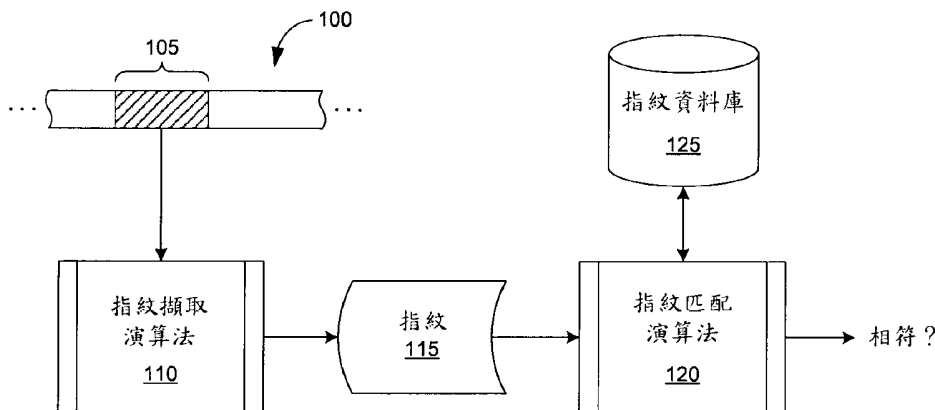
(54)名稱

聲音訊號之特徵指紋擷取與匹配

EXTRACTION AND MATCHING OF CHARACTERISTIC FINGERPRINTS FROM AUDIO SIGNALS

(57)摘要

從一聲音樣本擷取一聲音指紋，其中該指紋包含的資訊為該樣本中內容的特徵。該指紋可以利用計算代表該聲音樣本之一能量頻譜、對該能量頻譜再取樣、轉換該已再取樣能量頻譜以產生一連串特徵向量，並利用該特徵向量之差動編碼計算該指紋的方式所產生。所產生的指紋可以與一資料庫中的一組參考指紋比較，以識別該原始聲音內容。



100：聲音樣本

105：聲音訊框

110：指紋擷取演算法

115：指紋

120：指紋匹配演算法

125：指紋資料庫



(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 201246183 A1

(43)公開日：中華民國 101 (2012) 年 11 月 16 日

(21)申請案號：101102127

(22)申請日：中華民國 101 (2012) 年 01 月 19 日

(51)Int. Cl. : **G10L11/00 (2006.01)**
G10L15/10 (2006.01)

G10L15/02 (2006.01)

(30)優先權：2011/02/10 美國 13/025,060

(71)申請人：雅虎股份有限公司 (美國) YAHOO! INC. (US)
美國

(72)發明人：畢羅拉夫 社吉 BILOBROV, SERGIY (CA)

(74)代理人：陳傳岳；郭兩嵐

申請實體審查：有 申請專利範圍項數：21 項 圖式數：8 共 43 頁

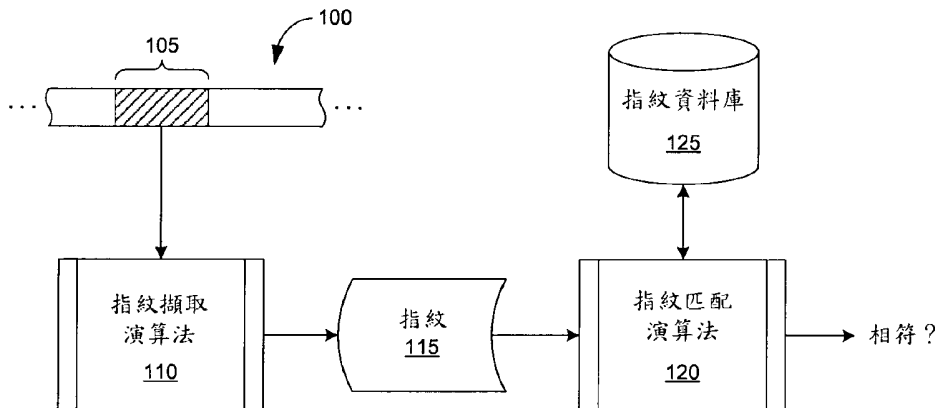
(54)名稱

聲音訊號之特徵指紋擷取與匹配

EXTRACTION AND MATCHING OF CHARACTERISTIC FINGERPRINTS FROM AUDIO SIGNALS

(57)摘要

從一聲音樣本擷取一聲音指紋，其中該指紋包含的資訊為該樣本中內容的特徵。該指紋可以利用計算代表該聲音樣本之一能量頻譜、對該能量頻譜再取樣、轉換該已再取樣能量頻譜以產生一連串特徵向量，並利用該特徵向量之差動編碼計算該指紋的方式所產生。所產生的指紋可以與一資料庫中的一組參考指紋比較，以識別該原始聲音內容。



100：聲音樣本

105：聲音訊框

110：指紋擷取演算法

115：指紋

120：指紋匹配演算法

125：指紋資料庫

六、發明說明：

【發明所屬之技術領域】

本發明一般來說與聲音訊號處理有關，且更特別的是與從聲音訊號取得特徵指紋以及搜尋所述指紋之一資料庫有關。

【先前技術】

由於檔案格式、壓縮技術與其他重現資料方法的多樣性，識別一資料訊號或是將其與其他資料相比較的問題，產生了顯著的技術難度。例如，在電腦上數位音樂檔案的情況中，有許多進行歌曲編碼及壓縮的格式。此外，該歌曲常以不同的頻率取樣為數位形式，並具有不同的特徵(例如，不同的波形)。紀錄類比聲音也會包含噪音及失真。這些顯著的波形差異使得在將所述檔案進行直接比較，對於有效的檔案或訊號辨識或比較時選擇性不多。直接的檔案比較也不允許對編碼成不同格式的媒體進行比較(例如，比較被編碼成為 MP3 及 WAV 的相同歌曲)。

由於這些理由，識別及追蹤像是散佈於網際網路上的媒體及其他內容時，常以附加超資料、浮水印或是某些其他包含代表該媒體之訊編碼的方式完成。但是此附加資訊常常不完整及/或不正確。例如，超資料常缺少資料，且檔名通常更不具有均一性。此外，像是浮水印方式為一種侵入行為，其利用所增加的資料或編碼改變了原始檔案。這些方法的另一種缺點為容易受到竄改攻擊。即使每一媒體檔案都包含像是超資料或浮水印的正確識別資料，如果該資訊被成功移除時該檔案便成為“解鎖”狀態(並因此被剽竊)。

為了避免這些問題，已經根據分析一資料訊號本身內容的蓋念發展其他的方法。在其中一類方法中，產生代表一聲音部分的聲音指紋，其中該指紋包含有關該聲音可以被使用於識別該原始聲音的特徵資訊。在一範例中，一聲音指紋包括識別一聲音片段的數位序列。產生一聲音指紋的處理則常根據於該指紋將被產生所代表之聲音的聲學及知覺性質。聲音指紋相對於原始聲音內容而言一般上具有較小的大小，並因此可以做為用於識別、比較及搜尋聲音內容的方便工具。聲音指紋可以在廣泛多種應用中使用，包含廣播監測、聲音內容組織、P2P 網路的內容過濾，以及歌曲或其他聲音內容的識別。當應用於這些多種領域時，聲音指印一般上與指紋擷取有關，也與指紋資料庫搜尋演算法有關。

現有的大多數指紋技術係根據於從頻率域中的聲音樣本擷取聲音特徵。首先將該聲音切割為訊框，並計算代表每一訊框的一組特徵。在該聲音特徵之中，可以被使用的是快速傅立葉轉換(FFT)係數、梅爾頻率倒頻譜係數(MFCC)、頻譜平坦度、尖銳度、熵能與調變頻率。所計算的特徵則被組合成為一特徵向量，其通常利用導出函數、平均或變異數進行轉換。該特徵向量則在量化之後利用像是主成分分析的演算法映射成一更壓縮的表現方式，以產生該聲音指紋。通常來說，利用處理一單一聲音訊框方式所獲得的指紋具有一相對小的尺寸，而可能無法以足夠的可靠度唯一識別該原始聲音序列。為了加強指紋獨特性並因此增加正確辨識的可能性(及減少主動錯誤率)，可以將小的次指紋結合成為代表大約 3 至 5 秒聲音的較大區塊。

一種由 Philips 所發展指紋技術是利用一種短時傅立葉轉換(STFT)於一聲音訊號每 11.8 微秒期間擷取一 32 位元(bit) 的次指

紋。該聲音訊號首先被切割為 0.37 秒長的重疊訊框，而該訊框以 31/32 的重疊因子利用一 Hamming 窗所加權，並將其利用 FFT 轉換至頻率域中。所獲得的頻率域資料可以以頻譜圖表現(例如時間頻率圖)，其中水平軸為時間而垂直軸為頻率。每一訊框的頻譜(頻譜圖欄)以對數間隔方式於 300Hz 至 2000Hz 範圍中被切割為 33 個非重疊頻帶。計算每一頻帶的頻譜能量，並利用與沿著該時間及頻率軸連續頻率中之能量差異的符號產生一 32 bit 次指紋。如果在一訊框兩頻帶之間的能量差異大於先前訊框中相同頻帶之間的能量差異，演算法便在該次指紋中對應位元處輸出“1”；否則便在該對應位元處輸出“0”。以結合 256 個連續 32 bit 次指紋成為單一指紋區塊的方式組合一指紋，其對應於 3 秒的聲音。

雖然此演算法的設計而言對於聲音處理、噪音以及失真的一般處理而言係健全的，但對於由於所形成頻譜尺度所引起的大速度改變而言則不健全。據此，便提出一種在尺度變化 Fourier-Mellin 域中擷取聲音指紋的修正演算法。此修正演算法包含在將該生音訊框轉換至頻率域之後所執行的額外步驟。這些額外步驟包含頻譜對數映射之後進行一第二次傅立葉轉換。因此對於每一訊框而言，應用一第一次快速傅立葉轉換，將該結果進行對數映射以獲得一功率頻譜，並應用一第二次快速傅立葉轉換。此過程可稱為對數再取樣傅立葉轉換的傅立葉轉換，並與已知在語音辨識中廣泛使用的 MFCC 方法類似。其主要差異在於 Fourier-Mellin 轉換使用完整頻譜的對數映射，而 MFCC 則係根據梅爾頻率的尺度(最大為 1kHz 的線性頻率，並於較高頻率具有對數間隔，以模仿人類聲音系統的性質)。

Philips 演算法則屬於一種所謂短期分析演算法的範疇，因為

其僅利用兩連續訊框的頻譜係數計算次指紋。也存在利用在頻譜圖中多個重疊快速傅立葉轉換訊框擷取頻譜特徵的其他演算法。某些根據於在時間域中評估多個訊框的所述方法則被稱為長期頻譜圖分析演算法。

一種在 Sukittanon, “Modulation-Scale Analysis for Content Identification,” IEEE Transactions on Signal Processing, vol. 52, no.10 (Oct. 2004) 中所描述之範例的長期分析演算法則係根據對於調變頻率的估計。在此演算法中，劃分該聲音並計算其頻譜圖。接著利用沿著該頻譜時間列(例如，水平軸)應用一第二次轉換的方式，計算代表每一頻譜圖頻帶(例如，該頻譜圖中一頻率範圍)的調變頻譜。這與所述修正 Philips 的方法不同，其中該第二次快速傅立葉轉換是沿著該頻譜圖頻率欄(例如，垂直軸)所應用。在此方法中，該頻譜圖被劃分為 N 個頻帶，並為每一頻帶計算相同數量 N 的連續小波轉換。

雖然此演算法的開發者主張相較於 Philips 演算法而言具有更好的效能，但現有的演算法仍存在一些不足之處。例如，該演算法對於可靠識別失真的語音與音樂而言並不具有足夠健全性，特別是利用 CELP 聲音編碼技術進行壓縮的聲音(例如，於像是 GSM 的手機音樂所相關的聲音)。此外，這些演算法一般而言對於噪音與類比失真而言較為敏感，像是與麥克風錄音關聯的情況之中。而即使該演算法可以在存在單一失真形式下識別聲音，也無法處理多數失真的組合情況，而此情況更加常見，且更接近於真實世界的情形(例如，在手機情況下於具有輕微反射的吵雜環境中利用麥克風進行錄音，並以 GSM 壓縮技術處理的聲音)。

因此當用於實際應用時，現有的擷取指紋技術則存在無法接

受的高錯誤比率(例如，主動錯誤與被動錯誤)、所產生的指紋太大而無法進行商業利用及/或速度太慢。據此，便存在必須克服目前聲音辨識技術所無法解決之現有限制的需求。

【發明內容】

據此，本發明能夠根據一聲音訊號的內容，從該訊號擷取一特徵指紋。此指紋可以對於一組參考指紋(例如，在一資料庫中)進行匹配，以決定該訊號的唯一性或兩訊號之間的相似性。由於該指紋擷取演算法的本質，並不需要承受像是目前解決方式中所遭遇的困難，且與此類解決方式相比之下，其具有快速、有效率、高正確性、可縮放性與強健性的特性。

在用於產生一聲音指紋的一方法實施例中，對一聲音訊號進行取樣並從該訊號計算頻譜資訊。該頻譜被劃分為多數個頻帶。並對該頻帶中的序列樣本進行再取樣。

在一實施例中，對該序列樣本再取樣包括以對數方式映射該樣本。在另一實施例中，對該序列樣本再取樣包括根據該對應頻帶的一中間頻率及/或頻率範圍，於時間域中對該序列樣本的大小進行縮放，並對該已縮放序列樣本進行再取樣。再另一實施例中，對該序列樣本再取樣包括根據該對應頻帶的一中間頻率及/或頻率範圍，於時間域中對該序列樣本進行偏移，並對該偏移序列樣本進行再取樣。在另一實施例中，對該序列樣本再取樣包括隨時間改變從不同序列樣本(也就是頻帶)進行取樣。

接著對該再取樣序列進行一第二次轉換，以獲得代表每一序列的一特徵向量。在一實施例中，該第二次轉換包括沿著該時間

軸進行轉換。在另一實施例中，該第二次轉換包括在沿著該時間軸進行轉換後，再沿著該頻率軸進行轉換。在另一實施例中，該第二次轉換包括一二維離散餘弦轉換(2D DCT)。該聲音指紋便根據該特徵向量加以計算。該聲音指紋可以儲存在一電腦可讀媒介或可以暫時儲存為一種可傳輸訊號。

對於該序列樣本所進行的不同再取樣形式使該演算法對於聲音撥放速度及時間壓縮與延伸的變化較不敏感。在此方法中，一聲音訊號的指紋在有關撥放速度或因時間壓縮或延伸的變化方面，應該具有極少或不具有變化。所述之再取樣也改善該第二次時頻轉換的低頻解析度。這使其能採用一種簡單的轉換，而取代在頻譜圖調變頻譜分析中所使用的複雜小波轉換，使該實作相對於先前方法而言更有效率也更快速。

此外，因為採用所述再取樣，該頻帶輸出訊框對大部分而言便包含表現該被分析聲音序列起始時的樣本。所形成的指紋便因此主要利用位於該序列起始處的樣本所產生。因為該聲音序列的一相對小部分對於所形成指紋具有大多的貢獻，該指紋便可用於匹配較短的聲音序列。例如在一實作中，由 5 秒原始聲音訊框所產生的指紋便可以與從兩倍短聲音片段所取得的樣本進行可靠匹配。

該指紋擷取技術的實施例也容忍噪音與訊號失真。一種實作可以在存在 100% 白色雜訊(也就是 0 db 的訊號雜訊比)的情況中偵測類語音的訊號。該技術對於濾波、壓縮、頻率等化與相位失真的情況也具有容忍度。

在另一實施例中，其中所產生的指紋訊框是利用一特定數量

頻帶所形成，其使用一種聲學模型標示不重要的頻帶。不重要的頻帶可以包含在分辨聲音樣本中並不實質增加任何可感知數值的頻帶。只處理相關頻帶提高訊號雜訊比，並改善整體指紋匹配處理的健全性。此外，排除不相關的頻帶可以大幅改善受限頻帶聲音內容的辨識效率，例如在以非常低位元率編碼的語音情況或是具有低速的類比紀錄情況。

本發明實施例也具備快速標記及大規模資料庫中的指紋效率搜尋。例如，可以從該指紋內容的一部份計算代表每一聲音指紋的索引。在一實施例中，使用來自一指紋的位元組作為該指紋索引，其中該位元對應於由於再取樣所形成的較穩定低頻係數。為了匹配一測試指紋與一資料庫中的一組指紋，該測試指紋便對於該索引進行匹配，以獲得一群候選指紋。接著該測試指紋便與該候選指紋匹配，藉此避免需要匹配該測試指紋與該資料中的每一指紋。

在另一實施例中，使用一邊緣偵測演算法決定一被分析聲音訊框或片段的精確邊緣。在某些應用中，特別是當聲音樣本只在該完整樣本的短期間內具有差異時，知道該聲音樣本之中被分析聲音訊框的邊緣位置是重要的。該邊緣偵測演算法使用線性迴歸技術識別一聲音訊框的邊緣。

該指紋技術的應用實施例很多，其包含聲音串流與其他內聲音內容(例如，串流媒體、無線電、廣告、網際網路廣播、CD 中的歌曲、MP3 檔案或其他的聲音內容形式)的即時識別。本發明實施例也提供效率、即時的媒體內容審核及其他報告。

【實施方式】

概論

本發明實施例能夠從一聲音樣本擷取特徵資訊(例如，聲音指紋)，也可以利用所擷取之特徵資訊匹配或識別該聲音。如第 1 圖描述，從一聲音樣本 100 所取得的聲音訊框 105 被輸入至一指紋擷取演算法 110 之中。該聲音樣本 100 可以由任何多種來源所提供。利用該聲音訊框 105 序列，該指紋擷取演算法 110 產生一或多個聲音指紋 115，其為該序列的特徵。做為一種辨別識別符，該聲音指紋 115 提供有關於該聲音樣本 100 訊框 105 序列之唯一性或其他特徵的資訊。實際上，代表該聲音樣本 100 的一或多個指紋 115 可以使該聲音樣本 100 被唯一識別。以下則詳細描述該指紋擷取演算法 110 的實施例。

在產生該指紋之後，所擷取之指紋 115 便可以被進一步處理或儲存在一媒體中稍後使用。例如，可以由一指紋匹配演算法 120 使用該指紋 115，其將該指紋 115 與一指紋資料庫 125(例如，來自已知來源的聲音指紋集合)中的項目相比，以決定該聲音樣本 100 的唯一性。以下也描述使用該指紋的各種方法。

該聲音樣本 100 可以根據該指紋擷取系統的應用來自於各種不同來源。在一實施例中，該聲音樣本 100 是從來自一媒體廣播器所接收的廣播所取樣並數位化。替代的，一媒體廣播器可以以數位形式傳輸該聲音，避免將其數位化的需要。媒體廣播器的形式包含，無線電傳輸器、衛星傳輸器以及纜線操作器，但並不限制於此。因此該指紋擷取系統可以用來審核這些廣播器，以決定在何時廣播何者聲音。這可以達成一種確保依循廣播限制、許可合約等等的自動化系統。因為該指紋擷取演算法 110 可以在不知道該廣播訊號精確開始與結束位置的情況下操作，因此其可以在

不與該媒體廣播器合作或不具備其相關知識下進行操作，以確保獨立性與不偏估的結果。

在另一實施例中，一媒體伺服器從一媒體圖書館取得聲音檔案，並透過網路(例如，網際網路)傳輸數位廣播，以供該指紋擷取演算法 110 使用。一串流網際網路無線電廣播為此結構形式的一種範例，其中媒體、廣告與其他內容被傳送至一個體或一群使用者。在這種實施例中，該指紋擷取演算法 110 與該匹配演算法 120 通常不具有任何關於在該聲音樣本 100 串流內容中所包含個別媒體之開始與結束時間的資訊；然而，這些演算法 110 及 120 並不需要此資訊以識別該串流內容。

在另一實施例中，該指紋擷取演算法 110 從一客戶端電腦接收該聲音樣本 100 或是其一訊框串 105，該客戶端電腦能夠對包含聲音檔案的一儲存裝置進行存取。該客戶端電腦從該儲存器取得一個別聲音檔案，並將該檔案傳送至該指紋擷取演算法 110，以從該檔案產生一或多個指紋 115。替代的，該客戶端電腦可以從儲存器 140 取得一批次檔案，並將其連續傳送至該指紋擷取演算法 110，以產生代表每一檔案的一組指紋。(當在此使用時，“一組”應被理解為在一群集中任意數量的項目，也可以僅有一單一項目)。該指紋擷取演算法 110 也可以由該客戶端電腦執行，或是由透過網路連接至該客戶端電腦的一遠端伺服器執行。

演算法

第 2 圖描述實作在第 1 圖中所顯示之該指紋擷取演算法 110 的一指紋擷取系統 200 實施例。該指紋擷取系統 200 包括一分析濾波器組 205，其連接至多個處理通道(每一個都包含一或多個處

理模組，在此以元件 210 及 215 標示)，接著連接至一差動編碼器 225，以產生一聲音指紋 115。該指紋擷取系統 200 係用於接收一聲音訊框 105，並由該聲音訊框 105 產生一聲音指紋。

以下將詳細描述對於每一輸入聲音訊框 105 而言，該一分析濾波器組 205 廣泛計算代表一以接收訊號跨及一頻率範圍的功率頻譜資訊。在一實施例中，每一處理通道都對應於該頻率範圍中的一頻帶，其中該頻帶可以重疊。據此，該通道劃分由該指紋擷取系統 200 所實作的處理動作，因此每一通道實作一對應頻道的處理。在另一實施例中，每一處理通道都處理多個頻帶(也就是多個頻帶皆與每一處理通道關聯)。在其他實施例中，用於多個頻道的處理可以由一單一模組於一單一通道中實作，或該處理可以以任何其他配置劃分為適合於該系統之應用與技術限制。

該分析濾波器組 205 接收一聲音訊框 105(像是第 1 圖描述中來自該聲音樣本 100 的訊框 105)。該分析濾波器組 205 將該聲音訊框 105 從該時間域轉換至該頻率域中，以計算代表該訊框 105 跨及一頻率範圍的功率頻譜資訊。在一實施例中，代表該訊號於大約 250 至 2250 Hz 範圍的功率頻譜被區分至多個頻帶之中(例如，Y 個頻帶，Y=13)。該頻帶可以具有一種線性或對數中間頻率分布(或任何其他尺度)，也可以重疊。該分析濾波器組 205 的輸出包含代表該多個頻帶每一個之該訊號能量的測量。在一實施例中，該平均能量的測量則利用該頻帶中平均頻譜能量的三次方根所取得。

根據該系統的軟體及硬體需求與限制，該分析濾波器組 205 也可能具有各種實作。在一實施例中，該分析濾波器組 205 包括多個帶通濾波器，其為該每一個頻帶隔離該聲音訊框 105 的訊號，

並接著進行能量估計及縮減取樣。在一實施例中，通過每一帶通濾波器的頻率可能隨時間改變。在另一實施例中，通過每一帶通濾波器的頻率則為固定(換言之並不隨時間改變)。第 7A 圖描述該實施例之一圖形化表現，其中通過每一帶通濾波器的頻率並不隨時間改變。圖式 702 中的每一個矩形都代表由一帶通濾波器所輸出之該聲音訊框 105 的訊號。另一方面，第 7B 圖描述該實施例之一圖形化表現，其中通過每一帶通濾波器的頻率則隨時間改變。如可在該圖式 704 中所見，在此範例中，通過每一帶通濾波器的頻率隨時間減少。在其他實施例中，通過的頻率隨時間增加。在對該聲音訊框 105 應用該帶通濾波器之後，每一頻帶都包含由其對應帶通濾波器所輸出的訊號。

在另一實施例中，該分析濾波器組 205 則利用一種短時快速傅立葉轉換(FFT)所實作。例如，於 8k Hz 所取樣的聲音 100 被劃分成為 64 ms 訊框 105(換言之，512 個樣本)。接著利用 Han 窗口計算兩聲音訊框 105 所組成之每一個 50%重疊部分的功率頻譜(換言之，1024 個樣本)，並實作快速傅立葉轉換，接著利用 M 個均一或對數間隔重疊三角窗口進行頻帶濾波。

可以使用各種時間頻率域轉換方式取代上述的快速傅立葉轉換。例如，可以使用改進離散餘弦轉換(MDCT)。改進離散餘弦轉換的優點之一是低複雜度，其可以只使用 $n/4$ 個快速傅立葉轉換點，以及該樣本的某些前後旋轉部分。據此，預期以改進離散餘弦轉換實作之該分析濾波器組 205 將比利用快速傅立葉轉換者為佳，例如，其具有兩倍計算轉換速度。

在另一實施例中，該分析濾波器組 205 則利用 MP3 混成濾波器組所實作，其包含一串聯多相位濾波器與一改進離散餘弦轉

換，並接著進行別名去除。該 MP3 濾波器組產生 576 個頻譜係數，其代表由 576 個樣本所組成之聲音的每一訊框 105。對於在 8k Hz 取樣的聲音而言，所形成的訊框率為 13.8 fps，而以上所述之 1024 點快速傅立葉轉換濾波器組為 15.626 fps。在該時間頻率分析期間，當進行資料再取樣時，該訊框率的差異將被凸顯，如以下所討論。該分析濾波器組 205 也可以利用正交鏡像濾波器(QMF)所實作。該 MP3 混成濾波器組的第一階段使用一種具有 32 個等寬度頻帶的正交鏡像濾波器。據此，11,025 Hz 聲音訊號的 250 至 2250 Hz 頻率範圍便可因此被劃分為 13 個頻帶。

該 MP3 濾波器組的優點之一為具有可攜性。對於不同的 CPUs 而言有存在高度最佳化的 MP3 濾波器組實作。據此，該指紋產生工作流程便可輕易的與該 MP3 編碼器整合，其可以從該 MP3 濾波器組獲得頻譜係數，而無須其他的處理。據此，該指紋產生工作流程可輕易的與該 MP3 解碼器整合，而其可以直接從一 MP3 位元串流獲得頻譜資料，而無須對其進行完整解碼。也可以與其他的聲音編碼方式整合。

一旦決定該次頻帶樣本之後，便對該次頻帶樣本進行緩衝並將其提供至一或多個再取樣器 210。該再取樣器 210 接收該次頻帶樣本，並對該次頻帶樣本進行再取樣，以產生再取樣序列。在一實施例中，該再取樣器 210 根據一種非均一次序對該次頻帶樣本進行再取樣，像是利用一種非連續或一種與該樣本被取樣之次序相反的順序。

在一實施例中，每一再取樣器 210 都對應於該 Y 個頻帶之一，並為該對應頻帶接收在時間上線性間隔的 S 個樣本序列(例如，S 是從 64 至 80 之間選擇，其與開濾波器組實作有關)。在一實施例

中，當接收該次頻帶樣本序列時，每一再取樣器都對其個別的次頻帶樣本序列執行對數再取樣、縮放取樣或偏移取樣。該再取樣的結果為該再取樣器 210 將產生代表每一聲音訊框的 M 個再取樣序列。

在一實施例中，對數再取樣包括一再取樣器 210 將其對應次頻帶樣本進行對數映射以產生在時間中以空間間隔之 T 個樣本(例如， $T=64$)的一再取樣序列。取代對數取樣方式，也可以執行其他的非線性取樣形式，像是一種指數再取樣。

在一實施例中，縮放再取樣包括一再取樣器 210 將其對應次頻帶樣本序列進行時間中的尺寸(換言之，其長度)的縮放。該次頻帶樣本序列則根據該中間頻率及/或該頻帶的頻率範圍進行縮放。例如，該縮放方式可以在較高的次頻帶中間頻率處，進行較大的次頻帶樣本序列尺寸縮放。做為另一範例，該縮放可以在較高的次頻帶中間頻率處，進行較小的次頻帶樣本序列尺寸縮放。該已縮放次頻帶樣本序列則由該再取樣器 210 進行再取樣，以產生具有 T 個樣本的一再取樣序列。

在一實施例中，偏移再取樣包括一再取樣器 210 將其對應次頻帶樣本序列進行時間中的偏移(也就是平移)。該次頻帶序列的偏移方式則根據該再取樣器頻帶的中間頻率及/或頻率範圍。例如，在較高的次頻帶中間頻率處可以具有較大的次頻帶樣本序列時間偏移。該已偏移次頻帶樣本序列則由該再取樣器 210 進行再取樣，以產生具有 T 個樣本的一再取樣序列。

在另一實施例中，每一再取樣器 210 都對應於多個頻帶。每一再取樣器 210 都接收多數頻帶的次頻帶樣本序列。由每一再取

樣器 210 所接收的次頻帶樣本序列數量則根據該實作所不同。在一實施例中，對應於每一再取樣器 210 的頻帶為連續的。

每一再取樣器 210 都對其對應次頻帶序列執行時間頻率再取樣。時間頻率再取樣包括一再取樣器 210 隨時間改變從不同的對應頻帶進行取樣，以產生具有 T 個樣本的一再取樣序列。在一實施例中，該再取樣器 210 所取樣的頻率隨著時間增加而減少。在另一實施例中，該再取樣器 210 所取樣的頻率則隨著時間增加而增加。該再取樣的結果為該再取樣器 210 將產生代表每一聲音訊框的 M 個再取樣序列。

第 8A 圖與第 8B 圖描述根據一實施例的時間頻率再取樣。在第 8A 圖中，圖式 802 的每一灰色外框矩形都代表一不同的頻帶(也就是一頻帶的樣本序列)。每一對角黑線都代表由一再取樣器 210 所產生的一再取樣序列，其為該時間頻率再取樣的結果。如在圖式 802 中所見，為了產生一再取樣序列，每一再取樣器 210 都隨時間改變對不同的對應頻帶進行取樣。在圖式 802 的實施例中，該再取樣器 210 所取樣的頻率隨時間增加而減少。第 8B 圖的圖式 804 則為第 8A 圖再取樣序列不繪製該頻帶時的描繪。

第 8C 圖的圖式 806 描述在該實施例中由該再取樣器 210 所產生的再取樣序列，其中每一再取樣器 210 對應於該 Y 個頻帶之一。與第 8A 圖相同，圖式 806 的每一灰色外框矩形都代表一不同的頻帶，而在一矩形中央處的每一黑線則代表一再取樣序列。如同在第 8C 圖中所見，在此實施例中由該再取樣器 210 所產生的再取樣序列數量與該頻帶的數量相同(也就是 $M=Y$)。這是因為每一再取樣器 210 都在其頻帶之中進行取樣。

然而，如在第 8A 圖中所見，在每一再取樣器 210 都對應於多個頻帶並執行時間頻率再取樣的實施例中，該再取樣序列的數量少於該頻帶的數量(也就是 $M < Y$)。在此實施例中需要更多的頻帶以確保每一再取樣器 210 都可以在相同時間期間獲得樣本，且每一再取樣序列都包含 T 個樣本。

在由該再取樣器 210 進行再取樣且產生 M 個再取樣序列之後，該再取樣序列便儲存為一 $[M \times T]$ 矩陣，其對應於具有一時間(水平軸)與一頻率(垂直)軸的已取樣頻譜。該 M 個再取樣序列也提供至一或多個轉換模組 215，以對該樣本進行轉換。

在一實施例中，對代表每一頻帶之樣本所執行的轉換是一種 T 點轉換，這是一種沿著時間軸(例如，該 $[M \times T]$ 矩陣的每一列)進行的轉換。在一實施例中，該 T 點轉換為一種 T 點快速傅立葉轉換。從該快速傅立葉轉換所形成的係數列則稱為特徵向量。在一實施例中，代表每一頻帶的特徵向量都包括為該頻帶所計算之快速傅立葉轉換的每間隔係數，其具有往上升的頻率。據此，每一特徵向量都包含 N 個係數(例如，其中 $N = T/2 = 32$)。在另一實施例中，取代 T 點快速傅立葉轉換，可以執行一種 T 點離散餘弦轉換(DCT)、 T 點離散哈特利轉換(DHT)或離散小波轉換(DWT)。所形成的特徵向量則提供至一差動編碼器 225。

在另一實施例中，於 T 點轉換之後進行一 M 點轉換。 T 點轉換則如上述對代表每一頻帶的樣本進行，在 T 點轉換之後，代表每一頻帶的樣本將在強度上進行縮放、視窗化及正規化。在所述縮放、視窗化及正規化之後，對該樣本進行一 M 點轉換，這是一種沿著頻率軸(例如，該 $[M \times T]$ 矩陣的每一欄)進行的轉換。在一實施例中，該 M 點轉換為沿著頻率軸進行的快速傅立葉轉換、離散

餘弦轉換、哈特利轉換或離散小波轉換。所形成的特徵向量則提供至一差動編碼器 225。

在另一實施例中，該轉換為一種二維離散餘弦轉換(2D DCT)。為了執行所述轉換，將代表每一頻帶的樣本正規化。一旦該樣本被正規化，便沿著時間軸進行一維離散餘弦轉換。在沿著時間軸進行一維離散餘弦轉換之後，沿著頻率軸進行一維離散餘弦轉換。所形成的特徵向量則提供至一差動編碼器 225。

該差動編碼器 225 產生代表該聲音樣本的指紋 115。在一實施例中，該差動編碼器 225 減去對應於每一鄰近頻帶對的特徵向量。如果具有 Y 個頻帶，便具有 Y-1 個鄰近頻帶對。兩特徵向量相減得到一具有 N 個差動值的向量。對於這些差動值的每一個而言，如果其差異大於等於 0 該差動編碼器 225 便選擇 1 的數值，如果其差異小於 0 該差動編碼器 225 便選擇 0 的數值。對於該序列中每一具有四個位元的群集而言，該編碼器根據一編碼表指定一位元數值。該最佳編碼數值則在該指紋擷取演算法的微調及訓練階段所計算。為該每一連續頻帶對的特徵向量重複此程序則形成一 $[(Y-1) \times N/4]$ 的位元矩陣。此矩陣以一線性位元序列所表示，則做為該聲音指紋 115。在 $Y=13$ 且 $N=8$ 的範例中，該指紋 115 具有 12 位元組的資訊。

在一實施例中，在特徵向量被量化之前利用該主成分分析以解除關聯性，並減少所獲得特徵向量的大小。也可以額外或替代的使用其他的解除關聯技術，像是數位餘弦轉換，以消除特徵向量的冗贅性並將其壓縮。

在一實施例中，該指紋擷取系統 200 為一特定聲音訊號中的

高度重疊聲音訊框序列產生多個指紋。在一範例中，由該系統 200 所處理的每一訊框 105 序列包含該聲音訊號 3 秒鐘，並在一先前序列開始之後啟動 64 微秒。在此方法中，為了多個每 64 微秒開始的聲音訊號 3 秒部分產生一指紋。為了實作所述結構，該指紋擷取系統 200 於該分析濾波器組 205 之前與之後可以具有緩衝記憶體，其中該緩衝在接收到次一聲音訊框 105 之後，便以該聲音訊號的次一 64 微秒進行更新。

第 6A 圖與第 6B 圖描述根據本發明多數替代實施例之一指紋擷取演算法 110 的圖形化表現。在此程序中，該分析濾波器組 205 接收一聲音訊框。圖式 602 為所接收聲音訊框於時間域中的表示。該分析濾波器組 205 對該聲音訊框執行快速傅立葉轉換，以將其從時間域轉換至頻率域，如圖式 604 所顯示。接著當該訊框於頻率域中時，計算代表該訊框的功率頻譜資訊。該分析濾波器組 205 應用多個帶通濾波器，以將代表每一頻帶的訊框訊號隔離，如圖式 606 所描述。

由該再取樣器 210 對該頻帶的次頻帶樣本序列進行再取樣。第 6B 圖顯示四種替代技術(以 A、B、C 及 D 標示)，其可以由該再取樣器 210 所執行以對該次頻帶樣本序列進行再取樣。在一實施例中，技術 A、B 及 C 為當每一再取樣器 210 對應於一頻帶時所實作的技術。在一實施例中，技術 D 可以於每一再取樣器 210 對應於多個頻帶，且該再取樣器 210 是用於執行時間頻率再取樣時所實作。

在技術 A 中，每一再取樣器 210 對其對應次頻帶樣本序列進行對數取樣(圖式 608)，以產生具有 T 個樣本的再取樣序列(圖式 616)。在技術 B 中，每一再取樣器 210 根據該次頻帶的中間頻率

及/或頻率範圍，對其對應次頻帶樣本序列進行尺寸縮放。如圖式 610 所顯示，在此範例中較高的次頻帶中間頻率與較寬的次頻帶頻率範圍，具有較小的次頻帶取樣序列尺寸。對該已縮放次頻帶樣本序列進行再取樣，以產生具有 T 個樣本的再取樣序列(圖式 618)。

在技術 C 中，每一再取樣器 210 根據該次頻帶的中間頻率及/或頻率範圍，於時間中偏移其對應次頻帶樣本序列。如圖式 612 所顯示，在此範例中較高的次頻帶中間頻率具有較大的次頻帶取樣序列偏移。對該已偏移次頻帶樣本序列進行再取樣，以產生具有 T 個樣本的再取樣序列(圖式 620)。

在技術 D 中，每一再取樣器 210 對其對應次頻帶樣本序列進行時間頻率再取樣。時間頻率再取樣係由一再取樣器 210 所實作，其隨時間改變從不同對應頻帶進行取樣。如圖式 614 所顯示，在此範例中該再取樣器 210 所取樣的頻率隨時間增加而減少。所述再取樣將產生具有 T 個樣本的再取樣序列(圖式 622)。

由該再取樣器 210 所產生的再取樣序列(M 個再取樣序列)則以 $[M \times T]$ 矩陣儲存。每一轉換模組 215 都對由其對應再取樣器 210(也就是如轉換模組 215 之相同通道中的再取樣器 210)所產生的再取樣序列執行轉換。第 6 圖描述三種替代技術(以 E、F 及 G 標示)，其可以由該轉換模組 215 實作以轉換該再取樣序列及產生一特徵向量。

在技術 E 中，該轉換模組 215 執行一種 T 點轉換，如在圖式 624 中所描述。在技術 F 中，該轉換模組 215 執行一種一維的 T 點轉換，接著執行另一維度的 M 點轉換，如在圖式 626 中所描述。在技術 G 中，該轉換模組 215 執行一種二維離散餘弦轉換或其他

適合的二維轉換，如在圖式 628 中所描述。

一旦轉換該次頻帶樣本而獲得特徵向量，該差動編碼器 225 便使用由該轉換模組 215 所產生的特徵向量以產生一指紋 115。

第 5 圖為一指紋擷取系統 200 的範例，其中該再取樣器 210 為對數再取樣器 210，該轉換模組 215 為 T 點轉換模組 215。該對數再取樣器 210 如以上所述(技術 A)實作對數取樣。然而，應該了解在其他實施例中可將該對數再取樣器 210 交換為實作其他再取樣技術(也就是技術 B、C 或 D)的再取樣器 210。

該 T 點轉換模組 215 如以上所述實作 T 點轉換(技術 E)。然而，應該了解在其他實施例中可將該 T 點轉換模組 215 交換為實作其他轉換技術(也就是技術 F 或 G)的轉換模組 215。

聲學模型

在該指紋擷取系統的各种應用中，因為用於該聲音樣本的一編碼處理移除該頻帶或是因為其他理由，使某些頻帶無法被感知而不顯著。因此在一實施例中，使用一聲學模型 235 以識別並標記一特定指紋的不顯著頻帶。聲學模型，像是心理聲學模型是在各種聲音處理領域中所熟知。可以在一指紋 115 建立與儲存於該資料庫 125 的期間，為了高品質參考樣本而計算代表該聲學模型 235 的一組模型參數。在該指紋 115 中的不顯著頻帶可以利用指定一特殊碼或將其對應數值(也就是位元值)定義為零的方式所標記。這可使該頻帶在任何後續匹配處理中被有效忽略，因此在一指紋與該資料庫紀錄匹配的程序中，只有具有非零數值的對應頻帶對被用來分辨該指紋 115。已遮罩頻帶(也就是具有數值零的頻帶)則從比較中一起排除。

在一實施例中，該聲學模型是一種代表人類聽覺系統的心理聲學模型。因為該指紋擷取系統為一種以人類聽覺系統為目標進行聲音是別的目的，因此所述模型相當有用。所述聲音可以利用一或多個知覺編碼器所壓縮，移除不相關的聲音資訊。使用人類心理聲學模型可從該指紋識別不相關的頻帶並將其排除。

但是該心理聲學模型只是一種適合於人類知覺編碼聲音的聲學模式形式之一。另一種聲學模式則用於消除一特定錄音裝置的性質。對於所述錄音裝置的每一頻帶而言，聲學模型可以根據其重要性指定該頻帶一權重因子。然而，另一種聲學模式則用於消除特定環境的性質，像是在車輛或房間中所發生的背景噪音。在所述實施例中，代表該聲學模型的每一頻帶都可以具有根據該系統所設計之環境中，該頻帶之重要性而對其指定一權重因子。

在一實施例中，該聲學模型 235 與該濾波器組 205 的參數則與該已分析聲音訊號 100 的形式與性質有關。包括一組次頻帶權重因子與多數濾波器頻帶的相異曲線以及其頻率分布則用於獲得該目標聲音訊號性質的一較佳匹配。例如對於類語音聲音而言，該訊號的功率主要集中於低頻帶，而音樂根據其類型則可能具有較高的相關成分。在一實施例中，該聲學模型的參數則從該參考聲音訊號所計算，並與其產生的指紋一起儲存於內容資料庫中。在另一實施例中，該聲學模型的參數則根據在匹配程序期間的已分析聲音訊號性質動態計算。

據此，該聲學模型 235 的可能應用包含微調代表特定環境及/或錄音裝置的聲音辨識參數，以及演算法性質編碼。例如，在已知該手機電話聲音路徑(麥克風特性、聲音處理及壓縮演算法等等)的聲學性質時，可使一聲學模型開發時便消除這些性質。在指紋

比較期間使用此模型可以明顯提高所產生指紋之匹配程序的強健性。

指紋索引及匹配

在一實施例中，一指紋索引編輯器 230 為每一指紋 115 產生一索引。接著該指紋 115 被儲存於該指紋資料庫 125 中，以進行該指紋資料庫 125 內容的效率搜尋及匹配。在一實施例中，依指紋 115 的索引包括該指紋 115 的一部分或其雜湊訊號。據此，在該指紋資料庫 125 中的指紋 115 便根據關於該指紋之有用的識別資訊進行索引編輯。

在以上所述每一指紋 115 都包括一 $[(Y-1) \times N/2]$ 位元矩陣的實施例中，該索引編輯器 230 使用該最左側欄位的位元作為索引。在每一指紋 115 都為 12×8 位元矩陣的範例中，代表該指紋 115 的索引可以是最左側兩欄的位元(總共為 24 位元)。在此方法中，作為代表每一指紋 115 之索引的位元為一種基於用於計算該指紋 115 之特徵向量的低頻頻譜係數之該指紋 115 的一子集合。因此這些位元對應於該再取樣及轉換頻譜頻帶的低頻成分，其較為穩定且對調變噪音及失真較不敏感。因此，有很高的可能性使得相似的指紋具有相同的索引數值。在此方法中，該索引可以用來標示相似的指紋進行標示並群集，並可能與該資料庫中的指紋進行匹配。

第 3 圖描述一種根據本發明一實施例，利用以上所述索引將一測試指紋與該指紋資料庫 125 進行匹配的方法。為了找出一測試指紋與該指紋資料庫 125 之間的相符結果，該匹配演算法從計算以上所述之代表該測試指紋的一索引數值開始(310)。利用此索引數值，可獲得一候選指紋群集(320)，例如該群集包含在該資料

庫 125 中具有相同索引數值的所有指紋。如以上所說明，因為計算該索引數值的方式，很可能與該資料庫 125 之間的相符結果係落於此候選指紋群集之中。

為了測試在該候選指紋群集中的任何相符結果，計算該測試指紋與每一候選指紋之間的位元錯誤率(BER) (330)。兩指紋之間的位元錯誤率為其對應位元但不相符的百分比。對於完全不相關的隨機指紋而言，預期該位元錯誤率為 50%。在一實施例中，兩指紋若相符合則其位元錯誤率大約低於 35%；然而，根據對於主動錯誤及/或被動錯誤的要求容忍度，也可以使用其他的數值限制。此外，可以使用其他不同於位元錯誤率的計算或條件，以進行兩指紋比較。例如，也可以使用位元錯誤率的相反測量，也就是相符率。此外，在進行兩指紋比較時對於某些位元對於其他位元可以具有較高的權重。

如果不存在預定匹配條件之中的任何相符結果(340)，或是如果不存在需要修改的索引(350)，該匹配演算法便無法在該資料庫 125 中找到任何與該測試指紋的相符結果。接著該系統可以繼續搜尋(例如，使用較小的限制條件以獲得候選指紋)或可以終止。如果存在一或多個相符指紋(340)，便回傳該相符指紋清單(360)。

在一實施例中，該系統可以在修正該已計算指紋索引之後，重複如以上描述的搜尋動作，以從該搜尋獲得相符的一相異候選指紋組(370)。為了修正該已計算指紋索引(370)，可以反轉該已計算指紋索引的一或多個位元。在該指紋索引具有 24 個位元的範例中，在無法利用原始指紋索引找到相符結果之後，便以每次翻轉該 24 個位元指紋索引一相異單一位元的方式，重複該搜尋步驟 24 次。也可以使用其他各種技術擴大該搜尋空間。

在一實施例中，該指紋索引編輯器 230 利用根據由該聲學模型 235 所計算並事先儲存在該資料庫 125 中的頻帶權重因子組從一或多個指紋選擇索引位元方式，產生一或多個索引。當使用多個索引時，包含使用利用位元反轉所獲得的索引，該候選指紋群集便包含對於每一計算索引所獲得所有候選指紋。

在另一實施例中，可以利用預先篩選窄化該搜尋範圍，並只選擇對於每一計算索引所獲得在大多數或所有候選群集之中所找到的候選指紋。所述利用多數索引對該多數候選指紋進行預先篩選的方式，包含針對利用位元反轉所獲得的索引進行，可以顯著改善該資料庫搜尋的效能。在一實施例中，將索引以及可能的候選指紋參照儲存在一電腦記憶體中，以進行該候選指紋的快速選擇與預先篩選。在該第二步驟(步驟 320)中，只有具有與該已知指紋之間具有最高可能相符結果的候選指紋被載入至電腦記憶體中並進行比較。此方法藉由只在電腦記憶體中保存小量索引的方式形成快速搜尋的結果，而在慢速裝置(例如硬碟或網路硬碟)儲存較大量的指紋。

聲音訊框邊緣偵測

在某些實施例中，可能需要偵測一相符聲音片段的邊緣。邊緣偵測使該系統可以預先知道即時於何處產生一特定相符聲音片段。根據該已分析聲音的品質，該邊緣偵測演算法的實施例可以偵測大約 0.1 至 0.5 秒精確性的一相符聲音片對邊緣。

如以上所描述，該指紋擷取技術的實施例累加次頻帶處理緩衝之中的聲音樣本。因為此緩衝效果，便造成指紋擷取演算法的輸出延遲並使聲音片段邊緣模糊。此現象於第 4 圖中描述，其為

代表聲音片段之一參考指紋與代表一來源樣本聲音串流隨時監所產生之一指紋序列之間，隨時間變化的位元錯誤率。在所描述實施例中，該次頻帶緩衝保有 3 秒的聲音，並在兩指紋具有小於 35% 的位元錯誤率時認為其之間相符。

首先在 T_0 時，該次頻帶處理緩衝為空白，而該已產生指紋因此與該原始聲音之間不具有任何相符結果(也就是預期該位元錯誤率近似等於 50%)。當增加聲音樣本至該次頻帶處理緩衝時，該位元錯誤率便降低，意指形成一較佳匹配。隨著有效的時間經過，該位元錯誤率於時間 T_1 時下降至門檻 35% 以下，其意指形成一相符結果。最後在時間 T_2 ，在該緩衝填滿樣本時該位元錯誤率便達到一種平坦變化。當在時間 T_3 該相符聲音片段的尾端通過該指紋擷取演算法時，其開始產生較差匹配指紋且該位元錯誤率開始增加，在時間 T_4 時達到辨識門檻 35%。獲得相符曲線的期間(T_1 - T_4)與該平坦階段的期間(T_2 - T_3)都小於該相符聲音片段的期間(T_0 - T_3)。

在一實施例中，使用一邊緣偵測演算法決定一相符聲音訊框或片段的精確邊緣。可獲得如第 4 圖描述的一位元錯誤率曲線。該位元錯誤率曲線被劃分為小區域，其對應於具有下降位元錯誤率的匹配開始階段(例如， T_1 - T_2)，具有大約固定位元錯誤率的平坦階段(例如， T_2 - T_3)以及具有上升為原錯誤率的匹配結束階段(例如， T_3 - T_4)。因為一真實位元錯誤率曲線的產生將帶有雜訊，因此可利用像是迴歸分析的適當技術進行劃分。在一實施例中，忽略所有產生大於 35% 以上位元錯誤率的樣本，因為其可能並不可信。接著可以利用線性迴歸法計算該聲音片段的匹配開始(也就是時間 T_1)，其以最佳方式匹配找出下降錯誤位元率區域(例如，

T1-T2)的一迴歸直線，而與一具有對應於 50% 位元錯誤率之水平直線相交。同樣的方法也可用於估計 T5，其以最佳方式找出與上升位元錯誤率區域(例如，T3-T4)與對應 50%位元錯誤率之水平直線的交叉位置。然而在此情況中，T5 對應於因為該次頻帶緩衝期間 B 所延遲之片段的尾端，並非該相符聲音片段的真實尾端。該片段的尾端位置(也就是時間 T3)變利用所獲得之估計時間 T5 減去該次頻帶緩衝期間 B 的方式所計算。

在另一實施例中，估計該相符聲音片段的尾端為該區域 T2-T3 的尾端，且該聲音片段的開始則利用從時間 T2 減去該次頻帶緩衝期間 B 的方式計算，T2 則對應於該區域 T2-T3 的開始處。

總結

雖然以上利用向量與矩陣的方式加以討論，但是為了任何指紋或次指紋所計算的資訊都可以以任何形式儲存及處理，並不限於向量或矩陣數值形式。因此所使用的向量與矩陣形式只是做為一種從一聲音樣本擷取資料的方便機制，應不意味以任何方式予以限制。此外，雖然以上以頻譜圖方式討論該功率頻譜，但要了解並不是只能利用頻譜圖表現代表一聲音訊號之功率頻譜或頻譜分析的資料，而是可以採用其他適合形式。

在一實施例中，利用一電腦程式產品實作一軟體模組，其包括包含電腦程式碼的一電腦可讀媒介，該程式碼可由電腦處理器所執行以實作任何或所有在此描述的步驟、操作或程序。據此，在此描述的任何步驟、操作或程序都可以利用一或多個軟體模組或硬體模組，以單獨或與其他裝置組合的方式所執行或實作。此外，以硬體元件方式所描述的任何系統部分也可以以軟體方式實

作，以軟體元件方式所描述的任何系統部分也可以以硬體方式實作，像是在一專用電路中的固定編碼。例如，用於實作所述方法的編碼可以被嵌入至一硬體裝置之中，像是一專用積體電路或其他客製化電路。這使的本發明具有與其他不同裝置相整合的優點。

在另一實施例中，該指紋擷取演算法則嵌入於各種任意聲音裝置之中，並於其上所執行，例如一行動電話、個人數位助理、MP3 撥放器及/或錄音器、機上盒、電視機、遊戲機或其他可儲存、處理或撥放聲音內容的裝置。將指紋擷取演算法嵌入於所述裝置上有許多優點。例如，直接在一行動電話上產生聲音指紋與從電話透過蜂巢式網路傳送壓縮聲音至一指紋擷取伺服器相比之下，可以提供更佳的匹配結果。在該行動電話上執行該演算法可消除因 GSM 壓縮技術所形成的失真，該壓縮技術是用於壓縮語音，但對於音樂撥放效能不佳。據此，此方法可以顯著改善由一行動電話所錄音的聲音辨識。這也可降低伺服器與網路流量的負載。

這種嵌入方式的另一種優點是可以監測聆聽經驗，而不會違法隱私權或使用者權利。例如，一錄音裝置可以記錄聲音、建立指紋，並接著指傳送指紋至一伺服器進行分析。所記錄的聲音絕不會離開該裝置。即使無法從該指紋回復成為原始聲音，但該伺服器接著仍可利用所傳送的指紋識別標的音樂或廣告。

本發明實施例的前述敘述是針對例證的目的所陳述，並不預期完全代表本發明，或限制本發明於所公開的形式之中。相關領域技術者可以了解在上述引導下可進行許多修改與變化。因此預期本發明的觀點將不受限於此描述的詳細細節中，而是由附加申請專利範圍所定義。

【圖式簡單說明】

第 1 圖 為根據本發明一實施例，用於擷取並使用來自一聲音樣本之指紋的程序結構圖式。

第 2 圖 為根據本發明一實施例，一指紋擷取系統的結構圖式。

第 3 圖 為根據本發明一實施例，一匹配演算法的流程圖。

第 4 圖 描述根據本發明一實施例的一邊緣偵測演算法。

第 5 圖 為根據本發明一實施例，包含對數取樣與 T 點轉換模組的一指紋擷取系統結構圖式。

第 6A 圖與第 6B 圖 描述根據本發明多數替代實施例之一指紋擷取演算法的圖形化表現。

第 7A 圖與第 7B 圖 描述根據本發明一實施例，應用於一聲音訊框之帶通濾波器的圖形化表現。

第 8A 圖至第 8C 圖 描述根據本發明多數替代實施例之再取樣次頻帶樣本序列的圖形化表現。

【主要元件符號說明】

100 聲音樣本

105 聲音訊框

110 指紋擷取演算法

115 指紋

120 指紋匹配演算法

201246183

125 指紋資料庫

200 指紋擷取系統

205 分析濾波器組

210 再取樣器/對數再取樣器

215 轉換/T 點轉換

225 差動編碼器

230 指紋索引編輯器

235 聲學模型

310~370 步驟

602~628、702~704、802~806 圖式

發明專利說明書

(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： 101102127
 ※申請日： 101.1.19
 ※IPC 分類： G10L 11/00 (2006.01)
 G10L 15/02 (2006.01)
 G10L 15/10 (2006.01)

一、發明名稱：(中文/英文)

聲音訊號之特徵指紋擷取與匹配

/ EXTRACTION AND MATCHING OF CHARACTERISTIC
 FINGERPRINTS FROM AUDIO SIGNALS

二、中文發明摘要：

從一聲音樣本擷取一聲音指紋，其中該指紋包含的資訊為該樣本中內容的特徵。該指紋可以利用計算代表該聲音樣本之一能量頻譜、對該能量頻譜再取樣、轉換該已再取樣能量頻譜以產生一連串特徵向量，並利用該特徵向量之差動編碼計算該指紋的方式所產生。所產生的指紋可以與一資料庫中的一組參考指紋比較，以識別該原始聲音內容。

三、英文發明摘要：

An audio fingerprint is extracted from an audio sample, where the fingerprint contains information that is characteristic of the content in the sample. The fingerprint may be generated by computing an energy spectrum for the audio sample, resampling the energy spectrum, transforming the resampled energy spectrum to produce a series of feature vectors, and computing the fingerprint using differential coding of the feature vectors. The generated fingerprint can be compared to a set of reference fingerprints in a database to identify the original audio content.

七、申請專利範圍：

1. 一種從一聲音訊框擷取一聲音指紋的方法，該方法包括：
 - 將該聲音訊框過濾至多個頻帶之中，以產生對應的多個已過濾聲音訊號；
 - 根據該對應頻帶的頻率，於時間域中縮放每一已過濾聲音訊號的大小；
 - 對該已縮放過濾聲音訊號進行再取樣，以產生已再取樣聲音訊號；
 - 轉換該已再取樣聲音訊號，以產生代表每一已再取樣聲音訊號的一特徵向量；以及
 - 根據該特徵向量計算該聲音指紋。
2. 如申請專利範圍第 1 項的方法，其中所述縮放所根據的頻率為至少該對應頻帶之一中間頻率與一頻率範圍。
3. 如申請專利範圍第 1 項的方法，其中所述轉換一已再取樣聲音訊號包括沿著一時間軸進行轉換。
4. 如申請專利範圍第 1 項的方法，其中所述轉換一已再取樣聲音訊號包括：
 - 沿著一時間軸轉換該已再取樣聲音訊號；以及
 - 沿著一頻率軸轉換該已再取樣聲音訊號。
5. 如申請專利範圍第 4 項的方法，進一步包括在強度、窗口中進行縮放，並對該已再取樣聲音訊號進行正規化。
6. 如申請專利範圍第 1 項的方法，其中所述轉換一已再取樣聲音訊號包括執行一二維離散餘弦轉換(2D DCT)。
7. 一種從一聲音訊框擷取一聲音指紋的方法，該方法包括：
 - 將該聲音訊框過濾至多個頻帶之中，以產生對應的多個已過濾聲音訊號；
 - 根據該對應頻帶的頻率，於時間域中偏移每一已過濾聲音訊號的大小；
 - 對該已偏移過濾聲音訊號進行再取樣，以產生已再取樣聲音訊號；

轉換該已再取樣聲音訊號，以產生代表每一已再取樣聲音訊號的一特徵向量；以及

根據該特徵向量計算該聲音指紋。

8. 如申請專利範圍第 7 項的方法，其中所述偏移所根據的頻率為至少該對應頻帶之一中間頻率與一頻率範圍。

9. 如申請專利範圍第 7 項的方法，其中所述轉換一已再取樣過濾聲音訊號包括沿著一時間軸進行轉換。

10. 如申請專利範圍第 7 項的方法，其中所述轉換一已再取樣聲音訊號包括：

沿著一時間軸轉換該已再取樣聲音訊號；以及

沿著一頻率軸轉換該已再取樣聲音訊號。

11. 如申請專利範圍第 10 項的方法，進一步包括在強度、窗口中進行縮放，並對該已再取樣聲音訊號進行正規化。

12. 如申請專利範圍第 7 項的方法，其中所述轉換該已再取樣聲音訊號包括執行一二維離散餘弦轉換(2D DCT)。

13. 一種從一聲音訊框擷取一聲音指紋的方法，該方法包括：
將該聲音訊框過濾至多個頻帶之中，以產生對應的多個已過濾聲音訊號；

對該已過濾聲音訊號執行時間頻率再取樣，以產生已再取樣聲音訊號；

轉換該已再取樣聲音訊號，以產生代表每一已再取樣聲音訊號的一特徵向量；以及

根據該特徵向量計算該聲音指紋。

14. 如申請專利範圍第 13 項的方法，其中每一已過濾聲音訊號都與一或多個處理頻道關聯，而所述時間頻率再取樣係對每一處理頻道的已過濾聲音訊號進行，以產生該已再取樣聲音訊號。

15. 如申請專利範圍第 14 項的方法，其中所述對一處理頻道之已過濾聲音訊號進行時間頻率再取樣包括：

隨著時間改變從該處理頻道的相異已過濾聲音訊號進行取樣，以產生一已再取樣聲音訊號。

16. 如申請專利範圍第 13 項的方法，其中所述轉換一已再取樣過濾聲音訊號包括沿著一時間軸進行轉換。

17. 如申請專利範圍第 13 項的方法，其中所述轉換一已再取樣聲音訊號包括：

沿著一時間軸轉換該已再取樣聲音訊號；以及

沿著一頻率軸轉換該已再取樣聲音訊號。

18. 如申請專利範圍第 17 項的方法，進一步包括在強度、窗口中進行縮放，並對該已再取樣聲音訊號進行正規化。

19. 如申請專利範圍第 13 項的方法，其中所述轉換該已再取樣聲音訊號包括執行一二維離散餘弦轉換(2D DCT)。

20. 一種從一聲音訊框擷取一聲音指紋的方法，該方法包括：

將該聲音訊框過濾至多個頻帶之中，以產生對應的多個已過濾聲音訊號；

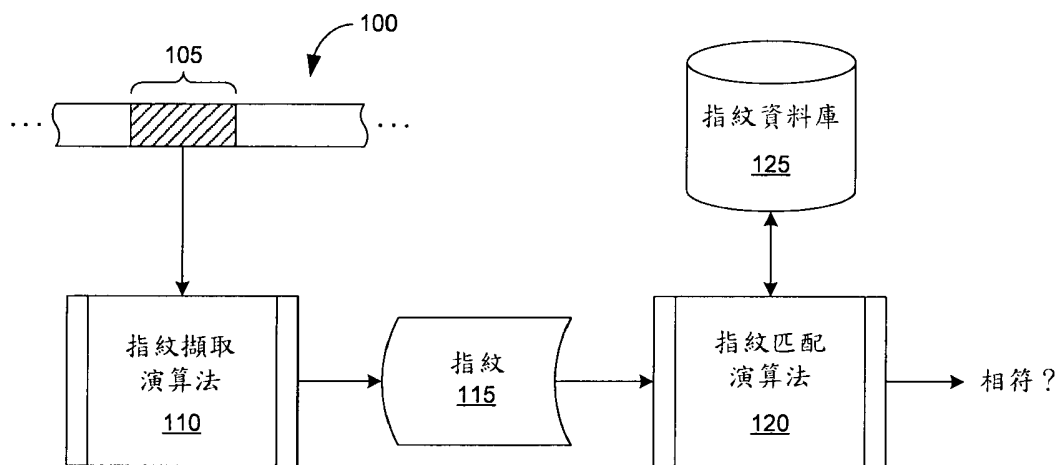
根據一非均一次序對該已過濾聲音訊號進行再取樣，以產生已再取樣聲音訊號；

轉換該已再取樣聲音訊號，以產生代表每一已再取樣聲音訊號的一特徵向量；以及

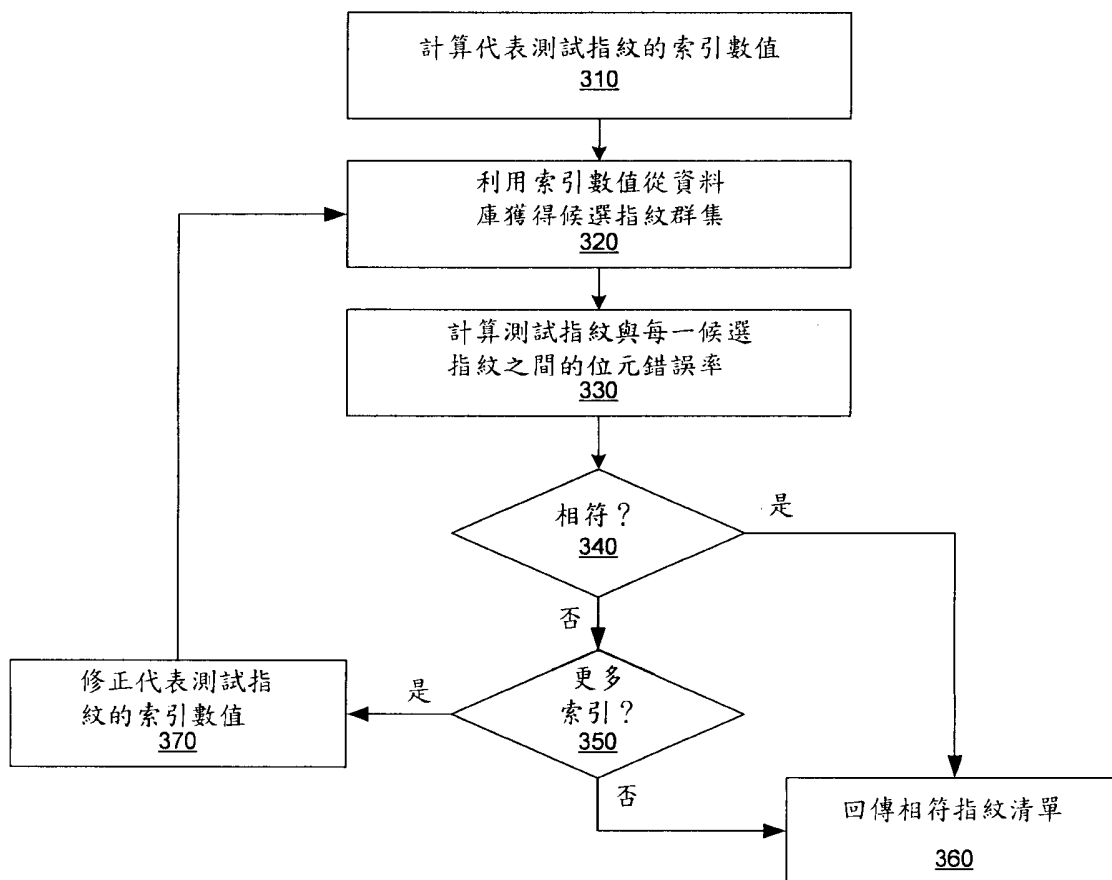
根據該特徵向量計算一聲音指紋。

21. 如申請專利範圍第 20 項的方法，其中所述根據一非均一次序對該已過濾聲音訊號進行再取樣包括以一種該已過濾聲音訊號首次被取樣之次序的一相反次序，對該已過濾聲音訊號進行再取樣。

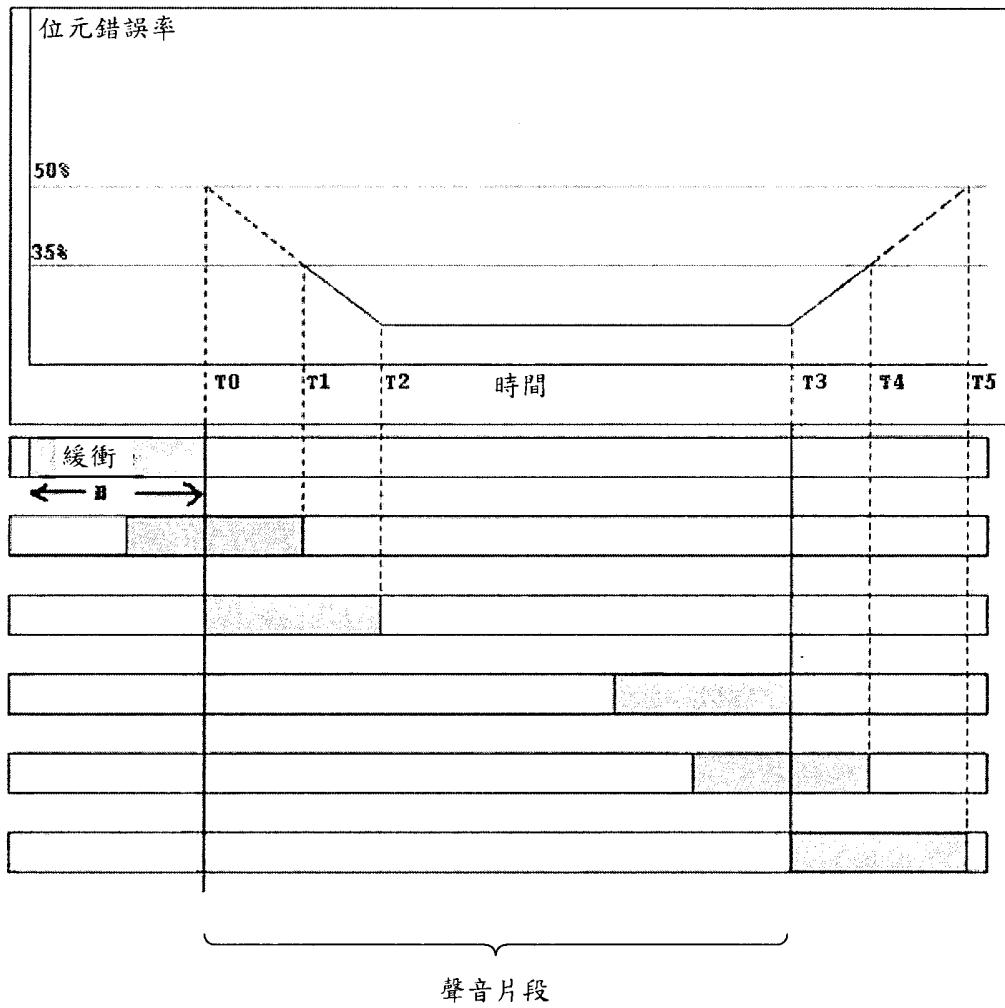
八、圖式：



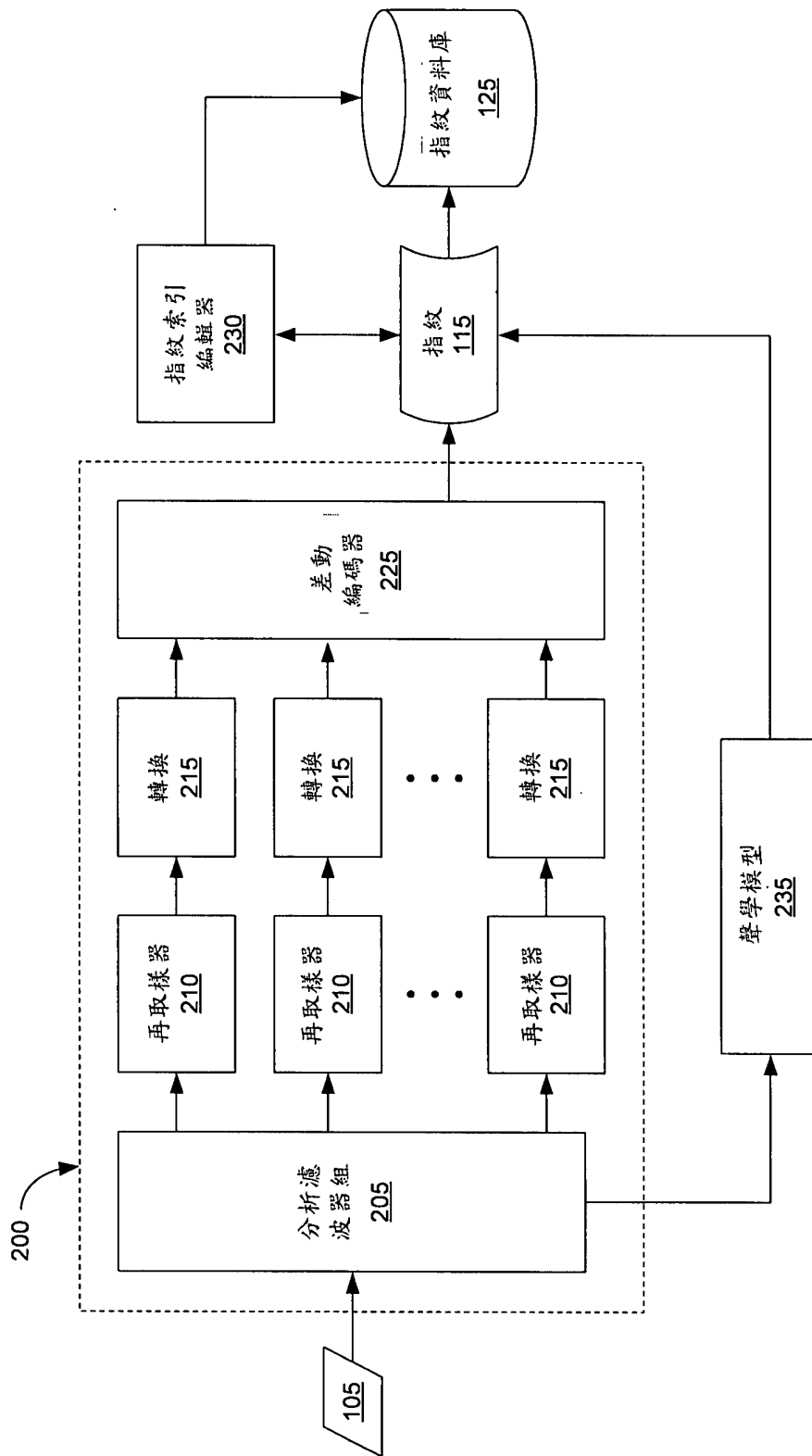
第1圖



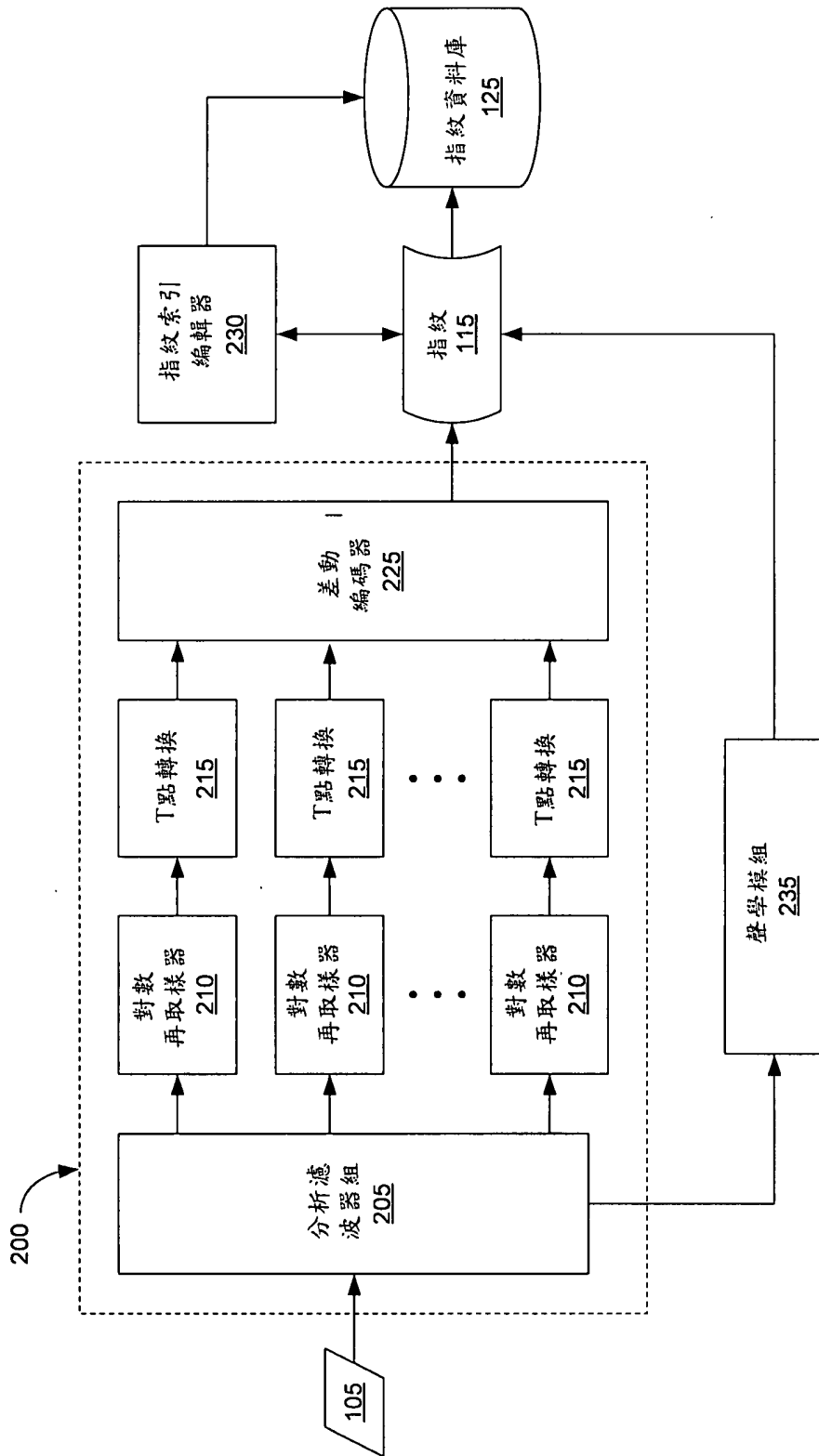
第3圖



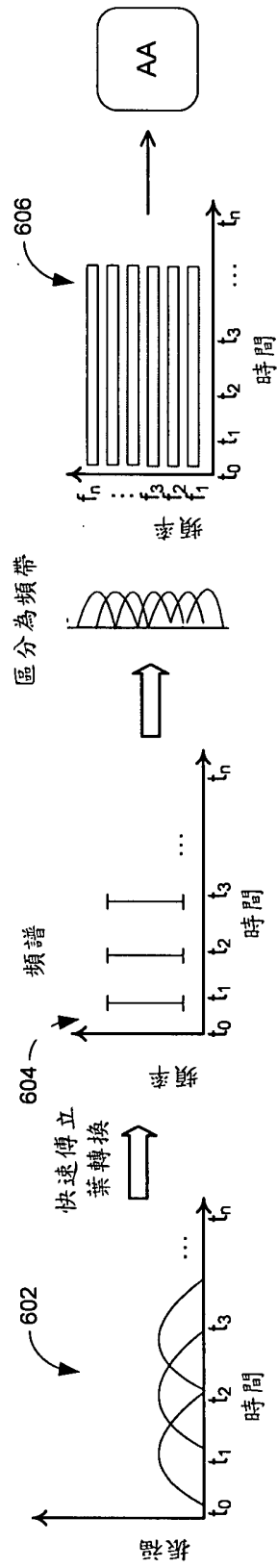
第4圖



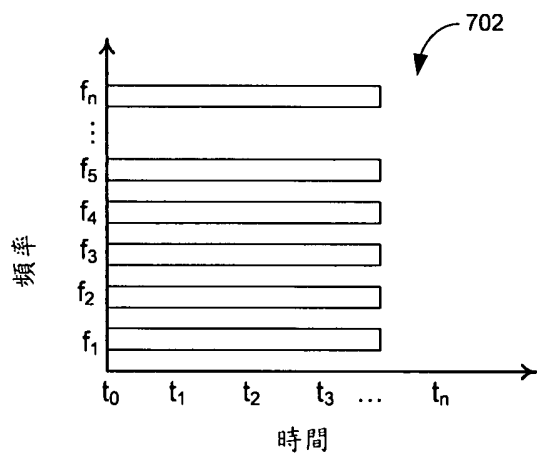
第2圖



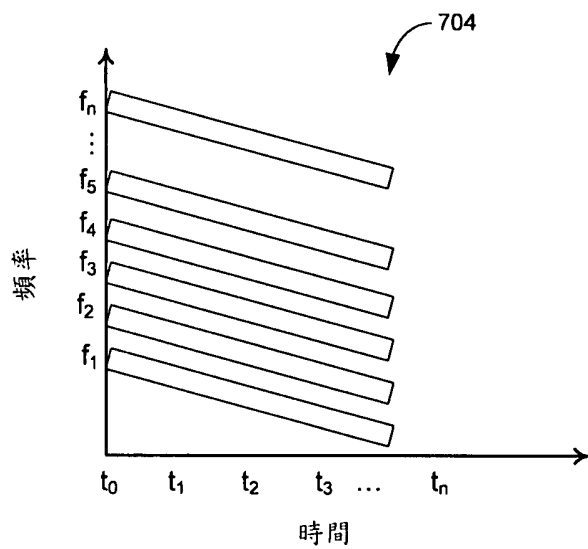
第5圖



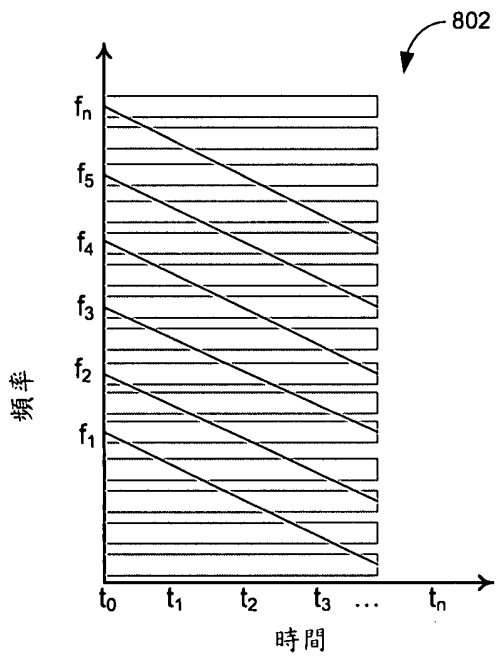
第6A圖



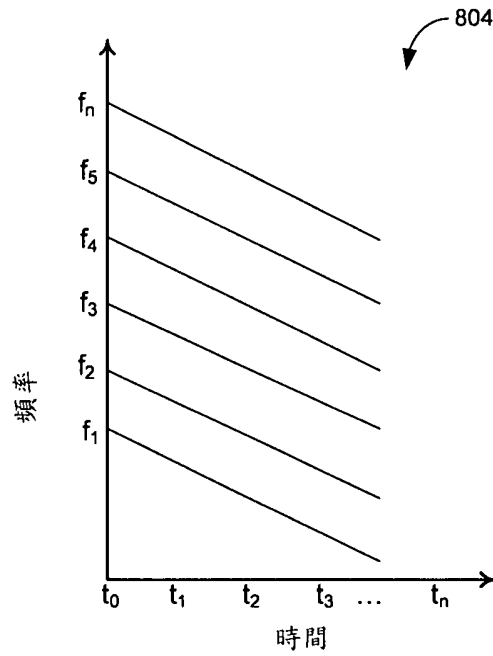
第7A圖



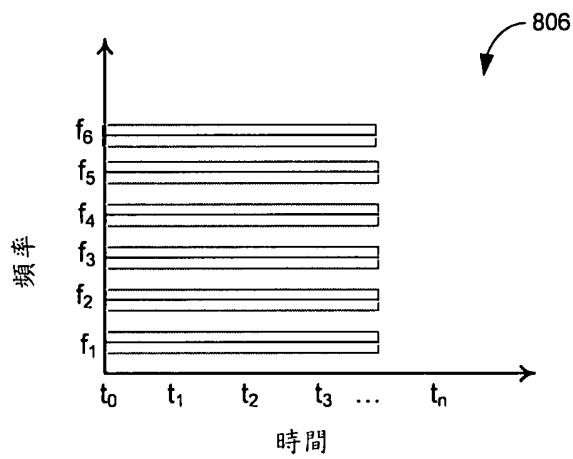
第7B圖



第8A圖



第8A圖



第8C圖

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：

100 聲音樣本

105 聲音訊框

110 指紋擷取演算法

115 指紋

120 指紋匹配演算法

125 指紋資料庫

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：