

## 發明專利說明書

200307205

(填寫本書件時請先行詳閱申請書後之申請須知，作※記號部分請勿填寫)

※申請案號：92109632 ※IPC分類：G06F 17/30  
 ※申請日期：92年04月24日

## 壹、發明名稱：

(中文) 強烈且不變音頻圖樣匹配(英文) Robust and invariant audio pattern matching貳、發明人(共2人)發明人 1姓名：(中文) 王禮俊(英文) Wang, Avery Li-Chun住居所地址：(中文) 美國加州巴洛艾托羅斯路二九二五號(英文) 2925 Ross Road, Palo Alto, CA 94303,  
U.S.A.參、申請人(共1人)申請人 1姓名或名稱：(中文) 莎琛娛樂有限公司(英文) Shazam Entertainment, Ltd.住居所地址：(中文) 英國倫敦卡芬迪西廣場二號(或營業所) (英文) c/o IDG Ventures Europe, 2 Cavendish  
Square, London, W1G 0PD, United  
Kingdom國籍：(中文) 英國(英文) UNITED KINGDOM代表人：(中文) 1. 王禮俊(英文) 1. Wang, Avery Li-Chun

說明書發明人續頁

發明人 2

姓名：(中文) 丹尼爾·庫伯特

(英文) Culbert, Daniel

住居所地址：(中文) 美國加州太陽谷艾斯特街一〇三五號

(英文) 1035 Aster Avenue #2196, Sunnyvale, CA  
94086, U.S.A.

## 捌、聲明事項

### ■主張專利法第二十四條第一項優先權：

【格式請依：受理國家（地區）；日期；案號 順序註記】

1.美國 \_\_\_\_\_ ; 2002/04/25 ; 60/376,055 \_\_\_\_\_

(1)

## 玖、發明說明

### 【發明所屬之技術領域】

本發明通常是有關於在一大音頻檔案資料庫上處理音頻訊號。更尤其是，它是有關於一種用以快速並準確地決定兩個音頻試樣是否匹配，以及是否免於含播放速度變動之各種變換之發明技術。本發明技術更能準確估算變換。

### 【先前技術】

用以快速且準確地自動辨識音樂及其它音頻訊號之需求持續成長。先前可用之音頻辨識技術常因準確性、或免於噪訊而犧牲速度。在某些申請案中，要在極度噪訊存在時估算時間對時間之分佈圖斜率時需要計算迴歸分析，這導引出許多難度及低效能之速度和準確性。先前既存之音頻辨識技術在重大播放速度變動存在中，因此無法實施快速及準確之辨識，例如，無法辨識以高於正常速度播放時之錄音。

增加問題複雜性的是 DJ 在無線電台、俱樂部和其它場所所使用之漸增受到歡迎種類之速度變動，修正音調之節奏變動。目前，不管播放速度變動及/或修正音調節奏變動的話，無能實施快速且準確音頻辨識之強力且可靠之技術。

### 【發明內容】

本發明提供一種說明兩音頻檔案之間關係特徵之快速

(2)

且不變動之方法，完成音頻辨識技術之要求。本發明方法克服既有技術之以上所提及缺點，甚至在極度噪訊存在時亦為準確。

根據本發明一觀點，兩音頻試樣間之關係具首先匹配得自各別試樣之某種指紋物件之特徵。對各音頻試樣產生一組指紋物件。各指紋物件是發生在各別音頻試樣內之一特別位置。各位置之決定依各別音頻試樣內容而定且各指紋物件在或接近各特別位置處具備各別音頻試樣之一或更多局部特性。在一實施例中，各指紋物件具備一變動成份和不變動成份之特徵。接著決定各對匹配指紋物件之一相對值。然後產生相對值之柱狀圖。如在柱狀圖中發現統計上之明顯尖峰，則兩音頻試樣具備，例如，實質上匹配之特徵。

根據本發明另一觀點，藉由以柱狀圖一軸上之一尖峯位置提供一全面相對值之估算可進而加強上述技術。接著，藉選取關注尖峯鄰近區並在所選取鄰近區計算一相對平均值可改進全面相對值。

還有在從柱狀圖尖峯決定一相對播放速度值之另一實施例中，對各對匹配之指紋物件計算一補償之相對時間偏置值。根據補償之相對時間偏置值產生另一柱狀圖。如在第二柱狀圖中發現統計上之明顯尖峯時，兩音頻試樣間之關係則進而具備尖峯特徵，進而加強本發明之準確性。

【實施方式】

(3)

本發明能在一大音頻檔案資料庫上作快速、強力、不變動、及比例化之編列索引及搜尋，並對音頻圖樣辨識應用特別有用。在某些實施例中，此處所發表之技術改進並增強在以上所參考之美國專利申請案中所發表之音頻辨識系統和方法。

兩音頻試樣檔案間之非常快速與有效比較運算在建立一商業上可存活之音頻辨識系統中是必要的。根據本發明一觀點，兩音頻試樣間關係具備得自如第 1 圖中所示，各別音頻試樣頻譜之第一匹配某種指紋物件之特徵。頻譜為一時間、頻率代表 / 分析，它是以滑動窗框中一次取樣  $2 * K$  並計算傅立葉 (Fourier) 變換產生的，因此在各音框中產生  $K$  頻箱。音框可重疊加以改進時間之解析分析。使用之特別參數依處理之音頻試樣種類而定。最好使用取樣率 8 KHz， $K=512$  之音框，和跨步為 64 試樣之分散時間音頻檔案。

指紋物件

產生各音頻試樣之音譜後，如第 2 圖中所示，篩檢求得局部特性，例如局部能量尖峯。匹配程序藉由從各音頻試樣之對應局部特性抽取一組指紋物件而開始。在一典範實施例中，一音頻試樣為一要加以辨識之未知聲音試樣而另一音頻試樣為一儲存在資料庫中之已知錄音。各指紋物件發生在各別音頻試樣內之一特別位置。在某些實施例中，各指紋物件位在一音頻檔案內之某些時間偏置位置並在

(4)

接近其各別時間座標位置，包含有關音頻檔案之一組敘述資訊。那就是，依接近各別時間偏置之音頻試樣而定加以計算各指紋物件中所包含之敘述資訊。這被編碼成一小資料結構。最好，以通常可再生，甚至存在噪訊，失真，及如變動播放速度之其它變換之方式，決定位置和敘述資訊。在這情況中，依各別音頻試樣之內容而定，決定各位置，且各指紋物件具備在或接近如第 1 圖中所示，例如，位置  $(t_1, f_1)$  或  $(t_2, f_2)$  之各別特別位置處各指紋物件具各別音頻試樣之一或更多局部特性。

在一典範實施例中，各指紋物件具備其位置，變動成份、和不變動成份之特徵。各局部特性為一音譜尖峯並從一對應音譜尖峯之頻率座標決定各頻率值。尖峯之決定是藉在各時間-頻率座標附近加以搜尋並選取比其鄰近具較大值之點。更明確地說，如第 2 圖中所示，將一音頻試樣 210 分析成在區域 221 和 222 表示高能量之頻譜代表 220。抽取與局部能量區 221 和 222 有關之資訊並將其摘要成一指紋物件 231，232 等之清單 230。各指紋物件選擇性地包含一位置欄 242，一變動成份 252，及一不變動成份 262。最好，選取鄰近區，使得各選取點在以其為中心之一  $21 \times 21$  單位區內為最大。讀者可參考以上所參考之美國專利申請案，更加討論鄰近區及點之選取。接著，對各對匹配之指紋物件，決定一相對值。在某些實施例中，相對值為各別音頻試樣參數值演算之商或差。然後產生一相對值之柱狀圖。如果在柱狀圖中發現一統計上之明顯尖峯，

(5)

則兩音頻試樣具實質上匹配之特性。

參考第 3 圖，分別如音頻試樣 1 和 2 之以上說明，分別備製指紋物件清單 310 和 320。從各清單比較各指紋物件 311 和 322。在步驟 351 中，例如使用各不變動成份  $1NV$  和  $1NV'$  將匹配指紋物件配成對，並在步驟 352 中將其放在一清單中。在步驟 353 中，計算各匹配對之相對值。接著，在步驟 354 中，產生一相對值之柱狀圖。在步驟 355 中，在柱狀圖中搜尋一統計上之明顯尖峯。在步驟 356 中，如找不到，則音頻試樣 1 和 2 不匹配，例如為第 4A 圖之柱狀圖 410。另外，如偵測到一統計上之明顯尖峯，則音頻試樣 1 和 2 匹配，例如為第 4B 圖之柱狀圖 420。

如第 361 步驟中之說明，藉由以柱狀圖一軸上之一尖峯位置提供一全面相對值  $R$  之估算可進而加強上述技術。在某些實施例中，首先選取所關注尖峯鄰近區能將  $R$  細調。在第 1 圖中，這以一特殊位置  $(t_1, f_1)$  附近之一關注區 110 表示。接著，計算所選取鄰近區中之平均相對值。這平均值可為在所選取鄰近區中以數點各相對值計算加權之平均值。在某些實施例中，能進而將  $R$  細調，對各匹配對產生相對時間偏置值  $t'-R*t$ 。以這些相對時間偏置值，步驟 362-364 表示產生一第二柱狀圖，允許計算一補償時間偏置。

例如，可實施其它種之時間-頻率分析，加以抽取指紋物件，例如為 Wigner-Ville 分佈或微波。而且，不用頻



(6)

譜圖尖峯，亦能使用例如為倒頻譜係數之其它特性。而且，可使用超解析技術，得到由頻譜尖峯所提供之時間-頻率座標之更細微頻率和時間估算。例如，可使用有關頻率箱之拋物線內插法增加頻率解析度。在朱利亞斯（史密斯三世（Julius O. Smith III）和薩比亞西拉（Xavier Serra）之"PARSHL：根據正弦波代表，對非諧和聲音之分析/合成程式"，國際電腦音樂會議錄（ICMC-87，東京），電腦音樂協會，1987，及 Prentice Hall 公司所出版由史提芬凱（Stereon M.kay）（1988 年元月）所著之"現代頻譜估算：理論與應用"中可發現相關之典範教義，此處將後兩者納入參考。

#### 匹配處理

在一匹配運算中，經由其各別指紋物件比較兩音頻試樣。如以前參考第 3 圖之討論，產生匹配指紋物件配對，各配對實質上包含匹配成份。備置資料，允許快速搜尋之一種方式為將指紋物件編碼成數值標識，如 32 位元無符號之整數，並使用數值標識作為儲存和搜尋之關鍵。例如在艾迪生衛斯理（Addison Wesley）公司所出版，由唐納歐文努斯（Donald Ervin Kmuth）（1998 年 4 月 7 所著之"計算機程式規劃技術，第 3 冊：儲存和搜尋（第 2 版）"中熟知有效資料處理技術，此處將其納入參考。

在一典範實施例中，各指紋物件包含一不變動成份和一變動成份。不變動成份指的是對應於頻譜尖峯之頻率值

(7)

比率，而且在時間延長下，頻譜尖峯間之時間差（即，時間差距）比率不變動。例如，參考第 5A 和 5B 圖，如音頻試樣頻譜在座標  $(t_1, f_1)$ ， $(t_2, f_2)$ ，和  $(t_3, f_3)$  是某些局部頻譜尖峯，則對於兩點之不變動量為  $f_2/f_1$ ，即  $f_2'/f_1'=f_2/f_1$ 。額外 3 點之不變動量指定為  $f_3/f_1$ ， $(t_3-t_1)/(t_2-t_1)$ ，或  $(t_3-t_2)/(t_2-t_1)$ ，或藉變更這些點及/或計算這些數量或其組合之函數加以產生任何其它組合。例如， $f_2/f_1$  除以  $f_3/f_1$  可產生  $f_2/f_3$ 。而且，如使音頻試樣線性延長，如只是快速播放，則頻率和時間差額外地享受交互關係，故如  $f_1*(t_2-t_1)$  之數量亦為不變動量。可使用這些數量之對數，以加減取代乘除。為探求頻率和時間延長之比，假設他們無相依性，故具有一頻率變動量和一時間變動量是必要的。為使匹配運算有效率，我們使用不變動部位編列指紋索引並使用近似或正確值加以搜尋。使用近似匹配加以搜尋允許某些特別強韌性，對抗失真及圓弧化誤差，但如果搜尋不變動成份變成多維範圍之搜尋則產生更多成本。在較佳實施例中，需要正確匹配各指紋物件之不變動成份，因此產生一非常快速之系統，為了噪訊存在之辨識而對敏感度有一些妥協。重要的是要注意甚至在對應之音頻試樣中，只有少數指紋物件正確匹配，則這方法亦運作良好。在柱狀圖尖峯偵測步驟中，甚至如果正確匹配並殘存少如 1-2% 之指紋物件則在統計上明顯有一尖峯。

除了，或不用不變動成份外，亦能使用變動成份，減

(8)

小匹配指紋物件之個數。例如，我們可能需要來自第一音頻試樣之一變動成份  $V$  在  $\pm 20\%$  內匹配來自第二音頻試樣之一對應成份  $V'$ 。在那情況中，我們可形成一數值標識代表，使得上部位（例如，最高有效位元）包含不變動成份，而下部位（例如，最低有效位元）包含變動成份。然後，搜尋一近似匹配變成在使用變動成份之最低和最高值組成之標識上作範圍搜尋。如使用一變動成份完成搜尋，則因此未必嚴格需要在匹配運算時使用不變動成份。然而，建議在匹配程序中使用不變動成份，因它有助降低疑似匹配之個數，因此使柱狀圖編程程序有效率並降低處理一般開銷量。

另一方面，新變動成份本身可能是或不是兩指紋物件間匹配準則之一部分。變動成份之代表值可因從一原始錄音至一取樣錄音之某些參數變換而失真。例如，可選取如  $f_1, f_2, f_3$  之頻率變動成份以及如  $(t_2 - t_1), (t_3 - t_1)$  或  $(t_3 - t_2)$  之時間變動成份作為播放速度之變動成份。假設第二音頻試樣，例如，授引自資料庫之匹配試樣有一座標為  $(t_1', f_1'), (t_2', f_2')$  和  $(t_3, f_3')$  之頻譜，這些座標對應於以第一音頻試樣所列之相同點。然後，頻率成份  $f_1'$  可能有一比例化之值  $f_1' = R_f \cdot f_1$ ，其中， $R_f$  為一線性延長參數，說明多快或多慢會將第一試樣錄音與第二錄音比較。可使用各兩匹配音頻試樣之變動成份藉兩頻率值  $R_f = f_1' / f_1$  間之比率加以計算說明一巨觀參數之全面延長值之估算。這指定兩匹配時間-頻率點之相對音調比；例

(9)

如， $R_f=2$  意為第一音頻試樣為第二音頻試樣音調（頻率）之半。另一可能性為使用  $R_t = (t_2' - t_1') / (t_2 - t_1)$ 。在這情況中，相對值  $R$  為相對播放速度比，即， $R_t=2$  意為第一音頻試樣播放速度為第二音頻試樣之兩倍。

如  $R_f=1/R_t$ ，即， $f'/f = (t_2 - t_1) / (t_2' - t_1')$ ，則由於這種音頻試樣之交互時間-頻率關係，兩音頻試樣有一線性時間延長關係。在這情況下，我們可使用此處所發表之柱狀圖編程法，形成估算利用對應變動頻率成份之相對頻率比  $R_f$ ，且再次形成估算相對播放速度  $R_t$ ，然後實施比較加以偵測播放關係是否為線性或非線性。

通常，利用來自第一和第二音頻試樣之對應變動成份，從所匹配之指紋物件加以計算一相對值。相對值可為頻率之簡單比或時間差，或造成估算用以說明第一與第二音頻試樣間映射之全面參數的某些其它函數。但通常可使用例如為  $R = F(v_1, v_1')$  之任何兩個輸入之函數  $F(\quad)$ ，其中， $v_1$  和  $v_1'$  各為變動量。最佳者為  $F(\quad)$  為一連續函數，使得測量  $v_1$  和  $v_1'$  時之小誤差在輸出  $R$  形成小誤差。

### 柱狀圖編程

如此處之說明，對從指紋物件之匹配配對清單所計算之相對值組產生一柱狀圖。然後在柱狀圖中搜尋一尖峯。柱狀圖中，統計上存在之明顯尖峯表示已發生可能之匹配。這種方法不用如  $(t_1' - t')$  之時間偏置差，而在柱狀圖

(10)

中特別搜尋相對值之集叢。

根據本發明之原理，柱狀圖之作用在形成計數值箱，各箱相當於沿著柱狀圖獨立軸之一特定值。為達本發明之目的，柱狀圖之產生可就對相對值清單之分類加以完成。因此，偵測相對值清單之柱狀圖尖峯之一種快速和有效方式為將清單由小至大分類，然後篩檢找出具相同或類似值之最大塊項目。

#### 統計意義

如本發明此處之討論，甚至假如只有少如 2% 之指紋物件倖免於所有失真並匹配無誤時，兩音頻試樣亦能匹配無誤。藉由記下兩音頻試樣間之比較刻痕，這是可能的。明確地說，在柱狀圖尖峯附近選取一鄰近區並計數落在鄰近區中之所有匹配配對，記下刻痕。此外，可計算權重點數，扣減離尖峯中心較遠之配對之貢獻。

估算截止準則之一種方式為假設非匹配音軌刻痕之概率分佈以指數末尾往下掉。將這模式套用在實際所測量之非匹配音軌刻痕分佈。接著，對於一 N 音軌資料庫，計算最高刻痕之累積概率分佈（例如，取一單一非匹配刻痕之累積概率分佈之第 N 階指數）。一旦知道概率曲線並選取偽正量之一最大位準時（例如，0.5%），則可選取一數字臨界值並用以決定柱狀圖尖峯之匹配配對是否有一統計上明顯之個數。

(11)

## 超精細估算

一旦找到一統計上明顯之柱狀圖尖峯，則可計算全面相對值（如相對播放速度）之高解析"超精細"估算。這種計算之完成是藉由在尖峯附近選取一鄰近區，例如，包含離尖峯柱狀圖箱，中心約 3 或 5 箱寬之間隔，並計算鄰近區中之平均相對值。使用這種技術，我們可發現準確性達 0.05% 內之相對播放速度。以此處所發表之偏置引申，可以優於 1ms 之準確性估算全面之時間偏置，該準確性比以上所討論之頻譜音框之時間解析更精細。

## 強力迴歸分析

如以上所參考之美國專利申請案中之討論，在試樣真正匹配之情況中，如第 6A 圖中所示，在匹配試樣之匹配指紋物件之對應時間座標（ $t'$ ， $t$ ）彼此相對所劃之分佈圖中可看到一斜線。難題是在高噪訊量存在中找尋迴歸方程式，它是由斜線之斜率和偏置所決定的。斜率表示相對播放速度，而偏置為一音頻試樣一開始對第二音頻試樣一開始之相對偏置。習知上有如最低均方調和之迴歸技術，例如，為威廉培斯（William Ho Press），布萊恩佛萊納利（Brian P. Flannery），沙烏提可夫斯基（Saul A. Tenkolsky），及威廉維特寧（William T. VeHerling）（1993 年元月）在劍橋大學校刊所著之"以 C 寫成之數值秘笈：科學計算之技術（第二版）"，此處將該文納入參考。不幸地，習知技術苦於不相稱之敏感度，其中，單一之

(12)

遠局外物可使所估算之迴歸參數急劇傾斜。實際上，相對點常由局外物主導，使其非常難以偵測正確斜線。強力迴歸分析之其它技術可用以克服局外問題，在噪訊存在之相對點之間找到線性關係，但這些傾向於緩慢與反覆且在局部最佳化中可能卡住。在找尋一未知線性迴歸變數之文獻中存在廣大各種技術。從數學作品 (Mathworks) 及此處所納入參考之 Matlab 工具幫手包含迴歸分析用之各種軟體常規。

本發明提供估算相對播放速度 (或，在線性播放關係情況下，對等地為相對音調之例數) 之發明方法，該方法解決問題，甚至假如匹配之斜率不等於 1 時，如第 6B 圖，在時間-時間分佈圖中找到一迴歸線。如此處之發表，使用局部相對播放速度之柱狀圖，利用先前未考慮之資訊並提供快速且有效解決迴歸分析問題之未預期優點。

為找尋遍偏置，假設對應之時間點具下列關係。

偏置 =  $t1' - R_t * t1$ ，其中， $R_t$  由先前之討論求得。這為補償之時間偏置且作用在使兩音頻試樣間之時間座標系統正常化。這在如構成第 7A 圖中未知斜率之斜線及第 7C 圖中垂直線之時間-時間分佈圖之剪力變換亦可看到。第 7B 圖之柱狀圖 720 說明表示全面相對播放速度比 R 之累積相對播放速度比尖峯。然後由偏置公式指定新相對值，如第 7D 圖中所見到的，產生一新的柱狀圖 740。新柱狀圖 740 之尖峯指定全面偏置之估算值，如上述，利用尖峯鄰近區之平均值，該估算值可會是陡峭的。

(13)

概要言之，第一柱狀圖編程階段提供一種方式加以估算相對播放速度，以及決定是否存在匹配。第二柱狀圖編程階段確信候選匹配音頻試樣有明顯個數之亦暫時對齊之指紋物件。第二柱狀圖編程階段亦作為一第二獨立篩檢準則並有助降低偽正量之概率，因此提供較有力準則加以決定兩音頻試樣是否匹配。只在第一柱狀圖中如有一統計上之明顯尖峯時可選擇實施第二柱狀圖之編程階段，因此節省計算資源和努力。可選擇實施進一步之最佳化，例如，降低計算上之混亂，不用對清單上匹配指紋物件之所有配對計算第二柱狀圖，第二柱狀圖可只使用對應於第一柱狀圖尖峯之匹配對加以產生。

#### 多重錄音之同步處理

本發明之執行可用以對非同步之音頻錄音加入旁白及時間校準。例如，假設在稍微不同位置或環境，以不同麥克風獨立操作一 DAT 錄音機和一卡帶錄音機。如稍後預期要從各別錄音機將兩段錄音組合成一段混音，則可使用此處說明之強力迴歸分析技術兩音軌同步化，得到時間偏置。照這樣，甚至假如非同步化之錄音機以稍微不同速度操作時可以高度準確性決定相對速度，允許參考另一段錄音補償一段錄音。如發現其中一段錄音已損毀且需從另一音源加以補遺時，這尤其有用。如此處說明之時間校準和同步化因此允許透通性混音。



(14)

## 資料庫搜尋

因比較方法極快速，可能要將一大資料庫之變頻試樣預先處理成各別之指紋物件清單。因一嫻熟此技術者會認知到，使用目前可用之資料處理技術因此可將一未知音頻試樣預先處理成指紋物件之其本身各別清單。使用資料庫中預先處理之指紋物件，然後可實施上述之匹配，柱狀圖編程，及尖峯偵測技術加以找尋匹配。

雖然已詳細說明本發明及其優點，應了解的是本發明並不限於或被界定成此處所表示者或所討論者。尤其是，此處所發表之圖示和說明以圖例解釋有關本發明之技術，表示本發明之實例，並提供利用本發明之實例且不可推斷為使本發明受到限制。已知之方法，技術，或系統可不詳細加以討論，故能避免模糊本發明之原理。因以技術中之其中一項平常技能將會認知到，只要不偏離本發明之原理和精神，對本發明可加以實施，修飾，或另外改變。例如，可以在電腦可讀取媒體中具體之電腦可執行指令之形式加以實施或另外實現此處所說明之方法，技術，和步驟。另外，本發明可在一具有客戶終端和伺服器之電腦系統中加以實施。客戶終端傳送第一和第二音頻試樣間關係特徵所需之，例如，為指紋物件之資訊至表現特徵之伺服器處。因此，應以下列請求項目及其法律上之等效請求項加以決定發明範圍。

【圖式簡單說明】

(15)

第 1 圖代表一分析音頻試樣之頻譜圖。

第 2 圖為一表示根據本發明一觀點，產生自一音頻試樣指紋物件之範例圖。

第 3 圖說明根據本發明原理所比較之兩音頻試樣。

第 4A-B 圖表示具有及不具一統計上明顯尖峯之典範柱狀圖。

第 5A-B 圖說明當播放速度變動時，時間-頻率點之運動。

第 6A-B 圖表示匹配混合標識之第一音頻試樣（試樣聲音）和第二音頻試樣（資料庫聲音）之對應時間。當試樣聲音之播放速度與資料庫聲音相同時，斜率為 1。

第 7A-D 圖說明找到並繪製本發明柱狀圖技術之快速及有效斜率。

#### 元件對照表

210：音頻試樣

220：頻譜

221，222：能量區

230：清單

231：，232：指紋物件

242：位置欄位

252：變動成份

262：不變動成份

310，320：指紋物件清單

(16)

1 , 2 : 音 頻 試 樣

311 , 322 : 指 紋 物 件

#### 肆、中文發明摘要

發明之名稱：強烈且不變音頻圖樣匹配

本發明提供一種用以快速並準確決定兩個音頻試樣是否匹配、以及是否免於如為播放速度變動之各種變動之發明技術。兩音頻試樣間之關係具首先匹配得自各別試樣之某些指紋物件之特徵。對各音頻試樣（210）產生各發生在一特別位置（242）之一組（230）指紋物件（231，232）。各位置（242）之決定依各音頻試樣（210）之內容而定且各指紋物件（232）在或接近各別特殊位置（242）處具備一或更多局部特性（222）。接著決定各對匹配指紋物件之一相對值。然後產生一相對值之柱狀圖。如發現一統計上之明顯尖峯，則兩音頻試樣具備實質上匹配之特徵。

#### 伍、英文發明摘要

發明之名稱：Robust and invariant audio pattern matching

The present invention provides an innovative technique for rapidly and accurately determining whether two audio samples match, as well as being immune to various kinds of transformations, such as playback speed variation. The relationship between the two audio samples is characterized by first matching certain fingerprint objects derived from the respective samples. A set (230) of fingerprint objects (231,232), each occurring at a particular location (242), is generated for each audio sample (210). Each location (242) is determined in dependence upon the content of the respective audio sample (210) and each fingerprint object (232) characterizes one or more local features (222) at or near the respective particular location (242). A relative value is next determined for each pair of matched fingerprint objects. A histogram of the relative values is then generated. If a statistically significant peak is found, the two audio samples can be characterized as substantially matching.

(1)

### 拾、申請專利範圍

1. 一種具備一第一和一第二音頻試樣之間關係特徵之方法，包含以下步驟：

產生第一音頻試樣之第一組指紋物件，各指紋物件發生在第一音頻試樣內之一各別位置，各別位置之決定依第一音頻試樣內容而定，且各指紋物件在或接近各別位置處具備第一音頻試樣之一或更多特性；

產生第二音頻試樣之第二指紋物件，各指紋物件發生在第二音頻試樣內之一各別位置，各別位置之決定依第二音頻試樣內容而定，且各指紋物件在或接近各別位置處具備第二音頻試樣之一或更多特性；

藉由使來自第一音頻試樣之第一指紋物件和來自第二音頻試樣，實質上類似於第一指紋物件之第二指紋物件相匹配而使指紋物件配成對；

根據配對步驟，產生所匹配指紋物件之配對清單；

決定各對匹配指紋物件之相對值；

產生一相對值之柱狀圖；以及

在柱狀圖中搜尋一統計上之明顯尖峯，該尖峯具備第一和第二音頻試樣間關係之特徵。

2. 如申請專利範圍第 1 項之方法，其中，如發現一統計上之明顯尖峯時，則第一和第二音頻試樣間之關係具備實質上匹配之特徵。

3. 如申請專利範圍第 1 或 2 項之方法，更包含以柱狀圖軸上一尖峯位置加以估算一全面相對值之步驟，全面

(2)

相對值更具備第一和第二音頻試樣間關係之特徵。

4. 如申請專利範圍第 3 項之方法，更包含超精細估算全面相對值之決定步驟，其中之決定步驟包含：

在尖峯附近選取一鄰近區，以及

在鄰近區中計算一平均相對值。

5. 如申請專利範圍第 1 項之方法，其中，各指紋物件具一不變動成份，而各對匹配之指紋物件中之第一和第二指紋物件具匹配之不變動成份。

6. 如申請專利範圍第 5 項之方法，其中，使用至少以下之一，產生不變動成份：

(i) 一第一和一第二頻率值間之比率，從接近各指紋物件各別位置之第一和第二局部特性分別決定各頻率值；

(ii) 一頻率值和一時間差值間之乘積，從第一局部特性決定頻率值，並在接近各指紋物件各別位置之第一局部特性與第二局部特性之間決定時間差值；以及

(iii) 一第一和一第二時間差值間之比率，從第一和第二局部特性決定第一時間差值，從第一和第三局部特性決定第二時間差值，各局部特性接近各指紋物件之各別位置。

7. 如申請專利範圍第 6 項之方法，其中，各局部特性為一頻譜尖峯並從一對應頻譜尖峯之一頻率座標決定各頻率值。

8. 如申請專利範圍第 1 或 5 項之方法，其中，各指

(3)

紋物件具一變動成份，並利用第一與第二指紋物件之各別變動成份加以決定各對匹配指紋物件之相對值。

9. 如申請專利範圍第 8 項之方法，其中，變動成份為從接近各指紋物件之各別位置之一局部特性所決定之頻率值，使得一對匹配指紋物件之相對值具備第一和第二指紋物件各別頻率值比率之特徵，且柱狀圖中之尖峯具備第一和第二音頻試樣間關係之特徵而第一和第二音頻試樣具備相對音調之特徵，或在線性延長之情況中為相對播放速度之特徵。

10. 如申請專利範圍第 9 項之方法，其中，各別頻率值之比率具備成除法或差法演算之特徵。

11. 如申請專利範圍第 9 項之方法，其中，各局部特性為一頻譜尖峯並從一對應頻譜尖峯之一頻率座標決定各頻率值。

12. 如申請專利範圍第 8 項之方法，其中，變動成份為從接近各指紋物件之各別位置之第一和第二局部特性所決定之時間差值，使得一對匹配指紋物件之相對值具備各別變動時間差值比率之特徵，且柱狀圖中之尖峯具備第一和第二音頻試樣間關係之特徵而第一和第二音頻試樣具備相對播放速度之特徵，或在線性延長之情況中為相對音調之特徵。

13. 如申請專利範圍第 12 項之方法，其中，各別變動差值之比率具備成除法或減法演算之特徵。

14. 如申請專利範圍第 12 項之方法，其中，各局部

(4)

特性為一頻譜尖峯並從一對應頻譜尖峯之一頻率座標決定各頻率值。

15. 如申請專利範圍第 8 項之方法，更包含以下步驟：

利用各別變動成份決定第一和第二音頻試樣之一相對音調，其中，各變動成份為從接近各指紋物件各別位置之一局部特性所決定之頻率值；

利用各別變動成份決定第一和第二音頻試樣之一相對速度，其中，各變動成份為從接近各指紋物件各別位置之一第二局部特性所決定之時間差值；以及

偵測相對播放速度之相對音調和倒數實質上是否相異，在這情況下，第一和第二音頻試樣間之關係具備線性之特徵。

16. 如申請專利範圍第 1 項之方法，其中， $R$  為一從相對值條狀圖之尖峯所決定之相對播放速度值，更包含以下步驟：

對於清單中之各對匹配指紋物件，決定一所補償相對時間偏置值， $t - R * t'$ ，其中， $t$  和  $t'$  為有關第一和第二指紋物件之時間位置；

產生所補償相對時間偏置值之第二柱狀圖；以及

在所補償相對時間偏置值之第二柱狀圖中搜尋一統計上之明顯尖峯，該尖峯更具備第一和第二音頻試樣間關係之特徵。

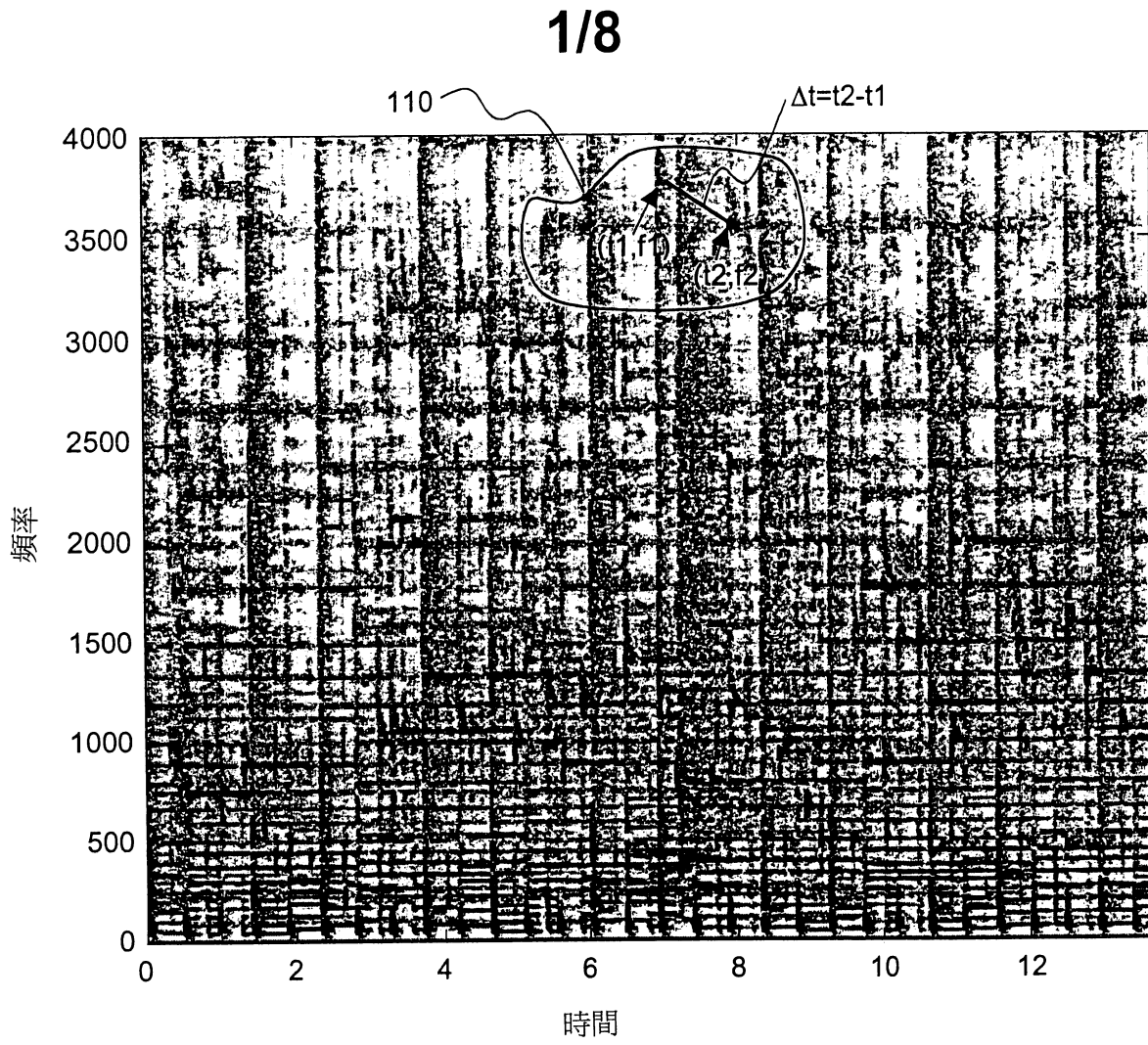
17. 一種用以實施如申請專利範圍中任一先前項之方



(5)

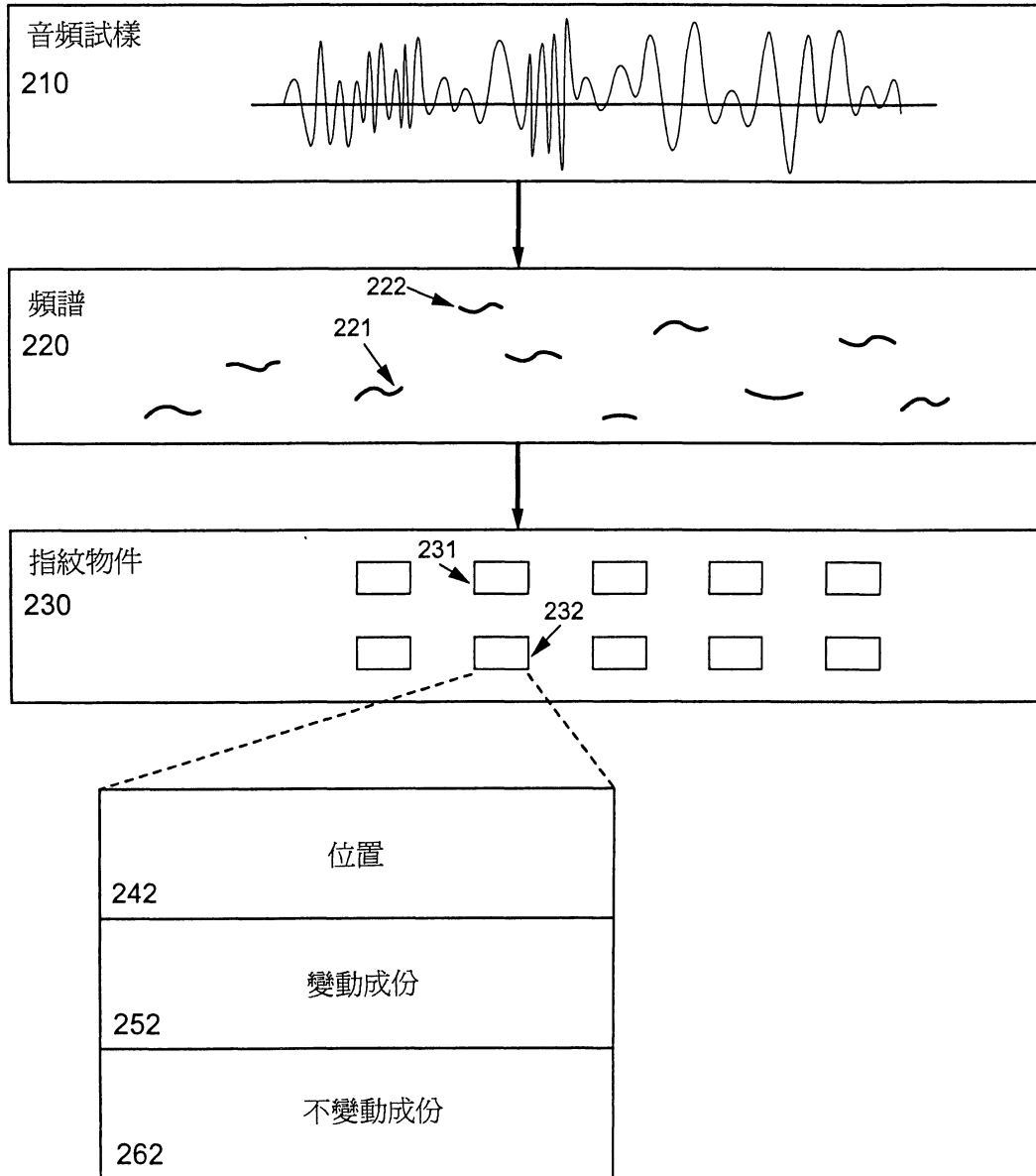
法之電腦程式產品。

18. 一種用以實施如申請專利範圍第 1 至 16 項中任一項方法之電腦系統，該電腦系統包含一客戶終端用以將第一與第二音頻試樣間關係特徵所需之資訊傳送至一表現該特徵之伺服器。

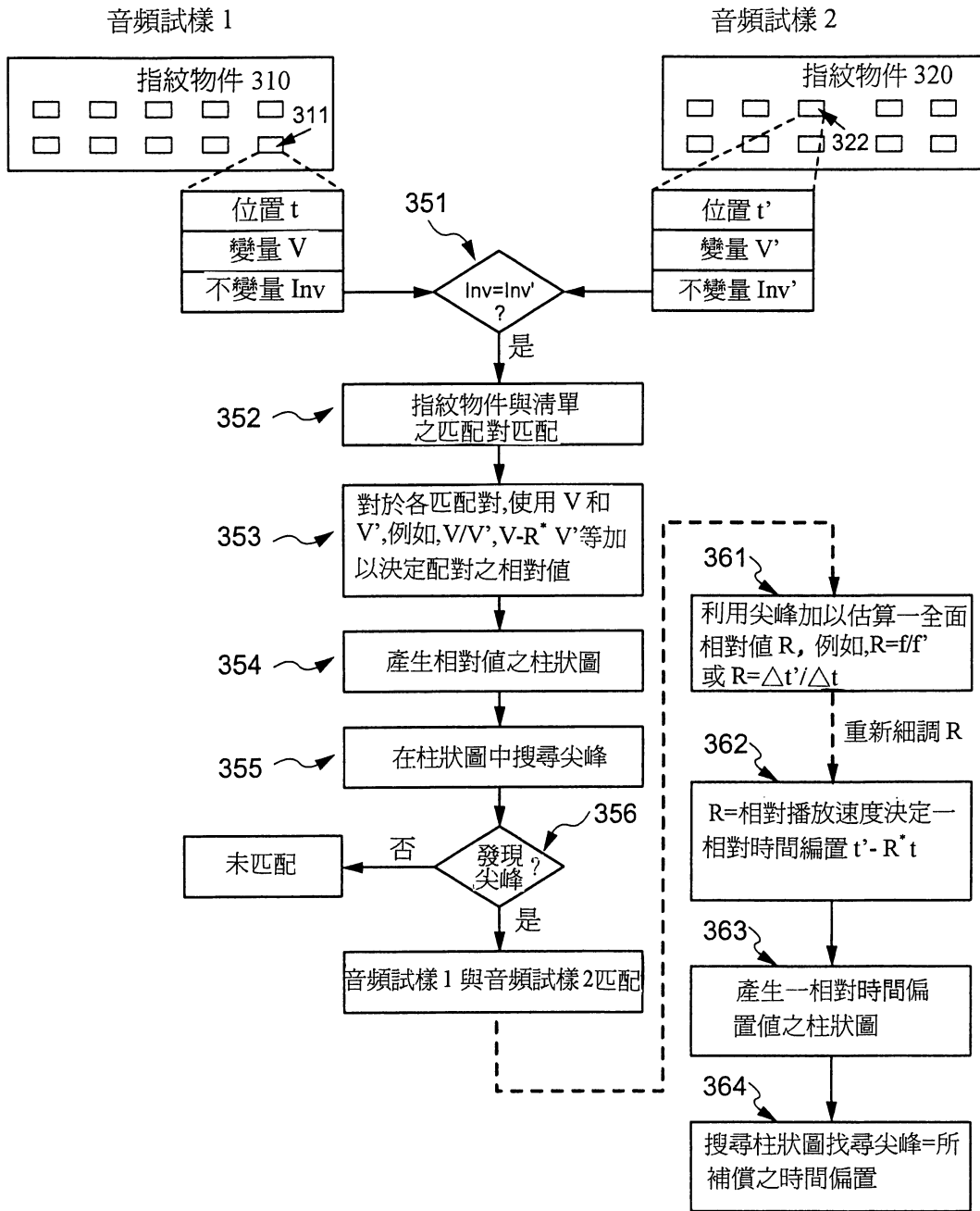


第 1 圖

2/8

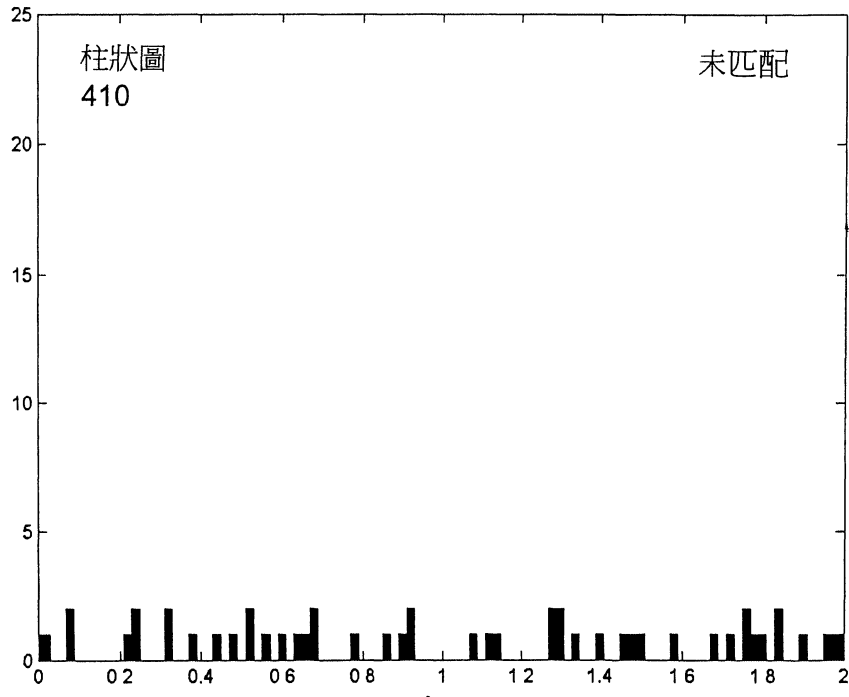


第 2 圖

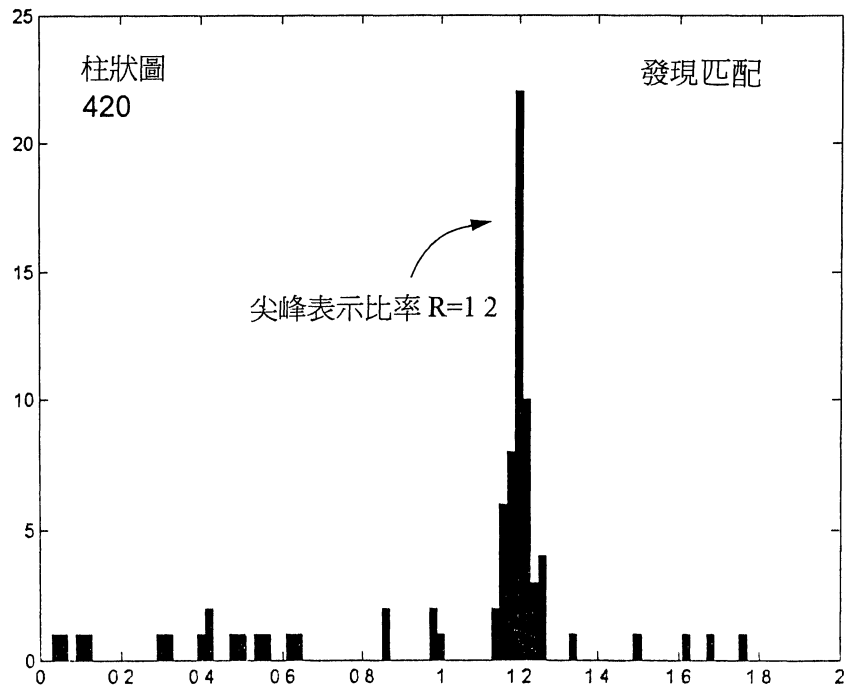


第 3 圖

4/8



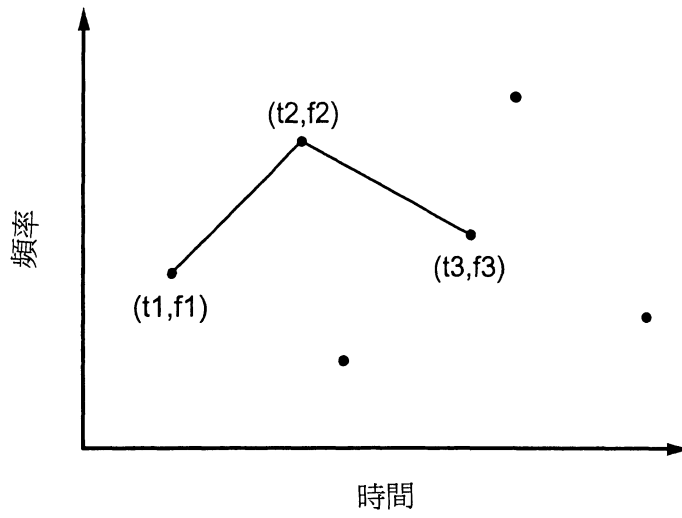
第4A圖



第4B圖

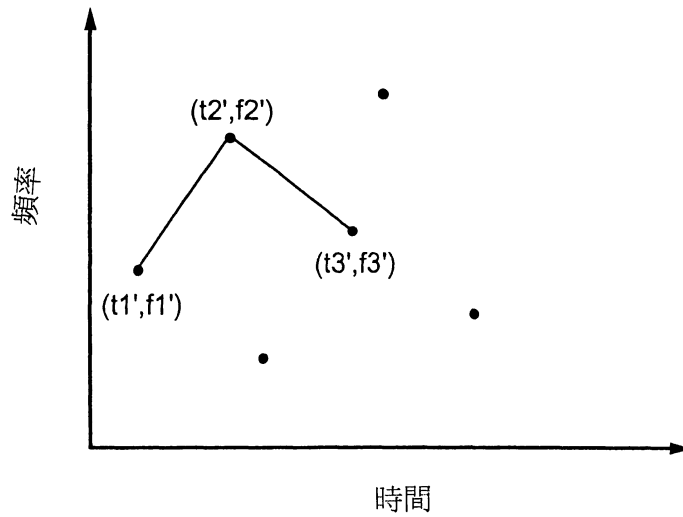
# 5/8

正常播放速度



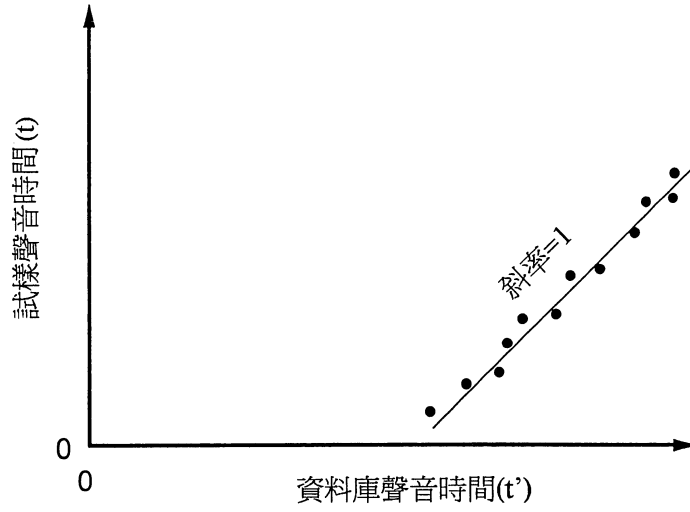
第5A圖

快速播放速度

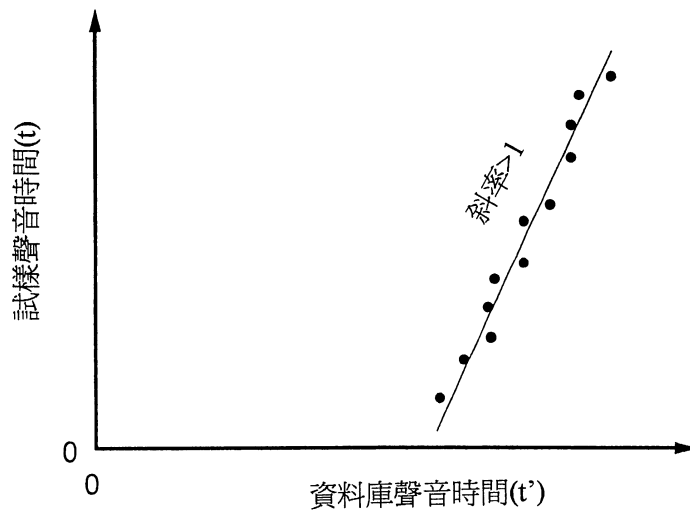


第5B圖

6/8

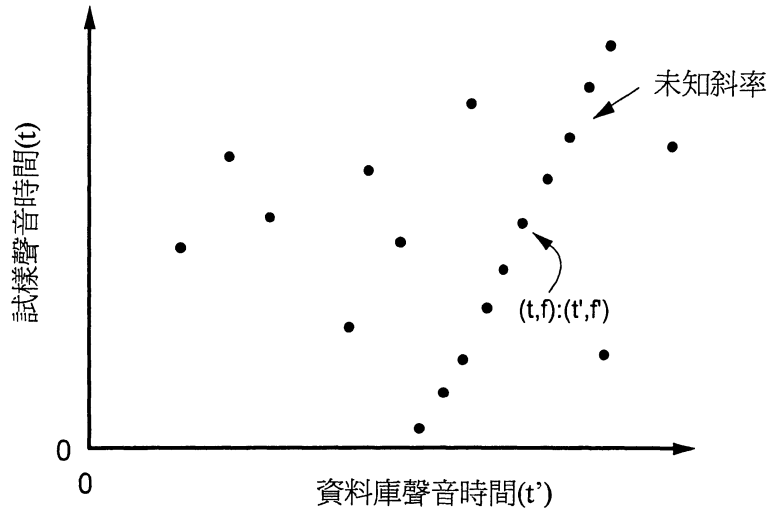


第6A圖

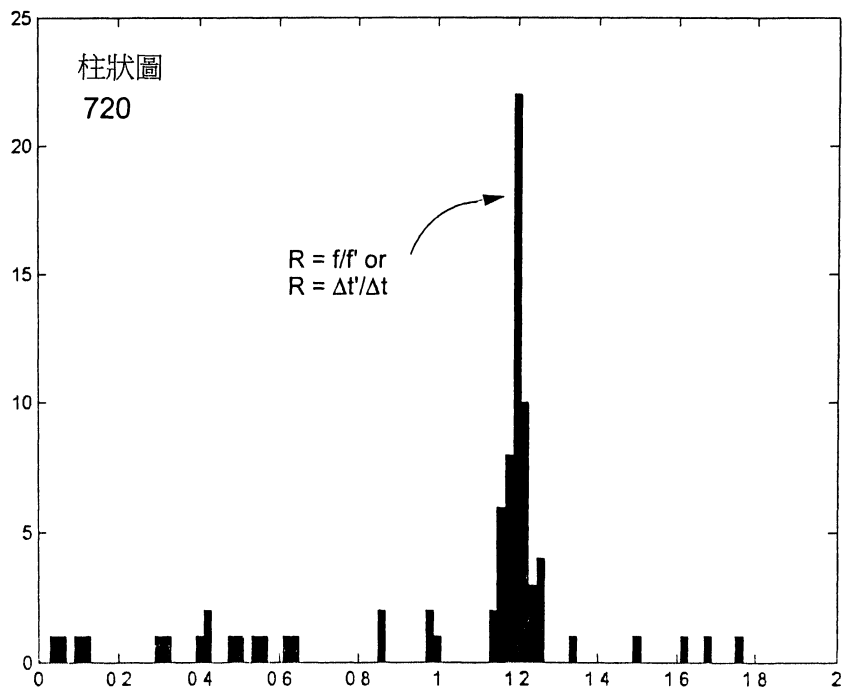


第6B圖

7/8

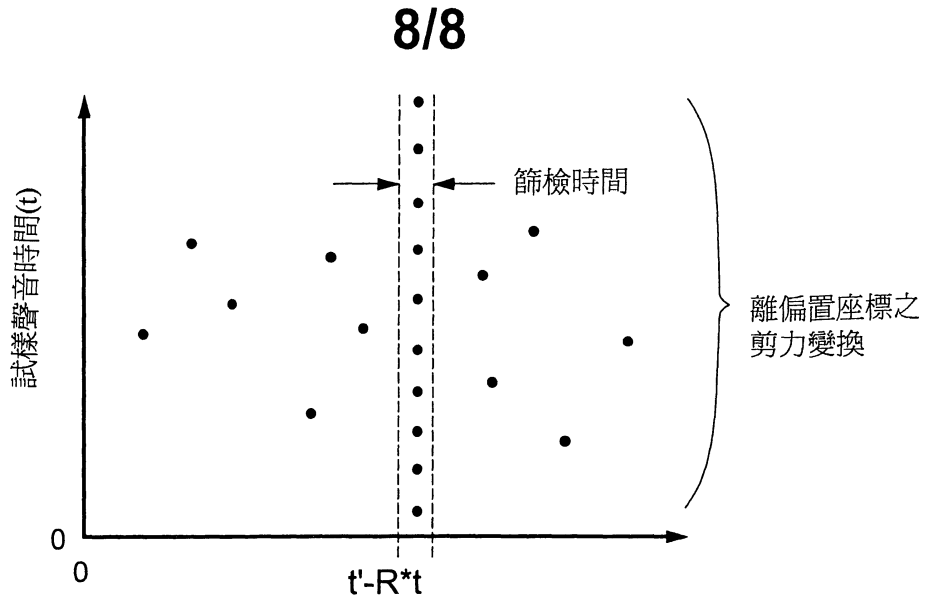


第7A圖

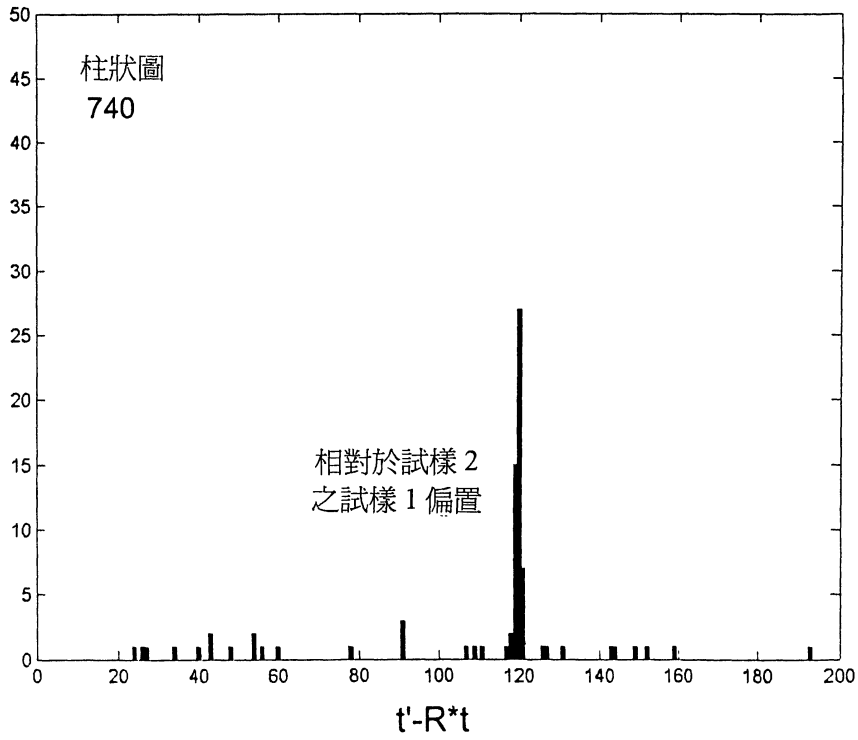


第7B圖





第7C圖



第7D圖

陸、(一)、本案指定代表圖為：第 3 圖

(二)、本代表圖之元件代表符號簡單說明：

無

柒、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

無