

四、聲明事項：

主張專利法第二十二條第二項 第一款或 第二款規定之事實，其事實發生日期為： 年 月 日。

申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

有主張專利法第二十七條第一項國際優先權：

無主張專利法第二十七條第一項國際優先權：

主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

主張專利法第三十條生物材料：

須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

九、發明說明：

【發明所屬之技術領域】

本發明是有關於多媒體資料(multimedia data)之語音註解(speech annotation)及檢索(retrieval)，特別是關於一種管理多媒體資料的方法與裝置，利用音節模擬影像(syllable-transformed image)於多媒體資料之語音註解及檢索。

【先前技術】

隨著網際網路的爆炸性成長與數位影像擷取裝置如數位攝影機(digital video camera)、數位相機(digital camera)與數位錄影機(digital video recorder)等的普及，使用者獲得多媒體資料的方法越來越多，也越來越簡單。使用者所擁有的多媒體資料的增加速度自然地越來越快。相對地，管理多媒體資料也變得困難與耗時。

雖然時下已有一些管理多媒體資料的系統或方法。但多數是利用文字(text)或語言轉文字(language-to-text)的方式來描述、索引(index)及搜尋(retrieval)多媒體資料。時下的多媒體資料的註解(annotation)或搜尋技術大致可分為下列四種：以文字註解與搜尋、語音轉文字(speech-to-text)的註解與搜尋、圖形分析搜尋、以及語音註解(speech annotation)與搜查。

文字註解與搜尋方式較不用特殊需求。但會有輸入文字較為冗長的麻煩，和註解搜尋時可能受限於系統提供的關鍵詞。以文字註解方式的文獻如美國專利 6,833,865，此專利文獻揭露一種在數位影像擷取裝置裡嵌入已編碼及結構化資料的引擎(embedded metadata engines in digital capture devices)。在數位影像擷取裝置裡加入分析影像內容的功能，透過此影像內容的分析，自動產生出內容相關的額外資訊並與原始影像資料一併儲存。此專利文獻的揭露雖能動態地為數位影像產生註解，但並未揭露影像的搜尋。

語音轉文字的註解與搜尋的方式需要語音辨識器，也造成語言相關的限制。以語音轉文字的註解與搜尋方式的文獻如美國專利 6,397,181。此專利文獻揭露一種對多媒體資料註解及取得的方法和裝置(method and apparatus for voice annotation and retrieval of multimedia data)。在註解時使用語音輸入的方式，並利用語音辨識技術將語音內容轉成文字，然後利用文字註解產生反向索引表(reverse index table)。在搜尋時使用語音輸入的方式，經語音辨識技術產生搜尋關鍵詞，再透過反向索引表找出符合的多媒體資料。

另一篇美國專利 6,499,016 文獻揭露一種利用以語音基礎指令語言而自動儲存和呈現影像(automatically storing and presenting digital images using a speech-based command language)的方法。在註解時使用語音轉文字的方式，在搜

尋時則利用文字來搜尋。使用者在拍照時可透過數位相機及時以語音錄下註解資訊，透過多段指令(command):敘述(statement)的語音輸入，為影像附加如地點和時間等描述，然後使用者將影像與註解資訊送到伺服器處理，由伺服器來辨識語音資訊，將語音轉成文字儲存。根據文字註解，使用者可以下達關鍵字動態產生不同的相簿以供瀏覽。

美國專利 6,813,618 文獻揭露一種在數位圖相簿裡尋找相關圖形材料的系統與方法(system and method for acquisition of related graphical material in a digital graphics album)。此專利透過對文字註解的搜尋來達到以圖找圖的目的。使用者可以透過網路找到與指定影像相關的影像。

利用圖形分析來進行搜尋的方式，則系統需要具備圖形分析的能力，使用者雖然不必對每張圖片附加註解，不過只能針對圖形搜尋，並且必須先找到作為搜尋條件的圖形，甚且也難以精確分析圖形的內容。以圖形分析方式的文獻如“An active learning framework for content-based information retrieval”, Multimedia, IEEE Transactions on Vol. 4, Issue 2, June 2002, pp 260-268. 此篇論文以內容為基礎之資訊檢索(content-based information retrieval)來建立樹狀屬性表(attribute tree)，以提供影像的標記。

以聲音搜尋的方式有多種，包括直接比對搜尋條件與註解的原始聲音資料、或者將聲音訊號經過辨識，將辨識結果以 N gram 組合來建立檢索向量，再進行語言檢索等。前者在資料量多時，逐一比對會耗費相當多的時間；後者則會受到語言相關性的限制。

雖然已存在上述四種對多媒體資料的註解或搜尋技術，但這些技術仍有前述可能發生的問題，並且都與語言相關(language dependent)，使用者因此無法自由自在地使用想要的語言或聲音。

因此，如何讓使用者可以簡化管理步驟，快速找到想要的多媒體資料，以及利用聲音以語言無關(language independent)的方式索引和搜尋是本發明的主要動機。

【發明內容】

本發明有效解決上述習知技術的缺憾。其主要目的是提供一種管理多媒體資料的方法與裝置。

本發明主要是根據由音節轉換而成的圖案 (syllable-transformed pattern) 與特徵影像分析 (eigen-image analysis)，來搜尋語音註解的圖像。對具有語音註解的數位圖像 (digital photo with speech annotation)，根據辨識出的音節後選者 (recognized syllable

candidate)轉換而成的模擬影像圖案(image-like pattern)，提供了新穎的索引與搜尋技術。在此技術中，提供了一個將聲音轉換為代碼時依據的機制，藉此達到與語言無關的搜尋。

本發明之管理多媒體資料的方法包括主要三個步驟。首先被擷取註解之聲音特徵透過一自動語音辨識器(speech recognizer)產生辨識結果，並轉換成一個模擬影像的資料。然後，利用一影像特徵處理方法來建立一索引資料庫。在搜尋時，以自然口述語音搜尋此多媒體資料。

依此，本發明之管理多媒體資料的裝置包括一個多媒體資料索引製作模組和一個多媒體資料檢索模組。多媒體資料索引製作模組將被擷取註解之聲音特徵轉換成一模擬影像資料後，擷取出此模擬影像資料的特徵參數並完成一索引資料庫的建立。多媒體資料檢索模組以自然口述語音來搜尋此多媒體資料。

將本發明實施在 1055 張具有語音註解的數位圖像，並相較於傳統僅使用音節、字元(character)和字(word)的索引方法，結果顯示本發明達到較佳的檢索準確性，並且檢索時間也較短。

茲配合下列圖示、實施例之詳細說明及申請專利範圍，將上述及本發明之其他目的與優點詳述於後。

【實施方式】

以下描述本發明之管理多媒體資料的方法與裝置。在描述中，本發明提供之管理多媒體資料的裝置以模組(module)的來描述其主要元件。一個模組，特別是它的功能性(functionality)，可以用硬體(hardware)、或軟體(software)、或硬體與軟體兩種模組的組合來實現。

第一 A 圖與第一 B 圖說明本發明之管理多媒體資料的裝置與運作流程。此管理多媒體資料的裝置包括第一 A 圖之為多媒體資料製作索引模組，以及第一 B 圖之搜尋此多媒體資料模組。參考第一 A 圖，此多媒體資料索引製作模組首先將被擷取註解之聲音特徵 101 透過一自動語音辨識器 103 產生一辨識結果，例如音節矩陣(syllable matrix)113。然後，轉換此辨識結果為一模擬影像資料 105，並利用一影像特徵處理方法來完成一索引資料庫的建立。參考第一 B 圖，在搜尋時，此多媒體資料檢索模組以自然口述語音搜尋此多媒體資料。

第二圖為多媒體資料索引製作模組的一個詳細方塊示意圖。此多媒體資料索引製作模組包括自動語音辨識器 103、一影像模擬模組 201、以及一特徵擷取與索引

建立模組 205。自動語音辨識器 103 辨識被擷取註解之聲音特徵 101，產生一辨識結果。影像模擬模組 201 首先計算出此辨識結果之相似度分數，然後根據此相似度分數，將此辨識結果轉換成一模擬影像資料。特徵擷取與索引建立模組 205 擷取此模擬影像資料的特徵參數，然後將此特徵參數作量化並完成索引資料庫的建立。此索引資料庫的建立解決了因為模擬影像資料量大，逐一比對相當耗費時間的問題。

第三圖詳細說明第一 B 圖之自然口述語音搜尋多媒體資料的步驟流程。參考第三圖，多媒體資料檢索模組首先以聲音輸入搜尋條件，並轉換此搜尋條件成一索引資料。然後比對此索引資料與索引資料庫裡的索引資訊。依此，找出符合此搜尋條件之相對應的多媒體資料

自動語音辨識的主要目的是希望自然口述語音能夠透過電腦來辨識，進而使電腦執行相對應的工作。當聲音訊號藉由類比到數位的轉換裝置輸入，並以數值方式儲存至電腦時，語音辨識器會將輸入的測試聲音樣本與事先建立好的語音辨識模型，進行比對以找出一個被認為最“像”該聲音樣本的辨識結果。如此，可以知道輸入的自然口述語音代表何意。

在使用自動語音辨識之前，利用機率與統計的方

法，配合收集到的語音和文字語料，事先建立好辨識時所需的辨識模型，包括語言模型和聲學模型。

第四圖說明自動語音辨識的方法。當辨識器收到聲音訊號時，如標號 401 所示，首先對此聲音訊號做參數的擷取，例如常用的梅爾倒頻譜參數。然後，利用動態搜尋方法，配合事先建立好的語言模型和聲學模型，進行自動語音辨識階段，如標號 403 所示。最後得到辨識結果。

基於此語音辨識結果，得以了解圖片的語音註解內容。但就時下的語音辨識技術，仍存在語音辨識誤差 (speech recognition error)。此誤差的產生包括統計方法上的偏差，以至於在辨識模型上發生混淆；以及一般人發聲上存在著一些容易混淆的發音，例如：勺和夕的發音混淆、一和七的發音混淆。

因此，本發明使用以發音音節為基礎之語音辨識器 (syllable-based recognizer)，有效地利用此語音辨識器定義出的所有發音音節，來建立索引資料庫，並且利用語音辨識器定義出的所有發音音節，建立一個混淆矩陣 (confusion matrix)，將音節結果轉換成對應的音節發音分數，用來量測各發音與其他發音的混淆。為避免誤差，本發明取辨識前 n 名最佳結果作為候選，稱之為

n-最佳音節候選者(n-best syllable candidate), n 為一自然數。

將每一個音節看成類似於影像上的像素，呈現的結果就如同是一張影像一般，在色彩上呈現深淺、相似度遠近的關係。但是，因為音節間彼此的距離並不線性的關係，所以本發明利用多元尺度量測方法(multidimensional scaling, MDS)，如第五圖所示，先建立資料間兩兩之距離矩陣，再將辨識音節結果根據多維度空間描述距離，透過矩陣空間的座標轉換，呈線性的距離關係，讓相似度(距離關係)以影像之色階深淺表示(gray level)，藉此轉換成模擬影像資料。

第六圖為一真實範例的示意圖，說明轉換後的模擬影像之間視覺上的性質(virtual property)。在第六圖裡，語音註解檔(speech-annotated file)裡，音節 geng 之已辨識 n-最佳音節候選者的模擬影像比音節 ge 之已辨識 n-最佳音節候選者的模擬影像更相似於查詢音節(query syllable)geng 之已辨識 n-最佳音節候選者的模擬影像。從第六圖，雖然對於音節 geng 在語音註解檔會發生語音辨識錯誤，然而本發明利用音節轉換成模擬影像作搜尋的技術仍然可以在查詢與語音註解檔之間，正確地匹配到查詢音節。換言之，利用本發明之搜尋技術，語音註解檔裡的音節 geng 可以正確地匹配到查詢音節

geng。

以下第七圖中，以一範例說明本發明從辨識結果(即第七 A 圖之發音音節)來建立索引的程序。如第七 B 圖所示，首先將選出的 n-最佳音節候選者轉換成一模擬影像資料，並以一個音節候選者為一張影像將影像作切割，如第七 C 圖所示。對切割後的影像資料再作影像特徵值分析(Eigen-Image)，所得的音節模擬影像特徵值為一個多維度向量資料，如第七 D 圖所示。最後將相似的影像特徵值的音節模擬影像聚成(cluster)同一串碼字(codeword)序列，如第七 E 圖所示。所有碼字序列則建立成碼本(codebook)。

語音註解及以聲音輸入的搜尋條件，經上述步驟皆可轉成一串碼字序列。本發明利用資訊檢索方法統計其詞頻及反轉詞頻(term frequency and inverse document frequency, $tf \times idf$)，建立向量式索引，此一向量維度的索引就代表每個註解內容的資訊。當使用者以聲音輸入搜尋條件，依同樣程序被轉換成一個向量維度(dimension)的資料，檢索時將此查詢語音向量與資料庫內索引的向量，兩兩作向量內積比對的運算，回覆給使用者符合搜尋條件的多媒體資料。

影像特徵值分析係對影像資料作主成分分析

(principle component analysis, PCA), 優點為擷取音節轉換後之模擬影像資料的具代表性特徵(extract significant feature), 以及降低雜訊資料干擾。主成分分析有多種實現方式。不失一般性, 以下之範例說明的影像資料作主成分分析的兩個主要步驟為共變異矩陣估計(covariance matrix estimation)和奇異值分解分析(singular value decomposition, SVD)。

奇異值分解分析的主要目的是將影像向量的所有維度投射(project)到單一且更具鑑別性資訊空間(single latent information space), 同時顯著降低資料維度(significantly reduced dimensionality)。

如第八圖所示, 共變異矩陣估計的運算包括正規化影像(image normalization)、取影像平均值(taking image mean)、取得減去影像平均值(obtaining differential image)、建立依影像取維度的矩陣(dimension-by-image matrix construction) W 和建立共變異矩陣(covariance matrix construction)。奇異值分解分析的運算包括求取 W 之自相關矩陣(autocorrelation matrix computation)、降低維度和資料雜訊(reducing dimensionality and data noise)、建立特徵影像的轉換基底(transform basis construction of eigen-image) U 、以基底 U 轉換新影像(eigen-image transformation)。

實驗結果顯示，本發明在儲存上有較少的資料維度、較快的檢索速度。檢索率約有 16.26% 的提升。

綜上所述，本發明提供一種管理多媒體資料的方法和裝置，包括對多媒體資料附加語音註解以及利用音節模擬影像之處理方式製作索引與檢索多媒體資料。將擷取語音資料的聲音特徵根據語音辨識器辨識出的音節矩陣。且利用音節相似度計算方法，估計出音節發音分數，將辨識結果之音節矩陣參考音節發音分數轉換為模擬影像資料。最後利用影像特徵處理方法來建立索引，經由比對註解與搜尋語音之索引，可以有較好的準確率找出想要的多媒體資料。

基於多媒體資料之語音註解，建立索引時，首先擷取註解之聲音特徵，經由自動語音辨識器產生 n 名音節辨識後選者。然後利用音節模擬影像資料轉換，將此 n 名音節辨識後選者轉換成模擬影像像素之資料。再透過影像特徵值分析，擷取影像內具代表性的特徵並且去除雜訊干擾。接著利用群集分析將音節之影像特徵值聚類，建立碼字(codeword)及碼本(codebook)。最後，每句輸入語音訊號都可經過上述流程轉換成一連串的編碼，利用資訊檢索上常用的 $tf \times idf$ 統計方式建立索引。

檢索時，輸入的語音查詢經同樣的處理流程來建立向量索引，再與資料庫內所建立的索引資訊做相似度比對，找出符合搜尋條件的語音註解及其對應之多媒體資料內容。

惟，以上所述者，僅為本發明之實施例而已，當不能依此限定本發明實施之範圍。即大凡本發明申請專利範圍所作之均等變化與修飾，皆應仍屬本發明專利涵蓋之範圍內。

【圖式簡單說明】

第一 A 圖與第一 B 圖說明本發明之管理多媒體資料的方法的步驟流程。

第二圖為多媒體資料索引製作模組的一個詳細方塊示意圖。

第三圖詳細說明第一 B 圖之自然口述語音搜尋多媒體資料的步驟流程。

第四圖說明自動語音辨識的方法。

第五圖是語音辨識器定義出之發音音節的一個範例。

第六圖說明本發明利用多元尺度量測方法，將音節資料間兩兩之距離矩陣為非線性關係，投影到線性可分的維度描述。

第七圖以一範例說明本發明從辨識結果來建立索引的程序。

第八圖進一步說明共變異矩陣估計和奇異值分解分析的運算。

【主要元件符號說明】

101 聲音特徵	103 自動語音辨識器
105 模擬影像資料	113 音節矩陣
201 影像模擬模組	205 特徵擷取與索引建立模組
401 參數的擷取	403 辨識階段(語言學解碼)

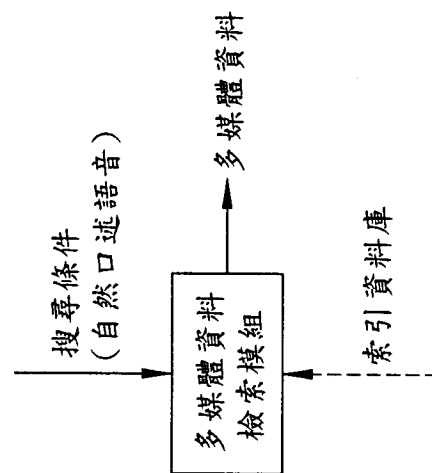
五、中文發明摘要：

一種管理多媒體資料的方法與裝置，包括對多媒體資料附加語音註解以及利用音節模擬影像之特徵分析處理來製作索引與檢索多媒體資料。基於多媒體資料之語音註解，被擷取註解之聲音特徵透過一自動語音辨識器產生辨識結果，並轉換成一個模擬影像的資料。然後，利用一影像特徵分析和量化處理來建立索引。在搜尋時，以自然口述語音搜尋此多媒體資料。實驗結果顯示，本發明在儲存上有較少的資料維度、較快的檢索速度。檢索的準確率約有 16.26% 的提升。

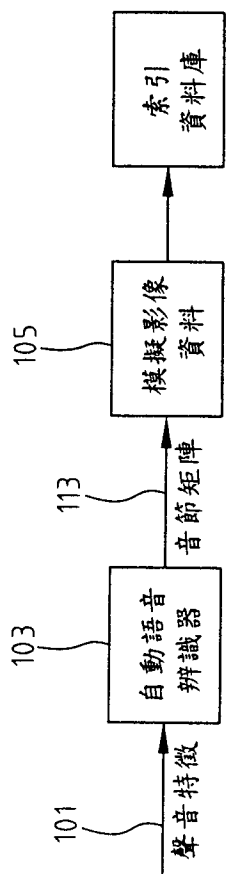
六、英文發明摘要：

A method and an apparatus for multimedia data management are disclosed. The method provides an indexing and retrieval scheme for digital photos with speech annotations based on image-like patterns transformed from the recognized syllable candidates. For annotated spoken content, the recognized n-best syllable candidates are transformed into a sequence of syllable-transformed patterns. Eigen-image analysis is further adopted to extract the significant information to reduce the dimensionality. Vector quantization is applied to quantize the syllable-transformed patterns into feature vectors for indexing. The invention of indexing scheme reduces the dimensionality and noise of data, and achieves better performance of 16.26% for speech annotated photo retrieval.

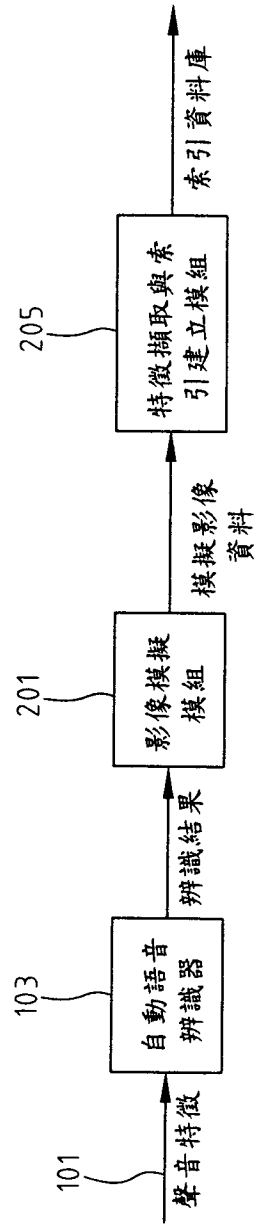
十一、圖式：



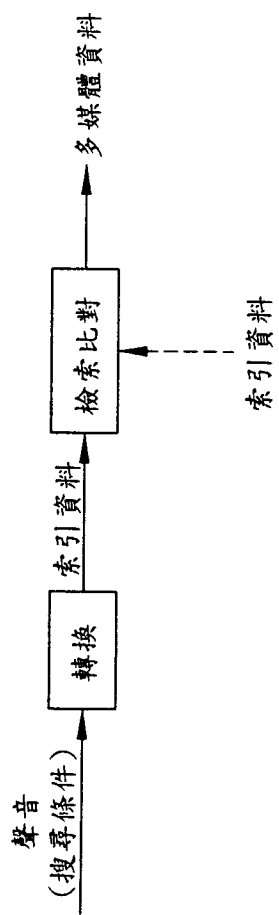
第一B圖



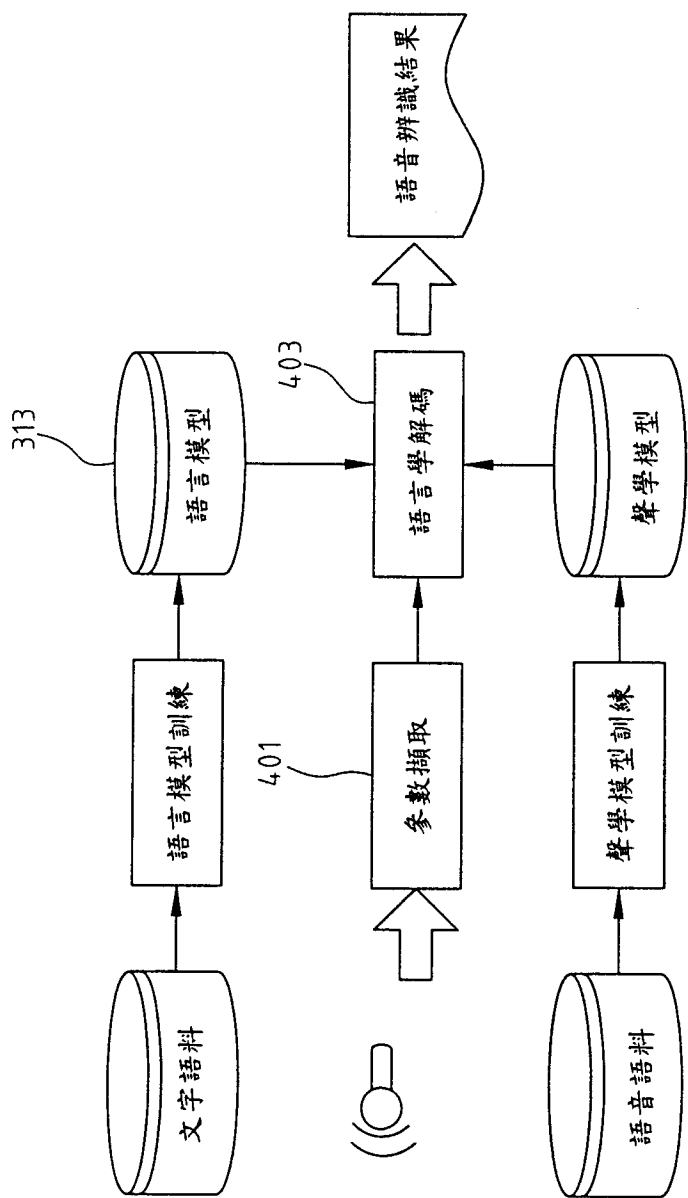
第一A圖



第二圖



第三圖



第四圖

以線性的距離關係象
徵影像之色階深淺的
關連性

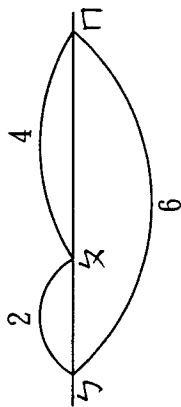
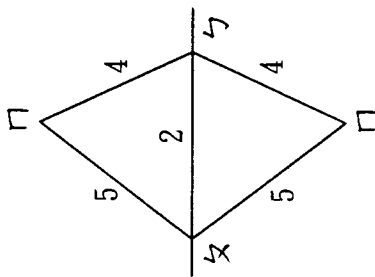
	ㄅ	ㄆ	ㄇ
ㄅ	0	2	6
ㄆ	2	0	4
ㄇ	6	4	0



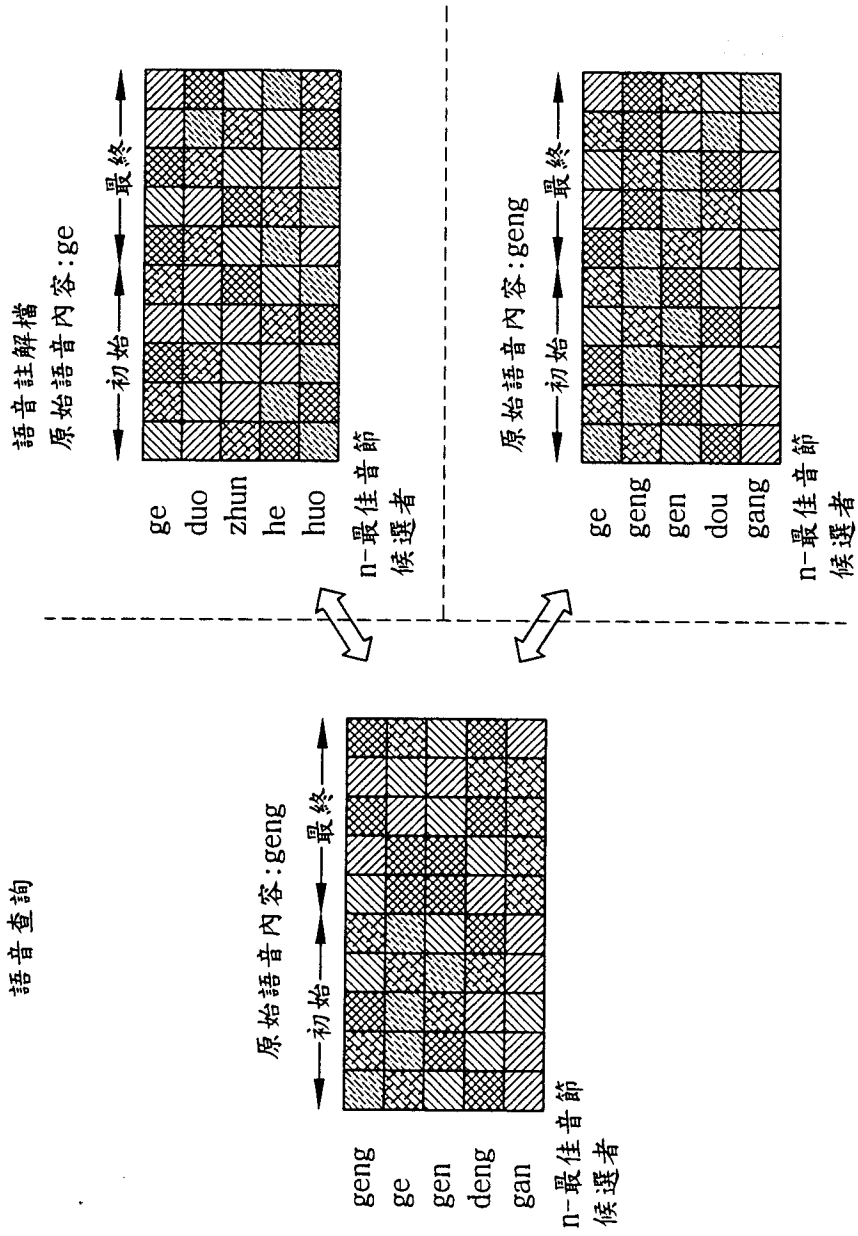
原本音節兩兩距離
為非線性關係

	ㄅ	ㄆ	ㄇ
ㄅ	0	2	4
ㄆ	2	0	5
ㄇ	4	5	0

空間座標轉換



第五圖



第六圖

第七A圖

xue

...

huo

jiao



第七B圖

xue
xuan
xu
que
xun

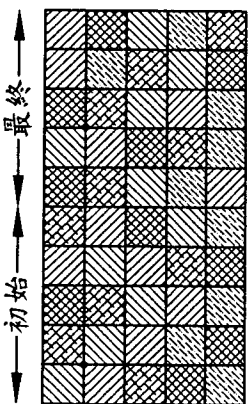
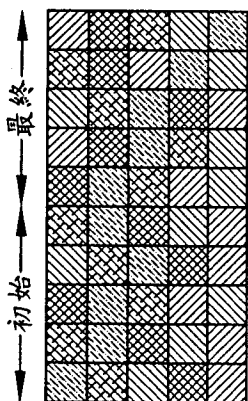
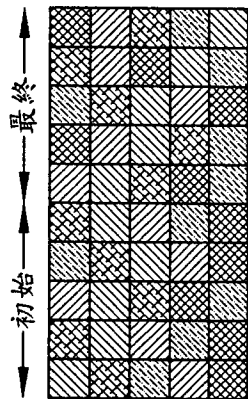
...

wo
huo
guo
wu
luo

jiao
qiao
jiang
diao
riao



第七C圖



第七D圖

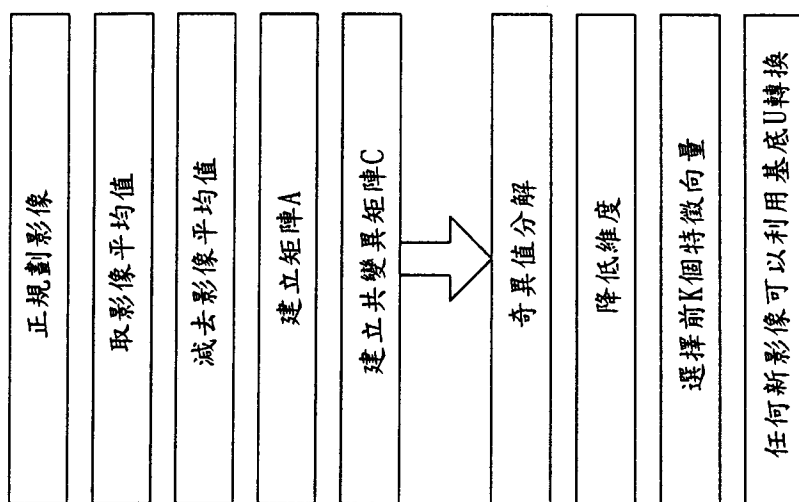


第七E圖

8

2

1



第八圖

七、指定代表圖：

(一)本案指定代表圖為：第（一 A 與一 B）圖。

(二)本代表圖之元件符號簡單說明：

101 聲音特徵	103 自動語音辨識器
105 模擬影像資料	113 音節矩陣

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

發明專利說明書

公告本

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※ 申請案號：95120209

※ 申請日期：95.6.7

※IPC 分類：G06F17/30 (2006.01)

G10L15/26 (2006.01)

G06T1/00 (2006.01)

一、發明名稱：(中文/英文)

管理多媒體資料的方法與裝置/

METHOD AND APPARATUS FOR MULTIMEDIA DATA MANAGEMENT

二、申請人：(共1人)

姓名或名稱：(中文/英文)

財團法人工業技術研究院/

INDUSTRIAL TECHNOLOGY RESEARCH INSTITUTE

代表人：(中文/英文)

張進福/CHANG, JIN-FUI

住居所或營業所地址：(中文/英文)

新竹縣竹東鎮中興路4段195號/

195, SEC. 4, CHUNG HSING ROAD, CHUTUNG, HSINCHU, TAIWAN

31040

國籍：(中文/英文)

中華民國/R.O.C.

三、發明人：(共4人)

姓名：(中文/英文)

1. 吳宗憲/WU, CHUNG-HSIEN

2. 賴育昇/LAI, YU-SHENG

3. 黃建霖/HUANG, CHIEN-LIN

4. 康家豪/KANG, CHIA-HUA

國籍：(中文/英文)

1. 中華民國/R.O.C.

I312945

2. 中華民國/R.O.C.
3. 中華民國/R.O.C.
4. 中華民國/R.O.C.

十、申請專利範圍：

1. 一種管理多媒體資料的方法，該方法包含下列步驟：
將被擷取註解之聲音特徵透過一自動語音辨識器產生一辨識結果；
轉換該辨識結果為一模擬影像資料，並利用一影像特徵處理方法來建立一索引資料庫；以及
在搜尋時，以自然口述語音搜尋該多媒體資料。
2. 如申請專利範圍第 1 項所述之管理多媒體資料的方法，其中該模擬影像資料的轉換包括下列步驟：
計算出該辨識結果之相似度分數；以及
根據此相似度分數，將該辨識結果轉換成該模擬影像資料。
3. 如申請專利範圍第 1 項所述之管理多媒體資料的方法，其中該索引資料庫的建立包括下列步驟：
擷取該模擬影像資料的特徵參數；以及
量化該特徵參數及完成該索引資料庫的建立。
4. 如申請專利範圍第 1 項所述之管理多媒體資料的方法，其中該自然口述語音搜尋該多媒體資料包括下列步驟：
以聲音輸入搜尋條件，並轉換該搜尋條件成一索引資料；
然後比對該索引資料與該索引資料庫裡的索引資訊；
以及
找出符合該搜尋條件之相對應的該多媒體資料。

5. 如申請專利範圍第 3 項所述之管理多媒體資料的方法，其中該擷取該模擬影像資料的特徵參數更包括：
從該辨識結果選出 n 名最佳音節候選者轉換成該模擬影像資料，並以一名音節候選者為一張影像將該模擬影像資料作切割， n 為自然數；以及
對切割後的影像資料再作影像特徵值分析。
6. 如申請專利範圍第 5 項所述之管理多媒體資料的方法，其中該量化該特徵參數及完成該索引資料庫的建立更包括：
量化該影像特徵值分析所得的音節模擬影像特徵值為一個多維度向量資料；以及
將相似的該影像特徵值資料聚成同一串碼字序列，所有碼字序列則建立成一碼本。
7. 如申請專利範圍第 1 項所述之管理多媒體資料的方法，其中該自動語音辨識器為一種以發音音節為基礎之語音辨識器。
8. 如申請專利範圍第 1 項所述之管理多媒體資料的方法，其中該影像特徵處理方法係對該模擬影像資料作主成分分析。
9. 一種管理多媒體資料的裝置，該裝置包含：
一多媒體資料製作索引模組，將被擷取註解之聲音特徵透過一自動語音辨識器產生一辨識結果，然後轉換該辨識結果為一模擬影像資料，並利用一影像特徵處理方法來完成建立一索引資料庫；以及

一搜尋此多媒體資料模組，以自然口述語音搜尋該多媒體資料。

10. 如申請專利範圍第 9 項所述之管理多媒體資料的裝置，其中該多媒體資料索引製作模組更包括：

該自動語音辨識器，辨識該被擷取註解之聲音特徵，產生該辨識結果；

一影像模擬模組，計算出該辨識結果之一相似度分數，然後根據該相似度分數，將該辨識結果轉換成一模擬影像資料；以及

一特徵擷取與索引建立模組，擷取該模擬影像資料的特徵參數，然後將該特徵參數作量化並完成該索引資料庫的建立。
11. 如申請專利範圍第 9 項所述之管理多媒體資料的裝置，其中該索引資料庫內每一索引係由一串碼字序列建立而成。
12. 如申請專利範圍第 9 項所述之管理多媒體資料的裝置，其中該自動語音辨識器為一種以發音音節為基礎之語音辨識器。
13. 如申請專利範圍第 12 項所述之管理多媒體資料的裝置，其中該模擬影像資料係經由該辨識結果選出 n 名最佳之該發音音節轉換成， n 為自然數。