

發明專利說明書

(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：95136283

※ 申請日期：95.9.29

※IPC 分類：G10L11/00

一、發明名稱：(中文/英文)

智慧型音訊處理器及其方法

INTELLIGENT CLASSIFICATION OF SOUND SIGNALS WITH
APPLICAATION AND METHOD

二、申請人：(共 1 人)

姓名或名稱：(中文/英文)

國立交通大學/NATIONAL CHIAO TUNG UNIVERSITY

代表人：(中文/英文)(簽章)張俊彥/CHANG CHUN-YEN

住居所或營業所地址：(中文/英文)

新竹市大學路 1001 號

No. 1001 Dasyue Road, Hsinchu, Taiwan, R.O.C.

國 籍：(中文/英文) 中華民國/TW

三、發明人：(共 2 人)

姓 名：(中文/英文)

白明憲/BAI MINGSIAN R

陳孟君/CHEN MENG-CHUN

國 籍：(中文/英文)

中華民國/TW

中華民國/TW

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。

五、中文發明摘要：

一種智慧型音訊處理器，包括：一特徵擷取單元接收一音訊訊號，並使用數個音訊描述子對音訊訊號擷取數個特徵值；一資料預處理單元對數個特徵值正規化，以作為智慧型音訊處理系統之分類資訊；以及一分類演算單元依分類資訊將音訊訊號分類成數種不同種類的音樂。

六、英文發明摘要：

A device that integrates various intelligent classification and preprocessing algorithms is provided. A feature extracting unit receives audio signals and extracts audio features for identification by using various descriptors; a preprocessing unit normalized the data for data consistency; a classification unit classifying audio signals into several categories according the audio features.

七、指定代表圖：

(一)、本案代表圖為：第 1 圖

(二)、本案代表圖之元件代表符號簡單說明：

10	智慧型音訊處理器
11	特徵擷取單元
12	資料預處理單元
13	分類演算單元
14	儲存裝置

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

九、發明說明：

【發明所屬之技術領域】

本發明係有關一種智慧型音訊處理器及其方法，特別是一種音訊的分類以及音訊的前處理器及其處理方法。

【先前技術】

目前，網路下載音樂盛行，各種音樂在網路流傳快速，為數愈來愈多的各種不同音樂存放在資料庫或硬碟中，一般在數量不多時，通常以人工整理分類多數的音樂檔，但是當數量增加到大量時，分類的工作便成為一種耗費人力的工作，尤其還必須依賴具有專業音樂技能人。是故，音樂和歌曲的分類工作愈來愈形重要。

習知，在音訊特徵擷取上，是藉由線性預估係數，梅爾頻率倒頻譜係數等方法，此類方法大多是在頻域上擷取到的特徵，無法充分表現出音訊的特性。

另外，在資訊分類上，類神經網路、模糊類神經網路、最近鄰居法則及隱藏式馬可夫模型應用於影像辨識上，可以有效的辨識影像內容。

美國專利號「140662」揭露一種以類神經網路為架構的國語語音撥號裝置，其使用類神經網路作為語音辨認，用於汽車電話中的語音撥號電路。音訊訊號之特徵擷取方法是利用線性預估係數法，無法全面性顯示語音訊號之特徵，尤其是與其它背景音樂混和時，其辨識會產生誤差。

美國專利號「US 5712953」揭露一種可分辨音訊屬於音樂或非音樂之系統，其音訊特徵擷取是以功率頻譜的分布作為區分之依據，其應用於廣泛之音樂或歌曲辨識時時會產生相當之誤差。

【發明內容】

為了解決音樂和歌曲的分類的問題，本發明之一實施例提供一種智慧型音訊處理器，其係利用頻率域、時域及統計值之特徵擷取方法，將音訊之特徵擷取出，取出聲音之特質，能更全方位代表聲音之特性。

為了解決音樂和歌曲的分類的問題，本發明之一實施例提供一種音訊分類處理方法，其係利用類神經網路、模糊類神經網路、最近鄰居法則及隱藏式馬可夫模型應用在歌者或樂器的辨識。利用辨識歌手姓名以自動將歌曲以歌手作為分類的標的，而音樂的辨識可以依照樂器的不同來做分類，使得整理音樂的工作變成相當容易。

為了解決音樂和歌曲的分類的問題，本發明之一實施例提供一種智慧型音訊處理器，其係利用獨立成份分析，以分離歌曲中的音樂部份，分離出之音樂可以作為音樂伴唱帶。另外本發明亦可以降低背景環境中之噪音，使得在吵雜環境中需要錄音時，突顯想要錄音之內容。

為達到上述目的，本發明之一實施例提供一種智慧型音訊處理器，包括：一特徵擷取單元接收一音訊訊號，並使用數個音訊描述子對音訊訊號擷取數個特徵值；一資料預處理單元對數個特徵值正規化，以作為智慧型音訊處理器之分類資訊；以及一分類演算單元依分類資訊將音訊訊號分類成數種不同種類的音樂。

另外，本發明之一實施例提供一種音訊分類處理方法，包括：一音訊分類器接收一第一音訊訊號，對第一音訊訊號擷取出第一組音訊特徵參數；對第一組音訊特徵參數正規化，以做為音訊分類器之數個分類項目；接收一第二音訊訊號，對第二音訊訊號擷取出第二組音訊特徵參數；對第二組音訊特徵參數正規化，以計算出分類資訊；以及使用人工智慧演算將第二音訊訊號分類至分類項目，並儲存至資料庫。

【實施方式】

第1圖為根據本發明之一實施例之智慧型音訊處理器之架構示意圖。一特徵擷取單元 11 接收音訊訊號，其使用數個音訊描述子對音訊訊號擷取數個特徵值。特徵擷取單元 11 可在頻率域、時域及統計值上擷取音訊訊號之特徵值；其中，在處理頻率域之特徵時，所用之計算法包括：線性預期編碼 (Linear Predictive Coding, LPC)、梅爾倒頻譜係數 (Mel-scale Frequency Cepstral Coefficients, MFCC)、響度 (loudness)、音高 (pitch)、自相關 (autocorrelation)、音訊頻譜重心 (Audio Spectrum Centroid)、音訊頻譜由重心決定的程度 (Audio Spectrum Spread)、音訊頻譜平坦程度 (Audio Spectrum Flatness)、音訊頻譜波封 (Audio Spectrum Envelope)、諧音頻譜重心 (Harmonic Spectral Centroid)、諧音頻譜偏差 (Harmonic Spectral Deviation)、諧音頻譜重心決定的程度 (Harmonic Spectral Spread) 及諧音頻譜變異 (Harmonic Spectral Variation)；另外，在處理時域之特徵時，所用之計算法包括：對數出擊時間 (log attack time)、節拍重心 (Temporal Centroid) 及過零率 (Zero Crossing Rate)；再者，在處理統計上之特徵時，所用之計算法包括偏態 (skewness) 及峰度 (kurtosis)。

一資料預處理單元 12 對特徵值做正規化計算，以作為智慧型音訊處理器 10 之分類資訊。

一分類演算單元 13 依分類資訊將音訊訊號分類成數種不同種類的音樂，分類演算單元 13 依類神經網路 (Artificial Neural Networks)、模糊類神經網路 (Fuzzy Neural Networks)、最近鄰居法則 (Nearest Neighbor Rule) 及隱藏式馬可夫模型 (Hidden Markov Models) 分類該音訊訊號。

根據上述，本發明用於音訊分類上，可作為歌手辨識及樂器之辨識。首先，輸入音樂訊號，利用特徵擷取方法擷取出音訊特徵，對

音訊特徵參數正規化作為音訊分類處理器之輸入，利用這些已知之輸入訓練辨識系統，訓練完成後以做為音訊分類之數個分類項目。

第 2 圖為根據本發明之一實施例之類神經網路架構示意圖。本發明之分類演算單元 13 所使用之類神經網路分為三層，第一層是輸入層 21，第二層是隱藏層 22，第三層是輸出層 23。輸入層 21 之輸入，即 $x_1 \dots x_i \dots x_{N_x}$ ，是正規化後的參數值，經過不同權重 ($v_{11} \dots v_{N_x N_x}$) 的加權及各節點之函式 ($g_1 \dots g_h \dots g_{N_x}$) 運算後可以得到隱藏層 22 之數值，即 $z_1 \dots z_h \dots z_{N_x}$ ，再經過不同權重 ($w_{11} \dots w_{N_x N_x}$) 的加權後於輸出層 23 經由函式 ($f_1 \dots f_0 \dots f_{N_y}$) 運算可以得到輸出值，即 $y_1 \dots y_0 \dots y_{N_y}$ ，輸出值和目標值的差利用倒傳遞演算法調整權重值，直到輸出和所設定之目標值相近時才停止。

第 3 圖為根據本發明之一實施例之模糊類神經網路架構示意圖。本發明之分類演算單元 13 所使用之模糊類神經網路分為五層，第一層是輸入層 31，第二層是歸屬度函數層 32，第三層是規則層 33，第四層是隱藏層 34，第五層是輸出層 35。輸入層 31 之輸入是正規化後的參數值，經過高斯歸屬度函數模糊化後可以得到歸屬度函數層 32，歸屬度函數層 32 再經由規則化後可以得到規則層 33，規則層 33 經過不同權重的加權後可以得到隱藏層 34，隱藏層 34 再經過不同權重的加權後可以得到輸出層 35，輸出值和目標值的差用於調整權重值，直到輸出和所設定之目標值相近時才停止。

第 4 圖為根據本發明之一實施例之最近鄰居法則之步驟示意圖。將訓練資料經過特徵擷取 S41 後，標示類別 S42，再將測試訊號經過特徵擷取 S43，計算測試資料與訓練資料分別的距離 S44，距離的的估算利用歐幾里得距離表示，將測試訊號的類別歸類至與其最近的點同一類別 S45。

第 5 圖為根據本發明之一實施例之隱藏式馬可夫模型之處理步驟示意圖。本發明使用隱藏式馬可夫模型之隨機過程，稱為觀測序列，

將訓練資料經過特徵擷取 S51 後，利用波氏演算法 (Baum-Welch method) 估算出隱藏式馬可夫模型，每一種特徵建立一種隱藏式馬可夫模型 S52，並產生隱藏式馬可夫模型資料庫 S53，再將測試訊號特徵擷取 S54 作為新的觀測序列，利用維特比 (Viterbi algorithm) 演算法 S55 計算出狀態觀測序列，最後計算資料庫中各種模型得到此觀測序列的機率，機率最大的就是最適合描述此觀測序列的模型，以分類儲存 S56 至一資料庫。

本發明用於測試三個不同的男歌手 (伍思凱、林志炫、阿杜)，訓練歌曲分別使用三人專輯中的六首不同之歌曲，而測試歌曲是不同於這六首之歌曲、也就是外部測試，得到之結果如表一所示：

分類方法	成功偵測機率
最近鄰居法則	64%
類神經網路	90%
模糊類神經網路	94%
隱藏式馬可夫模型	89%

表一

本發明用於測試四種不同的樂器 (小提琴、中提琴、大提琴、低音大提琴)、訓練歌曲和測試歌曲是同一首歌曲之不同部分、也就是內部測試，得到之結果如表二所示：

分類方法	成功偵測機率
最近鄰居法則	100%
類神經網路	98%
模糊類神經網路	99%
隱藏式馬可夫模型	100%

表二

對歌手識別而言，類神經網路、模糊類神經網路及隱藏式馬可夫模型可以識別歌手，達到良好之分類效果。

對樂器識別而言，類神經網路、模糊類神經網路、最近鄰居法則及隱藏式馬可夫模型可達到良好之識別及分類效果。

另外，在音訊前處理部份更包括一獨立成份分析單元，其可將歌曲中的人聲以及背景音樂分別獨立出來，或用於萃取背景噪音中之語音訊號，分離出數個音樂成份，最後輸入至特徵擷取單元。

綜合上述，本發明利用特徵擷取方法將擷取到之特徵參數正規化後做為輸入，利用這些輸入訓練辨識系統，訓練完成後便具有分類之功用，再將欲分類之音樂訊號輸入，便可將音樂做分類整理。歌曲部份依照歌手不同來區分，而純音樂之部份可依照樂器之不同而區分，如此使得整理音樂的工作變成相當容易。

以上所述之實施例僅係為說明本發明之技術思想及特點，其目的在使熟習此項技藝之人士能夠瞭解本發明之內容並據以實施，當不能以之限定本發明之專利範圍，即大凡依本發明所揭示之精神所作之均等變化或修飾，仍應涵蓋在本發明之專利範圍內。

【圖式簡單說明】

第 1 圖為根據本發明之一實施例之智慧型音訊處理器之架構示意圖。

第 2 圖為根據本發明之一實施例之類神經網路架構示意圖。

第 3 圖為根據本發明之一實施例之模糊類神經網路架構示意圖。

第 4 圖為根據本發明之一實施例之最近鄰居法則之步驟示意圖。

第 5 圖為根據本發明之一實施例之隱藏式馬可夫模型之處理步驟示意圖。

【主要元件符號說明】

10	智慧型音訊處理器
11	特徵擷取單元
12	資料預處理單元
13	分類演算單元
14	儲存裝置
21	輸入層
22	隱藏層
23	輸出層
31	輸入層
32	歸屬度函數層
33	規則層
34	隱藏層
35	輸出層
S41	特徵擷取
S42	標示類別
S43	特徵擷取
S44	計算距離
S45	分類儲存
S51	特徵擷取
S52	建立隱藏式馬可夫模型
S53	產生隱藏式馬可夫模型資料庫
S54	特徵擷取
S55	維特比演算法
S56	分類儲存

十、申請專利範圍：

1. 一種智慧型音訊處理器，包括：

一特徵擷取單元，其係接收一音訊訊號，並使用數個音訊描述子對該音訊訊號擷取數個特徵值；

一資料預處理單元，正規化該些特徵值以取得數個分類資訊；及

一分類演算單元，其係依該分類資訊將該音訊訊號分類成數種不同種類的音樂。

2. 如請求項 1 所述之智慧型音訊處理器，更包括一獨立成份分析單元接收該音訊訊號、自該音訊訊號分離出數個音樂成份，以及將該些音樂成份輸入至該特徵擷取單元。

3. 如請求項 2 所述之智慧型音訊處理器，其中該音訊訊號為一第一聲波與第二聲波所混合之訊號。

4. 如請求項 3 所述之智慧型音訊處理器，其中該第一聲波為一生物體發出的聲音訊號。

5. 如請求項 4 所述之智慧型音訊處理器，其中該第二聲波為一樂器之混合訊號。

6. 如請求項 4 所述之智慧型音訊處理器，其中該第二聲波為一環境之雜訊。

7. 如請求項 1 所述之智慧型音訊處理器，其中該音訊訊號為一人的聲波與樂器之音波所混合之訊號。

8. 如請求項 7 所述之智慧型音訊處理器，其中該特徵擷取單元係在頻率域、時域及統計值上擷取該音訊訊號之特徵值。

9. 如請求項 8 所述之智慧型音訊處理器，其中該特徵擷取單元處理該頻率域之特徵時，利用之計算法包括：線性預期編碼、梅爾倒頻譜係數、響度、音高、自相關、音訊頻譜重心、音訊頻譜由重心決定的程度、音訊頻譜平坦程度、音訊頻譜波封、諧音頻譜重心、諧音頻譜偏差、諧音頻譜重心決定的程度或諧音頻譜變異或上述之組合。

10.如請求項 8 所述之智慧型音訊處理器，其中該特徵擷取單元在處理該時域之特徵時，利用之計算法包括：對數出擊時間、節拍重心或過零率或上述之組合。

11.如請求項 8 所述之智慧型音訊處理系統，其中該特徵擷取單元在處理該統計上之特徵時，利用之計算法包括偏態及峰度。

12.如請求項 1 所述之智慧型音訊處理器，其中該分類演算單元係依類神經網路、模糊類神經網路、最近鄰居法則及隱藏式馬可夫模型分類該音訊訊號。

13.一種音訊分類處理方法，包括：

一音訊分類器接收一第一音訊訊號，對該第一音訊訊號擷取出一第一組音訊特徵參數；

正規化該第一組音訊特徵參數以取得數個分類項目；

接收一第二音訊訊號，對該第二音訊訊號擷取出第二組音訊特徵參數；

正規化該第二組音訊特徵參數以取得數個分類資訊；及

使用人工智慧演算將該第二音訊訊號分類至該等分類項目，並儲存至資料庫。

14.如請求項 13 所述之音訊分類處理方法，更包括以獨立成份分析該第二音訊訊號，並自該第二音訊訊號中分離出數個音樂成份。

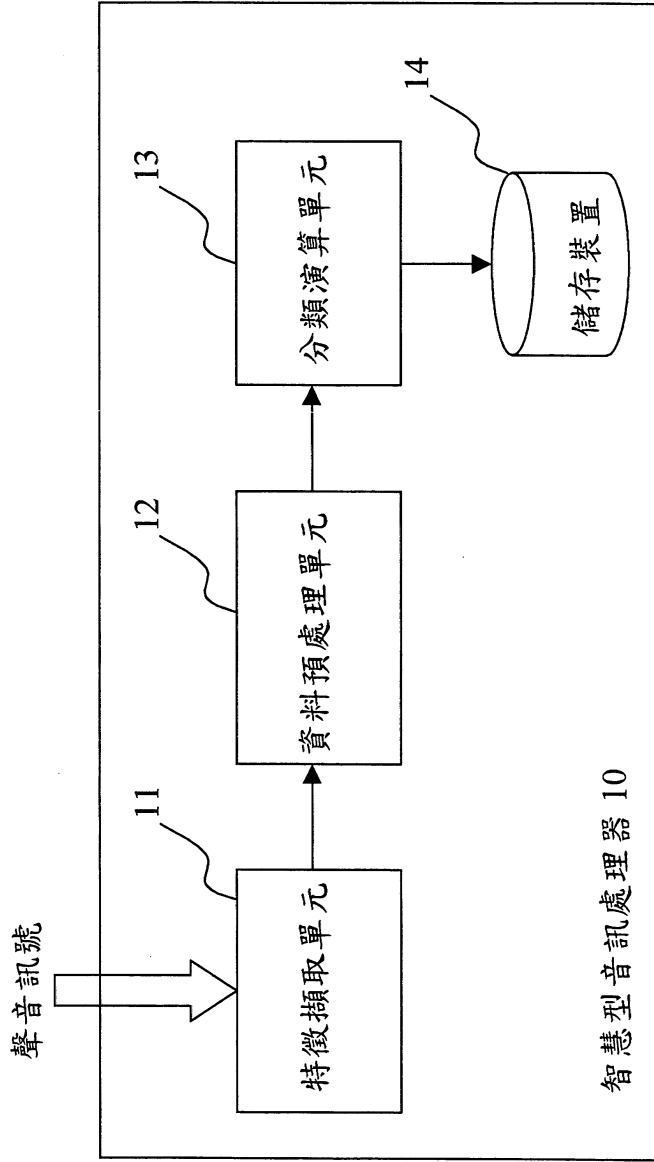
15.如請求項 13 所述之音訊分類處理方法，其中該第一音訊訊號為一欲分類之測試訊號，可產生該音訊分類器之分類項目。

16.如請求項 13 所述之音訊分類處理方法，其中該第二音訊訊號為數個聲波所混合之訊號。

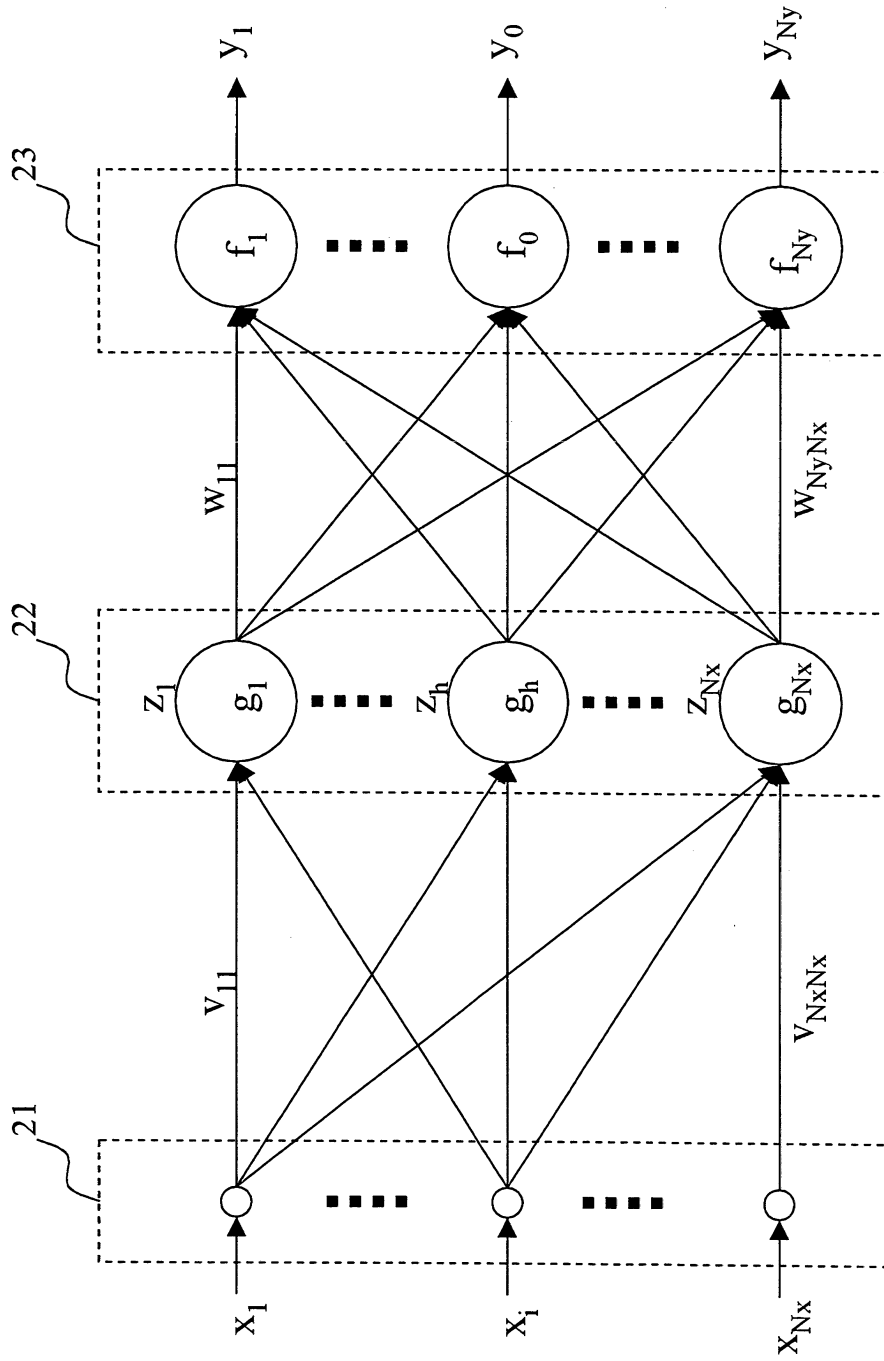
17.如請求項 13 所述之音訊分類處理方法，其中對該第一音訊訊號擷取出第一組音訊特徵參數，係在頻率域、時域及統計上擷取該音訊訊號之特徵。

18.如請求項 13 所述之音訊分類處理方法，其中對該第二音訊訊號擷取出第二組音訊特徵參數，係在頻率域、時域及統計上擷取該音訊訊號之特徵。

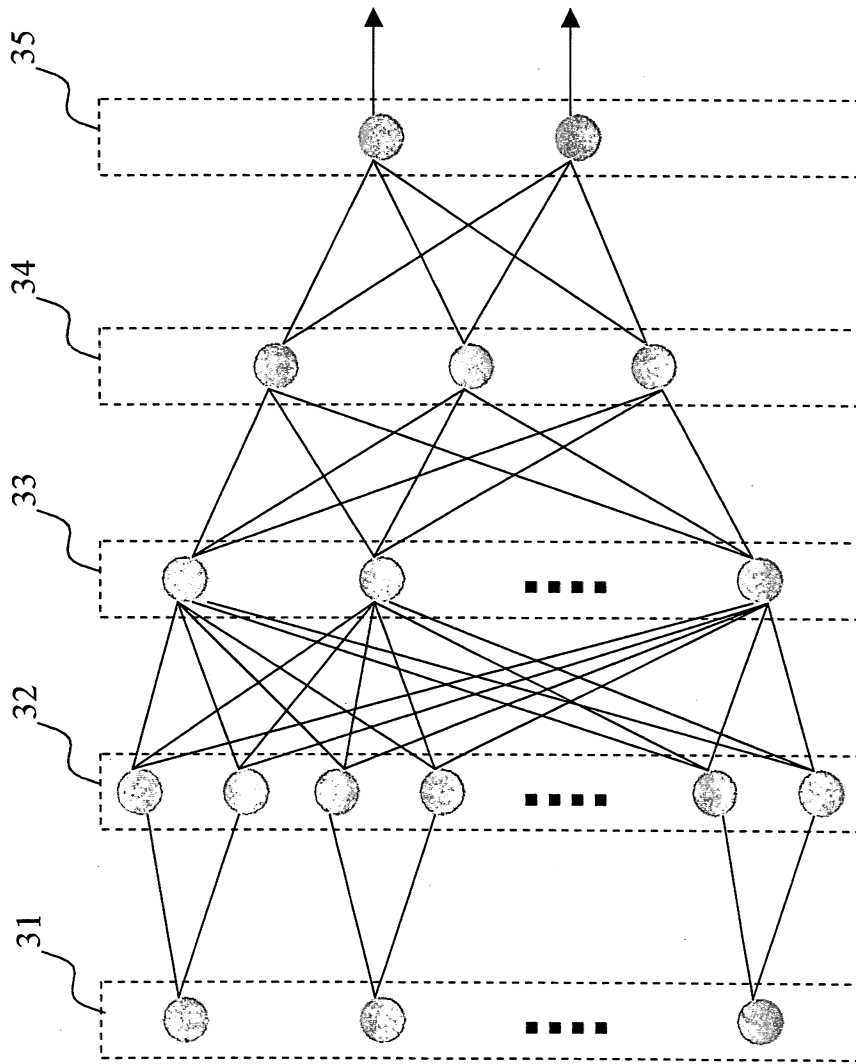
19.如請求項 13 所述之音訊分類處理方法，其中分類至該等分類項目係依類神經網路、模糊類神經網路、最近鄰居法則及隱藏式馬可夫模型分類該輸入訊號。



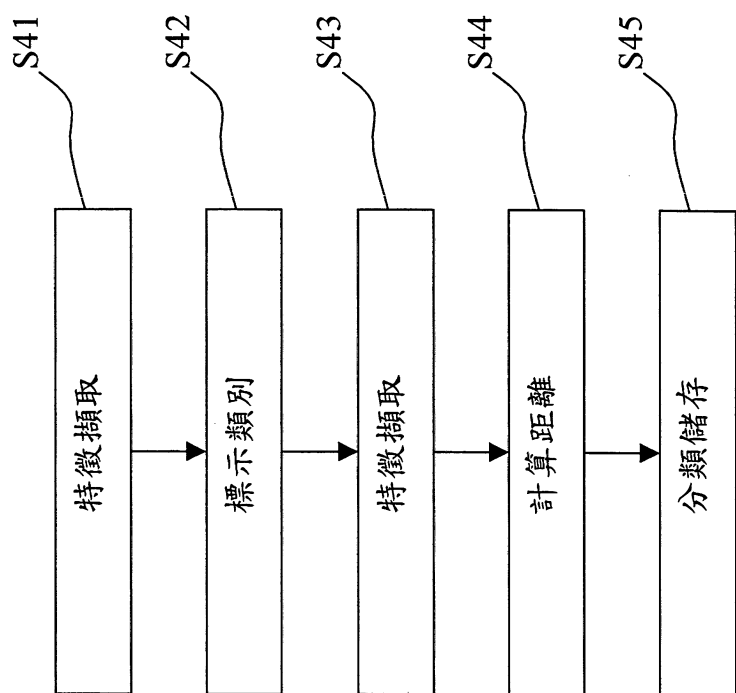
第1圖



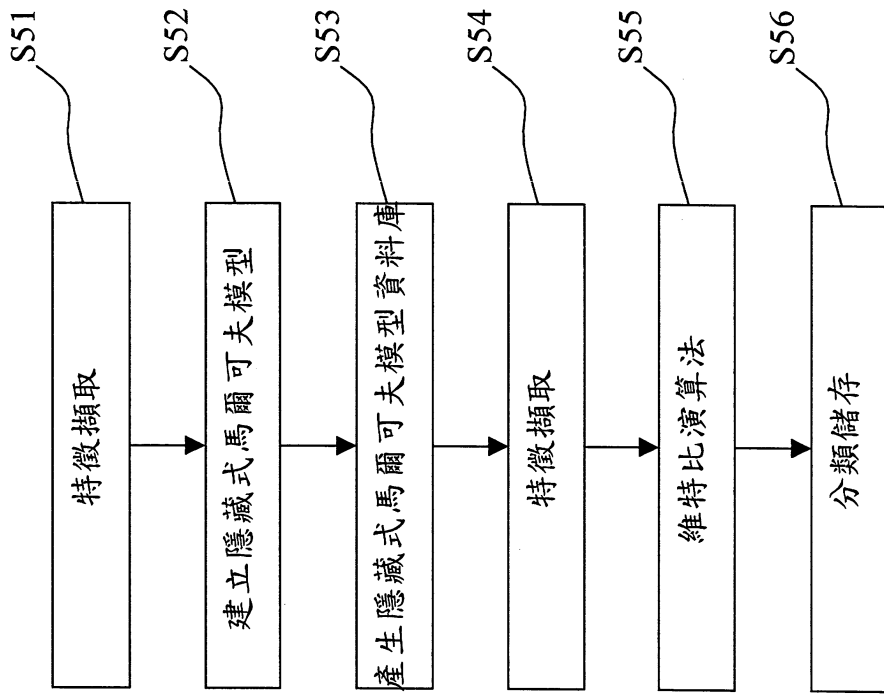
第2圖



第3圖



第4圖



第5圖