

公告本

申請日期	90.12.7
案 號	90130379
類 別	G10L 19/00

A4
C4

535141

(以上各欄由本局填註)

發明專利說明書		
一、發明 ~~新~~ 型名稱	中 文	健全語音分類之方法及裝置
	英 文	METHOD AND APPARATUS FOR ROBUST SPEECH CLASSIFICATION
二、發明 ~~創作~~ 人	姓 名	黃平仁 PENGJUN HUANG
	國 籍	中國
	住、居所	美國加州聖地牙哥市史普魯斯朗路11805-C
三、申請人	姓 名 (名 稱)	美商奎康公司 QUALCOMM INCORPORATED
	國 籍	美國
	住、居所 (事務所)	美國加州聖地牙哥市摩豪斯大道5775號
	代 表 人 名 姓 名	菲力普 R. 華德渥斯 PHILIP R. WADSWORTH

(由本局填寫)

承辦人代碼：
大 類：
I P C 分類：

A6
B6

本案已向：

國（地區） 申請專利，申請日期： 案號： ，☐有 ☐無主張優先權
美國 2000年12月08日 09/733,740 ☒有 ☐無主張優先權

有關微生物已寄存於： 寄存日期： ，寄存號碼：

裝
訂
線

五、發明說明(1)

發明背景

I. 發明範疇

揭示的具體實施例係有關於語音處理的領域。更明確而言，揭示的具體實施例係有關健全語音分類之一新改善方法及裝置。

II. 發明背景

藉由數位技術的聲音傳輸已廣泛使用，特別是在遠距離與數位無線電電話應用。其次，此已建立決定可在一頻道上傳送的最少資訊量，而維持重建語音的可感知品質。如果語音只藉由取樣及數位化傳輸，每秒64 kb (kbps)的資料率對於達成傳統類比電話的一語音品質是需要的。然而，藉著在適當編碼後的語音分析使用、傳輸、及接收器的再合成，資料率明顯降低是會發生。較正確語音分析執行，更正確資料可被編碼，如此便會減少資料率。

藉由擷取與人類語音產生的一模型有關參數而採用壓縮語音技術的裝置是稱為語音編碼器。一語音編碼器可將輸入的語音信號分成時間方塊、或分析訊框。語音編碼器典型包含一編碼器及一解碼器、或一寫解碼器。編碼器可分析輸入的語音訊框，以擷取某些相關參數，然後將參數量化成二進位表示，亦即，成為一組位元或二進位資料封包。資料封包可在通訊頻道傳送給一接收器及一解碼器。解碼器可處理資料封包，將他們解量化以產生參數，然後藉由使用解量化的參數而將語音訊框重新合成。

語音編碼器的功能可藉由移除語音的固有所有自然冗餘

五、發明說明(2)

而將數位化語音信號壓縮成一低位元率信號。數位壓縮可藉由使用一組參數表示輸入語音訊框、及採用量子化來表示一組位元的參數而達成。如果輸入語音訊框具有許多位元 N_1 ，而且由語音編碼器所產生的資料封包具許多位元 N_0 ，經由語音編碼器所達成的壓縮因素是 $C_1 = N_1/N_0$ 。此要求是要保有解碼語音的高聲音品質，而可達成目標壓縮因素。一語音編碼器的效率是因下列而定：(1)多好的語音模型，或上述分析與合成處理、執行的組合；及(2)參數量化處理少每訊框 N_0 位元的目標位元率執行程度。語音模型的目標是要捕捉語音信號的實體、或目標聲音品質、與每訊框的一小組參數。

語音編碼器能以時域編碼器實施，而可藉由採用高時間解析度處理每次將語音的小片段(典型5微秒(ms)子訊框)編碼而嘗試捕捉時域語音波形。對於每個子訊框而言，來自一編碼簿空間的高精密表示可經由在技藝中已知的各種不同搜尋演算法而找到。或者，語音編碼器能以頻域編碼器實施，以便使用一組參數(分析)來嘗試捕捉輸入語音訊框的短暫語音頻譜，及採用一對應合成處理，以便從頻譜參數重建語音波形。參數量化器可根據在 A. Gersho 和 R.M. Gray 所發表名稱“Quantization and Signal Compression (1992)”中描述的已知量子化技術而將參數以碼向量的儲存表示法表示而保存參數。

一眾所週知的時域語音編碼器是在由 L.B. Rabiner 和 R.W. Schafer 所發表名稱“Digital Processing of Speech Signals 396-

五、發明說明 (3)

453 (1978) 中描述的碼激發線性預測 (CELP) 編碼器，其在此僅列出供參考。在一 CELP 編碼器中，在語音信號的暫態關聯、或冗餘可藉由一線性預測 (LP) 分析移除，其中該線性預測分析可發現一暫態構形成分濾波器的係數。將短暫預測濾波器應用到輸入語音訊框可一 LP 殘餘信號，而可使用長期預測濾波器參數及一隨後的隨機編碼簿而進一步模型化及量化。因此，CELP 編碼可將時域語音波形編碼的工作分成 LP 暫態濾波器係數編碼及將 LP 殘餘編碼的分開工作。時域編碼能以一固定率 (亦即，對於每個訊框而言，可使用相同數目的位元 N_0)、或一可變率 (其中不同位元率可用於不同類型的訊框內容) 執行。可變率編碼器可嘗試只使用需要將寫解碼參數編碼成適於獲得一目標品質位準所需的位元量。一變數率 CELP 編碼器是在美國專利案號 5,414,796 中描述，其已轉讓給目前揭示的具體實施例，而且在此僅列出供參考。

例如 CELP 編碼器的時域編碼器典型係依賴每一訊框的大量位元 N_0 ，以保持時域語音波形的正確性。此編碼器典型可傳遞具相當大的每訊框位元量 N_0 的優良聲音品質 (例如，8 kbps 或以上)。然而，在低位元率 (4 kbps 及以下) 方面，時域編碼器不能保持高品質與健全效率，由於受限於可用的位元量。在低位元率上，受限的編碼簿空間可修剪能成功使用在高速率商業應用的傳統時域編碼器的波形匹配能力。

典型上，CELP 方法係採用一短暫預測 (STP) 濾波器及長

五、發明說明 (4)

期預測(LTP)濾波器。藉著合成(AbS)方法的一分析可使用在編碼器，以找到LTP延遲及最佳的隨機編碼簿增益與指標。例如增強可變率編碼器(EVRC)的目前最新發展CELP編碼器能以每秒大約8 kbps的一資料率達成良好品質的合成語音。

同時可知道無聲語音不會週期性出現。在傳統CELP方法中編碼LTP濾波器所消耗的頻寬並不如有聲語音可有效用於無聲語音，其中語音的週期性較強，而且LTP過濾是有意義。因此，一更有效率(亦即，較低位元率)編碼方法對於無聲語音是想要的。正確的語音分類對於選取最有效率的編碼方法是必需的，而且可達成最低的資料率。

對於在較低位元率編碼而言，語音的頻譜、或頻域、編碼的的各種不同方法已發展，其中語音信號可如同頻譜的時間變化發展而分析。可參考例如R.J. McAulay和T.F. Quatieri在Speech Coding and Synthesis第4章(W.B. Kleijn和K.K. Paliwal eds., 1995)的Sinusoidal Coding。在頻譜編碼器中，目的是要使用一組頻譜參數來模型化，或預測語音的每個輸入訊框的短暫語音頻譜，而不是正確模仿時間變化的語音波形。頻譜參數然後可編碼，而且語音的一輸出訊框可使用解碼的參數建立。結果的合成語音不匹配最初的輸入語音波形，但是可提供類似可感知的品質。在技藝中眾所周知的頻域編碼器範例包括多重頻帶激立編碼器(MBEs)，正弦轉換編碼器(STCs)、與調和編碼器(HCs)。此頻域編碼器可提供具有一小組參數的高品質參數模型，

五、發明說明(5)

其中該小組參數能以在低位元率可用的位元數量正確量化。

然而，低位元率編碼可使用在限制單編碼機構有效性的一受限編碼解析度的臨界限制、或一受限的編碼簿空間，以便在相等正確的各種不同背景情況下可提供不能表示各種不同類型語音片段的編碼器。例如，傳統低位元率頻域編碼器不能傳送語音訊框的相位資訊。或者，相位資訊可藉由使用一隨意人為產生的初始相位值與線性插入技術而重建。可參考H. Yang等人在Quadratic Phase Interpolation for Voiced Speech Synthesis in the MBE Model, in 29 Electronic Letters 856-57 (May 1993)的發表。因為相位資訊是人為產生，即使正弦的振幅可藉由量子化-解除量子化處理而保持，所以由頻域編碼器所產生的輸出語音將不能與最初輸入的語音(亦即，主要脈衝將不會同步)對齊。因此，證明不容易在頻域編碼器中採用任何閉迴路效率測量，例如，信號-雜訊比(SNR)、或可感知的SNR。

在低位元率可有效率將語音編碼的一有效技術是多重模式編碼。多重模式編碼技術已用來執行低率語音編碼、及一開迴路模式決定處理。此一多重模式編碼技術已在Amitava Dās等人發表的Multi-mode and Variable-Rate Coding of Speech, in Speech Coding and Synthesis ch. 7 (W.B. Kleijn & K.K. Paliwal eds., 1995)中描述。傳統多重模式編碼器可將不同模式、或編碼-解碼演算法應用到不同類型的輸入語音訊框。每個模式、或編碼-解碼處理可被自訂，而能能以

五、發明說明(6)

最有效率方式表示例如有聲語音、無聲語音、或背景雜訊(非語音)的某類型的語音片段。此多重模式編碼技術的成功是高度依賴正確的 mode 決定，或語音分類。一外部開迴路 mode 決定機構可檢查輸入語音訊框，並且決定何種 mode 可應用到訊框。開迴路 mode 決定係典型藉著從輸入訊框擷取許多參數、評估某暫時與頻譜特性的參數、及基於評估的一 mode 決定而執行。如此，mode 決定可達成，而無需預先知道輸出語音的正確情況，亦即，從聲音品質或其他效率測量的觀點輸出語音是如何接近輸入語音。一語音寫解碼器的開迴路 mode 決定是在美國專利案號 5,414,796 描述，其已讓與本發明之受讓人，而且在此僅列出供參考。

多重模式編碼可以是使用每訊框相同位元數量 N_0 的固定率、或可變率，其中不同位元率可用於不同 mode。可變率編碼的目標係只使用將寫解碼器參數編碼成獲得目標品質的一適當位準所需的位元量。結果，與一固定率相同的目標聲音品質，較高率編碼器能藉由使用可變位元率(VBR)技術而以一明顯較低平均率獲得。一可變率語音編碼器是在美國專利案號 5,414,796 描述。目前的研究興趣與強烈的商業需求在於發展可在中間到低位元率(亦即，2.4 到 4 kbps 及以下的範圍)上操作的一高品質語音編碼器。應用領域包括無線電話、衛星通訊、網際網路電話、各種不同多媒體及聲音流應用、語音郵件、及其他聲音儲存系統。推動力在封包損失情況下於高容量是需要，而且於健全效率是要求的。各種不同最近語音編碼標準化努力是用以推動低率

五、發明說明(7)

語音編碼演算法的研究與發展的另一直接推動力量。一低率語音編碼器可建立更多頻道、或使用者、允許的應用頻寬。一低率語音編碼器編碼器與適當頻道編碼的一額外層耦合可適合編碼器規格的所有位元安排，而且可在頻道錯誤情況下傳遞一健全效率。

多重模式VBR語音編碼因此是以低位元率編碼語音的一有效機構。傳統多重模式方法需要語音的各種不同片段(例如，無聲、有聲、暫態)的有效率編碼方法、或模式、及背景雜訊、或靜音的一模式設計。語音編碼器的整個效率是因模式分類的健全性及每個模式執行的狀況而定。編碼器的平均率是因語音的無聲、有聲、及其他片段的不同模式位元率而定。為了要以一低平均率達成目標品質，在條件改變下，需要正確決定語音模式。典型上，有聲與無聲語音片段能以高位元率捕捉，而且背景雜訊與靜音片段能以一相當低速率工作的模式表示。多重模式可變位元率編碼器需要正確的語音分類，以便藉由使用每訊框的最少位元數量而正確捕捉及編碼一高百分比的語音片段。更正確的語音分類可產生一較低平均的編碼位元率、與較高品質的解碼語音。先前的語音分類技術只考慮語音隔離訊框的最少參數量，而產生較少及不正確的語音模式分類。因此，在變化環境情況下需要可正確分類許多語音模式的一高效率語音分類器，為了要允許多重模式可變位元率編碼技術的最大效率。

發明概述

五、發明說明(8)

揭示的具體實施例是針對可評估語音的許多特性參數的一健全語音分類技術，以便在各種不同情況下可使用一高正確度將語音的各種不同模式分類。因此，在一觀點中，語音分類的一方法將揭示。該方法包括下列：將分類參數從外部元件輸入一語音分類器；在語音分類器中，從該等輸入參數之中至少一者產生內部分類參數；設定一正常化自動關聯係數功能臨界值；及根據一信號環境而選取一參數分析器；及分析輸入參數與內部參數，以產生一語音模式分類。

在另一觀點中，一語音分類器將揭示。該語音分類器包括下列：一產生器，用以從至少一外部輸入參數產生內部分類參數；一正常化關聯係數功能臨界值產生器，其可根據一信號環境而用以設定一正常化自動關聯係數功能臨界值、及選取一參數分析器；及一參數分析器，用以分析至少一外部輸入參數及內部參數，以產生一語音模式分類。

圖式之簡單說明

本發明的特徵、目的、與優點可從下列連同附圖的詳細描述而變得更顯然，其中相同的參考文字在圖中是表示類似元件：

圖1是語音編碼器在每端狀態結束的一通訊頻道方塊圖；

圖2是由圖1所示的編碼器所使用的一健全語音分類器方塊圖；

圖3係描述一健全語音分類器的語音分類步驟流程圖；

圖4A、4B、和4C是由揭示的具體實施例所使用有關語音

五、發明說明 (9)

分類的狀態圖；

圖5A、5B、和5C是由揭示的具體實施例所使用有關語音分類的決定表；及

圖6是具分類參數、及語音模式值的一語音信號具體實施例範例圖。

較佳具體實施例之詳細說明

揭示的具體實施例可提供用以改善聲碼器應用語音分類的方法及裝置。可分析新分類參數，以產生具比先前可用正確的更多語音分類。一新決策處理可在逐一訊框基礎上用來分類語音。從最初的輸入語音、SNR資訊、雜訊抑制輸出語音、聲音活動資訊、線性預測係數(LPC)分析、與開迴路音調評估的參數可由一以新狀態為基礎決定裝置用於語音的一正確分類各種不同模式。語音的每個訊框可藉由分析過去與未來訊構、及目前訊框而分類。經由揭示的具體實施例分類的語音模式包含暫態、主動語音的轉變、及字、聲音、無聲音、與靜音的結束。

揭示的具體實施例係在周圍雜訊變化位準的環境中提供各種語音模式的一語音分類技術。語音模式能以更有效率方式而更可靠及正確供編碼識別。

在圖1中，一第一編碼器10可接收數位化語音取樣 $s(n)$ 及編碼該等取樣 $s(n)$ ，用以在一傳輸媒體12、或通訊頻道12上傳輸給一第一解碼器14。編碼器14可將編碼的語音取樣解碼，並且合成一輸出語音信號 $s_{\text{SYNTH}}(n)$ 。對逆向傳輸而言，一第二編碼器16可編碼能在一通訊頻道18上傳輸的數

五、發明說明 (10)

位化語音取樣 $s(n)$ 。一第二解碼器20可接收及解碼該等編碼的語音取樣，而產生一合成輸出語音信號 $s_{\text{SYNTH}}(n)$ 。

語音取樣 $s(n)$ 係表示根據在技藝中包括例如脈衝碼調變(PCM)、壓擴 μ -律、或A-律的已知任何各種不同方法而數位化及量化的語音信號。如技藝所已知，語音取樣 $s(n)$ 可組成輸入資料訊框，其中每個訊框包含一預定數量的數位化語音取樣 $s(n)$ 。在一具體實施例中，8仟赫的一取樣比可採用，而且每20毫秒訊框包含160個取樣。在下述的具體實施例中，資料傳輸率會隨著不同訊框而從8 kbps (完全速率)到4 kbps (半速率)到2 kbps (四分之一速率)到1 kbps (八分之一速率)改變。或者，其他資料率可使用。如在此的使用，術語“完全速率”或“高速率”通常可視為大於或等於8 kbps的資料率，而且術語“半速率”或“低速率”通常可視為小於或等於4 kbps的資料率。改變資料傳輸率是有益的，因為較低位元率可選取性用於包含相當少語音資訊的訊框。如在技藝中熟諳此技者的瞭解，其他取樣率、訊框大小、與資料傳輸率可使用。

第一編碼器10與第二解碼器20係同時包含一第一語音編碼器、或語音寫解碼器。同樣地，第二編碼器16與第一解碼器14係同時包含一第二語音編碼器。在技藝中熟諳此技者可瞭解到語音編碼器能以一數位信號處理器(DSP)、一應用特殊積體電路(ASIC)、非連續閘邏輯、韌體、或任何傳統可程式軟體模組、及一微處理器實施。軟體模組可存在隨機存取記憶體(RAM)、快閃記憶體、暫存器、或在技藝

五、發明說明 (11)

中已知可寫儲存媒體的任何其他形式。或者，任何傳統處理器、控制器、或狀態機器可取代微處理器。為語音編碼特別設計的ASICs是在美國專案號5,727,123和5,784,532描述，其已讓與本發明之受讓人，而且在此僅列出供參考。

圖2係描述一健全語音分類器的具體實施例。在一具體實施例中，圖2的語音分類裝置可在圖1的編碼器(10、16)。在另一具體實施例中，健全語音分類器可單立式，用以提供輸出給例如圖1編碼器(10、16)裝置的語音分類模式。

在圖2中，輸入語音是提供給一雜訊抑制器(202)。輸入語音典型可由一聲音信號的類比-數位轉換產生。雜訊抑制器(202)可從輸入語音信號過濾雜訊元件，用以產生一雜訊抑制輸出語音信號、及目前輸出語音的SNR資訊。SNR資訊與輸出語音信號是輸入語音分類器(210)。雜訊抑制器(202)的輸出語音信號亦輸入聲音活動偵測器(204)、LPC分析器(206)、與開迴路輸入音調評估器(208)。SNR資訊可由語音分類器(210)使用，以週期性設定臨界值及在乾淨與雜訊語音之間區別。SNR參數以下稱為curr_ns_snr。輸出語音信號以下稱為t_in。在一具體實施例中，如果雜訊抑制器(202)不存在、或關閉，SNR參數curr_ns_snr應該預先設定成一預設值。

聲音活動偵測器(204)可將目前語音的聲音活動資訊輸出給語音分類器(210)。聲音活動資訊輸出係表示目前語音是否為主動或不活動。在一具體實施例中，聲音活動資訊輸出可以是二進位，亦即，主動或不活動。在另一具體實施

五、發明說明 (12)

例中，聲音活動資訊輸出可以是多重值。聲音活動資訊參數在此處稱為vad。

LPC分析器(206)可將目前輸出語音的LPC反映係數輸出給語音分類器(210)。LPC分析器(206)亦可輸出例如LPC係數的其他參數。LPC反映係數參數在此處稱為refl。

開迴路音調評估器(208)可將一正常化自動關聯係數功能(NACF)值、及音調值附近的NACF輸出給語音分類器(210)。NACF參數以下稱為nacf，而且在音調參數附近的NACF以下稱為nacf_at_pitch。一更週期性語音信號可產生nacf_at_pitch的一較高值。nacf_at_pitch的一較高值是更可能與一靜止聲音輸出語音類型有關。語音分類器(210)可維持一nacf_at_pitch的陣列值。nacf_at_pitch是在一子訊框的基礎上計算。在一具體實施例中，兩開迴路音調評估可藉由測量每訊框的兩子訊框而於輸出語音的每個訊框測量。nacf_at_pitch是從每個子訊框的開迴路音調評估計算。在具體實施例中，nacf_at_pitch值的五元素陣列(亦即nacf_at_pitch[5])包含輸出語音的二與二分之一訊框值。nacf_at_pitch陣列可於輸出語音的每個訊框更新。一nacf_at_pitch參數陣列的新使用可提供目前、過去、與前視(未來)信號資訊使用的語音分類器(210)，以便進行更正確與健全語音模式決定。

除了從外部元件輸入語音分類器(210)的資訊之外，語音分類器(210)可從輸出語音產生使用在語音模式決策處理的額外新參數。

五、發明說明 (13)

在一具體實施例中，語音分類器(210)可內部產生一零交錯率參數，以下簡稱zcr。目前輸出語音的zcr參數是定義成在語音每訊框的語音信號中符號改次數。在有聲語音，zcr值是較低，而無聲語音(或雜訊)具有一高zcr值，因為信號是非常任意。zcr參數可由語音分類器(210)用來分類有聲及無聲語音。

在一具體實施例中，語音分類器(210)可內部產生一目前訊框能量參數，以下簡稱E。E可由語音分類器(210)藉著將目前訊框能量與過去及未來訊框能量相比較而用來識別暫態語音。參數vEprev是從E取得的先前訊框能量E。

在一具體實施例中，語音分類器(210)可內部產生一前視訊框能量參數，以下簡稱Enext。Enext可包含來自一部分目前訊框的能量值及一部分的輸出語音的下一訊框。在一具體實施例中，Enext係表示在目前訊框的二分之一及在輸出語音下一訊框的前半部能量。Enext可由語音分類器(210)用來識別轉變語音。在語音結束，與目前訊框的能量相比較，下一訊框的能量會明顯降低。語音分類器(210)可將目前訊框能量與下一訊框能量相比較，以識別語音的結束及語音開始情況、或向上暫態及下暫態語音模式。

在一具體實施例中，語音分類器(210)可內部產生定義成 $\log_2(EL/EH)$ 的一頻帶能量比參數，其中EL是從0到2仟赫的低頻帶目前訊框能量，而且EH是從2仟赫到4仟赫的高頻帶目前訊框能量。頻帶能量比參數以下稱為bER。bER參數允許語音分類器(210)識別有聲語音及無聲語音模式，大體

五、發明說明 (14)

上，有聲語音係集中在低頻帶能量，而雜訊無聲語音係集中在高頻帶能量。

在一具體實施例中，語音分類器(210)可從輸出語音而產生三訊框平均聲音能量參數，以下簡稱 vE_{av} 。在其他具體實施例中， vE_{av} 可以是除了三之外的在許多訊框上的平均。如果目前語音模式是主動及有聲， vE_{av} 便可計算在輸出語音最後三個訊框中的能量平均。輸出語音的最後三訊框能量平均可在基礎語音模式決定上提供語音分類器(210)比獨自單訊框能量計算更穩定的統計。 vE_{av} 可由語音停止而目前訊框能量 E 與平均聲音能量 vE_{av} 相比較明顯降低時，音分類器(210)用來分類有聲語音的結束、或向下暫態模式。只有如果目前訊框是有聲的，或重新設定成無聲或非主動語音的一固定值， vE_{av} 便會更新。在一具體實施例中，固定的重新設定值是0.01。

在一具體實施例中，語音分類器(210)可內部產生一先前三訊框平均聲音能量參數，以下簡稱 vE_{prev} 。在其他具體實施例中， vE_{prev} 可以是除了三之外的許多訊框上的平均。 vE_{prev} 可由語音分類器(210)用來識別轉變語音。在語音開始上，目前訊框的能量與先前三個聲音訊框的平均能量相比較會明顯升高。語音分類器(210)可將目前訊框的能量與先前三個訊框相比較，以識別語音情況的開始、或向上暫態及語音模式。同樣地，在有聲語音結束上，目前訊框的能量會明顯消失。因此， vE_{prev} 亦可用來在語音結束上分類轉變。

五、發明說明 (15)

在一具體實施例中，語音分類器(210)可內部產生定義成 $10 \cdot \log_{10}(E/vE_{prev})$ 的一目前訊框能量與先前三訊框平均聲音能量比參數。在其他具體實施例中， vE_{prev} 可以是除了三之外的許多訊框上的平均。目前能量與先前三訊框平均聲音能量比參數以下稱為 vER 。 vER 可由語音分類器(210)用來分類有聲語音的開始、與有聲語音的結束、或向上暫態模式與下暫態模式，當語音重新開始時， vER 會較大，而且在有聲語音結束時， vER 會較小。 vER 參數可連同 vE_{prev} 參數使用在分類暫態語音。

在一具體實施例中，語音分類器(210)可內部產生定義成 $\text{MIN}(20, 10 \cdot \log_{10}(E/vE_{av}))$ 的一目前訊框能量與三訊框平均聲音能量比參數。目前訊框能量與三訊框平均聲音能量比以下稱為 $vER2$ 。 $vER2$ 可由語音分類器(210)用來分類在有聲語音結束時的暫態聲音模式。

在一具體實施例中，語音分類器(210)可內部產生一最大子訊框能量索引參數。語音分類器(210)可甚至將輸出語音的目前訊框分成數個子訊框，並且計算每個子訊框的均方根(RMS)能量值。在一具體實施例，目前訊框可分成10個子訊框。最大子訊框能量索引參數是在目前訊框、或在目前訊框第二半部中具有最大RMS能量值的子訊框索引。最大子訊框能量索引參數以下稱為 maxsfe_idx 。將目前訊框分成數個子訊框可在訊框內提供語音分類器(210)具有關峰值能量位置的資訊，包括最大峰值能量的位置。更大的解析度可藉著將一訊框分成更多子訊框而達成。雖然能量會

五、發明說明 (16)

在一暫態語音模式增加或逐漸減弱，但是當無聲或靜音語音模式的能量通常穩定時，maxsfe_idx可連同其他參數由語音分類器(210)用來分類暫態語音模式。

語音分類器(210)可使用從編碼元件直接輸入的新參數、及內部產生的新參數，以便比先前更正確及健全分類語音模式。語音分類器(210)可將一新決定處理應用到直接輸入及內部產生的參數，以產生改良的語音分類結果。決定處理是在下面參考圖4A-4C和5A-5C詳細描述。

在一具體實施例中，經由語音分類器(210)輸出的語音模式包含：暫態、向上暫態、向下暫態、有聲、無聲、與靜音模式。暫態模式是一有聲但是不是週期性語音，最佳是以完全速率CELP編碼。向上暫態模式在主動語音中是第一有聲訊框，最好是以完全速率CELP編碼。向下暫態模式典型是在一字結束的低能量有聲語音，最好是以半位元率CELP編碼。聲音模式是一高週期性有聲語音，主要是包含母音。聲音模式語音能以完全速率、半速率、四分之一速率、或八分之一速率編碼。用以編碼聲音模式語音的資料率選取可符合平均資料率(ADR)需求。主要包含子音的無聲音模式最好是以四分之一率雜訊激勵線性預測(NELP)編碼。靜音模式是不活動語音，最好是以八分之一速率CELP編碼。

在技藝中熟諳此技者可了解到參數與語音模式並未局限於揭示具體實施例的參數與語音。額外參數與語音模式可採用，而不致於違背揭示具體實施例的範圍。

五、發明說明(17)

圖3係描述一健全語音分類技術的語音分類步驟的一具體實施例流程圖。

在步驟300，從外部元件輸入的分類參數可於雜訊抑制輸出語音的每個訊框處理。在一具體實施例中，從外部元件輸入的分類參數包含從一雜訊抑制器元件輸入的curr_ns_snr和t_in、及從一開迴路音調評估器元件輸入的nacf和nacf_at_pitch參數、從一聲音活動偵測器元件輸入的vad、及從一LPC分析元件輸入的refl。控制流程執行步驟302。

在步驟302，額外內部產生的參數可從外部元件輸入的分類參數計算。在一具體實施例中，zcr、E、Enext、bER、vEav、vEprev、vER、vER2、和maxsfe_idx可從t_in計算。當在內部產生的參數於每個輸出語音訊框計算時，控制流成便執行步驟304。

在步驟304，NACF臨界值可被決定，而且一參數分析器可根據語音信號的環境而選取。在一具體實施例中，NACF臨界值可藉著將在步驟300輸入的curr_ns_snr參數與一SNR臨界值相比較而決定。從雜訊抑制器取得的curr_ns_snr資訊可提供一週期性決定臨界值的新適當控制。在此方式中，不同週期性臨界值可應用在具雜訊元件不同位準的語音信號分類處理。當最適當的nacf或語音信號雜訊位準的週期性臨界值於輸出語音的每個訊框選取時，一更正確的語音分類決定便可產生。決定一語音信號的最適當週期性臨界值允許語音信號的最佳參數分析器選取。

乾淨與雜訊語音信號的本質不同是週期性。當雜訊出現

五、發明說明 (18)

時，語音雜訊便會出現。當語音雜訊出現時，週期性測量、或nacf便會低於乾淨語音的週期性測量。因此，nacf臨界值便會於一雜訊環境降低補償，或於一乾淨信號環境提高。揭示具體實施例的新語音分類技術於所有環境並未週期性固定臨界值，產生一更正確與健全的模式決定，而與雜訊位準無關。

在一具體實施例中中，如果curr_ns_snr的值是大於或等於25分貝的一SNR臨界值時，乾淨語音的nacf臨界值可應用。乾淨語音的nacf臨界值可由下表定義：

類型的臨界值	臨界值名稱	臨界值
有聲	VOICEDTH	0.75
轉變	LOWVOICEDTH	0.5
無聲	UNVOICEDTH	0.35

表 1

在具體實施例中，如果curr_ns_snr的值小於25分貝的一SNR臨界值，雜訊語音的nacf臨界值便可應用。雜訊語音的nacf臨界值可由下表定義：

類型的臨界值	臨界值名稱	臨界值
有聲	VOICEDTH	0.65
轉變	LOWVOICEDTH	0.5
無聲	UNVOICEDTH	0.35

表 2

雜訊語音是與具增加雜訊的乾淨語音相同。隨著適當的週期性臨界值控制，健全語音分類技術可比先前更可能產生乾淨與雜訊語音的相同類決定。當nacf臨界值於每個訊構設定時，控制流程便執行步驟306。

裝
訂
線

五、發明說明 (19)

在步驟306，從外部元件輸入的參數及內部產生的參數可被分析，以產生一語音模式分類。根據信號環境所選取的一狀態機器、或分析的任何其他方法可運用在參數。在一具體實施例中，從外部元件輸入的參數及內部產生的參數可運用在圖4A-4C和5A-5C詳細描述的一以狀態為基礎的模式決定處理。決定處理可產生一語音模式分類。在一具體實施例中，暫態、向上暫態、向下暫態、有聲、無聲、或靜音的一語音模式分類可產生。當一語音模式決定產生時，控制流程便執行步驟308。

在步驟308，狀態變數與各種不同參數可更新，以包括目前訊框。在一具體實施例中， vE_{av} 、 vE_{prev} 、與目前訊框的有聲狀態可被更新。目前訊框能量 E 、 $nacf_at_pitch$ 、與目前訊框語音模式可更新用以分類下一訊框。

步驟300-308可於語音的每個訊框重複。

圖4A-4C係描述一健全語音分類技術具體實施例的模式決定處理的具體實施例。決定處理係基於語音訊框的週期性而選取語音分類的一狀態機器。對於語音的每個訊框而言，與週期性相容的一狀態機器、或語音訊框的雜訊元件可藉著將亦即 $nacf_at_pitch$ 值的語音訊框週期性測量與在圖3的步驟304中設定的NACF臨界值相比較而於決定處理選取。語音訊框的週期性位準係限制及控制模式決定處理的狀態轉變，以產生一更健全分類。

圖4A係描述當 vad 是1（有主動語音），而且 $nacf_at_pitch$ 的第三值（亦即 $nacf_at_pitch[2]$ ，零索引）非常高、或大於

五、發明說明 (20)

VOICEDTH的具體實施例中所選取的一狀態機器具體實施例。VOICEDTH是在圖3的步驟304中定義。圖5A係描述由每個狀態所評估的參數。

初始狀態是靜音。目前訊框將始終歸類為靜音，而不管先前狀態，如果vad=0（亦即沒有聲音活動）。

當先前狀態是靜音時，目前訊框便歸類為無聲或向上暫態。如果nacf_at_pitch[3]是非常低，zcr是高，bER是低，而且vER是非常低，或如果是上述這些條件的一組合符合，目前訊框便可歸類為無聲。否則，分類便會預設成向上暫態。

當先前狀態是無聲時，目前訊框便可歸類為無聲或向上暫態。如果nacf是非常低，nacf_at_pitch[3]是非常低，nacf_at_pitch[4]是非常低，zcr是高，bER是低，vER是非常低，而且E是小於vEprev，或如果是上述這些條件的一組合符合，目前訊框便可保持歸類為無聲。否則，分類會預設成向上暫態。

當先前狀態是有聲時，目前訊框便可歸類為無聲、暫態、向下暫態、或有聲。如果vER是非常低，而且E是小於vEprev，目前訊框可歸類為無聲。如果nacf_at_pitch[1]和nacf_at_pitch[3]是低，E是大於vEprev的一半，或上述這些條件的一組合符合，目前訊框便歸類為暫態。如果vER是非常低，而且nacf_at_pitch[3]具有一適度值，目前訊框便歸類為向下暫態。否則，目前分類便可預設成有聲。

當先前狀態是暫態或向上暫態時，目前訊框可歸類為無

五、發明說明 (21)

聲、暫態、向下暫態或有聲的。如果 vER 是非常低，而且 E 是小於 vE_{prev} ，目前訊框便歸類為無聲。如果 $nacf_at_pitch[1]$ 是低， $nacf_at_pitch[3]$ 具有一適度值， $nacf_at_pitch[4]$ 是低，而且先前狀態不是暫態，或如果上述這些條件的一組合符合，目前訊框可歸類為暫態。如果 $nacf_at_pitch[3]$ 具有一適度值，而且 E 是 0.05 倍小於 vE_{av} ，目前訊框便可歸類為向下暫態。否則，目前分類可預設成有聲。

當先前訊框是向下暫態時，目前訊框便可歸類為無聲、暫態、或向下暫態。如果 vER 是非常低，目前訊框將可歸類為無聲。如果 E 是大於 vE_{prev} ，目前訊框便可歸類為暫態。否則，目前分類便保持向下暫態。

圖 4B 係描述當 vad 是 1 (有主動語音)，而且 $nacf_at_pitch$ 的第三值是非常低，或小於 $UNVOICEDTH$ 時的具體實施例中所選取的一狀態機器具體實施例。 $UNVOICEDTH$ 是在圖 3 的步驟 304 中定義。圖 5B 係描述每個狀態所評估的參數。

初始狀態是靜音。如果 $vad=0$ (亦即沒有聲音活動)，目前訊框將始終歸類為靜音，而不管先前的狀態。

當先前狀態是靜音時，目前訊框便可歸類為或無聲或向上暫態。如果 $nacf_at_pitch[2-4]$ 出現增加趨勢， $nacf_at_pitch[3-4]$ 具有一適度值， zcr 是減到非常低， bER 是高，而且 vER 具有一適當值，或如果上述這些條件的一組合符合，目前訊框可歸類為向上暫態。否則，分類便會預設成無聲。

當先前狀態是無聲時，目前訊框可歸類為無聲或向上暫

五、發明說明 (22)

態。如果 $\text{naef_at_pitch}[2-4]$ 係出現增加趨勢， $\text{naef_at_pitch}[3-4]$ 具有一適度非常高， zcr 是適度非常低， vER 不是低， bER 是高， refl 是低， naef 具有適度值，而且 E 是大於 vEprev ，或如果上述這些條件的一組合符合，目前訊框便可歸類為向上暫態。這些條件的組合與臨界值便會因在參數 curr_ns_snr 中所反映語音訊框的雜訊位準而改變。否則，分類便會預設成無聲。

當先前狀態是有聲時、向上暫態、或暫態時，目前訊框便可歸類為無聲、暫態、或向下暫態。如果 bER 小於或等於零， vER 是非常低， bER 是大於零，而且 E 是小於 vEprev ，或如果上述這些條件的一組合符合，目前訊框可歸類為無聲。如果 bER 是大於零， $\text{naef_at_pitch}[2-4]$ 出現一增加趨勢， zcr 不是高， vER 不是低， refl 是低， $\text{naef_at_pitch}[3]$ 和 naef 是適度值，而且 bER 是小於或等於零，或如果上述這些條件的一組合符合，目前訊框便可歸類為暫態。這些條件的組合與臨界值會因在參數 curr_at_snr 中所反映語音訊框的雜訊位準而改變。如果 bER 是大於零， $\text{naef_at_pitch}[3]$ 是適度值， E 是小於 vEprev ， zcr 不是高，而且 vER2 是小於 -15，目前訊框可歸類為向下暫態。

當先前訊框是向下暫態時，目前訊框便可歸類為無聲暫態或向下暫態。如果 $\text{naef_at_pitch}[2-4]$ 出現增加趨勢， $\text{naef_at_pitch}[3-4]$ 是適度高， vER 不是低，而且 E 是兩倍大於 vEprev ，或如果上述這些條件的一組合符合，目前訊框將可歸類為暫態。如果 vER 不是低，而且 zcr 是低，目前訊框

五、發明說明 (23)

便將歸類為向下暫態。否則，目前分類便可預設成無聲。

圖 4C 係描述當 vad 是 1 (有主動語音)，而且 nacf_at_pitch 的第三值 (亦即 nacf_at_pitch[3]) 是適度值，亦即大於 UNVOICEDTH、及小於 VOICEDTH 的具體實施例中所選取的一狀態機器具體實施例。UNVOICEDTH 和 VOICEDTH 是在圖 3 的步驟 304 中定義。圖 5C 係描述由每個狀態所評估的參數。

初始狀態是靜音。如果 vad=0 (亦即沒有聲音活動)，目前訊框將始終歸類為靜音，而不管先前的狀態。

當先前狀態是靜音時，目前訊框便可歸類為無聲或向上暫態。如果 nacf_at_pitch[2-4] 顯示增加趨勢，nacf_at_pitch[3-4] 是適度高，zcr 不是高，bER 是高，vER 具有一適度值，zcr 很低，而且 E 是兩倍大於 vEprev，或如果上述這些條件的一組合符合，目前訊框便可歸類為向上暫態。否則，分類便預設成無聲。

當先前狀態是無聲，目前訊框可歸類為無聲或向上暫態。如果 nacf_at_pitch[2-4] 出現增加趨勢，nacf_at_pitch[3-4] 具有一適度非常高值，zcr 不是高，vER 不是低，bER 是高，refl 是低，E 是大於 vEprev，zcr 是很低，nacf 不是低，maxsfe_idx 指向最後子框架，而且 E 是兩倍大於 vEprev，或如果上述這些條件的一組合符合，目前訊框可歸類為向上暫態。這些條件的組合與臨界值可因在參數 curr_ns_snr 中所反映語音訊框的雜訊位準而改變。否則，分類便預設成無聲。

當先前狀態是有聲、向上暫態、或暫態時，目前訊框便

五、發明說明 (24)

可歸類為無聲、有聲、暫態、向下暫態。如果bER是小於或等於零，vER是非常低，E_{next}是小於E，nacf_at_pitch[3-4]是非常低，bER是大於零，而且E是小於vE_{prev}，或如果上述這些條件的一組合符合，目前訊框可歸類為無聲。如果bER是大於零，nacf_at_pitch[2-4]出現增加趨勢，zcr不是高，vER不是低，refl是低，nacf_at_pitch[3]和nacf不是低，或如果上述這些條件的一組合符合，目前訊框便可歸類為暫態。這些條件的組合與臨界值可因在參數curr_ns_snr中所反映語音訊框的雜訊位準而改變。如果bER是大於零，nacf_at_pitch[3]不是高，E是小於vE_{prev}，zcr不是高，vER是小於-15，而且vER₂是小於-15，或如果上述這些條件的一組合符合，目前訊框歸類為向下暫態。如果nacf_at_pitch[2]是大於LOWVOICEDTH，bER是大於或等於零，而且vER不是低，或如果上述這些條件的一組合符合，目前訊框可歸類為有聲。

當先前訊框是向下的-暫態時，目前訊框便可歸類為無聲、暫態、或向下暫態。如果bER是大於零，nacf_at_pitch[2-4]出現增加趨勢，nacf_at_pitch[3-4]是適度高，vER不是低，而且E是兩倍大於vE_{prev}，或如果上述這些條件的一組合符合，目前訊框便可歸類為暫態。如果vER不是低，而且zcr是低，目前訊框便可歸類為向下暫態。否則，目前分類便預設成無聲。

圖5A-5C是由語音分類的揭示具體實施例所使用決定表的具體實施例。

五、發明說明 (25)

圖 5A 係根據一具體實施例而描述當 naef_at_pitch 的第三值 (亦即 $\text{naef_at_pitch}[2]$) 是非常高、或大於 VOICEDTH 時，經由每個狀態所評估的參數、及狀態轉變。在圖 5A 描述的決定表可由圖 4A 描述的狀態機器使用。語音的先前訊框的語音模式分類是在最左欄顯示。當參數值是在與每個先前模式有關的列中顯示時，語音模式分類轉變成目前模式是在相關欄的頂端列中識別。

圖 5B 係根據一具體實施例而描述當第三值 (亦即 $\text{naef_at_pitch}[2]$) 是非常低時、或小於 UNVOICEDIE 時，經由每個狀態所評估的參數、與狀態轉變。在圖 5B 中描述的決定表是由描述的狀態機器使用，語音的先前訊框的語音模式分類是在最左欄顯示。當參數值是在與每個先前模式有關的列顯示時，語音模式分類轉變成目前模式是在相關欄的頂列中識別。

圖 5C 係根據一具體實施例而描述當 naef_at_pitch 的第三值 (即是 $\text{naef_at_pitch}[3]$) 是適度，亦即，大於 UNVOICEDTH 但是小於 VOICEDTH 時，經由每個狀態所評估的參數、與狀態轉變。在圖 5C 中描述的決定表可由在圖 4C 中描述的狀態機器使用。語音的先前訊框的語音模式分類是在最左欄顯示。當參數值是在與每個先前模式有關的列顯示時，語音模式分類轉變成目前模式是在相關欄的頂端列中識別。

圖 6 是與參數值與語音分類有關的一語音信號具體實施例的時間線圖。

在技藝中熟諳此技者可了解到語音分類器可使用一 DSP、

五、發明說明 (26)

一 ASIC、非連續開邏輯、韌體、或任何傳統可程式軟體模組、及一微處理器實施。軟體模組可存在RAM記憶體、快閃記憶體、暫存器、或在技藝中已知的可寫入儲存媒體的任何其他形式。或者，任何傳統處理器、控制器、或狀態機器可取代微處理器。

較佳具體實施例的先前描述的提供允許在技藝中熟諳此技者可製造或使用本發明。這些具體實施例的各種不同修改對於在技藝中熟諳此技者是顯然的，而且在此定義的一般原理可運用在其他具體實施例，而無需使用本發明之設施。因此，本發明並未侷限於在此顯示的具體實施例，而是符合在此揭示的原理及新特徵的廣泛範圍。

四、中文發明摘要(發明之名稱：健全語音分類之方法及裝置)

一種用以語音改變模式健全分類之語音分類技術，以允許多重模式可變位元率編碼技術的最大效率[5027-530]。一語音分類器可正確分類以最小位元率編碼之一高百分比的語音片段，而符合較低位元率需求。(該高度正確語音分類能產生一較低平均編碼位元率、與較高品質解碼語音)。該語音分類器係考慮語音每訊框的最大量參數，可於每個訊框產生許多及正確的語音模式分類。該語音分類器在可變環境情況下可正確地分類許多語音模式。該語音分類器可從外部元件輸入分類參數，從該等輸入參數產生內部分類參數，設定一正常化自動關聯係數功能臨界值、及根據該信號環境而選取一參數分析器，然後，將該等參數分析，以產生一語音模式分類。

英文發明摘要(發明之名稱：METHOD AND APPARATUS FOR ROBUST SPEECH CLASSIFICATION)

A speech classification technique [502-530] for robust classification of varying modes of speech to enable maximum performance of multi-mode variable bit rate encoding techniques. A speech classifier accurately classifies a high percentage of speech segments for encoding at minimal bit rates, meeting lower bit rate requirements. Highly accurate speech classification produces a lower average encoded bit rate, and higher quality decoded speech. The speech classifier considers a maximum number of parameters for each frame of speech, producing numerous and accurate speech mode classifications for each frame. The speech classifier correctly classifies numerous modes of speech under varying environmental conditions. The speech classifier inputs classification parameters from external components, generates internal classification parameters from the input parameters, sets a Normalized Auto-correlation Coefficient Function threshold and selects a parameter analyzer according to the signal environment, and then analyzes the parameters to produce a speech mode classification.

六、申請專利範圍

1. 一種語音分類之方法，其包含：

將分類參數從外部元件輸入一語音分類器；

在該語音分類器中，從該等輸入參數之中至少一者產生內部分類參數；

根據一信號環境而設定一正常化自動關聯係數功能臨界值，並且選取一參數分析器；及

分析該等輸入參數及該等內部參數，以產生一語音模式分類。

2. 如申請專利範圍第1項之方法，其中該等輸入參數包含一雜訊抑制語音信號。
3. 如申請專利範圍第1項之方法，其中該等輸入參數包含有關一雜訊抑制語音信號的信號-雜訊比資訊。
4. 如申請專利範圍第1項之方法，其中該等輸入參數包含聲音活動資訊。
5. 如申請專利範圍第1項之方法，其中該等輸入參數包含線性預測反映係數。
6. 如申請專利範圍第1項之方法，其中該等輸入參數包含正常化自動關聯係數功能資訊。
7. 如申請專利範圍第1項之方法，其中該等輸入參數包含音調資訊的正常化自動關聯係數功能。
8. 如申請專利範圍第7項之方法，其中音調資訊上的該正常化自動關聯係數功能是一陣列值。
9. 如申請專利範圍第1項之方法，其中該等內部參數包含一零交錯率參數。

六、申請專利範圍

10. 如申請專利範圍第1項之方法，其中該等內部參數包含一目前訊框能量參數。
11. 如申請專利範圍第1項之方法，其中該等內部參數包含一前視訊框能量參數。
12. 如申請專利範圍第1項之方法，其中該等內部參數包含一頻帶能量比參數。
13. 如申請專利範圍第1項之方法，其中該等內部參數包含三訊框平均聲音能量參數。
14. 如申請專利範圍第1項之方法，其中該等內部參數包含一先前三訊框平均聲音能量參數。
15. 如申請專利範圍第1項之方法，其中該等內部參數包含一目前訊框能量與先前三訊框平均聲音能量比參數。
16. 如申請專利範圍第1項之方法，其中該等內部參數包含一目前訊框能量與三訊框平均聲音能量比參數。
17. 如申請專利範圍第1項之方法，其中該等內部參數包含一最大子訊框能量索引參數。
18. 如申請專利範圍第1項之方法，其中該設定一正常化自動關聯係數功能臨界值包含將一信號雜訊比資訊參數與一預定信號雜訊比值相比較。
19. 如申請專利範圍第1項之方法，其中該分析包含將該等參數應用到一狀態機器。
20. 如申請專利範圍第19項之方法，其中該狀態機器包含每個語音分類模式的一狀態。
21. 如申請專利範圍第1項之方法，其中該語音模式分類包

六、申請專利範圍

含一暫態模式。

22. 如申請專利範圍第1項之方法，其中該語音模式分類包含一向上暫態模式。

23. 如申請專利範圍第1項之方法，其中該語音模式分類包含一向下暫態模式。

24. 如申請專利範圍第1項之方法，其中該語音模式分類包含一聲音模式。

25. 如申請專利範圍第1項之方法，其中該語音模式分類包含一無聲音模式。

26. 如申請專利範圍第1項之方法，其中該語音模式分類包含一靜音模式。

27. 如申請專利範圍第1項之方法，其進一步包含更新至少一參數。

28. 如申請專利範圍第27項之方法，其中該更新參數包含正在音調參數上的一正常化自動關聯係數功能。

29. 如申請專利範圍第27項之方法，其中該更新參數包含三訊框平均聲音能量參數。

30. 如申請專利範圍第27項之方法，其中該更新參數包含一前視訊構能量參數。

31. 如申請專利範圍第27項之方法，其中該更新參數包含一先前三訊框平均聲音能量參數。

32. 如申請專利範圍第27項之方法，其中該更新參數包含一聲音活動偵測參數。

33. 一種語音分類器，其包含：

六、申請專利範圍

一產生器，用以產生分類參數；

一正常化自動關聯係數功能，其可根據一信號環境而用以設定一正常化自動關聯係數功能臨界值，及選取一參數分析器；及

一參數分析器，用以分析至少一外部輸入參數及內部參數，以產生一語音模式分類。

34. 如申請專利範圍第33項之語音分類器，其中用以產生分類參數的該產生器可從一雜訊抑制語音語音產生參數。
35. 如申請專利範圍第33項之語音分類器，其中用以產生分類參數的該產生器可從信號到雜訊比資訊產生參數。
36. 如申請專利範圍第33項之語音分類器，其中用以產生分類參數的該產生器可從聲音活動資訊產生參數。
37. 如申請專利範圍第33項之語音分類器，其中用以產生分類參數的該產生器可從線性預測反映係數產生參數。
38. 如申請專利範圍第33項之語音分類器，其中用以產生分類參數的該產生器可從正常化自動關聯係數功能資訊產生參數。
39. 如申請專利範圍第33項之語音分類器，其中用以產生分類參數的該產生器可在音調資訊上從正常化自動關聯係數功能產生參數。
40. 如申請專利範圍第39項之語音分類器，其中在音調資訊上的該正常化自動關聯係數功能是一陣列值。
41. 如申請專利範圍第33項之語音分類器，其中該等產生的參數包含一零交錯率參數。

六、申請專利範圍

42. 如申請專利範圍第33項之語音分類器，其中該等產生的參數包含一目前訊框能量參數。
43. 如申請專利範圍第33項之語音分類器，其中該等產生的參數包含前視訊構能量參數。
44. 如申請專利範圍第33項之語音分類器，其中該等產生的參數包含一頻帶能量比參數。
45. 如申請專利範圍第33項之語音分類器，其中該等產生的參數包含三訊框平均聲音能量參數。
46. 如申請專利範圍第33項之語音分類器，其中該等產生的參數包含一先前三訊框平均聲音能量參數。
47. 如申請專利範圍第33項之語音分類器，其中該等產生的參數包含一目前訊框能量與先前三訊框平均聲音能量比參數。
48. 如申請專利範圍第33項之語音分類器，其中該等產生的參數包含一目前訊框能量與三訊框平均聲音能量比參數。
49. 如申請專利範圍第33項之語音分類器，其中該等產生的參數包含一最大子訊框能量索引參數。
50. 如申請專利範圍第33項之語音分類器，其中該設定一正常化自動關聯係數功能臨界值包含將一信號雜訊比資訊參數與一預定信號雜訊比值相比較。
51. 如申請專利範圍第33項之語音分類器，其中該分析包含將該等參數應用到一狀態機器。
52. 如申請專利範圍第51項之語音分類器，其中該狀態機器

六、申請專利範圍

包含每個語音分類模式的一狀態。

53. 如申請專利範圍第33項之語音分類器，其中該語音模式分類包含一暫態模式。
54. 如申請專利範圍第33項之語音分類器，其中該語音模式分類包含一向上暫態模式。
55. 如申請專利範圍第33項之語音分類器，其中該語音模式分類包含一向下暫態模式。
56. 如申請專利範圍第33項之語音分類器，其中該語音模式分類包含一聲音模式。
57. 如申請專利範圍第33項之語音分類器，其中該語音模式分類包含一無聲音模式。
58. 如申請專利範圍第33項之語音分類器，其中該語音模式分類包含一靜音模式。
59. 如申請專利範圍第33項之語音分類器，其進一步包含更新至少一參數。
60. 如申請專利範圍第59項之語音分類器，其中該更新的參數包含在音調參數上的一正常化自動關聯係數功能。
61. 如申請專利範圍第59項之語音分類器，其中該更新的參數包含三訊框平均聲音能量參數。
62. 如申請專利範圍第59項之語音分類器，其中該更新的參數包含一前視訊構能量參數。
63. 如申請專利範圍第59項之語音分類器，其中該更新的參數包含一先前三訊框平均聲音能量參數。
64. 如申請專利範圍第59項之語音分類器，其中該更新的參數包含一聲音活動偵測參數。

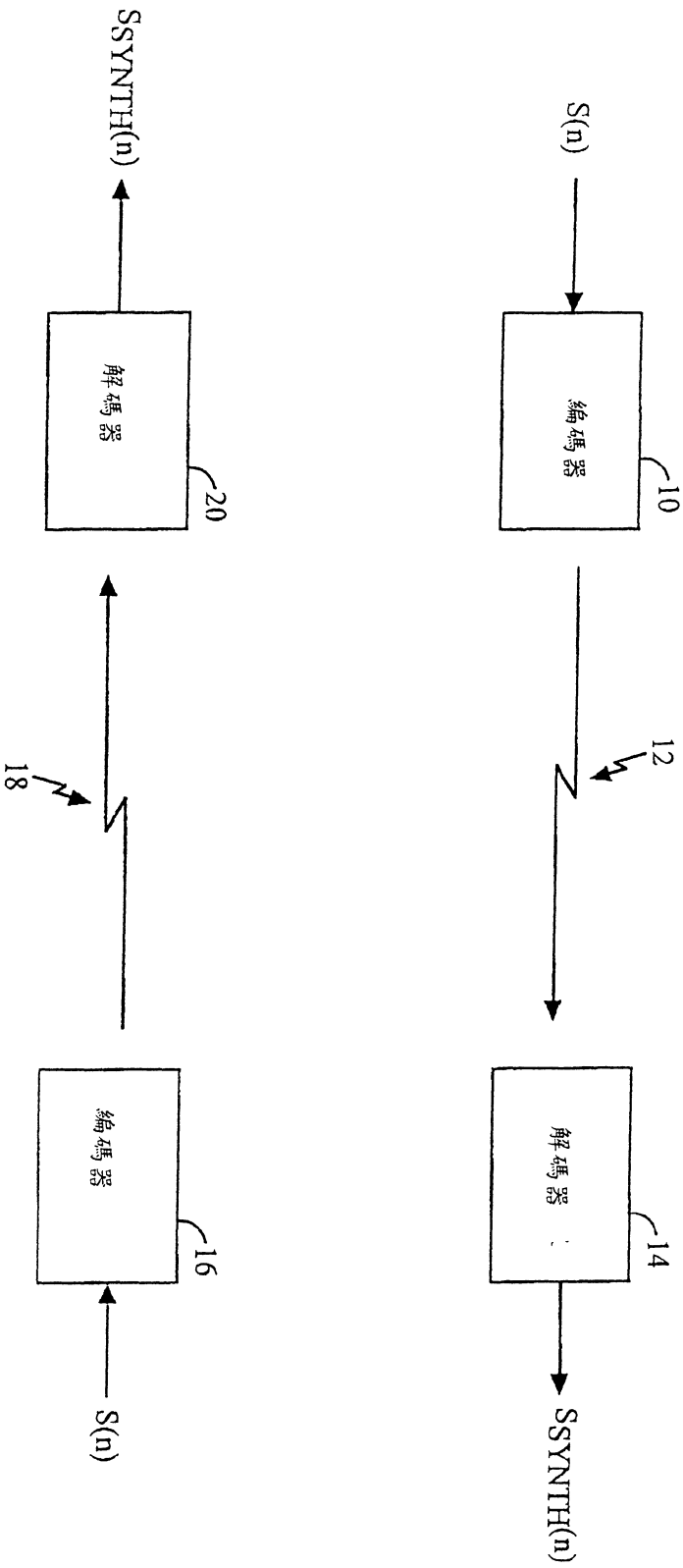


圖 1

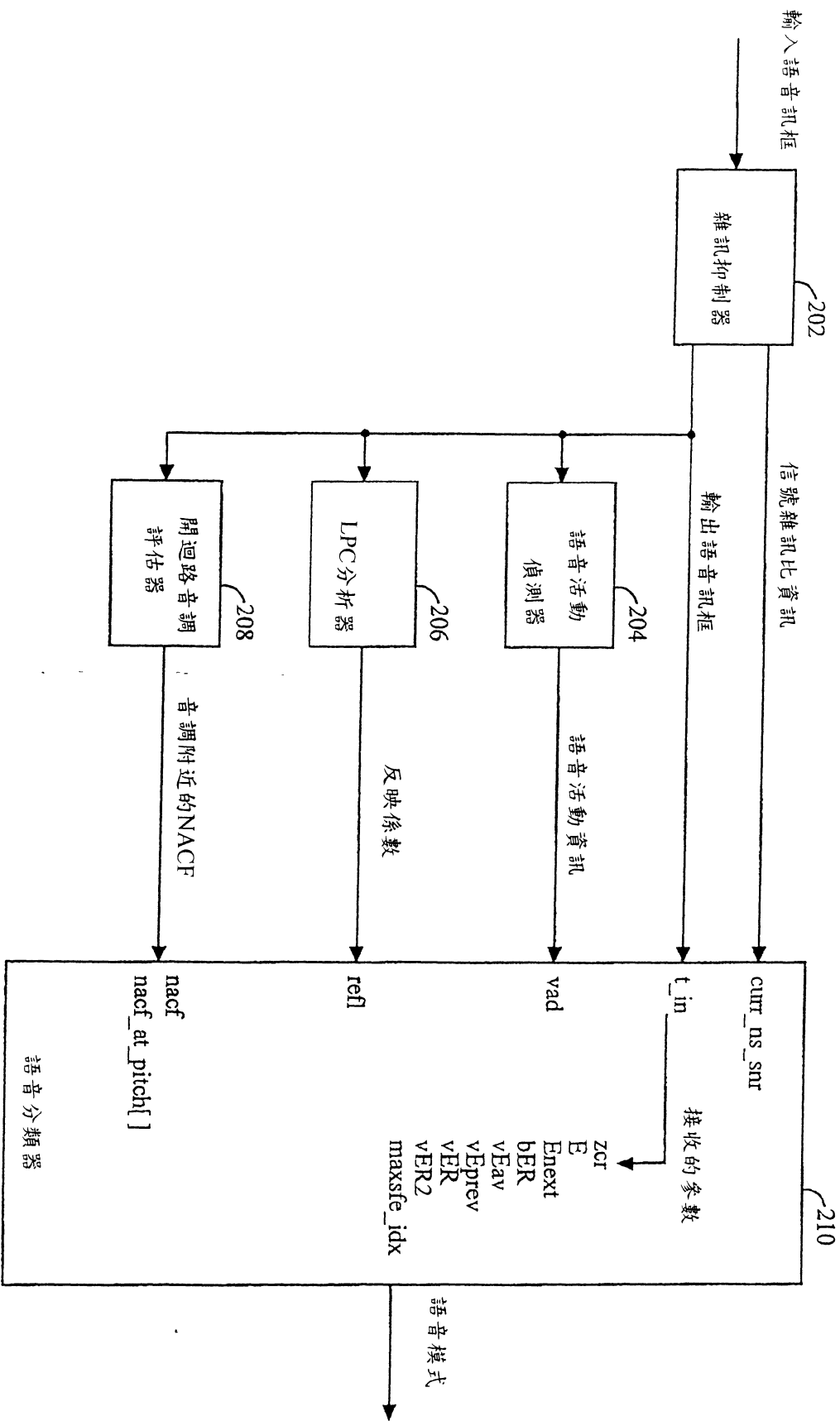


圖 2

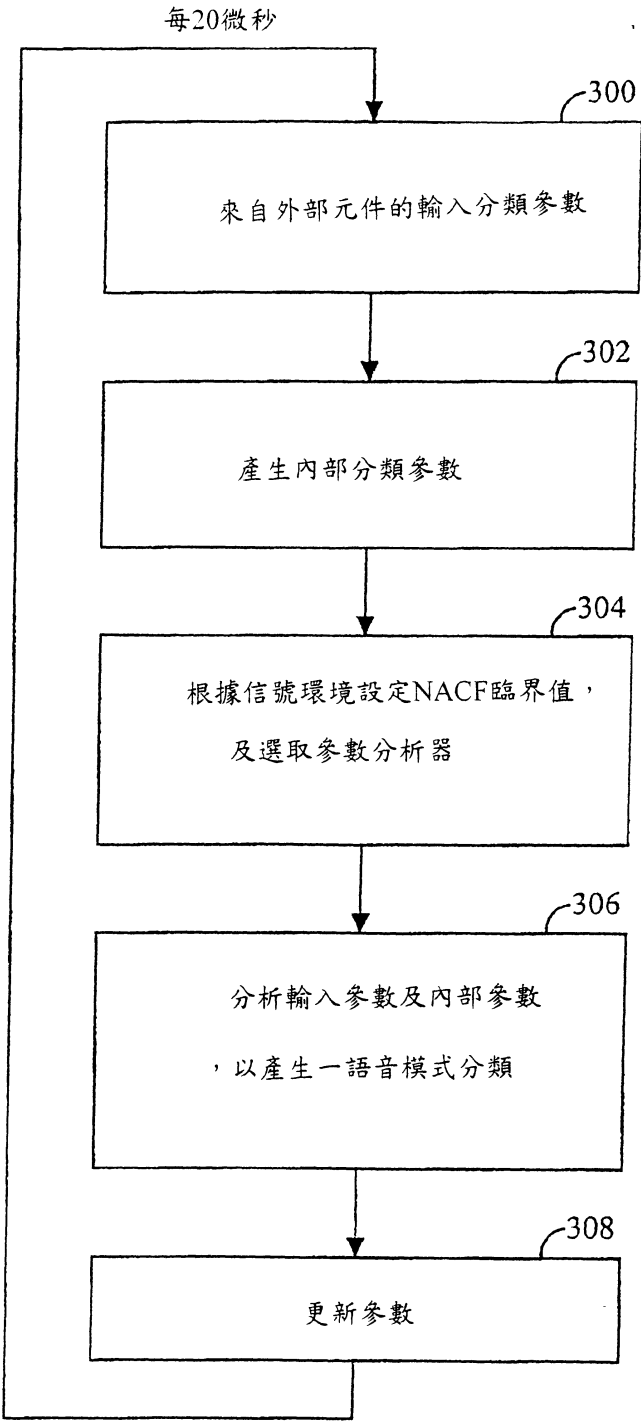


圖3

圖 4A

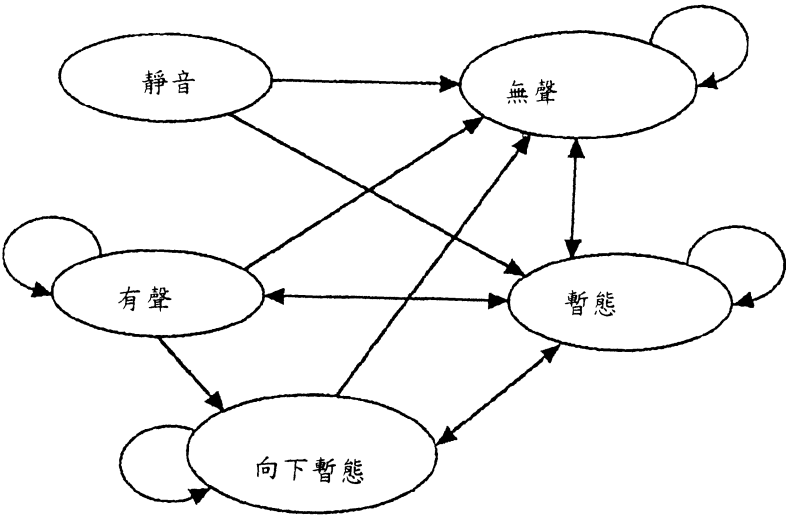


圖 4B

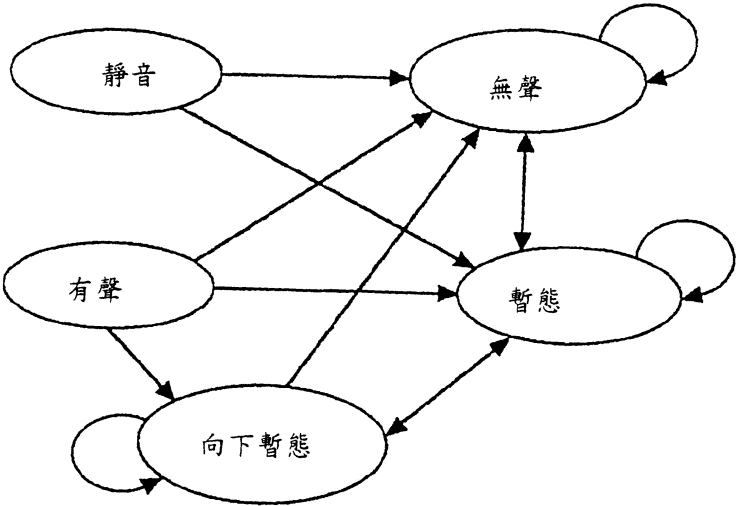
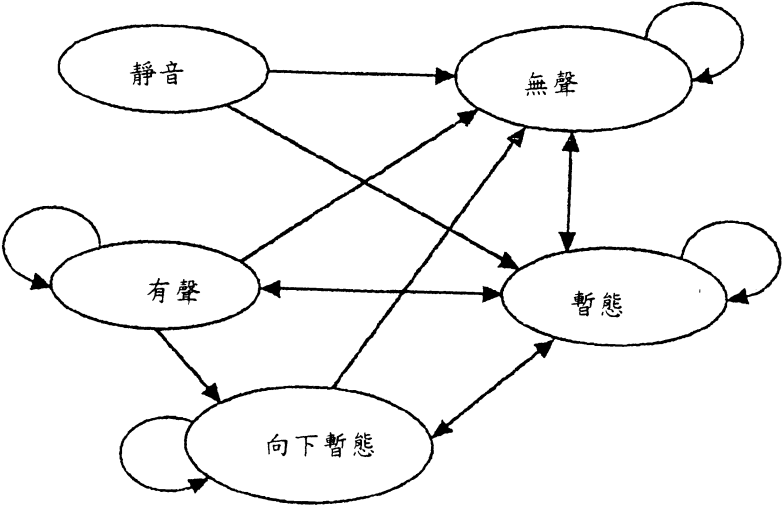


圖 4C



目前 先前	靜音	無聲	有聲	向上暫態	暫態	向下暫態
靜音	$V_{ad}=0$	$nacf_ap[3]$ 非常低， zcr 高， bER 低， VER 非常低	X	預設值	X	X
無聲	$V_{ad}=0$	$nacf_ap[3]$ 非常低， $nacf_ap[4]$ 非常低， $nacf$ 非常低， zcr 高， bER 低， VER 非常低， $E < E_{prev}$	X	預設值	X	X
有聲	$V_{ad}=0$	VER 非常低， $E < E_{prev}$	預設值	X	$nacf_ap[1]$ 低， $nacf_ap[3]$ 低， $E < 0.5 * E_{prev}$	VER 非常低， $nacf_ap[3]$ 不太高
向上暫態， 暫態	$V_{ad}=0$	VER 非常低， $E < E_{prev}$	預設值	X	$nacf_ap[1]$ 低， $nacf_ap[3]$ 不太高， $nacf_ap[4]$ 低， 先前分類不是暫態	$nacf_ap[3]$ 不太高， $E > 0.05 * VE_{av}$
向下暫態	$V_{ad}=0$	VER 非常低	X	X	$E > E_{prev}$	預設值

圖 5A

目前 先前	靜音	無聲	有聲	向上暫態	暫態	向下暫態
靜音	Vad=0	預設值	X	nacf_ap[2]，nacf_ap[3]及 nacf_ap[4]顯示增加傾向， nacf_ap[3]不太低，nacf_ap[4] 不太低，zcr不太高， VER不太低，bER高， zcr非常低	X	X
無聲	Vad=0	預設值	X	nacf_ap[2]，nacf_ap[3]及 nacf_ap[4]顯示增加傾向， nacf_ap[3]不太低，nacf_ap[4] 不太低，zcr不太高， VER不太低，bER高，zcr非常低， nacf_ap[3]非常高，nacf_ap[4] 非常高，reil低，E>Eprev， nacf不太低等	X	X
有聲，向上暫態，暫態	Vad=0	bER<=0， VER非常低， E<Eprev， bER>0	X	X	bER>0，nacf_ap[2]，nacf_ap[3] 及nacf_ap[4]顯示增加傾向， zcr不太高，VER不太低，reil低 nacf_ap[3]不太低， nacf不太低bER<=0	bER>0，nacf_ap[3]， 不太高，E<Eprev， zcr不太高，VER2<-15
向下暫態	Vad=0	預設值	X	X	nacf_ap[2]，nacf_ap[3]及 nacf_ap[4]顯示增加傾向， nacf_ap[3]相當高，nacf_ap[4] 非常高。VER不太低， E>2*Eprev等	VER不太低，zcr低

圖 5B

目前 先前	靜音	無聲	有聲	向上暫態	暫態	向下暫態
靜音	Vad=0	預設值	X	nacf_ap[2]，nacf_ap[3]及 nacf_ap[4]顯示增加傾向， nacf_ap[3]不太低，nacf_ap[4] 不太低，zcr不太高， VER不太低，bER高，zcr非高低， $E > 2 * E_{prev}$ ，nacf_ap[3]非常高， nacf_ap[4]非常高等	X	X
無聲	Vad=0	預設值	X	nacf_ap[2]，nacf_ap[3]及 nacf_ap[4]顯示增加傾向， nacf_ap[3]不太低，nacf_ap[4] 不太低，zcr不太高， VER不太低，bER高，zcr非高低， nacf_ap[3]非常高，nacf_ap[4] 非常高，refl低， $E > E_{prev}$ ， nacf不太低， $E > 2 * E_{prev}$ 等	X	X
有聲，向上暫態 ，暫態	Vad=0	bER≤0， VER非常低， $E < E_{prev}$ ， bER>0	nacf_ap[2]>低 VOICEDTH， bER≥0， VER不太低等	X	bER>0，nacf_ap[2]，nacf_ap[3] 及nacf_ap[4]顯示增加傾向， zcr不太高，VER不太低，refl低， nacf_ap[3]不太低，nacf不太低 bER≤0等	bER>0，nacf_ap[3] ，不太高， $E < E_{prev}$ ，zcr不太高， VER2<-15
向下暫態	Vad=0	預設值	X	X	nacf_ap[2]，nacf_ap[3]及 nacf_ap[4]顯示增加傾向， nacf_ap[3]相當高，nacf_ap[4] 非常高，VER不太低， $E > 2 * E_{prev}$ 等	VER不太低，zcr低

圖 5C

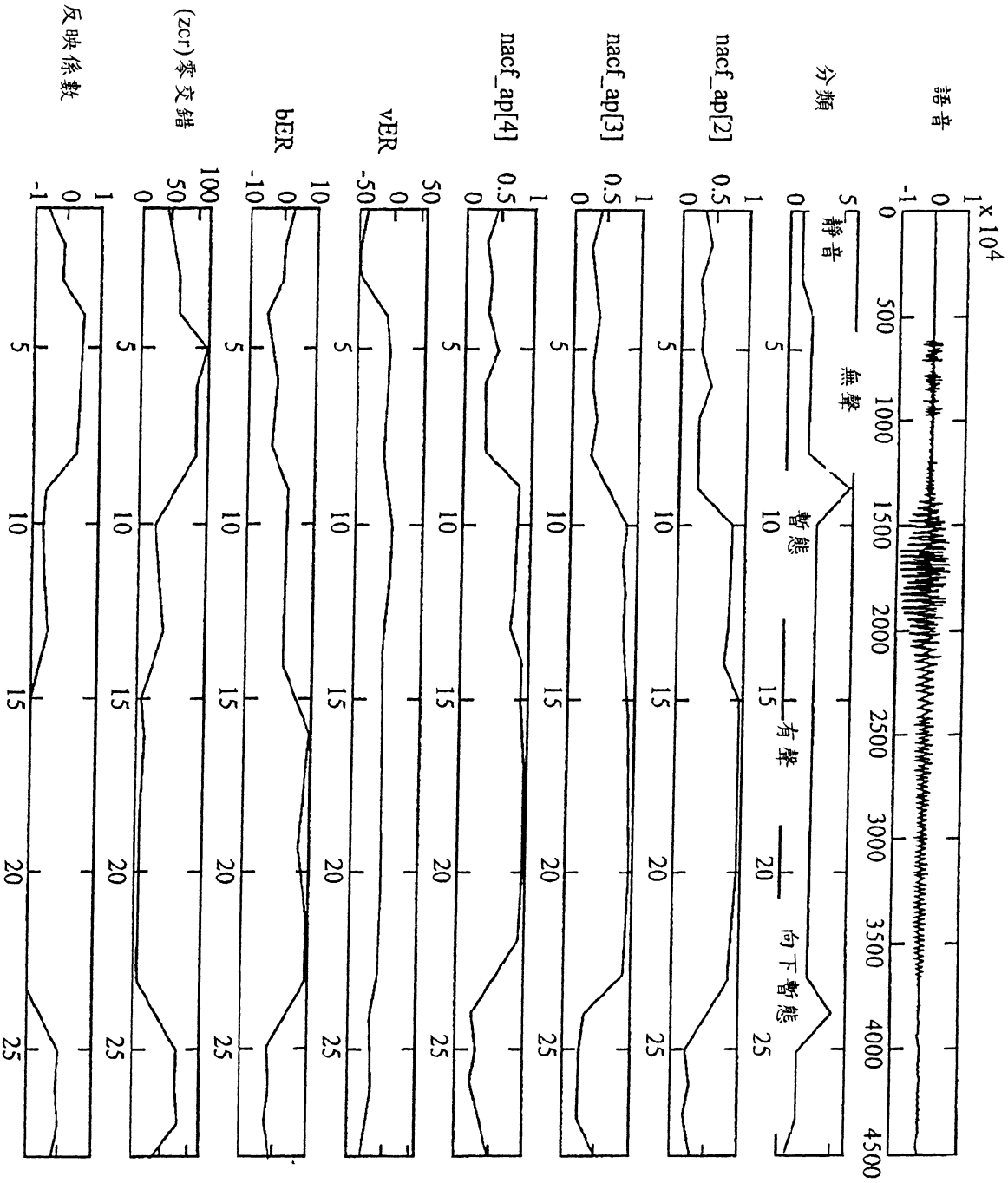


圖 6