

公告本

申請日期	91.1.11
案 號	91100301
類 別	G10L 15/00

A4
C4

(以上各欄由本局填註)

546632

發 明 專 利 說 明 書

一、發明 新型名稱	中 文	用以有效率儲存語音辨識模型之系統及方法
	英 文	SYSTEM AND METHOD FOR EFFICIENT STORAGE OF VOICE RECOGNITION MODELS
二、發明 創作人	姓 名	哈里納斯 葛拉達卓 HARINATH GARUDADRI
	國 籍	加拿大
	住、居所	美國加州聖地牙哥市歐維多街9435號 9435 Oviedo Street San Diego, California 92129 United States of America
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 5775 Morehouse Drive San Diego, California 92121-1714 United States of America
	代 表 人 姓 名	菲力普 R. 華德渥斯 PHILIP R. WADSWORTH

裝

訂

線

(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	2001年01月12日	09/760,076	☒ 有 ☐ 無主張優先權

有關微生物已寄存於：，寄存日期：，寄存號碼：

(請先閱讀背面之注意事項再填寫本頁各欄)

裝訂線

五、發明說明(1)

發明背景

I. 範疇

概略地說，本發明係有關該通信範疇；更特定言之，本發明係有關一種在一語音辨識系統中用以改良樣板儲存之系統及方法。

II. 背景

語音辨識(VR)代表該其中一項最重要之用以賦與一個機器模擬智慧、以'辨識使用者命令或使用者發聲命令"和"幫助人機介面"的技藝。VR亦代表一種人類語音瞭解的重要技藝將利用技藝、使一個聲音語音信號回復成一個語言訊息之系統稱為語音辨識器。於此利用該術語"語音辨識器"概括地意指任何口語使用者介面致能裝置。

基於安全的理由，導致使用VR(一般亦意指為語音辨識)逐漸變得越來越重要。例如，可利用VR代替在一個無線電話輔助鍵盤上壓按鈕的手動工作。當一個使用者於開車期間欲打電話時，此將係特別重要。當使用一個沒有VR的電話時，則該駕駛員於打電話時須將一隻手從方向盤上移開，且須看著該電話輔助鍵盤壓按鈕。該等行動增加了車禍的可能性。一個語音致能電話(即一個設計用作語音辨識的電話)容許該駕駛員一邊打電話、一邊繼續留意道路。此外，一個不必用手操作的汽車kit系統將允許該駕駛員於打電話期間把兩隻手同時保持在方向盤上。

將語音辨識裝置分類成語言依存(SD)裝置抑或語言無關(SI)裝置。語言依存裝置係較普通的，其中訓練語言依存

五、發明說明(2)

裝置使辨識特殊使用者的命令。對照之下，語言無關裝置能夠接受任何使用者的語音命令。為了增加一個特定VR系統的性能，故不論是語言依存抑或語言無關，均需接受訓練以使該系統裝備有效的參數。換言之，該系統需在其可最佳化作用之前先行學習。

一個不必用手操作之汽車kit其一個代表性的字彙可能包含：該輔助鍵盤上的數字；該等關鍵字"電話"、"傳送"、"撥號"、"取消"、"清除"、"增加"、"刪除"、"歷史"、"計劃"、"是"及"否"；及該等預設個數的名稱，通常稱為同事、朋友、或家庭成員。一旦訓練完成時，該使用者即可藉由說出該等訓練的關鍵字來打電話，其中該VR裝置藉由比較該等口語的語調與該等先前訓練的語調(儲存成樣板)和採取該最佳的匹配作辨識。例如，如該名稱"約翰"為該等訓練名稱中的其中一個名稱時，則該使用者可藉由講述該片語"打電話給約翰"以啟始一個電話給約翰。該VR系統將辨識該等字元"打電話"和"約翰"，及將撥接該使用者先前輸入成約翰的電話號碼之號碼。利用廢物樣板表示所有不在該字彙中的字元。

結合多個VR引擎將提高準確度，及於該輸入語音信號中利用更多的資訊量。於2000年7月18日提出之美國專利申請案號碼09/618,177(下文中稱為'177申請案)、定名為"語音辨識之結合的引擎系統及方法"中，及於2000年9月8日提出之美國專利申請案號碼09/657,760(下文中稱為'760申請案)、定名為"利用映射之自動化語音辨識系統及方法"

五、發明說明(3)

中說明一種用以結合VR引擎之系統及方法，其中指定給本發明的受讓人，及以引用的方式完全併入本文中。

雖然一個結合VR引擎的VR系統比一個利用一單一VR引擎的VR系統準確，但該結合之VR系統的每一個VR引擎可因一個嘈雜的環境而包含不準確度。可因背景雜訊而無法辨識一個輸入語音信號。背景雜訊可導致一個輸入語音信號與該VR系統字彙中的一個樣板不相匹配，或可導致一個輸入語音信號與該VR系統字彙中的一個樣板匹配錯誤。當該輸入語音信號與一個樣板不相匹配時，則拒絕該輸入語音信號。當該VR系統選擇了一個未與該輸入語音信號相對應的樣板時，則導致一個匹配錯誤。該匹配錯誤條件亦已知為取代，因以一個不正確的樣板取代一個正確的樣板。

一個就背景雜訊改良VR準確度之具體實施例係合適的。其中一個可導致一個拒絕或一個匹配錯誤的背景雜訊範例係當開車時利用一個行動電話作語音撥接時，及當附加的道路雜訊損毀該麥克風上所接收的輸入語音信號時。該附加的道路雜訊可降低語音辨識和準確度，及導致一個拒絕或一個匹配錯誤。

另一個可導致一個拒絕或一個匹配錯誤的雜訊範例係當一個放在該遮陽板上的麥克風或一個雙耳式耳機上所接收的語音信號受制於迴旋失真。因迴旋失真而導致的雜訊已知為迴旋雜訊和頻率匹配錯誤。迴旋失真取決於許多因素，像是該嘴巴與麥克風之間的距離、該麥克風的頻率回應

五、發明說明(4)

、該汽車內部的聲音特性...等等。該等條件可降低語音辨識的準確度。

傳統上，先前的VR系統已包含了一個RASTA濾波器，用以過濾迴旋雜訊。然而，該RASTA濾波器並未過濾背景雜訊。於美國專利號碼5,450,522中說明該一種濾波器。如是，需要一種過濾迴旋雜訊和背景雜訊之技藝。該一種技藝將改良該一個VR系統的準確度。

於一個VR系統中，不論其為一個語言依存的VR系統抑或一個語言無關的VR系統，均由一個記憶體的大小限制該可儲存在該等兩種VR系統型式該記憶體中的樣板總數。因受限於該可儲存的樣板總數，故該記憶體的有限大小限制了該VR系統的堅韌性。一種在該等VR系統其記憶體中增加該可儲存之樣板總數的系統及方法係合適的。

發明概要

該等說明之具體實施例係針對一種在一語音辨識系統中用以改良樣板儲存之系統及方法。一方面，一種語音辨識之系統及方法包含：記錄多個語調；擷取該等多個語調的特性，以產生該等多個語調其擷取之特性；依照該等多個語調其擷取之特性產生多個VR模型；及lossy壓縮該等多個VR模型，以產生多個lossy壓縮之VR模型。一方面，利用A-定律壓縮和擴充。另一方面，利用mu-定律壓縮和擴充。一方面，該等VR模型為隱藏馬可夫模型(HMM)。另一方面，該等VR模型為動態時間扭曲(DTW)模型。

一方面，一個語音辨識(VR)系統包括一個訓練模組。架

五、發明說明 (5)

構成：擷取多個語調的特性，以產生該等語調其擷取之特性；依照該等語調其擷取之特性產生多個VR模型；及lossy壓縮該等多個VR模型，以產生多個lossy壓縮之VR模型。一方面，該VR系統尚包含：一個特性擷取模組，架構成擷取一個測試語調的特性，以產生一個測試語調其擷取之特性；一個擴充模組，架構成擴充該等多個lossy壓縮VR模型之其中一個lossy壓縮之VR模型，以產生一個擴充之VR模型；及一個形態匹配模組，其中使該測試語調其擷取之特性與該擴充之VR模型相匹配，以產生一個辨識假設。

圖示簡單說明

圖1說明一個VR系統中的一個VR前端；

圖2說明一個語音片段其一個範例性之隱藏馬可夫模型(HMM)；

圖3說明一個根據一具體實施例之VR系統其一個HMM模組的一個前端；

圖4說明一個具有一 μ -定律壓縮擴張設計、而非對數壓縮的前端；

圖5說明一個具有一A-定律壓縮擴張設計、而非對數壓縮的前端；

圖6說明 $C=50$ 之一個 $\text{Log}_{10}()$ 函數和該 μ -對數函數之一個定點執行之標繪圖；

圖7說明一個根據一具體實施例之利用 μ -定律壓縮和 μ -定律擴充之前端；

五、發明說明 (6)

圖8說明一個根據一具體實施例之利用A-定律壓縮和A-定律擴充之前端；

圖9說明一個根據一具體實施例、用以產生模型之訓練處理其輸入、處理及輸出的方塊圖；

圖10說明一個根據一具體實施例的VR系統；及

圖11說明一個根據一具體實施例、於語音辨識期間利用擴充壓縮訓練之模型的VR系統。

較佳具體實施例詳細說明

一個VR系統包含一個執行前端處理、以便記述一個語音片段特徵的前端。圖1說明一個VR系統中的一個VR前端10。一個Bark振幅產生模組12於每隔T毫秒即將一個數位化脈衝碼調變(PCM)語音信號 $s(n)$ 轉換成k個bark振幅。於一個具體實施例中，T為10毫秒，k為16個bark振幅。如是，每隔10毫秒就有16個bark振幅。熟諳此藝者應瞭解，k可為任意的正整數，且T可使任何時間期間。

該Bark量化為相對應至人類聽覺感知上之關鍵頻帶的一個扭曲頻率量化。Bark振幅計算在此項技藝中已為吾人所知，且於Lawrence Rabiner和Biing-Hwang Juang之"語音辨識的原理"(1993年)中說明之，其中以引用的方式完全併入本文中。

Bark振幅模組12耦合至一個對數壓縮模組14上。對數壓縮模組14藉由取得該每一個bark振幅的對數以將該等bark振幅轉換成一個 $\log_{10}$ 的標度。對數壓縮模組14耦合至一個Cepstral轉換模組16上。Cepstral轉換模組16計算j個靜態

五、發明說明(7)

的 cepstral 係數和 j 個動態的 cepstral 係數。Cepstral 轉換係此項技藝中已為吾人所熟知的一個餘弦轉換。例如，參看先前以引用的方式併入本文中之 Lawrence Rabiner 和 Biing-Hwang Juang 的 "語音辨識的原理" (1993 年)。於一個具體實施例中， j 等於 8。熟諳此藝者應瞭解， j 可為其它任意的正整數。如是，則前端模組 10 於每隔 T 毫秒即產生 $2*j$ 個係數。由一個後台模組 (未顯示) (像是一個藉由使 HMM 模型與該等前端特性相匹配、以執行語音辨識之 HMM 模組) 處理該等特性。

藉由計算該 VR 前端中的 " j " 個靜態 cepstral 參數和 " j " 個動態 cepstral 參數訓練 HMM 模型。該訓練處理收集多個 (N 個) 相對應至一單一狀態上的訊框。接著，該訓練處理計算該等 N 個訊框的平均值和變異數，而導致一個長度為 $2j$ 的平均向量和一個長度為 $2j$ 的對角線協方差。將該等平均向量和變異數向量共同稱為一個高斯混合元件，或簡稱為 "混合"。由 N 個高斯混合元件表示每一種狀態，其中 N 為一個正整數。該訓練處理亦計算轉換機率。

於擁有少量記憶體資源的裝置中， N 等於 1 或其它某個小的數字。於一個最小型的足印 VR 系統 (即記憶體最少的 VR 系統) 中，一個單一的高斯混合元件係代表一種狀態。於較大型的 VR 系統中，利用多個 (N 個) 訊框計算多個平均向量和該等相對應的變異數向量。例如，如計算一組 "十二個平均值 / 十二個變異數" 時，則產生一種 12-高斯-混合-元件的 HMM 狀態。於一個分散式語音辨識 (DVR) 系統內的

五、發明說明 (8)

VR伺服器中，N可高達32。

一個HMM模型可為一個用以辨識一個輸入語音信號的或然性框架。於一個HMM模型中，利用暫時和光譜兩種特性記述一個語音片的特徵。由一連串的狀態和一組轉換機率表示每一個HMM模型(全字元或副字元)。圖2說明一個語音片其一個範例性的HMM模型。該HMM模型可代表一個字元"哦"，或代表一個字元"美國俄亥俄州"的一部分。利用Viterbi解碼比較該輸入語音信號與多個HMM模型。將該最佳匹配的HMM模型考慮成該辨識假設。該範例性HMM模型30擁有五種狀態-"開始32"、"結束34"及"一個代表性三重電話的三種狀態：狀態一36、狀態二38及狀態三40"。於一個典型的後台中，利用一個全字元模型、連同小型的字彙VR系統。

於中型~大型的字彙系統中，利用副字元模型。典型的副字元單元為內文無關(CI)的電話和內文相關(CD)的電話。一個內文無關的電話係與該等電話向右和向左無關。內文相關的電話亦稱為三重電話，因其係視該等電話向左和向右而定。內文相關的電話亦稱為全部電話。

該VR技藝中的一個電話為實現一個音素。於一個VR系統中，利用HMM或此項技藝中其它已知之VR模型型式建立內文無關的電話模型和內文相關的電話模型。一個音素為一種特定語言中該最小功能語音片段的一個摘要。在此，該字元功能的暗指在知覺上不同的聲音。例如，用"b"的發音取代"貓"中"k"的發音，則導致該英文語言中一個

五、發明說明 (9)

不同的字元。如是，"b"和"k"為英文語言中兩個不同的音素。

轉換 a_{ij} 為從狀態 i 轉換成狀態 j 的機率。 a_{s1} 係從開始狀態 32 轉換成第一種狀態 36。 a_{12} 係從第一種狀態 36 轉換成第二種狀態 38。 a_{23} 係從第二種狀態 38 轉換成第三種狀態 40。 a_{3E} 係從第三種狀態 40 轉換成結束狀態 34。 a_{11} 係從第一種狀態 36 轉換成第一種狀態 36。 a_{22} 係從第二種狀態 38 轉換成第二種狀態 38。 a_{33} 係從第三種狀態 40 轉換成第三種狀態 40。 a_{13} 係從第一種狀態 36 轉換成第三種狀態 40。

可依照所有的轉換/機率： a_{ij} 建構一個轉換機率矩陣，其中 n 為該 HMM 模型中的狀態總數， $i = 1, 2, \dots, n$ ， $j = 1, 2, \dots, n$ 。當狀態之間沒有轉換時，則轉換/機率等於零。該從一種狀態開始的累積轉換/機率為一致不變的，即等於一。

圖 3 說明一個根據一具體實施例之 VR 系統其一個 HMM 模組的一個前端。一個 Bark 振幅模組 12 耦合至一個對數壓縮模組 14 上。對數壓縮模組 14 耦合至一個 RASTA 濾波模組 18 上。RASTA 濾波模組 18 耦合至一個 Cepstral 轉換模組 16 上。利用一個頻帶通行濾波器 $h(i)$ 過濾該等 k 個通道中每一個通道的對數 Bark 振幅。於一個具體實施例中，該 RASTA 濾波器為一個中心頻率約莫 4 赫茲得的頻帶通行濾波器 $h(i)$ 。語音中，每秒概略約有四個字節。因此，一個中心頻率約為 4 赫茲的頻帶通行濾波器將保留像語音的信號和減弱不像語音的信號。如是，該頻帶通行濾波器導致改良了背

五、發明說明 (10)

景雜訊中的辨識準確度和頻率匹配錯物條件。熟諳此藝者應瞭解，該中心頻率可不等於4赫茲，端視該工作而定。

接著，該Cepstral轉換模組處理該等過濾之對數Bark振幅，以於每隔T毫秒即產生該等 $2*j$ 個係數。於1995年9月12日提出之美國專利號碼5,450,522、定名為"語音參數化的聽覺模型"中說明一個可用於該VR前端中之頻帶通行濾波器的一個範例，其中以引用的方式併入本文中。圖3中所示之該前端將減少該等通道匹配錯誤條件的影響，及改良VR的辨識準確度。

圖3中所描述的該前端對背景匹配錯誤條件而言並不怎麼健全，其中一個理由為該對數壓縮處理對。該等bark通道有一個非線性的振幅影響。對數壓縮導致將該等bark通道上的低振幅區域擴大地比高振幅區域多。因該背景雜訊典型地係在該等bark通道上的低振幅區域中，故VR性能將隨著信號/雜訊比(SNR)的減小而開始降低。如是，一個在該等bark通道上的低振幅區域中像線性的和高振幅區域中相對數的模組係較可取的。

藉由利用一個對數壓縮擴張設計有效地達成此，像是藉由該國際電訊聯盟(ITU-T)推薦書G.711 (11/88)-音頻的脈衝碼調變(PCM)中和G711.C/G.711-編碼/解碼功能中所說明的該G.711對數壓縮擴張(壓縮和擴充)。該ITU-T(該國際電訊聯盟的電傳通訊標準化部分)為該促進電傳通訊設備和系統其共同標準的主要國際團體。

有兩種G.711對數壓縮擴張設計：一個mu-定律壓縮擴張

五、發明說明 (11)

設計和一個 A-定律壓縮擴張設計。該 mu-定律壓縮擴張設計和該 A-定律壓縮擴張設計均為脈衝碼調變(PCM)方法。就是取樣一個類比信號，及量化該每一個取樣之信號的振幅，即指定一個數字的數值。該 mu-定律壓縮擴張設計和該 A-定律壓縮擴張設計藉由該取樣之信號其對數曲線的一個線性近似值量化該取樣之信號。

該 mu-定律壓縮擴張設計和該 A-定律壓縮擴張設計均在一個對數曲線上操作。因此，將該對數曲線分成線段，其中每一個連續線段的長度等於該先前線段長度的兩倍。因該 mu-定律壓縮擴張設計和該 A-定律壓縮擴張設計以不同的方式計算該線性近似值，故 A-定律及 mu-定律壓縮擴張設計具有不同的線段長度。

該 G.711標準包含一個近似如下面表 1 中所示之該 mu-定律線性近似值之 mu-定律查表。按照該 mu-定律壓縮擴張設計，則將以一個總計為 8,159 個區間趨近一個類比信號。

數值範圍	區間號碼	區間大小
0	1	1
1-16	15	2
17-32	16	4
33-48	16	8
49-64	16	16
65-80	16	32
81-96	16	64
97-112	16	128
113-127	16	256

表 1

裝
訂
線

五、發明說明 (12)

該 G.711 標準包含一個近似如下面表 2 中所示之該 A-定律線性近似值之 A-定律查表。按照該 A-定律壓縮擴張設計，則將以一個總計為 4,096 個區間趨近一個類比信號。

數值範圍	區間號碼	區間大小
0-32	32	2
33-48	16	4
49-64	16	8
65-80	16	16
81-96	16	32
97-112	16	64
113-127	16	128

表 2

該 G.711 標準指定一個 μ -定律壓縮擴張設計，代表 8 位元/樣品中以 14 位元/樣品量化之語音。該 G.711 標準亦指定一個 A-定律壓縮擴張設計，代表 8 位元/樣品中以 13 位元/樣品量化之語音。代表性的 8-位元資料為語音電話學。該 G.711 規格對具有一機率密度函數(pdf)的信號(像是語音)而言係最佳化的。

熟諳此藝者應瞭解，可利用其它的壓縮擴張設計。此外，熟諳此藝者應瞭解，可利用其它的量子化比率。

於一個具體實施例中，可於該前端中利用一個 μ -定律壓縮擴張設計 20，而不利用該對數壓縮設計，如圖 4 中所示。圖 4 說明一個利用一 μ -定律壓縮設計(即一個 μ -對數

五、發明說明 (13)

壓縮模組20)之具體實施例的前端。Bark振幅產生模組12耦合至mu-對數壓縮模組20上。mu-對數壓縮模組20耦合至一個RASTA濾波模組18上。RASTA濾波模組18耦合至一個Cepstral轉換模組16上。

一個包含迴旋失真的數位化語音信號 $s(n)$ 進入Bark振幅產生模組12中。於Bark振幅產生模組12將該數位化PCM語音信號 $s(n)$ 轉換成 k 個bark振幅之後，該迴旋失真則變成倍數的失真。mu-對數壓縮模組20對該等 k 個bark振幅執行mu-對數壓縮。該mu-對數壓縮使得該倍數失真增加。因迴旋失真元件為固定的，故RASTA濾波模組18過濾任何固定的元件、藉以移除該迴旋失真。Cepstral轉換模組16依照該RASTA過濾之輸出計算 j 個靜態的cepstral係數和 j 個動態的cepstral係數。

於另一個具體實施例中，可於該前端中利用一個A-定律壓縮設計21，而不利用一個對數壓縮設計，如圖5中所示。圖5說明一個利用一A-定律壓縮設計(即一個A-對數壓縮模組21)之具體實施例的前端。Bark振幅產生模組12耦合至A-對數壓縮模組21上。A-對數壓縮模組21耦合至一個RASTA濾波模組18上。RASTA濾波模組18耦合至一個Cepstral轉換模組16上。

Mu-對數壓縮和A-對數壓縮兩者均為lossy壓縮技藝。可利用任何lossy壓縮技藝壓縮該等 k 個bark振幅。當壓縮和擴充一個來源的結果與該來源不同時，則利用lossy壓縮。因將一個lossy壓縮之模型擴充所需的時間比擴充一個

五、發明說明 (14)

lossless模型所需的時間短，故lossy壓縮比一個lossless壓縮有效益。此外，lossy壓縮軟體比lossless壓縮軟體便宜。

一個使用G.711 mu-定律壓縮擴張的具體實施例包含兩個函數。一個稱為u定律_壓縮，用以壓縮bark振幅；另一個稱為mu定律_擴充，用以擴充濾波器的輸出、以產生bark振幅。於一個具體實施例中，mu-對數壓縮模組20利用下面的公式執行該壓縮：

對數_Bark (i) = {255 - mu定律_壓縮[Bark(i)]}*C，其中C為一個常數。

可將該C的數值調整成利用一個定點VR執行中該可利用的解答。

圖6說明該 $\text{Log}_{10}()$ 函數和該mu-對數函數之一個定點執行之標繪圖，其中 $C=50$ 。圖6顯示，就低振幅信號而言，該mu-對數函數比該 $\text{Log}_{10}()$ 函數線性化；反之，就高振幅信號而言，該mu-對數函數為對數的。如是，因該mu-對數函數以不同的方式處理低振幅信號和高振幅信號，故該mu-對數函數為一個非一致的量化器。

於某些辨識設計中，該後台操作該等bark通道振幅，而非操作靜態和動態cepstral參數。於該'177申請案和該'760申請案中所述之該結合之引擎設計中，該DTW引擎於時間群集和振幅量子化之後操作bark通道振幅。該DTW引擎係以樣板匹配為基礎。儲存之樣板與該輸入之語音信號的特性相匹配。

該'177申請案和該'760申請案中所述之該DTW引擎，其

五、發明說明 (15)

對背景匹配錯誤條件較之對通道匹配錯誤條件來得健全。圖7描述一個改良該等通道匹配錯誤條件其DTW引擎之具體實施例的一個前端。圖7說明一個根據一利用mu-定律壓縮(即mu-對數壓縮模組20)和mu-定律擴充(即mu-定律擴充模組22)之具體實施例的前端。Bark振幅模組12耦合至一個mu-對數壓縮模組20上。mu-對數壓縮模組20耦合至一個RASTA濾波模組18上。RASTA濾波模組18耦合至mu-定律擴充模組22上。mu-對數壓縮模組20對k個bark振幅執行A-對數壓縮。因迴旋失真元件為固定的，故RASTA濾波模組18於該等k個bark振幅中過濾任何固定的元件、藉以移除任何迴旋失真。mu-對數擴充模組22對該等RASTA過濾之bark振幅執行mu-對數擴充。

於一個具體實施例中，利用下面的公式執行該Mu-對數擴充：

$$\text{Bark}'(i) = \text{mu定律_擴充}\{255 - [R(i) * D]\}$$
，其中D為一個常數。

R(i)為該RASTA模組的輸出及 $D = 0.02$ (或 $1/C$)。於一個具體實施例中，該乘積 $[R(i) * D]$ 係在0~127的範圍內。該Mu-對數擴充將該 $\text{Bark}'(i)$ 放在該bark振幅的範圍中，且該RASTA處理將該等通道匹配錯誤條件的反效果移除。

圖8描述一個用以改良該等通道匹配錯誤條件其DTW引擎之具體實施例。圖8說明一個根據一利用A-定律壓縮(即一個A-對數壓縮模組24)和A-定律擴充(即一個A-定律擴充模組26)之具體實施例的前端。一個Bark振幅模組12耦合

五、發明說明 (16)

至 A-對數壓縮模組 24 上。A-對數壓縮模組 24 耦合至一個 RASTA 濾波模組 18 上。RASTA 濾波模組 18 耦合至 A-定律擴充模組 26 上。A-對數壓縮模組 20 對 k 個 bark 振幅執行 A-對數壓縮。因迴旋失真元件為固定的，故 RASTA 濾波模組 18 於該等 k 個 bark 振幅中過濾任何固定的元件、藉以移除任何迴旋失真。A-對數擴充模組 26 對該等 RASTA 過濾之 bark 振幅執行 A-對數擴充。

除了可在一個 VR 系統的前端中利用 Mu-定律和 A-定律壓縮和擴充、記述一個語音片的特徵之外，於一個具體實施例中，亦利用 mu-定律和 A-定律壓縮和擴充用以壓縮和擴充一個 VR 模型的狀態。於一個具體實施例中，該 VR 系統壓縮和擴充一個 HMM 模型的平均值和變異數。於另一個具體實施例中，該 VR 系統壓縮和擴充一個 DTW 模型的樣板。

於一個 HMM VR 系統中，由一序列的 HMM 狀態表示該等字彙字。每一種狀態包括一組平均值/變異數/轉換成該字彙字元中其它種狀態的轉換機率。

於其它未以 HMM 為基礎之 VR 系統(像是一個 DTW VR 系統)中，由一組自擷取一個訓練全集的特性而導出之平均值和/或變異數表示該等字彙字元。一個"訓練全集"由大量使用者所記錄之該等字彙字元的多個語調組成。於一個具體實施例中，由一組平均值表示該等字彙字元。於另一個具體實施例中，由一組平均值/變異數表示該等字彙字元。

五、發明說明 (17)

圖9說明該根據一具體實施例、用以產生模型(亦稱為樣板)之訓練處理其輸入50、訓練模組52及輸出模型54的流程圖。該訓練處理包括計算該VR字彙中所有字元其所有狀態的元件(包含廢物模型)，以於辨識期間代表字彙範圍以外(OOV)的語調。藉由記錄大量的使用者以產生一個由該等欲辨識之字彙字元其多個語調所組成的訓練全集。將該訓練全集輸入(50)訓練模組52中，其中訓練模組52處理該訓練全集，以產生訓練模組52的輸出、該等字彙字元的壓縮模型及OOV語調54。訓練模組52依照該訓練全集產生模型，及壓縮該等模型。於一個具體實施例中，訓練模組52包括一個依照訓練全集50以產生模型的特性擷取模組(未顯示)，及一個壓縮該等模型的壓縮模組(未顯示)。

於一個具體實施例中，訓練模組52產生HMM模型，及對該等HMM模型執行A-定律壓縮。就相同的儲存空間量而言，及與不對該儲存空間實行壓縮設計相較，則該A-定律壓縮將致能該儲存空間儲存更多的模型。於另一個具體實施例中，訓練模組52產生DTW模型，及對該等DTW模型執行mu-定律壓縮。就相同的儲存空間量(即記憶體)而言，及與不對該儲存空間實行壓縮設計相較，則該mu-定律壓縮將致能該儲存空間儲存更多的模型。如是，不論該等模型為語言無關抑或語言依存，A-定律和mu-定律壓縮均會減少一個VR系統中的記憶體需求。然而，於另一個具體實施例中，該訓練模組利用另一種與A-定律、Mu-定律壓縮不同之lossy壓縮技藝壓縮模型。熟諳此藝者應瞭

五、發明說明 (18)

解，可利用此項技藝中任何已知之模型VR技藝代替該等HMM和DTW模型。

圖10說明一個根據一具體實施例之VR系統。一個特性擷取模組60自一個測試語調 $s(n)$ 中擷取特性。將該等特性轉移給一個後台模組，即形態匹配模組62。形態匹配模組62使該等特性與字彙字元和OOV語調54的模型相匹配。於一個具體實施例中，形態匹配模型62為一個HMM模組，其使該等特性與HMM模型相匹配，藉以產生辨識假設。於另一個具體實施例中，形態匹配模組62為一個DTW模組，其使特性與DTW模型相匹配，藉以產生辨識假設。

圖11說明一個根據一於語音辨識期間利用擴充之壓縮訓練模型之具體實施例的VR系統。一個特性擷取模組70自一個測試語調 $s(n)$ 中擷取特性。將該等特性轉移給一個後台模組，即形態匹配模組72。形態匹配模組72使該等特性與字彙字元和OOV語調74的模型相匹配，其中已由一個擴充模組76處理了該等模型的平均值和變異數。形態匹配模組62利用一個VR應用(未顯示)所指定的一個文法(未顯示)自擴充模組76中獲得模型。一個VR應用典型地為致能使用者利用該VR系統完成一項工作的服務邏輯。可由一個用戶單元上的一個處理器執行該服務邏輯。該服務邏輯為該用戶單元中一個使用者介面模組(未顯示)的一個元件。

該文法利用小型字彙VR系統中的字元模型和大型字彙VR系統中的副字元模型指定該動態字彙。典型的文法包含7-位元的電話號碼、美元總額及一組名稱中一個城市的

五、發明說明 (19)

名稱。典型的文法規格包含一個代表下面條件的OOV條件：
：無法根據該輸入之語音信號確信地判定一個辨識。

該文法指定語法。該語法限制欲與該等擷取之特性相匹配之字元模型和副字元模型的順序。型態匹配模組72向擴充模組76要求該等欲擴充之模型。擴充模組76擴充形態匹配模組72所要求的壓縮模型72。於一個具體實施例中，擴充模組76擴充HMM模型。於另一個具體實施例中，該擴充模組擴充DTW模型。如以A-定律壓縮壓縮該等模型時，則以A-定律擴充擴充該等模型。如以mu-定律壓縮壓縮該等模型時，則利用mu-定律擴充擴充該等模型。

於一個具體實施例中，形態匹配模組72為一個使該等特性與擴充之HMM模行相匹配之HMM模組，藉以產生辨識假設。於另一個具體實施例中，形態匹配模組72為一個使特性與擴充之DTW模行相匹配之DTW模組，藉以產生辨識假設。

熟諳此藝者應瞭解到，可利用任何種種不同的述語和技藝表示該等資訊和信號。例如，可由電壓、電流、電磁波、磁場或質點、光學場或質點、或該等前面的任何組合表示上面整個說明中所參考的資料、指令、命令、資訊、信號、位元、符號及晶片。

熟諳此藝者將更進一步察知，可將就該等此中所揭示的具體實施例所說明之各種不同的例證性邏輯區塊、模組、電路及通則步驟執行成電子硬體、電腦軟體或該等兩者的組合。為了清楚地例證說明硬體與軟體的可交換性，故已

五、發明說明 (20)

於上就其機能概括地說明了不同的例證性元件、區塊、模組、電路及步驟。不管將該機能執行成硬體亦或執行成軟體，端視強加於該全面系統上的特殊應用和設計限制而定。熟諳的技工可就每一種特殊的應用而以不同的方式執行該說明之機能，然而，不應將該等執行決策解譯成導致脫離本發明的範疇。

如範例，可藉由一個執行一組軟體指令之處理器、一個特殊應用積體電路(ASIC)、一個欄位可程式行列閘(FPGA)或其它的可程式邏輯裝置、離散閘或電晶體邏輯、離散硬體元件(例如像是暫存器)、任何傳統的可程式軟體模組和一個處理器、或該等前面任何設計成執行該等此中所述之功能的組合實施或執行各種就該等此中所揭示之具體實施例而說明的例證性邏輯區塊、模組及映射。可於一個微處理器中有利地執行該等VR系統元件，像是Bark振幅產生12、RASTA濾波模組18、Mu-對數壓縮模組20、A-對數壓縮模組21、Mu-數擴充22、A-對數擴充26、Cepstral轉換模組16、訓練模組52、形態匹配模組62及擴充模組76。但是，或者可於任何傳統的處理器、控制器、微控制器或狀態機器中執行Bark振幅產生12、RASTA濾波模組18、Mu-對數壓縮模組20、A-對數壓縮模組21、Mu-數擴充模組22、A-對數擴充模組26、Cepstral轉換模組16、訓練模組52、形態匹配模組62及擴充模組76。該等模型/樣板可常駐在隨機存取記憶體(RAM)、快閃記憶體、唯讀記憶體(ROM)、電氣式可程式化唯讀記憶體(EPROM)、電氣可清除式可

五、發明說明 (21)

程式唯讀記憶體(EEPROM)、暫存器、硬式磁碟、一個可移式磁碟、一個唯讀光碟(CD-ROM)或此項技藝中已知之其它任何形式的儲存媒體。可將該記憶體(未顯示)整合至上述任何處理器(未顯示)中。一個處理器(未顯示)和記憶體(未顯示)可常駐在一個ASIC(未顯示)中。該ASIC可常駐在一個電話中。

於先前提供該等具體實施例的說明，以致能任何熟諳此藝者實行或利用本發明。熟諳此藝者將迅速顯見該等具體實施例其不同的修正，及可在未利用本發明的技能下將該等此中所定義的通用原則應用到其它的具體實施例上。如是，未意欲將本發明限制在該等此中所示之具體實施例上，而是使本發明適合與該等此中所揭示之原則和新穎特性符合之最寬廣範疇。

四、中文發明摘要（發明之名稱：用以有效率儲存語音辨識模型之系統及方法）

一種藉由改良語音辨識(VR)樣板的儲存、以改良語音辨識之方法及系統。該改良之儲存意指可將更多的VR模型儲存在記憶體中。當記憶體中儲存的VR模型越多時，則該VR系統就越堅韌，且該VR系統就因而越準確。利用Lossy壓縮技藝壓縮VR模型。於一個具體實施例中，利用A-定律壓縮和A-定律擴充壓縮和擴充VR模型。於另一個具體實施例中，利用Mu-定律壓縮和Mu-定律擴充壓縮和擴充VR模型。於一個訓練處理期間壓縮VR模型，及於語音辨識期間擴充該等VR模型。

英文發明摘要（發明之名稱：SYSTEM AND METHOD FOR EFFICIENT STORAGE OF VOICE RECOGNITION MODELS）

A method and system that improves voice recognition by improving storage of voice recognition (VR) templates. The improved storage means that more VR models can be stored in memory. The more VR models that are stored in memory, the more robust the VR system and therefore the more accurate the VR system. Lossy compression techniques are used to compress VR models. In one embodiment, A-law compression and A-law expansion are used to compress and expand VR models. In another embodiment, Mu-law compression and Mu-law expansion are used to compress and expand VR models. VR models are compressed during a training process and they are expanded during voice recognition.

六、申請專利範圍

1. 一種語音辨識(VR)之方法，包括：
記錄多個語調；
擷取該等多個語調的特性，以產生該等多個語調其擷取之特性；
依照該等多個語調其擷取之特性產生多個VR模型；
及
lossy壓縮該等多個VR模型，以產生多個lossy壓縮之VR模型。
2. 如申請專利範圍第1項之方法，尚包括擴充該等多個lossy壓縮之VR模型，以產生多個擴充之VR模型。
3. 如申請專利範圍第2項之方法，尚包括比較該等多個擴充之VR模型之其中一個擴充之VR模型與一個測試語調的特性。
4. 如申請專利範圍第2項之方法，尚包括擷取一個測試語調的特性。
5. 如申請專利範圍第4項之方法，尚包括使該測試語調其擷取之特性與該等多個擴充之VR模型之其中一個擴充之VR模型相匹配，以產生一個匹配。
6. 如申請專利範圍第5項之方法，尚包括產生一個假設給該匹配。
7. 如申請專利範圍第1項之方法，其中該lossy壓縮為A-定律壓縮該等多個VR模型，以產生多個A-定律壓縮之VR模型。
8. 如申請專利範圍第1項之方法，其中該lossy壓縮為mu-

六、申請專利範圍

定律壓縮該等多個VR模型，以產生多個mu-定律壓縮之VR模型。

9. 如申請專利範圍第1項之方法，其中該等多個產生之VR模型為隱藏馬可夫模型(HMM)。
10. 如申請專利範圍第1項之方法，其中該等多個產生之VR模型為動態時間扭曲(DTW)模型。
11. 如申請專利範圍第7項之方法，尚包括擴充該等多個A-定律壓縮之VR模型之其中一個A-定律壓縮之VR模型，以產生一個擴充之VR模型。
12. 如申請專利範圍第8項之方法，尚包括擴充該等多個mu-定律壓縮之VR模型之其中一個mu-定律壓縮之VR模型，以產生一個擴充之VR模型。
13. 如申請專利範圍第11項之方法，尚包括擷取一個測試語調的特性。
14. 如申請專利範圍第12項之方法，尚包括擷取一個測試語調的特性。
15. 如申請專利範圍第13項之方法，尚包括使該測試語調其擷取之特性與該等多個擴充之VR模型之其中一個擴充之VR模型相匹配，以產生一個匹配。
16. 如申請專利範圍第14項之方法，尚包括使該測試語調其擷取之特性與該等多個擴充之VR模型之其中一個擴充之VR模型相匹配，以產生一個匹配。
17. 如申請專利範圍第15項之方法，尚包括產生一個假設給該匹配。

六、申請專利範圍

18. 如申請專利範圍第16項之方法，尚包括產生一個假設給該匹配。
19. 一種包括一個訓練模組之語音辨識(VR)系統，其中將該訓練模組架構成：擷取多個語調的特性，以產生該等語調其擷取之特性；依照該等語調其擷取之特性產生多個VR模型；及lossy壓縮該等多個VR模型，以產生多個lossy壓縮之VR模型。
20. 如申請專利範圍第19項之VR系統，尚包括：
- 一特性擷取模組，架構成擷取一測試語調的特性，以產生一測試語調其擷取之特性；
 - 一擴充模組，架構成擴充該等多個lossy壓縮之VR模型之其中一個lossy壓縮之VR模型，以產生一擴充之VR模型；及
 - 一形態匹配模組，其中使該測試語調其擷取之特性與該擴充之VR模型相匹配，以產生一辨識假設。
21. 一種語音辨識(VR)系統，包括：
- 多個lossy壓縮之VR模型；
 - 一特性擷取模組，架構成擷取一測試語調的特性，以產生一測試語調其擷取之特性；
 - 一擴充模組，架構成擴充該等多個lossy壓縮之VR模型之其中一個lossy壓縮之VR模型，以產生一擴充之VR模型；及
 - 一形態匹配模組，其中使該測試語調其擷取之特性與該擴充之VR模型相匹配，以產生一辨識假設。

六、申請專利範圍

22. 一種語音辨識(VR)訓練系統，包括：

- 一特性擷取模組，架構成擷取多個語調的特性，及產生多個VR模型給該等多個語調其擷取之特性；及
- 一壓縮模組，架構成lossy壓縮該等多個VR模型。

裝

訂

線

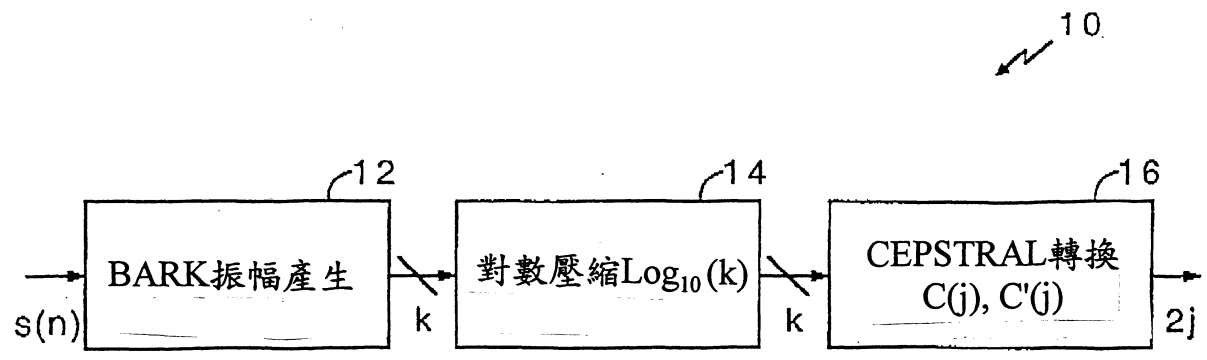


圖 1

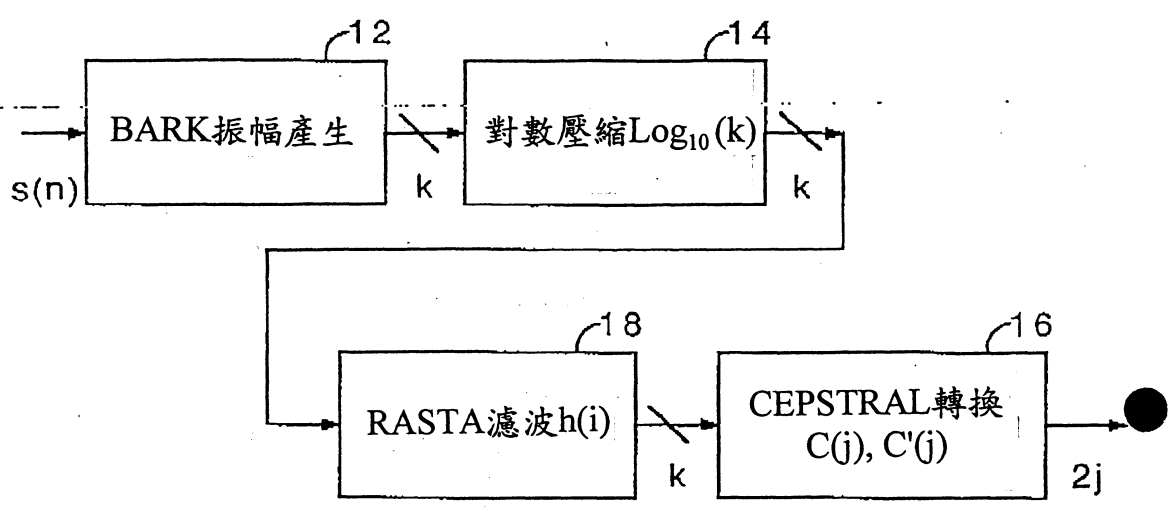


圖 3

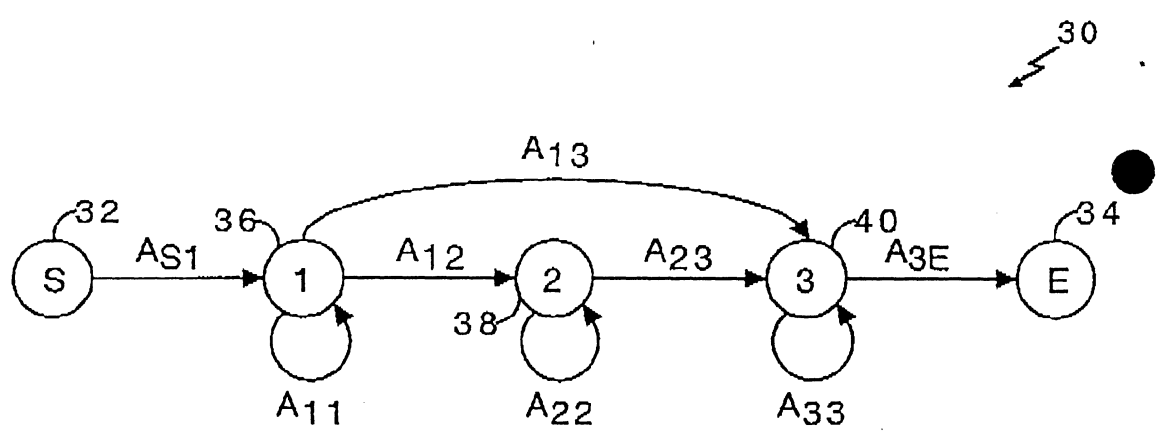


圖 2

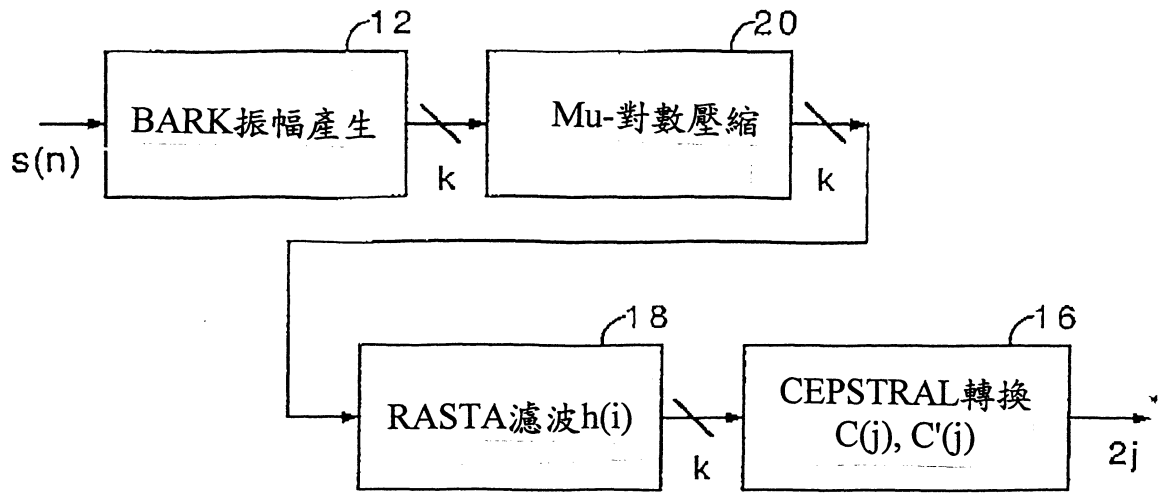


圖 4

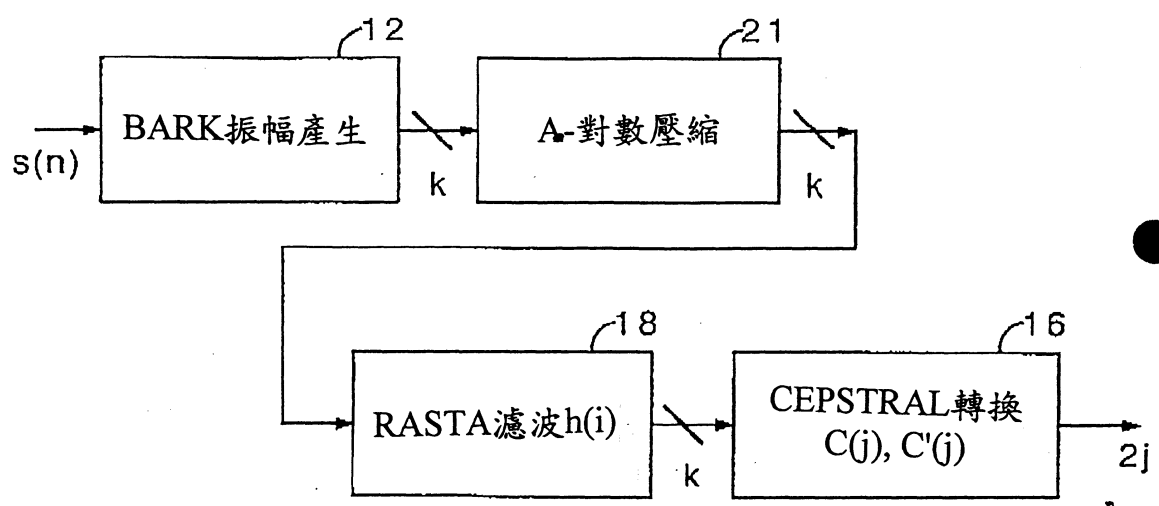


圖 5

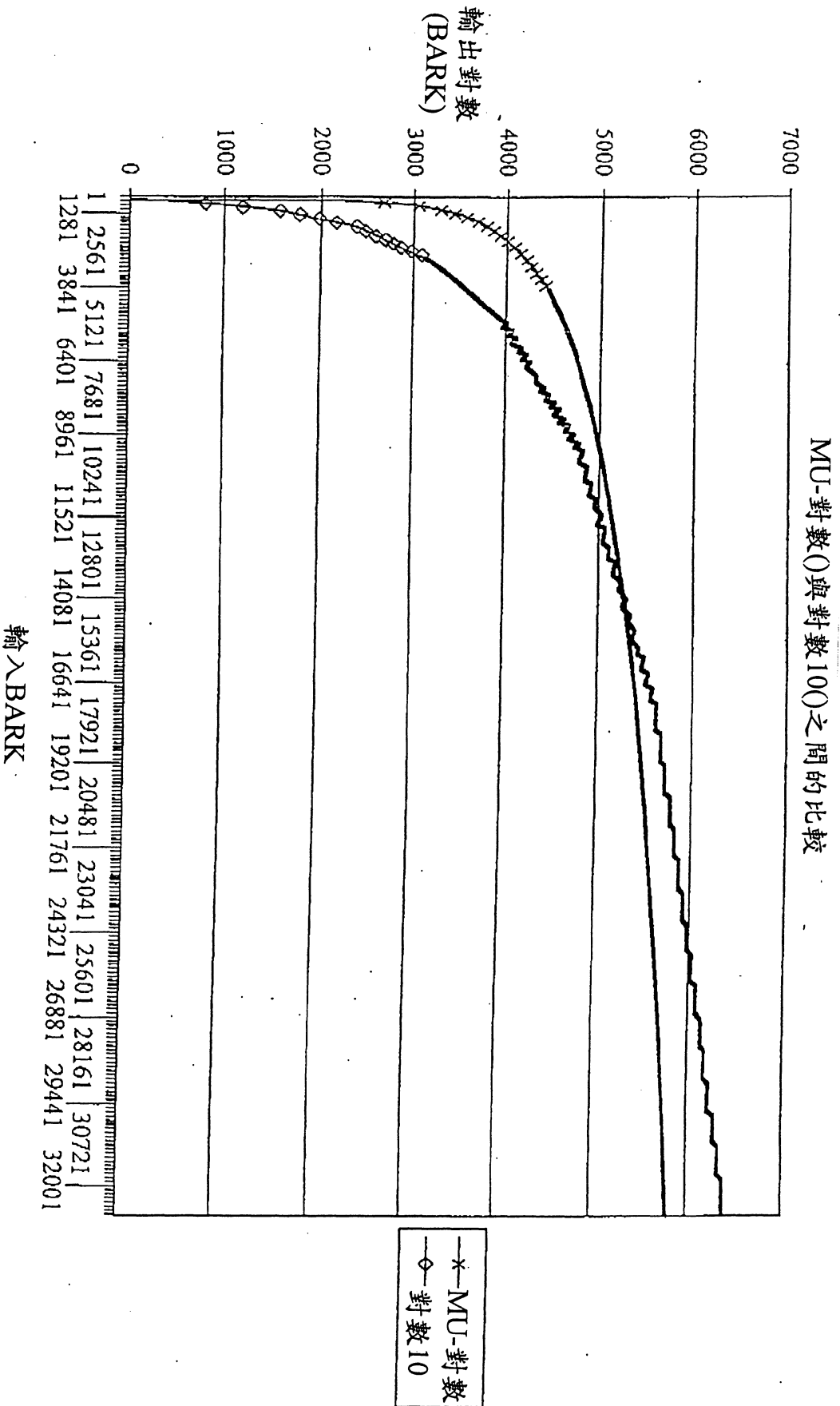


圖 6

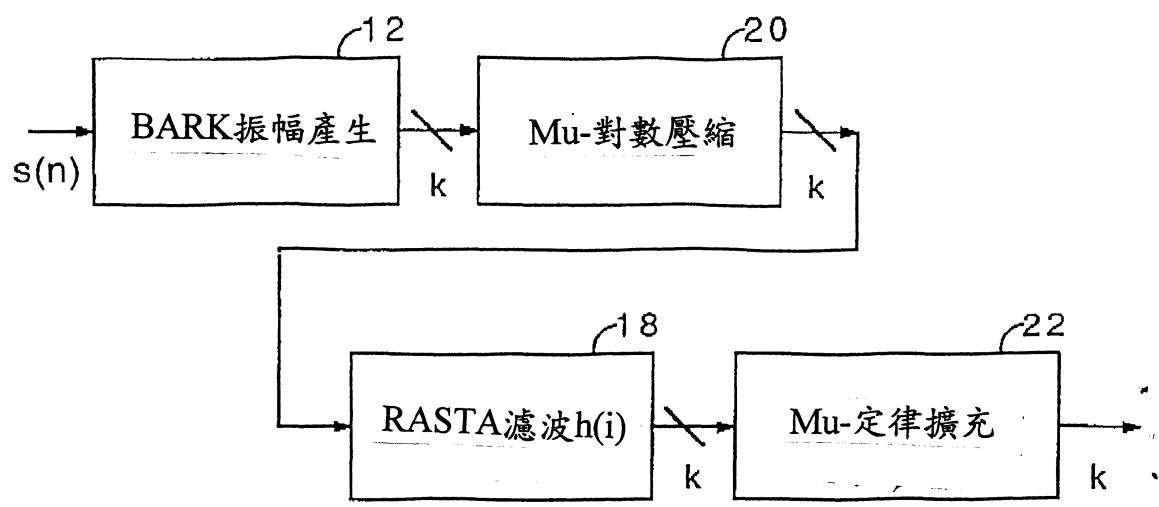


圖 7

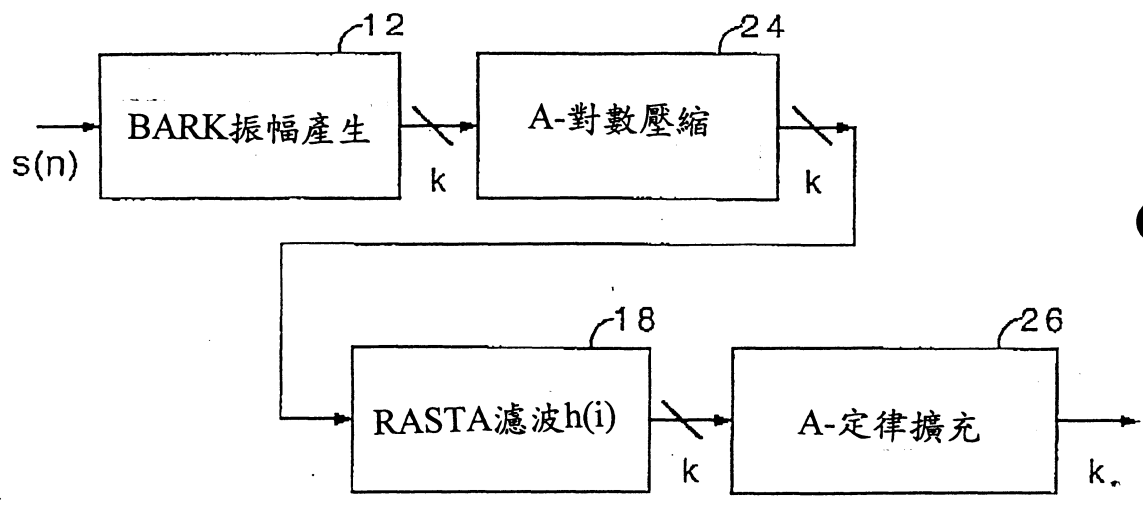


圖 8

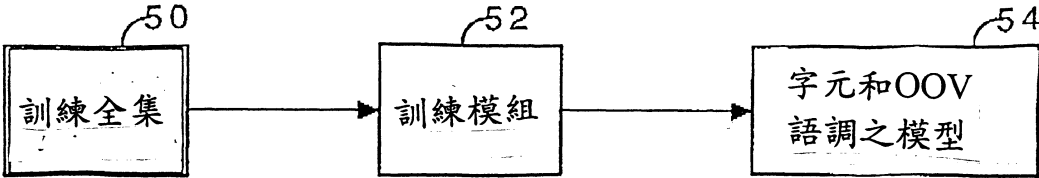


圖 9

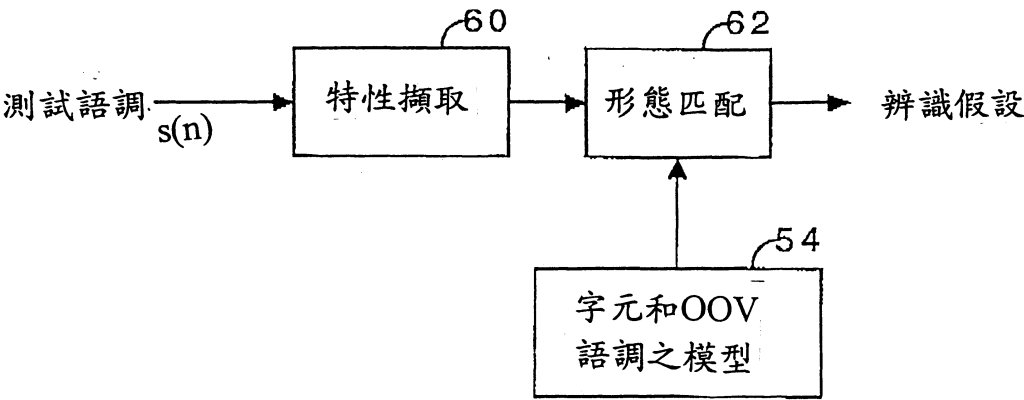


圖 10

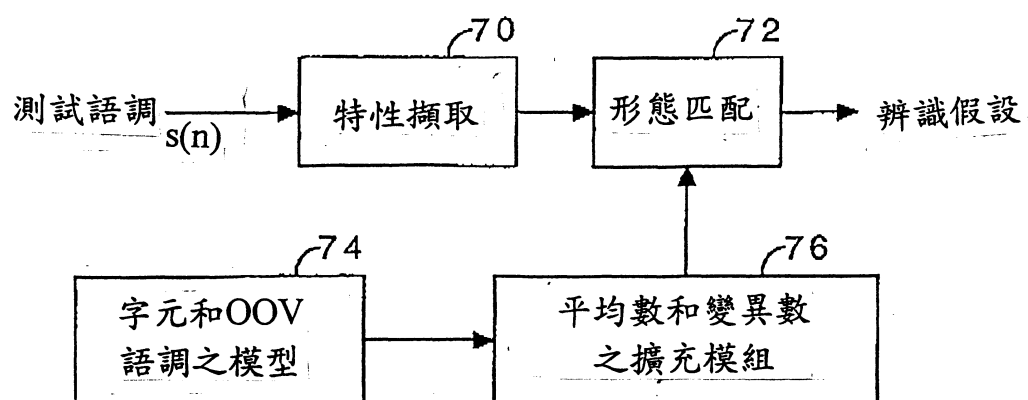


圖 11