

公 告 本

318239

申請日期	83. 11. 15
案 號	83110578
類 別	Int. Cl. 6 G10 L 7/08

A4
318239^{G4}

(以上各欄由本局填註)

發 明 專 利 說 明 書

一、發明 名稱	中 文	分散式語音辨識系統
	英 文	"DISTRIBUTED VOICE RECOGNITION SYSTEM"
二、發明 創作人	姓 名	1. 保羅·E·傑可伯 2. 張 建 忠
	國 籍	均美國
	住、居所	1. 美國加州聖地牙哥市托倫斯街1684號 2. 美國加州聖地牙哥市塞普瑞斯街11456號
三、申請人	姓 名 (名稱)	美商奎康公司
	國 籍	美國
	住、居所 (事務所)	美國加州聖地牙哥市拉斯克大道6455號
	代 表 人 姓 名	艾文·M·傑可伯

裝 訂 線

318239

(由本局填寫)

承辦人代碼：
大類：
IPC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 1993.12.22 08/173,247

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

經濟部中央標準局員工消費合作社印製

四、中文發明摘要 (發明之名稱：分散式語音辨識系統)

一種語音辨識系統，其中一特徵抽取裝置放置於一遠端工作站。此特徵抽取裝置從一輸入的語言架構中抽取特徵，然後將此抽取的特徵提供給一中央處理工作站。在此中央處理工作站中，此特徵提供給一文字解碼器，其決定此輸入語言架構的語法。

英文發明摘要 (發明之名稱：“DISTRIBUTED VOICE RECOGNITION SYSTEM”)

A voice recognition system is disclosed wherein a feature extraction apparatus is located in a remote station. The feature extraction apparatus extracts features from an input speech frame and then provides the extracted features to a central processing station. In the central processing station, the features are provided to a word decoder which determines the syntax of the input speech frame.

六、申請專利範圍

1. 一種語音辨識系統包含：

設置於一遠端站的特徵抽取裝置，用以接收語言取樣架構，並且根據一預定的特徵抽取格式，從該語言取樣的架構中抽取一組語言特徵，並用以提供該組語言特徵；以及

設置於一中央處理站的文字解碼器，用以接收該組語言特徵，並根據一預定的解碼格式來決定一語法。

2. 根據申請專利範圍第1項的系統，其中該組特徵是線性預測編碼參數。

3. 根據申請專利範圍第1項的系統，更包含一設置在該遠端站的局部文字解碼器，用以根據一預定的小型字彙解碼格式來決定一語法。

(請先閱讀背面之注意事項再填寫本頁)

訂

318239

第八三一〇五七八號專利申請案
中文修正圖式(八十五年十二月)

修正
補充

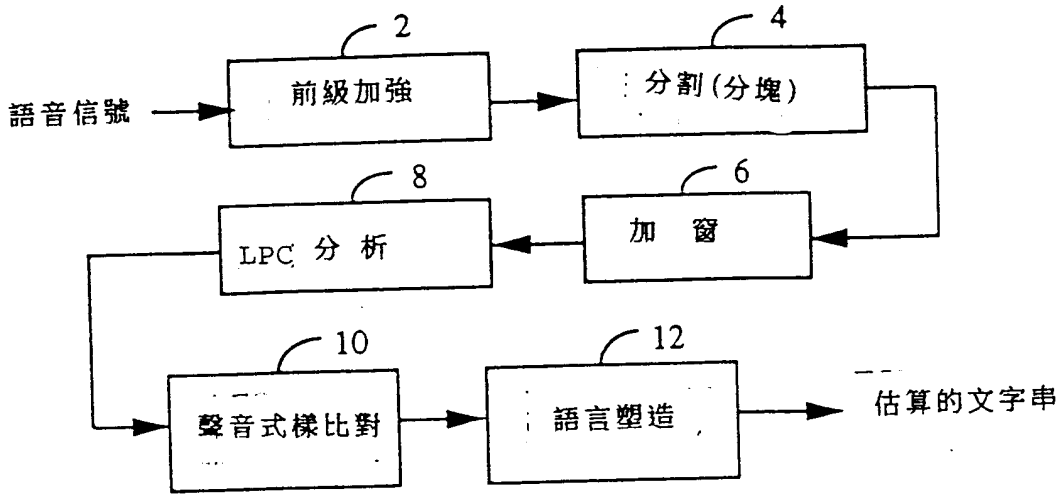


圖 1

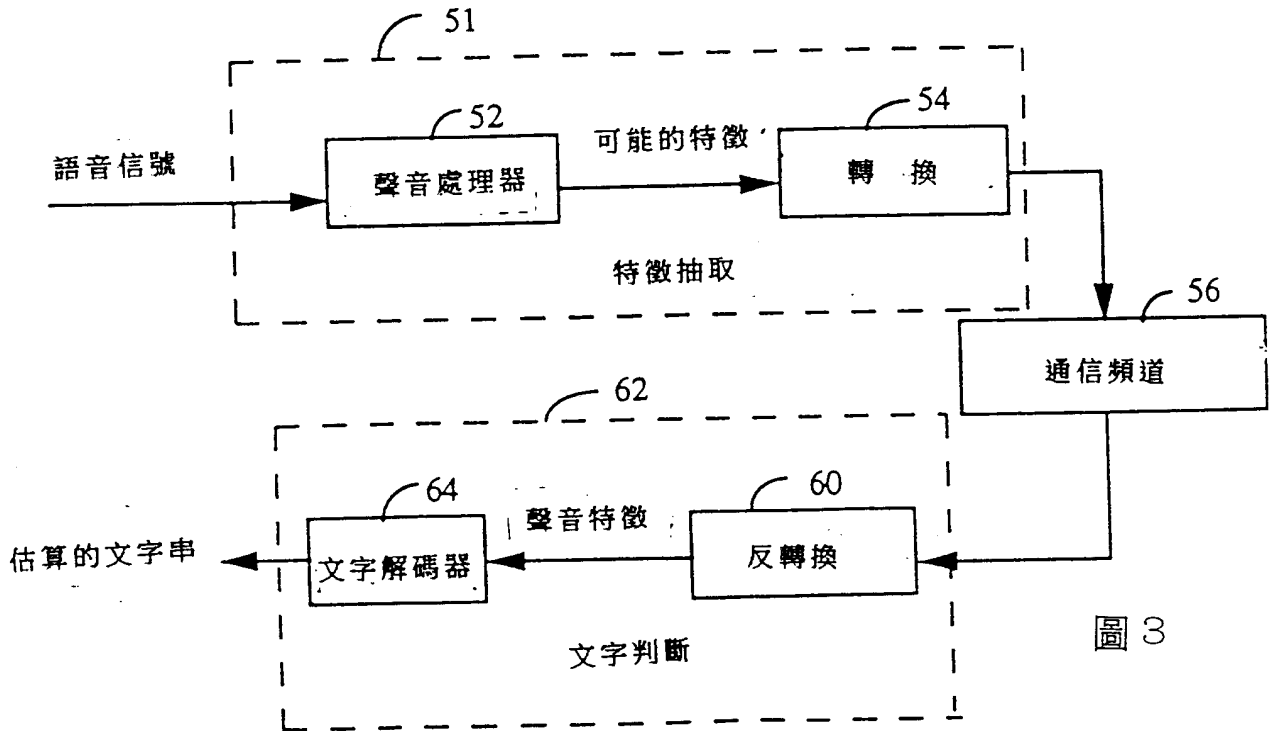


圖 3

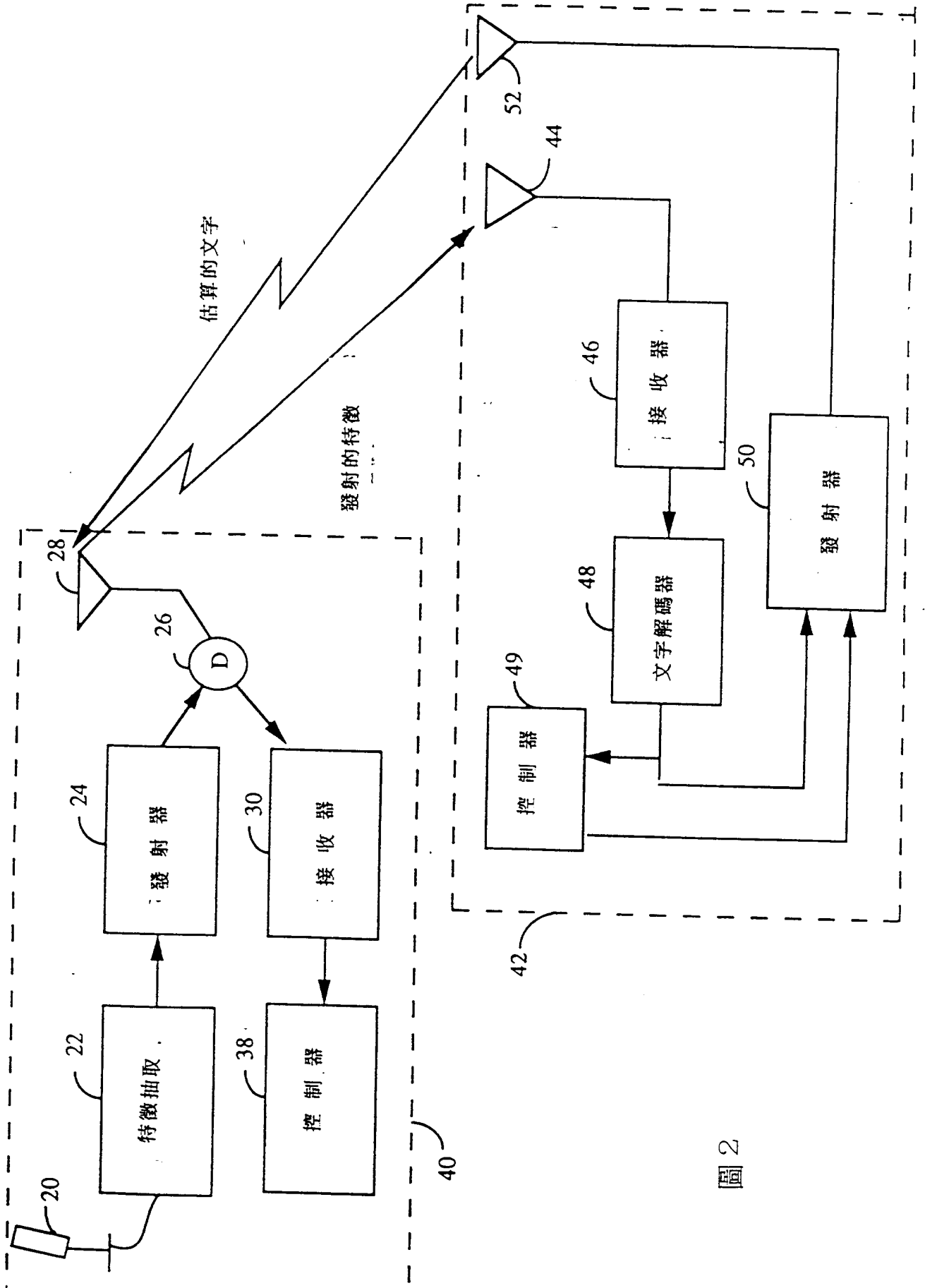


圖 2

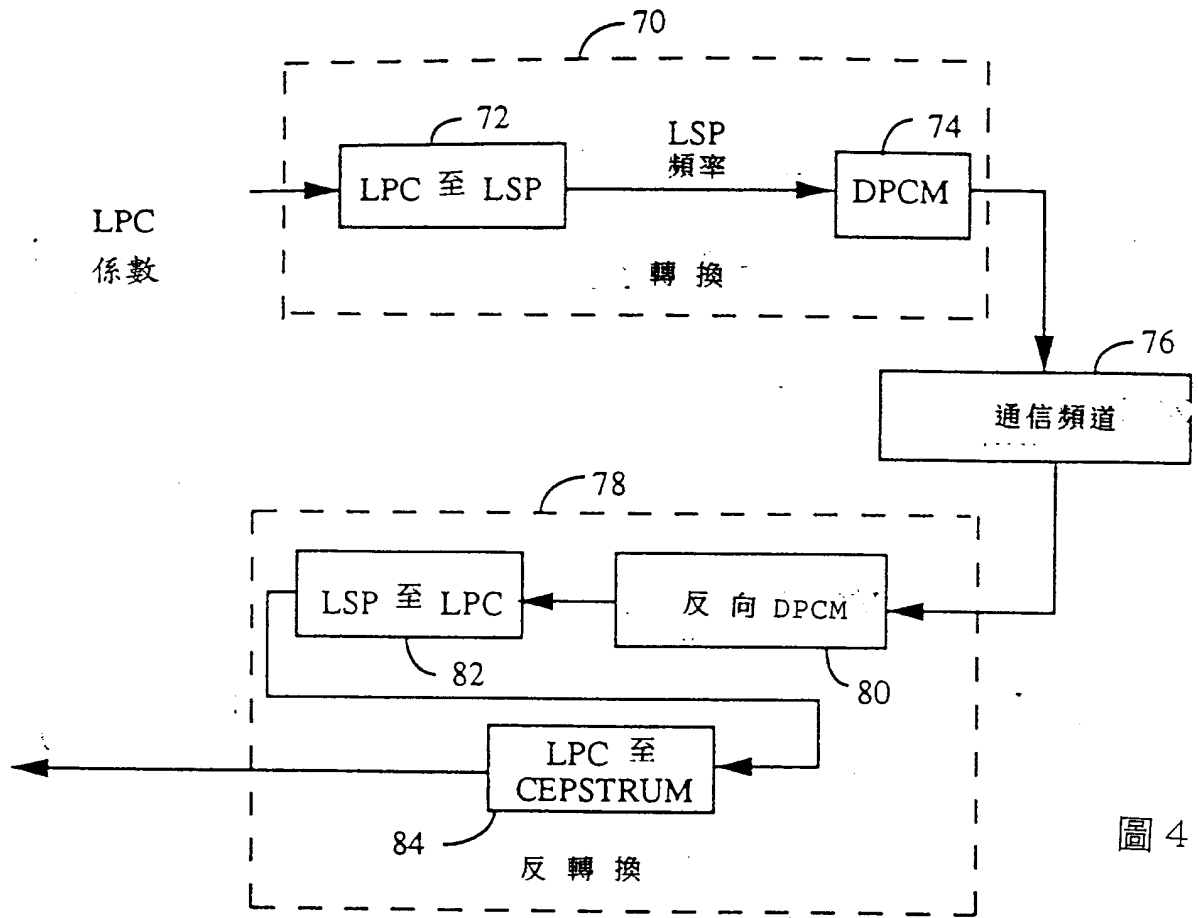


圖 4

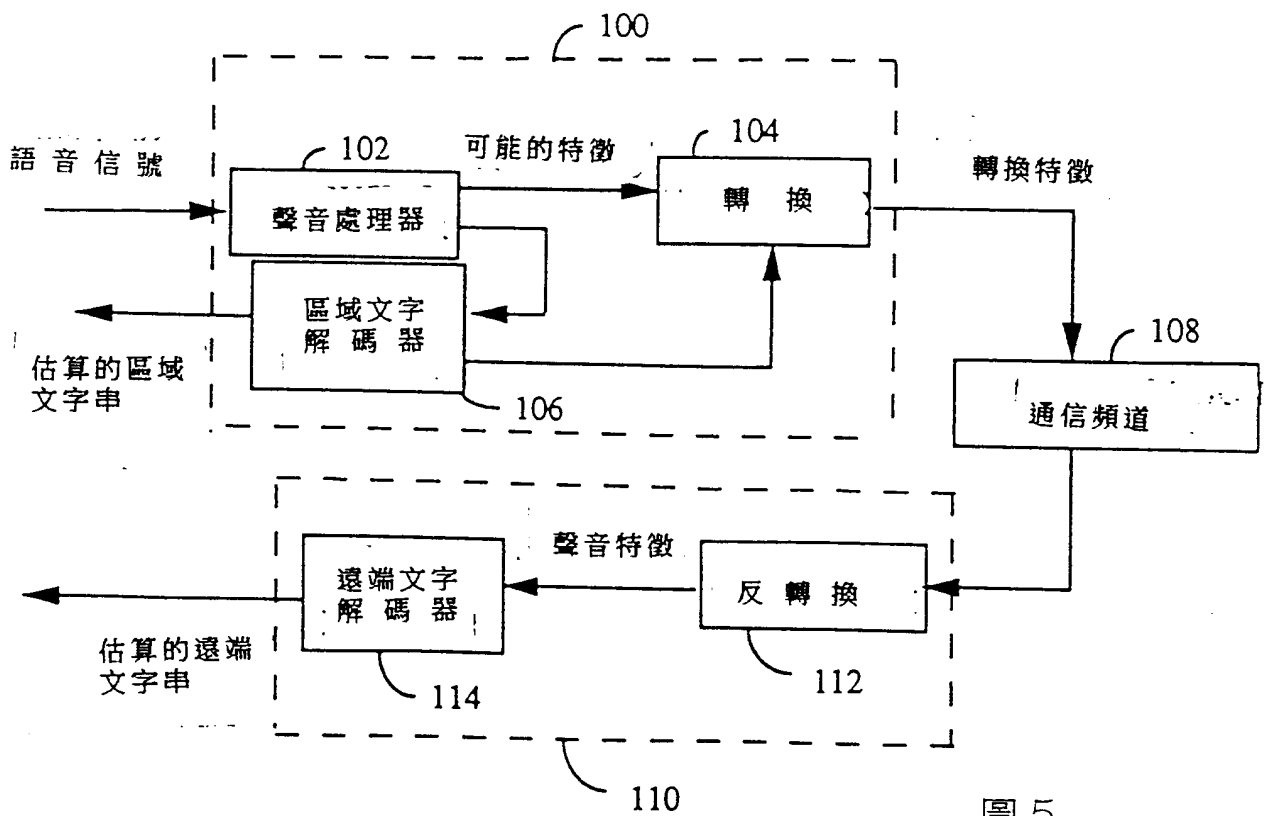


圖 5

83年12月31日 83110578號專利申請案
修正頁

五、發明說明()

發明範圍

本發明係關於語音信號的處理。更特別地是，本發明係關於一種無前例的方法和裝置，用來完成一標準語音辨識系統的一分散式施行。

相關技術的描述

語音辨識表示賦與一具有模擬智慧的機器辨認使用者或使用者說出的指令，以及使人類與機器的介面變得較容易的最重要的技術之一。其也表示人類語音理解的一種關鍵技術。使用從一聲音的語音信號恢復一語言的訊息之技術的系統稱為語音辨識器(VR)。一語音辨識器由一聲音處理器，其從輸入的原始語音中抽取作VR所必須之連續的資料關聯特徵(向量)和一解讀此連續的特徵(向量)以產生有意義的和想要的輸出格式，像一系列相當於輸入發聲的語言的文字之文字解碼器所組成。為增加一指定系統的成效，需要訓練用有效的參數來裝備此系統。換言之，此系統在其最佳地工作之前必須學習。

此聲音處理器在一語音辨識器中，表示一前端語音分析子系統。響應一輸入語音信號，其提供一適當的表示來記述此時變語音信號的特徵。應該捨棄無關的訊息，像背景雜訊，波道失真、說話者特徵和說話的習慣。有效率的聲音特徵將供給語音辨識器較高的識別能力。最有用的特性為短時間光譜包絡。在描述此短時間頻譜的包絡的特徵中，二種最平常使用的頻譜分析技術為線性預估碼(LPC)和

(請先閱讀背面之注意事項再填寫本頁)

表
訂

經濟部中央標準局員工消費合作社印製

五、發明說明()

以濾波器組為基底的頻譜分析模型。然而，其已經顯示（像1978年Prentice Hall, Rabiner, L.R.和Schafer, R.W.所討論的“語音信號的數位處理”），對於聲道頻譜波包，LPC不僅僅提供一良好的近似，也在所有數位程序中比濾波器組模型，在計算上要便宜很多。經驗也證明，LPC為基礎的語音辨識器的效能不比濾波器組為基礎的辨識器差或更好（1993年Prentice Hall, Rabiner, L.R.和Juang, B.H.“語音辨識的基礎”）。

參考圖1，在一LPC基礎的聲音處理器中，輸入的語音提供給一麥克風（未示出）並轉換成一類比電子信號。然後此一電子信號經由一A/D轉換器（未示出）而被數位化。此數位化的語音信號通過前級加強濾波器2，為使此信號藉由頻譜地弄平，並且使其較不容許在其後的信號處理中的有限精度效果。然後此前級加強濾波的語音送至分割元件4，在該處被分割或阻斷成不是暫時地重疊就是非重疊區塊。然後此語音數據的架構送至加窗元件6，在該處移去加框的直流分量，而一數位加窗的運算在每一架構上完成，以減少由於架構邊界上不連續性所造成的阻斷效應。在LPC分析中最常使用的窗函數為海明(Hamming)窗， $w(n)$ 定義為：

$$w(n) = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1 \quad 1$$

此加窗的語音供給LPC分析元件8。在LPC元件8中，以

(請先閱讀背面之注意事項再填寫本頁)

衣

訂

經濟部中央標準局員工消費合作社印製

五、發明說明()

加窗的取樣為基礎計算自相關函數，而直接從自相關函數中獲得相關的LPC參數。

通常來說，此文字解碼器將由聲音處理器所產生的聲音特徵序列翻譯成一說話者原始的文字串的估計。此被完成在以下二步驟：聲音式樣比對和語言塑造。語言塑造在分離文字辨識的應用中是無效的。從LPC元件8所得到的LPC參數供給聲音式樣比對元件10，以偵測和分類可能的聲音式樣；像音素、音節、字等等。候選的式樣送至語言塑造元件12，其塑造依照造句法的限制規則來決定什麼樣的文字序列較合乎文法且有意義。當單獨的聲音訊息是不明確的時，對於語音辨識，依照造句法的訊息是重要的指引。以語言塑造為基礎，VR連續地判斷聲音特徵比對結果，並提供估計的文字串。在文字解碼器中，聲音式樣比對和語言塑造二者都需要一數學模型（不是決定論的，就是推測的）來描述說話者音韻上的和聲音語音上的變化。一語音辨識系統的效能直接和此二模型的品質有關。在不同聲音式樣比對模型的種類中，基於樣型比對的動態時間校正（DTW）和推測的隱藏式馬可夫模型（HMM）是最常使用的二種。然而，已經證明DTW為基礎的近似可以看成HMM為基礎者的一特殊情況，其為一變數的雙重推測的模型。HMM系統是目前最成功的語音辨識的規則系統。在HMM中的雙重推測特性提供在吸收的聲音上和由語音信號所組成的暫時變化一樣較具質性。此通常造成辨識精確度的改善。考慮語言模型，一推測的模型稱為K文字語言模型（其細節

（請先閱讀背面之注意事項再填寫本頁）

裝

訂

五、發明說明(4)

描述於1985年IEEE會議第73卷1616至1624頁F. Jelinek所著“一實驗用不連續的口述辨識器的發展”中)已經成功地應用於實際的大型字彙語音系統。當在小的字彙情況中，一決定論的文法已經表示為一有限狀態網路(FSN)，應用於航空公司和預約以及資料系統(見Rabiner, L.R. 和Levinson, S.Z.所著“一種以隱藏式馬可夫模型和位準建造為基礎的與說話者無關，直接語法，連接的文字辨識系統”，1985年6月卷33，第3號，LASSP，IEEE翻譯。

在統計上，為了使辨識錯誤的機率最小，語音辨識問題可以形式化如下：用聲音痕跡觀察 O ，語音辨識的運算將發現最類似的文字串 W^* ，像：

$$W^* = \arg \max P(W|O) \quad (1)$$

其中最大化涵蓋所有可能的文字串 W 。根據貝斯(Bayes)規則，在上面的方程式中後面的機率 $P(W|O)$ 可以重寫為：

$$P(W|O) = \frac{P(W)P(O|W)}{P(O)} \quad (2)$$

因為 $P(O)$ 與辨識無關，此文字串估計可由以下獲得：

$$W^* = \arg \max P(W)P(O|W) \quad (3)$$

此處 $P(W)$ 表示文字串 W 將會被說出的先前機率，而 $P(O|W)$ 是說話者說出此文字序列 W 使聲音痕跡 O 被觀察的機率。

五、發明說明 (5)

$P(O|W)$ 由聲音式樣比對來決定，而先前機率 $P(W)$ 由使用語言模型來定義。

在連接的文字辨識中，如果字彙是小的（小於100），一決定性的文法可用來嚴格地支配在語言中，那些字可合乎邏輯地跟隨在其他文字之後以形成合法的句子。此決定性的文法可以與聲音比對規則系統結合，暗中地限制可能的字的搜尋空間，並減少引人注目的計算。然而，當字彙的大小不是中型的（大於100但是小於1000）就是大型的（大於1000）時，此文字列 $W=(w_1, w_2, \dots, w_n)$ 的機率，可以由推測的語言模型獲得。由簡單的機率理論，先前的機率， $P(W)$ 可以分解成：

$$P(W)=P(w_1, w_2, \dots, w_n)=\prod_{i=1}^n P(w_i | w_1, w_2, \dots, w_{i-1}) \quad (4)$$

其中 $P(w_i | w_1, w_2, \dots, w_{i-1})$ 是將文字列 $(w_1, w_2, \dots, w_{i-1})$ 加於 w_i 之前， w_i 將被說出的機率。 w_i 的選擇與整個過去輸入文字的歷史有關。對於一個大小為 V 的字彙，需要 V^i 個值來完全地指定 $P(w_i | w_1, w_2, \dots, w_{i-1})$ 。即使對於中型的字彙大小，也需要非常大的取樣數目來訓練此語言模型。由於不足的訓練數據所引起的 $P(w_i | w_1, w_2, \dots, w_{i-1})$ 的不正確估計，將會減損原始聲音比對的結果。

對於上述問題的一種實際的解答是假設 w_i 僅與先前的 $k-1$ 個文字有關 $(w_{i-1}, w_{i-2}, \dots, w_{i-k+1})$ 。一推測的語言

五、發明說明()

模型可以完全地以 $P(w_i | w_1, w_2, \dots, w_{i-1})$ 的項來描述，由此推導出 k 文字語言模型。因為如果 $k > 3$ ，大部分的文字串將不會發生在語言中，單一文字 ($k=1$)，雙文字 ($k=2$) 和三文字 ($k=3$) 是最有力的推測語言模型，其統計上地加以考慮文法。語言模型句含依照造句法的和與語意有關的訊息二者，其在辨識中是重要的，但是這些可能性必須從一個大的語音數據的聚集中被訓練。當可用以訓練的數據相對地被限制時，像 k 文字將不會發生在此數據中， $P(w_i | w_{i-2}, w_{i-1})$ 可以直接由雙文法機率 $P(w_i | w_{i-1})$ 來估計。此程序的細節說明可以在 1985 年 IEEE 會議卷 73 第 1616 至 1624 頁，F. Jelinek 的“一實驗用不連續的口述辨識器的發展”中找到。在連接的文字辨識器中，整個文字模型用作基本語音單元，而在連續的語音辨識中，次能帶單元，像音素，音節或半音節可以用作基本語音單元。此文字解碼器將據此修正。

慣用的語音辨識系統結合聲音處理器和文字解碼器並未考慮其可分離性，應用系統的限制（像功率消耗、記憶體可用性等）和通信頻道特性。此激起發明一分散式語音辨識系統使此二元件適當地分離的興趣。

發明摘要

本發明為一新奇的及改良的分散式語音辨識系統，其中 (i) 前端聲音處理器可以是 LPC 基礎或濾波器組基礎；(ii) 在文字解碼器中的聲音式樣比對可以用隱藏式馬可夫模型

五、發明說明(7)

(HMM) , 動態時間彎曲(DTW) 或甚至是神經網路(NN)為基礎; 以及(iii)對於連接的或連續的文字辨識目的, 語言模型可以用決定論的或推測的文法為基礎。本發明在藉由適當地分開元件(特徵抽取, 和文字解碼器)來改良系統效能上, 與常用的語音辨識器不同。像在下一個例子中說明的, 如果LPC為基礎的特徵(像Cepstrum係數)將被送過通信頻道, 一LPC和LSP之間的轉換可以使用於減輕在特徵序列上的雜訊效應。

圖例概述

本發明的特徵、目的和優點, 從敘述於下的細節描述會變得更明顯, 當與圖例同時看, 其中類似的參考字元在整個的說明中相對地相同; 而其中:

圖1 為一傳統的語音辨識系統的方塊圖;

圖2 為一本發明的實例成品在一無線通信環境中的方塊圖;

圖3 為本發明的一般方塊圖;

圖4 為一本發明的轉換元件和反轉換元件之一實例具體裝置的方塊圖; 以及

圖5 為一本發明的較佳具體裝置的方塊圖, 包含一區域文字解碼器加上一遠方文字解碼器。

較佳具體裝置的細節描述

在一個標準的語音辨識器中, 不是在辨識就是在訓練中

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

紉

五、發明說明()

，大部分計算的複雜性集中在此語音辨識器的文字解碼器子系統中。在具有分散系統結構之語音辨識器的執行中，通常希望將文字解碼工作放在此子系統，其可適當的吸收計算的負擔。然而此聲音處理器應該位於儘可能地靠近語音源，以減低由信號處理及／或頻道產生的誤差所引起之量化誤差的效應。

本發明一種實例的成品說明在圖2中。在此實例的具體裝置中，環境為一無線通信系統，包含一可攜式蜂巢電話，或個人通信裝置40和一中央通信中心當作一細胞基地站42。在此實例的具體裝置中，出現分散的VR系統。在此分散的VR中，聲音的處理器或特徵抽取元件22放在個人通信裝置40中，而文字解碼器48放在中央通信中心內。如果不用分散式的VR，VR僅在可攜式蜂巢電話中完成，其將不可能實現，即使對於中型大小字彙，連接的文字辨識造成高的計算成本。另一方面，如果VR只放在基地站，由於編碼的語音和頻道影響的結合之語音信號的衰減，精確度會明顯的降低。很明顯地，有三個設計分散式系統之目的優點。第一是由於文字解碼器的硬體部分不再放在電話40中，因此降低了行動電話的成本。第二是降低可攜式電話40在電池（未示出）上的消耗，其起因於局部地完成計算密集的文字解碼器運算。第三是在辨識精確度中預期的改善，以及分散式系統的彈性和可延展性。

語音供給麥克風20，其轉換此語音信號成為電子信號，並提供給特徵抽取元件22。由麥克風20來的信號可以是類

(請先閱讀背面之注意事項再填寫本頁)

表

訂

五、發明說明 (9)

比的或數位的。如果信號是類比的，則一類比至數位轉換器（未示出）需要放在麥克風20和特徵抽取元件22之間。此語音信號提供給特徵抽取元件22。特徵抽取元件22抽取輸入語音的相關特徵，將用於解讀此輸入語音的語言解譯。一個可以用來估算語音特徵的例子是一輸入語音架構的頻率特徵。此經常用作輸入語音架構的線性預測編碼參數。然後此語音抽取的特徵供給發射器24編碼、調制及放大此抽取的特徵信號，並將此特徵經由雙工器26提供給天線28，在該處，語音特徵被發射至蜂巢式基地站或中央通信中心42。已在論文中所熟知的不同形式的數位編碼、調制和傳輸技巧都可以使用。

在中央通信中心42上，發射的特徵在天線44上被接收，並供給接收器46。接收器46可以完成解調和解讀接收的發射特徵的功能，提供給文字解碼器48。文字解碼器48從語音的特徵來決定語音之語言的估算，並提供一作用信號給發射器50。發射器50完成此作用信號的放大、調制和編碼的功能，並提供此放大的信號給天線52，其發射此估算的文字或一指令信號給可攜式電話40。發射器50也可以使用熟知的數位編碼、調制或傳輸技術。

在可攜式電話40上，估算的文字或指令信號在天線28上被接收，其經由雙工器26將接收的信號供給接收器30，其解調、解讀此信號，然後將此指令信號或估算的文字提供給控制元件38。反應此接收的指令信號或估算的文字，控制元件38提供預定的反應（例如，撥一電話號碼、提供訊

（請先閱讀背面之注意事項再填寫本頁）

裝

訂

線

五、發明說明()

息給在可攜式電話上的顯示螢幕，等)。

表示於圖2的相同系統也可以適合於一些微不同的情形，其中由中央通信中心42送回的訊息不需要是此發射語音的翻譯；反而由中央通信中心42送回的訊息可以是一反應由可攜式電話至解碼的訊息。例如，某人可能詢問在一遠端答錄機(未示出)上的訊息，經由一通信網路連接至中央通信中心42，在此情況中，信號由中央通信中心42發射至可攜式電話40，在此實施例中，可以是來自於答錄機的訊息。一第二控制元件49將被放置於中央通信中心內。

將特徵抽取元件22放置在可攜式電話40中而不放在中央通信中心42的意義說明如下。如果聲音處理器放置在中央通信中心42上，與分散式VR相反，一低頻寬的數位無線電頻道可能需要一自動語音編碼器(在第一子系統上)，由於量化的偏差，其限制特徵向量的解析度。然而，將聲音處理器放在可攜式或行動電話中，可以供給整個頻道的頻寬給特徵傳輸。通常抽取的聲音特徵向量所需要的頻寬比傳輸此語音信號為小。因為辨識的精確度和輸入語音信號的衰減非常有關，應該將特徵抽取元件22儘可能的靠近使用者，使得特徵抽取元件22以麥克風的語音為基礎，抽取特徵向量，而不是(傳聲的)電話語音，其可能在傳輸中加入訛誤。

在實際的應用中，語音辨識器被設計在周圍的條件下操作，像背景雜訊。因此考慮在出現雜訊中，語音辨識的問題是重要的。已經證明，如果字彙的訓練(參考式樣)完

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

五、發明說明(11)

成在像測試條件完全(或近似)相同的環境中完成,即使在非常大雜訊的環境中,語音辨識不僅能提供良好的效能,也能降低由於有意義的雜訊所產生之辨識精確度的衰減。訓練和測試條件之間的不吻合認為是辨識效能的主要衰減因素之一。假設聲音特徵傳過通信頻道能比語音信號更可靠(因為聲音特徵對於如上所述的傳輸所需要的頻寬比語音信號少),提出的分散式語音辨識系統在提供比對的條件上是有利的。如果一語音辨識器被遙遠地完成,主要由於像在無線通信中碰到的衰減的頻道變化,比對條件可能非常地破碎。局部地使用VR可以避免這些影響,如果可以局部地吸收大量的訓練計算。很不幸地,在許多應用中,這是不可能的。很明顯地,分散式語音辨識裝置可以避免由頻道混亂所引起的不吻合的條件,並且補償集中完成的缺點。

參考圖3,數位語音取樣供給特徵抽取元件51,其提供特徵通過通信頻道給文字判斷元件62,在該處一估算的文字串被決定。此語音信號提供給聲音處理器52,其對於每一個語音架構決定可能的特徵。因為文字解碼器對於辨識和訓練的工作都需要聲音特徵序列當輸入,這些聲音特徵必須發射通過通信頻道56。然而,並不是所有用於典型的語音辨識系統的可能特徵都適合於經由雜訊頻道的傳輸。在某些情況中,需要轉換元件22來幫助音源編碼並且降低頻道雜訊的影響。一個LPC為基礎聲音特徵的例子而廣泛的應用於語音辨識中的是cepstrum係數, {ci}。其可以直

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明 (12)

接由 LPC 係數 $\{a_i\}$ 獲得，如下：

$$c_m = a_m + \sum_{k=1}^{m-1} \binom{k}{m} c_k a_{m-k} \quad m=1, \dots, P \quad (5)$$

$$c_m = \sum_{k=1}^{m-1} \binom{k}{m} c_k a_{m-k} \quad m=P+1, \dots, Q \quad (6)$$

其中 P 是 LPC 濾波器使用的級數，而 Q 是 Cepstrum 特徵向量的大小。因為 Cepstrum 特徵向量的快速變化，並不容易壓縮一系列 Cepstrum 係數的架構。然而，存在 LPC 們和線光譜對 (LSP) 頻率之間的某一轉換，其緩慢的改變且能用一 δ 脈衝編碼調制 (DPCM) 技術有效地編碼。因為 Cepstrum 係數可以直接地由 LPC 係數推導出，藉由轉換元件 54 將 LPC 們轉換成 LSP 們，然後編碼傳過通信頻道 56。在遠端文字判斷元件 62 上，轉換的可能特徵藉由反轉換元件 60 作反轉換，提供聲音特徵給文字解碼器 64，其相應地產生一估算的文字串。

轉換器元件 54 的一實例的具體裝置說明於圖 4 中，為一轉換器子系統 700 在圖 4 中，得自於聲音處理器 52 的 LPC 係數被供給 LPC 至 LSP 轉換元件 72。在 LPC 至 LSP 元件 72 中，LSP 係數可以如下方法決定。對於第 P 級的 LPC 係數，當下列方程式之 P 的根在介於 Q 和 π 之間時，可獲得相關的 LSP 頻率：

$$P(w) = \cos 5w + p_1 \cos 4w + \dots + p_5 / 2 \quad (7)$$

$$Q(w) = \cos 5w + q_1 \cos 4w + \dots + q_5 / 2 \quad (8)$$

五、發明說明 (13)

其中， p_i 和 q_i 可遞歸地計算如下：

$$p_0 = q_0 = 1 \quad (9)$$

$$p_i = -a_i - a_{P-i} p_{i-1}, 1 \leq i \leq P/2 \quad (10)$$

$$q_i = -a_i + a_{P-i} q_{i-1}, 1 \leq i \leq P/2 \quad (11)$$

此 LSP 頻率供給 DPCM 元件 74，在該處為了傳輸整個通信頻道 76，其被編碼。

在反轉換元件 78 中，從頻道接收的信號通過反向 DPCM 元件 80 和 LPC 至 LSP 元件 82，以重建語音信號的 LSP 頻率。此 LPC 至 LSP 元件 72 的反向程序由 LSP 至 LPC 元件 82 完成，其轉換此 LSP 頻率回入 LPC 係數，以用在推導 Cepstrum 係數。LSP 至 LPC 元件 82 完成以下的轉換：

$$P(z) = (1+z^{-1}) \prod_{i=1}^{P/2} (1 - 2\cos(\omega_{2i-1})z^{-1} + z^{-2}) \quad (12)$$

$$Q(z) = (1-z^{-1}) \prod_{i=1}^{P/2} (1 - 2\cos(\omega_{2i})z^{-1} + z^{-2}) \quad (13)$$

$$A(z) = 1 - \sum_{i=1}^P a_i z^{-i} = \frac{P(z) + Q(z)}{2} \quad (14)$$

然後此 LPC 係數提供給 LPC 至 Cepstrum 元件 84，其根據方程式 5 和 6，產生 Cepstrum 係數給文字解碼器 64。

五、發明說明(14)

因為文字解碼器完全依賴一聲音特徵列，如果直接發射通過通信頻道的話，其可能容易產生雜訊，一可能的聲音特徵列推導出並在圖3中描述的子系統51中轉換成另一表示，以方便傳輸。使用在文字解碼器中的聲音特徵列，其後經由反轉換而獲得。因此，在分散式VR的裝置中，送過空氣（頻道）的特徵列可與實際使用在文字解碼器中的不同。我們期望由轉換元件70的輸出更可以用文章中已知的任何錯誤保護的技術來編碼。

在圖5中，表示本發明的一種改良的具體裝置。在無線通信應用中，使用者可能由昂貴的通道存取緣故，不希望一些小量的簡單但常用的語音指令佔據通信頻道，此可以藉由在電話聽筒100和基地站110之間更分散此文字解碼功能來完成；實際上，具有相當小的字彙大小的語音辨識已局部地完成在聽筒上，而一具有較大字彙大小的第二語音辨識系統放在遠端的基地站。其共同使用聽筒上的聲音處理器。在局部文字解碼器的字彙表含有最廣泛使用的文字或文字串。另一方面在遠端文字解碼器中的字彙表包含規則的文字或文字串。基於在圖5中說明的下部結構，頻道忙線的平均時間可以縮短，而平均的辨識精確度會增加。

此外，將會存在二組可用的語音指令，一組稱為特殊語音指令，相當於可由局部的VR辨識的指令；另一組稱為規則語音指令，相當於那些不能用局部的VR辨識的。無論何時，當發出一特殊語音指令時，對於局部的文字解碼器，

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明 ()

真實的聲音特徵被抽取出來，而語音辨識功能未經由通信頻道而局部地完成。當發出一規則的語音指令時，轉換的聲音特徵向量經由頻道發射，而文字解碼在遠端的基地站中完成。

因為對於任何特殊語音指令，聲音特徵不需要被轉換，也不需要編碼，且對於局部的 VR，字彙的大小很小，因此需要的計算遠少於在遠端的（在整個可能的字彙中計算與搜尋正確的文字串，正比於字彙的大小）。此外，局部的語音辨識器也可用一簡化的 HMM 版本來作模型（像具有較低數目的陳述，對於陳述輸出機率具有較低的混音分量，等）。因為聲音特徵將直接送給局部的 VR，因此在頻道中沒有可能的干擾。此使得在聽筒（子系統 1）上 VR 的局部裝置可行，雖然限制了字彙，其計算的負擔也被限制了。想像此分散式 VR 結構也可用於不同於無線通信系統的其他目的的應用。

參考圖 5，語音信號送給語音處理器 102，然後抽取特徵（例如，LPC 為基礎的特徵參數），從語音信號中。然後這些特徵提供給局部的文字解碼器 106，其由其小型字彙中搜尋以辨識此輸入語音信號。如果不能解讀此輸入文字串並判斷遠端 VR 應該可以解讀，其信號轉換元件 104 會準備將此特徵傳輸。然後此轉換的特徵發射通過通信頻道 108 至遠端文字解碼器 110。在反向轉換元件 112 上接收此轉換的特徵，其完成轉換元件 104 的反向運算並提供聲音特徵給遠端文字解碼器元件 114，其相應地產生估算的

（請先閱讀背面之注意事項再填寫本頁）

衣

訂

五、發明說明 (16)

遠端文字串。

先前此較佳具體裝置的描述提供給任何有技術的人能夠製造或使用本發明。此具體裝置之不同的修正及那些文章中的技巧已經都知道，而此處所定義的一般性原則，可以應用於其他具體的裝置，沒有發明許可的使用。因此，本發明並不限於在此說明的具體裝置，也適合包含此處揭示的原則和新奇的特徵的最寬廣的視野。

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線