

申請日期：87.12.1	案號：87119918
類別：G10L 5/06	

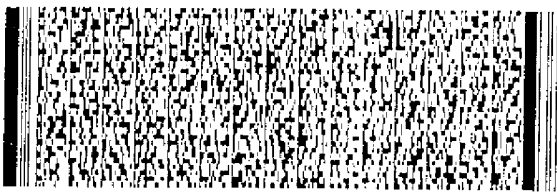
(以上各欄由本局填註)

# 公告本

## 發明專利說明書

462037

一、發明名稱	中文	編碼語音資訊之上下文相關音素網路
	英文	CONTEXT DEPENDENT PHONEME NETWORKS FOR ENCODING SPEECH INFORMATION
二、發明人	姓名(中文)	1. 斯利藍 巴拉克 2. 史堤芬 澳斯汀
	姓名(英文)	1. SREERAM BALAKRISHNAN 2. STEPHEN AUSTIN
	國籍	1. 英國 2. 英國
	住、居所	1. 美國加州三番市密蘇里路319號 2. 美國加州聖馬諦歐市南格蘭街701號
三、申請人	姓名(名稱)(中文)	1. 美商摩托羅拉公司
	姓名(名稱)(英文)	1. MOTOROLA INC.
	國籍	1. 美國
	住、居所(事務所)	1. 美國伊利諾州史堪伯市東阿崗崑路1303號摩托羅拉中心
	代表人姓名(中文)	1. F. 強 莫辛格
	代表人姓名(英文)	1. F. JOHN MOTSINGER



462037

本案已向

國(地區)申請專利

申請日期

案號

主張優先權

美國 US

1997/12/01 08/980,954

有

有關微生物已寄存於

寄存日期

寄存號碼

無



## 五、發明說明 (1)

## 1、發明之領域

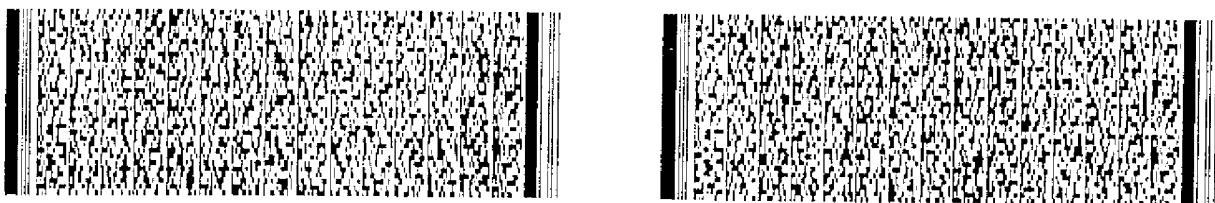
本發明通常係有關於電腦語音辨認。

## 2、發明之背景

最近在電腦硬體和軟體之進步係已允許電腦語音辨認 (CSR) 來跨越使用上之限制界線。現在做為可有用於高價位個人電腦之系統係可使用做為大量字彙、及連續語音檢測。為了得到適當之操作性，此些系統係必需適應特定使用者之聲音和使用環境。而且，此些系統係只能辨認從某些字彙中抽出之單字並且經常被結合成特別之語言模型，其係以捕捉不同之連續單字之相對機率。若無此些限制，則係很難從CSR系統得到適當之操作性。

在大部分之CSR系統中，使用者和環境指定部分、或聲音模型係經常被分開成字彙和語言模型。無論如何，因為上述之限制，需要語音辨認之任何應用係必需對使用者/環境之指定之聲音模型與應用之指定之字彙和語言模型兩者加以存取。

主要之障礙係對於許多不同之使用者需要存取各種不同之應用之系統、可能以相互平行並且經常是在網際網路或區域網路 (LAN) 上，要使CSR系統移動超越獨立檢測。理由為：(a) 每一應用將必需對每一使用者/環境保持分開之聲音模型；或 (b) 每一使用者係將需要為他們所希望使用之每一應用予以維護不同組之字彙和語言模型。因為聲音和語言模型之大小係針對一從中等至大型字彙之應用而言典型上皆需從數百萬位元到數千萬位元之等級，因此



## 五、發明說明 (2)

在情境 (a) 或 (b) 中，系統之資源均係很容易地被即克服。

一種可能係可將在不同引擎上之聲音模型予以儲存至字彙和語言模型，並且以經由區域網路 (LAN) 或網際網路予以連結引擎。無論如何，巨大數量之網路交通係可被產生來將數百萬資料之資料移至目標辨認器。

如此，存在於CSR系統上之需求係與應用之字彙和語言模型為無關而可不用犧牲相關於最後辨認精確度之操作性。

## 3、圖式之簡單說明

現在僅藉由例子來敘述本發明之理想實施例，並參考附隨之圖式，其為：

圖1係敘述簡單之音素網路；

圖2係根據本發明之理想實施例來敘述等於圖1之三音網路；

圖3係根據本發明之理想實施例來敘述做為圖2之三音網路之編碼；

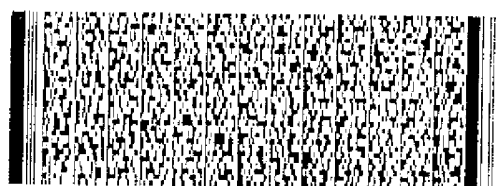
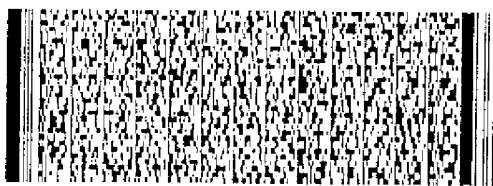
圖4係根據本發明之理想實施例之整個系統之通常方塊圖；

圖5係根據本發明之理想實施例之整個系統之詳細方塊圖；及

圖6係根據本發明之理想實施例之圖5之詳細方塊圖。

## 4、較佳實施例之詳細說明

上下文相關音素網路 (CD音素網路) 係對以上之技術思



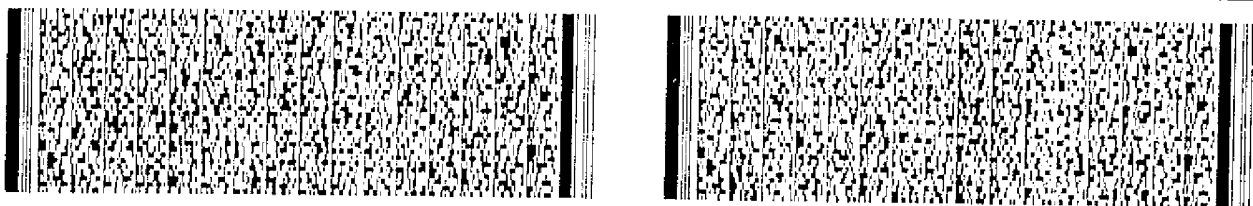
## 五、發明說明 (3)

想上之問題提供一個解決之道。而用以產生CD音素網路做為編碼語音資訊之中間的步驟之方法和裝置係被敘述於此。CD音素網路係導源於將語音辨認分成兩部分：係為使用者之客戶與可連續地傾聽和編碼使用者之語音以做為CD音素網路之特定環境；及可使用此些CD音素網路連同應用之特定的字彙和語言模型來操作最後之辨認之一套應用。

CD音素網路係提供一方法可用來編碼從與第一次（即，目標）應用之字彙和語言模型為無關之客戶之輸出：譬如為可使用不需對起初被使用來產生CD音素網路之使用者/環境音之特定之響模型做存取之此些CD音素網路之應用。重要的是，不像簡單之音素網路，CD音素網路係以一可不犧牲最後之以辨認精確度來看之操作性；及甚至可加強操作性之方式來編碼語音資訊。

有關於在此所提議之硬體，需要使用者之知識及環境之辨認方面係產生在使用者所在地點。因而從使用者之對於任何應用欲使自己所接受之投資達到槓桿作用之看法，此係具有有利之點。而從應用發展者之觀點，此係免除了必需維護使用者特定模型之頭痛問題，並能證明對於具有好幾千個使用者之網際網路服務之提供者而言為非常有吸引力。更進而，需要對較大語言模型和字彙做存取之辨認方面，其結果是硬碟空間之大容量，係產生在集中化之應用，因此，集中化應用之語言模型和字彙係以不需要下載大的資料檔案即能自動地成為可應用。

在詳細地敘述CD音素網路之前，先將簡單之音素網路敘



## 五、發明說明 (4)

述為用做語音辨認之中間階段係為有用的。如圖1之所  
示，音素網路係包含一組節點1、2、3及4；弧10、12、  
14、16、18及19。節點1、2、3及4係表示可能之音素之開  
始和結束時間。而以"從節點" 2 來鏈結"至節點" 3之弧  
10、12、14、16、18及19，係以在時間上對應於"從"和"  
至"之節點來表示特別音素之開始和結束。而關聯至每一  
弧10、12、14、16、18及19係音素之點數20或可能性。

簡單之音素網路之主要不完善之處和其為什麼未能使用  
成在大多數語音辨認之中間階段係弧20和22為上下文無  
關。換言之，在節點2 ( $t = 10$ ) 與3 ( $t = 35$ ) 間之點數  
22係無關於弧為被當成到達節點2或弧為從節點3取得。而  
CD音素網路係以使弧之點數為與上下文相關來加以改正。  
而弧之上下文係可被分成："從上下文"，係以被當成到達  
"從節點"之路徑來決定；及"至上下文"係以被取自"至節  
點"後之路徑來決定。

在更精巧之語音辨認網路中，"從"和"至"上下文係影響  
關聯於特殊之弧的音素之點數。而上下文被使用之方式係  
隨著系統的複雜程度而變化。在無上下文被使用之所在的  
簡單之音素網路係位在頻譜之一端。在一個三音之網路  
中，例如，當"至上下文"係關聯於取自"至節點"之第一弧  
之音素時，"從上下文"係關聯於被當成到達"從節點"之最  
後之弧的音素。甚至在更複雜之系統中，諸如五音網路，  
"從"和"至"上下文係包含被當成到達"從節點"和"至節點"  
之路徑之最後兩個和最先兩個弧。

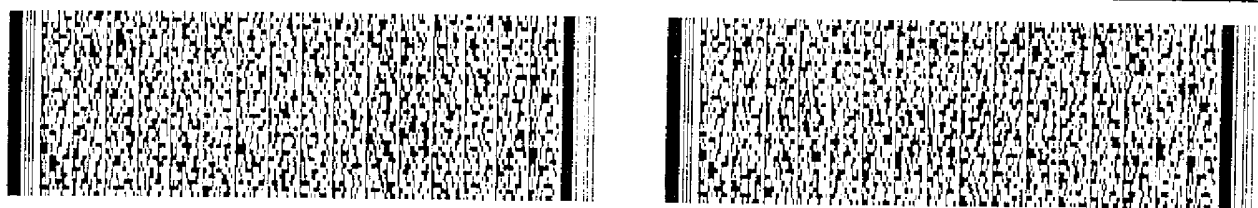


## 五、發明說明 (5)

為了確定一個弧之"從"和"至"上下文係符合關聯於弧之點數的"從"和"至"上下文，弧之"從"和"至"節點係也需要被分開。而圖2係做為相等於圖1所示之簡單之音素網路之三音之特定例子之此種分開之說明。在圖2之情形，節點已被分開以便於進入任意節點（即，節點3a）之弧24和26兩個係具有相同之音素標記（即，AH），並且來自任意節點（即，節點3a）之弧28和30係具有相同之音素標記（即，K）。而因為"從"和"至"音素上下文係應可唯一地被指定，所以弧點數32和34係現在即可表示三音點數。例如，在連結節點2b至節點3d之弧38之點數36係以TH之左邊上下文和G之右邊上下文來表示三音點數。

圖3係說明做為圖2中所示之三音網路之編碼。圖3係包含關聯於時間之一串節點。對於每個節點，則具有包含源自於那特殊之節點之所有的弧之一子串弧。關聯於每一弧係音素標記、三音點數及"至節點"。因為三音網路係具有使屬於每一弧之三音點數，所以它們可服務做輸入至任意之縱橫字謎語音辨認器。然後音素網路產生器係可對具有被給特殊之字彙和語言模型之最佳點數之路徑來搜尋三音網路。音素網路產生器並不需要具有任意使用者或環境特定聲音模型或甚至可對原來語音做存取。所有需要用來解碼在語音訊號中之單字之所有資訊係其本身被捕捉在三音網路中。

參考圖4和圖5，根據本發明之理想實施例，此處所敘述之資料記錄媒體40係在其上已經儲存指令和資料，當載入



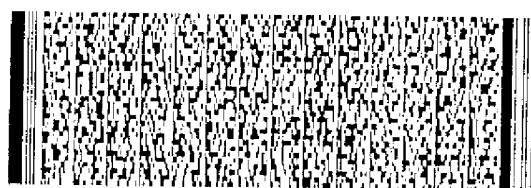
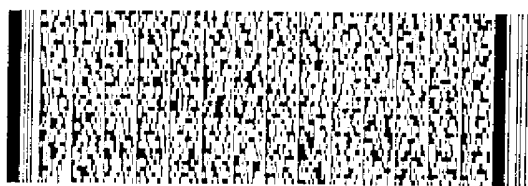
## 五、發明說明 (6)

具有操作系統44之至少第一通用目的微處理器42時，係引起至少第一通用目的微處理器42予以執行用以產生CD音素網路做為編碼語音資訊之中間步驟之方法。理想的是，第一通用目的微處理器42係包含具有音素網路產生器48（理想為CD音素網路產生器）之操作系統44之第一部分46；及具有至少第一應用52之操作系統44之第二部分50。

音素網路產生器48和第一應用52係與操作系統44並無相關且為動態連結庫之工具。操作系統44之第一和第二部分46和50，係也可形成在一單獨通用目的微處理器或另一電腦上之單位操作系統。另外，操作系統44之第二部分50係可被儲存在第二通用目的微處理器72上，而與第一部分46和第一通用目的微處理器42分開。

當操作系統44之第二部分50係分別至少具有第一和第二應用52和58，而兩個應用52和58之每一個係分別可具有不同之字彙60和62；及語言模型64和66。另外，第一和第二應用52和58係分別享有共通字彙68和共通語言模型70。

音素網路產生器48和第一應用52係彼此之位置可以遠隔地相分離並以連結56相連接，為無線連結或網路連結（即，操作系統44之第一部分46係可遠隔地與操作系統44之第二部分50相分離）。在操作上，音素網路產生器48係產生具有可定義CD音素網路之輸出54之一CD音素網路。音素網路產生器48之輸出54係在格式上可被儲存做為以後對第一應用52之輸入。第一應用52係適合於接收音素網路產生器48之輸出54。



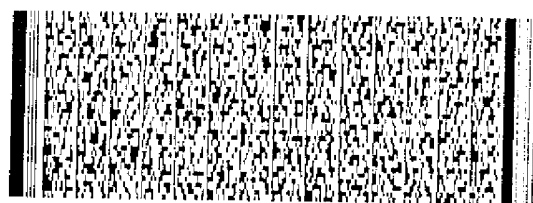
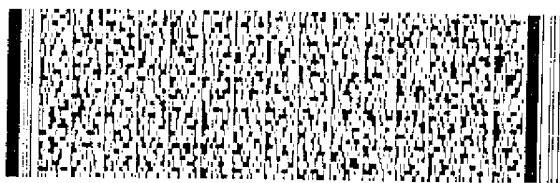
## 五、發明說明 (7)

參考圖6，音素網路產生器48係詳細地予以顯示具有：微音器74，係連接至可輪流地連接至操作系統44之第一部分46之輸入之類比數位轉換器74。具有音素之文字80之資料記錄媒體40係也被連接至操作系統44之第一部分46之輸入。

操作系統44之第一部分46係與使用者和環境特定聲音模型為相關。操作系統44之第一部分46係包含很多以軟體形成之元件。此些係含有：特色抽取元件76，係具有輸入和輸出；及維特比 (Viterbi) 解碼器78，係具有連接至特色抽取元件76之輸出之輸入並具有連接至第二部分50之輸入。另外，元件78 (維特比解碼器) 係也是鮑姆-威爾斯 (Baum-Welsh) 所建立之解碼器或元件60和62 (分別是特色抽取和維特比解碼器) 係也可以神經網路所建立之解碼器來加以結合並替代。

操作系統44之第二部分50係與可獨立地選擇自操作系統44之第一部分46之第一應用為相關。操作系統44之第二部分50係包含具有被耦合至維特比解碼器78之輸出82、84及86之輸入之第一應用52。維特比解碼器78之輸出82、84及86係分別為音素標記、點數和"從"節點-"至"節點對，它們係定義CD音素網路之弧。第一應用52係被耦合至字彙60和語言模型64。電腦程式之形式之圖形使用者介面88係設置成具有連接至第一應用52之輸入及具有耦合至顯示器90之輸出。

在操作上，操作系統44之第一部分46係產生CD音素網路

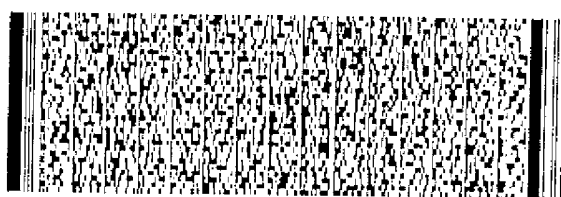
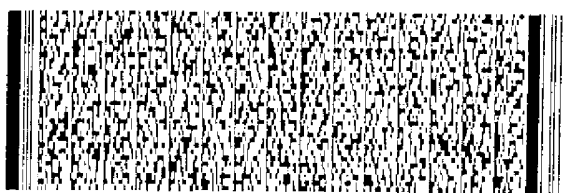


## 五、發明說明 (8)

用來捕捉使用者和環境特定聲音模型。因此，操作系統44之第一部分46係以可連續地適應於使用者之語音、環境及語言使用之使用者和環境特定聲音模型之輔助而將粗糙之語音轉換至CD音素網路。模型係可在個人電腦上、蜂巢式無線電話、呼叫器或任何其他合適之硬體配置件上運轉。操作系統44之第一部分46係從語音訊號來抽取語音辨認所要求之需要的資訊，將在字彙、語言模型、及使用者和環境無相關之格式之需要的資訊予以封包並予以傳送需要的資訊至操作系統44之第二部分50。

然後操作系統44之第二部分50係使用合適之搜尋引擎裝置從CD音素網路來抽取相關於使用第一應用52之字彙60和語言模型64之第一應用52之需求之資訊以便於擷取候選人辨認之結果並傳遞候選人辨認之結果至第一應用52。此種搜尋引擎裝置之變化係可將其涵蓋範圍想像為從偵測裝置、形成檔案裝置、SQL詢問裝置乃至於自然語言詢問裝置。

理想地，在第一搜尋引擎中嘗試辨認（即，操作系統44之第二部分50）之後，而在第一搜尋引擎中一有不滿意之辨認，CD音素網路係向前往第二引擎100。第二引擎100係接連地擷取候選人之辨識結果並傳遞候選人之辨識結果至第一應用52。將音素網路從第一引擎傳送至第二搜尋引擎之執行以達到正確率之可信水準，係可將其加以延伸至任意數目之搜尋引擎。例如，假若目前之搜尋引擎之可信水準尚未超過預定之臨限，則操作系統44之第二部分50係可



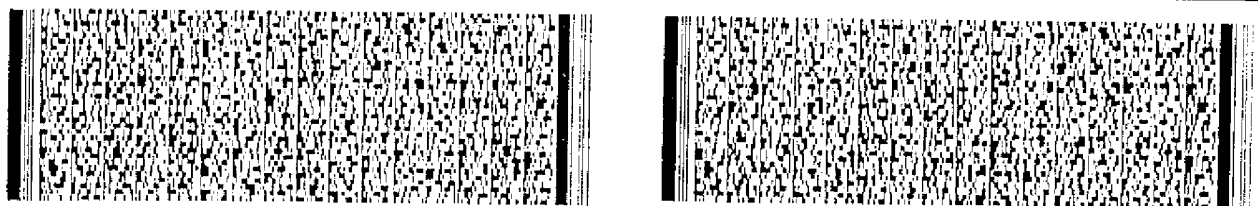
## 五、發明說明 (9)

建立來將CD音素網路傳送至一接連搜尋引擎。將CD音素網路從一搜尋引擎傳送至另一個係會繼續發生直至搜尋引擎可產生超過相對於正確率之可信水準之預定之臨限之候選人辨認結果。此外，操作系統44之第二部分係可同時地將CD音素網路送至至少兩個搜尋引擎。至少兩個搜尋引擎之正確率之可信水準係被比較且具有正確率之最高可信水準係被選擇來操作第一應用52。

以適切的搜尋引擎裝置來結合適當之字彙和語言模型，則發展者係可快速地語音化成為其應用。許多語言係可活動成如膠般地用來將裝置結合在一起，其係包含視窗培基語言和超文件置標語言(HTML)。在HTML之情形裏，假若裝置係設計做超文件鏈結、下拉選單、結合框及內文登錄欄，則很容易即了解到幾乎任何之網頁係可使語音化。

根據本發明之理想的具體例之此項趨近，係可以比起三音網路為多或少之上下文來延伸至網路。即將改變之所有係為需要來確定對於每一弧之正確的"從"和"至"上下文之節點分開之程度。音素網路產生器48本身係不需要知道用來決定每一弧點數之上下文之真正特性，因為網路本身係將加強可導引具有適於做為弧點數之上下文之"從"和"至"一弧之路徑之限制。

本發明之一方面之做為編碼語音資訊之方法係包含：從關聯於作業系統之音素網路產生器之語音來產生一上下文相關音素網路；將上下文相關音素網路傳輸至第一(或複數的)應用；及抽取在第一(或複數的)應用上之來自使



## 五、發明說明 (10)

用第一（或複數的）應用之字彙和語言模型之上下文相關音素網路所需要的資訊以便於操作第一（或複數的）應用。

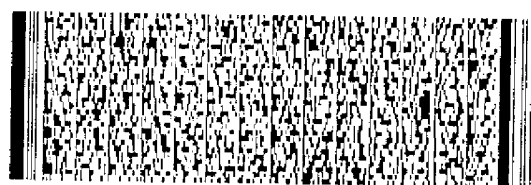
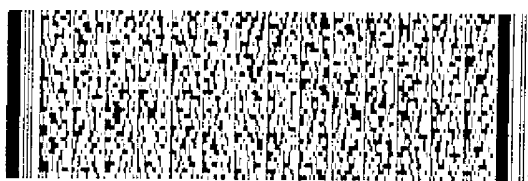
發明之另一方面之做為編碼語音資訊之方法係包含：提供含有兩部分之操作系統，其中，操作系統之第一部分係相關於使用者和環境特定聲音模型而操作系統之第二部分係相關於可獨立地選擇從操作系統之第一部分來之第一（或複數的）應用，其中，操作系統之第二部分係抽取來自使用第一（或複數的）應用之字彙和語言模型之上下文相關音素網路所需要的資訊。

發明之尚有另一方面之做為編碼語音資訊之裝置係包含：操作系統；產生具有可定義上下文相關音素網路之輸出之上下文相關網路之音素網路產生器；及適於接收音素網路產生器之輸出之第一（或複數的）應用並抽取來自使用第一（或複數的）應用之字彙和語言模型之所需要的資訊，其中，音素網路產生器和第一（或複數的）應用係與操作系統並無關聯。

音素網路產生器之輸出係在格式上可被儲存成對於第一（或複數的）應用之以後之應用。

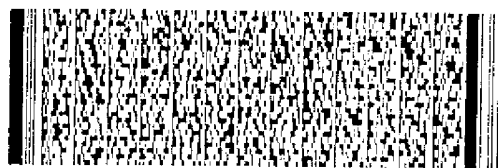
音素網路產生器和第一（或複數的）應用係可被遠隔地分離，並以無線連結或網路連結。

當發明係用與特定之實施例相連結般地來加以敘述，所加上之有利點和修正係很快地就會為熟習此技藝之人士所運用，本發明在其寬廣方面而言，並不因此便被限制在所



## 五、發明說明 (11)

示並敘述之指定之細節、代表之裝置及說明例上。不同之更改、修正及變化係對於熟習前面敘述所啟發出之技藝之人士而言為顯而易見的。而且，應可了解到本發明並不被限制在前面所述，除了在根據所附上之申請專利範圍之精神和觀點所做之所有如此之更改、修正及變化之外。

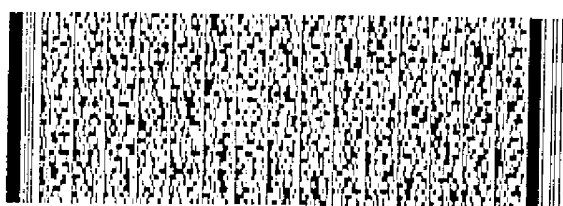


## 四、中文發明摘要 (發明之名稱：編碼語音資訊之上下文相關音素網路)

一種可產生上下文相關音素網路以做為編碼語音資訊之中間步驟之方法和裝置。上下文相關音素網路係從在關連於操作系統(44)之音素網路產生器(48)中之語音所產生。然後上下文相關音素網路係被傳輸至第一應用(52)。

## 英文發明摘要 (發明之名稱：CONTEXT DEPENDENT PHONEME NETWORKS FOR ENCODING SPEECH INFORMATION)

A method and apparatus for generating a context dependent phoneme network as an intermediate step of encoding speech information. The context dependent phoneme network is generated from speech in a phoneme network generator (48) associated with an operating system (44). The context dependent phoneme network is then transmitted to a first application (52).



## 六、申請專利範圍

## 1、一種編碼語音資訊之方法，包含：

使用適於使用者之聲音之聲音模型從音素網路產生器中之語音在本地使用者位置產生上下文相關音素網路，以作為語音辨認之中間步驟，其中上下文相關音素網路係代表以節點及弧表示之語音輸入，每一弧代表一具有以節點表示之開始及結束時間之音素點數，音素網路致使語音輸入以弧及節點表示，俾使語音輸入被封裝成一與字彙、語言模型、使用者及環境無關之中間格式；及

將上下文相關音素網路傳輸至一或多個遠離於本地使用者之應用程式，以致使遠端應用程式使用一由應用程式選擇之字彙或語言模型在每一應用程式執行語音辨認，俾使本地使用者位置不須執行語音辨認。

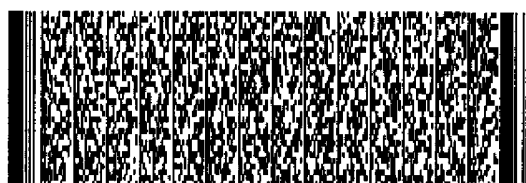
2、如申請專利範圍第1項之方法，其中上下文相關音素網路係與字彙和語言模型為無關。

## 3、如申請專利範圍第1項之方法，更包含：

在一第一應用上予以擷取需要來自使用第一應用之字彙和語言模型之上下文相關音素網路之資訊以便於操作第一應用。

## 4、一種編碼語音資訊之方法，包含：

提供包含兩部分之操作系統，其中操作系統之第一部分係與可連續適應於使用者之聲音之使用者和環境之特定聲音模型為相關；及操作系統之第二部分係相關於從操作系統之第一部分無關地選擇之第一應用，其中操作系統之第二部分係擷取從使用第一應用之字彙和語言模型所需要之



## 六、申請專利範圍

資訊。

5、如申請專利範圍第4項之方法，其中操作系統之第一部分係產生上下文相關音素網路，以便於捕捉使用者和環境之特定聲音模型並傳輸上下文相關音素網路至操作系統之第二部分。

6、如申請專利範圍第4項之方法，更包含第二應用，其中第一應用和第二應用係包含不同之字彙和不同之語言模型。

7、如申請專利範圍第4項之方法，更進而包含第二應用，其中第一應用和第二應用分享共同之字彙和共同之語言模型。

8、一種編碼語音資訊之方法，包含：

藉由使用者和環境之特定聲音模型和音素語言模型之輔助而轉換語音至上下文相關音素網路；

基於使用者和環境之特定聲音模型和音素語言模型而連續地接收上下文相關音素網路；

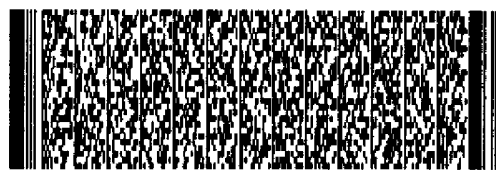
由語音訊號予以擷取做為語音辨認之所需要之資訊；及封裝在字彙、語言模型、及使用者和環境之獨立格式。

9、一種編碼語音資訊之裝置，係包含：

操作系統；

產生具有定義上下文相關音素網路音素網路之輸出之上下文相關音素網路音素網路之音素網路產生器；及

適於接收音素網路產生器之輸出之第一應用並從使用第一應用之字彙和語言模型之輸出予以擷取所需之資訊，其



## 六、申請專利範圍

中音素網路產生器和第一應用係與操作系統無關。

10、一種電腦可讀取之資料記錄媒體，包含指令和資料，被載入具有操作系統之第一通用目的微處理器，該微處理器具有一作業系統以使第一通用目的微處理器包含：在本地使用者位置產生具有定義上下文相關音素網路之輸出之上下文相關音素網路產生器，其中上下文相關音素網路致使語音輸入以節點及弧之形式表示，其中每一弧代表一具有以節點表示開始及結束時間之音素點數，俾使語音輸入被封裝成一中間格式；及適於接收音素網路產生器之輸出之複數個遠離於本地使用者位置之應用程式，並從使用複數個應用程式之字彙和語言模型之輸出以擷取所需之資訊，俾使不須自本地使用者位置擷取資訊，音素網路產生器和複數個應用程式與操作系統無關。

11、如申請專利範圍第10項之資料記錄媒體，其中資料記錄媒體係包含具有儲存在其上之音素網路產生器之第一部及具有存在其上之複數個應用之第二部。

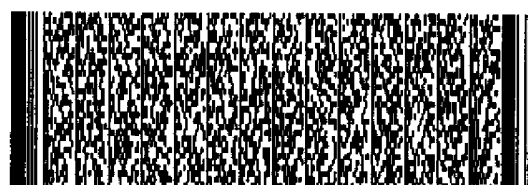
12、一種方法，包含：

產生上下文相關音素網路以做為編碼語音資訊之中間步驟；

在第一搜尋引擎中嘗試辨認；

在第一搜尋引擎中辨認不滿意之下，使上下文相關音素網路前進至第二搜尋引擎；及

從第二搜尋引擎擷取候選者之辨認結果。



圖式

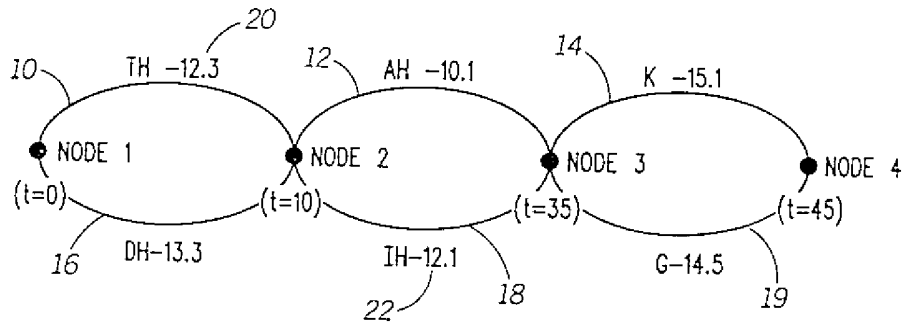


圖 1

NODE 1a, TIME=0, ARCS=2  
 TH-12.1, TO 2a  
 TH-12.5, TO 2b  
 NODE 1b, TIME=0, ARCS=2  
 DH-13.1, TO 2c  
 DH-13.5, TO 2d  
 NODE 2a, TIME=10, ARCS=2  
 AH-9.8, TO 3a  
 AH-9.6, TO 3b  
 NODE 2b, TIME=10, ARCS=2  
 IH-12.3, TO 3c  
 IH-12.6, TO 3d  
 ...etc...

圖 3

圖式

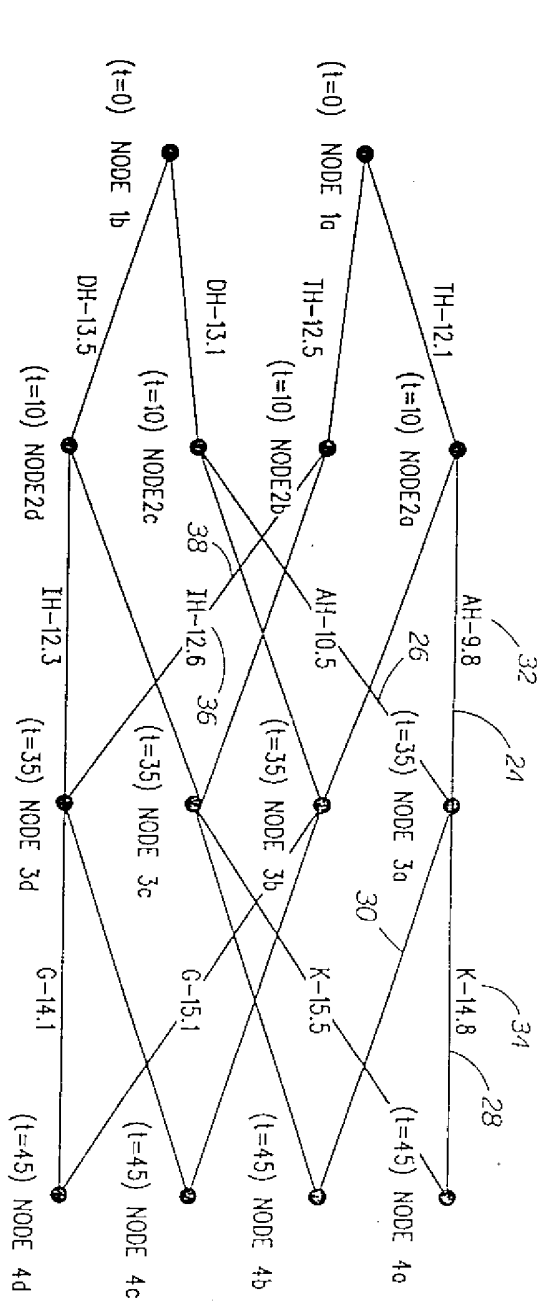


圖 2

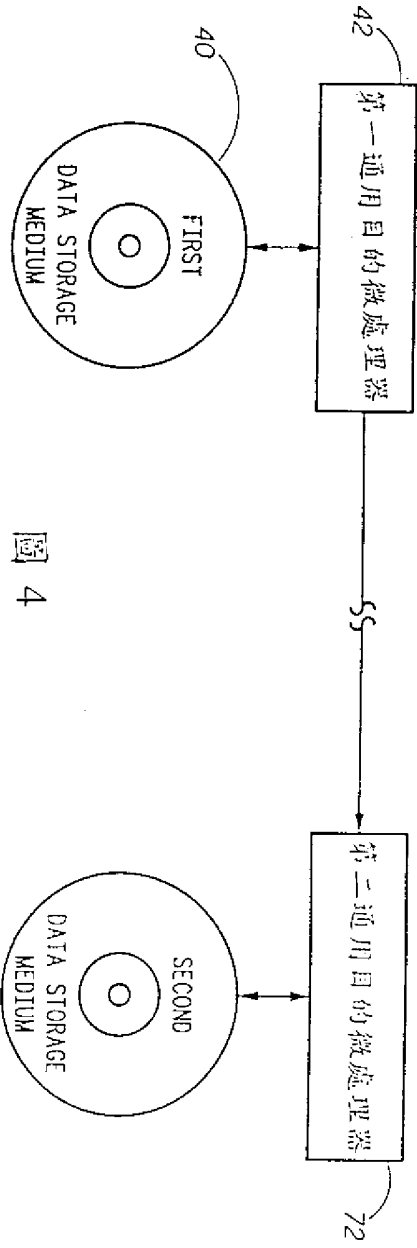


圖 4

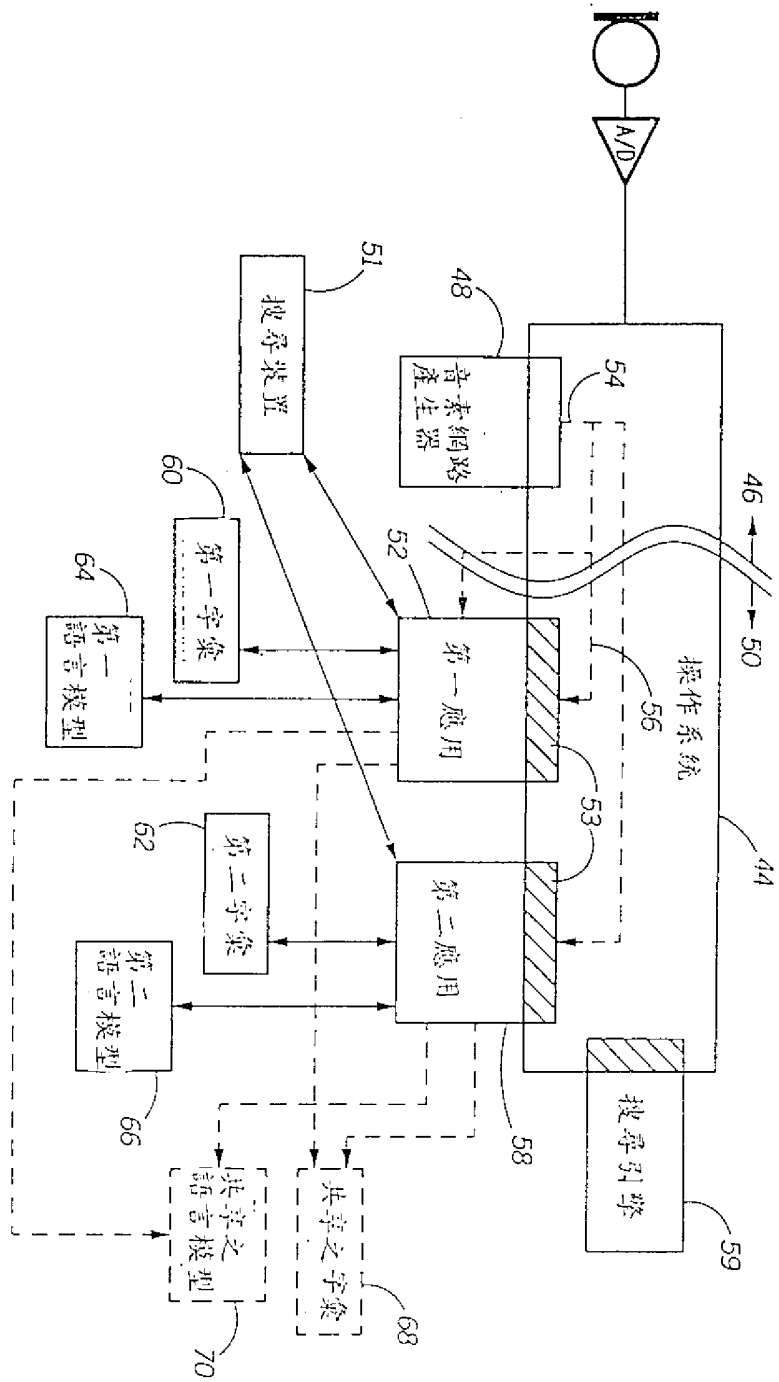
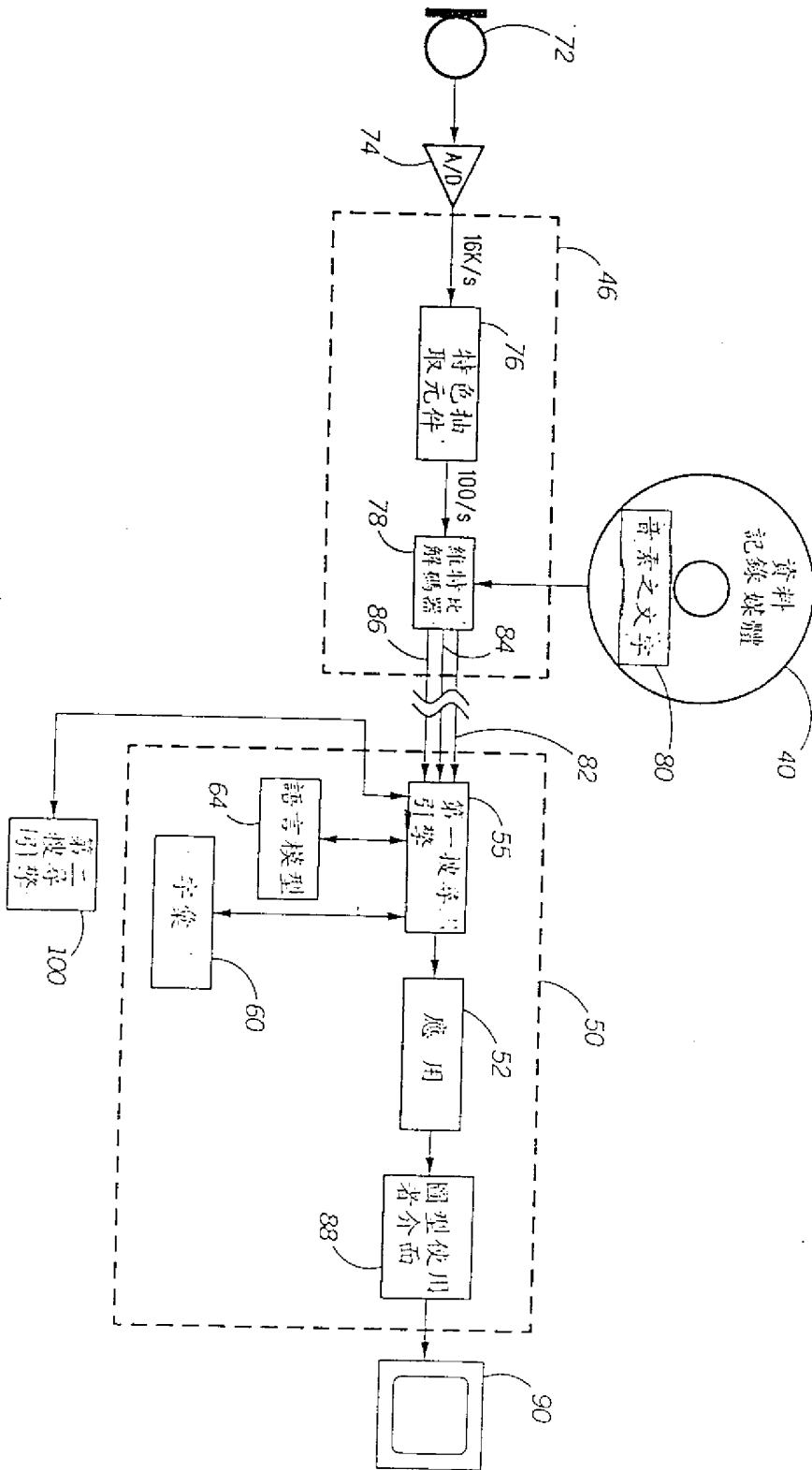


圖 5

圖 6



## 六、申請專利範圍

## 1、一種編碼語音資訊之方法，包含：

使用適於使用者之聲音之聲音模型從音素網路產生器中之語音在本地使用者位置產生上下文相關音素網路，以作為語音辨認之中間步驟，其中上下文相關音素網路係代表以節點及弧表示之語音輸入，每一弧代表一具有以節點表示之開始及結束時間之音素點數，音素網路致使語音輸入以弧及節點表示，俾使語音輸入被封裝成一與字彙、語言模型、使用者及環境無關之中間格式；及

將上下文相關音素網路傳輸至一或多個遠離於本地使用者之應用程式，以致使遠端應用程式使用一由應用程式選擇之字彙或語言模型在每一應用程式執行語音辨認，俾使本地使用者位置不須執行語音辨認。

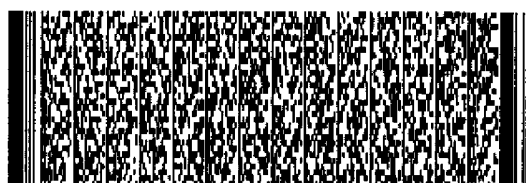
2、如申請專利範圍第1項之方法，其中上下文相關音素網路係與字彙和語言模型為無關。

## 3、如申請專利範圍第1項之方法，更包含：

在一第一應用上予以擷取需要來自使用第一應用之字彙和語言模型之上下文相關音素網路之資訊以便於操作第一應用。

## 4、一種編碼語音資訊之方法，包含：

提供包含兩部分之操作系統，其中操作系統之第一部分係與可連續適應於使用者之聲音之使用者和環境之特定聲音模型為相關；及操作系統之第二部分係相關於從操作系統之第一部分無關地選擇之第一應用，其中操作系統之第二部分係擷取從使用第一應用之字彙和語言模型所需要之



## 六、申請專利範圍

資訊。

5、如申請專利範圍第4項之方法，其中操作系統之第一部分係產生上下文相關音素網路，以便於捕捉使用者和環境之特定聲音模型並傳輸上下文相關音素網路至操作系統之第二部分。

6、如申請專利範圍第4項之方法，更包含第二應用，其中第一應用和第二應用係包含不同之字彙和不同之語言模型。

7、如申請專利範圍第4項之方法，更進而包含第二應用，其中第一應用和第二應用分享共同之字彙和共同之語言模型。

8、一種編碼語音資訊之方法，包含：

藉由使用者和環境之特定聲音模型和音素語言模型之輔助而轉換語音至上下文相關音素網路；

基於使用者和環境之特定聲音模型和音素語言模型而連續地接收上下文相關音素網路；

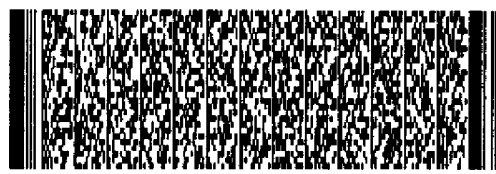
由語音訊號予以擷取做為語音辨認之所需要之資訊；及封裝在字彙、語言模型、及使用者和環境之獨立格式。

9、一種編碼語音資訊之裝置，係包含：

操作系統；

產生具有定義上下文相關音素網路音素網路之輸出之上下文相關音素網路音素網路之音素網路產生器；及

適於接收音素網路產生器之輸出之第一應用並從使用第一應用之字彙和語言模型之輸出予以擷取所需之資訊，其



## 六、申請專利範圍

中音素網路產生器和第一應用係與操作系統無關。

10、一種電腦可讀取之資料記錄媒體，包含指令和資料，被載入具有操作系統之第一通用目的微處理器，該微處理器具有一作業系統以使第一通用目的微處理器包含：在本地使用者位置產生具有定義上下文相關音素網路之輸出之上下文相關音素網路產生器，其中上下文相關音素網路致使語音輸入以節點及弧之形式表示，其中每一弧代表一具有以節點表示開始及結束時間之音素點數，俾使語音輸入被封裝成一中間格式；及適於接收音素網路產生器之輸出之複數個遠離於本地使用者位置之應用程式，並從使用複數個應用程式之字彙和語言模型之輸出以擷取所需之資訊，俾使不須自本地使用者位置擷取資訊，音素網路產生器和複數個應用程式與操作系統無關。

11、如申請專利範圍第10項之資料記錄媒體，其中資料記錄媒體係包含具有儲存在其上之音素網路產生器之第一部及具有存在其上之複數個應用之第二部。

12、一種方法，包含：

產生上下文相關音素網路以做為編碼語音資訊之中間步驟；

在第一搜尋引擎中嘗試辨認；

在第一搜尋引擎中辨認不滿意之下，使上下文相關音素網路前進至第二搜尋引擎；及

從第二搜尋引擎擷取候選者之辨認結果。

