公告本

I241555

發明專利說明書

（填寫本書件時請先行詳閱申請書後之申請須知，作※記號部分請勿填寫）

※ 申請案號：92100771 ※IPC分類：G10L 15/00

※ 申請日期：92.1.15

壹、發明名稱

(中文)連續聲音辨識裝置及連續聲音辨識方法及程式記錄媒體

(日文)連続音声認識装置および連続音声認識方法、並びに、プログラム記録媒体

貳、發明人(共 1 人)

發明人 1 (如發明人超過一人，請填**說明書發明人續頁**)

姓名：(中文)鶴田 彰

(英文)AKIRA TSURUTA

住居所地址：(中文)日本國奈良縣大和郡山市天井町 167-3

(英文)167-3, TENJO-CHO, YAMATOKOORIYAMA-SHI, NARA-KEN, JAPAN

國籍：(中文)日本 (英文)JAPAN

參、申請人(共 1 人)

申請人 1 (如申請人超過一人，請填**說明書申請人續頁**)

姓名或名稱：(中文)日商夏普股份有限公司

(英文)SHARP KABUSHIKI KAISHA

住居所或營業所地址：(中文)日本國大阪府大阪市阿倍野區長池町 22 番 22 號

(英文)22-22 NAGAIKE-CHO ABENO-KU OSAKA-SHI OSAKA 545-8522 JAPAN

國籍：(中文)日本 (英文)JAPAN

代表人：(中文)町田 勝彥

(英文)KATSUHIKO MACHIDA

捌、聲明事項

☐ **本案係符合專利法第二十條第一項☐第一款但書或☐第二款但書規定之期間，其日期為：** _____

☑ **本案已向下列國家(地區)申請專利，申請日期及案號資料如下：**

【格式請依：申請國家(地區)；申請日期；申請案號　順序註記】

1.日本；2002 年 01 月 16 日；特願 2002-007283

2._____

3._____

☑ **主張專利法第二十四條第一項優先權：**

【格式請依：受理國家(地區)；日期；案號　順序註記】

1.日本；2002 年 01 月 16 日；特願 2002-007283

2._____

3._____

4._____

5._____

6._____

7._____

8._____

9._____

10._____

☐ **主張專利法第二十五條之一第一項優先權：**

【格式請依：申請日；申請案號　順序註記】

1._____

2._____

3._____

☐ **主張專利法第二十六條微生物：**

☐國內微生物　【格式請依：寄存機構；日期；號碼　順序註記】

1._____

2._____

3._____

☐國外微生物　【格式請依：寄存國名；機構；日期；號碼　順序註記】

1._____

2._____

3._____

☐熟習該項技術者易於獲得，不須寄存。

玖、發明說明

(發明說明應敘明：發明所屬之技術領域、先前技術、內容、實施方式及圖式簡單說明)

發明所屬之技術領域

本發明係關於可利用音素環境依存型音響模型，高精確度地施行辨識之連續聲音辨識裝置及連續聲音辨識方法及記錄連續聲音辨識程式之程式記錄媒體。

先前技術

一般，作為在大詞彙連續聲音辨識中所使用之辨識單位，從辨識對象詞彙之變更及擴展成大詞彙較為容易之觀點而言，多半使用音節及音素等小於單字之所謂子字之辨識單位。另外，為了顧慮到諧音耦合等之影響，已知使用依存於前後之環境(上下文)之模型較為有效。例如，依存於前後各一個音素之所謂三音素模型之音素模型已廣為一般所利用。

又，作為辨識連續發出之聲音之連續聲音辨識方法之一，有依照以子字之網路及樹形結構等描述詞彙中之各單字之子字標記辭典、與描述單字之連接限制之文法或統計的語言模型之資訊，連結單字而獲得辨識結果之方法。

有關此等以子字作為辨識單位之連續聲音辨識技術，例如在古井貞熙監譯之刊物「聲音辨識之基礎(下)」中，有詳細之說明。

如上所述，使用依存於環境之子字施行連續聲音辨識時，已知使用音素環境依存型之音響模型，不僅在單字內，即使在單字與單字間，其辨識精確度也較為良好。但，由於使用於單字之始終端之音響模型需要依存於前後所連接

之單字，故與使用不依存於音素環境之音響模型之情形相比，處理會變得更為複雜，且處理量也會大幅增加。

以下，參照單字辭典、語言模型與音素環境依存型音響模型，具體地說明依各單字之連接歷程動態產生狀態系列樹之方法。

例如，對「asanotenki（日語「朝の天氣(早晨的天氣)」）・・・」之發音，考慮「朝(a；s；a)」這個單字之最後音素/a/時，有必要針對得自圖3所示之單字辭典之資訊之單字「朝日(a；s；a；h；i)」之第3個音素/a/與其前後連接之音素組成之三音素"s；a；h"、與得自圖4所示之語言模型之資訊之單字「の(n；o)」與接在其前面之單字「朝(a；s；a)」之連接詞「朝の(a；s；a；n；o)之第3個音素/a/與其前後連接之音素組成之三音素"s；a；n"，展開假設。在本例之情形，只要展開2個假設即可，但若使用更複雜之文法及統計的語言模型時，在單字之終端有可能連接許多不同的單字，而在該情形下，有必要依存於此等領先音素而例如利用圖2B所示之領先音素、中心音素及後續音素組成之三音素之狀態系列如圖5B所示一般，展開多種假設。

對於此問題，特開平5-224692號公報曾經揭示在單字內使用音素環境依存型之音響模型，另一方面，在單字字界中，使用不依存於環境之音響模型之連續聲音辨識方式。依據此連續聲音辨識方式，可抑制在單字間之處理量之增加。又，特開平11-45097號公報所揭示之方式為：在辨識對象詞彙之各單字中，使用描述不依存於前後單字而決定之音

響模型系列作為辨識單字之辨識單字辭典，在單字字界中使用依存於前後單字而描述之字間字辭典而予以對照之連續聲音辨識方式。依據此連續聲音辨識方式，即使在單字字界中使用音素環境依存型音響模型，也可抑制處理量之增加。

但，在上述以往之連續聲音辨識方式中，卻有以下之問題。即，在特開平5-224692號公報所揭示之連續聲音辨識方式中，在單字內使用音素環境依存型之音響模型，在單字字界中，使用不依存於環境之音響模型，因此，雖可抑制在單字間之處理量之增加，但在另一方面，由於使用於單字字界中之音響模型之精確度較低，尤其在施行大詞彙之連續聲音辨識時，有導致降低辨識性能之虞。

相對地，在特開平11-45097號公報所揭示之連續聲音辨識方式中，使用描述不依存於前後單字而決定之音響模型系列作為辨識單字之辨識單字辭典，在單字字界中使用依存於前後單字而描述之字間字辭典而予以對照之方式。因此，在單字字界中也可利用音素環境依存型音響模型，一面確保精確度，一面在大詞彙之情形時，也可抑制在單字字界中之處理量之增加。但一般而言，單字之得分及字界會受到前面之單字之影響，故多數辨識單字共有一個字間字時，如圖9A所示，並未考慮到辨識單字"k；o；k"及"s；o；k"之字間字"o"之字界之連接歷程，故如圖9B所示，與有考慮到單字之連接歷程之情形相比，有招致性能降低之虞。另外，例如如日語之助詞の"(發音為/o/)"等，並未揭示對無

法在辨識單字辭典與字間字辭典中分割之單字。

發明內容

因此，本發明之目的在於提供即使在單字字界也可利用音素環境依存型音響模型，一面確保精確度，一面在大詞彙之連續聲音辨識時，也可抑制在單字字界之處理量之增加之連續聲音辨識裝置及連續聲音辨識方法及記錄連續聲音辨識程式之程式記錄媒體。

為達成上述之目的，本發明之連續聲音辨識裝置之特徵在於：以依存於鄰接之子字而決定之子字作為辨識單位，並利用依存於子字環境之環境依存型音響模型，辨識連續發音之輸入聲音者；且包含：音響分析部，其係分析輸入聲音而獲得特徵參數之時間系列者；單字辭典，其係儲存詞彙中之各單字，以作為子字之網路或子字之樹形結構者；語言模型儲存部，其係儲存表示單字間之連接資訊之語言模型者；環境依存型音響模型儲存部，其係儲存上述環境依存型音響模型，以作為在該環境依存型音響模型之狀態系列中，彙集多數子字模型之狀態系列而樹形結構化所形成之子字狀態樹者；對照部，其係參照屬於上述環境依存型音響模型之子字狀態樹、上述單字辭典及語言模型，展開上述子字之假設，並施行上述特徵參數之時間系列與上述展開之假設之對照，以輸出包含有關符合單字終端之假設之單字、累積得分及始端開始幀之單字資訊，以作為單字點陣者；及探索部，其係施行對上述單字點陣之探索而產生辨識結果者。

依據上述構成，可參照將依存於子字環境之環境依存型音響模型樹形結構化之子字狀態樹、單字辭典及語言模型，展開子字之假設，因此，與後續之單字之領先子字無關地，只要展開1個假設即可，故可減少全部假設之狀態之總數。即，可大幅減少假設之展開處理量，與單字內及單字字界無關地，容易施行假設之展開。另外，利用對照部施行來自上述音響分析部之特徵參數系列與上述展開之假設之對照之際之對照處理量也可大幅減少。

又，在一實施例中，上述發明之連續聲音辨識裝置中，儲存於上述環境依存型音響模型儲存部之環境依存型音響模型係在中心子字依存於前後子字之環境依存型音響模型中，將領先子字及中心子字相同之子字模型之狀態系列樹形結構化之子字狀態樹。

依據此實施例，由於係利用將領先子字及中心子字相同之子字模型之狀態系列樹形結構化之子字狀態樹展開上述假設，因此，欲展開其次之假設時，只要僅注目於終端假設之中心子字而展開具有對應之領先子字之子字狀態樹即可。也就是說，即使有多數後續子字，也只要展開更少之假設即可，故假說之展開較為容易。

又，在一實施例中，上述發明之連續聲音辨識裝置中，上述環境依存型音響模型係狀態由多數子字模型所共有之狀態共有模型。

依據此實施例，由於狀態由多數子字模式所共有，故在樹形結構化之際，可將共有之狀態彙總成一個，並可減少

節點數。因此，可大幅減少上述對照部在對照時之處理量。

又，在一實施例中，上述發明之連續聲音辨識裝置中，上述對照部在參照上述子字狀態樹而展開假設之際，利用得自上述單字辭典及語言模型之可連接之子字資訊，在構成上述假設之子字狀態樹之狀態中，在可互相連接之狀態附加標記。

依據此實施例，由於在構成上述展開之假設之子字狀態樹之狀態中，將標記僅附在可互相連接之狀態，故在上述對照之際，可限定有必要施行簡易飛點計算之狀態，而將對照處理量進一步減少。

又，在一實施例中，上述發明之連續聲音辨識裝置中，上述對照部在施行上述對照之際，可依據上述特徵參數之時間系列，算出上述展開之假設之得分，並依據此得分之臨限值或包含假設數之基準，施行上述假設之剔除(剪掉無效的樹枝)。

依據此實施例，由於在施行上述對照之際，可施行假設之剔除，故可剔除成為單字之可能性較低之假設，大幅減少其後之對照處理量。

本發明之連續聲音辨識方法之特徵在於：以依存於鄰接之子字而決定之子字作為辨識單位，並利用依存於子字環境之環境依存型音響模型，辨識連續發音之輸入聲音者；且利用音響分析部，分析上述輸入聲音而獲得特徵參數之時間系列；利用對照部，參照將上述環境依存型音響模型之狀態系列樹形結構化而形成之子字狀態樹、描述詞彙中

之各單字，以作為子字之網路或子字之樹形結構之上述單字辭典、及表示單字間之連接資訊之語言模型，而展開上述子字之假設，並施行上述特徵參數之時間系列與上述展開之假設之對照，產生包含有關符合單字終端之假設之單字、累積得分及始端開始幀之單字資訊，以作為單字點陣；利用探索部，施行對上述單字點陣之探索而產生辨識結果者。

依據上述構成，與上述本發明之連續聲音辨識裝置之情形同樣，可參照將環境依存型音響模型樹形結構化之子字狀態樹，展開子字之假設，因此，與後續之單字之領先子字無關地，只要展開1個假設即可，並與單字內及單字字界無關地，容易施行假設之展開。另外，施行特徵參數系列與上述展開之假設之對照之際之對照處理量也可大幅減少。

又，本發明之程式記錄媒體之特徵在於記錄有本發明之連續聲音辨識程式，其係具有作為上述發明之連續聲音辨識裝置之音響分析部、單字辭典、語言模型儲存部、環境依存型音響模型儲存部、對照部及探索部之機能。

依據上述構成，與上述本發明之連續聲音辨識裝置之情形同樣，與後續之單字之領先子字無關地，只要展開1個假設即可，並與單字內及單字字界無關地，容易施行假設之展開。另外，施行特徵參數系列與上述展開之假設之對照之際之對照處理量也可大幅減少。

實施方式

以下，依據圖式之實施形態，詳細說明本發明。圖1係本

發明之實施形態之連續聲音辨識裝置之區塊圖。本連續聲音辨識裝置係由音響分析部1、向前對照部2、音素環境依存型音響模型儲存部3、單字辭典4、語言模型儲存部5、假設緩衝器6、單字點陣儲存部7及向後探索部8所構成。

在圖1中，輸入聲音被音響分析部1變換成特徵參數之系列而輸出至向前對照部2。在向前對照部2中，參照儲存於音素環境依存型音響模型儲存部3之音素環境依存型音響模型、儲存於語言模型儲存部5之語言模型及單字辭典4，而在假設緩衝器6上展開音素假設。而，使用上述音素環境依存型音響模型，利用幀同步簡易飛點光束搜尋施行上述展開之音素假設與特徵參數系列之對照，產生單字點陣而將其儲存於單字點陣儲存部7。

作為上述音素環境依存型音響模型，係使用考慮到前後各一個音素環境之所謂三音素模型之音素環境之隱式馬爾可夫模型(HMM)。即，上述子字模型為音素模型。但，在本實施形態中，係將以往如圖2B所示，以3狀態之狀態系列(狀態號列)表現考慮到中心音素前後各一個之領先音素與後續音素之三音素模型之情形改為如圖2A所示，將領先音素與中心音素相同之三音素模型之狀態系列彙集而成為樹形結構(以下稱音素狀態樹)。如圖2B所示，狀態由多數三音素模型所共有之狀態共有模型可利用將狀態系列樹形結構化之方式，削減狀態數，以減少計算量。

作為上述單字辭典4，係使用針對辨識對象詞彙之各單字，以音素系列標記該單字之讀音，並如圖3所示，將上述音

素系列樹形結構化之辭典。在語言模型儲存部5中，例如如圖4所示，儲存有利用文法所設定之單字間之連接資訊，以作為語言模型。又，在本實施形態中，雖係使用將表示單字之讀音之音素系列樹形結構化之辭典作為單字辭典4，但使用網路化之辭典也無妨。又，作為語言模型，雖係使用文法模型，但改用統計的語言模型也無妨。

在上述假設緩衝器6上，如上所述，利用上述向前對照部2，參照音素環境依存型音響模型儲存部3、單字辭典4、語言模型儲存部5，而依次展開如圖5A所示之音素假設。向後探索部8係一面參照儲存於語言模型儲存部5之語言模型及單字辭典4，一面例如利用A＊算法探索儲存於單字點陣儲存部7之單字點陣，以獲得對輸入聲音之辨識結果。

以下，依照圖6所示之向前對照處理動作流程圖，說明利用上述向前對照部2，參照音素環境依存型音響模型儲存部3、單字辭典4、語言模型儲存部5，而在假設緩衝器6上展開假設，以產生單字點陣之方法。

在步驟S1，首先在開始對照前，施行假設緩衝器6之初始化。而後，將由無音至接續在各單字之始端之"-；-；＊"之音素狀態樹設定於假設緩衝器6，以作為初始假設。在步驟S2，使用上述音素環境依存型音響模型，施行處理對象之幀中之特徵參數與假設緩衝器6內之圖7A所示之音素假設之對照，計算各音素假設之得分。在步驟S3，如圖7B所示，依據上述得分之臨限值或假設數等，如假設1及假設4所示，施行音素假設之剔除，藉此防止音素假設之不必要之

增加。在步驟S4，在殘留於假設緩衝器6內之音素假設中，針對單字終端有效之音素假設，將單字、累積得分及始端開始幀等之單字資訊保存於單字點陣儲存部7，藉此產生並保存單字點陣。在步驟S5，如圖7B之假設5及假設6所示，參照音素環境依存型音響模型儲存部3、單字辭典4及語言模型儲存部5之資訊，延伸殘留於假設緩衝器6內之音素假設。在步驟S6，判別該處理對象幀是否為最終幀。其結果，若為最終幀時，結束向前對照處理動作。另一方面，若非為最終幀時，返回步驟S2，轉移至次一幀之處理。其後，重複施行上述步驟S2至步驟S6之動作，在上述步驟S6，判別為最終幀時，結束向前對照處理動作。

以下，說明在上述向前對照處理動作之際，使用將領先音素及中心音素相同之三音素模型之狀態系列樹形結構化之音素狀態樹之情形之效果。

例如，對「asanotenki（日語「朝の天氣（早晨的天氣）」）・・・」之發音，考慮「朝(a；s；a)」這個單字之最後音素/a/時，可以針對得自圖3所示之單字辭典4之資訊之單字「朝日(a；s；a；h；i)」之第3個音素/a/與其前後連接之音素組成之三音素"s；a；h"、與得自圖4所示之語言模型之資訊之單字「の(n；o)」與接在其前面之單字「朝(a；s；a)」之連接詞「朝の(a；s；a；n；o)之第3個音素/a/與其前後連接之音素組成之三音素"s；a；n"，展開音素假設。此時，只要展開2種音素假設即可，但若使用更複雜之文法及統計的語言模型時，在單字之終端有可能連接許多不同的單字，如圖5B

所示，需要依照單字之領先音素，展開多種音素假設。對此，如本實施形態所示，在展開音素狀態樹之音素假設時，與次一單字之領先音素無關地，如圖5A所示，只要展開1個如圖2A所示之音素狀態樹"s；a；＊"即可。又，在圖5A中，係以仿照「樹」之三角形作為音素狀態樹符號。

而，如圖5B所示，就各個音素展開假設時，將後續之單字之領先音素之種類設定為全部27種時，新展開之音素假設數為27種，其新展開之音素假設之狀態總數為81(=27×3)種。

對此，如圖5A所示，使用上述音素狀態樹展開音素假設時，新展開之音素假設數為1種，其狀態總數為29(1+7+21)種，因此，可大幅削減假設之展開處理及對照處理之處理量。

又，將文法應用於上述語言模型時，後續之音素被單字辭典4及語言模型限定之情形相當多。因此，如圖8所示，音素狀態樹"s；a；＊"之各狀態中，僅在依據單字辭典4之音素行"s；a；h"及依據語言模型之音素行"s；a；n"所需要之狀態才附加標記(在圖8中，為橢圓形記號)，與音素狀態樹"s；a；＊"之全部狀態數29相比，可將對照之全部狀態數削減至狀態數5，因此，對照之處理量可進一步加以削減。

如以上所述，在本實施形態中，係將領先音素與中心音素相同之三音素模型之狀態系列彙集而樹形結構化之音素狀態樹儲存於音素環境依存型音響模型儲存部3中，其結果，在狀態由多數三音素模型所共有之狀態共有模型之情形

，可將樹形結構化之際共有之狀態歸納成一種，並可削減節點數。因此，就各個音素展開假設之情形，將上述音素狀態樹作為音素假設之用時，與後續之單字之領先音素無關地，只要展開1種音素假設即可，因此，假定後續之單字之領先音素之種類為全部27種時，以往為了新展開27種音素假設，其全部音素假設之狀態總數為81種。對此，在本實施形態中，新展開之音素假設只有1種，故可將全部音素假設之狀態總數削減至29種。

即，依據本實施形態，可大幅削減利用上述向前對照部2，參照儲存於音素環境依存型音響模型儲存部3之音素環境依存型音響模型、儲存於語言模型儲存部5之語言模型及單字辭典4，而展開音素假設之際之音素假設之展開處理量。因此，與單字內及單字字界無關地，容易施行假設之展開。且可大幅削減利用向前對照部2，使用上述音素環境依存型音響模型，以幀同步簡易飛點光束搜尋施行來自音響分析部1之特徵參數系列與上述展開之音素假設之對照之際之對照處理量。

其時，上述向前對照部2在施行與上述音素假設之對照之際，可計算各音素假設之得分，並依據得分之臨限值或假設數之臨限值，施行無效假設之剔除。因此，可剔除成為單字之可能性較低之音素假設，大幅減少對照處理量。另外，向前對照部2在展開上述音素假設之際，可參照語言模型儲存部5及單字辭典4，在構成上述音素假設之音素狀態樹之狀態中，將標記僅附在可互相連接而與上述對照有關

之狀態上，故在該情形下，在樹形結構化之狀態中，與上述對照無關之狀態無必要施行簡易飛點計算，因此，可使對照處理量進一步減少。

又，在上述之說明中，上述音素環境依存型音響模型係使用考慮到前後各一個音素環境之所謂三音素模型之HMM，但依存於鄰接之子字而決定之子字則不受此限定。

而，作為上述實施形態之音響分析部1、向前對照部2及向後探索部8之上述音響分析手段、對照手段及探索手段之機能可利用記錄於程式記錄媒體之連續聲音辨識程式加以實現。上述實施形態之上述程式記錄媒體既可使用與RAM(隨機存取記憶體)個別獨立設置之ROM(唯讀記憶體)所構成之程式媒體，也可使用裝定於外部輔助記憶裝置而被讀出之程式媒體。又，在其中任何一種情形，由上述程式媒體讀出連續聲音辨識程式之程式讀出手段均可具有可在上述程式媒體直接存取而讀出之構成，或可具有可下載至設於上述RAM之程式記憶區(未圖示)，而在上述程式記憶區存取而讀出之構成。又，由上述程式媒體下載至RAM之上述程式記憶區用之下載程式係事先儲存於本體裝置。

在此，所謂上述程式媒體係構成可與本體側分離之狀態，且包含磁帶或卡帶等磁帶系、軟碟、硬碟等磁碟或CD(音碟)-ROM、MO(光磁)碟、MD(小型光碟)、DVD(數位多用途光碟)等光碟之碟片系、IC(積體電路)卡、光卡等卡片系、光罩ROM、EPROM(紫外線消除型ROM)、EEPROM(電消除型ROM)、快閃ROM等半導體記憶體系而可固定地持有程

式之媒體。

又，上述實施形態之連續聲音辨識裝置具有數據機而構成可連接於包含網際網路之通訊網路時，上述程式媒體也可使用可利用由通訊網路下載等方式而流動性地持有程式之媒體。又，在該情形下，由上述通訊網路下載程式用之下載程式可事先儲存於本體裝置，或由別的記錄媒體安裝。

又，記錄於上述記錄媒體者並僅不限定於程式，資料也屬可記錄之對象。

圖式之簡單說明

圖1係本發明之連續聲音辨識裝置之區塊圖。

圖2A、圖2B係音素環境依存型音響模型之說明圖。

圖3係圖1之單字辭典之說明圖。

圖4係語言模型之說明圖。

圖5A、圖5B係利用圖1之向前對照部展開假設之說明圖。

圖6係利用向前對照部執行之向前對照處理動作之流程圖。

圖7A、圖7B係利用上述向前對照部執行假設對照及假設剔除之說明圖。

圖8係僅在音素假設之音素狀態樹中之有需要之狀態附加標記之情形之說明圖。

圖9A、圖9B係辨識單字與字間字之字界之連接歷程未被考慮到之情形與有被考慮到之情形之比較圖。

圖式代表符號說明

1　　音響分析部

2　　向前對照部

3 音素環境依存型音響模型儲存部

4 單字辭典

5 語言模型儲存部

6 假設緩衝器

7 單字點陣儲存部

8 向後探索部

肆、中文發明摘要

本發明係關於連續聲音辨識裝置及連續聲音辨識方法、連續聲音辨識程式及程式記錄媒體。在單字(即電腦字)之字界中，也可利用音素環境依存型音響模型，一面確保精確度，一面抑制在施行大詞彙之連續聲音辨識時之處理量之增大。彙集領先音素及中心音素相同之三音素模型而將領先音素之狀態、中心音素之狀態及後續音素之狀態之狀態系列樹形結構化而構成音素狀態樹，將此音素狀態樹儲存於音素環境依存型音響模型儲存部3中。因此，在利用向前對照部2，參照上述音素狀態樹、儲存於語言模型儲存部5之語言模型及單字辭典4而展開音素假設之際，與後續之單字之領先音素無關地，只要展開1個音素假設即可，故展開動作較為容易，不必顧慮到與單字內及單字字界之關係。且可大幅減少在施行與來自音響分析部1之特徵參數系列之對照之際之對照處理量。

伍、日文發明摘要

単語境界にも音素環境依存音響モデルを用いて精度を確保しつつ大語彙の連続音声認識時にも処理量の増大を抑える。音素環境依存音響モデル格納部（３）には、先行音素および中心音素が同じトライフォンモデルをまとめて先行音素の状態と中心音素の状態と後続音素の状態との状態系列を木構造化した音素状態木を格納している。したがって、前向き照合部（２）によって、上記音素状態木，言語モデル格納部（５）に格納された言語モデルおよび単語辞書（４）を参照して音素仮説を展開する際には、次に続く単語の先頭音素に関係無く１つの音素仮説を展開すればよく、単語内および単語境界に関係なく仮説の展開が容易になる。また、音響分析部（１）からの特徴パラメータ系列との照合を行う際における照合処理量を大幅に削減できる。

陸、(一)、本案指定代表圖為：第 1 圖

(二)、本代表圖之元件代表符號簡單說明：

1 音響分析部

2 向前對照部

3 音素環境依存型音響模型儲存部

4 單字辭典

5 語言模型儲存部

6 假設緩衝器

7 單字點陣儲存部

8 向後探索部

柒、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

拾、申請專利範圍

1. 一種連續聲音辨識裝置，其特徵在於：以依存於鄰接之子字而決定之子字作為辨識單位，並利用依存於子字環境之環境依存型音響模型，辨識連續發音之輸入聲音者；且包含：

音響分析部，其係分析上述輸入聲音而獲得特徵參數之時間系列者；

單字辭典，其係儲存詞彙中之各單字，以作為子字之網路或子字之樹形結構者；

語言模型儲存部，其係儲存表示單字間之連接資訊之語言模型者；

環境依存型音響模型儲存部，其係儲存上述環境依存型音響模型，以作為在該環境依存型音響模型之狀態系列中，彙集多數子字模型之狀態系列而樹形結構化所形成之子字狀態樹者；

對照部，其係參照屬於上述環境依存型音響模型之子字狀態樹、單字辭典及語言模型，而展開上述子字之假設，並施行上述特徵參數之時間系列與上述展開之假設之對照，以輸出包含有關符合單字終端之假設之單字、累積得分及始端開始幀之單字資訊，以作為單字點陣者；及

探索部，其係施行對上述單字點陣之探索而產生辨識結果者。

2.如請求項1之連續聲音辨識裝置，其中

儲存於上述環境依存型音響模型儲存部之環境依存型音響模型係在中心子字依存於前後子字之環境依存型音響模型中，將領先子字及中心子字相同之子字模型之狀態系列樹形結構化之子字狀態樹者。

3.如請求項2之連續聲音辨識裝置，其中

上述環境依存型音響模型係狀態由多數子字模型所共有之狀態共有模型者。

4.如請求項1之連續聲音辨識裝置，其中

上述對照部在參照上述子字狀態樹而展開假設之際，利用得自上述單字辭典及語言模型之可連接之子字資訊，在構成上述假設之子字狀態樹之狀態中，在可互相連接之狀態附加標記者。

5.如請求項1之連續聲音辨識裝置，其中

上述對照部在施行上述對照之際，可依據上述特徵參數之時間系列，算出上述展開之假設之得分，並依據此得分之臨限值或包含假設數之基準，施行上述假設之剔除者。

6.一種連續聲音辨識方法，其特徵在於：以依存於鄰接之子字而決定之子字作為辨識單位，並利用依存於子字環境之環境依存型音響模型，辨識連續發音之輸入聲音者；且

利用音響分析部，分析上述輸入聲音而獲得特徵參數之時間系列；

利用對照部，參照將上述環境依存型音響模型之狀態系列樹形結構化而形成之子字狀態樹、描述詞彙中之各單字，以作為子字之網路或子字之樹形結構之上述單字辭典、及表示單字間之連接資訊之語言模型，而展開上述子字之假設，並施行上述特徵參數之時間系列與上述展開之假設之對照，產生包含有關符合單字終端之假設之單字、累積得分及始端開始幀之單字資訊，以作為單字點陣；

利用探索部，施行對上述單字點陣之探索而產生辨識結果者。

7. 一種程式記錄媒體，其特徵在於記錄有連續聲音辨識程式者，該程式係使電腦具有作為如請求項1之音響分析部、單字辭典、語言模型儲存部、環境依存型音響模型儲存部、對照部及探索部之機能者。

拾壹、圖式

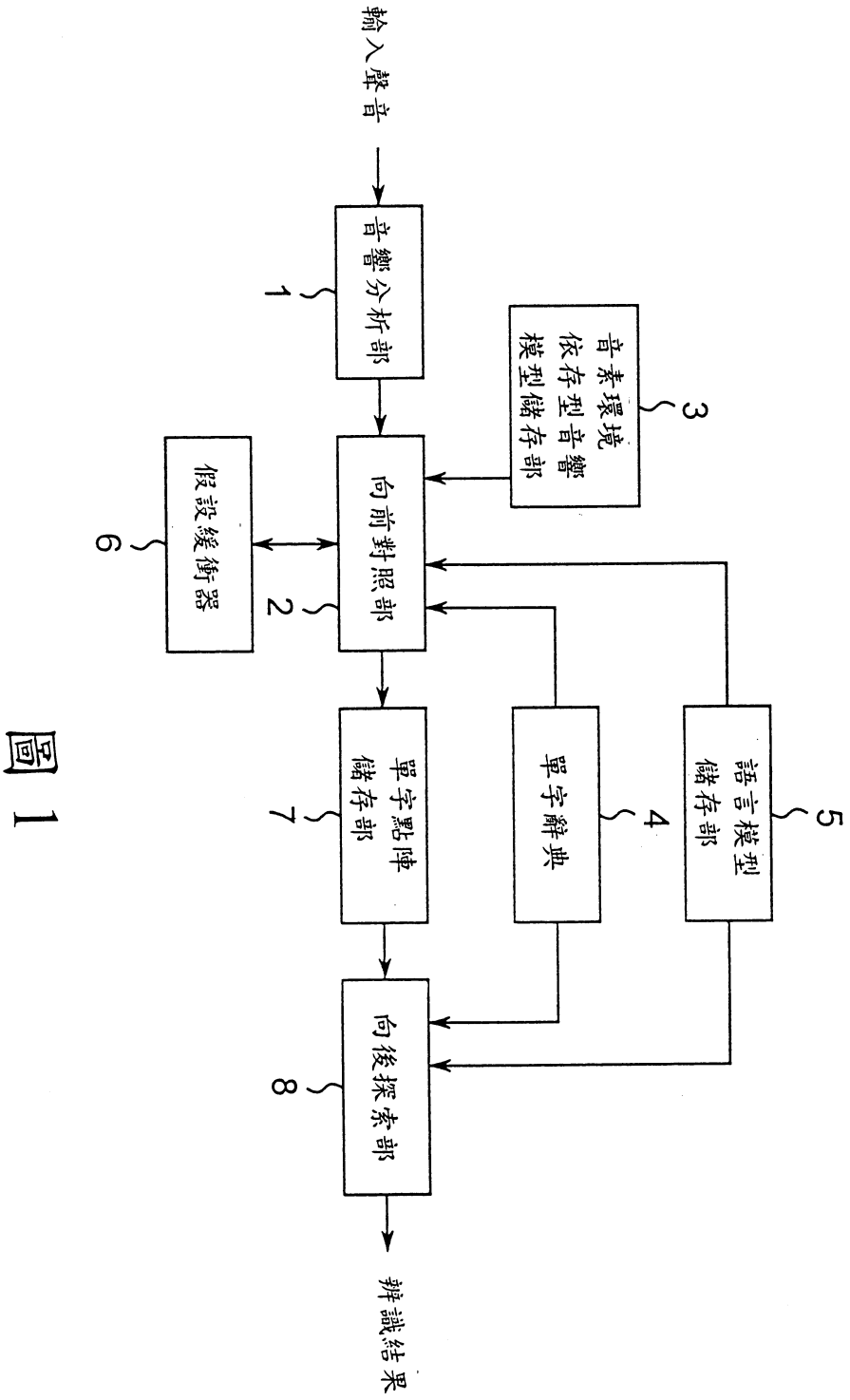

輸入聲音
音響分析部
1
音素環境依存型音響模型儲存部
3
向前對照部
2
假設緩衝器
6
語言模型儲存部
5
單字辭典
4
單字點陣儲存部
7
向後探索部
8
辨識結果

圖 1

圖 2A

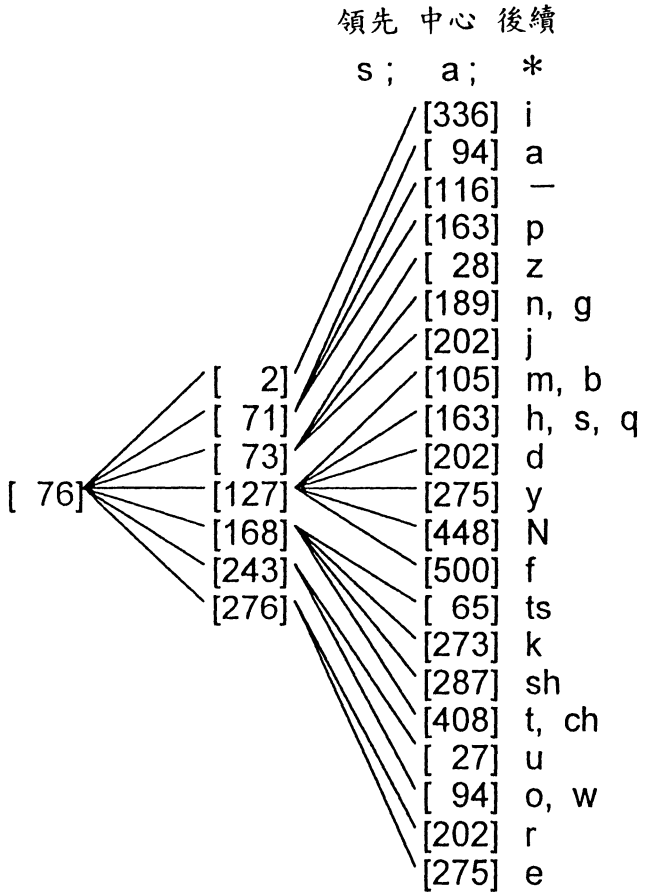
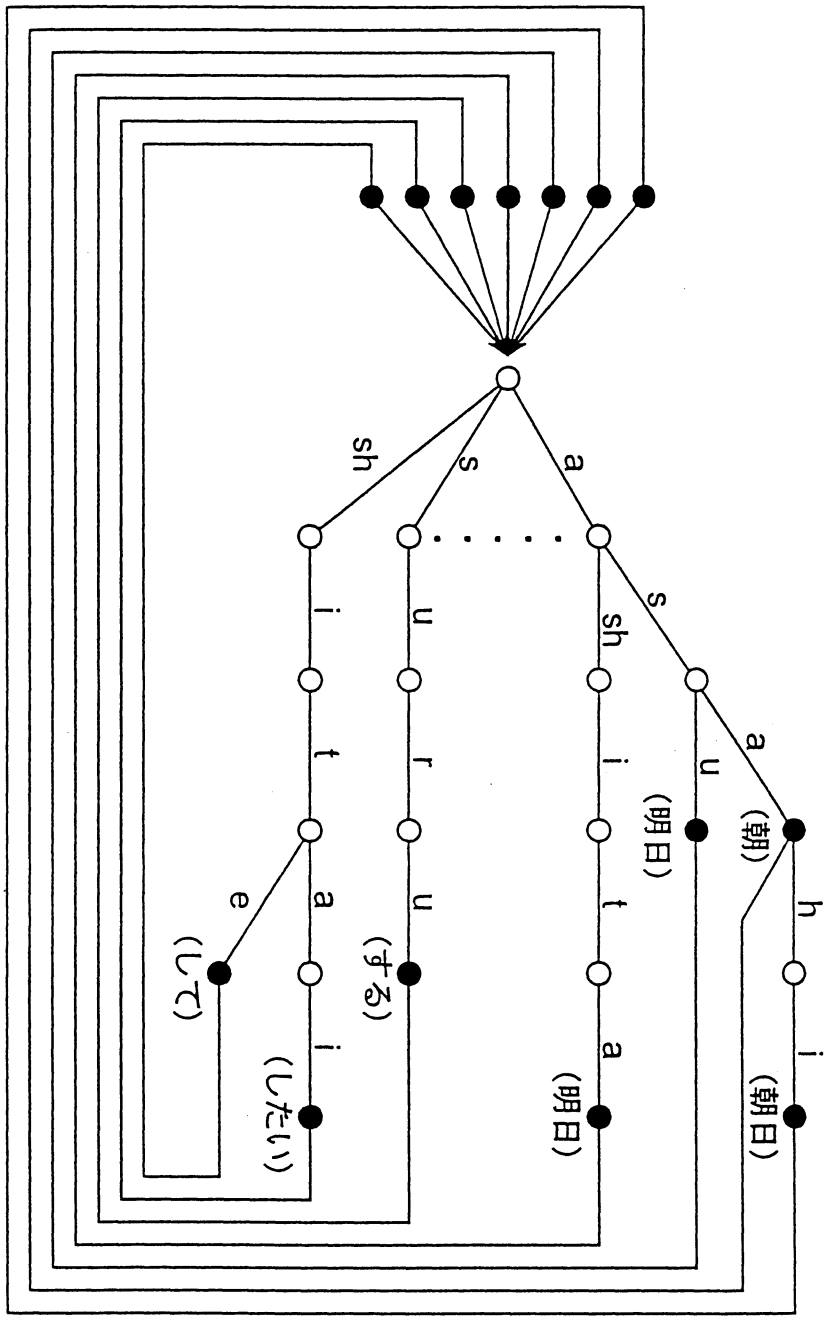


圖 2B

領先	中心	後續	狀態號列
s ;	a ;	—	[76]—[71]—[116]
s ;	a ;	a	[76]—[71]—[94]
s ;	a ;	i	[76]—[2]—[336]
s ;	a ;	u	[76]—[243]—[27]
s ;	a ;	e	[76]—[276]—[275]
s ;	a ;	o	[76]—[243]—[94]
s ;	a ;	k	[76]—[168]—[273]
s ;	a ;	s	[76]—[127]—[163]
s ;	a ;	sh	[76]—[168]—[287]
s ;	a ;	t	[76]—[168]—[408]
s ;	a ;	ch	[76]—[168]—[408]
s ;	a ;	ts	[76]—[168]—[65]
s ;	a ;	n	[76]—[73]—[189]
s ;	a ;	h	[76]—[127]—[163]
s ;	a ;	f	[76]—[127]—[500]
s ;	a ;	m	[76]—[127]—[105]
s ;	a ;	y	[76]—[127]—[275]
s ;	a ;	r	[76]—[276]—[202]
s ;	a ;	w	[76]—[243]—[94]
s ;	a ;	N	[76]—[127]—[448]
s ;	a ;	q	[76]—[127]—[163]
s ;	a ;	g	[76]—[73]—[189]
s ;	a ;	z	[76]—[73]—[28]
s ;	a ;	j	[76]—[73]—[202]
s ;	a ;	d	[76]—[127]—[202]
s ;	a ;	b	[76]—[127]—[105]
s ;	a ;	p	[76]—[71]—[163]

圖式續頁

圖 3



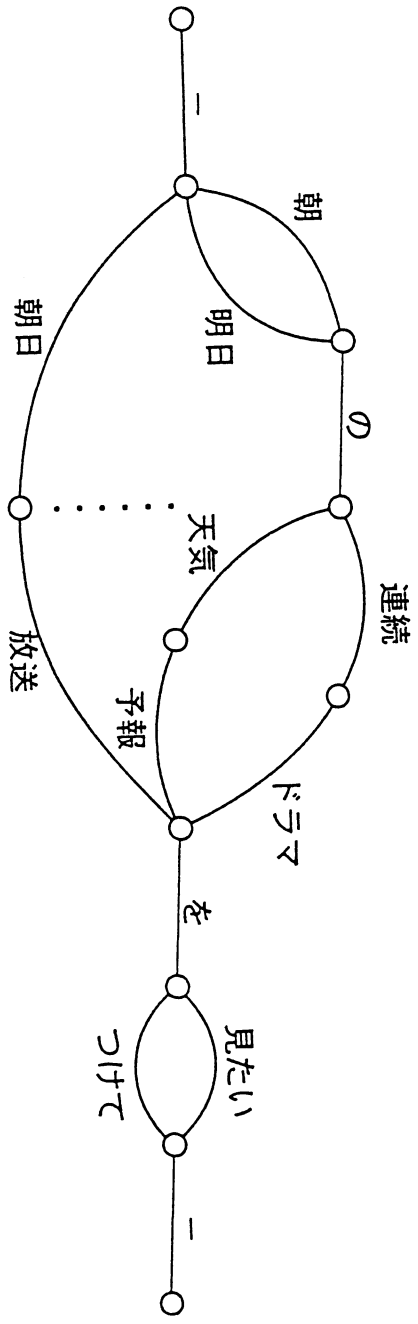


圖 4

圖式續頁

圖 5A

-;-;*
-;a;*
a;sh;*
a;s;*
s;a;*

圖 5B

-;-;a
-;a;sh
-;a;s
a;sh;i
a;s;u
a;s;a
a;s;h
s;a;;n
s;a;;p
由單字辭典選擇
「朝日(a;s;a;h;i)」
由語言模型選擇
「の(n;o)」
次一單字之領先音素

圖式續頁

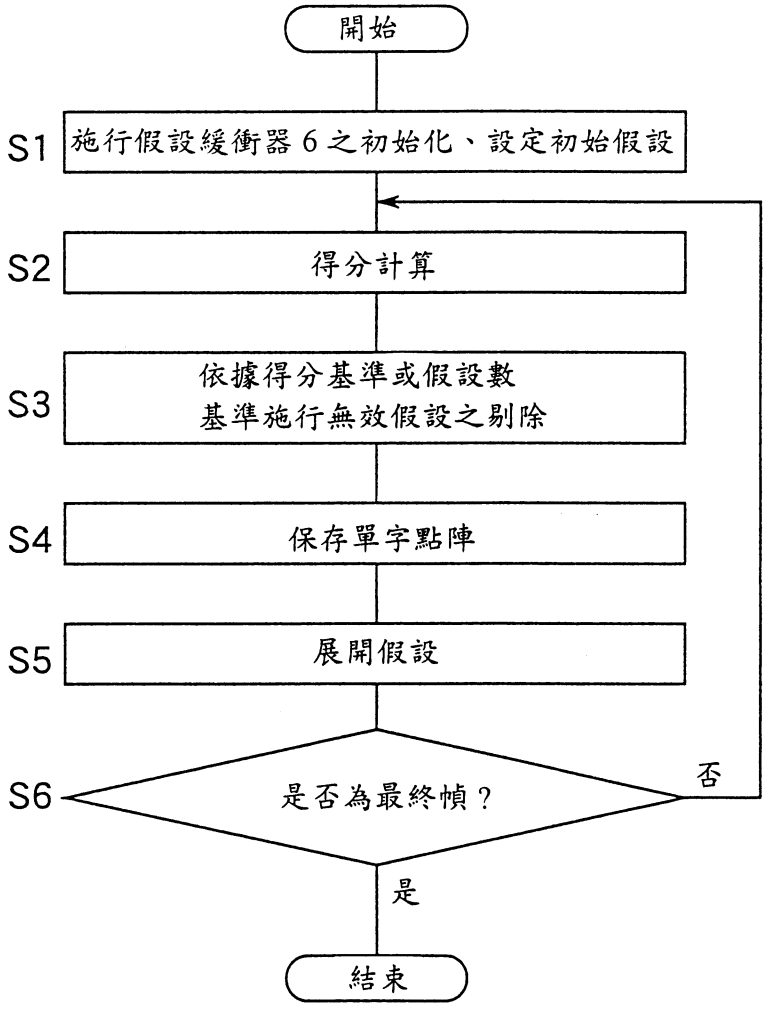

開始
S1
施行假設緩衝器 6 之初始化、設定初始假設
S2
得分計算
S3
依據得分基準或假設數
基準施行無效假設之剔除
S4
保存單字點陣
S5
展開假設
S6
是否為最終幀？
否
是
結束

圖 6

圖式續頁

圖 7A

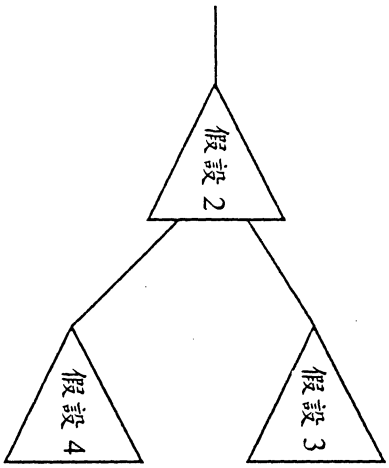
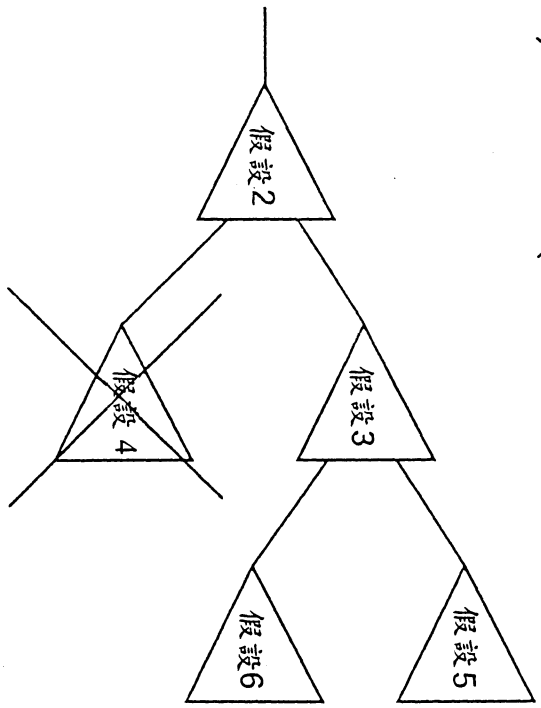


圖 7B



圖式續頁

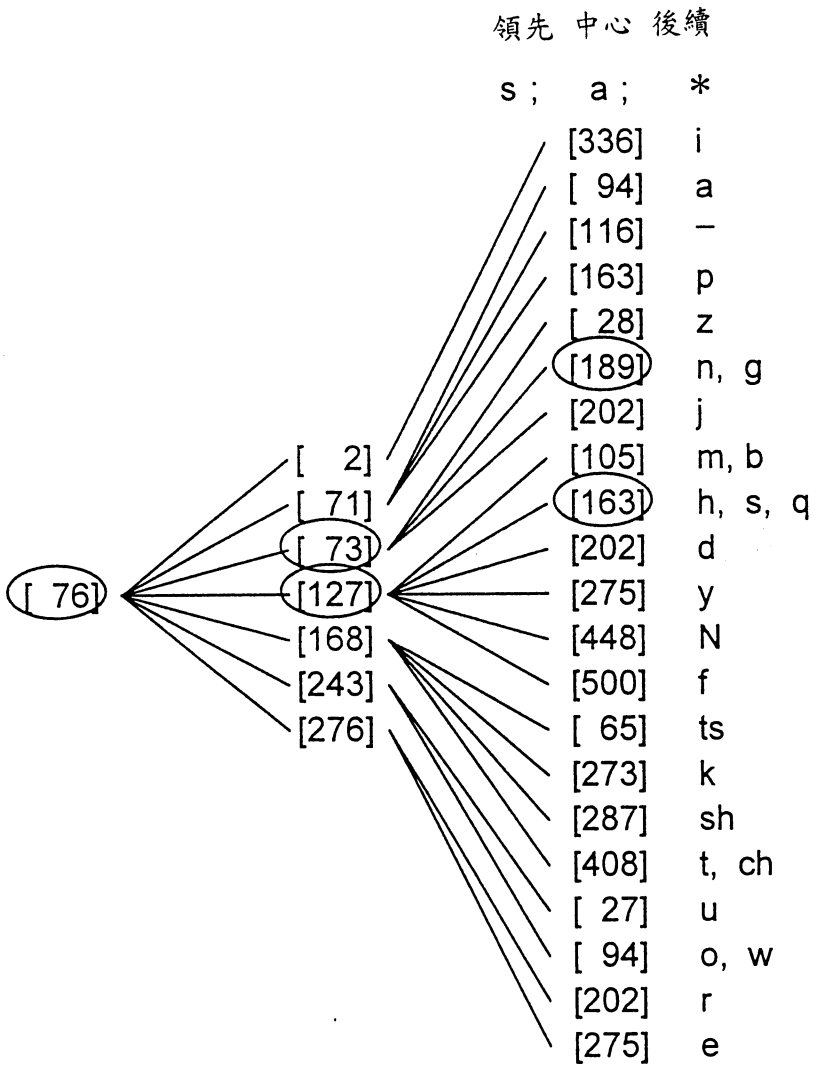

領先 中心 後續
s ; a ; *
[76]
[2]
[71]
[73]
[127]
[168]
[243]
[276]
[336] i
[94] a
[116] –
[163] p
[28] z
[189] n, g
[202] j
[105] m, b
[163] h, s, q
[202] d
[275] y
[448] N
[500] f
[65] ts
[273] k
[287] sh
[408] t, ch
[27] u
[94] o, w
[202] r
[275] e

圖 8

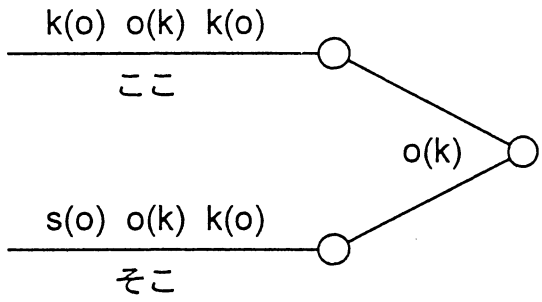


圖 9A

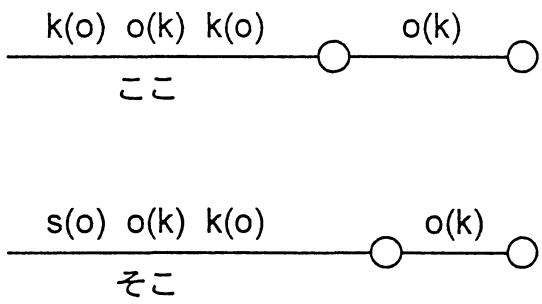


圖 9B