

公告本

申請日期	89 年 11 月 1 日
案 號	89123011
類 別	G10L 15/06

A4
C4

487903

(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	用於語音辨識系統之聲音模式之鑑別性訓練的方法與設備
	英 文	Method and apparatus for discriminative training of acoustic models of a speech recognition system
二、發明 創作人	姓 名	(1) 艾坦·巴納得 Barnard, Etienne (2) 珍-蓋·達罕 Dahan, Jean-Guy
	國 籍	(1) 南非 (2) 加拿大
住、居所		(1) 南非普利托里亞理查街四八五號 485 Richard Street, Pretoria 0083, South Africa
		(2) 美國麻州布魯克朗卡伯恩克里森十二號 12 Colbourne Crescent #3, Brookline, MA 02445-4522, U.S.A.
三、申請人	姓 名 (名稱)	(1) 語音作業國際股份有限公司 Speechworks International, Inc.
	國 籍	(1) 美國
	住、居所 (事務所)	(1) 美國麻州波士頓大西洋大道六九五號 695 Atlantic Avenue, Boston, Massachusetts 02111, U.S.A.
	代 表 人 姓 名	(1) 瑞克·歐林 Olin, Rick

裝

訂

線

(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

美國 1999年11月2日 09/433,609 有主張優先權

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

經濟部智慧財產局員工消費合作社印製

五、發明說明（ 1 ）

發明部份

本發明一般係有關自動語音辨識系統之改善性能，且更特別有關語音辨識系統之聲音模型之改善之鑑別訓練方法。

發明背景

許多自動語音辨識系統使用發音字典來辨認所接收之發話中所含之特定字。”發話”一辭在此處用以指由人或機器所產生之一或更多聲音。一發話之例包括，但不限於一單個聲音，任何二或更多聲音，一單個字，或二或更多字。一般言之，發音字典包含界定發話之預期發音之資料。每一發音包含一組音素，每一音素使用多個聲音模型界定，每一模型包含一音素所屬之各種聲頻及語音特徵值。

當收到一發話時，所接收之發話，或所接收之發話之至少一部份與發音字典中所含之預期發音比較。當所接收之發話或其一部份符合發音字典中所含之預期發音時，辨識出該發話。辨識包括決定該發話中所辨認之音素符合一特定字彙之對應音素之聲音模型。

聲音模型常根據實際接收之發話修改或”訓練”，以提高語音辨識器區別不同聲音單位之能力。雖每一聲音模型與一特定音素關聯，但以此聲音模型為基礎之字典可具有聲音相似或包含相似音素組之登記。此等字彙可能使語音辨識器難以區別。此等字間之混淆會在語音辨識器所用之應用軟體中引起誤差。

五、發明說明 (2)

會發生此混淆之一理由為聲音模型通常使用通用資訊訓練，而不參考語音辨識器或有關應用軟體所用之內容。結果，語音辨識器缺乏可用以鑑別音素或其他聲音單位之資訊，此可能與語音辨識器所用之特定工作有關。

例如，英文字 A U S T I N 及 B O S T O N 聲音相似，且語音辨識器可能難以區別。如語音辨識器使用於航空訂票系統，且 A U S T I N 及 B O S T O N 在字彙中，則 A U S T I N 及 B O S T O N 之混淆可能導致訂票錯誤或使用者困腦。

在另一例，考慮說數字 F I F T Y 及 F I F T E E N。如語音辨識器使用於証卷交易系統，則 F I F T Y 及 F I F T E E N 會導致錯誤訂單或使用者困腦。

使用通用模造之先行方法之例包括例如：

B.Juang等,"最少誤差分類之鑑別學習",IEEE信號處理會議記錄40:12(1992年12月),3043;Q,Huo等,"CDHMM 基礎之語音辨識之線上準貝士適應之研究",IEEE語音及聲音處理會議記錄,卷2,705(1966);

A.Sankar等,"聲音適應演算法之試驗研究",IEEE語音及聲音處理會議記錄,卷2,713(1996);

L.Bahl等,"大型字彙語音辨識系統之 Gaussian 混合模型之鑑別訓練",IEEE語音及聲音處理會議記錄,卷2,613(1996)。

此等參考文件中所概述之方法及其他先行方法具有明顯之缺點。例如，先行方法僅應用於框基礎之語音辨識系

(請先閱讀背面之注意事項再填寫本頁)

訂
線

五、發明說明 (3)

統之範圍中，此使用隱蔽之 Markov 模型。先行方法無一以分段基礎之語音辨識系統工作。此等方法之一基本假設為使用相同之聲音特色來匹配辨識器之字典中之每一片語。在分段基礎之系統中，分段方法產生一分段網路，在此，每一假設獨立選擇一最佳徑路通過該網路。結果，不同之假設依分段特色之不同序列記分，而非所有與框特色之公共序列相對記分。

在分段基礎之系統中，有時使用另外聲音特色，此等與上述之主特色不同，且自一假設至次一假設不同。需要一種方法，此能鑑別各副特色之聲音單位，此為先行方法所不能者。

而且，先行方法大體依賴手動介入，以達成訓練或調諧。先行方法並不自動根據由系統所經驗之實際發話，執行鑑別訓練。

根據上述，需要一種自動方法，用以根據有關語音辨識系統所用之特定應用軟體之資訊，訓練聲音模型。

特別需要一種訓練聲音模型之方法，其中，根據有關語音辨識器所用之特定應用軟體之資訊，訓練語音辨識器來鑑別聲音單位。

發明概要

以上需要，及由以下說明可明瞭之其他需要及目的由本發明達成，在一方面，此包含一種方法，用以自動訓練或修改語音辨識系統中之字之一或更多聲音模型。根據語

(請先閱讀背面之注意事項再填寫本頁)

訂

線

五、發明說明（4）

音辨識器所用之特定應用軟體有關之資訊，包括至少一正確對齊序列及至少一錯誤對齊序列之語音分段對齊資料，修改聲音模型。正確對齊序列正確代表發話者所發之一片語。錯誤對齊序列代表語音辨識系統辨識為不正確之一片語。逐段比較分段對資料，以辨認競爭之分段及引起辨識誤差之分段。

當辨認出一錯誤分段時，修改正確對齊序列中之音素之聲音模型，由移動其平均值更接近該分段之聲音特色執行。同時，修改錯誤對齊序列中之音素之聲音模型，由移動其平均值更遠離錯誤對齊序列之分段之聲音特色執行。結果，根據代表辨識誤差之經驗發話，聲音模型收斂至更佳之值。

附圖簡述

本發明以附圖中之實例圖解，而非限制，且在附圖中，相同之參考編號指相似之元件，且其中：

圖 1 顯示其中所用之一系統，以說明本發明之各層面及特色。

圖 2 A 為鑑別訓練之方法之流程圖。

圖 2 B 為圖 2 A 之方法之另外步驟之流程圖。

圖 3 為可由語音辨識器使用分段方法產生，且由圖 2 A，圖 2 B 之方法所接收之實例發話及實例分段對齊序列。

圖 4 顯示使用上述技術之示範聲音模型之移動。

五、發明說明 (5)

圖 5 為本實施例可使用之一電腦系統之方塊圖。

元件對照表

2	電話機
4	公共交換電話網路
1 0	語音應用模組
1 0 0	語音應用系統
1 0 2	應用軟體
1 0 4	自動語音辨識系統
1 0 8	自動語音辨識器
1 1 0	發音字典
1 1 2	聲音模型
1 1 4	分段對齊資料
1 1 6	量度資料
1 1 8	鏈路
3 0 2	發話波形
3 0 8	音素
3 1 0	對齊序列
3 1 2	音素記分值
5 0 0	電腦系統
5 0 2	匯流排
5 0 4	處理器
5 0 6	主記憶器
5 0 8	僅讀記憶器

(請先閱讀背面之注意事項再填寫本頁)

訂

線

五、發明說明 (6)

- 5 1 0 儲存裝置
- 5 1 2 顯示裝置
- 5 1 4 輸入裝置
- 5 1 6 遊標控制器
- 5 1 8 通訊介面
- 5 2 0 網路鏈路
- 5 2 2 本地網路
- 5 2 4 主電腦
- 5 2 6 網際網路服務提供器
- 5 2 8 網際網路

較佳實施例之詳細說明

說明一種方法及裝置，提供自動語音辨識系統中之聲音模型之改良之鑑別訓練。

在以下說明中，在說明上，提出許多特定細節，俾澈底明瞭本發明。然而，精於本藝之人士明瞭，可實施本發明而不用此等細節。在其他情形，熟悉之結構及裝置以方塊圖之形態顯示，以避免不必要模糊本發明。

系統綜覽

說明用以自動訓練或修改語音辨識系統中之字之一或更多聲音模型之方法。一般言之，根據有關語音辨識器所用之特定應用軟體或內容之資訊，修改聲音模型。應用軟體特定之資訊包含至少一正確對齊序列及至少一錯誤對齊

(請先閱讀背面之注意事項再填寫本頁)

訂

線

五、發明說明 (7)

序列之語音分段對齊資料。正確對齊序列代表一字彙，此正確表示發話者所發之話。錯誤對齊序列代表語音辨識系統根據發話者之發話所辨識之字彙，但此不正確表示發話者所發之話。

分段對齊資料逐段交互比較，以辨認引起辨識誤差之競爭分段。當辨認一錯誤分段時，修改正確對齊序列中之音素之聲音模型，移動其平均值更接近該分段之聲音特色。同時，修改錯誤對齊序列中之音素之聲音模型，移動其平均值更遠離錯誤對齊之分段之聲音特色。

結果，依據由特定語音辨識應用軟體之實際使用所獲得之經驗資訊，逐漸改正特定音素之聲音模型。故此，該應用軟體之語音辨識器之性能經由此自行改正機程隨時間而大為改善。

而且，並非使用通用聲音模型，該語音辨識器根據調諧至是時使用之特定應用軟體之聲音模型，實施辨識。

圖 1 顯示此處所用之語音應用系統 1 0 0，以說明本發明之各方面及特色。

在一實施例中，一發話人使用電話機 2，經由公共交換電話網路 (P S T N) 4 呼叫系統 1 0 0。該呼叫終止於並由應用軟體 1 0 2 回答，此與一自動語音辨識 (A S R) 系統 1 0 4 互動。或且，一發話者與使用適當電腦 (此執行應用軟體) 之應用軟體 1 0 2 直接互動。

應用軟體 1 0 2 為使用 A S R 1 0 4 之語音辨識服務器之任何元件。應用軟體 1 0 2 之例包含一語音發動系

(請先閱讀背面之注意事項再填寫本頁)

訂
線

五、發明說明 (8)

統或一電話機基礎之服務。以一或更多電腦程式或程序之形狀實施，用以製造航空飛行訂位，發送股票報價，提供公司資訊等。應用軟體 1 0 2 連接至 A S R 1 0 4，並使用鏈路 1 0 6 與其通訊。

在一實施例，應用軟體 1 0 2 在處理所接收之發話及決定邏輯步驟，以採取應用軟體之進一步功能之過程中，與一或更多語音應用模組 1 0 互動。語音應用模組 1 0 包含一或更多軟體元件，此執行預訂定之高階語音處理。例如，應用軟體 1 0 2 可呼叫語音應用模組 1 0，以產生及翻譯一是／否回答至發話者。適用作語音應用模組 1 0 之一商業產品為”對話模組”，由麻州波斯頓城之”語音工作國際公司”供應市面。

A S R 1 0 4 包含一自動語音辨識器（”辨識器”）1 0 8，一發音字典 1 1 0，聲音模型 1 1 2，分段對齊資料 1 1 4，及量度資料 1 1 6。辨識器 1 0 8 在通訊上分別由鏈路 1 1 8，1 2 0，1 2 2，1 2 4 連接至此等元件。鏈路 1 1 8，1 2 0，1 2 2，1 2 4 可使用任何機構實施，以提供其所連接之各別實體間之資料交換。鏈路 1 1 8，1 2 0，1 2 2，1 2 4 之例包含網路連接，線，光纖鏈路，及無線通訊鏈路等。

不滅失儲存器可用以儲存發音字典，聲音模型，及量度資料。不滅失儲存器例如可為一或更多碟片。在辨識器 1 0 8 之操作期間中，發音字典，聲音模型，及分段對齊資料儲存於不滅失記憶器中。

（請先閱讀背面之注意事項再填寫本頁）

訂

線

五、發明說明 (9)

辨識器 1 0 8 為一機構，其構造在使用發音字典 1 1 0 來辨識發話者之發話。使用鏈路 1 0 6 接收由應用軟體 1 0 2 所提供之資料形態之發話。辨識器 1 0 8 亦可能需要與 A S R 1 0 4 中未顯示或未說明於此處之其他組成件互動，以避免模糊本發明之各種特色及層面。辨識器 1 0 8 宜提供依賴發話者之連續語音辨識。適合用作辨識器 1 0 8 之一商業產品為”語音工作之核心辨識器” 5 . 0 ，由語音工作國際公司供應市面。

發音字典 1 1 0 包含資料，此界定可由 A S R 1 0 4 辨識之發話之預期之發音。發音字典 1 1 0 之一例更詳細說明於 1 9 9 9 年 6 月 2 4 日所提出之同待核定之申請書序號 0 9 / 3 4 4 1 6 4 號，題為”自動決定語音辨識系統中之發音字典之精確度”中，此具名 Etienne Barnard 為發明者，且其整個內容列作參考，如整個說明於此處。

辨識器 1 0 8 對其自應用軟體 1 0 2 所接收之每一發話製造及儲存分段對齊資料 1 1 4 。分段對齊資料 1 1 4 表示發話者之一完整發話內之各分立之聲音分段之界線。分段對齊資料用以決定何聲音分段相當於靜音或一或更多音素。對所接收之每一發話，辨識器 1 0 8 製造並儲存多個對齊序列，作為分段對齊資料 1 1 4 之一部份。每一對齊序列代表發話之一正確分段之假設序列。

聲音模型 1 1 2 儲存資料，此界定構成字之多個音素之每一個之多個聲音特徵。通常儲存多個預訂定之聲音模型 1 1 2 於不滅失儲存器中，在辨識器 1 0 8 之運轉時間

(請先閱讀背面之注意事項再填寫本頁)

訂
線

五、發明說明 (10)

上載入，並在決定如何辨識音素及字中用作參考。而且，辨識器 1 0 8 製造並儲存其在所接收之發話中所辨識之每一語音分段之聲音模型，與收到及處理發話同時。

依據一實施例，構造發音診斷工具 1 1 4，用以自動決定發音字典 1 1 0 之精確度，並辨認未滿足特定精確度標準之特別預期之發音。然後可更新未滿足特定精確度標準之預期之發音，俾更精確反映所接收之發話之實際發音。

A S R 1 0 4 可包含此處未顯示及說明之其他組件，以避免模糊本發明之各層面及特色。例如，A S R 1 0 4 可包含各種軟體發展工具及應用測試工具，可用以協助發展程序。此種工具之一為可自市面獲得之可再用之語音軟體模組工具包，稱為”對話模組”，由麻州波士頓城之語音工作國際公司供應。

鑑別訓練

現說明可用於以上系統中之一鑑別訓練機構。在較宜之實施例中，根據特定應用軟體 1 0 2 之特定資訊，修改及儲存（”訓練”）聲音模型 1 1 2。明確言之，由在辨識器 1 0 8 使用應用軟體 1 0 2 時普通混淆之字之例上之訓練，加強訓練。

在先行之方法，通常使用最大可能性技術訓練聲音模型。廣泛言之，最大可能性技術包括接收一特定音素，例如”a”（長音”a”）之發音之所有可獲得之例，並計

（請先閱讀背面之注意事項再填寫本頁）

裝

訂

線

五、發明說明 (11)

算該音素所屬之多個聲音參數之平均值及變異值。該系統然後對另一音素，例如“a h”獨立實施相似之處理。當辨識器108需鑑別“a”及“a h”時，此比較現假定之對齊序列之一音素之聲音模型及在過去已處理之發話之聲音模型。此法之缺點為每一音素獨立模造，而不顧及其聲音模型與一相似音素者之區別。

在一較宜之實施例，不獨立計算音素聲音模型，而是計算相似之聲音或相關之音素之聲音模型一起。例如，一較宜之方法考慮辨識器108所實際混淆之“a”音素之聲音模型及“a h”音素之聲音模型。亦考慮其反向。在反應中，該方法依據所經驗之混淆，修改聲音模型。

明確言之，使用強迫分段對齊序列技術，由應用軟體所接收之每一發話分割為多個音素，產生該發話之一“現對齊序列”。該方法並自辨識器接收代表音素之至少一正確辨識序列（“正確對齊序列”）之分段對齊資料，及代表音素之至少一不正確辨識之序列（“錯誤對齊序列”）之分段對齊資料。對正確對齊序列中之每一音素，該方法決定錯誤對齊序列中之最接近之對應音素（“競爭音素”）。對每對正確及錯誤之競爭音素，修改聲音模型內之平均值，轉移錯誤音素之平均值離開其記分所用之聲音特色向量。而且，正確音素之聲音模型之平均值向其記分所用之聲音特色轉移。在一實施例，僅移動最接近之正確平均值及最接近之錯誤平均值。

在使用應用軟體102之期間中，該方法對所接收之

（請先閱讀背面之注意事項再填寫本頁）

訂
線

五、發明說明 (12)

許多發話重複執行。結果，聲音模型逐漸收斂於代表改正值之值上，提供現應用軟體 1 0 2 之內容。部份根據錯誤發話之例之此訓練有利於協助提高區別及精確度。

該方法可在應用軟體 1 0 2 開閉且不處理有效呼叫時離線實施，或在線實施。例如，在應用軟體 1 0 2 執行之期間中，但在一特定呼叫終止後，可根據在該呼叫期間中所值得之辨識誤差，執行在線處理。此使系統可根據每一特定呼叫之內容資訊，諸如使用者確認回覆，修改其性能。

圖 2 A 為鑑別訓練方法之流程圖。圖 2 A 假定在執行圖 2 A 之第一方塊（方塊 2 0 2）之步驟之前，發話者已開始連接至語音辨識應用軟體，該應用軟體已執行，及與該應用軟體關聯之自動語音辨識系統已接收數位譯碼及儲存發話者之至少一發話。

在方塊 2 0 2，接收 n 最佳假設序列之一表。n 最佳假設序列各為多個音素，此等代表該發話中之一字之可能辨識。例如，辨識器 1 0 8 根據現聲音模型，產生音素之 n 最佳假設序列之一表。

在方塊 2 0 4，對 n 最佳表中之假設序列記分，並選擇最佳者。記分包括根據假設序列與所接收之數位化語音資料之符合良好程度，指定一值給每一假設序列。例如，記分可包括比較每一音素及聲音模型 1 1 2，並計算每一聲音模型之變異值。變異值為現音素之聲音模型之平均值與預訂定之聲音模型 1 1 2 之一之平均值之距離。一假設

（請先閱讀背面之注意事項再填寫本頁）

訂 線

五、發明說明 (13)

序列中所有音素之變異值之和為該假設序列之記分。亦可使用其他已知之記分技術。

在方塊 2 0 6，接收 n 最佳表中之最高記分假設序列之分段資料。每一假設序列之分段資訊普通由語音辨識器自動產生並儲存。分段包括分割代表一發話之數位化資料流為分段，其中，每一分段代表一分立之聲音。宜使用一強迫對齊方法執行分段。一實例分段方法之詳細發表於同待核定之申請書序號 [N U M B E R]，題為”語音辨識系統之分段方法”，具名之發明者為 MarkFanty 及 Michael Phillips，其整個內容列作參考，如全部說明於此處。

在方塊 2 0 8，接收正確對齊序列及以使用者確認或其他資料為基礎之錯誤對齊序列之分段資料。

明確言之，假定應用軟體 1 0 2 接收發話者所發出之字” A U S T I N ”。應用軟體 1 0 2 傳送代表該字之語音資料至辨識器 1 0 8。辨識器 1 0 8 錯誤地辨識該字為” B O S T O N ”，但亦回覆一低可信值給應用軟體

1 0 2，表示辨識器 1 0 8 並不確信” B O S T O N ”之辨識為正確。於回覆時，應用軟體 1 0 2 產生一提示給發話者，”你是否說” B O S T O N ？”。發話者回答”否”，此經正確辨識。於回覆時，應用軟體 1 0 2 提示發話者，”你是否說 A U S T I N ？”，及發話者回答”是”。故此，應用軟體 1 0 2 儲存一資訊，指示 A U S T I N 為正確之字，但需要使用者確認，以完成辨識。所儲存之資訊可包含一使用者確認旗標或等效之資料。在方塊

(請先閱讀背面之注意事項再填寫本頁)

訂

線

五、發明說明 (14)

2 0 8 ，該程序接收 B O S T O N 之錯誤假設序列及 A U S T I N 之正確假設序列二者之分段對齊資料。

在一些實施例，以上步驟可由語音應用模組 1 0 中所實施之功能協助。

圖 3 為一實例發話及實例分段對齊序列圖，此可由語音辨識器使用分段方法產生，並由圖 2 之方法在方塊 2 0 6 及方塊 2 0 8 處接收。

圖 3 包含一波形圖 3 0 0 ，此顯示一發話波形 3 0 2 ，一第一對齊序列 3 1 0 ，及一第二對齊序列 3 1 6 。提供此等元件之概要圖，以供說明。在實際上，發話波形及對齊序列以數位資料之形態實施，且儲存於電腦記憶器中。而且，可有任何數之對齊序列，且顯示二個僅為例解。

波形 3 0 2 為所發之話之圖形代表，其中，垂直軸線 3 0 4 代表幅度，及水平軸線 3 0 6 代表時間。

第一對齊序列 3 1 0 包含一水平線由分段界線 3 1 4 分割，此使用蔭蔽記號顯示。各分段落於分段界線之間，並與波形 3 0 2 之各部份關聯。每一分段與一音素記分值 3 1 2 及一音素 3 0 8 關聯。同樣，第二對齊序列 3 1 6 包含多個分段，各與第二音素 3 1 7 及第二音素記分值 3 1 8 關聯。

例如，第一對齊序列 3 1 0 代表 B O S T O N 之一 n 最佳假設序列，及第二對齊序列 3 1 6 代表 A U S T I N 之假設序列。故此，第一對齊序列 3 1 0 分割為代表 # h ， b ， a a ， s ， t 關閉， t ， i x ， n ， h # 之音素

(請先閱讀背面之注意事項再填寫本頁)

訂 · 線

五、發明說明 (15)

3 0 8 之分段。此等音素聯合代表 " B O S T O N " 之普通發音。音素 # h 為界線分段，此指示靜音，且為一片語之假設開始。音素 h # 同樣為一片語終止處之假設靜音。第二對齊序列包含 # h , a a , s , t 關閉 , t , i x , n , h # 之音素。在第二對齊序列中，分配於第一對齊序列中之音素 " b " 之一些能量分佈於音素 " a a " 中。而且，分配於第一對齊序列中之音素 " b " 之能量之前部假設為第二對齊序列中之靜音。

如以上有關圖 2 A 之方塊 2 0 8 所述，當收到此分段資料時，S R S 亦有其他資訊可用於鑑別上。例如，

S R S 可知道在發出由波形 3 0 2 所代表之話後，S R S 要求發話者確認該發話為 " B O S T I N " ，及發話者說 " 是 " 。在使用對話模組之系統中，自動產生及儲存此確認資訊。

或且，S R S 可知道該發話先已被辨識為 B O S T O N ，但此辨識錯誤，且正確之發話為 A U S T I N 。假定發生此，致知道第一對齊序列 3 1 0 為錯誤對齊序列，而知道第二對齊序列 3 6 1 為正確。另假定 S R S 指定一假設記分值 " - 2 0 0 0 " 給第一對齊序列 3 1 0 及一錯誤記分值 " 2 2 0 0 " 給第二對齊序列 3 1 6 。換言之，假定該假設記分值不正確指示何對齊序列為錯誤。

每一假設記分值為一特定對齊序列中之所有分段記分值 3 1 2 ， 3 1 8 之和。每一分段記分值代表一特定音素

(請先閱讀背面之注意事項再填寫本頁)

訂

線

五、發明說明 (16)

變異離開由聲音模型 1 1 2 所代表之音素之平均值。分段記分值可由比較一分段及所有預訂定音素值，選擇最接近預訂定值，並計算變異離開平均值來製造及儲存。

再參考圖 2 A，在方塊 2 1 0，該方法辨認在正確對齊序列及錯誤對齊序列中成對之競爭音素，以決定一誤差之最可能位置。例如，該方法分別檢查第一對齊序列 3 1 0 及第二對齊序列 3 1 6 之分段記分值 3 1 2，3 1 8。執行逐段比較。對每一分段，辨認所有之競爭分段。然後，該方法辨認在錯誤對齊序列中並不符合正確對齊序列之競爭分段之分段。

各分段及對應之競爭分段可以成對之旗標之形態表示及儲存。

例如，符號 (S_{1c} , S_{1w}) 指示正確對齊序列之分段 1 對應錯誤對齊序列之分段 1。再參考圖 3，正確對齊序列 3 1 6 之各分段與錯誤對齊序列 3 1 0 之各分段之垂直比較指示以下各點。第一，正確對齊序列 3 1 6 之分段 1 (" h # ") 重疊錯誤對齊序列 3 1 0 之分段 1 (" h # ")。第二，正確對齊序列 3 1 6 之分段 1 (" h # ") 亦重疊錯誤對齊序列 3 1 0 之分段 2 (" b ")。第三，正確對齊序列 3 1 6 之第二分段 1 (" a a ") 重疊錯誤對齊序列 3 1 0 之分段 2 (" b ")。第四，正確對齊序列 3 1 6 之第二分段 (" a a ") 重疊錯誤對齊序列 3 1 0 之分段 3 (" a a ")。以上資訊可表示如：(S_{1c} , S_{1w}) , (S_{1c} , S_{2w}) , (S_{2c} , S_{2w}) , (S_{2c} ,

(請先閱讀背面之注意事項再填寫本頁)

訂
線

五、發明說明 (17)

S_{3w})。

該方法然後比較每對之每一分段之分段記分值，並決定正確對齊序列之分段之分段記分值是否高於錯誤對齊序列之分段記分值。

例如，比較 (S_{ic} , S_{iw}) 之分段記分值。如 S_{ic} 之分段記分值高於 S_{iw} 者，則需要改正行動。在一實施例，僅當記分值之差低於一預定臨限時，方採取改正行動。一實例臨限值為 60 點。臨限值之使用辨識大於臨限值之記分值之差表示充分區別，無需改善，而小於臨限之差表示區別不充分，需要改善，否則可能發生辨識誤差。

在另一實施例，如正確競爭音素較之錯誤競爭音素記分低 500 點以上，則該辨識視為不良，不作改善，亦不採取改正行動。此反應辨識在此情形中，所選之 n 最佳假設可能無效。故此，不想根據不良資料來改善，而更宜僅當系統確信改正行動為適當時，方施行此改正行動。

當辨認出在臨限值內之一對分段時，該方法檢查分段有關之聲音模型，並嘗試決定聲音模型之何部份出錯。修改聲音模型有關之參數值，俾使聲音模型產生更接近正確值及更離遠錯誤值之分段記分。視方法是改正錯誤或正確者而定，朝向或離遠量度值移動。

如顯示於圖 2 B 之方塊 2 1 2，該方法修改錯誤競爭音素之聲音模型之平均值，移動此等值離開其記分所依據之特色向量。同樣，在方塊 2 1 4，該方法修改正確競爭音素之聲音模型之平均值，移動其更接近其記分所依據之

(請先閱讀背面之注意事項再填寫本頁)

訂 · 線

五、發明說明 (18)

特色向量。

在一實施例，修改行動包含由記分所依據之特色向量減一音素之辨認所產生之模型之平均值。該移動量宜約為此差之 2 %。其他實施例可移動模型更大或更小之量。移動量可作為預訂定之常數值實施。移動量亦稱為”學習率”。宜使用較小之值與較多之發話，例如以百計之發話聯合。使用此法，在鑑別中之累積效果逐漸但高度精確改正。該方法導致在大但非頻常之錯誤上順利進行。

圖 4 顯示一示範聲音模型之移動，使用以上技術。

例如，圖 4 顯示有關二分段 (S_{ic} , S_{iw}) 之結構。聲音模型 402 代表”#h”，並具有由點 402 所代表之平均記分。正確分段記分值 S_{ic} 由點 406 代表，及錯誤記分值 S_{iw} 由點 404 代表。為提高 #h 音素之鑑別，聲音模型 402 移動更接近點 406，如由向量 408 所示，且離點 404 更遠，如由向量 410 所示。結果，聲音模型 402 之淨位移由向量 412 表示。

結果，下次遇到同一音素，正確分段 S_{ic} 之記分值提高，因為點 402 至點 406 之距離縮短，且故此，變異離開平均值之較小。

反之，錯誤分段 S_{iw} 之記分值減小，因為點 402 至點 404 之距離加大，且故此，變異離開平均值加大。此改善可導致由聲音模型之改善鑑別訓練產生正確辨識。

圖 4 以圖解之形態顯示以上技術，然而，在實際上，每一聲音模型包含宜為 3 2 Gaussian 成份依以下關係混合：

(請先閱讀背面之注意事項再填寫本頁)

訂
線

五、發明說明 (19)

$$N(x, m, \theta) = \sum_{n=1}^N w_n \sum_{i=1}^n e^{-\frac{(x-m_i)^2}{\sigma_i^2}} \frac{1}{\sqrt{2\pi}\sigma_i}$$

其中，每一 Gaussian 成份依以下關係訂定：

$$\frac{1}{\sqrt{2\pi}\sigma} (e^{-\frac{(x-m)^2}{\sigma^2}})$$

此關係在觀念上可為一準鈴形曲線，其中，曲線之高度為平均值，及曲線之中心部份之寬度為變異。該 3 2 Gaussian 成份置於多維空間中，與許多不同之量度值關聯。8 成份模造多種內容中之一特定音素，包含雜音及其他因素。在一實施例，以上有關圖 4 所述之修改對所有量度值執行。分段記分值以所有量度值為基礎。

再參考圖 2 B，在方塊 2 1 6 中，該方法對正確對齊序列及錯誤對齊序列之每對競爭音素重覆執行方塊 2 1 2，2 1 4，直至所有各對之競爭音素考慮完為止。在圖 3 之例中，有四對競爭音素，如以上所辨認。其餘音素（" s "，t 關閉，" t "，" i x "，" n "，# h）在選擇作為正確音素上相符合，且不競爭。故此，其餘音素之音聲模型不修改。

在方塊 2 1 8，對由應用軟體所接收之許多發話重覆執行整個方法。此確保根據充足之資料量改正。

自以上說明可明瞭，此處所發表之方法較之先行方法有許多優點。此包括至少以下：

（請先閱讀背面之注意事項再填寫本頁）

訂 · 線

五、發明說明 (20)

1 . 根據發話者及語音辨識系統間之對話中之實際發生執行改正。如此，根據有關在實際使用自動語音辨識器中所發生之誤差之資訊或假設，而非根據人為介入執行改正。此改正自動執行，且可在語音辨識應用軟體之在線處理期間中由語音辨識器執行。

2 . 鑑別訓練之方法首次施行於分段基礎之語音辨識系統上。先行方法僅操作於框基礎之系統上，以隱蔽之 Markov 模型為基礎。先行方法不涉及辨認競爭分段，或根據分段之辨認處理誤差。

硬體綜覽

圖 5 為方塊圖，此顯示一電腦系統，其上可實施本發明之實施例。電腦系統 5 0 0 包含一匯流排 5 0 2 或其他通訊機構，用以交通資訊，及一處理器 5 0 5 連接於匯流排 5 0 2，用以處理資訊。電腦系統 5 0 0 亦包含一主記憶器 5 0 6，諸如隨機進出記憶器 (R A M) 或其他動態儲存裝置連接於匯流排，用以儲存資訊及欲由處理器 5 0 4 執行之指令。在執行欲由處理器 5 0 4 所執行之指令之期間中，主記憶器 5 0 6 亦可用以儲存暫時變數或其他中間資訊。電腦系統 5 0 0 另包含一僅讀記憶器 (R O M) 5 0 8 或其他靜態儲存裝置，連接至匯流排 5 0 2，用以儲存處理器 5 0 4 用之靜態資訊及指令。設置一儲存裝置 5 1 0，諸如磁碟或光碟，且連接至匯流排 5 0 2，用以儲存資訊及指令。

(請先閱讀背面之注意事項再填寫本頁)

訂
線

五、發明說明 (21)

電腦系統 5 0 0 可經匯流排 5 0 2 連接至顯示器

5 1 2，諸如陰極射線管 (C R T)，用以顯示資訊給電腦之使用者。一輸入裝置 5 1 4 連接至匯流排 5 0 2，包含字母數字及其他鍵，用以連通資訊及命令選擇至處理器 5 0 4。另一種使用者輸入裝置為遊標控制器 5 1 6，諸如滑鼠，軌跡球，或遊標方向鍵，用以連通方向資訊及命令選擇至處理器 5 0 4，並用以控制顯示器 5 1 2 上之遊標之移動。此輸入裝置普通具有在二軸線，一第一軸線 (例如 X) 及一第二軸線 (例如 Y) 上之二自由度，使裝置可指定在一平面上之位置。

本發明係有關使用電腦系統 5 0 0 來實施鑑別訓練之方法。依據本發明之一實施例，由電腦系統 5 0 0 提供鑑別訓練之方法，反應處理器 5 0 4，執行主記憶器 5 0 6 中所儲存之一或更多指令之一或更多序列。此等指令可由另一電腦可讀之媒體，諸如儲存裝置 5 1 0 讀出至主記憶器 5 0 6。主記憶器 5 0 6 中所儲存之指令序列之執行導致處理器 5 0 4 執行此處所述之方法步驟。在另一實施例，可使用硬線電路取代或結合軟體指令，以實施本發明。故此，本發明之實施例不限於硬體電路及軟體之任何特定組合。

如此處所用，" 電腦可讀之媒體 " 一辭指參與提供指令給處理器 5 0 4 以供執行之任何媒體。此一媒體可採取許多形狀，包括，但不限於不滅失媒體，可滅失媒體，及傳輸媒體。不滅失媒體例如包含光或磁碟，諸如儲存裝置

(請先閱讀背面之注意事項再填寫本頁)

訂
線

五、發明說明 (22)

5 1 0。可滅失媒體包含動態記憶器，諸如主記憶器

5 0 6。傳輸媒體包含同軸電纜，銅線，及光纖，包含構成匯流排 5 0 2 之線。傳輸媒體亦可採取聲波或光波之形狀，諸如在無線電波及紅外線資料通訊期間所產生者。

電腦可讀之媒體之普通形狀包括例如軟碟，可撓性碟，硬碟，磁帶，或任何其他磁性媒體，C D - R O M，任何其他光學媒體，打孔卡，紙帶，具有孔圖案之任何其他物理媒體，R A M，P R O M，及 E P R O M，快閃 E P R O M，任何其他記憶片或載體，諸如以後所述之載波，或電腦可讀之任何其他媒體。

各種形狀之電腦可讀之媒體可包括攜帶一或更多指令之一或更多序列至處理器 5 0 4，以供執行。例如，指令可先攜帶於一遠地電腦之磁碟上。遠地電腦可載入指令於其動態記憶器中，並使用數據機發送該指令於電話線上。電腦系統 5 0 0 之本地數據機可接收電話線上之資料，並使用紅外線發射機變換該資料為紅外線信號。一紅外線偵測器可接收外線信號所攜帶之資料，及適當之電路可置該資料於匯流排 5 0 2 上。匯流排 5 0 2 攜帶該資料至主記憶器 5 0 6，處理器可自此取出並執行指令。由主記憶器 5 0 6 所接收之指令在由處理器 5 0 4 執行之前或後，可選擇儲存於儲存裝置 5 1 0 中。

電腦系統 5 0 0 亦包含一通訊介面 5 1 8 連接至匯流排 5 0 2。通訊介面 5 1 8 提供雙向資料通訊連接至一網路鏈路 5 2 0，此連接至本地網路 5 2 2。例如，通訊介

(請先閱讀背面之注意事項再填寫本頁)

訂

線

五、發明說明 (23)

面 5 1 8 可為整合服務數位網路 (I S D N) 卡或一數據機，以提供資料通訊連接至對應型式之電話線。在另一例，通訊介面 5 1 8 可為本地區域網路 (L A N) 卡，以提供資料通訊連接至可相容之 L A N。亦可實施無線鏈路。在任何此實施中，通訊介面 5 1 8 發送及接收電，電磁，或光信號，此攜帶代表各種資訊之數位資料流。

網路鏈路 5 2 0 普通提供經由一或更多網路交通資料至其他資料裝置。例如，網路鏈路 5 2 0 可提供經本地網路 5 2 2 連接至主電腦 5 2 4，或至由網際網路服務提供者 (I S P) 5 2 6 所操作之資料裝備。I S P 5 2 6 轉而提供資料通訊服務經現通稱為”網際網路” 5 2 8 之世界訊包資料通訊網路。本地網路 5 2 2 及網際網路 5 2 8 二者使用電，電磁，或光信號，此攜帶數位資料流。通過各種網路之信號及在網路鏈路 5 2 0 上及通過通訊介面 5 1 8 (此等攜帶數位資料往來於電腦系統 5 0 0) 之信號為輸送資訊之載波之一示範形態。

電腦系統 5 0 0 可經由網路，網路鏈路 5 2 0，及通訊介面 5 1 8 發送訊息及接收資料，包括程式碼。在網際網路之例，服務器 5 3 0 可經由網際網路 5 2 8，I S P 5 2 6，本地網路 5 2 2，及通訊介面 5 1 8 發送應用軟體程式所要求之程式。依據本發明，此一下載之應用軟體提供此處所述鑑別訓練之方法。

所接收之程式在接收時可由處理器 5 0 4 執行，及／或儲存於儲存裝置 5 1 0 中，或其他不滅失儲存器中，供

(請先閱讀背面之注意事項再填寫本頁)

訂
線

五、發明說明 (24)

其後執行。如此，電腦系統 5 0 0 可獲得載波形態之應用軟體程式。

在以上說明中，已參考特定之實施例說明本發明。然而，顯然，其中可作各種修改及更改，而不脫離本發明之廣泛精神及範圍。故此，說明及附圖視為例解，而非限制意義。

(請先閱讀背面之注意事項再填寫本頁)

訂

線

四、中文發明摘要 (發明之名稱： 用於語音辨識系統之聲音模式之鑑別性訓練的方法與設備)

提供一種方法及裝置，用以自動訓練或修改語音辨識系統中之聲音單位之一或更多模型。根據有關語音辨識器所用之特定應用軟體之資訊，包含至少一正確對齊序列及至少一錯誤對齊序列之語音分段對齊資料，修改聲音模型。正確對齊序列正確代表發話者所發之一片語。錯誤對齊序列代表語音辨識系統認不正確之一片語。逐段比較分段對齊資料，以辨認競爭分段及引起辨認誤差者。當認出一錯誤分段時，修改正確對齊序列中之音素之聲音模型，由移動其平均值更接近該分段之聲音特色執行。同時，修改錯誤對齊序列中之音素之聲音模型，由移動其平均值更遠離錯誤對齊序列之分段之聲音特色執行。結果，根據代表辨識誤差之經驗發話，聲音模型收斂至更適宜之值。

(請先閱讀背面之注意事項再填寫本頁各欄)

裝

英文發明摘要 (發明之名稱：

METHOD AND APPARATUS FOR DISCRIMINATIVE TRAINING OF ACOUSTIC MODELS OF A SPEECH RECOGNITION SYSTEM)

A method and apparatus are provided for automatically training or modifying one or more models of acoustic units in a speech recognition system. Acoustic models are modified based on information about a particular application with which the speech recognizer is used, including speech segment alignment data for at least one correct alignment and at least one wrong alignment. The correct alignment correctly represents a phrase that the speaker uttered. The wrong alignment represents a phrase that the speech recognition system recognized that is incorrect. The segment alignment data is compared by segment to identify competing segments and those that induced the recognition error. When an erroneous segment is identified, acoustic models of the phoneme in the correct alignment are modified by moving their mean values closer to the segment's acoustic features. Concurrently, acoustic models of the phoneme in the wrong alignment are modified by moving their mean values further from the acoustic features of the segment of the wrong alignment. As a result, the acoustic models will converge to more optimal values based on empirical utterance data representing recognition errors.

訂

線

六、申請專利範圍

1. 一種分段基礎之自動語音辨識系統之訓練聲音模型之方法，包括步驟：

接收正確對齊資料，此代表由語音辨識系統所接收之一發話之正確分段對齊序列；

接收錯誤對齊資料，此代表根據自語音辨識系統所接收並說明該發話之資訊知道為不正確之發話之對齊序列；

辨認與正確對齊資料中之一第二音素相對應之錯誤對齊資料中之一第一音素；

修改第一音素之聲音模型，由移動其至少一平均值更遠離用以對第一音素記分之特色值執行。

2. 如申請專利範圍第 1 項所述之方法，另包括步驟：

接收正確對齊資料，此代表根據自語音辨識系統所接收並說明該發話之資訊知道為正確之發話之一對齊序列；

辨認與錯誤對齊資料中之第一音素相對應之正確對齊資料中之一第二音素；

修改第二音素之第二聲音模型，由移動其至少一平均值更接近用以對第二音素記分之特色值執行。

3. 如申請專利範圍第 1 項所述之方法，其中，接收正確對齊資料包括步驟：接收正確對齊資料，此代表由語音辨識系統所接收之一發話之 n 最佳假設中所選出之一最高記分假設之對齊序列之一分段對齊序列。

4. 如申請專利範圍第 1 項所述之方法，其中，接收錯誤對齊資料包括步驟：接收錯誤對齊資料，此代表根據

六、申請專利範圍

自語音辨識系統所接收之反應提示發話者確認該發話之使用者確認資訊知道為不正確之發話之一對齊序列。

5 . 如申請專利範圍第 1 項所述之方法，其中，接收正確對齊資料包括步驟：接收正確對齊資料，此代表根據自語音辨識系統所接收之反應提示發話者確認該發話之使用者確認資訊知道為正確之發話之一對齊序列。

6 . 如申請專利範圍第 1 項所述之方法，另包括步驟：對與正確對齊資料中之一或更多音素相對應之錯誤對齊資料中之所有音素重覆該辨認及修改步驟。

7 . 如申請專利範圍第 2 項所述之方法，另包括步驟：對與錯誤對齊資料中之一或更多音素相對應之正確對齊資料中之所有音素重覆該辨認及修改步驟。

8 . 如申請專利範圍第 1 項所述之方法，其中，移動至少一平均值更遠離第二音素之第二聲音模型之一對應平均值之步驟包括自第二聲音模型之平均值減第三聲音模型之平均值。

9 . 如申請專利範圍第 1 項所述之方法，其中，移動至少一平均值更遠離第二音素之第二聲音模型之一對應平均值之步驟包括減小第三聲音模型之平均值約百分之 2 (2 %) 。

10 . 如申請專利範圍第 1 項所述之方法，其中，修改一第一聲音模型另包括步驟：修改第一音素所屬之所有聲音模型，由移動其所有平均值更遠離第二音素所屬之所有第二聲音模型之對應平均值執行。

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

六、申請專利範圍

1 1 . 如申請專利範圍第 2 項所述之方法，其中，修改一第三聲音模型另包括步驟：修改第三音素所屬之所有聲音模型，由移動其所有平均值更接近第二音素所屬之所有聲音模型之對應平均值執行。

1 2 . 一種改善分段基礎之自動語音辨識系統（A S R）之性能之方法，由使用自 A S R 所用之特定應用軟體所獲得之資訊訓練其聲音模型執行，包括步驟：

接收由 A S R 所接收之一發話之一正確分段對齊序列；

接收發話之一對齊序列，此根據自特定應用軟體之內容中之語音辨識系統所接收之資訊知道為不正確；

辨認與正確分段對齊序列中之一第二音素相對應之已知不正確對齊序列中之一第一音素；

修改第一音素之第一聲音模型，由移動其至少一平均值更遠離第二音素之第二聲音模型之一對應平均值執行。

1 3 . 如申請專利範圍第 1 2 項所述之方法，另包括步驟：

接收發話之一對齊序列，此根據自特定應用軟體之內容中之語音辨識系統所接收之資訊知道為正確；

辨認與正確對齊序列中之第二音素相對應之已知正確對齊序列中之一第三音素；

修改第三音素之一第三聲音模型，由移動其至少一平均值更接近第二音素之第二聲音模型之對應平均值執行。

1 4 . 一種電腦可讀之媒體，攜帶一或更多指令序列

（請先閱讀背面之注意事項再填寫本頁）

裝

訂

線

六、申請專利範圍

，用以訓練分段基礎之自動語音辨識系統之聲音模型，其中，由一或更多處理器執行一或更多指令序列導致一或更多處理器執行步驟：

接收正確對齊資料，此代表由語音辨識系統所接收之一發話之正確分段對齊序列；

接收錯誤對齊資料，此代表根據自語音辨識系統所接收並說明該發話之資訊知道為不正確之發話之一對齊序列；

辨認與正確對齊資料中之一第二音素相對應之錯誤對齊資料中之一第一音素；

修改第一音素之聲音模型，由移動其至少一平均值更遠離第二音素之一第二聲音模型之對應平均值執行。

15. 如申請專利範圍第14項所述之電腦可讀之媒體，其中，該等指令另含用以執行以下步驟之指令：

接收發話之一對齊序列，此根據自特定應用軟體之內容中之語音辨識系統所接收之資訊知道為正確；

辨認與正確對齊序列中之第二音素相對應之已知正確對齊序列中之一第三音素；

修改第三音素之第三聲音模型，由移動其至少一平均值更接近第二音素之第二聲音模型之對應平均值執行。

16. 一種分段基礎之自動語音辨識系統，此由依據有關該系統所用之應用軟體之資訊訓練其聲音模型提供改善之性能，該系統包含：

一辨識器，此包含一或更多處理器；

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

六、申請專利範圍

不減失儲存器，連接至辨識器，並包含多個分段對齊資料及多個聲音模型；

一電腦可讀之媒體，連接至辨識器，並攜帶一或更多指令序列，用以訓練聲音模型，其中，由一或更多處理器執行一或更多指令序列導致一或更多處理器執行步驟：

接收正確對齊資料，此代表由語音辨識系統所接收之發話之正確分段對齊序列；

接收錯誤對齊資料，此代表根據自語音辨識系統所接收並說明該發話之資訊知道為不正確之發話之一對齊序列；

辨認與正確對齊資料中之第二音素相對應之錯誤對齊資料中之一第一音素；

修改第一音素之第一聲音模型，由移動其至少一平均值更遠離第二音素之第二聲音模型之對應平均值執行。

17. 如申請專利範圍第16項所述之語音辨識系統，其中，該等指令另包含用以執行以下步驟之指令：

接收發話之一對齊序列，此根據自特定應用軟體之內容中之語音辨識系統所接收之資訊知道為正確；

辨認與正確對齊序列中之第二音素相對應之已知正確對齊序列中之一第三音素；

修改第三音素之第三聲音模型，由移動其至少一平均值更接近第二音素之第二聲音模型之對應平均值執行。

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

487903
89123011

圖 1

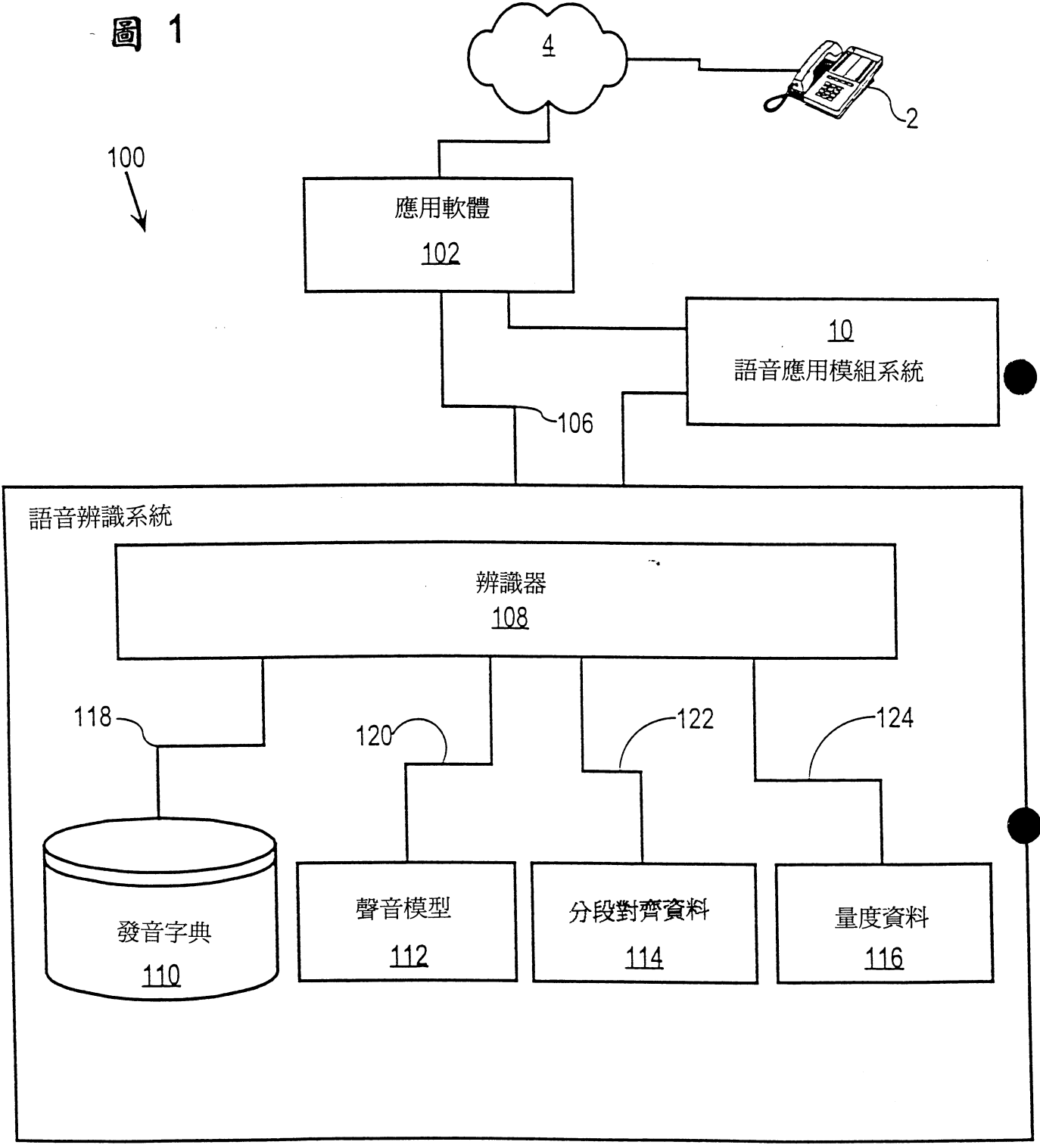


圖 2A

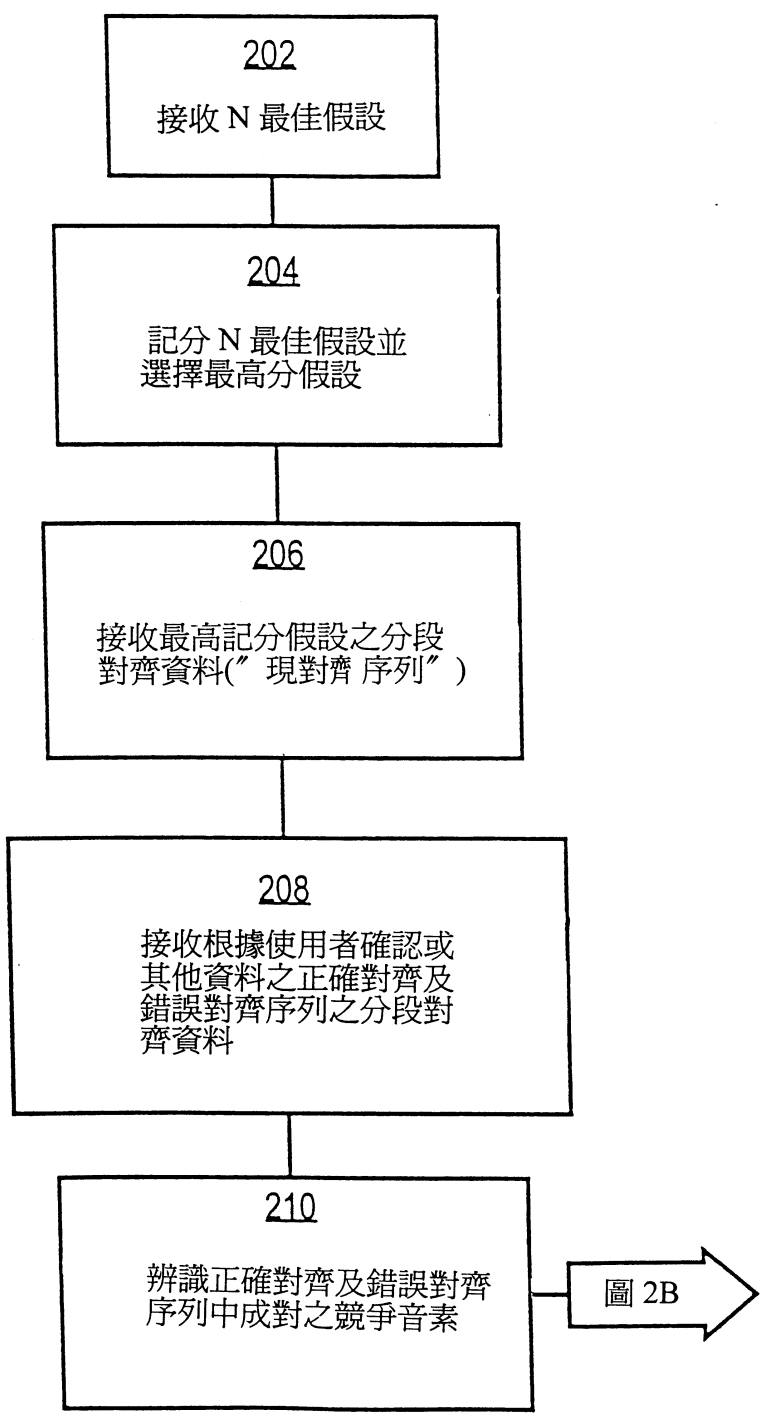
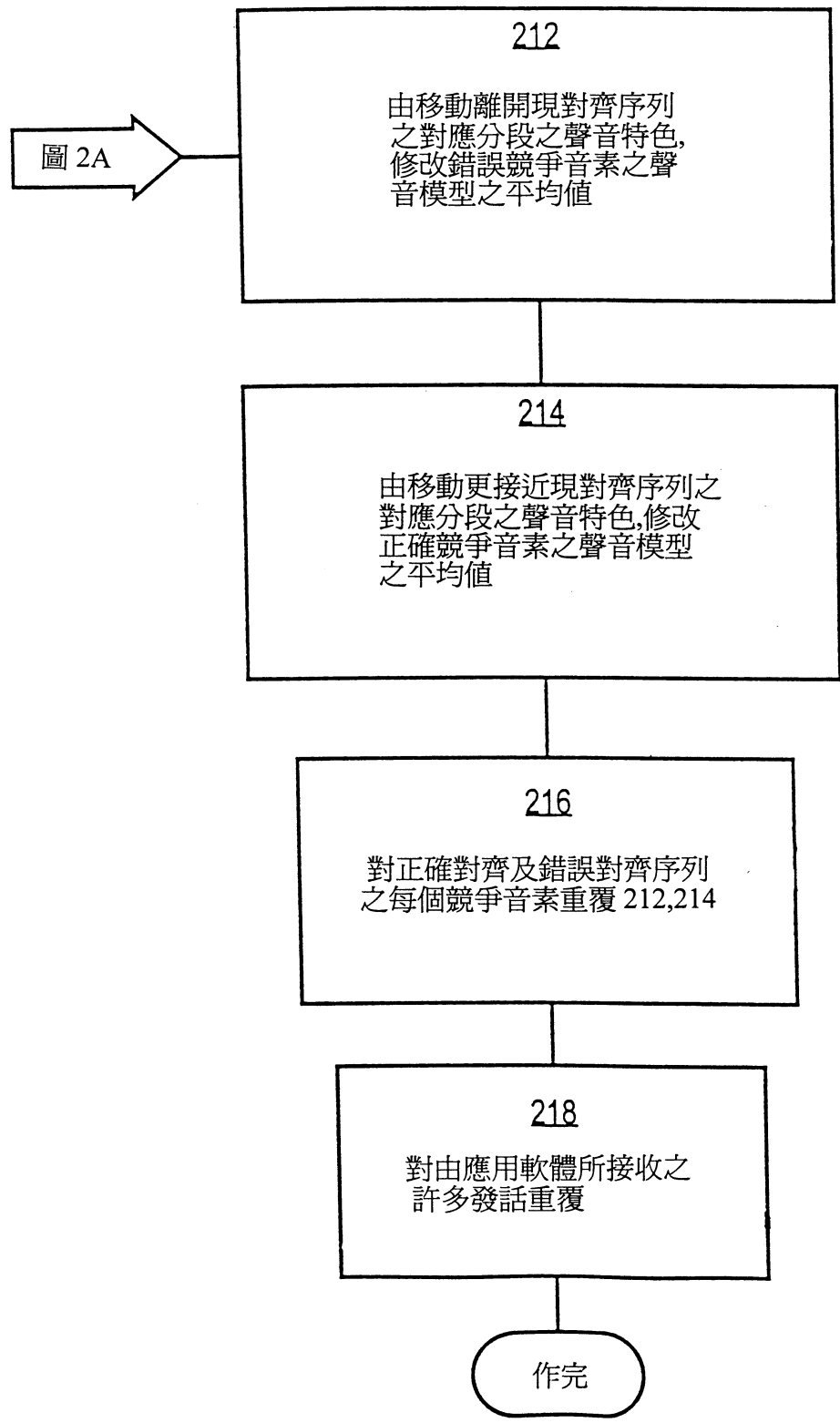


圖 2B



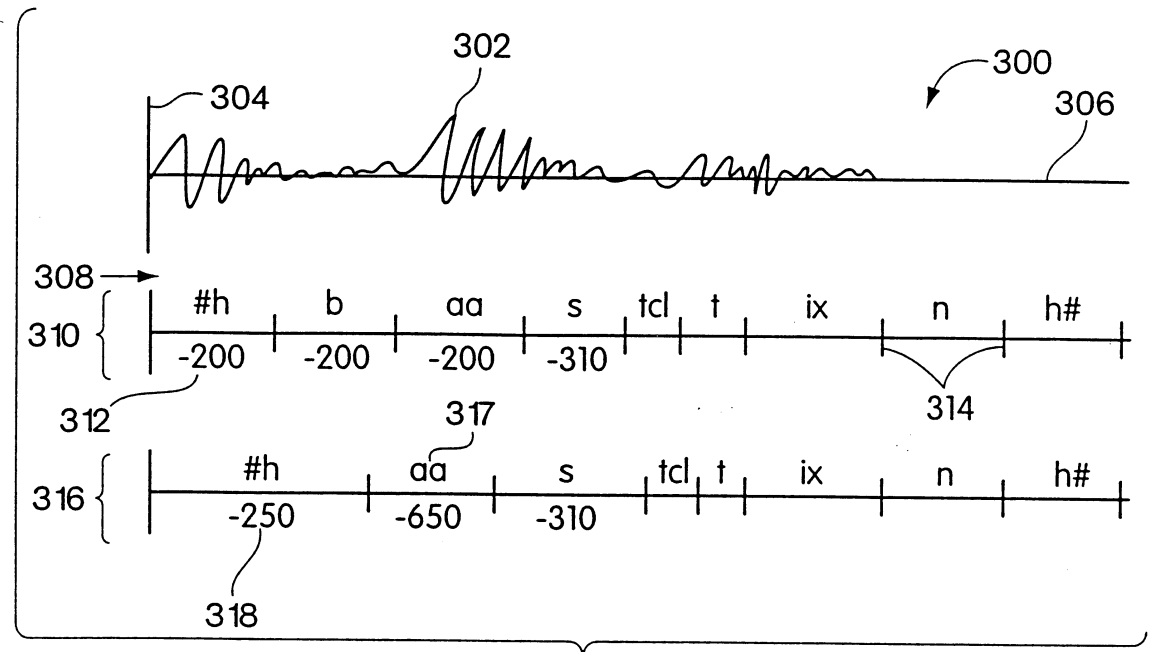


圖 3

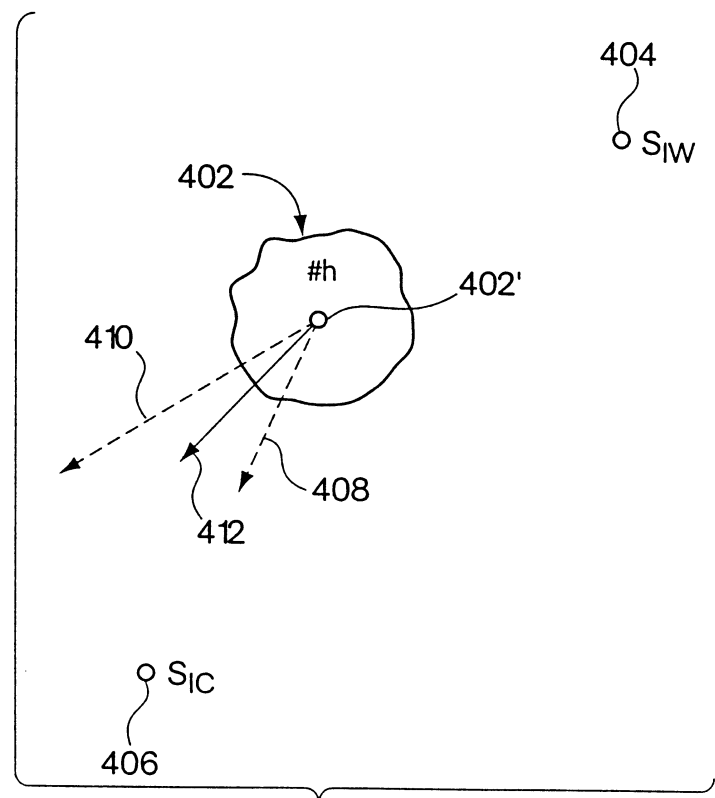


圖 4

圖 5

