

母音列을 利用한 韓國語 單語音聲認識

洪 光 錫*

Korean Word Speech Recognition Using Vowel String

*Kwang-Seok Hong**

Summary

A large vocabulary speech recognition system needs to use data driven word hypothesis, since it is difficult to achieve that accuracy in an acceptable amount of time. Pattern matching using spectrum distance measure has been used mainly in existing methods of the selection of candidate words. Using such approach, processing time increases, moreover, it is difficult to cope with variation due to individuality of a speaker and effect of coarticulation between phonemes.

In this paper, therefore, the speech recognition of the continuous word using rules is proposed. The input signal of speech is segmented according to word and syllable using the rule by the property of LSP-formant. The two method for deminishing the number of candidatures is proposed. One, which determine the number of syllable, and the other, which use a serial of vowel that has recognized as vowels in a word before. The effectiveness of the vowel string is confirmed by applying the above proposal in a station name recognition system of the subway.

序 論

音聲認識에 關한 研究는 音聲合成과 함께 오래전 부터 研究가 활발히 進行되어 最近에는 限定된 用途 이지만 音聲에 의해 機械와 對話하기 위한 音聲入力 裝置가 實現되고 있다. 音聲認識이란 音聲信號로 부터 意思傳達를 위한 內容을 抽出하는 作業을 컴퓨터 나 電子回路 등에 의해 自動으로 遂行하는 技術을 말

한다. 이러한 音聲認識에 關한 研究의 窮極의인 目標 은 不特定多數의 사람에 의해 自然스럽게 發聲된 音聲을 語彙의 制限없이 認識 하는데 있다. 이를 위해 서는 音聲信號를 適切히 表現하는 方法과 音韻境界 의 正確한 檢出方法 및 單語認識을 통한 連結單語音聲認識 方法 등이 要求된다.

單語認識은 音聲의 發聲길이의 變化를 補償하기 위해서 DTW[Sakoe와 Chiba, 1978]方法을 使用하 는 境遇 入力패턴과 標準패턴 사이의 最大 類似性을

* 공과대학 정보공학과 (Dept. of Information Engineering, Cheju Univ., Cheju-do, 690-756, Korea)

갖는 經路를 다이나믹하게 決定함으로써 좋은 認識率을 갖기 때문에 音聲認識 시스템에서 많이 使用하고 있지만 DP(Dynamic Programming)의 많은 計算量으로 實時間 音聲認識 및 連續音聲認識으로의 擴張이 어렵다. 또한 VQ(Linde 등, 1980)와 統計的手法을 利用한 HMM(Rabiner와 Jung, 1986)方式 등은 認識率은 떨어지는 反面에 認識時間이 DTW보다 빨라서 이를 利用한 大容量 隔離單語 認識시스템 開發에 대한 研究등에서 實用化의 可能性을 보여주지만 트레이닝시 많은 데이터와 時間을 必要로 한다.

또한 大語彙의 認識 시스템에 있어서 許容된 時間안에 원하는 目標을 達成하기 위해서는 認識對象 語彙중 入力音聲이 될 可能性이 있는 單語를 미리 候補로 選擇하는것이 必要하다. 候補單語의 豫備選擇을 위해서는 言語情報등의 高次元的 知識을 利用하는 top-down方法과 入力單語의 音響의 特性과 같은 低次元的 知識을 利用하는 bottom-up方法이 使用되고 있는데 處理 語彙數가 많을때는 後者の 方法이 특히 有利하다. bottom-up의 候補選擇에는 音素 또는 音素連鎖의 發生順序를 利用하는 方法(Matsunagay와 Kohda, 1978)과 音素들 사이의 混同을 防止하기 위하여 混同을 일으키기 쉬운 音素들을 그룹화 하여 利用하는 方法(Itahashi와 Yokoyama, 1984), 量子化 分布들 사이의 類似度를 利用하는 方法(Aoki 등, 1987)등 여러가지 方法이 提案되어 있다. 이상에서 列擧한 方法들은 모두 特徵벡터들 간의 類似度 比較를 통해 候補를 選擇하기 때문에 根本적으로는 音素 認識 過程과 다를 바 없고 다만 處理 基準을 緩和하여 處理時間을 短縮하는데 不過하다.

따라서 本 論文에서는 單語認識時 發聲된 單語중에서 單語를 構成하고 있는 音節數와 母音의 系列을 利用하여 候補單語를 줄인후 DTW로 單語를 認識하는 方法을 提案한다. 이 方法은 連續 發聲된 文章을 單語單位로 分離하고 그 單語內的 音節數가 決定되면 單語辭典의 單語중 音節數가 같은 單語를 候補單語로 하여 認識하는 方法과 單語內的 音節에 있는 母音을 세그먼트하여 LSP-formant의 特徵을 利用하여 母音認識을 먼저 한 후 그 母音의 系列을 보고 認識 單語 候補를 줄이는 方法을 適用하여 같은 候補에 있는 單語를 認識率이 좋은 DTW를 이용하여 認識하는 方法으로써 빠른 時間에 認識할 수 있다는 長點을 갖

는다. 本 論文에서는 이상에서 提案된 內容을 電鐵驛名을 認識對象으로 하여 實驗을 하였다.

母音列을 利用한 單語認識

大容量 單語認識을 하는 境遇 認識候補單語의 縮소는 認識時間의 短縮을 意味한다. 이장에서는 單語 認識時 候補單語의 縮小를 위하여 比較의 抽出해 내기가 쉬운 單語의 母音列을 利用한다.

音節單位로 세그멘테이션 되어있는 單語를 認識한 境遇 그 單語에 있는 母音列을 推定할 수 있다면 單語 認識을 하는데 있어 候補單語를 大幅으로 縮小시킬 수 있다. 이를 위하여는 音節에 포함되어 있는 母音을 正確하게 推定하여야 한다. 여기에서는 規則에

Table 1. Word data.

당산, 사당, 강남, 방학, 한남, 안산, 한장, 삼각지, 안양, 한양대, 합정, 삼성, 낙성대, 삼선교, 아현, 강변, 남영, 창동, 상록수, 안국, 쌍문, 압구정, 3구역, 4구역, 방배, 잠실, 가리봉, 아시오, 잠원, 반월, 망월사, 상왕십리, 상계, 서초, 성수, 성북, 석수, 서우역, 서울대입구, 성내, 성대앞, 건대입구, 선릉, 석계, 청량리, 서빙고, 신당, 신사, 미아, 신답, 미아삼거리, 신천, 시청, 인천, 신설동, 신촌, 신도림, 이대, 신대방, 시흥, 길음, 신림, 기지, 신의문, 신왕산, 1구역, 2구역, 주안, 중앙, 두정, 구파발, 무악재, 독섬, 부천, 충정로, 문래, 숙대입구, 구로, 부곡, 군포, 구로공단, 충무로, 구의, 부평, 수원, 불광, 수유, 주세요, 대방, 개봉, 백운, 대림, 대야미, 동암, 동작, 종합운동장, 봉천, 녹번, 녹천, 송내, 동대문, 동대문운동장, 홍대입구, 동대입구, 도봉, 도봉산, 종로3가, 종로5가, 고속터미널, 옥수, 오류, 홍제, 노원, 동인천, 독립문, 충신대, 노량진, 금호, 용봉, 금정, 을지로4가, 을지로3가, 을지로입구, 재물포, 제기동, 세장, 역삼, 명학, 역곡, 명동, 경복궁, 영등포, 영등포구청, 연신내, 용산, 교대, 회기, 회룡, 회현, 관악, 화서, 화물터미널, 왕십리, 의정부, 휘경, 월계, 양재, 예, 혜화, 중각, 돈암

의한 方法(홍과 박, 1990)으로 母音列 抽出 및 母音 認識方法을 適用한다. 이는 LSP-formant의 性質을 利用하여 각 母音의 規則에 의하여 認識하는 方法이다.

Table 1에서 보는 바와 같이 本 論文에서 電鐵驛名 認識시스템에 使用한 151개의 單語를 母音列로 分類하여 候補單語를 縮小한다. 이 境遇에는 113種類의 母音列이 생기는데 相當한 縮小效果를 거두지는 못하고 있다. 그러나 單語의 수가 많아질수록 그 效果를 發揮하는데 實際로 13만 2천單語인 國語辭典에서 /ㅏ-ㅓ-ㅗ/系列의 單語數는 213개로서 單語候補數는 全體의 0.16%에 그친다. [이, 1989]

또한 單語를 構成하고 있는 모든 母音列을 찾지 않고 첫번째 母音만을 찾아낸다 하더라도 本 論文의 電鐵驛名 認識시스템에서는 17 그룹으로 壓縮되고 특히 151개 單語중에서 90개 單語는 母音列을 찾는 即時 認識單語가 確定된다는 長點을 갖는다.

Fig. 1에 母音列을 利用한 單語認識에 關한 全體 流程도를 나타내었다. 音聲이 入力되면 始作點과 끝점을 抽出한후 LSP-formant 分析을 한다. 文章을 單語 單位, 音節單位로 分離한후 母音列을 抽出하고 該當하는 母音列을 參照하여 單語辭典에 있는 候補單語와 DTW를 遂行하여 單語認識을 한다. 특히 候補單語가 1개인 境遇는 即時 認識하게 된다.

Fig. 1에서 全體 單語의 單語辭典은 모든 認識候補單語의 群이고 音節數에 의한 單語辭典은 音節數가 같은 單語의 群으로 分類한 것이다. 또한 母音列에 의한 單語辭典은 單語를 構成하는 母音列이 같은 單語의 群이다.

母音抽出 및 認識段階에서는 母音列을 抽出 한 후 規則에 의하여 母音을 認識하게 된다. 그러나 母音중에서 /ㅏ/, /ㅓ/, /ㅗ/, /ㅛ/, /ㅜ/, /ㅠ/, /ㅣ/, /ㅡ/를 除外한 나머지 母音은 二重母音의 性格을 갖고 있어서 母音의 포먼트 特性이 變化하기 때문에 認識하지 못하는 結果를 가져오게 된다. 이 境遇에는 規則에 의한 方法보다 패턴매칭에 의한 認識方法을 導入하는 것이 좋다. 이때 母音列 抽出 및 認識의 成功 與否는 다음과 같은 네가지 條件에 의해 決定된다. 첫번째 條件은 母音의 길이가 60ms 이상인 境遇에만 認識한다. 두번째는 母音의 存在區間이 변하지 않아야 한다. 세번째는 規則에 의하여 母音을 認識할 때

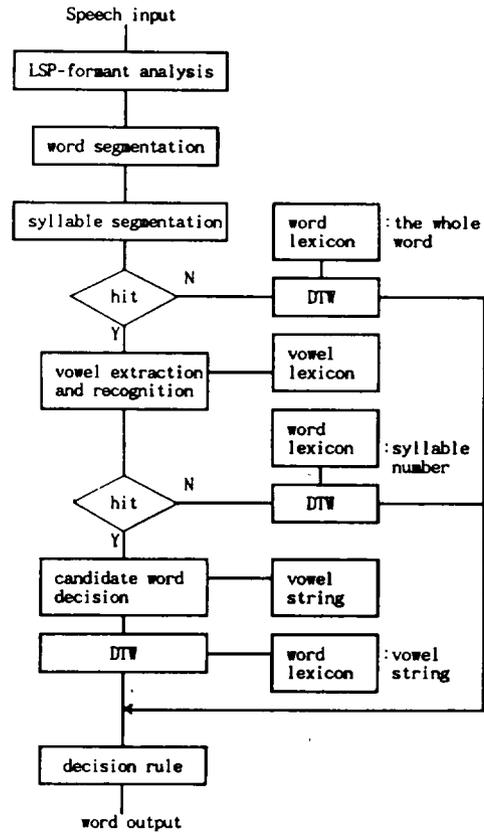


Fig. 1. Word Recognition System

均값과 주어진 基準值 以上 差異가 나지 않아야 한다. 네번째는 첫번째 母音을 認識하지 못하면 失敗한 것으로 한다. 여기에서 첫번째와 두번째 條件은 母音의 安定區間을 確保하기 위한 것 이고, 세번째 條件은 母音認識時 좀더 正確히 하기위함 이고, 네번째 條件은 單語 標準패턴이 너무 複雜해 지지 않게 하기 위함이다. 총 151개의 單語를 모음열로 구분하면 113개의 母音列로써 區分된다.

實驗 및 考察

實驗을 위해 發聲된 音聲信號는 Fig. 2에 나타낸 바와 같이 處理하였다. 마이크를 통해 들어온 音聲信號는 4KHz 低域通過 필터에 의해 帶域制限한 후, IBM-PC와 interface한 ADC0800을 거쳐 8bit 10KHz로 샘플링하였고, power와 ZCR 그리고 時間길이 情報

를 利用하여 音聲信號의 始作點과 끝점을 抽出한 후 하드디스크에 貯藏하였다. digital화된 音聲 샘플들은 差分필터 (1-Z⁻¹)에 의해 高域強調한 후 10차 LSP계수, power, LSP-formant를 抽出하였는데 한 프레임은 10ms로 分析하였고, 프레임사이에는 重疊이 되지 않도록 하였다.

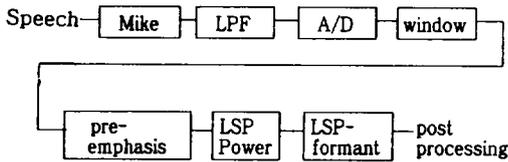


Fig. 2. Procedure of speech analysis.

세그멘테이션 實驗과 單語認識 實驗을 위해서 1인 話者가 1회씩 發聲한 20개의 文章과 11회씩 發聲한 2音節單語 91개, 9회씩 發聲한 1, 3, 4, 5, 6音節 單語 60개를 使用하였다. 이는 文章 20개, 單語 1541개에 該當한다.

또한 文章과 單語의 發聲에 있어서는 實驗室에서 周邊의 雜音에 介意치 않고 音聲을 받았는데 이는 音聲認識 시스템의 實際狀況에 조금이라도 가까이 하려 함이다.

1. 세그멘테이션 實驗 및 結果

本 實驗에서는 LSP-formant[홍과 박, 1990]를 利用하여 單語單位, 音節單位 세그멘테이션과 세그멘트된 單語의 母音列 推定을 電鐵驛名 認識 시스템에 適用하여 實驗을 한다.

實驗 데이터는 電鐵驛名 認識시스템의 實効性을 위하여 雜音이 存在하고 있는 實驗室에서 마이크를 통하여 直接 컴퓨터에 入力하였다. 여기에서는 單語單位, 音節單位 세그멘테이션과 母音列 抽出에 關한 實驗을 한다.

1) 單語單位 세그멘테이션 實驗 및 結果

提案한 單語單位 세그멘테이션 알고리즘의 有効性을 確認하기 위하여 Table 2에 있는 文章을 發聲하여 實驗 하였다. 20개의 文章音聲 으로서 여기에 속하는 單語數는 총 54개의 單語이고, 音節數는 150개 이다.

Table 2. Continuous speech data.

방배 한장 주세요.
 신사 두장 주세요.
 홍제 세장 주세요.
 삼각지 한장 주세요.
 노량진 두장 주세요.
 사구간 세장 주세요.
 종로오가 주세요.
 영등포구청 주세요.
 부평 한장 주세요.
 송내 세장 주세요.
 교대 한장 주세요.
 월계 두장 주세요.
 동대문운동장 주세요.
 상계 두장 주세요.
 울지로 삼가 주세요.
 화물터미널 주세요.
 서울역 두장 주세요.
 상왕십리 주세요.
 양재 세장 주세요.
 신도림 한장 주세요.

發聲된 文章으로부터 單語를 分離해 내는 實驗은 다음과 같은 세 段階로 하였다.

1段階 : LSP-formant와 파워 利用.

1차 LSP-formant가 1인곳과 PW가 40 이상.

2段階 : 時間길이 情報를 利用한 1 段階補完.

單語와 單語 境界部 時間길이 90ms 보다 적은곳.

3段階 : 文章形式을 利用한 2 段階補完.

4單語 以上인 境遇 뒤의 2單語 除外한 나머지는 1單語 處理.

20개의 發聲된 文章을 使用하여 세그멘테이션한 結果는 Table 3과 같다. Table 3에서 S는 段階이고, E는 에러이다.

Table 3. Segmentation results for sentence speech

S \ E	insertion error	defection error
step 1	6/54 (11.11%)	0/54 (0%)
step 2	1/54 (1.85%)	0/54 (0%)
step 3	0/54 (0%)	0/54 (0%)

이 實驗結果로 보아 文章에서 單語를 分離하는 境遇 특징파라미터 뿐만아니라 時間길이 및 文章形式을 利用한 境遇가 보다 正確한 세그멘테이션이 可能하였다.

또한 獨立의으로 發聲한 151개의 單語에 대하여는 Table 4와 같다. 이 境遇 역시 모든 單語가 正常的으로 세그멘테이션이 되고 있음을 나타내고 있다.

Table 4. Segmentation results for word speech

S \ E	insertion error	defection error
step 1	38/151 (25.2%)	0 (0%)
step 2	5/151 (3.3%)	0 (0%)
step 3	0/151 (0%)	0 (0%)

2) 音節單位 세그멘테이션 實驗 및 結果

音節單位 세그멘테이션 實驗을 하기 위하여 使用된 데이터는 文章音聲 20개와 單獨發聲된 單語 151개를 利用하여 實驗을 實施 하였다. 音節單位 세그멘테이션에서 使用된 파라미터는 LSP-formant와 파워, 時間길이 情報를 利用하였다. 또한 本 論文에서는 正確한 音節의 境界보다는 單語를 構成하고 있는 正確한 音節의 數를 必要로 하기 때문에 音節境界에서의 正確한 區分을 要求하지는 않는다.

1段階 : 1차 LSP-formant를 利用한 세그멘테이션.

1개 프레임이 不連續인 境遇 連續處理.

2段階 : 時間길이와 2차 LSP-formant를 利用한 세그멘테이션.

音節의 長이가 300ms보다 長면 2차 LSP-formant의 不連續部分離. 2차가 連續이면 3차이용. (단, 여러곳이 不連續이면 中止)

Table 5에서는 文章을 發聲하여 實驗하였는데 총

Table 5. Segmentation results for sentence speech

S \ E	insertion error	defection error
step 1	20/150 (13.3%)	22/150 (14.7%)
step 2	4/150 (2.7%)	10/150 (6.7%)

Table 6. Segmentation results for word speech

S \ E	insetion error	defection error
step 1	59/392 (15.1%)	49/392 (12.5%)
step 2	11/392 (2.8%)	23/392 (5.9%)

150개 音節중 最終인 插入에러가 4개, 脫落에러가 10개 發生하였고, Table 6에서는 單語를 發聲하여 實驗하였는데 총 392개 音節중 插入에러 11개, 脫落에러 23개가 發生하였다. 이는 LSP-formant를 利用한 簡單한 規則으로써 相當히 좋은 結果를 나타내고 있다.

3) 母音列 抽出 實驗 및 結果

母音列을 利用하여 候補單語를 選擇후 單語認識을 할 境遇 母音列을 얼마나 正確히 抽出하여 認識하는 냐가 關鍵이 되고있다. 여기에서 使用된 데이터는 音節單位 세그멘테이션에서 使用된 데이터를 그대로 利用하는데 文章이나 單語에서 얻어진 音節에서 安定된 母音部를 推定한다. 세그멘테이션된 音節에서 1차 LSP-formant의 位置가 1인값(LSP係數의 1차와 2차사이)에 存在하는 區間만을 母音列로 定한다.

/방배 한장 주세요. / 라고 發聲하여 處理한 Fig. 3을 보면 제 1차 LSP-formant의 境遇 /방배/에서 /ㅏ/, /ㅑ/가 明確하게 分割되어 있으며 2차와 3차의 境遇도 分割이 되어 있는것을 볼 수 있다. /한장/에서도 역시 /ㅏ/, /ㅑ/가 明確히 分離되어 있으나 장의 /ㅏ/는 調音結合의 影響으로 若干 不安定하게 되어있다. /주세요/에서는 /주/와 /세요/로 分離되어 있는데 /세/에서의 /ㅑ/와 /오/에서의 /ㅓ/가 母音꺼리 만났기 때문에 連續으로되어 있다. 그러나 2차와 3차 LSP-formant가 分離되어 있기 때문에 3차 까지의 포먼트를 利用하여 音節을 抽出할 수 있게 된다. 또한 音節單位로 分離된 곳을 보면 母音部를 中心으로 세

그멘테이션 되어 있는데 安定된 母音區間을 얻기 위해서는 分離된 音節로 부터 1차 LSP-formant가 連續인 區間 중에서 2차와 3차 LSP-formant가 安定되어 있는 區間만을 別途로 抽出하여 母音列로 한다. Fig. 3에서 파워부분의 가로막대 3단은 단어단위 세그먼트된 곳이고, 그아래 2단은 음절중 모음부가 추정된 곳이다.

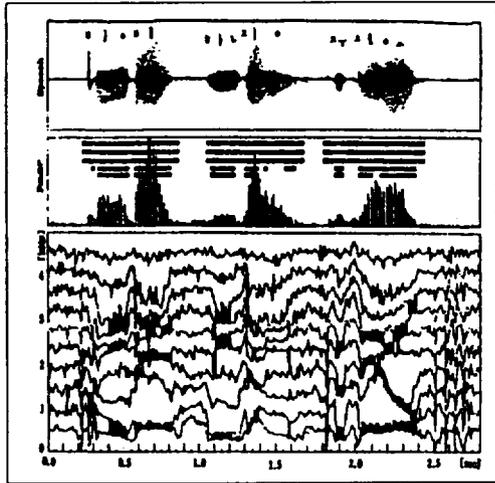


Fig. 3. Continuous Word Speech.

2. 連續單語音聲認識 實驗 및 結果

連續單語 音聲認識은 電譯辭名 認識 시스템을 假定하여 實施였다. 認識實驗 段階는 LSP-formant를 利用한 音節單位 세그멘테이션 후 한單語에 속한 音節의 갯수를 구하여 그에 該當하는 音節數와 같은 標準패턴내의 單語와 認識 함으로써 候補單語를 줄이는 方法을 使用하고, 또한 LSP-formant를 利用한 音節單位 세그멘테이션 過程으로 부터 小區間으로 區分된 母音列을 받아들여 각 母音區間에 대하여 規則에 基礎한 3개의 LSP-formant의 存在位置를 調查하여 母音列을 認識한다. 다음段階로 調查된 母音열을 利用하여 單語辭典으로 부터 調查된 母音列과 一致하는 候補單語를 設定한다. 그 다음 段階로는 候補單語와 入力單語를 DTW 計算에 의해 가장 類似도가 높은 單語를 認識 單語로 決定한다. 이러한 方法을 反復 하여 連續單語 音聲認識을 遂行한다. 그러나 LSP-formant를 利用하여 調查된 母音列이 單語 辭

典에 없는 境遇는 全體 電譯辭名에 대하여 DTW方式을 適用하여 單語 認識을 遂行한다. DTW 計算時 파라미터는 LSP係數를 使用하였다.

1) 音節數와 DTW에 의한 單語認識 實驗 및 結果

音節數와 DTW에 의한 單語認識은 單語를 認識할 때 認識單語의 候補를 縮小하여 認識時間을 短縮하고자 하는것 이다. 이장에서는 音節數를 利用하지 않고 모든 候補와 距離計算을 하여 認識하는 境遇(方法 1)와 單語內의 音節數가 理想的으로 구해졌다고 假定했을 때의 音節數를 利用하여 候補를 縮小하여 認識하는 境遇(方法 2)에 대하여 實驗을 하였다.

方法 1의 境遇 2 音節 單語는 標準패턴을 10개씩 使用하였으며 나머지는 8개씩의 標準패턴을 使用하였다. 이는 2音節 單語數가 많기때문에 다른것 보다 2개의 標準패턴을 더 使用한 것이다. 方法 2의 경우 모든 單語의 標準패턴을 3개씩 使用하여 實驗하였다. 決定單語는 두 方法 모두 KNN法을 使用하였는데 K=2를 適用하였다.

Table 7. Word recognition results for word speech

M \ R	recognition rate
method 1	150/151 (99.3%)
method 2	149/151 (98.7%)

Table 7에서 M은 方法이고 R은 認識率인데, 方法 1은 標準패턴의 數를 總 1390개 使用하였고 方法 2는 453개의 標準패턴을 使用하였다. 그러나 認識率의 境遇는 거의 差異가 생기지 않는다는 것을 알 수 있다. 文章音聲에서 分離된 單語를 實驗데이터로 한 境遇의 結果를 Table 8에 나타내었다.

Table 8. Word recognition results for sentence speech

M \ R	recognition rate
method 3	54/54 (100%)
method 4	54/54 (100%)
method 5	52/54 (96.3%)

方法 3은 모든 候補單語와 比較한 境遇로서 각 單語 3개씩 총 453개의 標準패턴과 比較하였다. 方法 4는 單語內的 音節數가 理想的으로 구해졌을 때의 境遇로서 1音節의 標準패턴은 총 3개, 2音節은 273개, 3音節은 123개, 4音節은 24개, 5音節은 27개, 6音節은 3개의 標準패턴을 使用하였다. 方法 5는 實際로 音節數를 구하여 適用한 境遇이다. 音節數 구할때 成功與否의 初期值는 音節의 길이는 300ms로 하였고 音節과 音節사이의 間隔은 40ms로 하였다. 이때 誤認識된 單語는 /부경/과 /사구간/인데 音節 세그멘테이션시 音節數에 插入에러가 發生한것을 認知하지 못하고 候補를 選擇했기 때문이다.

2) 母音列을 利用한 單語認識 實驗 및 結果

母音列을 利用한 單語認識의 有效性을 確認하기 위한 實驗에서 使用된 데이터는 20개의 文章인데 文章內的 單語를 合하면 總 54개의 單語를 使用하였다. 이 實驗에서는 母音列만을 利用하여 實驗하는 境遇(方法 1)와 音素의 數와 母音列을 結合한 境遇에 대한 境遇(方法 2)에 대하여 實驗한다.

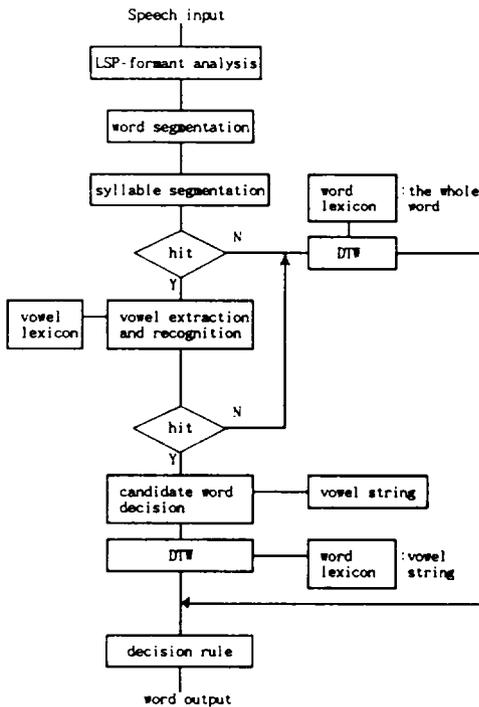


Fig. 4. Word Recognition System

方法 1의 境遇는 Fig. 4에 의한 節次에 따라서 實驗하고, 方法 2의 境遇는 Fig. 1에 의한 節次에 따라서 實驗한다. 方法 1에서는 連續音聲을 LSP-formant分析하여 單語單位, 音節單位로 세그멘테이션하였다. 다음 段階로 音節單位로 세그멘테이션한 後에 母音列을 抽出하여 母音列 單語辭典과 DTW를 利用하여 單語認識 하였다. 이때 母音抽出이 不可能할 때 全體 單語辭典과 比較하였다.

Table 9. Word recognition results for sentence speech

M	R	recognition rate
method 1		51/54 (94.4%)
method 2		49/54 (90.7%)

Table 9에 單語認識 實驗 結果를 나타내었다. 方法 2에서는 音節數를 구할때 에러가 發生한 境遇 그 것을 發見하지 못하면 致命的인 認識 에러가 發生하는데 여기에서는 2개가 發生하였다.

實驗條件은 母音의 길이 60ms 以上인 境遇만 認識하도록 하였고 平均값과 30Hz以上 差異가 나면 母音 認識이 失敗하도록 하였다. 또한 母音의 存在區間이 1, 2, 3차 LSP-formant중 하나라도 변하거나 母音列의 첫번째 母音을 認識하지 못하면 reject하여 전단에서 單語認識을 하도록 하였다. 이 實驗에서는 音節數를 구할 때의 成功與否의 條件과 母音認識時 成功與否의 條件에서 認識率을 크게 左右하는데 條件이 緩和되면 認識時間은 빠르나 認識率이 떨어지고, 條件이 強化되면 認識率이 좋아지나 認識時間이 더 길린다는 特性을 갖는다.

各 段階 認識單語별 認識時間 정규화를 Table 10에 나타내었다. 1段階는 單語辭典內的 모든 單語와 比較한 境遇이고, 2段階는 音節數에 의한 單語辭典과 比較한 境遇이다. 또한 3段階는 單語의 母音列에 의한 單語辭典과 比較하는 境遇에 대한 것이다. 특히 3段階에서 /ㅏ, ㅑ/系列에 속하는 것은 7개의 單語이고, /ㅣ, ㅓ/系列에 속하는 것은 4單語 인데 /ㅓ, ㅑ/系列도 4單語가 있다. 마찬가지로 하여 /ㅜ, ㅠ/, /ㅗ, ㅋ/, /ㅛ, ㅠ, ㅋ/, /ㅣ, ㅓ/, /ㅛ, ㅑ/, /ㅛ, ㅑ/, /ㅓ, ㅑ/系列은 각각 세 單語씩 包含하고 있다. 또한 /

Table 10. Normalized word recognition time

	comparison word	comparison word number	normalized recognition time
step 1	whole word	151×3(453)	453/453(1.0)
step 2	1 syllable	1×3(3)	3/453(0.007)
	2 syllable	91×3(273)	273/453(0.603)
	3 syllable	41×3(123)	123/453(0.272)
	4 syllable	8×3(24)	24/453(0.053)
	5 syllable	9×3(27)	27/453(0.060)
	6 syllable	1×3(3)	3/453(0.007)
step 3	/ㅏ, ㅑ/ group	7×3(21)	21/453(0.046)
	/ㅣ, ㅓ/ group+1	4×3(12)	12/453(0.026)
	/ㅓ, ㅕ/ group+6	3×3(9)	9/453(0.020)
	/ㅏ, ㅑ/ group+12	2×3(6)	6/453(0.013)
	the rest group+90	1×3(3)	3/453(0.007)

ㅏ, ㅑ/, /ㅏ, ㅑ/, /ㅏ, ㅣ, ㅓ/, /ㅏ, ㅓ/, /ㅣ, ㅓ/, /ㅣ, ㅣ/, /ㅣ, ㅓ, ㅕ/, /ㅓ, ㅕ/, /ㅓ, ㅕ, ㅣ, ㅓ/, /ㅓ, ㅕ, ㅣ, ㅓ, ㅕ, ㅑ/, /ㅓ, ㅕ, ㅣ, ㅓ, ㅕ, ㅑ, ㅏ, ㅑ/, /ㅓ, ㅕ, ㅣ, ㅓ, ㅕ, ㅑ, ㅏ, ㅑ/ 각각 2單語씩 包含하고 있다. 그외의 나머지는 모두 한 單語 이기 때문에 母音列이 確認되는 대로 認識이 可能하다. Table 10에서 보는 바와 같이 認識時間은 音聲分析時間을 除外하고 全體單語와 比較하는 境遇에 時間을 1로 하여 나머지 境遇와 比較하였다.

摘 要

LSP-formant의 特徵을 利用하여 規則에 의한 單語단위, 音節단위 세그멘테이션, 音節數와 DTW를 利用한 單語音聲認識, 母音列을 利用한 單語音聲認識 實驗을 하였다. 특히 規則을 利用한 母音認識의 境遇 각 母音을 表現할 수 있는 LSP-formant 存在區間이 母音별로 서로 다른 位置에 存在하기 때문에 그 位置 別로 母音을 分類하여 簡單한 規則으로 認識 實驗을

할 수 있었다.

單語 認識 實驗에서는 電鐵驛名 認識 시스템을 構成 하였는데, 특히 候補單語를 줄이기 위해 앞에서 處理된 單語중의 音節數를 利用하여 認識 候補單語數를 相當히 減縮할 수 있었고, 音節로부터 認識된 母音의 列을 利用하여 再次 候補單語數를 줄일 수 있다는 長點을 갖고있다. 특히 母音列을 利用할 境遇 DTW 計算을 하지않고 認識單語가 決定되는 境遇도 發生하기 때문에 相當한 認識時間의 短縮을 가져오게 된다. 또한 實驗에 使用한 單語는 151개를 使用하였는데 이 시스템에서는 單語가 어느정도 더 追加가 되어도 認識時間에는 크게 影響을 주지 않는다는 長點을 갖는다.

文후으로 부터 規則을 利用하여 單語를 세그먼트한 후, 單語認識時 候補單語를 줄이기 위하여 母音列을 認識하여 使用하였는데 이 境遇 復母音에 대하여는 簡單한 規則만으로 解決하기는 어렵고 패턴認識 등의 方法을 거쳐서 解決해야 할 것으로 생각된다.

參 考 文 獻

- Aoki, M., T. Sato, K. Shirai, 1987, Speaker-Independent Isolated Word Recognition using Vector Quantization, IECE Tech. Rep., Vol. IT87-37, pp. 61-63.
- 홍광석, 박병철, 1990, LSP의 次數間 距離정보를 이용한 formant 抽出과 韓國語 母音認識, 한국음향학회지 9권5호.
- Itahashi, S. and S. Yokoyama, 1984, Vocabulary Reduction Effect by Specifying Phoneme Sequences in Word, Trans. IECE, Vol. J67-D, 8, pp. 869-876.
- 이회정, 1989, 音素單位 音聲認識을 위한 自動세그멘테이션과 單語假說, 성균관대학교 박사학위논문.
- Linde, Y., A. Buzo, R. H. Gray, 1980, An algorithm for vector quantizer design, IEEE Trans. Commumun., Vol. COM-28.
- Matsunaga, S. and M. Kohda, 1978, Reduction of Word and Minimal Phrase Candidates for Speech Recognition on Phoneme Recognition, Trans. IECE Jpn., Vol. J70-D, 3, pp. 592-600.
- Rabiner, L.R. and B.J. Juang, 1986, An introduction to hidden Markov Models, IEEE ASSP Magazine.
- Sakoe, H. and S. Chiba, 1978, Dynamic programming algorithm optimization for spoken word recognition, IEEE Trans. Acoust., Speech and Signal Proc., vol. ASSP-26.