필터뱅크 出力을 利用한 실시간 隔離 單語 認識에 관한 연구

A Study on the Real Time Recognition of Korean Isolated Words with Filter Bank Output

김 계 국* 이 종 악* 강 성 진**

(Kye Kook Kim, Jong Arc Lee, Seong Jin Kahng)

概 要

本 硏究에서는 韓國語 10개 都市名을 認識 對象으로 하였다. 各 單語는 男性 話者 10人에 의하여 5번씩 反復 發音한 500單語를 對象으로 하여 필터뱅크 出力을 抽出하여 認識 파라메터로 使用하였다.

필터뱅크는 RC 能動素子를 利用하여 200[Kz]부터 1/3 octave 間隔으로 15채널로 構成하였다. 基準音은 集團化 알고리즘에 의해 設定하였으며 類似度 比較를 위해 DTW 알고리즘을 利用하였다. 距離 計算式에 따른 認識結果를 把握하기 위하여 유크라드 式과 체비셔브 式을 使用하여 距離를 計算하였으며 認識 結果 각각 16.4[%], 15.0[%]의 誤認識率을 얻었다.

ABSTRACT

In this paper, 10 city names of Korean were recognized. The name are articulated each 5 times by 10 male speakers. Filter bank output on total 500 words were extracted and they were used as feature parameters. Filter bank was constructed of 15 channels with 1/3 octave spacing from 200[Hz], using RC active circuit. Reference templates were created by clustering algorithm. DTW algorithm was used to compare similarity between reference templates and input words.

Euclidean distance equation and Chebyshev distance equation were used to know the distinction between the recognition results obtained by the method of distance caculation, error rates are 16.4[%], 15.0[%], respectively.

I. 序 論

1970年代 初부터 많은 硏究陳들은 人間과 通話할 수 있는 能力을 갖춘 機械를 設計하는데 專念해 왔으며, 이를 實現하기 위해 音聲 認識이 硏究되어 왔다. 音聲 認識의 目的은 機械로 하여금 人間의 自然的인 音聲을 理解하여 理解된 音聲에 따라 業務를 遂行토록 하는데 있다. 이를 위해 많은 硏究陳들이 音聲 認識, 音聲 合成, 話者 識別, 映像認識등을 硏究해 왔다. 音聲은 똑 같은 言語라 할지라도 單獨으로 發聲할 때와 [illegible] 發生되었을 때 그 音色이 다르며 그 意味를 지닌 內容이 發聲者의 性別, 나이, 發聲時의 感情, 氣分에 의하여 複雜하게 변한다.

이와 같은 原因때문에 音聲 認識 裝置의 實用化가 어렵게 되었음에도 불구하고 18세기 後半부터 시작

*전국대학교 전자공학과
**대유전문대학 산사통신과

된 信號 處理에 관한 硏究는 1939년 Dulley의 Vocoder 에 관한 硏究가 發表되면서 부터 音聲信號의 基礎 硏究가 시작되었다. 사람은 發音할 때마다 그 速度가 多樣하기 때문에 類似度 比較를 위해 時間軸의 變化를 考慮해야 한다. 그러므로 類似度 比較를 위한 方法은 DTW, VQ, HMM 등이 있다. 이들 中에서 DTW 利用한 方法은 VQ나 HMM을 利用한 方法보다 計算量이 많고 認識 遂行時間이 많이 걸린다는 短點이 있으나 認識率이 높은 것으로 알려져 있다. 그러므로 이러한 理由 때문에 一般的으로 DTW 方法을 많이 使用하고 있으며, 本 硏究에서도 基準音과 入力音 사이의 類似度 比較를 위해 이 方法을 使用하였다.

특히 필터 뱅크를 導入한 理由는 人間의 音聲을 통해서 工場 시스템을 制御하거나 電話 番號, 버스 路線 案內 등 自動 應答이 可能하게 될 自動化時代에는 신속한 應答이 要求되기 때문이다. 이처럼 實時間 處理를 要하는 境遇는 소프웨어的으로 特徵 파라미터를 抽出하면 많은 時間이 걸리기 때문에 필터 뱅크 出力을 利用한다면 遂行時間을 크게 줄일 수 있다.

1983년 Rabiner 등은 필터 뱅크시스템 構成에 대해 硏究한 바 있으나, 國內에서의 硏究 實情은 미진한 狀態이며 필터 뱅크를 하나의 칩으로 小型化하여 우리의 技術로 實時間 認識 시스템을 開發할 必要가 있다고 생각 되므로, 이를 위하여 本 硏究에서는 RC 能動 素子를 使用하여 필터 뱅크 認識 시스템을 構成하였다.

Ⅱ. 音聲 信號 處理

一般的으로 音聲 認識은 音聲 信號의 特徵을 抽出하고 이 音聲 信號들 중에서 基準音을 設定하여 컴퓨터에 貯藏하고 入力 信號와 類似度를 比較하는 3가지 基本 過程을 通해서 이루어진다.

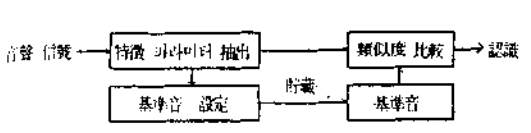


그림 1. 音聲認識의 基本 過程

오늘날 音聲 認識에서 入力 音聲을 表現하기 위해서 線形 豫測 係數(LPC)를 使用하고 있다. 一般的으로 이 線型 豫測 係數를 計算하기 위해서는 比較的 값 비싼 裝置가 必要하다. 이러한 音聲 認識 裝置의 價格을 줄이기 위해서 提案된 方法이 필터뱅크 分析에 의한 것이다. 즉 LPC 係數 대신 필터뱅크 出力 에너지를 利用하여 音聲 信號를 表現하고 있다. 그것은 값싼 集積回路를 利用하여 필터 뱅크를 構成할 수 있기 때문이다. 이러한 經濟的側面外에도 필터 뱅크 認識 裝置를 實驗에 使用하는 두가지 理由가 있다.[7]

첫째, 人間의 귀는 필터 뱅크와 類似 한 構造로 音聲 信號를 理解하기 때문이다.

둘째, LPC 認識 시스템과 廣帶域 音聲(0~10 [KHz])을 使用하는 필터뱅크 認識 시스템은 標準 音聲(Standard word)을 對象으로 하여 認識 할 경우 認識率이 거의 類似하기 때문이다.

또한, 필터 뱅크 시스템에서 LPF의 차단 주파수는 일반적으로 20~30Hz이므로 샘플률은 40~60Hz 정도로 설정 할 수 있다. 만일 필터의 창수가 5일 때, 샘플률을 40Hz로 하면 1초 동안의 신호는 총 200개의 특징파라미터로 표현할 수 있다. 8KHz로 샘플링한 신호에 비해 데이타량을 약 40:1로 감소시킬 수 있다.

音聲 認識을 위해 提案된 필터 뱅크의 構造는 多樣하지만 特定한 用途에 어떠한 필터 뱅크를 選擇해야 하는가에 대한 指針은 없다. 現在까지 音聲 認識에서는 構造가 다른 필터 뱅크의 效能(單語의 誤差率)을 比較해 왔을뿐이며, 필터 帶域幅, 필터의 數, 필터의 類型 등은 體系的으로 調査된 바가 없다. 필터 뱅크의 特徵 파라미터를 從來의 DTW 알고리즘을 使用하여 認識하는 硏究는 關心의 對象이 되고 있으며, 그림 2에서 알 수 있는 바와 같이 音聲 信號는 Q개의 帶域 通過 필터로 出力시키고 있다. 이를 위해 모든 필터는 아날로그 필터의 나지딘 필터로 構成하고 있다. RC Active 필터를 사용한 것은 회로 구조가 간단하고 쉽게 구성할 수 있으며 Digital 필터에 비해 고속 처리가 가능하므로 실시간 처리에 좋기 때문이다.

이러한 帶域 필터 뱅크는 周波數 스펙트럼을 여러 帶域으로 分散시키고 있으며, 旣存의 시스템(商業用이나 實驗用 單語 認識 裝置)에서 필터의 數는 5-32個까지 選擇하고 있다. 그 理由는 音聲 認識을 위해 使用하는 大部分의 필터 뱅크 시스템은 보코더에 使用된 設計에 基礎를 두고 있기 때문이다. 이들 필터 帶域은 一般的으로 모든 周波數 스펙트럼이 연속적이며 특히, 모든 필터 뱅크의 混成스펙트럼은 平坦하다. 이것은 모든 周波數스펙트럼에 대하여 同一한 값(Weighting)이 주어졌음을 立證하고 있다. 代表的인 方法은 周波數 스펙트럼을 均一하게 나누고 필터 帶域의 周波數 範圍를 一定하게 決定하고 있다. 또한 周波數의 範圍를 代數的으로 均等하게 하여(즉, 옥타브 또는 $\frac{1}{3}$ 옥타브) 필터 帶域을 정해주거나 音의 明瞭度(Articulation)와 같은 音聲 情報 抽出에 關係된 周波數 範圍에 따라 決定하고 있다.

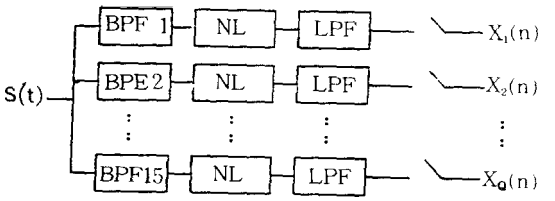


그림 2. Q次 필터 뱅크에서의 特徵抽出 블럭도

그림 2에 나타나 있는 바와 같이 各 帶域 通過 필터의 出力은 一般的으로 自乘 檢波器나 全波 整流器와 같은 非線型 回路를 通過하고 있다. 이 非線型 回路는 모든 周波數 스펙트럼에 걸쳐서 帶域이 制限된 信號 에너지를 不均一하게 分配하는 效果를 갖고 있다.

그러나 低周波에서의 信號 에너지는 一般的으로 制限된 全 帶域의 信號 에너지에 比例한다. 그러므로 音聲 信號가 非線形 回路와 低域 필터를 通過할 때 低域 필터의 出力은 特定한 周波數 領域에서의 音聲 信號 에너지가 된다. 各 瞬間에서의 에너지 값은 Q次 特徵 벡터로 表現할 수 있다. 이들 特徵 벡터의 時間的 變化(Time variation)는 音聲 패턴을 定義할 수 있다. 그러므로 時間 n인 瞬間의 信號를 $X_I(n)$으로 表現할 때 特徵 벡터는 式(1)과 같이 表現할 수 있다.[5]

$$X_I(n)=\{x_1(n),\ x_2(n),\ \cdots\cdots,\ x_Q(n)\} \quad (1)$$

n=1,2, ……, N일때 音聲 패턴 X는 다음과 같이 定義 할 수 있다.

$$x=\{X(1),\ X(2),\ \cdots\cdots,\ X(N)\} \quad (2)$$

이렇게 定義된 音聲 패턴은 基準音 設定과 DTW 알고리즘에 의한 類似性比較에 使用된다. 基準音 設定을 위해 K-means 알고리즘을 利用했다.

特徵 파라미터가 N개로 構成되어 있을때

$$\Omega=\{\chi(1),\ \chi(2),\ \cdots\cdots,\ \chi(N)\} \quad (3)$$

으로 表現 할 수 있으며 集團은 다음 式에 準하여 分類하였다.

$$\chi(j)\ \in\ \omega i \text{ if } \delta(\chi_j,\ \chi_p^{(i)}) \leqq \delta(\chi_j,\ \chi_p^{(k)})\ k \neq i \quad (4)$$

ω_i는 i번째 集團을 나타내며 各 集團內의 基準音은 모든 單語들과의 거리값이 最大인 單語들 中 그 값이 가장 작은 單語를 選擇하였다.

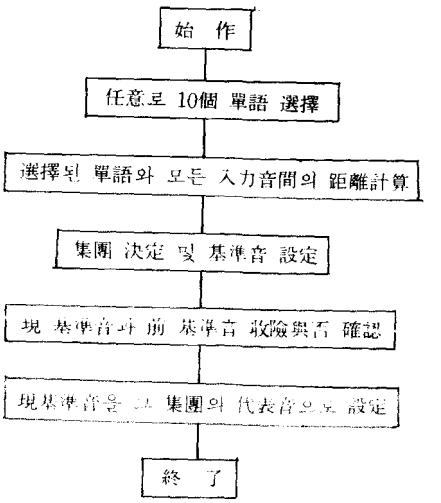


그림 3. 본 硏究에서 使用된 集團化 흐름도

基準音 設定과 類似度 比較를 위해 DTW 알고리즘을 使用하였다.

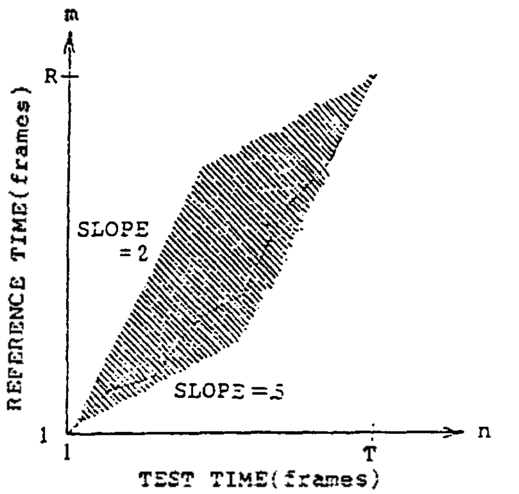


그림 4. 기울기 2와 1/2로 制約된 境遇 最適經路의 進行範圍

點(1,1)에서 (n,m)까지 最適 經路를 따라 累積된 距離는 Itakura의 制約條件에 準하였으며 두 音聲 패턴 T(n)과 R(n)間의 距離는 체비세브 式과 유크리드 式을 使用하였다.

$$d(n, m) = \sum_{i=1}^{15} |Tn(i) - Rm(i)| \quad (5\text{-}1)$$

$$d(n, m) = \sum_{i=1}^{15} [(Tn(i) - Rm(i))^2]^{1/2} \quad (5\text{-}2)$$

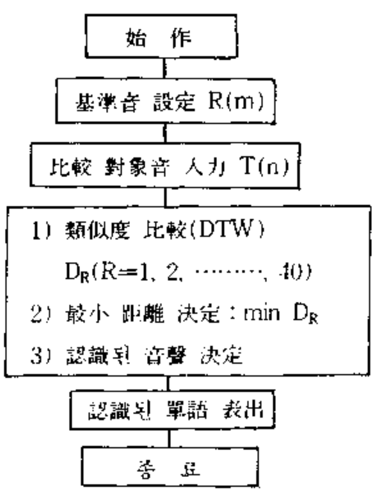


그림 5. 認識흐름도

Ⅲ. 認識 시스템 概要

필터 뱅크 시스템은 필터 뱅크 칩을 利用하여 하드웨어的 方法으로 構成하거나 소프트웨어的인 方法으로 構成할 수 있다. 그러나 소프트웨어的 具現은 遂行時間에 수반되는 時間的인 問題 때문에 實時間 次元에서 볼 때 바람직하지 못하다. 또한 필터 뱅크 칩을 使用하여 하드웨어的으로 構成할 수도 있으나 本 研究에서는 필터 뱅크 認識 시스템을 自體的으로 構成하는데 目的이 있기 때문에 RC能動素子를 使用하여 構成하였다. 필터의 數를 決定하는 것은 正確히 밝혀진 것은 없으나 5-32個 程度로 알려져 있으며 本 研究에서는 RC 能動 素子를 使用하여 15個 채널로 構成하였다. 필터의 帶域幅은 1000[Hz] 까지는 線型的으로 區分하고 1000[Hz] 以上의 周波數 範圍에서는 로그 單位로 帶域을 決定하고 있다. 그러나 本 研究에서는 좀 더 精密한 周波數 分析을 위해 200[Hz]부터 6000[Hz]까지 1/3 옥타브씩 15個 帶域으로 分離하였다. 本 研究를 위해 設定한 帶域幅은 그림 6과 같다.

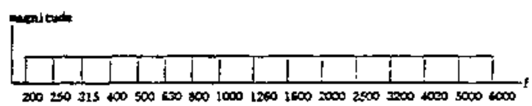


그림 6. 필터뱅크 周波數 帶域幅

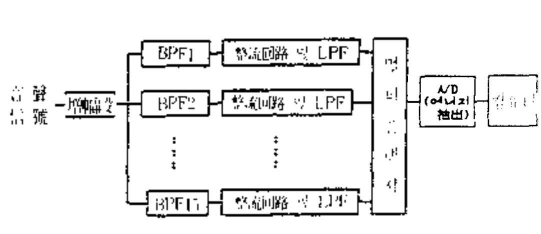
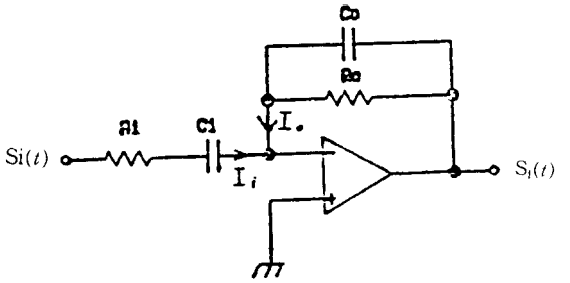


그림 7. 필터 뱅크 認識 시스템의 全體 [illegible]

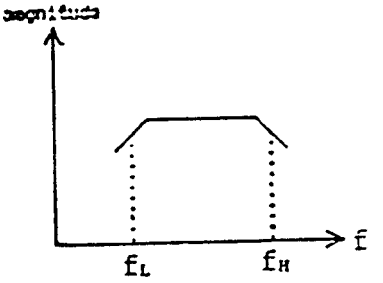
그림 7은 필터 뱅크 認識 시스템의 全體 [illegible] 나타내고 있다. 一般的으로 마이크의 出力은 [illegible] 微弱하기 때문에 入力된 音聲을 [illegible] 增幅할 [illegible] 있다. 增幅器는 [illegible] 및 [illegible] 增幅되 [illegible] 雜音에 대한 것을 考慮하여 [illegible] 構成 [illegible].

人間의 音聲은 좁은 입안에서 갑자기 넓은 空間으로 放射될때 높은 周波數 成分이 크게 弱化되므로 音聲認識 遂行에 편리하도록 4번째 增幅段에서 高域을 強調하였다. 高域 強調 過程에서 全般的으로 信號가 減衰함으로 이 信號를 增幅시키고 BPF를 通過시켰다.

BPF의 入出力 關係는 그림 8로 부터



(a)



(b)

그림 8. BPF 回路圖 및 周波數 特性

$$H(S)=\frac{Sj(s)}{Si(s)}=\frac{SCiRo}{(1+SCoRo)(1+SCiRi)} \quad (6)$$

여기서 各 帶域의 上, 下段 周波數는 各各

$$f_H=\frac{1}{2\pi RoCo} \quad (7\text{-}1)$$

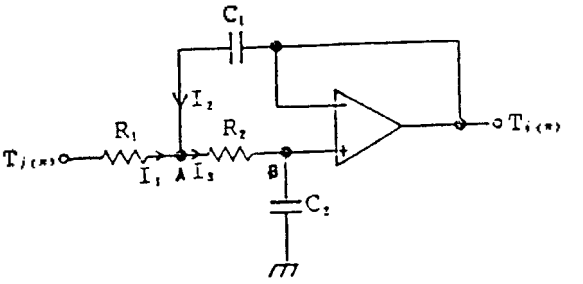
$$f_L=\frac{1}{2\pi RiCi} \quad (7\text{-}2)$$

이며, 中心 周波數는 다음과 같이 表現할 수 있다.

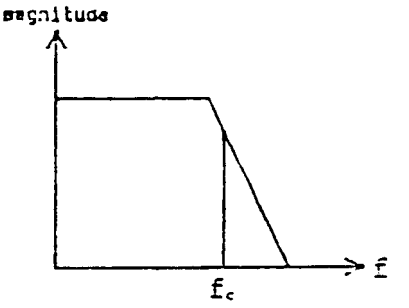
$$f_O=\sqrt{f_H f_L} \quad (8)$$

필터를 通過하여 나온 信號는 特定 周波數 成分이지만 正負로 振動하게 된다. 여기서는 波形의 振幅을 高麗해야 하므로 全波 整流 回路를 使用하여 負의 信號를 正의 信號로 反轉시켰다. 이렇게 하여 整流된 波形도 若干의 振動 信號로 存在하기 때문에 底域通過 回路로 出力시켰다.

LPF의 入出力 關係는 그림 9로 부터



(a)



(b)

그림 9. LPF의 回路圖 및 周波數 特性

$$H(S)=\frac{1}{S^2C_1C_2R_1R_2+SC_2(R_1+R_2)+1} \quad (9)$$

簡單히 表現하기 위해 $R_1=R_2=R$과 $C_1=2C_2$라고 하면 LPF의 遮斷 周波數는 다음과 같다.

$$fc=\frac{\sqrt{2}}{2\pi RC_1} \quad (10)$$

필터 뱅크에 使用된 各 필터는 2次 버터워스型으

로 構成하였다.

본 實驗에 使用한 A/D 變換器는 15채널을 變換할 수 있도록 멀티 플렉서를 수반하여 構成하였다. 이때 低域 通過 필터의 遮斷 周波數는 一般的으로 20~30[Hz] 이므로 本 硏究에서는 100[Hz]로 샘플링하여 出力 에너지를 抽出하였다.

音聲信號의 샘플링을 위한 變換器는 便利한 데이타 處理를 위해 컴퓨터에 接續하여 使用할 수 있도록 製作하였다. 變換器는 充分한 解象度를 提供하기 위해 比較的 高速 作用이 可能한, 12비트 축차 比較型인 ADC574를 使用하였으며 IBM-PC의 슬롯을 통해 PPI 8255를 인터페이스하고 8255를 利用하여 ADC574를 制御할 수 있도록 하였다.

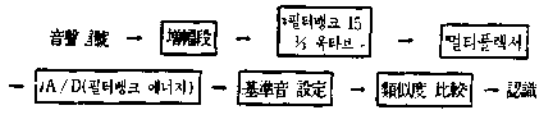


그림 10. 音聲 認識 시스템 블럭도

Ⅳ. 認識 結果

本 硏究에서는 不特定 話者의 音聲 認識을 위해 첫 音節과 둘째 音節이 類似한 인천, 춘천, 충무, 충주, 대구, 대전, 제주, 경주, 마산, 부산, 등 10個 都市名을 選擇하여 認識하였다. 各 都市名에 대하여 10名의 男性 話者가 5番反復 發音하였으며 總 500單語를 認識 對象으로 하였으며, 周圍 雜音을 考慮하지 않은 日常 生活 環境下에서 認識하기 위하여 一般 硏究室에서 전혀 發音 練習이 되지 않은 話者들로 하여금 直接 發音토록 하여 디지탈化 하였다.

필터 뱅크의 出力은 全波 整流된 信號이므로 信號의 에너지를 利用하여 끝점을 檢出했다. 마이크의 入力信號가 없을때에도 白色雜音이 存在하기 때문에 이 雜音 信號는 各 BPF出力에 나타난다. 즉 雜音信號가 周波數 帶域別로 BPF에서 에너지 形態로 存在한다. 이때 BPF의 오프셋 電壓이 零이 되도록 調節하였다. 여기서 認識 시스템의 機械的인 雜音이 가해졌을때 出力의 平均 에너지를 檢出하였다. 이 信號는 實際 音聲 信號보다 에너지가 작다는 假定下에 모든 채널을 同時에 追跡하여 信號 에너지가 처음으로 雜音의 平均値 以上되는 信號를 音聲이 始作되는 點으로 또 逆으로 追跡하여 그 以上이면 音聲이 끝나는 點으로 簡單히 始終點을 決定하였다. 10名의 話者가 5번씩 發音한 同一音 50個 대하여 最高 10個의 基準音을 設定하였으며 基準音을 除外한 나머지 單語를 入力 對象話로 使用하였다. 잘못 認識된 單語는 대전과 대구, 충주와 제주, 충무와 춘천 등이 었으며, 첫 音節뿐만 아니라 두번째 音節이 같은 境遇가 대부분 誤認識 되었다. 또한 춘천, 인천 등과 같이 破擦音이 수반된 單語는 集團化 結果 大部分 隔離音(outlier)이 存在하는 것으로 나타났으며 隔離音은 基準音 設定에서 除外시켰다. 대전이 대구로, 대구가 대전으로 잘못 認識된 것은 첫 音節이 같기 때문이며, 충주와 제주의 경우는 첫 音節이 確實히 다른데도 잘못 認識된 것은 두번째 音節이 같고 충주를 發音할때 /ㅊ/ 部分에서의 끝점 檢出의 問題가 아닌가 생각된다.

本 硏究에서는 距離 計算 方式에 따라 認識 結果가 어떻게 다른가를 考察하기 위하여 체비세프 距離式과 유크리드 距離式을 利用하였으며 誤認識結果는 各各 16.4, 15.0[%]를 얻었다.

유크리드 距離式은 체비세프 距離式보다 計算量이 많은 短點이 있으나 平均的으로 認識 結果가 더 좋은 것으로 나타났다.

表 1. 音聲 데이타의 集團化 結果

認識音聲	集團의 數	隔離音의 數	最大集團의 單語數
인 천	7	3	18
춘 천	7	3	11
경 주	9	1	18
제 주	9	1	13
대 구	8	2	17
대 전	10	0	11
부 산	10	0	8
마 산	9	1	11
충 무	9	1	11
충 주	9	1	15

表 2. 不特定 話者別 誤認識 結果

話者 \ 距離式	체비세브	유클리드
1	12.0	20.0
2	20.0	18.0
3	14.0	12.0
4	12.0	10.0
5	16.0	22.0
6	28.0	12.0
7	20.0	12.2
8	6.0	12.0
9	18.0	14.0
10	18.0	18.0
平 均	16.4	15.0

V. 결 론

本 硏究에서는 類似性이 많은 韓國語 10個 都市名을 對象으로 하여 認識하였으며 필터 뱅크 出力에너지를 特徵 파라미터로 使用하였다. 表 2에 나타낸 바와 같이 認識 結果가 多少 低調하다. 이것은 필터의 遮斷 特性의 問題가 큰 것으로 생각된다. 그러므로 필터 뱅크는 精巧하게 構成할 必要가 있으며 이를 얼마나 精巧하게 構成하느냐에 따라 認識結果가 달라질 수 있다. 특히, 필터 뱅크 시스템은 其他 다른 認識 시스템보다 데이타量 을 크게 줄일 수 있으며 모든 遂行 時間을 줄일 수 있기 때문에 필터의 特性이 理想型에 가깝도록 確固하게 構成된다면 더 좋은 認識 結果를 얻을 수 있다고 생각된다. 또한, 필터 뱅크 시스템은 複雜한 소프트웨어 處理를 必要로 하지 않고 방대한 데이타量을 處理하거나 瞬間的인 應答을 要하는 實時間 認識 次元에서 볼 때 반드시 硏究 되어야 한다고 생각된다.

參考文獻

1. 박현수 "[illegible] 인식에 관한 연구", 건국대학교 대학원 전자공학과 박사학위 논문 1989. 11.
2. 김계국, 고영덕, 이종악 "한국어 단독음 인식을 위한 표준 패턴 설정에 관한 연구" 한국 음향학회 논문집 Vol-6, No. 1, 1987. 3.
3. L.R.RABINER, S.E.LEVINSON, Isolated and Connected word Recognition Theory and Selected Application, IEEE Vol. Com-19. No5, May 1981.
4. L.R.RABINER, C.E.SCHIMIT Application of Dynamic Time Warping to Connected Digital Recognition, IEEE Vol.ASSP-28.No.4, August 1980.
5. G.M.WHITE, R.B.NEELY "Speech Recognition Experiments with Linear Predication, Bandpass Filtering, Dynamic Programming." IEEE Vol.ASSP-24.pp.183-188, Apr 1976.
6. S.E.LEVINSON, L.R.RABINER, A.E.ROSENBERG, J.G.WILPON Interactive Clustering Techniques for Selecting Speaker Independent Reference Templates for Isolated Word Recognition,IEEE Vol.ASSP-27, No. 2, Apr 1979.
7. B.A.DAUTRICH, L.R.RABINER, T.B.MATRIN "On the Effect of Varing Filter Bank Parameters of Isolated word Recognition," IEEE Vol.ASSP-31, No. 4, Aug 1983.
8. H. SKACE et, Dynamic programming Algorithm Optimization for Spoken Word Recognition,IEEE Vol-.ASSP-26, pp.43-49, Feb 1987.
9. MARTIN VETTERU, "A Theory of Multirate Filter Banks," IEEE Vol.ASSP-35, No. 3, Mar 1987.
10. M.J.T SMITH AND T.P.BARNWELL, "A New Filter Bank Theory for Time-Frequence Repeatation."IEEE Vol.ASSP-35, Mar 1987.
11. B.A.DAUTRICH, L.R.RABINER, and T.B.MARTIN "The Effects of Selected Signal Processing Tecnniques on the Performance of a Filter-Bank-Based Isolated Word Recognizer."The Bell System Technical Journal Vol.62, No.5, May 1983.
12. B.J.HOSTICXA, D.HERBST, B.HOEFFLINGER, U.KLEINE, J.PANDEL, AND R.SCHEWEER " Real-time Programmable Low Power SC Bandpass Single-Chip 20-Channel Speech Spectrum Analayzer Using a Multiplexed Swithed-Capacitor Filter Bank."

▲강 성 진(정회원)

1959년 12월 21일생

1982. 2 : 건국대학교 전자공학과 졸업

1986. 2 : 건국대학교 전자공학과 졸업(석사)

1991. 2 : 건국대학교 전자공학과 졸업(공학박사)

1991년 현재 : 대유전문대학전자통신과 전임강사

• 주관심분야 : 반도체신호처리

▲김계국 건국대학교 전자공학과(6권 1호 참고)

▲이종악 건국대학교 전자공학과(6권 1호 참고)