

線形豫測에 의한 숫자音聲의 自動認識 (A Spoken Korean-Digits Recognition System Based on Linear Prediction Spectra)

吳 永 煥* , 安 居 院 猛**
(Oh, Yung Hwan and Agui, Takeshi)

要 約

本 論文에서는 線形豫測스펙트럼을 주로 利用한 숫자音自動認識시스템과 그를 適用한 認識實驗結果에 關해 記述했다. 시스템은 파라메터抽出, 有聲 無聲音分析 및 分割, formant 追跡部와 認識部の 네 段階로 構成되어 있다. 無聲音과 有聲音의 接合部에서의 파라메터의 變化에 着目해 有聲·有聲音을 分割한 후, 有聲音 部分에 獨自의 追跡알고리즘을 適用해서 제 3 까지의 formant 周波數를 自動的으로 推定했다. 推定formant에 立脚해, DP matching手法에 의한 最終認識을 行했다.

成人男性 3 名의 150 개의 숫자음에 대한 認識實驗의 結果, 平均 94 %의 認識率을 얻어, 本 認識시스템의 有効性을 確認한 수 있었다.

Abstract

A speech recognition system for separately pronounced Korean digits is described. The system is composed of four stages ; parameter extraction, segmentation by voiced-unvoiced analysis, formant tracking and pattern matching. Digit speech is segmented into an unvoiced segment and/or a voiced one using ZCR and energy measurements, then to estimate the first three formant frequencies a relatively simple formant tracking scheme is applied to the raw formant data extracted from linear prediction spectra. Finally, pattern matching is made using dynamic programming method.

Recognition experiment is carried out for 150 digit utterances spoken by three male speakers, and recognition rate 94 % is obtained.

1. 序 論

筆者는 以前, 숫자音聲의 컴퓨터에 의한 自動認識의 可能性을 찾기 위한 基礎實驗의 結果에 關해 報告한 바 있다.^[1] 그러나 實驗對象으로 한 데이터數가 적어, 숫자音의 音聲學的 性質을 全般的으로 把握하기 어려웠으며, 따라서 그에 基礎를 둔 認識시스템의 適用範圍는 좁게 限定될 수 밖에 없었다. 또한, 音聲의 特性파라메터 (feature parameter)로 使用한 formant

周波數는, 一般的으로 發聲者의 性別, 나이, 發聲時期등에 따라 變化하며 그 存在範圍가 바뀐다. 특히, formant 周波數를 線形豫測(linear prediction)에 의해 求하는 境遇, 極의 脫落이나 追加등에 의해 正確한 周波數를 얻기가 어려우므로, 複數의 發聲者를 對象으로 한 認識시스템을 構成할 때는 以上의 諸問題點을 解決하지 않으면 안된다.

本 論文에서는 複數의 發聲者가 따로 떨어져 發聲한 單獨숫자음을 對象으로 構成한 認識시스템과 그를 利用한 認識實驗結果에 關해 記述한다. 國語의 單獨數字音은 다음과 같은 性質을 지니고 있다. 즉, (1) 모든 숫자음은 單音節語이며, (2)無聲音音(unvoiced consonant)은 머리部分에만 存在하며, (3)모든 숫자음은 有聲音音(voiced consonant)또는 母音, 즉 有聲音으로 끝난다.

* 正會員, 大田工業專門大學 電子科

(Daejeon Technical Junior College)

** 非會員, 東京工業大學

(Tokyo Institute of Technology, Japan)

接受日字: 1979年6月12日

以上の性質을 利用해서 有聲音部分을 中心으로 한 숫자음認識시스템을 다음과 같이 構成했다. 우선, 零交叉回數(Zero crossing rate)와 에너지를 利用해 말머리에서 有聲·無聲音의 判別 및 分割(segmentation)을 한 후, 線形豫測法에 의해 抽出한 有聲音部分의 formant周波數 데이터에, 比較的 簡單한 獨自의 formant 追跡 알고리즘(formant tracking algorithm)을 適用해, 第3까지의 formant 주파수를 自動적으로 推定했다. 推定된 formant 주파수에 DP(dynamic programming)手法를 利用한 matching에 의해 最終認識을 行했다.

한편, 本 研究의 意義는 다음과 같은 점에서 찾을 수 있다.

가) 숫자음의 音聲學的 性質이 어느 程度 明確히 돼, 認識시스템의 適用範圍의 擴張에 밝은 展望을 주고 있다.

나) 本 研究에서 作成해 使用한 自動formant 追跡 알고리즘의 有効性을 確認할 수 있었다.

다) 複數發聲者의 숫자음에 대해 좋은 認識結果를 얻어, 여기서 提案한 認識시스템의 有効性을 確認했다는 點등이다.

以下, 認識시스템의 詳細와 實驗結果에 關係 記述한다.

2. 認識시스템

숫자음認識시스템의 系統圖를 그림 1에 보이며, 以下 그 內容에 대해 記述한다.

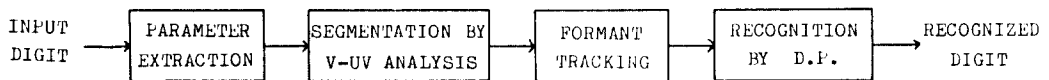


그림 1. 숫자음 인식시스템의 구성

Fig. 1. Block diagram of spoken digit recognizer.

2.1 파라메터抽出

入力音聲으로부터 抽出해 實驗에 使用한 特性파라메터는 다음의 세 種類다.

(1) 零交叉回數; 分析frame內에서 音聲波가 zero軸과 交叉하는 回數

(2) 에너지; 分析frame內의 音聲信號의 振幅의 제곱합

(3) formant 데이터; 線形豫測法으로 求한 音聲의 스펙트럼상의 頂點(以後, 極이라 부른다)의 周波數와 各極의 power 및 帶域幅이며, 本 시스템에서는 peak

-picking法에 의해서 抽出했다. 詳細한 抽出方法은 文獻(4)에 있으므로 여기서는 省略하기로 한다.

2.2 有聲·無聲音의 分割

2.1에서 求한 에너지와 零交叉回數를 利用해, 말머리에서의 有聲·無聲音分析(voiced-unvoiced analysis: V-UV analysis)을 해, 無聲音의 存在 如否와 無聲·有聲音이 結合하는 境界點을 抽出해서 最終적으로 分割을 行한다. 本 實驗에서는 有聲音區間內에서의 音素別分割은 하지 않는다. 無聲破裂音 뒤 에 有聲音이 오는 境遇, 그 境界의 前後에서 零交叉回數가 急激히 減少함과 同時에 에너지가 急增하는 點에 着眼해, 一定의 判斷條件에 의해 無聲破裂音의 有無 및 그 境界를 抽出한다.

여기서 i 번째 frame의 零交叉回數를 Z_i , 에너지를 E_i , 숫자음의 길이(持續時間)를 ℓ frame이라 할 때,

$$DZ_{i-1} = Z_i - Z_{i-1} \quad (i = 2, \dots, \ell) \dots (1)$$

$$PE_i = 100E_i / E_{max} \quad (i = 1, \dots, \ell)$$

$$E_{max} = \max_i [E_i] \quad (i = 1, \dots, \ell) \dots (2)$$

$$DE_{i-1} = PE_i - PE_{i-1} \quad (i = 2, \dots, \ell) \dots (3)$$

을 使用해,

$$S_i = DE_{i+1} - KDZ_i \quad (i = 1, \dots, \ell - 2) \dots (4)$$

를 計算한다. 여기서 K 는 conversion factor이며 本 實驗에서는 $K = 1$ 로 했다. S_i 는 無聲音音에서 有聲音으로 바뀔 때에 큰 값을 가지게 된다.

$$S_m = \max_i [S_i]$$

$$m \in i \quad i = 1, \dots, \ell - 2 \dots (5)$$

일때, m 번째 frame을 無聲·有聲音의 境界候補로 해, m 의 前半部가 無聲音區間인지를 다음의 條件에 의해 判定한다.

$$M_f = \frac{1}{m} \sum_{j=1}^m Z_j \dots (6)$$

$$M_b = \frac{1}{\ell - m - 1} \sum_{j=m+2}^{\ell} Z_j \dots (7)$$

라 할 때, 條件式(8)또는 (9)를 滿足하면 말머리에 無聲音音이 存在한다고 判定하며, 그 以外의 境遇는 有聲音音으로 構成된 숫자음으로 判定한다.

即, M_f 와 M_b 의 비가 threshold α_4 보다 큰 (差가 顯著한) 境遇(式(9))와, 비가 α_1 보다 작으나 α_2 以上이며, M_f 와 S_m 이 각각 α_1, α_3 以上인 境遇(式(8))에 m frame 의 前半部를 無聲子音區間으로 判定했다. 이와 같이 두 條件을 設定함으로써, 無聲子音等の 特性파라메터의 分散, 時間的 變動等에 對해서 安定된 segmentation을 할 수 있었다.

$$\left. \begin{aligned} M_f &\geq \alpha_1 \\ M_f &\geq \alpha_2 M_b \\ S_m &\geq \alpha_3 \end{aligned} \right\} \dots\dots (8)$$

$$M_f \geq \alpha_4 M_b \dots\dots\dots (9)$$

2.3 formant 追跡

2.2의 分析에 의해 有聲音으로 判定된 部分의 formant 데이터에 대해 tracking을 해, 第1부터 第3까지의 formant 周波數를 自動的으로 推定한다.

線形豫測法(LPC)에 의한 音聲의 推定 spectrum上에서 抽出하는 formant 情報은, 豫測次數나 音素間의 調音(coarticulation) 등에 의해 spurious pole이 餘分으로 追加되나든지, 隣接한 두 極이 하나로 merge되어 脫落極이 생기는 境遇가 있다. 그러므로 抽出한 formant 情報을 그대로 使用해 認識이나 合成등을 하면 error의 原因이 된다. 이 問題를 解決하기 위해, 音聲學的 知識을 利用해서 formant 주파수를 컴퓨터에 의해 自動的으로 追跡해 決定하는 tracking 手法은 이미 開發되어 몇몇의 報告가 나와 있다. 本論文에서는, 音素의 固有 formant 周波數를 精密하게 抽出하는 代身에 安定된 formant의 軌跡을 求解 認識 實驗에 使用할 目的으로, 比較的 簡單한 追跡 알고리즘을 獨自的으로 作成해 使用했으므로 以下 그 內容에 대해 記述한다.

本 追跡 알고리즘은 다음과 같은 前提條件下에서 成立한다.

1. 모든 音素(phoneme)는 3.5KHz 以內에 적어도 3개의 formant를 가진다.
2. 特殊한 境遇(예를 들면 母音과 鼻音의 結合時)를

除外하고, frame 間隔(12.8 ms)內에 各 formant가 變化할 수 있는 範圍는 最大 ± 500 Hz다.

3. 追加된 spurious pole은 正常極에 비해, power가 작고 帶域幅이 넓다.

4. 한 分析 frame內에서 生길 수 있는 spurious pole은 한 개 以下다.

이상의 條件을 前提로 作成한 追跡 알고리즘의 系統圖를 그림 2에, 順序圖를 그림 3에 보인다. 그림 3에서, 각각

N_j ; 第 j 番 frame 의, 周波數 3.5KHz 以下에 存在하는 極의 數

F_{ij} ; j frame 의 i 번째 極의 周波數(Hz)

P_{ij} ; F_{ij} 에 對應하는 極의 power (dB)

B_{ij} ; F_{ij} 에 對應하는 極의 帶域幅(Hz)

EF_{ij} ; 追跡 알고리즘에 의해 推定한 formant 周波數($i = 1, 2, 3$)다.

2.3.1 始點의 決定

本 시스템에서는 音聲信號의 始點·終點分析(end-point analysis)을 따로 행하지 않으므로 有聲音部의 첫번째 frame 의 formant 패턴이 正常이라고만은 할 수 없다. formant 追跡의 境遇, 現在의 frame의 formant가 正常인가를 前frame 과의 比較에 의해 判定하기 때문에, 始點의 formant를 正確하게 定할 必要가 있다.

本 實驗에서는, 최대에너지 frame (但, $N_j \geq 3$ 인 境遇로, $N_j < 3$ 인 때는 $j + 1$ 번 frame)의 formant를 基準으로 하여, 基準 formant 와의 差가 어느 값(여기서는 500 Hz) 以下인, 有聲音部分中 先頭에 가장 가까운 frame을 始點으로 해 追跡을 開始한다. 한편, N_j 가 4인 境遇는 그림 3의 $N_j = 4$ 의 處理를 해 spurious pole의 有無를 check한다. 위의 條件에 合致하는 frame이 一定 frame 以上 存在하지 않을 때와, 音聲의 길이가 一定 frame 以下の 짧은 音聲일 때에는, 先頭로부터 安定된 formant 패턴이 두 frame 以上 繼續되는 곳에서 追跡을 開始한

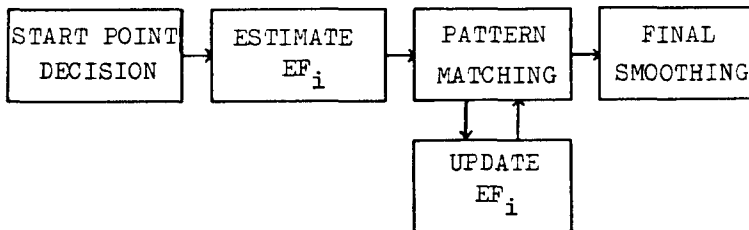


그림 2. Formant 追跡 알고리즘의 구성

Fig. 2. Block diagram of formant tracking.

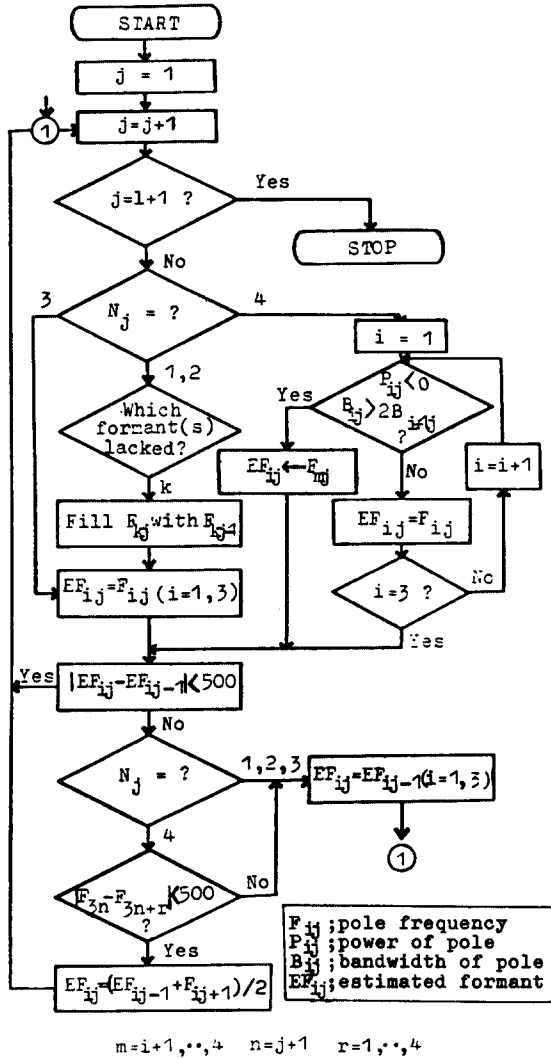


그림 3. Formant 추적의 順序圖
Fig. 3. Flow chart of formant tracking.

다. 이 條件은 最大에너지 frame이 先頭frame과 다른 音素일 境遇에 對處하기 위해 設定했다.

2. 3. 2 脫落極의 對策

極이 脫落한 境遇. 各 formant 別로 前frame 과의 pattern matching을 해, 距離 D_{kl} 의 합이 最小가 되도록 脫落極을 前 frame의 該當周波數로 채운다. formant 間 距離를

$$D_{kl} = |EF_{k,j-1} - F_{l,j}| \quad (\ell = 1, \dots, N_j) \quad (10)$$

로 定義했을 때,

(가) $N_j = 1$ 인 境遇

$$D_{m1} = \min_k [D_{k1}] \quad \dots \dots \dots (11)$$

$$m \in k, \quad k = 1, 2, 3$$

$$\left. \begin{aligned} EF_{mj} &= F_{1j} \\ EF_{nj} &= EF_{n,j-1} \end{aligned} \right\} \dots \dots (12)$$

$$n \neq m, \quad n \in k$$

(나) $N_j = 2$ 인 境遇

Step 1. $D_{11} < D_{21}$ 이면 $EF_{1j} = F_{1j}$ 로 한 뒤 Step 3에. 그 以外는 Step 2에.

Step 2. $EF_{1j} = EF_{1,j-1}, EF_{k+1j} = F_{k1} (k = 1, 2)$ 로 한 뒤 Step 5에.

Step 3. $D_{22} < D_{32}$ 이면 $EF_{2j} = F_{2j}, EF_{3j} = EF_{3,j-1}$ 로 한 뒤 Step 5에. 그 以外는 Step 4에.

Step 4. $EF_{2j} = EF_{2,j-1}, EF_{3j} = F_{21}$ 로 한 뒤 Step 5에.

Step 5. 끝. 다음의 處理에.

2. 3. 3 追加極의 對策

주로 母音의 鼻音化(nasalization) 現象에 의해 나타나는 spurious pole은 一般的으로 power가 작고 帶域幅이 큰 性質이 있으므로, 이를 利用해 顯著한 追加極은 除去할 수 있다. 本 實驗에서는 power가 0 dB以下로, 帶域幅이 다음 formant의 帶域幅의 2倍 以上 넓을 때 spurious pole로 보아 除去한다.

2. 3. 4 큰 變化에의 對應

上述한 前提條件 2는 大部分의 有聲音에 有効하다. 그러나, 母音과 鼻音의 接合部에서는 短時間內에 대단히 큰 變化를 觀察할 수 있다. 特別比較的 낮은 frame rate에 의한 分析에서는 이 傾向이 세계 나타나, 條件의 限界를 넘는 境遇도 자주 일어난다. 本 시스템에서는, 鼻音은 3.5 KHz 以內에 대개 4개의 極을 지니며, 相當히 安定된 formant 패턴을 보이는 點에 着眼해서, 다음과 같은 條件을 設定했다. 즉, 그림 3의 $N_j = 4$ 의 處理를 해도 matching 條件을 滿足하지 않는 境遇에 限해, 뒤의 一定frame을 調査해 formant 패턴일 때, 條件 2를 넘는 큰 變化를 認定함으로써 좋은 結果를 얻고 있다. 한편, 이때의 推定 formant 周波數는 前後frame의 平均値를 취했다.

2. 3. 5 Smoothing

追跡 알고리즘에 의한 formant의 推定이 끝나면, 最終的으로 式 (13)^[1]의 smoothing을 해, 부드러운 formant 軌跡을 얻는다.

$$EF_{ij} \leftarrow \frac{1}{4}EF_{i,j-1} + \frac{1}{2}EF_{ij} + \frac{1}{4}EF_{i,j+1} \quad \dots \dots (13)$$

2. 3. 6 他追跡手法과의 比較

線形豫測法에 의해 구한 spectrum上에서 peak - pickinge에 의해 抽出한 極情報로부터 제 3 formant까지 推定하는 代表的인 手法으로 문헌(3)과 (4)를 들 수 있다. 脫落極을 찾기 위한 反復處理³⁾를 하지 않는 比較的 簡單한 手法인 문헌(4)와 比較해, 本 手法은 다음과 같은 特徵이 있다.

1. 本 시스템에서는 比較的 낮은 frame rate와 豫測次數로 分析했다 (frame interval 12.8ms, 豫測次數 12). frame rate가 높으면 formant追跡의 精度가 높아지며, 豫測次數가 어느 程度 높으면 脫落極이 줄어 追跡error가 적어진다. 그러나, 認識을 目的으로 하는 境遇, 高精度로 formant周波數를 推定하기 보다는 短時間에 安定한 軌跡를 求하는 것이 有利하다. 文獻(3)에서는 各各 5 ms, 14次이며 (4)에서는 6.3 ms, 15次로 本 시스템에 비해 2倍以上의 frame rate로 分析하고 있다.

2. 本 시스템에서는 formant의 存在範圍를 넓게 잡고 있으므로 (3.5 KHz), 單獨으로 發聲된 제 3formant가 높은 音素(예를 들면 /이/)도 充分히 處理할 수 있다.

3. 本 시스템에서는 一定條件下에서 큰 變化를 許諾함으로써, 母音 - 鼻音의 結果等에서 볼 수 있는 急激한 formant의 變動도 追跡할 수 있다.

4. 帶域幅과 power를 考慮함으로써 spurious pole을 効果적으로 除去할 수 있다.

5. 追跡에 있어서 始點의 初期值로 固定值를 使用하지 않고, 에너지 最大frame의 formant周波數를 利用함으로써 始點의 不正確한 決定에 의한 追跡error를 줄일 수 있다.

2. 4 認識

以上の 알고리즘에 의해 말머리에서의 有聲·無聲音 分析과, 有聲音에 對해 formant 追跡알고리즘을 適用해 推定한 formant周波數를 利用해 最終的인 認識을 行한다. 本 實驗에서는 持續時間이 서로 다른 숫자音間의 時間軸을 正規化(normalization)하기 위해 DP matching⁽⁵⁾⁽⁶⁾ 手法을 使用했다. 또한 無聲音音間의 識別은 하지 않으며, 모두 formant周波數가 零인 同一音素로 看做했다. DP matching의 動作을 그림 4에 보인다. 一般的으로 널리 쓰이는 linear matching은 그림 4의 ②의 線上에서 mathing을 行함으로 두 pattern의 長이가 같아야 하며, 도중의 少數pattern의 不一致가 最終結果에 크게 影響을 미

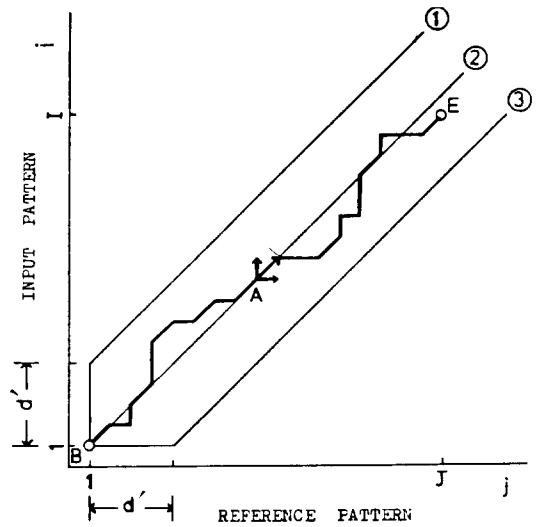


그림 4. DP에 의한 時間軸의 正規化

Fig. 4. Time-axis normalization by dynamic programming method.

친다. DP는 注目點의 前後 d' 의 範圍內에서 (①과 ③사이), 出發點B와 終點E와의 距離를 最小(또는 類似度를 最大)로 하는 matching path를 求하는 手法이다. 한편, 注目點A에서 취할 수 있는 path는 좌표로 表示한 水平, 垂直, 對角線의 세 方向에 限定된다. 實際認識은, 初期條件 (15)下에서 漸化式(recursive equation) (14)를 利用해, 두 패턴間의 類似度の 尺度로서 (16)式의 S를 利用해서 行했다(문헌(6)). 즉, 類似度の 尺度(距離) S를 最小로 하는 숫자音을 認識숫자音으로 했다.

$$g_{ij} = \min \begin{bmatrix} d_{ij} + g_{i-1} \\ d_{ij} + g_{i, j-1} \\ 2d_{ij} + g_{i-1, j-1} \end{bmatrix} \dots \dots (14)$$

$$i = 1, \dots, I \quad j = 1, \dots, J$$

$$g_{i1} = d_{i1} \dots \dots (15)$$

$$S = \frac{1}{I + J - 1} g_{IJ} \dots \dots (16)$$

I는 入力패턴의, J는 標準패턴의 長이 (frame數)로서 $|I - J| \leq d'$ 의 條件을 滿足한다. 以上の 條件下에서 (14)式을 連續的으로 풀어, 最終的으로 두 패턴間의 最小距離(最大類似度) S를 (16)式에 의해 얻는다. 한편, (14)式에서, $g_{ij} = \min [\dots]$ 는 g_{ij} 로서 괄호안의 세 式중의 最小直를 취하는 것을 뜻하며, 두 frame i, j 間의 距離 d_{ij} 는

$$d_{ij} = \sum_{k=1}^3 |EF_{ki} - LF_{kj}^{(t)}| \dots\dots\dots (17)$$

로 定義했다. 여기서 EF_{ki} 는 追跡 알고리즘에 의해 推定한 入力숫자음의 i frame의 formant 周波數이며, $LF_{kj}^{(t)}$ 는 숫자음 t 의 標準패턴의 j 번째 frame의 formant 周波數다.

3. 實驗

以上的 認識시스템을 利用해서 實際 숫자음성에 關해 認識實驗한 結果에 대해 以下 記述한다.

3. 1 實驗데이터

實驗에 使用한 데이터는 成人男性 3인이 5회씩 反復해서 發聲한 總 150個의 숫자음이다. 조용한 방에서 tape recorder에 錄音한 音聲을 cut-off frequency 5KHz, -12 dB/oct의 low-pass filter에 通過시킨 후, sampling 周波數 10KHz, 12bit로 A-D變換했다. Formant 情報抽出時는 25.6ms의 Hamming window를 進 후에 分析했다. 分析frame의 長이는 256點 (25.6ms)이며, frame 間隔은 128點 (12.8ms) 豫測次數는 12다.

3. 2 實驗結果

3. 2. 1 分割

2. 2의 手法에 의한 말머리에서의 有聲·無聲音分

析에 使用한 常數는 $\alpha_1 = 100$, $\alpha_2 = 1.3$, $\alpha_3 = 60$, $\alpha_4 = 1.9$ 다. 常數는 認識實驗과 同一發聲者의 데이터를 使用한 豫備實驗에 의해 定했으며, 全데이터에 대해 有聲·無聲音의 判定 및 接合部의 抽出에 error가 없었다. 여기서 使用한 無聲音과 有聲音의 境界에서의 파라미터의 變化에 注目하는 分割手法는 短時間內에 有聲音과 無聲音이 交代로 바뀌지 않는 音聲데이터에 대해서, 有效한 一手法이라 볼 수 있는 것이다.

3. 2. 2 formant追跡

여기서 提案한 formant追跡 알고리즘의 適用結果의 例를 그림 5에 보인다. 그림은 CRT display 表示된 結果이며, 왼쪽이 原formant情報이며, 오른쪽이 自動追跡에 의해 推定된 formant 周波數의 軌跡이다.

그림 5-(a)의 境遇, 單獨母音 /i/의 제3 formant가 相當히 높고, spurious pole이 많이 나타났으나 好結果를 얻었다. 그림 5-(b)의 境遇는, 母音과 鼻音의 結合에 의해 接合部에서의 極의 變化가 빠르며, 隣接極의 融合에 의한 極의 脫落이 보이나, 부드러운 formant의 軌跡을 얻었다. 그림 5-(e)의 境遇는, 二重母音 /여/와 鼻音 /ㅇ/이 結合해 spurious pole, 脫落極 및 빠른 變化등 複雜한 움직임을 보이고 있으나 效果的으로 極의 推定이 行해졌다.

本 알고리즘에 의한 追跡에 失敗한 例를 그림 6에 보인다. 그림 6-(a)와 (b)는 始點決定의 基準이 되는 最大에너지 frame의 formant 패턴의 不備(極의

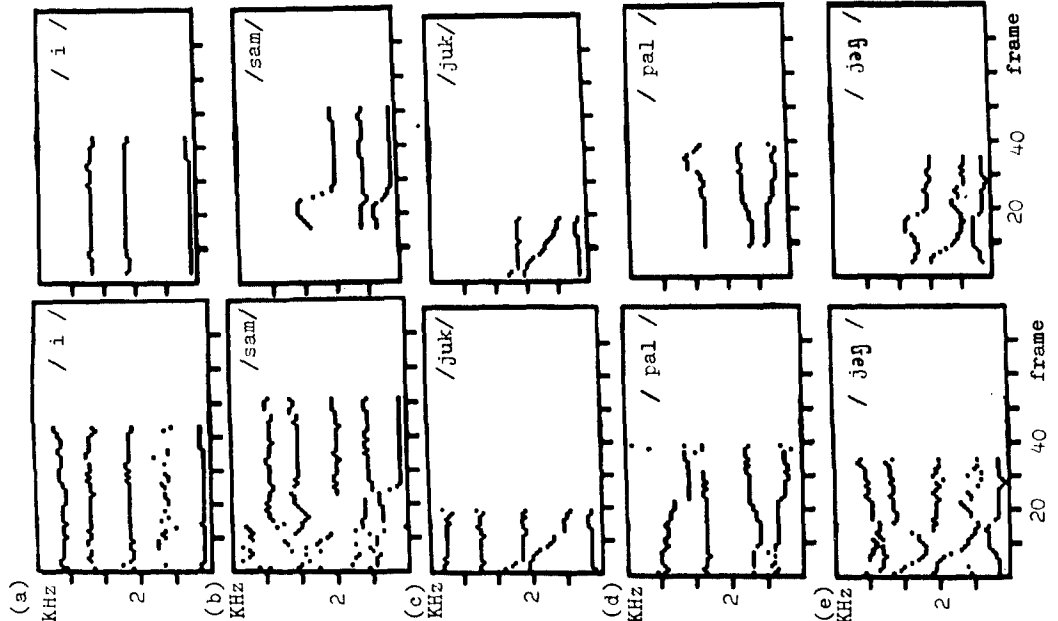


그림 5. 原formant 데이터 (왼쪽)와 追跡 알고리즘에 의해 推定한 formant 軌跡 (오른쪽)

Fig. 5. Raw formant trajectories (left-hand side) and estimated ones by tracking (right-hand side).

線形豫測에 의한 숫자音聲의 自動認識

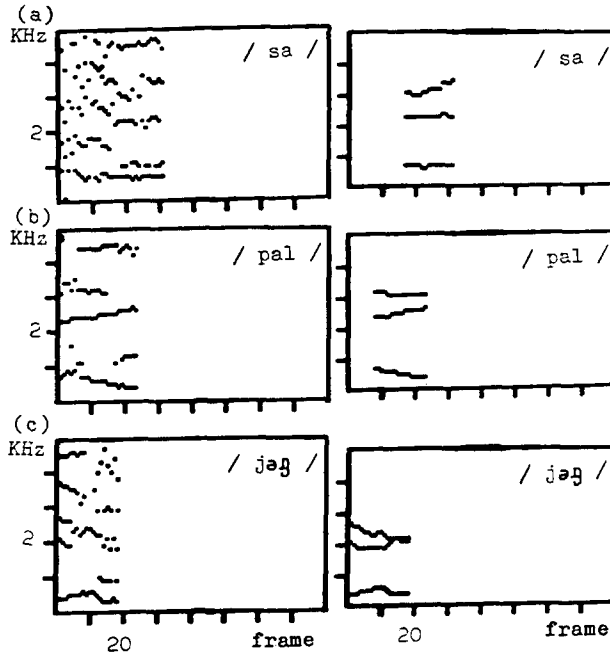


그림 6. Formant 追跡에 失敗한 例

Fig. 6. Examples of failures in tracking.

脫落)에 失敗의 原因이 있으며, 그림 6-(c)의 境遇는 여러 frame에 걸쳐 極의 脫落이 繼續되고, 그 사이에 formant의 變化가 커서 제 2 formant의 追跡이 不可能한 例다.

3. 2. 3 認識結果 및 檢討

표 1. 인식결과

Table 1. Confusion matrix.

OUT \ IN	1	2	3	4	5	6	7	8	9	0
1	15									
2		15								
3			14				1			
4				14			1			
5		2			13					
6					1	14				
7							15			
8							2	13		
9									15	
0	1	1								13

認識實驗에 의해 最終적으로 얻어진 認識結果를 표 1에, 發聲者別 認識率等을 표 2에 보인다. 여기서 標

표 2. 發聲者別 認識結果

Table 2. Correct rate for each speaker.

發聲者	평균 길이	표준패턴수	認識率 %
# 1	約 0.5 秒	10 개 (# 1)	100 (50/50)
# 2	0.3	10 개 (# 2)	88 (44/50)
# 3	0.3	10 개 (# 2)	94 (47/50)
全體	0.37	20 개	94 (141/150)

準패턴으로는, 이상의 formant 追跡 알고리즘에 의해 빠르게 推定된 認識實驗 데이터 중의 任意的 10개를 골라 使用했다. 發聲者 3名의 平均認識率은 94%이며, 本 論文에서 提案한 追跡 알고리즘과 그를 採用한 숫자음認識시스템이 有效함을 確認할 수 있었다. 本 시스템의 誤認識의 最大의 原因은 脫落極에 있으며, 그 影響을 줄이기 위해서는 frame rate나 豫測次數를 높일 必要가 있으나, 分析時間이 길어지는 缺點이 있다. 또한, 極의 脫落이 大部分 제 2 軌道에서 일어나 音素 /이/로 誤認識되는 性質을 利用해, 判 判定基準을 導入한다면 本 시스템으로 더욱 높은 認識率을 期待할 수 있을 것이다. 한편, 本 實驗에서는 發聲者에 따라 認識率에 큰 差가 보여, 特히 發聲者 1은 全 데이터에 대해 正解가 얻어졌다. 그 理由로서는 他發聲者에 비해 發聲速度가 느린(持續時間이 긴) 點

을 들 수 있다. 즉, 持續時間이 길어지면 分析 frame rate를 높힌 것과 同等의 效果가 나타나, formant 追跡의 精度가 높아졌기 때문에 볼 수 있을 것이다

4. 結 論

本 論文에서는 主로 formant 周波數에 依存한 國語 숫자음認識시스템과 그를 適用한 認識實驗 結果에 對해 記述했다. 말머리에서의 有聲·無聲音分析에 의한 分割(segmentation)을 한 후, 有聲音의 formant 情報에 比較的 簡單한 formant 追跡 알고리즘을 適用해서 推定한 제 3까지의 formant 周波數를 使用해, DP matching 手法에 의해 最終認識을 行했다. 成人男性 3명이 發聲한 150개의 숫자음에 對해 認識實驗한 結果 平均 94%의 認識率을 얻었다. 本 實驗에 의해, 無聲·有聲音의 接合部에서의 파라메터의 變化에 注目한 分割手法, formant 推定을 위한 追跡 알고리즘 및 그를 利用한 숫자음認識시스템의 有効성을 確認할 수 있었다. 한편, 追跡 알고리즘에 있어서의 보다 正確한 始點의 決定方法 및 脫落極이 繼續되는 境遇의 處理方法등이 남겨진 問題點이다.

앞으로, 보다 많은 데이터를 使用한 보다 効果的인 國語의 單語認識시스템의 開發을 目標로 한 研究를 해 나갈 豫定이다.

參 考 文 獻

1. 吳, 安居院: "숫자音聲自動認識에 關한 一實驗", 大韓電子工學會誌, 제 15권 제 6호, 1978년 12월.
2. 安居院, 吳; "韓國語數字音의 認識시스템에 對해서", 日本電子通信學會技術研究報告 PRL 79-1~11, 1979년 (日本語).
3. McCandless, S. S.: "An algorithm for automatic formant extraction using linear prediction spectra", IEEE Trans. ASSP, Vol. ASSP-22, No. 2, Apr. 1974.
4. Markel, J. D. and Gray, Jr., A. H.: Linear prediction of speech, Springer-Verlag, N.Y., 1976.
5. Velichko, V. M. and Zagoruyko, N. G.: "Automatic recognition of 200 words", Int. J. Man-Machine Studies, Vol. 2, pp. 223-234, 1970.
6. Sakoe, H. and Chiba, S.: "Recognition of continuously spoken words based on time-normalization by dynamic programming", J. Acous. Socie. of Japan, Vol. 27, No. 9, 1971. (In Japanese)