論文
2-10-1

한국어 숫자음을 이용한 화자식별
(Speaker Identification Using Korean Digits)

정 의 붕*
(Eui-Bung Jeong)

요 약

본 연구에서는 한국어의 숫자음을 대상 어휘로 이용하여 화자 식별을 하였다. 화자 식별을 위해서는 음성파에 포함된 여러 가지 정보로부터 화자 개개인이 갖는 개인적 특정 정보를 추출하여 화자 식별을 위한 특징 파라메타로 이용한다. 실험에서 사용한 특징 파라메타로는 음성파의 피치주기 사이에 존재하는 멀티펄스가 화자에 의존하는 개인의 특정 정보를 포함하고 있음에 착안하여 개개인의 음성파로부터 멀티펄스를 추출하고 추출된 멀티펄스로부터 화자별 특징 파라메타를 탐색하여 화자 식별 실험을 하였다.

ABSTRACT

In this paper, we have identified speakers who give digits in Korean. In order to identify speakers, we have utilized the specifie feature parameters which extracted from sound wave. We have noticed that multipulses are present in pitch periods of sound wave, which containes the personal information and depends on the speakers. In this experiment, we have extracted multipulses, and have attempted to identify the speaker by investigating the specific feature parameters of each speaker based on the extracted multipulses.

1. 서론

우리가 입으로 말하는 음성에는 많은 정보가 포함되어있다. 그중에서 음성파에 포함되어 있는 音韻性 情報를 抽出하여 認識하는 것을 音聲 認識이라 하고, 個人性 情報를 추출하여 화자를 추정하고 認識하는 것을 話者 認識이라 한다[1]. 화자 인식은 話者 照合과 話者 識別의 두 가지로 分類된다[2].

話者 認識에 사용되는 특징 파라메타로는 음성인식에서와 같이 음성파로 부터 스펙트럼 또는 그것을 표현하는 특정 파라메타 형태로 특징을 추출하여 인식 시 사용한다. 따라서 화자 사이에서는 分散의 변동이 크고 話者內에서는 변동이 적은 특징을 사용하여야 하며 또 최소의 계산으로 最大의 認識 結果를 얻을 수 있도록 선택하여야 한다.

특징 파라메타로는 에너지, 零交差率, 피치 주기 및 최대 진폭점의 위치등과 같이 음성파에 포함된 時間的 특징을 그대로 이용하는 방법과 線形 豫測 係數, 켑스트럼 係數나 PARCOR 係數 및 短時間 스펙트럼 등과 같이 시간적인 특징을 統計的인 특징으로 변형시켜 이용하는 방법이 있다. 대부분의 화자인식에서는 계산의 간략을 위하여 12~14차의 선형 예측 계수나 17~20차 대역 통과 필터 뱅크 에너지를 특징 파라메타로 주로 사용한다[3][4].

* 정회원 : 호원대학교 전기전자정보공학부 부교수

논문접수 : 2001. 9. 24.
심사완료 : 2001. 10. 10.

※ 본 논문은 2001년 호원대학교 교내학술연구조성비에 의해 연구되었음.

본 연구에서는 피치 주기 사이에 존재하는 멀티 펄스(multipulse) 列들이 話者에 依存하는 個人性 情報를 포함하고 있음[8][9]을 이용하여 음성 신호로부터 화자 식별을 위한 특징 파라메타로 음성신호에서 멀티펄스 열을 추출하고, 추출된 멀티펄스 열로부터 개인성 특징을 구하고 이를 사용하여 話者 識別 實驗을 수행하였다.

2. Multipulse열에 의한 個人性 情報의 抽出

B. S. Atal등은 音聲의 合成시 구동 음원 신호로 임펄스 열이나 백색 잡음 신호 대신 피치 주기 사이에 존재하는 복수개의 펄스(multipulse)를 이용하므로 원 음성에 가까운 합성음을 얻을 수 있다고 보고하였다[5]. 또한 J. N. Holmes에 의하면 음원 펄스로는 聲帶가 닫길 때 생기는 주(main) 펄스 이외에도 성대가 열릴 때나 열려 있는 동안에도 2차적인 구동 파형이 존재하고 있다고 한다[10]. 이 사실로 類推해 보면 有聲音의 驅動 音源을 모델화하는데 있어서 피치 주기에 따른 임펄스 열 이외에도 피치 주기의 주위에 있는 여러 개 펄스로 驅動 音源을 모델화 할 수 있을 것으로 생각된다.

멀티펄스를 구동 음원으로 사용할 경우 펄스의 위치나 크기는 합성된 음성 신호가 가능한 원 음성 신호에 近似되도록 順次적으로 얻어져야 한다. 다시 말하면 합성 음성 신호과 원 음성 신호 사이에서 얻어지는 誤差 信號의 총 에너지가 최소가 되도록 하여야 한다.

따라서 펄스의 位置나 크기를 決定하는 알고리즘으로는 하나의 펄스가 얻어지면 얻어진 펄스의 影響을 오차 신호에서 제거한 후 다음 펄스를 탐색하는 과정을 되풀이 한다[11].

2.1 Multipulse 列의 探索

원 음성신호를 S_n이라 하고 합성된 음성 신호를 $\tilde{S}_n$이라 하면 誤差 信號는 식(1)과 같다.

$$e(n)=S_n-\tilde{S}_n \tag{1}$$

음성 신호의 에너지가 포르만트 주파수 주변에 집중되어 있으므로 오차 신호도 포르만트 주파수 주변에 비교적 큰 오차 에너지를 갖고 있다.

Atal 등에 의해 提案된 멀티펄스법은 포르만트 주파수 지역에서 오차 신호 에너지를 억제시키기 위하여 오차 신호에 가중치 함수를 넣고, 이를 z-변환한 것이 식(2)이다.

$$E(z)=[S(z)-\tilde{S}(z)]W(z) \tag{2}$$

여기서 가중치 함수 $W(z)$는 식(3)과 같이 선택한다.

$$W(z)=\frac{H(\gamma^{-1}z)}{H(z)} \tag{3}$$

$H(z)$는 線形豫測 聲道 모델의 傳達 函數이므로 $W(z)$는 다음과 같이 나타낼 수 있다

$$W(z)=\frac{H(\gamma^{-1}z)}{H(z)}=\frac{1-\sum_{k=1}^{p}a_k z^{-k}}{1-\sum_{k=1}^{p}a_k\gamma^k z^{-k}} \tag{4}$$

식(4)를 noise weighting 필터라 하고 a_k는 p차 선형예측 계수이고 γ는 $0\le\gamma\le1$ 範圍를 갖는 정수로 음성 스펙트럼의 最大값을 抑制 시키는 계수이다. γ=0인 경우 억제 效果가 가장 크고 γ=1인 경우는 억제 효과가 없다.

식(3)을 식(2)에 대입하면 다음과 같다.

$$\begin{aligned}E(z)&=[S(z)-\tilde{S}(z)]\frac{H(\gamma^{-1}z)}{H(z)}\\&=\left(\frac{S(z)}{H(z)}-\frac{\tilde{S}(z)}{H(z)}\right)H(\gamma^{-1}z)\end{aligned} \tag{5}$$

그러므로 식(5)의 前項은 선형 예측 모델의 誤差 信號이고 後項은 합성음의 생성을 위한 펄스열 구동 신호이다. 그러므로 여기서는 펄스 열의 驅動 信號를 탐색해야 한다. 探索해야 할 펄스를 위치 m인 곳에 진폭의 크기가 A인 펄스라 하고 A_m으로 표시한다.

$H(\gamma^{-1}z)$의 임펄스 應答을 $h_w(n)$이라 하고 이를 식(2)에 대입하여 總誤差 에너지를 구하면 總誤差 에너지는 식(6)과 같다.

$$\sum_n e^2(n)=\sum_n[d(n)-A_m h_w(n-m)]^2 \tag{6}$$

이 식에서 신호 $d(n)$는 전달 함수 $H(\gamma^{-1}z)$를 통과한 선형 예측 오차 신호이고 $A_m h_w(n-m)$은 전달 함수 $H(\gamma^{-1}z)$를 통과한 임펄스이다.

탐색된 임펄스 A_m에 의해 합성된 음성에 대한 總 誤差 에너지가 最小가 되기 위해서는 식(6)을 A_m에 대해 微分하여 그 값이 0(零)가 되어야 하므로 이를 미분하여 0(零)으로 하고 A_m에 대해 整理하면 식(7)과 같다.

$$A_m = \frac{\sum_n d(n)h_w(n-m)}{\sum_n h^2{}_w(n-m)} = \frac{C_m}{\Phi} \tag{7}$$

$$C_m = \sum_n d(n)h_w(n-m)$$

: $d(n)$과 $h_w(n-m)$의 相互 相關

$$\Phi = \sum_n h^2{}_w(n-m) \text{ : } h_w(n)\text{의 에너지}$$

따라서 식(7)을 식(6)에 대입하면 總 誤差 에너지는 식(8)과 같이 얻어진다.

$$E = \sum_n e^2(n) = \left(\sum_n d^2(n)\right) - \left(\frac{C_m^2}{\Phi}\right) \tag{8}$$

식(8)에서 펄스의 위치 m과 관련된 값은 $d(n)$과 $h_w(n-m)$의 相互 相關인 C_m 뿐이므로 總誤差 에너지를 최소로 하는 임펄스는 $d(n)$과 $h_w(n-m)$의 相互 相關이 最大인 지점에 놓이게 되고 이때 진폭은 식(7)에서 얻어진다.

2.2 特徵 벡타(vector)의 抽出

앞에서 구한 펄스들의 위치와 크기만으로 화자의 개개인의 특징을 나타내는 특징 벡타로 取扱하기에는 어려움이 있으므로 펄스들의 位置와 振幅을 2차원 空間에 配置하고 배치된 펄스들의 分布에서 話者別 特徵 벡타를 구한다.

이를 위해 펄스들의 위치와 진폭을 피치주기의 兩側에 있는 피치 펄스로 正規化하여 2차원 空間에 配置한다. 음성 신호로 부터 피치 週期를 구하고 얻어진 펄스열로 부터 피치 주기 마다 最大의 振幅을 갖는 피치 펄스를 探索한다. 다음 隣接하는 피치 펄스로 멀티펄스 열을 피치 週期 마다 끊어 낸다.

s번째 끊어 낸 주기를 s-피치 주기라 하고 s-피치 주기에서 i번째 발생한 펄스의 時刻을 $t_{s,i}$, 진폭을 $a_{s,i}$라 하고, s, s+1번째 피치 펄스가 發生한 時刻을 T_s, T_{s+1}이라 하면 피치 주기 사이에서 발생한 각 펄스를 피치 펄스가 발생한 時刻을 基準으로 하여 正規化한다.

각 펄스의 정규화한 位置를 $x_{s,i}$라 하고 다음과 같이 決定한다.

$$\begin{aligned} t_{s,i} &= T_s \text{ 인 경우} \quad x_{s,i} = -1 \\ t_{s,j} &= T_{s+1} \text{ 인 경우} \quad x_{s,i} = 1 \\ x_{s,i} &= \frac{1-(-1)}{T_{s+1}-T_s}(t_{s,i}-T_s)+(-1) \\ &= \frac{2t_{s,i}-T_s-T_{s+1}}{T_{s+1}-T_s} \end{aligned} \tag{9}$$

또 隣接하는 두 피치 펄스의 振幅을 A_s, A_{s+1}라 할 때 진폭의 크기가 서로 다르기 때문에 이를 고려하여 인접한 피치 주기의 振幅 差인 기울기 $(A_{s+1}-A_s)/(T_{s+1}-T_s)$로 각 펄스의 진폭 $a_{s,i}$를 正規化 하고, 정규화 된 振幅을 $y_{s,i}$라 하면 $y_{s,i}$는 다음 식과 같다.

$$y_{s,i} = \frac{a_{s,i}}{\frac{A_{s+1}-A_s}{T_{s+1}-T_s}(t_{s,i}-T_s)+A_s} = \frac{a_{s,i}}{A_{s+1}\frac{t_{s,i}-T_s}{T_{s+1}-T_s}+A_s\frac{T_{s+1}-t_{s,i}}{T_{s+1}-T_s}} = \frac{2a_{s,i}}{A_{s+1}(1+x_{s,i})+A_s(1-x_{s,i})} \quad (10)$$

正規化된 펄스들의 座標를 音源 特徵이라 부른다. 또 어떤 分析 프레임으로 부터 얻어진 m개의 펄스들의 座標를 음원 特徵 벡타群 P_E라 부르고 2차원 空間上의 m개의 벡타 $\overrightarrow{E_j}$의 集合體로 한다.

$$P_E=(\overrightarrow{E_1},\ \overrightarrow{E_2},\ \cdots,\ \overrightarrow{E_j},\cdots,\ \overrightarrow{E_m}) \quad (11)$$

$\overrightarrow{E_j}$는 j번째 펄스의 座標이다.

2.3 特徵 벡타의 集團化

정규화된 펄스들의 분포는 몇 개의 묶음으로 분류되어 集團을 形成하고 있음을 알 수 있다.

따라서 화자가 마다 형성된 集團의 분포가 다르기 때문에 이것을 話者 識別시의 특징 벡타로 사용한다. 얻어진 집단의 분포를 개인성을 나타내는 변수로 취급하기 위해 집단을 이루고 있는 집단의 中心 좌표를 구하고 이들 특징 벡타로 하여 등록화자의 표준 패턴과 식별하고자 하는 미지화자의 패턴을 만든다. 집단화를 위한 알고리즘으로는 최대최소 거리, A simple Cluster -Seeking, K-means, ISODATA등 여러 가지 방법[12][13]들이 알려져 있지만 방대한 資料의 처리를 위해서는 가능한 처리 시간이 짧은 알고리즘을 사용하여야 한다. 따라서 본 연구에서는 K-means 알고리즘을 바탕으로 사용하였으며 그 過程은 다음과 같다.

집단화 알고리즘

과정 1 : 음성 구간(T)에서 추출된 피치 주기의 개수를 N으로 한다.

과정 2 : i번째 피치 주기에서 얻은 펄스의 수를 J개로 하고 P_{i1}, P_{i2}, ···, P_{ij}로 나타낸다.

과정 3 : 집단의 수를 펄스 수와 같은 J개로 하고, 각 집단의 집합을 S_1, S_2, ···, S_J라 한다. 또 각 집단의 초기 중심값 $\overline{P}_1$, $\overline{P}_2$,···, $\overline{P}_J$을 다음과 같이 결정한다.

$$\overline{P}_j=\frac{1}{N}\sum_{i=1}^{N}P_{ij}\ (j=1,2,\cdots,J) \quad (12)$$

과정 4 : 펄스 P_{i1}, P_{i2}, ···, P_{ij}의 각 좌표와 초기 중심값 $\overline{P}_1$, $\overline{P}_2$,···, $\overline{P}_J$의 각 좌표 사이에서 거리를 계산하고 다음과 같은 행열을 정의 한다. A_{ij}는 i번째 피치에 대한 행열임을 나타낸다.

$$A_{ij}=\begin{bmatrix} d_{i11} & d_{i12} & \cdots & d_{i1J} \\ d_{i21} & d_{i22} & \cdots & d_{i2J} \\ \vdots & \vdots & \cdots & \vdots \\ d_{ij1} & d_{ij2} & \cdots & d_{ijJ} \end{bmatrix} \quad (13)$$

단, $d_{1kl}=\|P_{1k}-\overline{P}_l\|$ 이다.

과정 5 : 미리 臨界값 ε_1을 정해 놓고, 행열 A_{ij}의 각 요소 가운데서 가장 작은 요소의 값 d_{1kj}가 다음 식을 만족하면 $P_{1k}\in S_j$로 한다.

$$d_{1kj} < \varepsilon_1 \quad (14)$$

다음 A_{ij}에서 행 k와 열 l를 제외한 행열 A_{i-1j-1} 를 구한 다음 그 요소의 값 중에서 가장 작은 요소의 값이 ε_1 미만이면 상기와 같은 방법으로 작은 행열 A_{i-2j-2}를 구한다. 식(14)가 만족될때까지 위의 過程을 반복한다.

과정 6 : 모든 피치 주기에 대해 4, 5의 과정을 수행한다.

과정 7 : 구해진 집단 S_j의 새로운 중심을 $\overline{P}_j'$라 하고 다음과 같이 구한다.

$$\overline{P}_j=\frac{1}{N_j}\sum_{P_{ij}\in S_j}P_{ij}\ :\ j=1,2,\cdots,J \quad (15)$$

여기서 $N_j(\leq N)$는 집합 S_j의 원소의 수이다.

과정 8 : 왜곡(distortion) 臨界값을 ε_2라 하고 모든 j에 대해

$$\| \overline{P}_j' - \overline{P}_j \| < \varepsilon_2 \quad (16)$$

이 되면 알고리즘은 收斂하는 것으로 하고 終了한다. 그렇지 않으면 $\overline{P}_j'$를 $\overline{P}_j$로 하고 과정 4로 가서 다시 모든 過程을 되풀이 한다.

2.4 距離 計算

集團化 알고리즘을 이용하여 登錄 話者와 未知 話者에 대하여 집단의 중심 좌표를 구한 후 이를 특징으로 하여 標準 패턴과 未知 음성 패턴 사이의 距離 計算으로 話者 識別을 행한다.

미지 화자와 등록 화자에서 추출된 集團의 中心을 각각 $\overline{P}_{uj}$, $\overline{P}_{rk}$라 하면 $\overline{P}_{rl}$과 Euclid 距離가 가장 가까운 집단의 중심 $\overline{P}_{un}$을 탐색하고 이 집단의 중심을 發聲한 發聲者를 찾아 화자 식별을 한다. 계산된 거리를 D_m으로 할 때 D_m은 식(17)과 같다.

$$D_m = \min \| \overline{P}_{ri} - \overline{P}_{ui} \| \ (i=1,2,\cdots,J) \quad (17)$$

여러 개의 숫자음을 조합하여 識別 實驗을 할 경우 등록 화자와 미지 화자의 각 숫자음에 대한 距離를 각각 $d_0, d_1, \cdots, d_9$라 하면 화자간 거리 D는 식(18)과 같이 정의하고 계산된 거리를 최소로하는 발성자를 찾아 식별을 한다.

$$D = \sum_{i=0}^{9} d_i \quad (18)$$

3. 實驗 結果 및 考察

3.1 識別 實驗

話者 識別을 위한 식별 對象 語彙로는 韓國語 숫자음 "영, 일, 이, 삼, 사, 오, 육, 칠, 팔, 구"의 10개를 선정하였다.

음성 데이터는 건강한 20대에서 30대의 남자 20명이 방음이 되어 있는 방에서 각 4회씩 발성한 것을 녹음하여 사용하였으며 발성을 위해 특별히 훈련을 한 것은 아니다.

음성 신호는 차단 주파수 7KHz, 표본화 주파수 16KHz로 표본화하였으며, signed 16Bit로 A/D 변환을 하였다.

실험에 사용한 음성 파형은 전체 음성 파형에서 비교적 파형의 에너지가 安定된 구간의 음성을 끊어내어 음성 데이터로 사용하였으며, 개인적인 특징을 추출하는 것이기 때문에 음성 인식에서 요구되는 끝점의 처리의 과정은 별도로 수행하지 않았다.

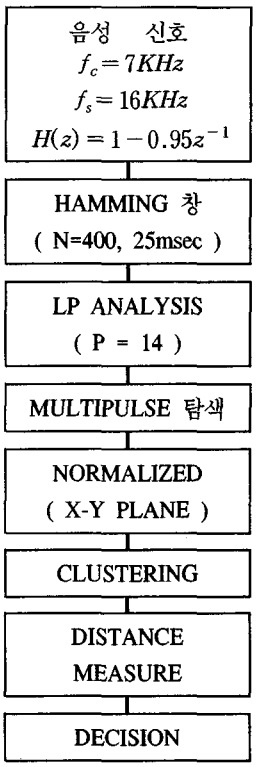
신호의 분석을 위하여 양자화된 음성 데이터는 전달 함수 $H(z)=1-0.95z^{-1}$로 고역 강조를 하였으며, Hamming 창을 이용하여 프레임의 길이 25msec로 음성 신호를 잘라내었으며 5msec씩 중복하여 분석하였다.

線形豫測 合成 필터의 傳達 函數는 14차의 선형 예측 係數를 사용하여 구성하였으며, 포르만트 주변에서의 誤差 에너지의 억제를 위해 사용한 noise weighting 필터에서 γ값은 0.8로 하였을 경우 가장 좋은 결과[6]를 실험적으로 얻고 있기 때문에 본 실험에서도 γ값을 0.8로 하여 펄스들을 탐색 하였다.

펄스들의 탐색을 위해 먼저 피치 주기를 찾아내고 탐색된 피치 주기에서 일정 수의 펄스를 구하였다. 얻어진 펄스는 피치 주기의 양 끝에 있는 주 펄스로 位置와 振幅에 대해 正規化를 하여 2차원 空間上의 座標로 나타내었다.

펄스 열의 분포로부터 펄스의 集團을 구하고 구해진 집단으로부터 집단의 中心 座標를 구하였다. 이들 중심 좌표를 特徵 벡타로 하여 화자 識別 實驗을 수행하였다.

본 실험에서는 특징 벡타의 수를 나타내는 집단의 수는 LPC 합성 필터의 차수와 같은 14개를 中心으로 8, 11, 17 및 20로 변화 시키면서 화자 식별 실험을 하였다. 識別을 위한 시스템의 불럭도는 [그림 1]과 같다.



[그림 1] 화자인식시스템의 불럭도

[Fig. 1] Block diagram of speaker identification system

3.2 實驗 結果 및 考察

본 연구에서는 B. S. Atal 등에 의해 音源 符號化의 한 방법으로 제안된 멀티펄스 열에 개인성 정보가 포함되어 있는 것에 着眼하여 화자별로 발성한 음성 신호로부터 멀티펄스 열을 추출하고 추출된 멀티펄스 열로부터 個人性 特徵 파라메타를 抽出하여 이를 화자 식별를 위한 특징 벡타로 하였다.

識別을 위한 對象 어휘로는 2 , 30대의 건강한 남성 20명이 4회씩 발성한 10개의 한국어 숫자음을 사용하였다.

실험은 피치 주기 사이에 존재하는 특징 벡타의 수, 즉 펄스의 수를 변화 시키면서 숫자별 식별 실험을 수행하였다. 특징 벡타의 수는 LPC 합성 필터의 線形豫測 計數의 차수와 같은 14개를 중심으로 하여 8, 11, 14, 17 및 20개로 결정하였다.

<표 1>은 피치 주기 당 펄스의 개수에 따른 10개의 숫자음에 대한 識別 結果를 나타내고 있다.

10개의 숫자음의 경우 평균 51%의 식별 결과를 얻었다. 식별율은 / pal(8) /에서 최고 70%의 식별율을 얻었다. 10개의 숫자음을 대상으로 할 경우 만족스런 식별 결과를 얻지 못하였기 때문에 복수개의 숫자음을 組合하여 식별 실험을 수행 하였다. 조합한 숫자음의 수가 두개인 경우는 "영과 일"을 세개인 경우는 "영, 일 및 이" 와 같이 다섯 개의 숫자음 까지를 組合하였다.

<표 2>는 복수개의 숫자음을 조합하여 식별을 할 경우 피치 주기 당 펄스의 개수와 결합한 숫자음의 개수에 따른 식별 결과를 나타내고 있다.

<표 1> 피치 주기당 서브펄스의 수에 따른 자인식율의 결과

<Table 1> The results of speaker identification rate as a function of the number of subpulses per pitch period(%)

펄스수 / 숫자음	8	11	14	17	20
yeong(0)	60	55	65	45	50
il(1)	50	45	55	45	55
i(2)	45	40	50	55	50
sam(3)	55	40	50	60	50
sa(4)	45	65	50	55	50
o(5)	50	45	50	65	50
yuk(6)	50	55	60	50	45
chil(7)	50	45	45	45	40
pal(8)	55	50	55	70	55
gu(9)	50	45	45	50	55
평 균	51.0	48.5	52.5	54.0	50.0

<표 2> 서브펄스의 수와 결합한 숫자음의 개수에 따른 화자인식율의 결과

<Table 2> The results of speaker identification rate as a function of the number of combined digits and the number of subpulses(%)

조합한 숫자의 수 / 펄스의 수	2	3	4	5
8	75	80	80	80
11	75	80	90	90
14	75	85	95	95
17	80	85	95	90
20	75	85	90	95

식별 결과 피치 주기 당 펄스 개수와 조합한 숫자음의 개수의 증가에 따라 식별율이 증가함을 보여주고 있다. 또 조합한 숫자음의 개수가 증가하면 펄스의 개수가 감소하여도 같은 식별 결과를 얻을 수 있었다. 식별 결과 네 개의 숫자음을 조합하고 피치 주기 당 펄스 수를 14개로 할 경우 최고 95%의 식별율을 얻을 수 있었다. 실험에서는 조합한 숫자음의 개수를 다섯 개로 할 경우에도 95%의 식별율을 얻었으며 조합하는 숫자음의 개수를 증가시킬수록 식별율이 증가할 것으로 예상되나 데이터를 처리함에 있어서 처리 시간이 더욱 증가할 것으로 판단된다.

따라서 화자 식별을 위한 最適 條件은 피치 주기 당 펄스의 수를 LPC 합성 필터의 차수와 같은 수인 14개에서 17개 정도로하고 組合하는 숫자음의 수를 4개로 하면 좋은 결과를 얻을 것으로 생각된다.

4. 결론

실험 결과 다음과 같은 몇 가지 결론을 얻었다.

첫째, 펄스들의 개수를 8개, 11개, 14개, 17개, 20개로 각각 變化시킬때 열개의 숫자음에 대한 화자 식별은일정한 傾向을 갖지 않고 對象으로 하는 숫자음에 따라 결과가 서로 달랐다.

식별율은 최고 70%, 최저 40%의 식별 결과를 얻었으며 평균 51%%의 식별율을 얻었다.

둘째, 숫자음을 2개에서 5개 까지 순서대로 組合하여 화자 식별을 할 경우 식별율은 조합한 숫자의 증가에 따라 증가하였다. 4개의 숫자음을 조합하고 펄스들의 개수를 14개로 하였을 때 95%의 식별율을 얻었다.

셋째, 숫자음을 組合할 경우 最適의 條件은 4개의 숫자음을 조합하고 14~17개의 펄스를 사용할 때임을 알 수 있었다.

넷째, 멀티펄스 열의 특징을 이용한 화자 식별에서 하나의 숫자음을 사용할 경우는 식별율이 50% 前後로 한계가 있다. 이는 피치 주기 사이에 존재하는 펄스가 聲帶의 共振 特性 즉 피치 주기와는 다른 음원 특징을 나타내고 있으며 음원 특징만으로는 개인성 정보를 나타내는데 限界가 있음을 의미한다. 그러나 여러 개의 숫자음을 조합할 경우는 90% 이상의 식별율을 얻을 수 있었으며 멀티펄스 열의 특징만으로도 화자 식별을 위한 개인성 특징 파라메타로 사용할 수 있다고 판단된다.

※ 참고문헌

[1] 古井貞熙, "ケプストラムの統計的特徵による話者認識", 信學論 vol. J63-A, no. 2 pp. 183~190, 1982.

[2] Douglas O'Shaughnessy, "Speaker recognition", IEEE. ASSP. Magzine, pp. 4~17.1986.

[3] M. Sambur, "Selection of acoustic features for speaker identification", IEEE Trans. Acoust., Speech, Signal Processing, ASSP-23, 1975.

[4] Furui S., "Cepstral analysis technique for automatic speaker verification", IEEE Trans. Acoust., Speech, Signal Processing., ASSP-29, 1981.

[5] 半谷精一郎 他, "話者認識に對するマルチパルス驅動型聲帶音源情報の有效性の檢討", 信學技報, SP88-20,1988.

[6] Atal B. S. and Remde J. R., "A new model of LPC excitation for producing natural sounding speech at low bit rates", IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 614~617 ,1982.

[7] Song F. K. and Rosenberg A. E., "On the use of instantaneous and transitional spectral information in speaker recognition", IEEE Trans. Acoust., Speech, Signal Processing., ASSP-36, 1988.

[8] 小澤一範, 荒關 卓, 小野 茂, "マルチパルス驅動型音源符號化法の檢討", 信學技報, CS82-161,1982.

[9] 武田 昌一, 잔川 吉章, 市川 熹, "殘差音源利用分析合成方式とマルチパルス法の基本特性の比較檢討", 信學論 A, vol J73-A, no. 11 pp. 1735~1742 ,1990-11.

[10] J. N. Holmes, "Formant excitation before and after glottal closure", IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 39~42 , 1976.

[11] P. E. Papamichalis, "Practical approaches to speech coding", Prentice-hall, Inc., Englewood Cliffs, New Jersey

[12] J. T. Tou and R. C. Gonzalez, "Pattern recognition principles", Addison-Welsey Publishing Company, Inc., pp. 86~104

[13] L. R. Rabiner and J. G. Wilpon, "Consideration in applying clustering techniques to speaker independent word recognition", J. Acoust. Soc. Amer., 66-3 pp. 663~673, 1979.

정 의 붕

1984. 원광대학교 전자공학과 졸업(공학사)

1986. 광운대학교 컴퓨터공학과 졸업 (공학석사)

1992. 건국대학교 전자공학과 졸업(공학박사)

현재 : 호원대학교 전기전자정보공학부 부교수

관심분야 : 신경망, 음성인식, 로보틱스