퍼지추론을 이용한 한국어 자음분류에 관한 연구

박 경 식*°, 백 재 영*, 지 규 인*, 박 인 갑*, 김 영 권*, 김 형 래*
* : 건국대학교 전자공학과

A Study on the Consonant Classification Using Fuzzy Inference

Kyung-Sik Park*°, Jae-Yeong Baeg*, Gyu-In Jee*, In-Gap Park*, Yung-Kwon Kim*, Hyung-Lae Kim*
* : Kon Kuk University Electronics Engineering

ABSTRACT

This paper proposes algorithm in order to classify Korean cosonant phonemes same as polosives, fricatives affricates into lax sounds, glottalized sounds, aspirated sounds. This three kinds of sounds are one of distinctive characters of the Korean languge which don't exist in language same as English.

This is thesis on classfication of 14 Korean consonants(k,t,p,s,c,k',t',p',s',c',k^h,t^h,p^h,c^h) as a previous stage for Korean phone recognition.

As feature sets for classfication, LPC cepstral coefficients induced from the 18'th LPC and cepstral analysis.

The experiments are two stages. First, using short-time speech signal analysis and Mahalanobis distance, consonant segments are detected from original speech signal, then the cosonants are classified by fuzzy inference.

As the results of computer simulations, the classification rate of the speech data was come to 93.75%.

제 1 장 서 론

음성의 인식단위에는 음소(phonemes), 음절 (syllables), 단독어(isolated words), 연결단어 (connected words), 연속음 (continuous speech) 등이 있다. 최근까지 음성인식에서 이용되는 인식단위는 주로 단독어로서, 단독어는 음성의 특징점들을 추출하기가 쉽고, 시간적 변화에 따른 매칭의 어려움을 쉽게 극복할 수 있으므로 주연구대상이 되어 왔다. 그러나, 인식대상 어휘가 증가함에 따라, 차지하는 기억용량이 증가하고 처리시간이 늘어나는 등, 처리에 한계가 있으므로, 음성을 보다 세밀한 단위로 나누어 처리할 필요가 있다.

특히, 임의 어휘를 대상으로 하는 인식시스템의 개발에는 음소단위에 의한 인식기술의 개발이 필수적이다. 그러나, 연속음성 중의 음소는 발성자에 의한 영향, 즉 개인차 (individual variation)뿐 아니라, 그때그때의 감정, 즉 정서성(emotion)과 전후에 발성되는 음소의 영향, 즉 조음결합(articulation combination) 등으로 인하여 공통된 특징점을 파악하기가 힘들다. 따라서, 음소의 특징점을 제대로 파악하기 위해서는 대량의 음성 데이터와 그에 알맞는 알고리즘의 개발이 필요하다.

본 논문에서는 음성신호가 궁극적으로 그 특징점을 파악하기가 어려운 반면, 추출된 특징점들이 반드시 어떤 음성, 또는 음소를 가리키는 것이 아닌 애매성을 지닌 값이라는 것에 근거하여 퍼지이론 (fuzzy theory)을 도입하고, 추출된 특징파라메타에 대하여 퍼지추론의 합성규칙 (the combinational rule of fuzzy inference)을 적용하여 음소 구분화를 수행한다.

제 2 장 전처리 과정

2-1. 음성구간의 검출

정확한 음성구간을 검출하기 위하여 음성신호를 처리의 기본단위인 프레임단위로 구분하고, 각각의 프레임에 대하여 에너지와 영교차율(Zero Crossing Rate)을 측정함으로써 음성의 시작프레임과 끝프레임을 검출한다

n번째 프레임구간의 에너지와 영교차율는 아래의 식과 같다.

$$E(n) = \sum_{i=0}^{N-1} |s(nN+i)| \quad n = 0,1,2,\cdots \tag{2-1}$$

$$ZCR(n) = \sum_{i=0}^{N-1} 0.5*|sign(s(n+i))-sign(s(n+i-1))| \tag{2-2}$$

여기서, s(n)은 음성샘플이다.

여기서, $sign(s(n)) = \begin{cases} 1 & \text{for } s(n) \geq 0 \\ -1 & \text{otherwise} \end{cases}$

2-2. 창함수 및, 프리엠퍼시스

음성은 성도의 변화에 따라 시간적으로 변화하나, 짧은 시간 구간만을 고려하면, 그 변화가 정적이라고 할 수 있다. 따라서 대부분의 음성분석기법은 음성신호가 상대적으로 느리게 변화한다는 특성을 이용하여, 고정된 길이의 단시간 창함수를 이용하여 프레임 단위로 구분하고 관련 파라메타를 추출한다. 또한, 음성신호의 진폭을 보다 작게하고, 고주파수 대역에서 보다 많은 정보를 가지도록 하기 위하여 분석전에 프리엠퍼시스(pre-emphasis)를 행한다.

2-2-1. 창함수

단시간 음성 분석을 행하기 위해서는, 창함수의 도입이 필수적이다. 창함수의 선택시 고려사항은 다음과 같다.

1. 음성의 특성이 창내에서 크게 변화하지 않도록 충분히 짧아야 한다.
2. 바라는 파라메타를 추출하기에 충분한 샘플을 포함하도록 충분히 길어야 한다.

본 논문에서는 위의 2가지 사항을 고려하고, 본 연구의 특성에 맞게 Hamming창을 선택하였다. Hamming창은 다음 수식에 의하여 정의된다.

$$w(n) = \begin{cases} 0.54 - 0.46\cos(2\pi n/N) & \text{for } 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases} \tag{2-3}$$

2-2-2. pre-emphasis

$$y(n) = s(n) - \alpha * s(n-1) \tag{2-4}$$

즉,

$$H(z) = 1 - \alpha * z^{-1} \qquad (2\text{-}5)$$

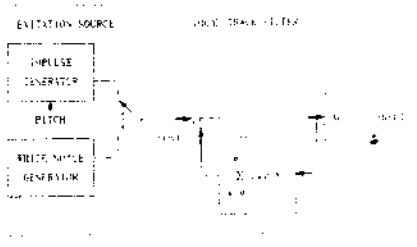
여기서, y(n) : 프리엠퍼시스를 통하여 구해진 신호
H(z) : 전달 함수
s(n) : 음성 샘플
α : 강조의 정도를나타내는 상수(0.95)

제 3 장 단시간 음성신호의 분석

이 장에서는 본 논문의 중심이 되는 유·무성음 구분화와 자음 구분화(평음, 경음, 격음)에서 주로 사용되는 LPC 분석과 켑스트럼 분석(cepstral analysis)에 대하여 주로 언급하였다.

3-1. LPC 분석

LPC 분석은 성도 공진을 나타내는 극점을 구하는 것으로, 분석법으로는 최소자승법(least square method), 격자법(lattice method)등이 있다. 상대적으로 계산이 간편한 최소자승법이 주로 사용되고 있으며, 본 논문에서도 역시 최소자승법을 사용하였다.



[그림 3-1] 디지탈 음성생성 모델(전극 성도모델)

3-1-1. LPC 모델

$$s(n) = \sum_{k=1}^{P} a_k z^{-k} \qquad (3\text{-}1)$$

LPC 계수 $a_k(k=1,2,\cdots,P)$는 LPC 분석에서 구하고자 하는 음성파라메타이고, P는 LPC의 차수이다.

3-1-2. 최소자승법에 의한 LPC 계수의 추출

최소자승법은 음성데이타에 대하여 LPC 분석을 행하여 오차신호 e(n)을 구하고, 그 오차신호내의 평균에너지 E를 최소화시키는 LPC계수 a_k를 선택한다는 개념이다. i번째 프레임의 오차신호 E_i내의 에너지는 E_i가 최소값을 가지기 위해서는

$$\frac{\partial E_i}{\partial a_k} = 0 \qquad \text{for } i = 1,2,3,\cdots,P \qquad (3\text{-}2)$$

$$R(i) = \sum_{n=0}^{N-1} x(n-1)\, x(n-k) \quad \text{for } i = 1,2,3,\cdots,P \qquad (3\text{-}3)$$

$$\sum_{k=1}^{P} a_k R(i-k) = R(i) \qquad (3\text{-}4)$$

즉, $\Phi \cdot a = R$ (3-5)

$$\begin{bmatrix} R(0) & R(1) & R(2) & \cdots & R(P-1) \\ R(1) & R(0) & R(1) & & \cdot \\ R(2) & R(1) & R(0) & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ R(P-1) & R(P-2) & R(P-3) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \cdot \\ a_P \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ R(3) \\ \cdot \\ R(P) \end{bmatrix} \qquad (3\text{-}6)$$

3-1-3. Levinson-Durvin 회귀절차를 이용한 LPC계수 추출

초기치 E_0, a_0가 아래의 값을 갖을때

$$E_0 = R(0), \quad a_0 = 0 \qquad (3\text{-}7)$$

$m = 1, 2, 3, \cdots, P$에 대하여

$$K_m = \frac{R(m) - \sum_{k=1}^{m-1} a_{m-1}(k)\, R(m-k)}{E_{m-1}} \qquad (3\text{-}8)$$

$$a_m(m) = K_m \qquad (3\text{-}9)$$

$$a_m(k) = a_{m-1}(k) - K_m\, a_{m-1}(m-k) \quad \text{for } 1 \le k \le m-1 \qquad (3\text{-}10)$$

$$E_m = (1 - K_m^2)\, E_{m-1} \qquad (3\text{-}11)$$

각 차이를 m에서 식(3-10)에서 구한 계수 $a_m(k)(k=1,2,3,\cdots,m)$는 최적의 m차 선형예측계수를 의미한다. 따라서, 구하고자 하는 P차 선형예측계수는

$$a_k = a_P(k) \quad \text{for } k = 1,2,3,\cdots,P \qquad (3\text{-}12)$$

3-2. 켑스트럼 분석(cepstral Analysis)

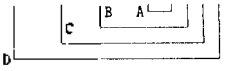
음성 X(z)는 여기원 E(z)(exitation source : randomnoise (무성음), quisiperiodic pulse train (유성음))와 성도필터(vocal track filter)의 전달함수 V(z)의 곱으로 나타낼 수 있고, 이것은 시간영역에서는 콤볼류션으로 변환될 수 있다. 즉,

$$X(z) = E(z) \cdot V(z) \qquad (3\text{-}13)$$

음성신호에 있어서 분리콤볼류션(decombolution)에 의하여 여기원과 성도의 임펄스 응답을 분리하는 것이 실제로는 불가능하다고 할지라도, 켑스트럼 분석과정을 거치면 하나의 신호성분을 감쇄시키고, 고찰하고자 하는 성분을 얻을 수 있다.

음성신호 x(n)을 푸리에 변환의 일종인 이산퓨리에변환(DFT)하여 구한 스펙트럼 신호를 X(k)라 할 때, x(n)의 켑스트럼 c(n)은 다음과 같다.

$$c(n) = \mathrm{IDFT}\,[\log |\, \mathrm{DFT}\,[\, x(n)\,]\,|\,] \qquad (3\text{-}14)$$



식(3-14)의 결과 C를 고찰하여 보면

$$C = \log |\, \mathrm{DFT}\,[\, x(n)\,]\,| \qquad (3\text{-}15)$$

$$= \log |\, V(z)\, P(z)\,| \qquad (3\text{-}16)$$

$$= \log |\, V(z)\,| + \log |\, P(z)\,| \qquad (3\text{-}17)$$

이제까지의 과정을 거친 성도임펄스 응답의 켑스트럼은 다음과 같다.

$$c(n) = \frac{1}{N} \sum_{K=0}^{N-1} \log|X(K)| \exp\left(\frac{2\pi K n}{N}\right) \quad \text{for } n=0,1,2,\cdots,N-1 \qquad (3\text{-}18)$$

3-2-1. LPC를 이용한 LPC켑스트럼 계수의 추출

LPC계수 $a_i(i=1,2,\cdots,P)$를 이용하여 LPC스펙트럼은 아래의 식과 같이 구할 수 있다.

$$X(K) = \sum_{N=0}^{N-1} x(n) \exp\left(-j\frac{2\pi K n}{N}\right) \quad K=0,1,2,\cdots,N-1 \qquad (3\text{-}19)$$

여기서, $x(0)=1$, $x(1)=a_1$, $x(2)=a_2$, $\cdots$ $x(P)=a_P$,
$x(P+1)=0$, $x(P+2)=0$, $\cdots$, $x(N-1)=0$

(3-19)에서 구한 X(K)를 식 (3-18)에 적용하여 LPC켑스트럼 계수를 구한다. LPC는 DFT에 의하여 구한 스펙트럼이 평활화된 형태로서 계산이 간단한 반면, 음성의 주요한 특성을 양호하게 나타내고 있으므로, 본 논문에서는 자음 구분화를 위한 파라메타의 하나로서 18차 LPC분석을 이용한 LPC켑스트럼 계수를 이용하였다.

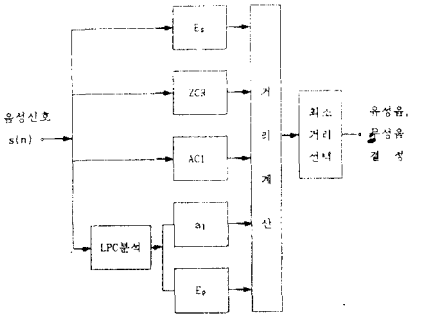
제 4 장 자모음 구분

본 논문에서 실험의 주대상은 다음과 같은 자음음소이다.

1) 평음(laxes) : ㄱ, ㄷ, ㅂ, ㅅ, ㅈ
2) 경음(glottalized sounds) : ㄲ, ㄸ, ㅃ, ㅆ, ㅉ
3) 격음(aspirated sounds) : ㅋ, ㅌ, ㅍ, ㅊ

자음구간의 추출에 사용된 특징파라메타는 아래와 같다.

1) 대수에너지 E_s
2) 영교차율 ZCR.(식 2-2 참조)
3) 1샘플 지연된 자기상관계수 AC1.
4) 10차 선형예측계수의 제1계수 a_1.(식 3-12 참조)
5) 예측오차의 대수에너지 E_p



[그림 4-1] 유성음, 무성음 결정 모형도

4-1. 특징파라메타의 추출

$$E_s = 10 * \log_{10}\left\{ \varepsilon + \frac{1}{N} \sum s^2(n) \right\} \qquad (4-1)$$

ε : $\log_{10}(0)$의 계산을 막기위한 작은 값의 양수

$$AC1 = \frac{\sum_{n=0}^{N-1} x(n)\,x(n-1)}{\left\{ \left\{ \sum_{n=0}^{N-1} x(n)x(n-1) \right\} \left\{ \sum_{n=1}^{N-1} x(n)x(n-1) \right\} \right\}^{1/2}} \qquad (4-2)$$

$$E_p = 10*\log_{10}R(0) - 10*\log_{10}\left\{ R(0) - \sum_{k=1}^{p} a_k R(k) \right\} \qquad (4-3)$$

4-2. 정상분포를 이용한 결정규칙

앞절에서 언급한 5가지의 측정값을 이용하여 음성신호의 해당 프레임을 유성음, 또는 무성음으로 분류한다. 유·무성음의 분류는 고전적인 오차의 최소확률 규칙(a classical minimum probability-of-error decision rule)이 사용된다. 평균(mean)벡터를 $m_i(i=1,2)$라 하고, 공분산(covariance)행렬을 W_i라 할 때, m_i와 W_i는 아래의 식을 통하여 구해진다.

$$m_i = \frac{1}{N_i} \sum x_i(n) \qquad (4-4)$$

$$W_i = \frac{1}{N_i} \sum x_i(n)\, x_i^t(n) - m_i\, m_i^t \qquad (4-5)$$

여기서, $x_i(n)$: i번째 음성프레임의 L차원 특징벡터(L=5)
$(x_k(k=0,1,\cdots L-1)$는 측정된 k번째 특징 파라메타.)
N_i : 음성신호의 프레임수

$$d_i(x) = (x-m_i)^t\, W_i^{-1} (x-m_i) \qquad (4-6)$$

W_i^{-1} : W_i의 역행렬(inverse matrix)
t : 벡터의 전치(transpose)를 나타내는 첨자

측정벡터 x를 각부류에 대하여 $d_i(x)$를 계산하고, 그 값이 최소인 i 를 해당프레임의 부류(1:유성음, 또는 2:무성음)로 결정한다.

4-3. 평활화 알고리즘(smoothing algorithm)

유·무성음 구분화를 통한 자음구간의 추출을 위하여, 마지막 단계로서 자음구간의 시작프레임과 끝프레임을 검출한다. 즉, 무성음 + 유성음 + 무성음 이면 유성음은 무성음으로 되고, 무성음(3개이상) + 유성음(2개) + 무성음(2개이상)이면 중간에 2개의 유성음 또한 무성음으로 한다. 이와같은 과정을 거쳐, 음성신호구간내의 자음구간을 검출하고, 시작프레임과 끝프레임을 결정함으로써, 결정된 자음구간을 중심으로 자음구분화를 수행한다.

제 5 장 퍼지추론을 응용한 자음분류

본 논문에서는 먼저 음성구간으로부터 자음구간을 추출하고, 추출된 자음구간과 자음구간 이후 혼합음구간(hybrid segments)에 대하여 각각 LPC캡스트럼계수를 추출하고 퍼지추론(fuzzy inference)을 이용한 분류를 수행하였다.

5-1. 퍼지추론

퍼지추론은 소속도함수 μ를 이용하여 수행된다. 퍼지추론과 퍼지관계를 소속도함수로 나타내는 방법은 여러가지가 있으나, 현재 가장 많이 사용되는 대표적인 방법인 max-min연산을 이용한다.

$$\mu_{R'}(y) = \vee [\, \mu_{A'}(x) \wedge \mu_R(x) \,] \qquad (5-1)$$

여기서, "$\vee$ (a, b)"는 a, b에서 최대값을 취하며, "$\wedge$ (a, b)"는 a, b가운데 최소값을 취함을 말한다.

5-2. 퍼지집합의 생성

먼저 각 파라메타에 대하여 벡터 양자화(vector quantization)를 행하고, VQ 코드북을 생성시킨 후에, 각 파라메타 값에 대하여 퍼지집합의 전체집합 (universal set)의 원소에 해당하는 퍼지화 값을 할당한다. VQ 코드북과 퍼지집합의 전체집합 U_c가 생성되면, 이 2가지를 이용하여 기준패턴(reference pattern)과 시험패턴 (test pattern)에 대하여 소속도 함수를 구하고 퍼지집합을 생성한다.

5-2-1. 파라메타의 퍼지화

단시간 분석(short time analysis)을 통하여 검출된 파라메타인 LPC캡스트럼 계수 C_0-C_4에 대하여 소속도 함수를 구하기 위해서, 먼저 입력패턴 (input pattern)에 대한 퍼지화 패턴(fuzzified pattern)을 정의한다.

LPC 캡스트럼 계수 C_0-C_4에 각각에 대하여 벡터 양자화를 시행하면, 그 값은 -2부터 2까지의 범위에 걸쳐 있음을 알 수 있다. 따라서 0.01간격으로 수치를 구분하여 표(5-1)과 같이 퍼지화 값을 부여하였다. 따라서, LPC 캡스트럼 계수의 퍼지집합에 대한 전체집합 U_c는 다음과 같다.

$$U_c = \{0, 1, 2, \cdots, 399\} \qquad (5-2)$$

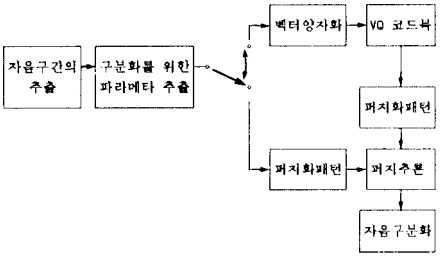
[표 5-1] LPC캡스트럼 계수의 전체집합 생성

LPC 캡스트럼 계수 C_0-C_4	퍼지화 값
-2.00 ~ -1.99	0
-1.99 ~ -1.98	1
-1.98 ~ -1.97	2
⋮	⋮
-0.02 ~ -0.01	198
-0.01 ~ 0.00	199
0.00 ~ 0.01	200
0.01 ~ 0.02	201
⋮	⋮
1.97 ~ 1.98	397
1.98 ~ 1.99	398
1.99 ~ 2.00	399

5-3. 퍼지추론에 의한 자음 구분화

앞절의 과정을 이용하면 벡터양자화 코드북의 기준패턴(reference pattern)과 시험패턴(test pattern)의 퍼지집합

F_{crk},F_{ctk}를 구할 수 있다. 본절에서는 이러한 퍼지집합을 근간으로 하여, 2단계의 퍼지추론을 수행하고, 마지막 단계에서 자음 구분화를 수행한다(그림 5-1 참조).



[그림 5-1] 자음 구분화

5-3-1. 파라메타 단위의 퍼지추론

LPC 캡스트럼 계수에 대하여 퍼지추론의 합성규칙(the compositional rule of fuzzy inference)을 적용하면 다음과 같다.

$$\mu_{ck} = \vee(\mu_{crk} \wedge \mu_{ctk}) \quad \text{for } k=0,1,2,\cdots,6 \qquad (5\text{-}3)$$

여기서, μ_{crk} : LPC 캡스트럼 계수 C_k에 대한 기준패턴의 소속도 함수
μ_{ctk} : LPC 캡스트럼 계수 C_k에 대한 시험패턴의 소속도 함수
μ_{ck} : LPC 캡스트럼 계수 C_k에 대하여 추론결과 구해진 소속도 함수

연산($\mu_{crk} \wedge \mu_{ctk}$)를 표로 나타내면 표5-2와 같다.

본 논문에서는 퍼지집합을 실제로 구성하지 않고, 단지 파라메타 값에 대응하는 퍼지집합의 원소, 즉 퍼지화 패턴의 차를 이용하여 식(5-4)와 같이 간편하게 max-min연산의 소속도 함수 μ_{ck}를 구한다. 즉,

$$\mu_{ck} = \frac{R_c - |f_{crk} - f_{ctk}|}{R_c} \qquad (5\text{-}4)$$

여기서, R_c : 소속도 함수 μ_k가 적용되는 퍼지화 값의 범위 (위의 예에서는 400)
f_{crk} : 기준패턴에 대한 퍼지화 값
f_{ctk} : 시험패턴에 대한 퍼지화 값

[표 5-2] $\mu_{crk} \wedge \mu_{ctk}$에 대한 결과를 나타내는 도표

퍼지화 값	소속도함수 μ_{crk}	소속도함수 μ_{ctk}	$\mu_{crk} \wedge \mu_{ctk}$
0	0.00	0.00	0.00
1	0.00	0.00	0.00
⋮	⋮	⋮	⋮
180	0.01	0.00	0.00
181	0.02	0.00	0.00
182	0.03	0.01	0.01
183	0.04	0.02	0.02
⋮	⋮	⋮	⋮
227	0.95	0.95	0.95
228	0.99	0.97	0.97
229	1.00	0.98	0.98
280	0.99	0.99	0.99
281	0.98	1.00	0.98
282	0.97	0.99	0.97
283	0.96	0.98	0.96
⋮	⋮	⋮	⋮
378	0.01	0.03	0.01
379	0.00	0.02	0.00
380	0.00	0.01	0.00

끝점	0.00	0.00	0.00
끝점	0.00	0.00	0.00
퍼지화패턴의 소속도 함수 μ_{ck}			0.99

5-3-2. 프레임 단위의 퍼지추론

LPC 캡스트럼 계수에 대하여 퍼지추론을 행하여, 그 소속도 함수 μ_k(k=0,1,⋯,5)를 구한 후에, 그중에서 최소의 값을 취하고, 코드북내의 M개의 코드워드와의 연산중에서 최대의 소속도 함수를 취함으로써 프레임단위의 퍼지추론을 행한다. 즉,

$$\mu_{fi} = \max_{m}[\min_{k}[\mu_{ck}]] = \vee_{m}[\wedge_{k}\mu_{ck}] \quad \text{for } m=0,1,\cdots,M-1,\ k=0,1,\cdots,6 \qquad (5\text{-}5)$$

여기서, k : LPC 캡스트럼 계수를 나타내는 첨자
m : VQ 코드북의 코드워드를 나타내는 첨자
i : 프레임 번호를 나타내는 첨자

5-3-3. 자음구간

자음구간의 길이는 자음에 따라 다르고, 또한 동일한 자음일지라도 화자에 따라 다를 수 있다. 따라서, 식(5-5)을 수행한 후에는 자음구간의 길이의 영향을 감안하도록 한다. 특정 자음구간이 L개의 프레임으로 구성되어 있다면, 각각의 프레임에 대하여 프레임단위 퍼지추론을 수행한 후, 이를 누적하고, 프레임수 L로 나누어 줌으로써, LPC 캡스트럼계수의 자음구간에 대한 소속도 함수 μ_c를 구한다. 즉,

$$\mu_c = \frac{1}{L}\sum_{i=0}^{L-1}\mu_{fi} \qquad (5\text{-}6)$$

5-3-4. 혼합음 구간

실험을 통하여 구한 자음구간은 눈으로 관찰한 실제의 자음구간에 비하여 짧게 나타나는데, 이것은 후속모음의 영향으로 유성음화되어 나타나는 현상으로 이러한 구간은 유성음과 무성음이 혼재되어 있는 구간으로 볼 수 있으므로 자음구간 이후 5개 프레임(L=5로 변환)에 대하여 식 5-4부터 5-6까지의 과정을 반복함으로써 혼합음구간에 대한 소속도함수 μ_{cc}를 구한다.

5-3-5. 자음부류의 결정

$$\mu_l = \frac{\mu_c + 0.7\mu_{cc}}{2} \quad \text{for } l = 0,1,2,\cdots,I-1 \qquad (5\text{-}7)$$

여기서, I는 각 파라메타에 대하여 존재하는 코드북의 수이다.
μ_l을 존재하는 모든 표준패턴에 대하여 구한 후에 최대의 소속도 함수를 구하면, 최대의 소속도 함수를 취하는 표준패턴의 부류가 해당자음의 부류(평음,또는 경음,격음)가 된다.

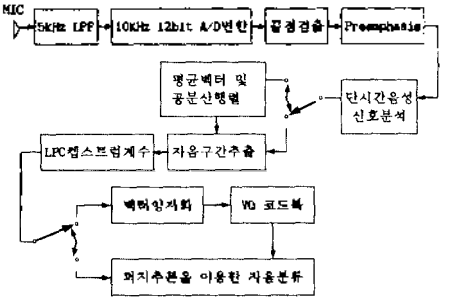
제 6 장 실험 및 고찰

6-1. 음성 데이타베이스

본 연구에서 사용된 음성 데이타는 방음장치가 되어있지 않은 연구실에서 콘덴서 마이크를 사용하여, 20대 화자 3명과 30대 화자 2명등 총 5명의 화자가 한국어 자음 14개(ㄱ,ㄲ,ㅋ,ㄷ,ㄸ,ㅌ,ㅂ,ㅃ,ㅍ,ㅅ,ㅆ,ㅈ,ㅉ,ㅊ)와 단모음 8개(ㅏ,ㅣ,ㅗ,ㅜ,ㅡ,ㅔ,ㅐ,ㅟ)등을 결합하여 만든 총 112개 단음을 5명의 화자가 각각 5회씩 발음한 총 2800개 단음에 대하여 실험을 수행하였다. 그 가운데 처음 2회 발음한 1120개 단음을 기준음으로 설정하여 자음구간의 추출에 이용되는 평균벡터(mean vector), 공분산행렬(covariance matrix), 그리고 자음분류에 이용되는 벡터양자화에 사용하였다. 기준패턴을 생성시킨 후에 나머지 3회의 단음을 시험데이타로 사용하여 자음분류를 위한 실험을 수행하였다.

6-2. 시스템의 구성

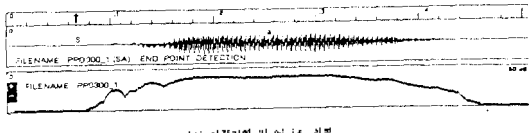
본 연구의 시스템 구성도는 그림 6-1과 같다.



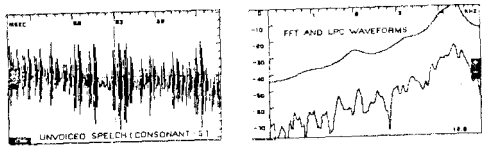
[그림 6-1] 분류시스템 구성도

6-3. 컴퓨터 시뮬레이션

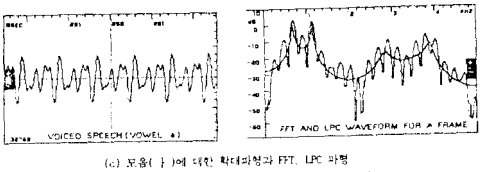
시뮬레이션은 크게 2단계로 분류된다. 1단계는 음성신호 구간으로부터 자음구간을 추출하는 것으로, 자음당 24개의 어휘등, 총 336개의 어휘로부터 구한 평균벡터와 공분산행렬을 기준패턴으로 사용하고, 음성데이타를 단시간 분석하여 구한 파라메타들을 시험패턴으로 삼아, 확률이론을 이용한 Mahalanobis거리를 비교하여 각각의 프레임을 유성음과 무성음으로 구분한다. 이때, 유성음은 모음프레임으로 간주하고, 무성음은 자음프레임으로 간주한다. 다음에 평활화 알고리즘을 수행하여 음성신호로부터 자음구간을 결정하였다. 그런데, 자음구간 이후 대략 5개 프레임(50ms)구간은 앞의 자음과 후속



(a) 시간파형 및 에너지 파형
(a) Time waveform and Energy waveform



(b) 자음(s)에 대한 확대파형과 FFT, LPC 파형
(b) Enlarged waveform and FFT, LPC waveforms for a consonant('s')



(c) 모음(a)에 대한 확대파형과 FFT, LPC 파형
(c) Enlarged waveform and FFT, LPC waveforms for a vowel('a/')

[그림 6-2] 끝점검출이 이루어진 음성데이타('사')에 대한 각종 분석 파형

모음의 영향으로 유성음과 무성음의 성격이 동시에 나타나는 혼합음구간으로 2단계 실험에서 이 구간 또한 자음분류에 이용하였다.

2단계는 퍼지추론을 이용한 자음분류로서 1단계 실험에서 구한 자음구간과 혼합음구간 5개프레임에 대하여 18차 LPC분석을 행하고 이로부터 7차의 LPC켑스트럼계수를 구함으로써 퍼지추론을 통한 자음분류(평음, 경음, 격음)를 수행하였다.

시뮬레이션은 Coprocessor가 설치된 80386컴퓨터를 이용하여 수행하였다. 시뮬레이션 결과 기준패턴으로 이용된 음성데이타에 대한 분류율은 93.75%였다

제 7 장 결 론

음성인식에 대한 연구는 지난 수십년간 지속되어 왔으나, 그 인식대상 어휘가 주로 단독어에 한정됨으로써 실제응용에 있어서는 많은 문제점이 발생한다. 연속음인식의 경우도 단독어 인식의 확장에 불과한 것으로 실제 음성의 인식에 이르기에는 매우 한정적이다. 만약 인식대상 어휘수를 늘인다면 그만큼 기억용량이 증가하고 처리속도가 저하되므로, 이를 해결하기 위하여는 고속, 대용량 시스템의 개발이 필요하다. 그런데, 만약 음소단위의 인식이 가능해진다면 소규모의 기억용량을 사용하면서도 임의 어휘에 대한 인식이 가능해지므로 이에 대한 연구가 반드시 필요하다. 그러나, 실제에 있어서는 특정음소에 대한 특징점을 제대로 파악하기도 힘들뿐더러, 그 음소가 위치하여 있는 장소에 따라서 음소의 특징이 변화하므로 음소에 대한 인식을 위해서는 상황에 따라 방대한 양의 데이타가 필요하다. 본 논문에서는 이를 감안하여 음소인식의 전단계로써 한국어 자음의 특징중 하나인 평음, 경음, 격음을 분류하는 연구를 수행하였다. 또한 인간의 음성 생성 시스템이 성도 필터로 모델화 될 수 있으므로 LPC켑스트럼 분석을 통하여 성도 특성을 분리해낼 수 있다는 것에 근거하여 18차 LPC분석을 수행한 후, 7차의 켑스트럼을 구하고 자음의 분류를 수행하였다.

실험에 사용된 분류자는 퍼지추론으로서, 음성의 특징파라메타가 기본적으로 어떤 특정음소를 가리키는 것이 아닌, 애매성을 지닌 값이라는 것에 근거하여 이를 이용하였다. 실험의 결과를 보면, 평음, 경음, 격음의 구분화가 만족할만하게 이루어졌음을 알 수 있다. 그러나, 대상 자음이 단음의 초성으로 한정됨으로써 주변음소에 의한 조음결합등의 영향이 고려되지 않았으므로 앞으로 이에 대한 연구가 진행되어야 할 것이다.

참 고 문 헌

[1] NIST : "Speech Copora Produced on CD-ROM Media by The National Institute of Standards and Technology(NIST)" ,April 1991.

[2] KAIST, "디지탈 음성처리 기술개발연구 최종보고서", 한국과학기술원 전기 및 전자공학과 통신연구실, 1984.12

[3] L.F.Lamel, L.R.Rabiner, A.E.Rosenberg, J.G.Wilpon, "An Improved Endpoint Detection for Isolated Word Recognition", IEEE Trans., on Acoustics, Speech and Signal Processing, Vol. ASSP-29, NO. 4, pp777-785, Aug 1981.

[4] D. O'Shaughnessy, "Speech communication, Human and Machine", Addison-Wesley Publishing Company, 1987.

[5] Sadaoki Furui, M, M, Sondhi, "Advances in Speech Signal Processing", NTT Human Interface Laboratoies, Tokyo Japan, 1991

[6] R, M, Gray, "Vector Quantization", IEEE Trans., on Acoustics, Speech and Signal Processing, Vol. ASSP-32, 1984.

[7] Kai-Fu Lee, Hai So-Wuen Hon, "Speaker-Independent Phone Recognition Using Hidden Marcov Models", IEEE Trans., on Acoustics, Speech and Signal Processing, Vol. ASSP-37, NO. 11, November 1989.

[8] A. Waibel, Kai-Fu Lee, "Speech Recognition", Morgan Kaufmann Publisher, Inc. California, 1989.

[9] L. J. Siegel, A. C. Bessey, "Voiced/Unvoiced/Mixed Exitation Classification of Speech", IEEE Trans., on Acoustics, Speech and Signal Processing, Vol. ASSP-30, NO. 3, June 1982.

[10] 이기문, 김진우, 이상역, "국어 음운론", 학연사, 1984.3

[11] W. Verhelst, O. Steenhaut, "A New Model for thr Short-Time Complex Cepstrum of Vloiced Speech", IEEETrans., on Acoustics, Speech and Signal Processing, Vol. ASSP-34, NO. 1, February 1986.

[12] 이창형, 오길록, "퍼지 이론 및 응용", 홍릉과학출판사, 1991.3.

[13] S, B, Davis, P, Mermelstein, "Comparison of Parametric Representation for Monosyllabic Word Recognition in Continuously Spoken Sentences", IEEE Trans., on Acoustics, Speech and Signal Processing, Vol. ASSP-28, February 1980.