음소판별필터를 이용한 한국어 단음절 음성인식

Speech Recognition on Korean Monosyllable using Phoneme Discriminant Filters

허 성 필*, 정 현 열**, 김 경 태***
(Sung Phil Hur*, Hyun Yeol Chung**, Kyung Tae Kim***)

요 약

선형판별함수를 이용하여 음소단위의 판별필터를 구성하였다. 음소판별필터를 이용한 음성인식 시스템은 발성구간의 검출에 유용하고, 음성의 구분과 식별을 동시에 시행할 수 있으며 모든 음소를 동일한 인식모델로 취급하는 것이 가능하였다. 이 때 전문가의 경험적 지식을 이용하지 않고 수리적인 반복학습방법으로 시스템을 구성한 것이 특징이다.

모든 음소판별필터는 독립적으로 동작하므로 하나의 음소구간에 대해 복수필터 출력이 발생될 수 있으며, 발성구간의 음소가 탈락하는 경우도 있다. 따라서 본 연구에서는 무게벡터와 패턴벡터와의 내적에 통합계수를 이용하여 최대값을 선택하는 방법으로 다수개의 경합출력을 하나로 통합하였으며, 동시에 시간적인 정보와 중간값필터를 이용하여 탈락과 오인식되는 음소를 보상하므로써 인식율을 향상시켰다.

인식실험결과 모음의 경우 학습용자료에서는 96.5%, 평가용자료에서는 87.6%의 인식율을 얻었고, 자음은 각각 84.0%, 70.8%의 음소인식율을 얻었다.

ABSTRACT

In this paper, we have constructed phoneme discriminant filters [PDF] according to the linear discriminant function. These discriminant filters do not follow the heuristic rules by the experts but the mathematical methods in iterative learning. Proposed system is based on the piecewise linear classifier and error correction learning method.

The segmentation of speech and the classification of phoneme are carried out simutaneously by the PDF. Because each of them operates independently, some speech intervals may have multiple outputs.

Therefore, we introduce the unified coefficients by the output unification process. But sometimes the output has a region which shows no response, or insensitive. So we propose time windows and median filters to remove such problems.

We have trained this system with the 549 monosyllables uttered 3 times by 3 male speakers. After we detect the endpoint of speech signal using threshold value and zero crossing rate, the vowels and consonants are separated by the PDF, and then selected phoneme passes through the following PDF. Finally this system unifies the outputs for competitive region or insensitive area using time window and median filter.

* 한국통신 통신망연구소
** 영남대학교 전자공학과
*** 한남대학교 정보통신공학과
접수일자 : 1994년 11월 16일

Ⅰ. 서 론

일상생활에서 인간은 음성을 의사전달과 정보전달의 수단으로 사용하고 있고, 궁극적으로 기계와의 통신이나 제어도 음성을 통해 하기를 원하고 있다. 그러므로 사람과 기계 사이에 효과적인 통신(man-machine communication)을 하는데 있어, 음성이 가장 자연스러운 정보교환 수단이라는 점에서 자동음성인식에 대한 연구가 활발히 진행되고 있다.

실제로 단어인식 및 연속음성인식에서 화자를 한정하지 않은 대어휘를 사용하는 경우 인식의 단위를 단어나 음절로 할 때 음소나 음절간의 조음결합(co-articulation)이 가장 큰 문제로 지적되고 있다. 또한 대량의 기억용량이 요구되고 처리시간이 비교적 많이 걸린다는 문제점이 있다. 그러므로 음운분류상 가장 작은 단위인 음소를 인식의 기본단위로 하는 것이 유리하리라 생각되며, 이러한 경우 고정밀도의 간결한 음소인식계와 그 설계기법이 요구된다.

음소단위의 음성인식 방법은 Bayes rule[1]이나 HMM (hidden markov model)[2]을 이용한 통계적 방법에서, 현재의 Neural Network[3] 및 Fuzzy System[4]에 이르기까지 여러 가지의 인식알고리즘이 제안되고 있으나, 실시간 처리의 차원에서는 어려움이 많다. 그리고 기존의 음성인식계는 음성의 구분화와 식별화라는 두 단계로 구성되어 있으며 특징파라메터의 선택과 인식규칙에 전문가의 경험적 요소가 많이 포함되어 있고, 분석계 및 특징추출계의 변경에 따른 인식계의 재설계가 비교적 어려운 것으로 알려져 있다.[5]

본 연구에서는 이러한 문제점을 극복하기 위하여 음소마다 독립된 검출구조를 가진 선형판별필터를 이용하여 수리적인 반복학습에 의한 음소 단위의 인식계를 구성하고[6,7], 학습과정에서는 비교적 많은 시간이 요구되나 인식단계에서는 곱셈과 덧셈에 의한 단순한 음소단위의 판별필터를 사용함으로써 실시간 처리가 가능해지고 동시에 보다 높은 인식율을 얻는 방법을 제안한다. 또한 본 시스템은 다단계 식별기구를 반복적 학습에 의해 설계한 점에서 perceptron 모델이나 다층 신경회로망(multilayer neural network) 모델과 유사하지만, 모든 계층의 파라메터가 변할 수 있고 각 음소마다의 재조정과 음성학적인 지식도입이 가능한 점에서 perceptron 모델과 다르며, 다층 신경회로망은 은닉층(hidden layer)에서 일어나는 기능 파악이 어려운 반면, 제안한 방법은 음소의 검출과 통합에 있어 각 부의 기능파악이 명확한 장점이 있다. 그리고 본 시스템의 평가는 한국어 역순 사전중의 단어를 출현 누적빈도[13]를 고려한 총 549개의 단음절에 대해 남성화자 3인이 방음실에서 랜덤하게 3회씩 발성한 4941개 음성자료를 사용하였다.

Ⅱ. 음소판별필터의 구성 및 설계

2.1 선형판별함수[8]

2개의 카테고리를 가진 선형판별함수 g(X)는

$$g(X)=\sum_{i=1}^{d} w_i x_i + w_{d+1} = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d + w_{d+1} \quad (1)$$

이고, 여기서 x_i는 입력패턴을 w_i는 무게벡터를 나타낸다. 이 때 TLU(threshold logic unit)의 출력을 i_0라고 두면, $g(X)>0$ 일 때 $i_0=1$ 이 되고, $g(X)<0$ 일 때 $i_0=-1$ 인 경우 $g(X)=0$ 인 초평면(hyperplane)에 의해 R_1과 R_2 영역으로 분리되어 진다. 그림 1은 R=2인 경우의 패턴분류기의 기본모형이다.

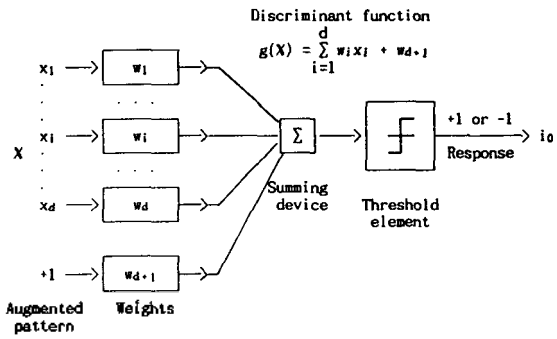


그림 1. TLU의 기본 모형
Fig. 1 Basic model for the threshold logic unit (TLU).

2.2 음소판별필터의 설계

각 프레임별 특징벡터를 입력패턴으로 하고

$$X=(x(n+K_1), x(n+K_1+1), \cdots, x(n+K_2), 1) \quad (2)$$

이때의 무게벡터가

$$W_i = (w(K_1), w(K_1+1), \cdots, w(K_2), w_0) \quad (3)$$

일 때, 입력패턴이 선형적으로 분리 가능하다면 오차정정 학습법(error correction learning)[9]을 이용한 음소판별필터의 무게벡터는 반복적인 학습방법으로 구할 수 있으며 그 알고리즘은 그림 2와 같다.

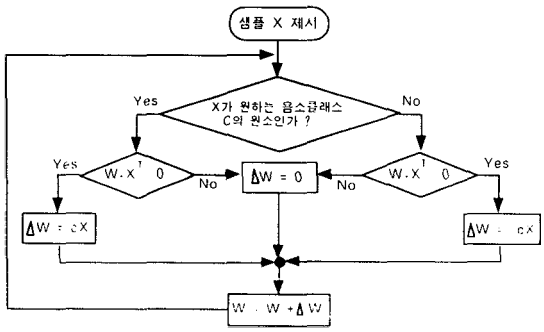


그림 2. 음소판별필터 설계를 위한 흐름도
Fig. 2 Flow chart of PDF design.

여기서 원하는 음소패턴클래스를 C, 그 외 모든 패턴클래스는 $\overline{C}$, 그리고 $c(>0)$는 학습계수를 나타낸다.

음소클래스 C_i에 대한 판별필터의 제 n번째 프레임의 출력을 $d_i(n)$, 그 때의 입력을 $y_i(n)$ 이라고 하면, 오차정정학습에 의해 구한 무게벡터와 임의의 입력과의 관계식은

$$y_i(n) = \sum_{k=K_{i1}}^{K_{i2}} W_i(k) \cdot X(n+k) + w_{i0} \quad (4)$$

그 출력응답은

$$d_i(n) = \mathrm{sgn}\,(y_i(n)) \quad (5)$$

이다. 이 때

$$\mathrm{sgn}(z) = \begin{cases} 1 & (z > 0) \\ -1 & (z \le 0) \end{cases} \quad (6)$$

되며, $X(n)$은 특징벡터, $W_i(n)$과 w_{i0}는 무게벡터와 문턱 값(threshold value)을, 그리고 $[K_{i1}, K_{i2}]$는 시간 창(time window)의 길이를 $(K_{i1} \le 0 \le K_{i2})$ 나타낸다. 이 때 $d_i(n) = 1$ 이 되는 구간에서 음소 C_i가 인식되어 진다.

그림 3은 발성음 "가(ka)"에 대해 미리 구해진 모음별 무게벡터를 적용한 후 threshold element를 통과한 각 판별필터의 출력결과를 보여 주고 있다. 여기서 모음 /a/인 판별필터에서는 음성 "가"의 'ㅏ' 영역이 양(+)의 응답을 보여 주며 나머지 판별필터에서는 음(−)의 응답을 나타내므로 양의 영역을 취함으로써 모음 'ㅏ'를 구분할 수 있으며 동시에 식별도 가능하다.

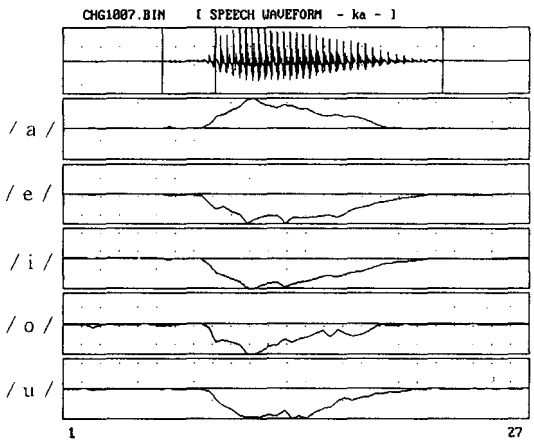


그림 3. 각 음소별 판별필터를 통해 나온 출력
Fig. 3 Temporal patterns of discriminant filter's output.

Ⅲ. 통합계수설정과 출력통합

3.1 판별필터 출력통합의 필요성

각 음소판별필터는 전부 독립적으로 동작하기 때문에 하나의 음소구간에 대해 복수필터에서의 출력이 나올 수 있다. 또한 음소판별필터는 음소클래스에 대한 유사도에 상당하는 무게벡터와 패턴벡터의 내적으로 표현되며, 각 무게벡터의 성분은 상대치로써 주어져 있으므로 필터 사이의 내적 값을 그대로 비교할 수 없다. 그래서 내적치 $y_i(n)$에 통합계수 $a_i(>0)$를 이용하여 $\{a_i y_i(n)\}$ 중에서 최대 값을 선택하는 방법으로 다수의 출력을 통합할 수 있다.

즉 복수출력은

$$i = \arg\max_i a_i y_i(n) \quad (7)$$

일 때 음소클래스 C_i에 유일하게 된다. 이 때 모든 $i \neq j$ 경우

$$d_j(n) = 0 \qquad (8)$$

으로 치환할 수 있다.

따라서, 판별필터와 출력통합을 포함한 기법은 구분적 선형(piecewise linear)인 식별경계를 구할 수 있다.

3.2 통합계수 설정법

통계적 수단에 의한 통합계수의 설정은 입력 패턴 분포의 가정을 필요로 하므로 여러 클래스의 판별을 최적으로 설정한다는 것은 어렵다. 그러므로 통합계수의 설정에는 반복학습을 사용한다. 본 실험에서는 모든 i에 대해 $a_i > 0$을 유지하므로 학습계수 c가 입력패턴에 따라서 변화하는 학습알고리즘을 제안하였으며, 그 방법은 다음과 같다.

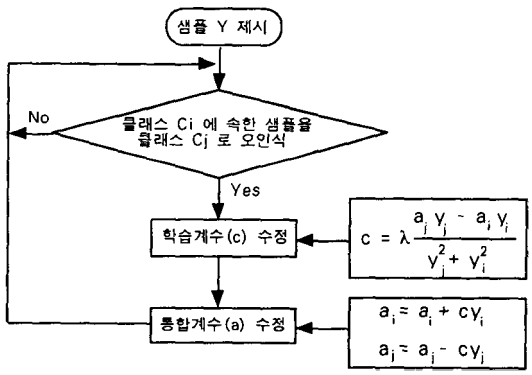


그림 4. 통합계수 결정방법
Fig. 4 Desicion method of unified coefficients.

이 알고리즘에서 $0 < \lambda \leq 1$ 이고, 통합계수의 초기값(initial value)이 양(+)의 값이면, $y_i > 0$, $y_j > 0$인 어떠한 패턴이 입력되어도 a_i, a_j는 항상 양의 값을 유지함으로 출력통합에 적합하다.

3.3 출력통합 방법

경합출력이 존재하는 경우의 출력통합 방법으로는 간단히 판별필터를 통해 나온 출력의 최대 값을 선택하는 방법이 있으며, 절대적인 비교를 위해 통합계수라는 가중치를 적용해 하나의 값으로 유일화하는 방법이 있다. 그러나 여전히 불감영역과 오인식되는 영역이 존재하므로 시간정보를 이용하였으며 이 때 중간값필터를 사용하여 오인식되는 음소를 완전히 제거할 수 있다.

3.3.1 시간정보를 이용한 출력통합 및 중간값필터

불감영역이나 시간적인 불일치를 동반하는 출력을 통합하기 위해서는 현재 관측하고 있는 프레임에서 전후 판별필터의 출력을 이용해 시간축 상의 억제를 가하여 출력을 통합한다. 그래서 가중치최대값 선택법을 시간축 상에 확장한 출력통합법을 제안한다. 이것은 분석프레임을 중심으로 전후 수 프레임에 걸친 시간창(time window) [−K, K]에 가중치(weight)를 가해

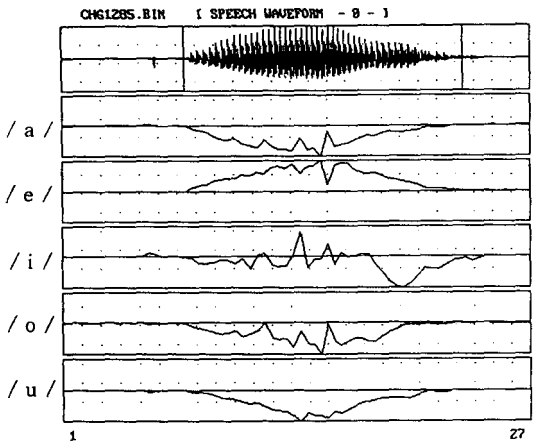
$$\max_{-K \leq k \leq K} \{a_i(k)\, y_i(n+k)\} \qquad (9)$$

인 i를 찾고, 이 때 $i \neq j$ 모든 경우

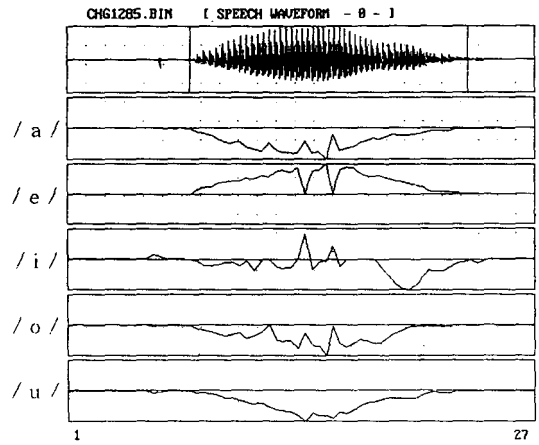
$$d_j(n) = 0 \qquad (10)$$

으로 바꾼다.

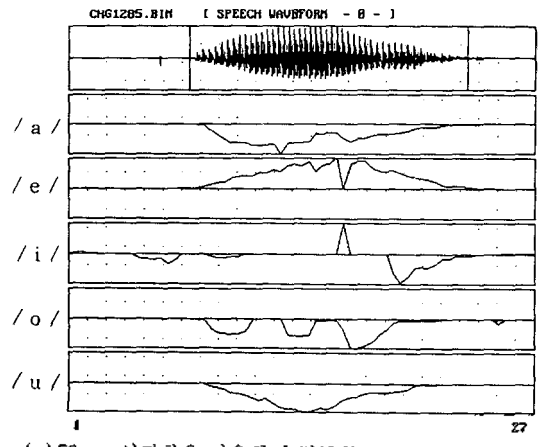
그림 5는 출력통합 과정의 일반적인 예로서 음절 "애"에 대해 출력통합을 하지 않은 경우[그림 5-(a)] 판별필터 /e/와 /i/에서 복수출력이 발생하였다. 이 때 (b)와 같이 통합계수를 이용한 출력중 일부 구간에서는 오인식이 발생하였으며, (c)의 시간창을 적용한 경우에도 여전히 일부 구간에서 오인식이 되는 음소영역이 존재하는 경우가 있었다. 음성발성에 있어 모음은 정상적인 발성인 경우 최소한 수십 ms 이상 동일 음소가 지속되므로 관측점에서 전후 정보를 이용하여 탈락이나 오인식 되는 음소를 보상할 수 있다. 또한 다른 음소가 삽입되어 오인식되는 것을 방지하기 위해 여러 프레임의 출력값을 크기 순으로 정렬한 후 중간값을 취하는 방법을 제안한다. 이 때 중간값 필터 (median filter)를 적용한 인식방법이 그림 6에 나타나 있다. 통합된 출력결과에 대해 시간정보를 포함한 중간값필터를 중첩되게 1 프레임씩 이동시키고, 크기 순으로 정열(sorting)한 후 가운데 값을 취함으로 임펄스성 잡음(impulse noise)을 제거하는 방법과 유사하게, 삽입된 음소가 제거되어 최종적으로 원하는 음소가 인식된다.



(a) 출력통합을 하지 않은 경우
case A : non-unified output.



(b) 통합계수를 이용한 출력통합
case B : unified output using weighted maximum-value for 10ms.



(c) 50 ms 시간창을 이용한 출력통합
case C : output unification using 10ms time windows.

그림 5. 판별필터를 통해 나온 출력 통합 과정
Fig. 5 The process of output unification

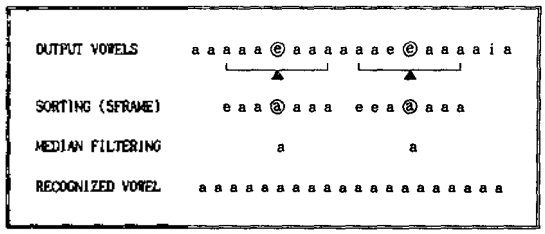


그림 6. 중간값 필터를 이용한 모음 인식 예
Fig. 6 An Example of vowel recognition using median filter(/a/).

Ⅳ. 음성자료 및 분석

4.1 음성자료

한국어 역순 사전에서 단어를 음절별 출현 누적빈도를 고려한 총 549개의 단음절에 대해 한국인 성인 남성화자 3인이 방음실에서 랜덤하게 각 3회씩 자연스럽게 발성한 총 4941개 음성자료를 학습 및 인식에 사용하였다. 이러한 음성자료는 음성파형, 에너지, 유/무성, 비성, 마찰성 및 스펙트럼 등의 음성특징 변화를 이용하여 시찰에 의해 음소단위로 레이블링하였다.[10]

4.2 분석방법

음성분석은 그림 7에 보인 바와 같이 4.5 KHz LPF를 통과한 음성신호는 10 KHz, 12 bit의 A/D 변환기를 통해 이산데이터로 변환된다. 특징추출은 10 ms를 1 프레임으로 한 25 Channel BPF (Q=6, 1/6 Octave, 중심주파수 250 Hz~4000 Hz)[11]를 통과시켜 나온 출력의 대수에너지 값을 특징파라메터로 하여 각 음소별 판별필터를 설계한 후, 입력음성에 대해 각각의 음소를 검출해 내며 판별필터를 통해 나온 출력에 대해서는 출력통합을 한 후 음성을 인식하게 된다.

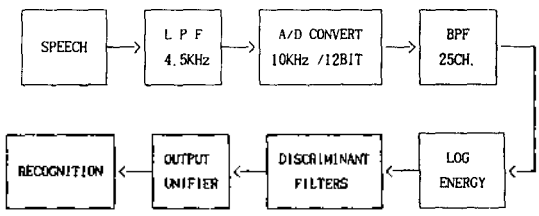


그림 7. 분석방법에 대한 흐름도
Fig. 7 The block diagram of analysis.

V. 인식실험 및 결과

인식실험은 음성신호에서 우선 발성구간을 검출한 후, 모음과 자음으로 분리하였다. 모음의 경우 7개의 판별필터를 설계하였으며, 자음에 대해서는 조음양식과 조음위치에 따라 7개의 클래스로 분류한 후 세부음소인식에서는 다단계 인식방법을 사용하였다. 3인 화자의 2회 발성을 표준패턴으로 사용하여 각 음소별 판별필터를 구성하였으며, 나머지 1회 발성을 입력으로 하여 인식실험을 실시하였다.

본 연구에서 사용한 음소판별필터와 출력통합기구를 조합한 음소인식계는 그림 8과 같이 구성한다.

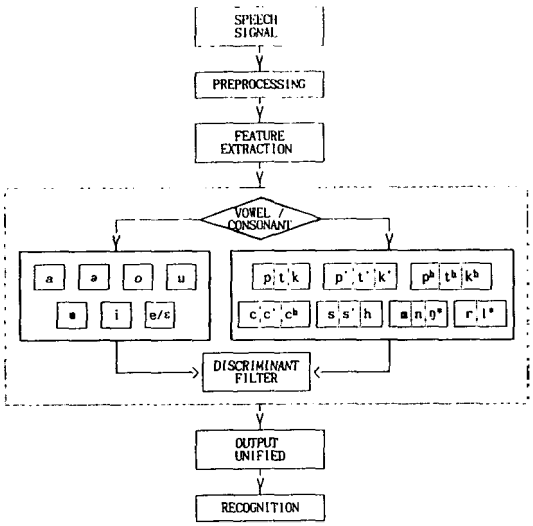


그림 8. 음소인식에 대한 전체 흐름도
Fig. 8 The schematic diagram of phoneme recognition.

5.1 음성구간 검출 및 자음과 모음분리

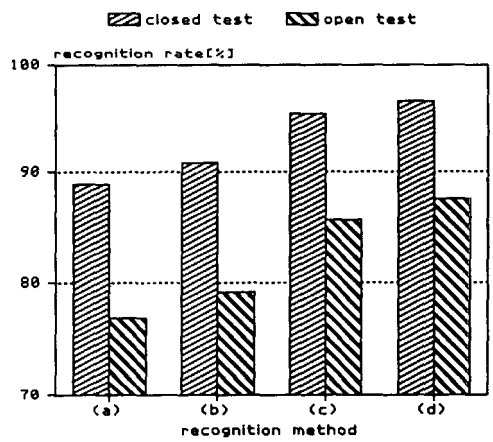
인식의 전단계로 모음과 자음구간의 구별은 먼저 문턱 값(threshold value)에 의한 영교차 율(zero crossing rate)을 이용하여 음성구간을 검출하였다. 그리고 선형판별필터를 이용하여 자/모 구간을 분리하여 각 음소에 대한 인식실험을 실시하였다.

5.2 모음인식

모음인식실험은 '아, 어, 오, 우, 으, 이, 에, 애' 8개의 모음에 대해 '에'와 '애'는 보통 사람이 뚜렷하게 구별하여 발성하지 않는 모음이므로 같은 발성으로 하며 자주 사용되는 '에'를 택하여 7개의 판별필터를 설계하였다.

각 프레임에서 최대 값을 취하는 간단한 출력통합 방법을 하여 얻은 결과는 그림 9의 (a)와 같이 88.9%의 인식율을 얻었다. (b)는 1 프레임에 대해 통합계수를 적용한 결과를 나타내는데 90.8%의 인식율을 얻었으나, 오인식되는 음소나 어느 클래스에도 식별되지 않는 음소가 많이 발생하였다. 이러한 경우 시간적인 정보를 이용하여 인식율을 향상시켰으며, 불감영역도 제거하였다.

모음의 경우 정상적인 발성인 경우 최소한 수 백 ms 이상 동일음소가 지속되므로 관측점에서 전후 정보를 이용하여 탈락이나 오인식되는 음소를 보상할 수 있다. 또한 다른 음소가 삽입되어 오인식되는 것을 방지하기 위해 5 프레임의 출력값에 대해 중간값 필터를 적용한 결과, 50ms의 시간창을 이용해서 얻은 인식율 [그림 9의 (c)]보다, (d)와 같이 96.5%로 인식율을 향상시켰으며, 음소의 탈락도 완전히 제거하였다.



(a) maximum value (b) weighted value
(c) time windows + (b) (d) median filter + (c)

그림 9. 출력통합 방법에 대한 인식율 비교
Fig. 9 Comparison of recognition scores.

출력통합 방법을 다르게 했을 때의 인식율 변화에서, 제안한 시간정보와 중간값필터를 동시에 적용한 경우의 인식율이 가장 좋게 나타남을 알 수 있다.[12]

인식실험결과 모음의 경우 제안된 방법을 적용한 경우 표 1에서와 같이 학습용 자료에서는 96.5%, 평

표 1. 중간값 필터를 이용한 모음 인식율
Table 1. Recognition scores by median filtering for vowels.

(50 msec) Closed test / Open test

IN.\OUT.	ㅏ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ	ㅔ	OMIT
ㅏ (a)	97.7	1.9	0.1	0.1	0.0	0.0	0.2	0.0
	91.8	6.7	0.7	0.2	0.1	0.1	0.6	0.0
ㅓ (ə)	0.5	96.8	2.5	0.1	0.1	0.0	0.0	0.0
	1.9	90.7	6.5	0.3	0.3	0.2	0.2	0.0
ㅗ (o)	0.0	3.4	95.9	0.4	0.2	0.1	0.0	0.0
	0.2	3.6	85.9	7.9	2.3	0.0	0.2	0.0
ㅜ (u)	0.3	0.1	4.2	94.4	0.4	0.5	0.1	0.0
	1.0	0.3	11.7	82.9	3.0	1.2	0.0	0.0
ㅡ (w)	0.1	0.2	0.6	2.0	96.6	0.2	0.3	0.0
	1.5	0.3	6.5	10.4	79.7	1.3	0.4	0.0
ㅣ (i)	0.3	0.1	0.1	0.3	0.5	96.0	2.7	0.0
	0.1	0.1	0.6	9.6	2.8	84.9	2.0	0.0
ㅔ (e)	0.2	0.8	0.3	0.2	0.0	1.4	97.1	0.0
	0.1	2.6	1.9	0.2	0.3	3.4	91.5	0.0

overall recognition scores : 96.5% / 87.6%

가용 자료에서는 87.6%의 인식율을 얻었다. 여기서 모음 'ㅗ'와 'ㅜ'간의 상호 오인식이 발생하였으며, 'ㅡ'의 인식은 다소 낮게 나타났다. 이 때 판별필터의 설계에 사용한 학습용자료 양의 부족으로 불특정화자에 대한 인식율이 전체적으로 낮게 나타났다.

5.3 자음인식

자음에 대해서는 조음양식에 따라 파열음(plosive)[/ㅂ, ㄷ, ㄱ/, /ㅃ, ㄸ, ㄲ/, /ㅍ, ㅌ, ㅋ/], 파찰음(affricate)[/ㅈ, ㅉ, ㅊ/], 마찰음(fricative)[/ㅅ, ㅆ, ㅎ/], 유음(liquid)[/ㄹ, ㄹ*/], 비음(nasal)[/ㅁ, ㄴ, ㅇ*/]군으로 나누었으며, 파열음인 경우 조음위치에 따라 연음, 경음, 기음으로 분류하여 음소판별필터를 구성하였다.[13]

음소판별필터 설계시 문제점은 가급적 학습시간을 줄이면서 인식율을 높이는 것이다. 학습시간을 줄이는 방법으로는 여러 가지가 있다. 전체 오차가 가장 적은 상태, 즉 최적해를 찾아가는 수렴속도를 증가시키는 방법으로 모멘텀 항(momentum term)을 사용하거나, 학습자료를 선별하는 방법이 있다. 본 실험에서는 학습시간을 결정하는 학습횟수를 줄이기 위해 자음의 인식에 있어 비슷한 특성을 가지는 데이터를 묶어서 여러 개의 부분집합으로 나누고, 각 부분집합에 속한 자료만을 학습하는 방법을 사용하였다. 이 때 각 그룹에 속하는 자료만을 학습시키므로 학습시간을 감소시킬 수 있고, 새로운 어휘에 대해서도 손쉽게 확장할 수 있다는 장점이 있다. 동시에 선형분리도를 높이고 무게벡터의 수렴성을 향상시킬 수

표 2.자음에 대한 인식율
Table 2.Recognition scores for consonants.

Closed test / Open test

consonant	PLOSIVE								
	ㅂ (p)	ㄷ (t)	ㄱ (k)	ㅃ (p')	ㄸ (t')	ㄲ (k')	ㅍ (p^h)	ㅌ (t^h)	ㅋ (k^h)
Rec.(%)	88.8	84.5	71.7	81.7	89.2	69.0	82.7	71.0	72.7
	74.5	74.6	72.4	78.3	87.3	67.0	89.7	69.8	90.5

AFFRICATE			FRICATIVE			LIQUID		NASAL		
ㅈ (c)	ㅉ (c')	ㅊ (c^h)	ㅅ (s)	ㅆ (s')	ㅎ (h)	ㄹ (r)	ㄹ (l*)	ㅁ (m)	ㄴ (n)	ㅇ (η*)
90.2	95.8	82.1	80.2	86.4	73.1	87.9	91.2	93.6	91.3	81.3
87.5	75.6	71.4	68.2	71.0	68.2	63.0	70.7	86.1	72.2	64.6

Total recognition scores : $\frac{84.0\%}{70.8\%}$

있다. 다단계 인식방법에 의해 7개의 소규모 망으로 분류한 후 각 그룹에 속하는 음소를 분류한 후 세부 음소로 인식하는 실험에서 음소판별필터 설계는 각 클래스의 분리도를 90% 이상을 만족하는 무게벡터를 이용하였으며, 학습용자료와 평가용자료에 대한 인식결과를 표 2에 나타내었다. 여기서 자음군의 분류에 있어서는 유음, 비음, 마찰음군은 비교적 검출이 잘 되었으나, 파열음과 파찰음군은 오인식이 심하게 나타났으며 특히 파열음은 군 내부에서의 오인식이 크게 발생하였다.

Ⅵ. 결 론

BPF의 출력을 특징파라메터로 하여 선형판별함수에 의한 음소판별필터를 구성하였다. 음소판별필터를 이용한 음성인식 시스템은 발성구간의 검출에 유용하고, 음성의 구분과 식별을 동시에 시행할 수 있으며 모든 음소를 동일한 인식모델로 취급하는 것이 가능하였다. 이 때 전문가의 경험적지식을 이용하지 않고 수리적인 반복학습방법으로 시스템을 구성한 것이 특징이다.

인식실험은 모음과 자음으로 분리한 후 모음에 대해서는 7개의 판별필터를 설계하였으며, 자음에 대해서는 음성학적 특성인 조음양식과 조음위치에 의한 집합으로 분류한 후 세부음소로 인식하는 방법을 사용하였다. 성인 남성화자 3인이 각 3회씩 발성한 549단음절에 대해 음소단위의 인식실험결과 모음의 경우 학습용자료에서는 96.5%, 평가용자료에서는 87.6%의 인식율을 얻었고, 자음의 경우 세부음소인식에서는 각 각 84.0%, 70.8%의 음소인식율을 얻었다.

이러한 인식계는 모음의 경우 다소 높은 인식율을 얻을 수 있었으나 자음군에 있어서는 유음, 비음, 마찰음군은 비교적 검출이 잘 되었으나, 파열음과 파찰음군은 오인식이 심하게 나타났으며 특히 파열음은 군 내부에서의 오인식이 크게 발생하였다.

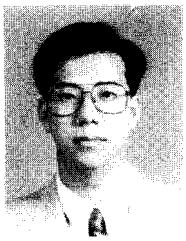
향후 음소환경을 적극적으로 고려하여 음소군을 분류하고, 자음특징을 보다 잘 반영한는 복수개의 파라메터로 판별필터를 설계한 후, 시간정보를 충분히 이용하여 단어인식 및 연속음성인식에 적용해 보고자 한다.

참 고 문 헌

1. 中川聖一, 確率モデルによる音聲認識, 電子情報通信學會, pp. 12-17, 1988.
2. L. R. Rabiner and B. H. Jung, "An Introduction to Hidden Markov Models," IEEE ASSP, Vol. 3, No. 1, pp. 4-16, Jan, 1986.
3. B. Widrow and M. A. Lehr, "30 Years of Adaptive Neural Networks : Perceptron, Madaline, and Back-propagation," Proc. IEEE Vol. 78, No. 9, Special Issue on Neural Networks 1.
4. N. Hataoka, A. Amano, T. Aritsuka and A. Ichikawa, "Large Vocabulary and Speech Recognition using Neural-Fuzzy and Concept Networks," IC-ASSP '90, pp. 513-516, 1990.
5. C. J. Weistein, S. S. Mccandless, L. F. Monsshein and V. W. Zue, "A System for Acoustic Phonetic Analysis of Continuous Speech," IEEE, ASSP, Vol. 23, No. 1, pp. 54-67, 1975.
6. S. Moriai, S. Makino and K. Kido, "Phoneme Recognition in Continuous Speech using Phoneme Discriminant Filters," ICASSP '86, pp. 2251-225.
7. 盛合 敏, 數理的な知識表現に基づく音聲認識系の構成法に關する研究, 東北大學博士學位論文, 1988.
8. Nils J. Nilsson, *The Mathematical Foundation of Learning Machines*, Morgan Kaufman Publishers, pp. 15-41, 1990.
9. J. M. Zurada, *Introduction to Artificial Neural Systems*, Wese Publishing Company, pp. 93-155, 1992.
10. H. Y. Chung, S. Makino and K. Kido, "韓國語語頭破裂子音の認識," 日本音響學會 1-2-21, Mar, 1989.
11. K. T. Kim, J. Miwa and K. Kido, "Recognition of

Isolated Korean Digits using Bandpass Filters Based on FFT," J. Acoust. Soc. Japan E Vol. 4, No. 4, 1983.
12. 허성필, 정현열, 김경태, "음소판별필터를 이용한 한국어 모음 인식에 관한 연구," 제10회 음성통신 및 신호처리 워크샵 논문집, pp. 305-310, 1993. 8.
13. 許熊, 國語音韻學, 正音社, pp. 190-226, 1985.

▲허 성 필 1966년 6월 18일 생



1991년 2월 : 영남대학교 전자공학과
1993년 8월 : 영남대학교 대학원 전자공학과(석사)
1993년 11월 : 한국통신 통신망 연구소

▲정 현 열 1951년 11월 26일 생
1975년 2월 : 영남대학교 전자공학과
1981년 2월 : 영남대학교 대학원 전자공학과(석사)
1989년 4월 : 일본 東北대학 대학원 정보공학과(공학박사)
1989년 3월~현재 : 영남대학교 전자공학과 부교수
1992년 7월~1993년 8월 : 미국 Carnegie Mellon Uni. Robotics 연구소 Visiting Research Scholar

▲김 경 태 1949년 5월 9일 생
1972년 2월 : 경북대학교 전자공학과
1980년 8월 : 연세대학교 대학원 전자공학과(석사)
1985년 3월 : Tohoku Univ., Japan 전기 및 통신 전공(박사)
1986년 2월~1991년 2월 : 한국전자통신연구소 신호처리 연구실
1991년 2월~현재 : 한남대학교 정보통신공학과