

유, 무성음 및 묵음 식별에 관한 연구

A Study on the Voiced, Unvoiced and Silence Classification

* 김 명 환 (M. H. Kim)
 ** 김 순 협 (S. H. Kim)

ABSTRACT

This paper describes on the voiced-unvoiced-silence Classification of the speech for speech recognition. Pattern recognition technique used to classify a given speech segment. The measurement parameters are the zero crossing rate, log energy, the correlation between adjacent speech samples, the first predictor coefficient from a 12-pole linear predictive coding (LPC) analysis, the energy in the prediction error.

The speech segment is assigned to a particular class based on a minimum distance rule obtained under the assumption that the measured parameters are distributed according to the multidimensional Gaussian probability density function.

Finally, optimum result is obtained with the combination using the zero crossing rate, the first predictor coefficient, the energy in the prediction error and classification error rate is less than 1%.

要 約

본 논문은 한국어 음성 인식을 위한 유성음, 무성음, 묵음 식별에 관한 연구이다. 주어진 음성 구간을 3가지 음성 신호 부류로 식별하기 위하여 패턴 인식 방법을 사용하였다.

여기에 사용한 분석 파라메타는 음성 신호의 영교차율, 대수 에너지, 정규화 된 첫번째 자동 상관 계수, 선형 예측 분석에서 얻은 첫번째 예측 계수, 그리고 예측 오차의 에너지이다.

한편 측정된 파라메타들이 다차원 가우스 확률 밀도 함수에 따라 분산되었다는 가정하에서 얻어진 최소 거리 법칙에 기본을 두고 음성 구간을 결정하였다. 측정된 파라메타들을 여러가지 방법으로 조합하여 식별한 결과 영교차율, 첫번째 예측 계수, 예측 오차의 에너지를 측정 파라메타로 사용했을때 1% 보다 적은 식별 오차율을 얻었다.

I. 서 론

주어진 음성 구간을 유성음, 무성음, 묵음으로 식별하기 위하여 일반적으로 피치 분석으로 수행되어져 왔으나 몇가지 단점이 있다.¹⁾

첫째, 유·무성음을 결정하는 데 단순한 특징에 기본을 두었다. 즉 유성음은 주기적인 음성 파형이라는 것이다. 일반적으로 유성음은 거의 주기적이지만 성대 진동의 특이성이나 분절음에서의 갑작스러운 변화는 유성음일지라도 비주기

* 대우전자 중앙연구소 연구원
 ** 광운대학 전자계산기 공학과 교수

적인 음성 파형을 만든다.

둘째, 음성 합성 분야에서는 피치 검출로 유·무성음을 결정한다는 것³⁾은 가능하나 음성 분할, 음성 인식과 같은 응용 분야에 있어서는 불필요한 복잡성 뿐만 아니라 유·무성음 사이의 경계 부분에서 나쁜 수행을 초래한다.

따라서 본 논문은 한국어 음성 인식을 위하여 주어진 음성 구간을 유성음, 무성음, 묵음 식별에 패턴 인식 방법을 사용하였다. 또한 이 방법에 필요한 파라메타는 영교차율, 대수에너지, 자동상관 계수, 12차 선형 예측 분석에서 얻은 첫

번째 예측 계수, 그리고 예측 오차의 에너지이다. 이 방법에서 3가지 음성 신호 부류의 각 파라메타에 관하여 Mahalanobis 거리 측정법⁴⁾은 음성 신호를 구분하기 위하여 만든 파라메타들의 집합으로 계산할 수 있고 이때 3가지 음성 신호 부류는 계산된 거리가 최소일 때를 선택하여 결정하였다. 그리고 본 연구에 사용된 음성 데이터는 광운대학 디지털 신호 처리 연구실의 시스템에 의해 전처리된 음성 신호이며 이에 대한 전체적인 시스템은 그림 1과 같다.

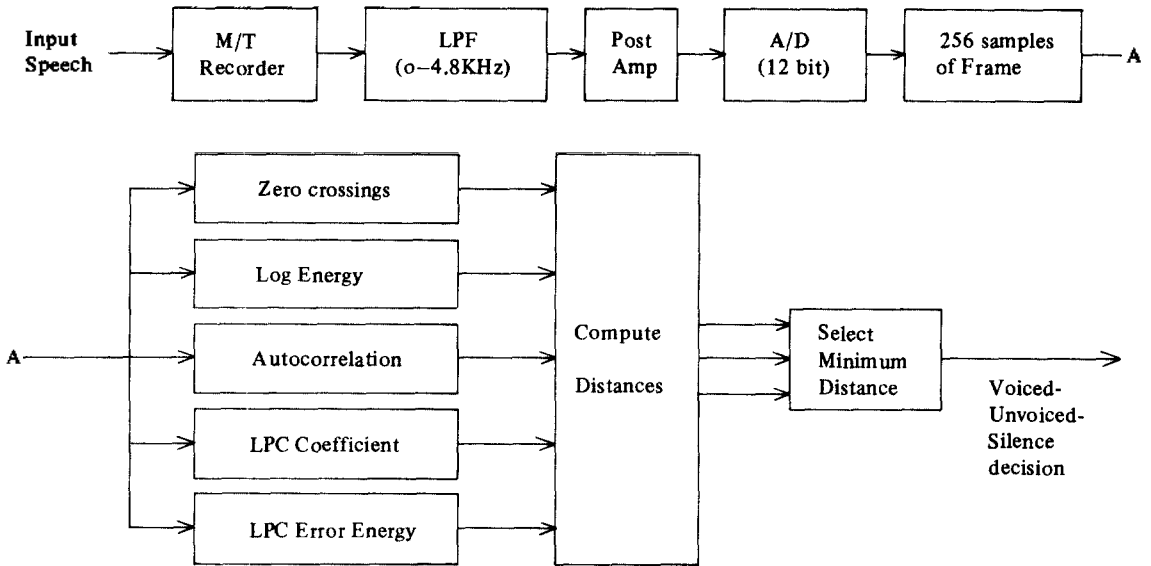


그림 1. 한국어 음성 식별에 관한 시스템 블록도
 Fig. 1. A block diagram of the Korean speech classification system.

II. 음성 신호의 분석 이론

2-1. 영교차율

영교차율에 대한 일반수식⁵⁾은

$$W(n) = \sum_{m=0}^{n-1} |Sgn\{x(m)\} - Sgn\{x(m-1)\}| \quad (1)$$

$$Sgn\{x(m)\} = \begin{cases} 1; & x(m) \geq 0 \\ -1; & x(m) < 0 \end{cases}$$

$$W(n) = \begin{cases} \frac{1}{2N}; & 0 \leq n \leq N-1 \\ 0; & \text{otherwise} \end{cases}$$

이고 음성 신호에 응용해 보면 유성음은 성문에서 주기적인 공기 흐름에 의하여 상대의 여기로써 생성되며 유성음의 에너지는 성문의 파형에 의해 나타나는 스펙트럼의 하향 때문에 약3(KHz) 이하에 집중되어 있다. 반면 무성음은 성도 내부에서 잡음원에 의해 상대의 여기 때문에 생성되며 잡음원의 스펙트럼은 평편한 반면 에너지의 대부분이 높은 주파수에 집중되어 있다.

높은 주파수는 높은 영교차율을, 낮은 주파수는 낮은 영교차율을 가지며 각 음성 신호에 대한 영교차율의 확률 밀도 함수는 그림 2에 나타내었다.

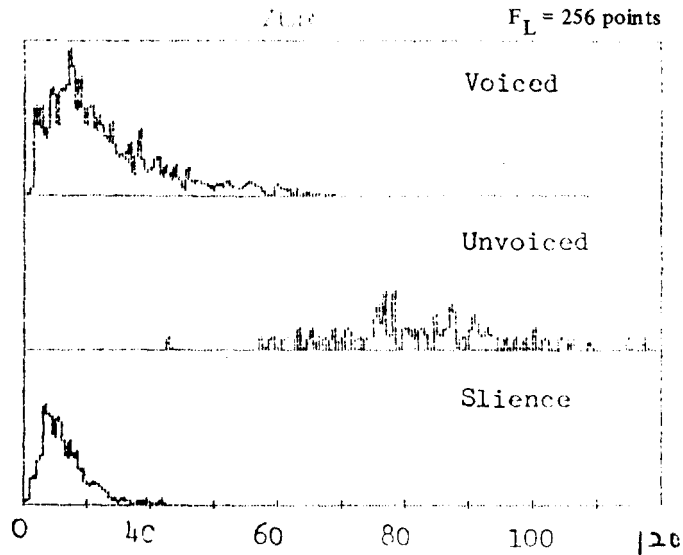


그림 2. 영교차율에 대한 확률밀도함수.
 Fig. 2. The probability density function for the zero-crossing rate.

2-2. 대수 에너지

대수 에너지는 샘플값 자승의 합에 대수를 취한 것으로 다음과 같다".

$$LE = 10 \log \sum_{m=1}^N x^2(m) \dots \dots \dots (2)$$

에너지 파라메타는 마이크로폰이나 증폭기의 감

도라든지 A/D변환기의 양자화 특성과 같은 요소에 의해 영향을 받지만 일반적으로 유성음의 에너지는 무성음의 에너지보다 높고 무성음의 에너지는 묵음의 에너지보다는 높다.

그림 3은 각 음성 신호에 대한 대수 에너지의 확률 밀도 함수를 나타내었다.

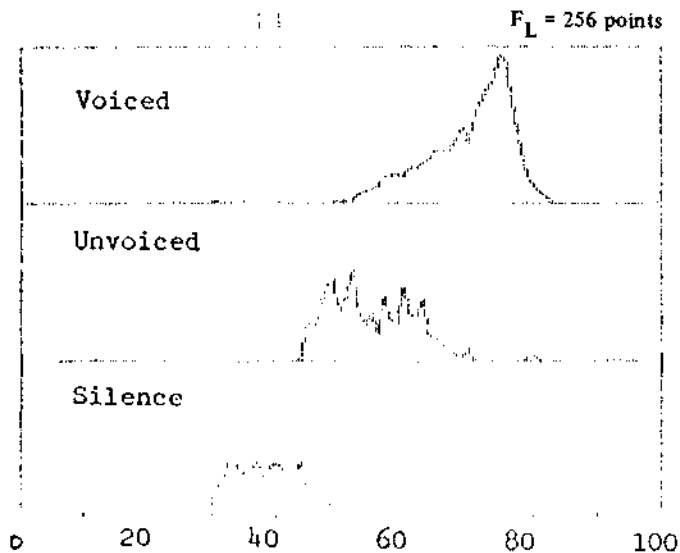


그림 3. 대수에너지에 대한 확률밀도 함수.
 Fig. 3. The probability density function for the log energy.

2-3. 자동상관법
이산 시간 신호의 자동상관 함수"는

$$R(k) = \sum_{m=-\infty}^{\infty} x(m) \cdot x(m+k) \dots\dots\dots(3)$$

이고, 정규화된 자동상관 계수는

$$C_1 = \frac{\sum_{m=1}^N (x(m) \cdot x(m-1))}{\sqrt{(\sum_{m=1}^N x^2(m)) (\sum_{m=0}^{N-1} x^2(m))}} \dots\dots(4)$$

이다. 자동상관 계수는 인접한 샘플들 사이의 상관관계이며 정의에 의하여 -1에서 1 사이에 변한다.

유성음은 낮은 주파수에서 에너지 집중으로 유성음의 인접한 샘플들은 높은 상관관계를 가지며 1에 근접해 있는 반면 무성음에 대해서는 0에 근접해 있다. 그림 4는 자동상관 계수에 대한 각 음성 신호의 확률 밀도 함수를 나타내었다.

2-4. 선형 예측법

선형 예측 분석의 기본 개념은 한 음성 샘플을 과거 음성 샘플들의 선형 결합으로 근사화할 수 있다"는 데 있으며 실제 음성 샘플과 선형 예측된 샘플들 사이의 차의 제곱의 합을 최소화 하는 조건으로 예측 계수의 특징 집합으로 결정한다.

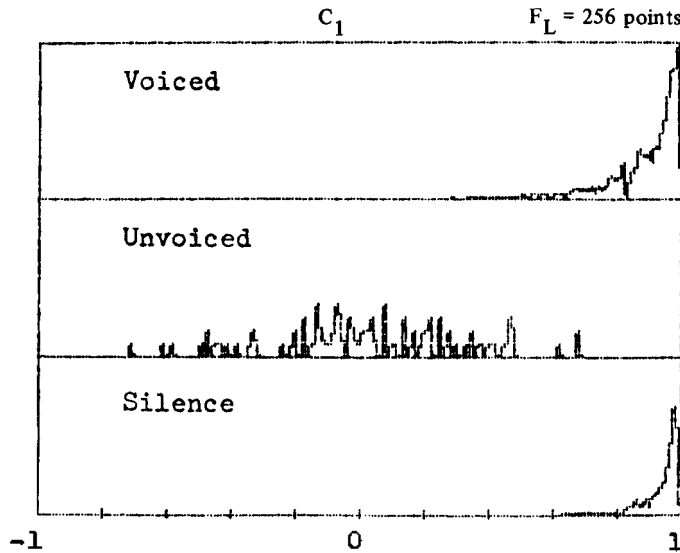


그림 4. 첫번째 자동상관 계수에 대한 확률밀도 함수.
Fig. 4. The probability density function for the auto correlation coefficient.

또 음성 합성 모델로부터 음성의 성대, 성도, 방사 특성의 정상 상태 함수를 시변 디지털 필터로 표현하면

$$S(n) = \sum_{k=1}^p a_k \cdot S(n-k) + G \cdot \sum_{l=0}^q b_l \cdot U(n-l),$$

$$b_0 = 1 \dots\dots\dots(5)$$

이고, 주파수 평면에서 해석하기 위하여 Z변환을 하고 입력과 출력의 비인 전달 함수로 표시하면 다음 식과 같다.

$$H(Z) = \frac{S(Z)}{U(Z)} = G \cdot \frac{1 + \sum_{l=1}^q b_l \cdot Z^{-l}}{1 - \sum_{k=1}^p a_k \cdot Z^{-k}} \dots\dots(6)$$

식(6)에서 $b_l = 0$; all-pole model (AR 모델)로 써,

$$S(Z) = \frac{G}{1 - \sum_{k=1}^p a_k \cdot Z^{-k}} \cdot U(Z) \dots\dots\dots(7)$$

이고, 식(6)에서 $a_k=0$; all-zero model (MA 모델)로써,

$$S(Z) = G \cdot (1 + \sum_{i=1}^p b_i \cdot Z^{-i}) \cdot U(Z) \dots\dots\dots (8)$$

이다. 또한 식(6)과 같이 $a_k \neq b_i \neq 0$; pole-zero model (ARMA 모델)은 음성 생성과정시 비음이나 마찰음 등에 반드시 존재하는 영점의 영향까지도 고려하는 분석 방법이다.⁷⁾ 그러나, all-pole 모델의 차수를 크게 하면 모든 음성에 거의 정합적인 표현을 할 수 있으므로 본 연구에서는

all-pole 모델만 다루며 그때의 차수는 12차로 하였다.

한편, 선형 예측 분석에서 얻은 첫번째 예측 계수와 예측 오차의 에너지를 음성 분석에서 응용하면 첫번째 예측 계수는 각 음성에서의 spectra가 서로 다르기 때문⁸⁾이며 무성음에 대해서는 첫번째 예측 계수의 평균값은 -0.02의 값, 유성음에 대해서는 -1.6의 값을 가진다. 그림5는 첫번째 예측 계수의 확률 밀도 함수를 나타내었다.

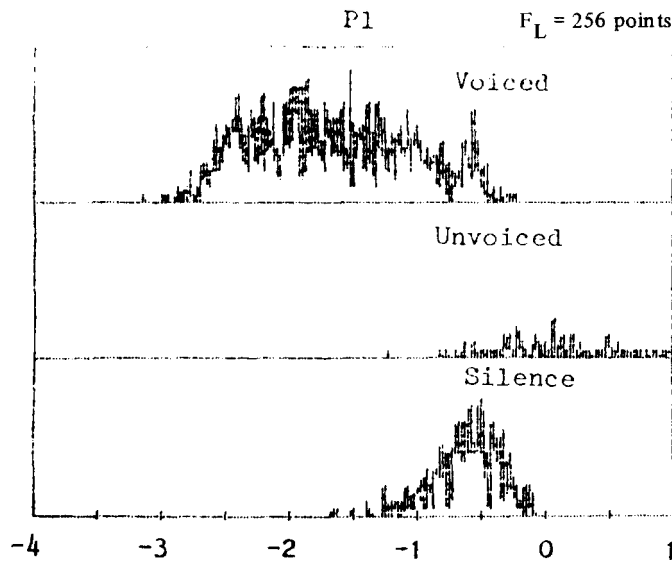


그림 5. 첫번째 예측계수에 대한 확률밀도함수.
Fig. 5. The probability density function for the first prediction coefficient.

예측 오차의 에너지는 각 음성의 스펙트럼에 대한 변화량을 측정하는 것⁹⁾이다.

이 파라메타는 각 음성을 구분하는데 무성음은 높은 주파수에 유성음은 낮은 주파수에 에너지가 분포하고 있다는 일반적인 관계를 사용하지 않은 파라메타이고 단지 어떤 주파수에서 다른 주파수의 변화에만 의존하며 예측 오차의 에너지에 대한 확률 밀도 함수는 그림6에 나타내었다.

III. 식별 알고리즘

그림1에서 5가지 파라메타는 음성 신호를 유성음, 무성음, 묵음으로 식별하기 위하여 사용하였다.

이런 결정을 하기 위하여 오차 결정 법칙에서 최소 확률을 사용하였다.¹⁰⁾ 측정치 벡터 X 가 M 패턴 부류 (W_1, W_2, \dots, W_M) 중의 어느 한 부류에 속해 있다고 가정하고 또 각각의 부류가 발생하는 데 그 이전의 확률들이 동일하다 (즉, X 가 어떤 부류가 될 확률이나 또 다른 부류가 될 확률은 동일하다)고 가정하면 측정치 벡터 X 가 실제로 W_i 부류로 되는 확률은

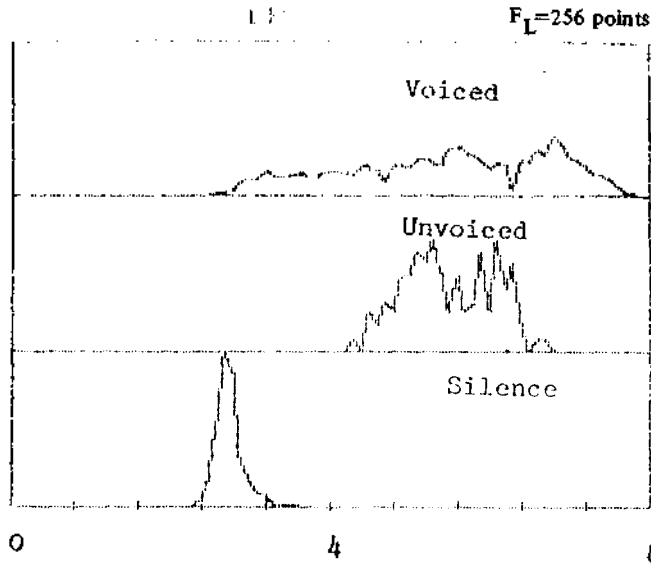


그림 6. 예측 오차의 에너지에 대한 확률밀도함수.
 Fig. 6. The probability density function for the prediction error energy.

$$P_j = \frac{P(x/W_j)}{\sum_{k=1}^M P(x/W_k)} \dots\dots\dots (9)$$

이다.
 여기서 $P(x/W_i) = P_i(x)$; 추정치 벡터 x 가 W_i 부류로 될 때의 확률 밀도 함수.

그리고, x 가 W_i 부류로 되지 않을 확률은

$$1 - P_j = 1 - \frac{P(x/W_j)}{\sum_{k=1}^M P(x/W_k)} \dots\dots\dots (10)$$

로써 오차 확률이다.
 결정 함수는 추정치 벡터 x 를 M패턴 부류들 중에서 정확하게 하나로 지정해 주는 함수 $d(x)$ 이며 최적 결정 함수는 오차 확률이 모든 x 값에 대하여 가장 적을 때이다.

즉, 식(10)에서 $1 - P_j$ 가 가장 적게 되는 j 값은 $P(x/W_i)$ 가 가장 크게 되는 j 값으로써 최적 결정 함수 $\hat{d}(x)$ 는

$$P(x/W_i) > P(x/W_j), \quad \forall j \neq i \dots\dots\dots (11)$$

또는 $\frac{P(x/W_i)}{P(x/W_j)} > 1, \quad \forall j \neq i$

일 때이다.

또한 확률 밀도 함수 $P(x/W_i)$ 의 식은

$$P(x/W_i) = \frac{1}{(2\pi)^{n/2} |C_i|^{1/2}} \exp\left\{-\frac{1}{2} (x - m_i)' C_i^{-1} (x - m_i)\right\} \dots\dots (12)$$

이다.
 여기서 m_i 는 평균 벡터이며 $|C_i|$ 는 행렬 C_i 의 행렬식이다.

그리고 식(12)에서 확률 밀도 함수의 지수형태(exponential form)를 자연 대수로 취하면 더욱 간단하게 된다. 달리 말하면 자연 대수는 단조 증가 함수이기 때문이며 그리고 최적 결정 함수에 대입하면 다음과 같다.

$$\hat{d}_i(x) = P(x/W_i) \cdot P(W_i) \dots\dots\dots (13)$$

$$\hat{d}_i(x) = \ln(P(W_i)) - \frac{1}{2} \ln|C_i| - \frac{n}{2} \ln 2\pi - \frac{1}{2} [(x - m_i)' C_i^{-1} (x - m_i)] \dots (14)$$

한편 식(14)의 처음 3항들은 추정치 벡터 x 와 무관하므로 다음식과 같이 간단하게 쓸 수 있다.

$$\hat{d}_i(x) = (x - m_i)' C_i^{-1} (x - \hat{m}_i) \dots\dots\dots (15)$$

윗 식은 추정치 벡터 x 와 평균 벡터 m_i 사이의

거리인데 통계적인 특징이 고려되고 유사도를 측정하는데 유용한 식이며 이를 Mahalanobis 거리 측정법이라고 한다.

마지막으로 식(19)에 의해 계산된 거리를 기본으로 하여 각 음성 신호 부류의 거리가 최소인 것을 선택하여 각 음성 신호를 결정한다.

IV. 식별 실험 및 결과

본 연구의 식별 실험에 사용된 음성은 성인 남성 3인에 의해 발생된 숫자음(1영~1구) 음성으로써 매회 5번씩 반복 발생하여 5회에 걸쳐 녹음한 것 중 임의로 선정한 154개 숫자음이다.

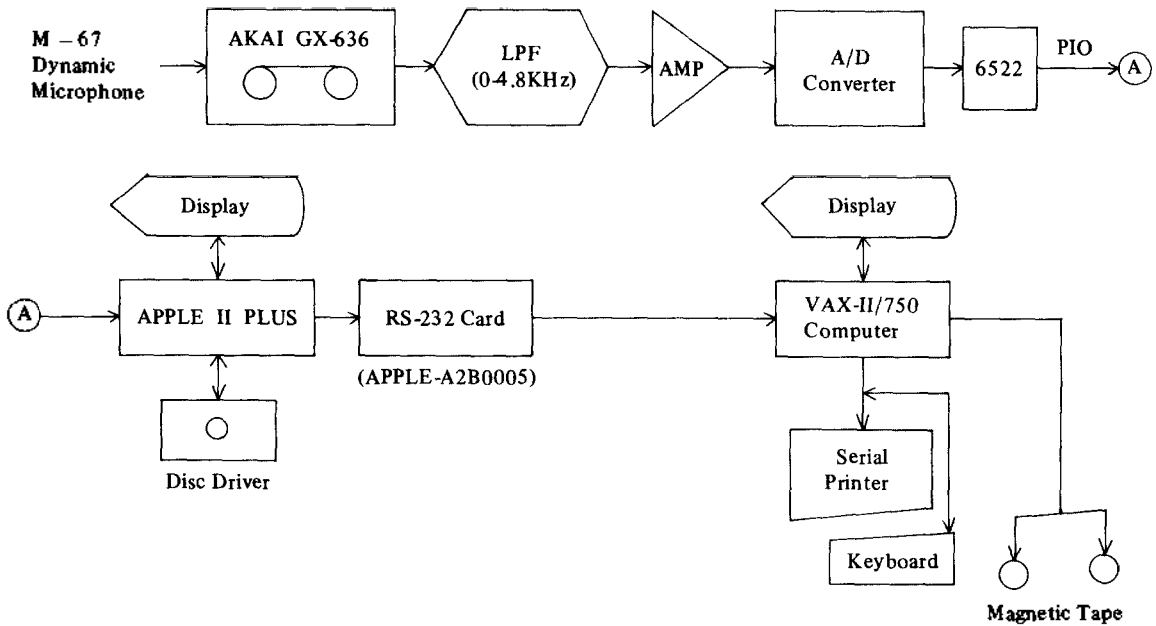


그림 7. 한국어 음성 분석 및 식별을 위한 시스템 블록도.
Fig. 7. A system block diagram for Korean speech analysis and classification.

4-1. 식별 시스템

식별 실험에 대한 전체 시스템은 그림7과 같다.

4-2. 표준 집합 설정

본 연구에서는 화자 독립 방법으로 음성 식별을 하였기 때문에 표준 집합 설정은 2명의 남성 화자가 발생한 숫자음 파형을 손수 식별하여 음성음, 무성음, 묵음 구간으로 설정하고 3가지 음성 신호에 대한 평균값, 표준 편차, 공분산 행렬을 통계적인 방법을 사용하여 구하였다.

그림 8은 |영|~|구|까지의 음성을 손수 식별한 구간을 나타내며 음성음 구간은 "V", 무성음 구간은 "U", 묵음 구간은 "S"로 표시하였다.

이러한 방법으로 손수 식별한 음성 구간에 대하여 2장에서 서술한 분석 파라메타를 구하고 각 음성 신호 부류에 대한 각 파라메타의 평균값과

분산값을 구하였다.

표 1은 표준 집합에 사용한 2명의 화자에 대한 각 음성의 평균치, 표준 편차, 공분산 행렬을 나타내었고 표에서 행은 2장에서 논의한 5가지 측정 파라메타이며 공분산 행렬의 대각선이 아닌 항(off-diagonal term)들은 다른 파라메타들 사이의 상관 관계를 측정한 것이다. 만일 앞에서 논의한 5가지 측정 파라메타들이 서로 독립적이고 상관 관계가 없다면 그에 off-diagonal term은 0일 것이다.

그러나, 모든 파라메타들은 서로 상관관계가 있으며 off-diagonal term의 값들은 0.07~0.96까지 변화한다.

유. 무성음 및 무성음 식별을 위한 연구

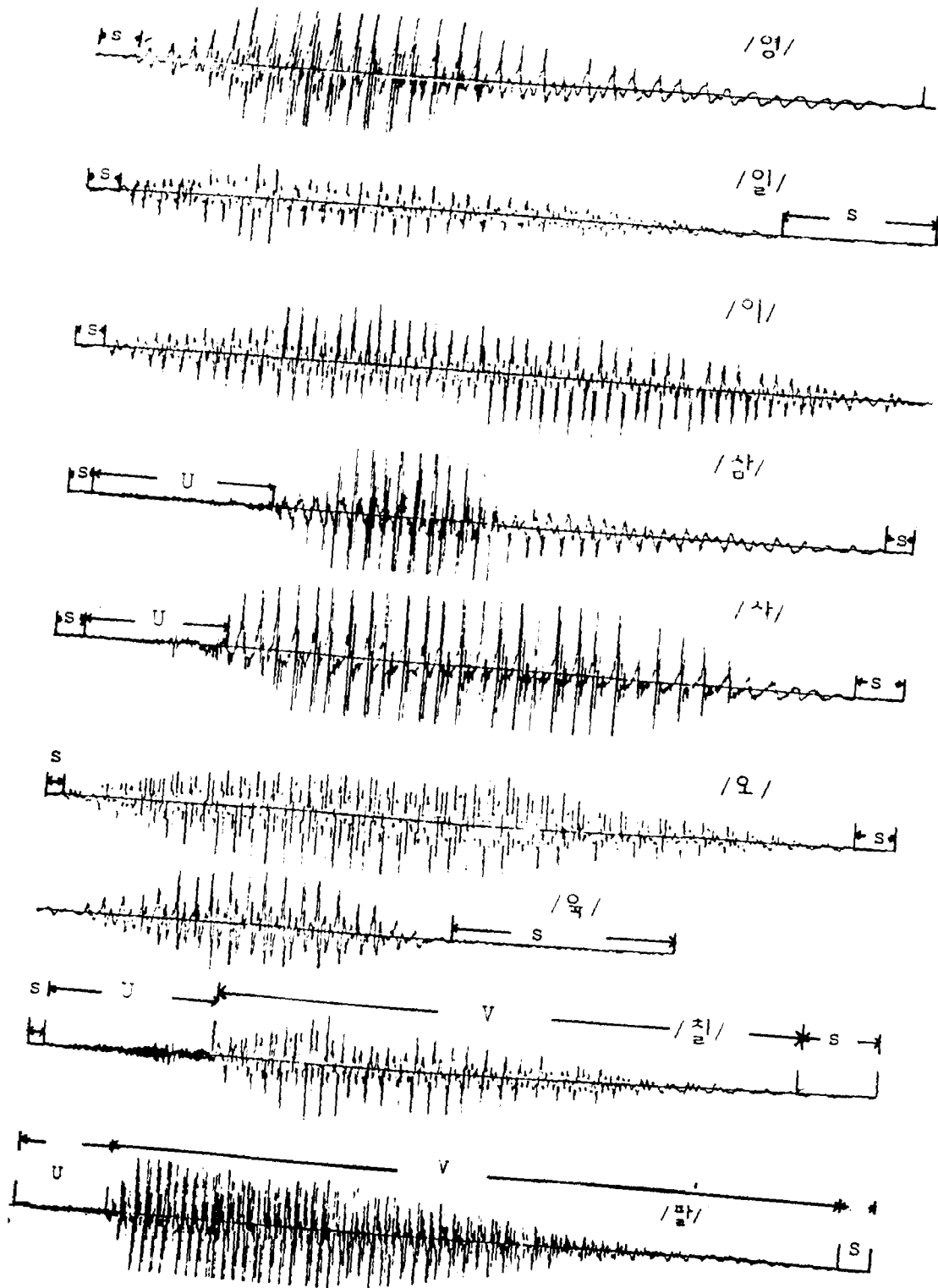




그림 8. 표준 집합에 사용한 음성 파형.

Fig. 8. Speech waveform using the reference set.

표 1. 표준집합에서의 평균값, 표준편차, 공분산 행렬.

Table 1. Means, Standard deviation, Covariance matrix in reference set.

VOICED (i = 1)	ZERO CROSSING	LOG ENERGY	FIRST AUTOCORR	FIRST LPC	LPC LOG ERROR
MEAN	25.5012	72.6845	0.887	-1.653	5.46539
STANDARD DEVIATION	17.9149	8.02614	0.12053	0.6084	1.26083
COVARIANCE MATRIX	1 0.3784 -0.9215 0.5210 0.67563	0.3784 1 -0.3358 0.0510 0.8747	-0.9215 -0.3358 1 -0.622 -0.6409	0.5210 0.0510 -0.622 1 0.38242	0.67563 0.8747 -0.6409 0.38242 1
UNVOICED(i = 2)					
MEAN	122.179	58.0179	0.05174	-0.0246	5.48372
STANDARD DEVIATION	28.7241	6.1372	0.0306	0.41339	0.56652
COVARIANCE MATRIX	1 0.5687 -0.8381 0.5290 0.6838	0.5687 1 -0.3149 -0.097 0.9681	-0.8381 -0.3149 1 -0.7553 -0.3737	0.5290 -0.097 -0.7553 1 -0.0561	0.6838 0.9681 -0.3737 -0.0561 1
SILENCE(i=3)					
MEAN	13.2985	41.3268	0.9401	-0.5879	2.7581
STANDARD DEVIATION	8.00351	3.7942	0.0715	0.25377	0.1936
COVARIANCE MATRIX	1 -0.210 -0.560 0.03720 0.6511	-0.210 1 0.6403 -0.6106 0.3603	-0.560 0.6403 1 -0.4056 -0.07784	0.03720 -0.6106 -0.4056 1 -0.30902	0.6511 0.3603 -0.07784 -0.30902 1

4-3. 거리 측정 및 결정 알고리즘

본 연구에서는 통계적인 특징을 고려하고 유사도를 측정하는데 유용한 거리 측정법인 Mahalanobis 거리 측정법을 사용하였다.

$$d_i(X) = (X - \hat{m}_i)^t C_i^{-1} (X - \hat{m}_i) \dots\dots\dots (10)$$

여기서 $i=1$ (유성음), $i=2$ (무성음), $i=3$ (묵음)을 나타내며,

\hat{m}_i 는 평균 벡타로써 수학적인 표현은

$$\hat{m}_i = \frac{1}{N_i} \sum_{m=1}^{N_i} x_i(m) \dots\dots\dots (17)$$

이다.

또, C_i 는 공분산 행렬을 나타내며

$$C_i = \frac{1}{N_i} \sum_{m=1}^{N_i} x_i(m) \cdot x_i(m)^t - \hat{m}_i \cdot \hat{m}_i^t \dots\dots\dots (18)$$

결정 알고리즘은 식(10)으로 거리 계산을 했을 때 $d_i(X)$ 가 최소인 것을 선택하여 결정하였으며 식별을 위한 전체 흐름도를 살펴보면 그림 9와 같다.

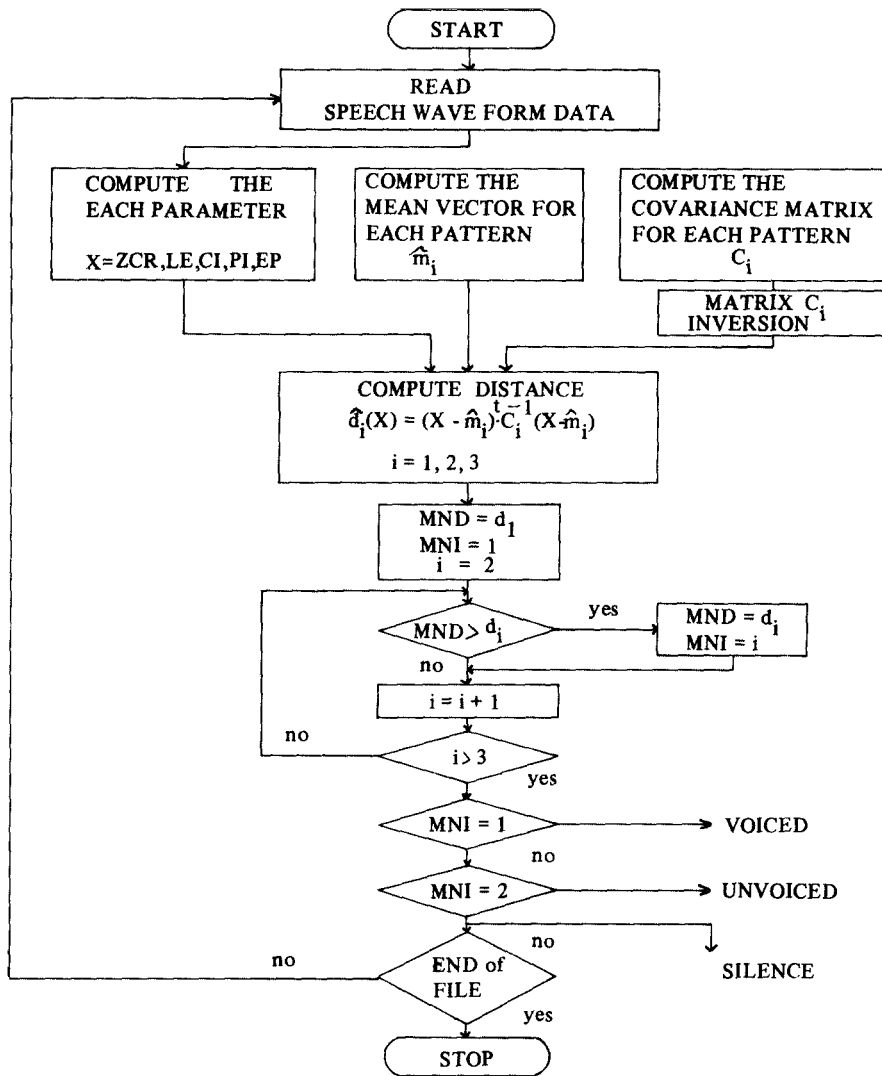


그림 9. 한국어음성식별을 위한 흐름도.

Fig. 9. Flow Chart for Korean speech classification.

4-4. 결과 및 고찰

전 장의 식별 시스템을 이용하고 성인 남성 3 명이 발성한 154개 숫자음을 대상으로 측정 파라메타들을 여러가지 방법으로 조합하여 식별한 결과와 고찰을 살펴보면 다음과 같다.

(1) 각각의 파라메타로 식별하였을 때,

영교차율은 유성음과 무성음, 무성음과 묵음 구간을 식별하는데 좋은 파라메타이며 대수 에너지는 음성과 비음성을 구별하는데 좋으며 첫 번째 자동상관 계수도 영교차율과 거의 비슷한 특징을 가지며 예측 계수의 파라메타는 유성음과 무성음을 구분하는데 좋다. 마지막으로 예측

표 2. 각각의 파라메타로 식별하였을 때의 전체 식별 오차율

Table 2. Total classification error rate with the each parameter.

MEASURED PARAMETER	V - U	V - S	U - S	TOTAL	ERROR RATE
ZCR	12	490	26	520	14.3%
LE	118	78	32	298	8%
CI	4	677	51	732	19.8%
PI	65	473	61	599	16.2%
EP	455	88	0	543	14.7%

오차의 에너지는 대수 에너지와 마찬가지로 음성과 비음성을 구분하는데 좋은 파라메타이다.

표 2는 각각의 파라메타로 식별하였을 때의 전체 식별 오차를 나타내었다.

(2) 2개의 파라메타를 조합하여 식별하였을 때,

전반적으로 각각의 파라메타로 식별하였을 때보다도 적은 오차가 발생하였으며 전체적으로 볼 때 가장 적은 식별 오차는 대수에너지와 첫 번째 자동 상관 계수를 조합하였을 때 2%의 식별 오차가 발생하였는데 오차의 대부분은 유성

표 3. 2개의 파라메타로 식별했을 때의 전체 식별 오차율.

Table 3. Total classification error rate with the two parameters.

MEASURED PARAMETER	V - U	V - S	U - S	TOTAL	ERROR RATE
ZCR, LE	54	46	5	105	2.8%
ZCR, CI	20	565	1	586	15.9%
ACR, PI	18	233	16	317	8.6%
ZCR, EP	6	81	2	89	2.4%
LE, CI	31	43	2	76	2%
LE, PI	296	37	8	341	9.2%
LE, EP	60	42	4	106	2.8%
CI, PI	32	271	16	319	8.6%
CI, EP	9	105	3	117	3.1%
PI, EP	560	45	2	607	16.4%

음과 무성음 구간의 경계 부분에서 발생하였으며 끝점(ending point)에서도 발생하였다. 표 3은 2개의 파라메타를 조합하여 식별하였을 때의 전체 식별 오차를 나타낸 것이다.

(3) 3 개의 파라메타를 조합하여 식별하였을 때.

유성음과 무성음 구간을 식별할 때 ZCR, CI, EP를 파라메타로 사용하였을 때 단1개의 식별 오차가 발생하였으며 유성음과 목음 구간은 ZC

R, CI, PI 을 파라메타로 사용하였을 때, 무성음과 목음구간은 LE, CI, EP를 파라메타로 사용하였을 때 적은 식별 오차가 발생하였으며 ZC R, PI, EP를 식별 파라메타로 선정하여 식별하였을 때 가장 적은 식별 오차율을 얻었으며 오차의 대부분은 음성의 끝점 부분에서 발생하였다.

3개의 파라메타를 조합하여 식별하였을 때의 전체 식별 오차율은 표 4에 나타내었다.

표 4. 3개의 파라메타로 식별했을 때의 전체 식별 오차율

Table 4. Total classification error rate with the three parameters.

MEASURED PARAMETER	V - U	V - S	U - S	TOTAL	ERROR RATE
ZCR, LE, CI	27	56	3	86	2.3%
ZCR, LE, PI	61	45	6	112	3%
ZCR, LE, EP	2	54	2	58	1.5%
ZCR, CI, PI	39	17	19	75	2%
ZCR, CI, EP	8	70	2	80	2.1%
ZCR, PI, EP	1	31	3	35	0.95%
LE, CI, PI	38	54	4	96	2.6%
LE, PI, EP	532	32	3	566	15.3%
LE, CI, EP	7	47	2	56	1.5%
CI, PI, EP	7	58	3	68	1.8%

표 5. 4, 5개 파라메타로 식별하였을 때의 전체 식별 오차율

Table 5. Total classification error rate with the four or five parameters.

MEASURED PARAMETER	V - U	V - S	U - S	TOTAL	ERROR RATE
ZCR, LE, CI, PI	39	56	3	98	2.6%
ZCR, LE, CI, EP	3	81	2	86	2.3%
LE, CI, PI, EP	0	56	0	56	1.5%
ZCR, CI, PI, EP	5	84	1	90	2.4%
ZCR, LE, PI, EP	9	78	3	90	2.4%
ZCR, LE, CI, PI, EP	3	94	1	98	2.6%

(4) 4, 5개 파라메타를 조합하여 식별하였을 때.

많은 파라메타를 조합하여 식별하였을 때 전반적으로 3%보다도 적은 오차를 얻었으며 오차의 대부분은 유성음 구간을 끝점 구간에서 묵음으로 식별하였으며 유성음과 무성음 구간의 경계 부분에서도 오차가 발생하였다.

표5는 4, 5개의 파라메타를 조합하여 식별하였을 때의 전체 식별 오차를 나타내었다.

V. 결 론

본 논문은 한국어 음성 인식을 위하여 음성 구간을 유성음, 무성음, 묵음으로 식별하기 위하여 패턴 인식 방법을 사용하였다.

특징 집합은 음성 분석에서 일반적으로 사용되는 음성 신호의 영교차율, 대수에너지, 자동상관 계수, 선형 예측 분석에서 얻은 첫번째 예측 계수와 예측 오차의 에너지이며 표준 집합은 3가지 음성 신호 부류에 대하여 손수 식별한 구간에서 평균 벡터와 분산을 통계적으로 처리 구성하였다.

그리고 특징 집합과 표준 집합간의 유사도 거리 측정법은 각 파라메타들의 상호간의 상관관계 및 통계적인 특징이 고려되는 Mahalanobis 거리 측정법을 사용하였으며 마지막으로 3가지 음성 신호 부류 판정은 각 음성에 대한 특징 집합과 표준 집합간의 계산된 거리가 최소일 때를 선택하여 결정하였다.

그리고, 각 파라메타들을 여러가지 방법으로 조합하여 식별한 결과는 영교차율, 선형 예측 분석에서 얻은 첫번째 예측 계수, 예측 오차의 에너지를 파라메타로 수행하였을 때 다른 조합방법보다도 좋은 결과를 얻었으며 그때의 식별 오차율은 1%보다도 적었다.

REFERENCE

1. L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signal, Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1978.
2. J.J. Dobnowski, R.W. Schafer and L.R. Rabiner, "Realtime Digital hardware pitch detector," IEEE. Trans. Acoust., Speech, Signal Processing, Vol. ASSP-24, pp. 2-8, Feb. 1976.
3. B.S. Atal and L.R. Rabiner, "A Pattern recognition approach to Voiced-Unvoiced-Silence classification with application to speech recognition," IEEE. Trans. Acoust., Speech, signal processing, Vol. ASSP-24, pp. 201-212, June, 1976.
4. L.R. Rabiner, C.E. Schmidt, and B.S. Atal, "Evaluation of a statistical approach to voiced-unvoiced-silence analysis for telephone quality speech," Bell Syst. Tech. J., Mar. 1977.
5. L.R. Rabiner and M.R. Sambur, "Application of an LPC distance measure to the voiced-unvoiced-silence detection problem," IEEE. Trans. Acoust., Speech, Signal Processing, Vol. ASSP-25, No.4, pp. 338-343, August 1977.
6. L.J. Siegel and A.C. Bessay, "Voiced/Unvoiced/Mixed excitation classification of speech," IEEE. Trans. Acoust., Speech, Signal Processing, Vol. ASSP-30, No.3, pp. 451-460, June, 1982.
7. 김준협, "한국어 음성의 분석과 자동 인식에 관한 연구," 박사학위논문, 연세대학교 대학원, 1982.
8. J.T. Tou and R.C. Gonzalez, "Pattern recognition principles," Addison-Wesley Publishing Co., Advanced Book Program, Reading, Mass., 1974.

1. L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signal, Prentice-