

## 음성신호 적응분할방법에 의한 특징분석

Features Analysis of Speech Signal  
by Adaptive Dividing Method장 승 관\* · 최 성 연\* · 김 창 석\*\*  
(S.-K. Jang, S.-Y. Choi, C.-S. Kim)

## ABSTRACT

In this paper, an adaptive method of dividing a speech signal into an initial, a medial and a final sound of the form of utterance utilized by evaluating extreme limits of short term energy and autocorrelation functions. By applying this method into speech signal composed of a consonant, a vowel and a consonant, it was divided into an initial, a medial and a final sound and its feature analysis of sample by LPC were carried out.

As a result of spectrum analysis in each period, it was observed that there existed spectrum features of a consonant and a vowel in the initial and medial periods respectively and features of both in a final sound. Also, when all kinds of words were adaptively divided into 3 periods by using the proposed method, it was found that the initial sounds of the same consonant and the medial sounds of the same vowels have the same spectrum characteristics respectively, but the final sound showed different spectrum characteristics even if it had the same consonant as the initial sound.

**Keywords:** extreme limits, speech signal, transient period, short-term energy

## 1. 서 론

음성은 인간이 기계와 의사전달을 위한 가장 편리한 수단으로서, 최근의 음성인식 기술은 특정화자인식시스템의 연구개발에서 불특정화자 인식시스템의 연구개발과 한정된 단어의 인식 시스템에서 대용량의 연속음성인식시스템의 연구로 전환되고 있다. 이와 같은 연구 내용은 불특정화자인식을 위한 시스템개발로서 화자를 군으로 분류(clustering)하고 변환(transformation)시키는 화자적응(speaker adaptation) 방법이 연구되고 있다.[1-4]

또한 연속음성의 인식을 향상시키기 위한 방법으로 음성신호의 동적특징(dynamic

---

\* 한국산업기술대학교 전기전자공학과

\*\* 명지대학교 전자정보통신공학부

feature) 파라미터를 일반화시키는 연구와 음소(音素) 모델링에 기초로 하는 연속음성인식 방법들이 연구되고 있다.[5-8]

화자적응의 기본 개념은 훈련된 특정화자의 통계학적인 특성을 갖는 군으로 훈련되지 않은 불특정화자의 특징파라미터를 이동시키는 것으로서, 음성신호의 동적 특성을 분석하는 연구방법과 유사한 처리절차를 갖는다. IBM 대용량 단어인식 시스템은 이 방법을 이용하여 개발되었으며 다음과 같은 두가지 처리절차를 갖고 있다. 첫 번째는 입력음성으로부터 10 ms 간격으로 특징벡터를 추출하는 과정으로 24차 멜캡스트럼(mel cepstrum)을 추정한 후 이 특징들을 24차원의 공간으로 변환시키는 과정이다. 두 번째는 음향학적인 모델을 설정하는 과정으로 단어를 음소의 열을 나타내는 나무(tree) 구조로 모델링한 후 각 가지를 다시 나무잎의 형태로 나타내며, 각 나무잎에 해당하는 음소는 다시 시작, 중간, 끝 부분에 해당하는 3개의 부음소열로 나누어서 모델의 특징을 추출하는 과정으로 되어 있다.[1],[8]

국내에서도 화자적응을 통한 대용량 단어 인식 시스템을 개발하려는 연구가 최근에 진행되고 있으며, 이와 같은 인식시스템 구현의 개념은 음성을 음소단위로 분할해서 음성인식을 수행하려는 데 기초하고 있다.[9-10] 특히 음성 발생기관의 천이 특성을 분석한 김창석[9]이 제안한 특징분석방법에 의하면 음성신호의 극점평가에 의하여 음성의 무성음과 유성음을 분할하고 각 구간별 특징 분석에 의하여 음성의 천이 특성을 분석이 가능함을 보여주었다. 효율적인 음성인식을 위한 또 다른 연구방향은 음성신호의 동적 특징 파라미터를 추정하는 방법으로서, 주로 음성의 주파수 특징을 추출하여 효과적인 음성인식을 할 수 있는 알고리즘의 연구가 많이 진행되고 있다.[11-14]

Rathinaelu[5]는 그의 논문에서 일반적인 음성특징은 정적특징 파라미터 즉 음성의 스펙트럼의 간단한 변환에 의한 캡스트랄 계수(cepstral coefficients)들이 많은 부분을 차지하고 있는 것으로 지적하고 있다. 특히 이와 같은 정적특징추정은 델타캡스트럼(delta-cepstrum), 델타델타캡스트럼(delta-delta-cepstrum)추정으로서 음성의 전처리과정에 한정되는 특징을 갖고 있다. 이와 같은 제한을 제거하기 위하여 Rathinaelu는 음성의 동적특징을 모델링할 수 있는 직접HMM(integrated Hidden Markov model)을 제안하였다. 직접HMM은 효율적인 음성인식을 위하여 가장 효과적인 음성모델인 즉 출력이 연속 혼합형 가우시안 출력 밀도(continuous mixture Gaussian output density)함수를 갖는 음소 HMM에 의한 음성인식방법이다.[14]

유사하게 국내에서도 음성의 동적특징도 분석할 수 있고, 음성의 특징의 손실이 적게 특징을 압축할 수 있는 OFEC알고리즘이 발표되었다.[15-16] 이 OFEC알고리즘은 음성신호를 적응중첩함수에 의하여 일정한 간격으로 프레임 분할한 후 각 프레임별로 LPC(linear prediction coefficients) 계수를 추출한다. 이렇게 추출된 LPC는 음성신호의 특징을 갖고 있는 특징파라미터로서 최적특징매핑함수에 의하여 추정평면으로 각 프레임별 LPC 계수를 매핑시킨다. 이렇게 매핑된 추정평면의 데이터는 원음성신호의 시간적인 변화에 대한 특징을 대표하고 있는 것으로 이 추정평면에서 원음성신호의 동적 특징을 분석할 수 있는 방법을 제안하였다. 또한 음성신호의 발생 시 발음의 시작부분과 정상상태가 되기 전인 천이구간에서의 성도파라미터 추정에 관한 연구이다.[17]~[20] 이러한 연구방법들은 성

도를 AR(auto-regressive)모델로 표시한 후 천이구간에서 LPC 추정에 의한 특징분석을 수행한다.

· 상기 서술한 바와 같이 현재 진행되고 있는 음성인식 연구분야는 국내외적으로 음성의 동적특성 분석에 의한 대용량 연속음성인식 구성에 관한 연구가 많이 진행되고 있으며, 이와 같은 추세에 따라 본 논문에서는 남자 2명과 여자 1명이 발성상의 단음절 음성을 대상으로 초성, 중성, 종성을 적응적으로 분할하는 방법을 제안하고, 단음절 단어의 구간별 동적 특징분석을 하고자 한다.

## 2. 음성신호의 적응 3분할방법

일반적으로 음성신호분석은 신호의 시작점과 끝점검출로부터 시작이 되며 천이구간의 분석에 의하여 무성음과 유성음의 분할 및 특징을 분석한다.[20] 음성성도모델에서 유성음의 에너지는 성문에 의하여 유기되는 에너지의 감소로 약 3 kHz 이하에 집중되고, 무성음의 경우는 대부분 높은 주파수성분을 가지고 있다. 따라서 영교차율이 높은 경우는 무성음이고, 영교차율이 낮은 경우는 유성음이다. 즉 10 ms당 평균 영교차율이 무성음에서는 약 49개, 유성음의 경우는 14개 정도이다.[11] 그러나 영교차율에 의한 음성분석은 여러 한계가 있어서 모든 음성신호 분석에 사용할 수 없는 문제가 있다.[20]

본 논문에서는 음성의 동적 특징을 분석하기 위하여 음성신호를 10 ms인 단구간으로 분할한 후 자기상관함수의 극점평가와 단구간 에너지에 의한 적응 3분할 방법을 제안한다.

### 2.1 단구간 자기상관함수의 극점 평가

본 논문에서는 음성신호의 동적특성을 분석하고, 음성의 발성학적 특성에 의한 초성, 중성, 종성으로 음성신호를 분할하기 위하여, 단음절의 음성신호를 시작점과 끝점을 검출한 후 검출된 음성신호를 10 ms인 단구간으로 분할한 후 단구간별로 다음의 단구간 자기상관함수를 적용하였다.

통계적 이산신호인 음성신호  $x(n)$ 에 대한 자기상관함수  $R(k)$ 는 다음과 같다.

$$R(k) = \sum_{n=-\infty}^{\infty} x(n) \cdot x(n-k) \quad (1)$$

식 (1)에서  $R(k)$ 는 무한주기를 가지는 신호  $x(n)$ 의 자기상관함수로서 일반적으로 음성신호의 분석에는 무한주기를 갖는 신호에 대한 자기상관함수보다는 음성신호를  $N$ 개씩 단 구간으로 분할한 후 데이터 개수가  $N$ 인 신호  $R_s(n)$ 의 단구간 자기상관함수를 이용하는 경우가 일반적인 분석방법이며, 각 구간별 음성신호의 특징을 추출하기 위한 단구간 자기상관함수  $R_s(k)$ 는 다음과 같다.

$$R_s(k) = \frac{N}{N-k} \sum_{n=0}^{N-1} x(n) \cdot x(n-k) \quad (2)$$

여기서  $x(n)$ 은 프레임의 음성데이터이고,  $\frac{N}{N-k}$ 은 식 (1)의 무한주기를 갖는 신호의 자기상관함수 즉  $-\infty$ 에서  $+\infty$ 까지인  $R(k)$ 함수가  $N$ 개로 제한됨으로서 생기는 오차를 보상시켜 주는 항이다. 본 논문에서는 음성신호의 동적 특성을 분석하기 위하여 식 (2)의 단구간의 데이터 개수  $N$ 을 100개로 하였으며 이 단구간 자기상관함수에 의해서 처리는 결과 데이터에 식 (3)에 의한 극점평가를 수행하였다.

데이터 개수가  $N$ 개인 신호  $x(n)$ 에서 단구간 자기상관함수를 적용한 후 극점을 평가하는 식은 다음과 같이 표현한다.

$$E(n) = \frac{1}{N} \sum_{m=n-N+1}^n \text{sgn}[R_s(n)] w(n-m) \quad (3)$$

$$\begin{cases} \text{sgn}[R_s(n)] = 1 & \text{if } R_s(n) \text{과 } R_s(n-1) \text{의 기울기 극성이 다른 경우} \\ \text{sgn}[R_s(n)] = 0 & \text{if } R_s(n) \text{과 } R_s(n-1) \text{의 기울기 극성이 같은 경우} \end{cases}$$

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

즉 식(3)은  $R_s(k)$ 의 기울기가 양의 값에서 음의 값으로 또는 그 반대의 경우로 변하는 점의 수를 평가하는 식이다. 이와 같은 평가방법은 자기상관함수의 특성에 따라 음성신호에 직류성분이 포함되어 있어도 직류성분에 의한 영향을 받지 않을 뿐만 아니라 단구간 자기상관함수를 이용함으로써 음성신호에 잡음이 포함되어 있어도 극점평가에 의하여 음성신호의 주파수특성을 평가할 수 있다는 장점이 있다.

음성신호의 동적 특성을 분석하는데 이용하는 또 다른 관계식은 단구간 에너지 함수이다. 단구간 에너지 함수는 음성신호를 분석하고자 하는 데이터 개수로 분할시킨 후 각 구간별 에너지를 평가하는데 사용하며, 음성신호를  $N$ 개씩인 단구간으로 분할시켜서 각 구간별 에너지를 평가하기 위한 식은 다음과 같다.[13][21]

$$E(n) = \sum_{m=n-N+1}^n x^2(m) \quad (4)$$

여기서,  $x(m)$ 은  $m$ 번째 음성데이터이고,  $n$ 은 전체 음성데이터수를 나타낸다. 일반적으로 식 (3)은 음성신호의 크기에 너무 민감하기 때문에 평가하는데 어려운 점이 있다. 따라서 이와 같은 문제점을 해결하기 위하여 식 (5)와 같은 평균크기함수(average magnitude function)를 적용한다.

$$M(n) = \sum_{m=n-N+1}^n |x(n)| w(n-m) \quad (5)$$

단구간프레임의 크기를  $N$ 으로 한정하면 식 (5)는 다음과 같은 단구간 평균크기함수(short-term average magnitude function)로 표현된다.

$$M_S(n) = \sum_{m=n-N+1}^n |x(n)| w(n-m) \quad (6)$$

여기서,  $x(n)$ 은 평가하고자 하는 음성신호이고,  $w(n)$ 은 단구간 창함수이다. 그러나 식 (6) 역시 평가된 음성신호에 직류성분이 있는 경우 음성신호만의 에너지를 평가할 수 없으므로 식 (7)과 같은 변형된 단구간 에너지평가함수를 적용하였다.

$$M_p(n) = \frac{1}{N} \sum_{m=n-N+1}^n |x_p(n) \cdot w(n-m)| \quad (7)$$

$$x_p(n) = |x(n) - x(n-1)|$$

여기서  $x_p(n)$ 은 이웃한 두 음성신호차의 절대값 크기를 나타낸다. 즉, 식 (6)을 직류성분이 포함된 음성신호에 적용함으로써 직류성분에 의한 에너지를 제거하고 음성신호만의 단구간 에너지를 평가할 수 있다.

그림 1은 10 kHz로 샘플링한 음성신호 “각”에 대하여 10 ms인 단구간으로 분할시켜서 식 (3)의 단구간 극점평가식과 식 (7)의 단구간 에너지 평가식의  $N$ 을 10 ms의 데이터인 100개로 하여 평가한 결과이다. 여기서 그림 1의 (a)는 음성신호 “가”이고 (b)는 음성신호의 단구간별로 평가한 극점 평가결과이며 (c)는 단구간별로 평가한 에너지 값이다. 특히 그림 1의 (b)에서 ㉠은 음성신호 “각”에서 평가한 극점 평가의 평균값을 나타낸다.

그림 1의 평가결과로부터 알 수 있듯이 음성신호 “가”는 가운데 부분에 에너지가 많이 집중됨을 알 수 있고 단구간 자기상관함수에 의한 극점 평가값은 초반부와 중반부에서 평균값보다 높게 평가되었음을 알 수 있다.

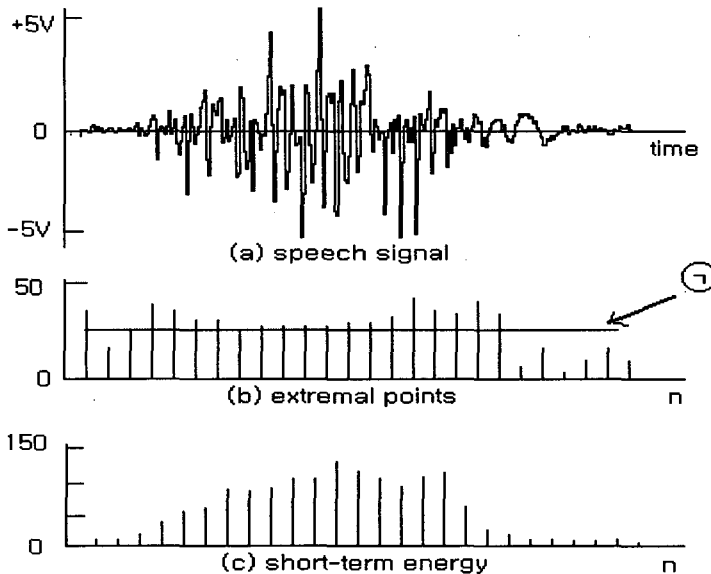


그림 1. 음성“가”에서 단구간 에너지와 극점 평가

## 2.2 음성신호의 적응 3분할

음성의 동적 특성을 분석하고, 음소의 특성 고찰을 하기 위하여 음성의 발생상의 초반부인 초성, 중반부인 중성, 후반부인 종성을 분할하였으며 본 논문에서는 식 (3)의 극점

평가값과 식 (7)의 단구간 에너지값을 이용하여 음성의 시작점과 끝점을 검출하고 또한 발성상의 초성, 중성, 종성 분할을 제안한다.

3분할을 하기 위하여 10 kHz로 샘플링한 음성 신호를 약 10 ms의 단구간으로 분할하였다. 이렇게 분할된 각 단구간별 데이터에 식 (2)의 단구간 자기상관함수를 적용한 후 각 단구간별 결과에 식 (3)의 극점평가함수를 적용하여 단구간별 극점 평가를 수행하였으며 음성 “낙”에 적용한 결과를 그림 2의 (b)에 나타내었다. 음성신호의 분할을 위한 또 다른 평가는 단구간별로 식 (7)의 단구간 에너지평가함수를 적용하여 단구간별 에너지를 평가하는 것으로서 그림 2의 (c)가 그 결과이다.

그림 2의 (b)에서 ⊕은 음성신호 “낙”에 평가한 극점의 평균 평가값으로서, 음성의 초반부에서 극점 평가값이 급격하게 증가하여 평균 극점 평가값 보다 크게됨을 알 수 있고, 종반부에서는 극점 평가값이 평균치보다 2배 이상 높은 값에서 급격하게 감소됨을 알 수 있다. 그림 2의 (c)의 단구간 에너지 평가에서는 초성 부분이나 중성 부분보다 중성부분에 에너지가 높게 평가되었음을 알 수 있고, 초성 부분보다 중성부분에서 에너지 변화가 급격히 감소됨을 알 수 있다.

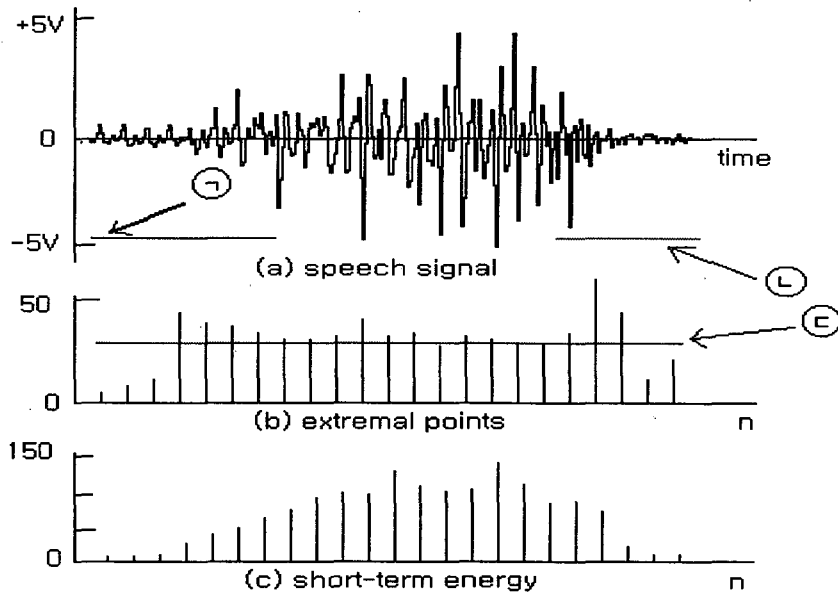


그림 2. 음성“낙”에서 단구간 에너지와 극점 평가

발성상의 초성, 중성, 종성을 분할을 위하여 먼저 본 논문에서는 음성신호의 시작점과 끝점을 검출하였으며 그 기준은 음성신호의 피크치가 -5V에서 +5V 사이의 신호를 12 bit의 A/D변환시켜서 얻은 데이터를 이용하여 모든 계산을 수행하였다. 음성신호의 시작점은 단구간 에너지가 200이상이고 단구간 자기상관함수의 극점 평가값이 5에서 40 사이의 단구간을 검출한 후 그 구간의 처음데이터를 음성신호의 시작점으로 하였고, 끝점은 단구간 에너지가 800 이하이고 단구간 자기상관함수의 극점 평가값이 5에서 40 사이의 단

구간을 검출하여 그 구간의 마지막 데이터를 음성신호의 끝점으로 하였다. 다음으로 발생상의 초성, 중성, 종성의 검출 및 분석을 위하여 시작점과 끝점을 검출한 음성신호로부터 단구간 에너지가 평균 단구간 에너지의 50 % 이상인 단구간을 검출한 후 단구간 에너지와 단구간 자기상관함수의 극점을 평가하였으며, 각 단구간별 에너지와 자기상관함수의 극점 평가값으로부터 각각 평균 단구간 에너지와 평균 극점을 평가하였다. 초성의 시작점은 음성신호의 시작점과 같은 점이고, 초성의 끝점은 단구간별 극점 평가값과 평균 단구간 극점 평가값의 차가 평균 단구간 극점 평가값보다 5배 이상인 구간에서 단구간 에너지의 증가가 최대인 구간을 검출한 후 검출한 구간의 마지막 데이터를 초성의 끝점으로 하였다. 중성의 시작점은 초성의 끝점과 동일하며 중성의 끝점 검출은 단구간 극점 평가값과 평균 단구간 극점 평가값의 차가 평균 단구간 극점 평가값보다 약 5배 이상인 구간을 검출하여 단구간 에너지의 감소가 최대인 구간을 검출한 후 검출한 구간의 마지막 데이터를 중성의 끝점으로 하였다. 종성의 시작점은 중성의 끝점과 동일하며 종성의 끝점은 음성신호의 끝점과 동일하다.

이상과 같이 음성신호의 시작점과 끝점, 발생상의 초성, 중성, 종성을 검출하기 위한 전 처리절차는 그림 3에 나타내었으며, 음성신호 “낙”에 대하여 본 논문에서 제안한 방법을 적용하여 발생상의 초성, 중성, 종성을 검출한 결과는 그림 2와 같고, 그림 2의 (a)에서 ㉠은 음성 “낙”의 초성구간이고 ㉡은 음성 “낙”의 중성구간이고 ㉢과 ㉣의 사이가 중성구간이다.

이렇게 분할한 각 구간별 특성분석의 결과는 그림 4와 같으며 본 논문에서는 동적 특성을 분석하기 위하여 분할된 초·중·종성구간을 2등 분할한 후 분할된 구간 데이터에 FFT를 적용하여 스펙트럼을 평가하고, 동일 데이터에서 LPC 스펙트럼을 추정하여 두 스펙트럼을 비교 분석하였다.

그림 4는 제안한 방법에 의하여 음성신호 “낙”을 초·중·종성으로 분할시켜서 추정한 결과들로서, 여기서 (a)는 초성의 데이터를 2등 분할하여 첫 번째 구간과 두 번째 구간의 데이터에 FFT 시켜서 평가한 파워 스펙트럼이고 (b)는 동일 데이터에 대하여 LPC를 구한 후 주파수 특성을 분석한 결과이다. 마찬가지로 (c), (d)는 중성구간에서 평가한 FFT 스펙트럼과 LPC 스펙트럼이며 (e)와 (f)는 종성에서 추정한 스펙트럼 결과이다.

그림 4로부터 알 수 있듯이 초성, 중성, 종성의 구간별 스펙트럼특징 분석에서는 초성의 주파수 특성은 약 1 kHz 이하의 주파수에서 높은 특성을 갖고있으며 약 2.7 kHz에도 높은 특징을 갖으며, 중성구간에서 추출한 주파수 특성은 1 kHz 부근의 주파수에서만 높은 특성을 갖고 있음을 확인하였고, 종성의 경우 약 1.3 kHz와 2.9 kHz 부근의 주파수에서 높게 나타남을 알 수 있다. 즉 초성과 중성, 중성과 종성의 스펙트럼 특징이 매우 차이가 있음을 알 수 있었다.

초성, 중성, 종성 구간별 유사성의 평가에서는 동일 발생 구간에서 2등분하여 추정한 스펙트럼 특성분석으로부터 초성의 경우 약 2.7 kHz의 고주파 성분에서 저주파 성분으로 천이가 이루어짐을 분석할 수 있었고, 중성의 경우 2구간의 주파수 특성이 유사하게 추정되었음을 평가하였고, 종성에서는 초성과 반대로 저주파 성분에서 고주파 성분으로 변화됨을 확인할 수 있었다. 위와 같은 구간별 주파수 특성 변화는 단구간 자기상관함수 극점평

가 결과인 그림 3의 (b)로부터 초성과 종성구간의 극점 평가값이 종성보다 크게 추정되었음을 알 수 있고 극점 평가값이 크면 원음성신호의 주파수 특징도 4 kHz 부근의 고주파 성분이 많음을 알 수 있었다.

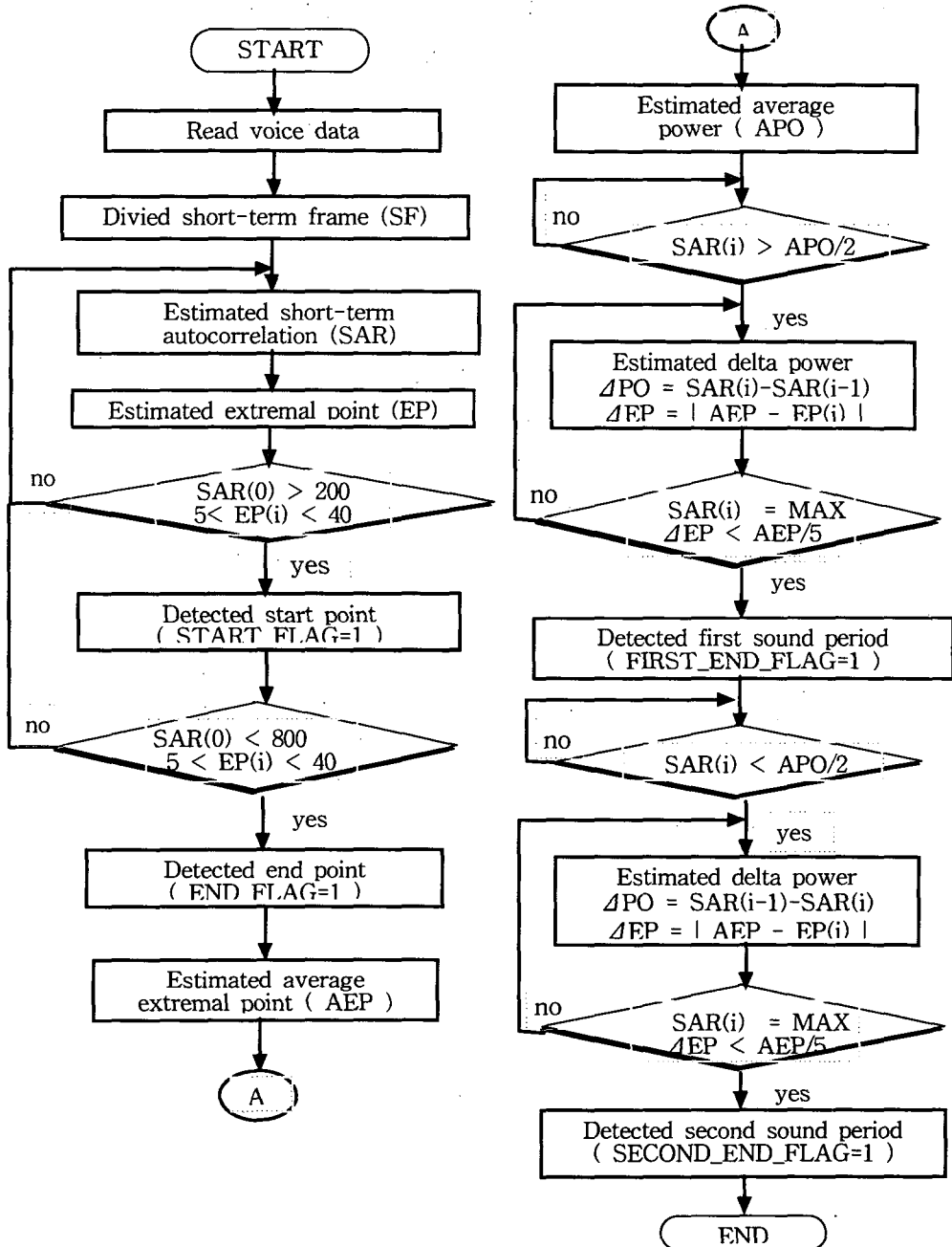


그림 3. 음성신호의 초·중·종성 3분할을 위한 전 처리과정



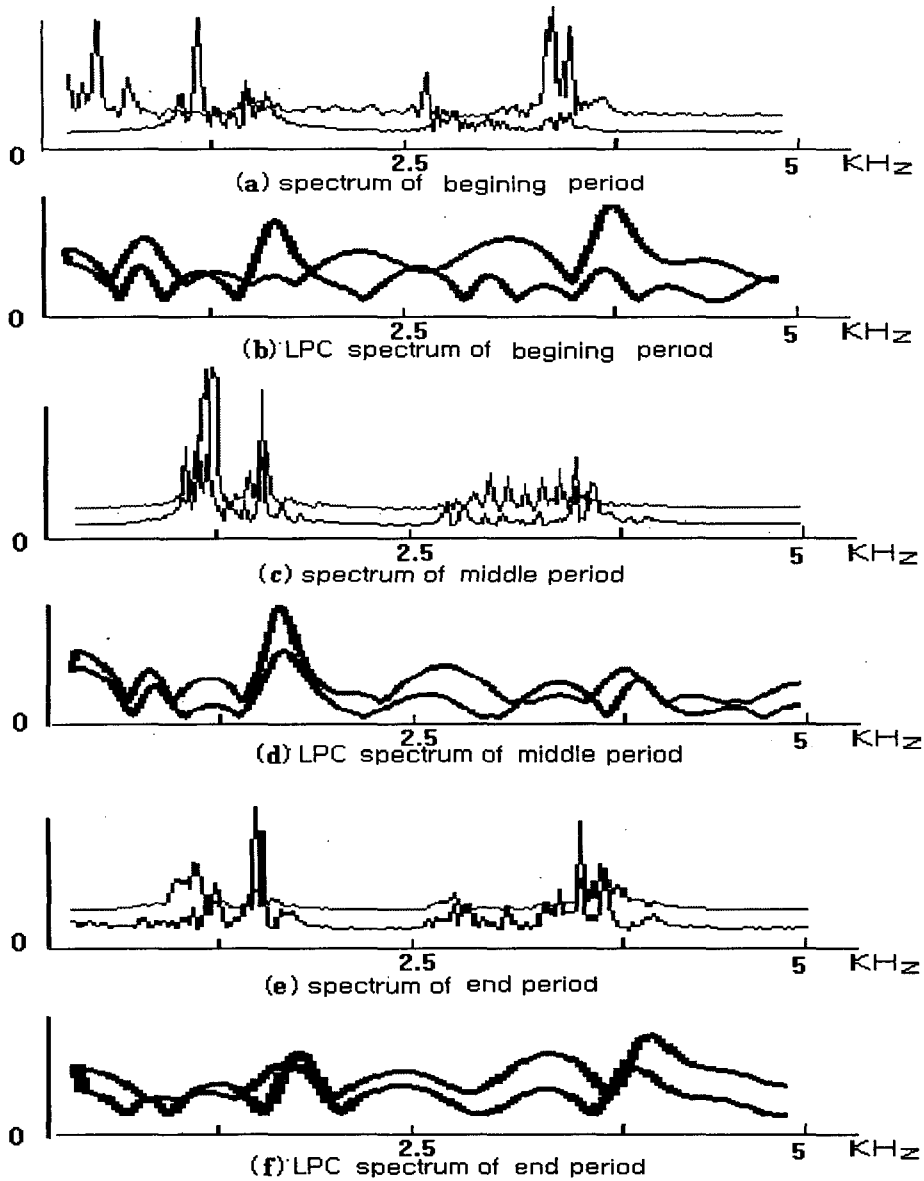


그림 4 음성 "나"에서 3분할에 의한 스펙트럼 분석

### 3. 실험 및 고찰

제안한 방법의 실효성을 평가하기 위하여 본 논문에서는 밀폐된 실내에서 3인이 발성한 초·중·종성이 있는 음성 신호 중 초성이 "ㄱ", "ㄴ", "ㄷ", "ㄹ", "ㅅ"과 중성은 "ㅏ", "ㅑ", "ㅓ", "ㅕ"와 종성은 "ㄱ", "ㄴ", "ㄷ", "ㄹ", "ㅁ", "ㅇ"으로 되는 음성을 발생해서 실험을 수행하였으며, 음성데이터는 12비트 A/D변환기에서 10 kHz 샘플링을 시켜서 얻은 후 IBM

PC/pentium 계열을 사용하여 수행하였다. 또한 음성신호의 발음상 초성, 중성, 종성을 분할한 후 각각의 특성을 분석하기 위하여 음성신호를 약 100개 단위의 구간으로 분할한 후 단구간별로 수행하였고, 천이구간의 특징추출은 분할된 음성데이터로부터 LPC 계수를 추출하고 이 LPC 계수에 의한 스펙트럼분석과 음성데이터에 FFT(fast Fourier transform)를 적용한 주파수 특성분석을 수행하였다.

### 3.1. 극점 평가에 의한 음성신호의 3분할

음성신호 “난”에 대하여 본 논문에서 제안한 방법으로 발음상 초·중·종성을 분할한 결과는 그림 5와 같다. 이 처리과정을 보면 먼저 음성신호를 시작점과 끝점을 검출한 후 검출된 음성신호의 동적 특성을 고찰하기 위하여 약 10 ms 구간 즉 100개의 단구간으로 분할하였다. 이렇게 분할된 각 단구간별 데이터에 식 (2)의 단구간 자기상관함수를 적용한 후 각 단구간별 결과에 식 (3)의 극점평가함수를 적용하여 단구간별 극점평가를 수행하였으며 음성 “난”에 적용한 결과를 그림 2의 (b)에 나타내었다. 특히 극점의 평가에서는 원 음성신호의 주파수 특징을 분석할 수 있으며, 본 논문에서는 주파수의 변화특성을 분석하기 위하여 음성신호를 2개로 분할하여 초반부의 극점평가결과는 ㉠이고 후반부의 극점평가결과는 ㉡으로서 초반부보다 후반부의 극점 평가값이 크게 평가되었고 그 결과로부터 초반부보다 후반부의 주파수 특징이 높다는 것을 알 수 있다.[9]

음성신호의 분할을 위한 또 다른 평가는 단구간별로 식 (7)의 단구간 에너지평가함수를 적용하여 단구간별 에너지를 평가하는 것으로서 음성신호 “난”에서 평가한 결과는 그림 2의 (c)와 같으며 단구간 에너지 평가에서는 초성 부분이나 중성 부분보다 중성부분에 에너지가 높게 평가되었음을 알 수 있고, 초성 부분보다 중성부분에서 에너지 변화가 급격히 감소됨을 알 수 있다.

그림 5의 3분할 처리절차에 의하여 음성 “난”에서 초·중·종성 분할결과는 그림에서 (a)와 같으며 여기서 ㉠은 초성이고 ㉡은 중성이며 ㉢과 ㉣사이가 중성구간이다.

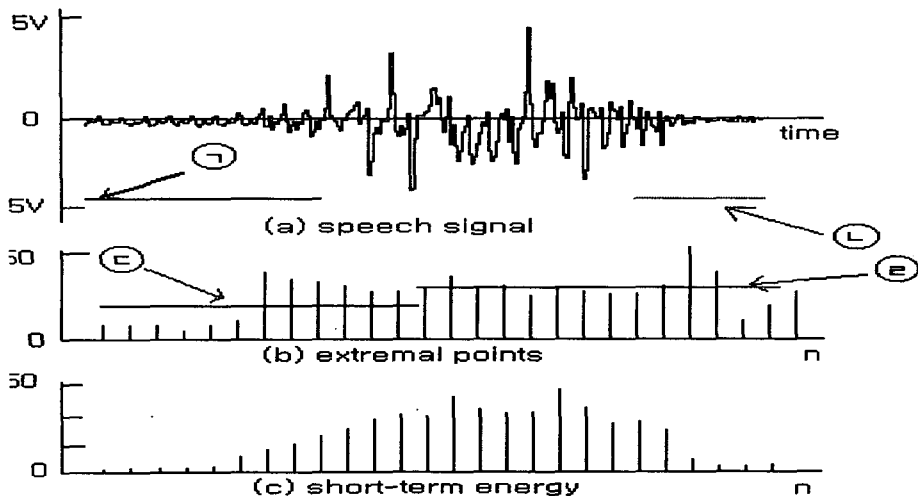


그림 5 음성 “난”에서 단구간 에너지와 극점 평가

제안한 방법의 타당성을 검토하기 위하여 발성상의 초성, 중성, 종성구간을 분할한 후 분할구간별로 D/A보드를 이용하여 청취하였으며 청취결과 분할구간에서 음성의 천이가 이루어짐을 확인하였다. 또한 분할 구간별로 추정된 스펙트럼 특성 분석을 통하여, 초성과 중성, 중성과 종성의 스펙트럼 특성을 비교하여 세 구간의 분할이 적합한 것을 검증하였고, 동일 구간을 다시 2개로 등분할 하여 각 구간별로 추정된 스펙트럼이 유사한가를 평가하였다. 그림 6는 음성 “간”과 “난”에 제안한 방법을 적용하여 초·중·종성으로 구간분할 한 후 구간별로 스펙트럼 특성을 분석한 결과이다.

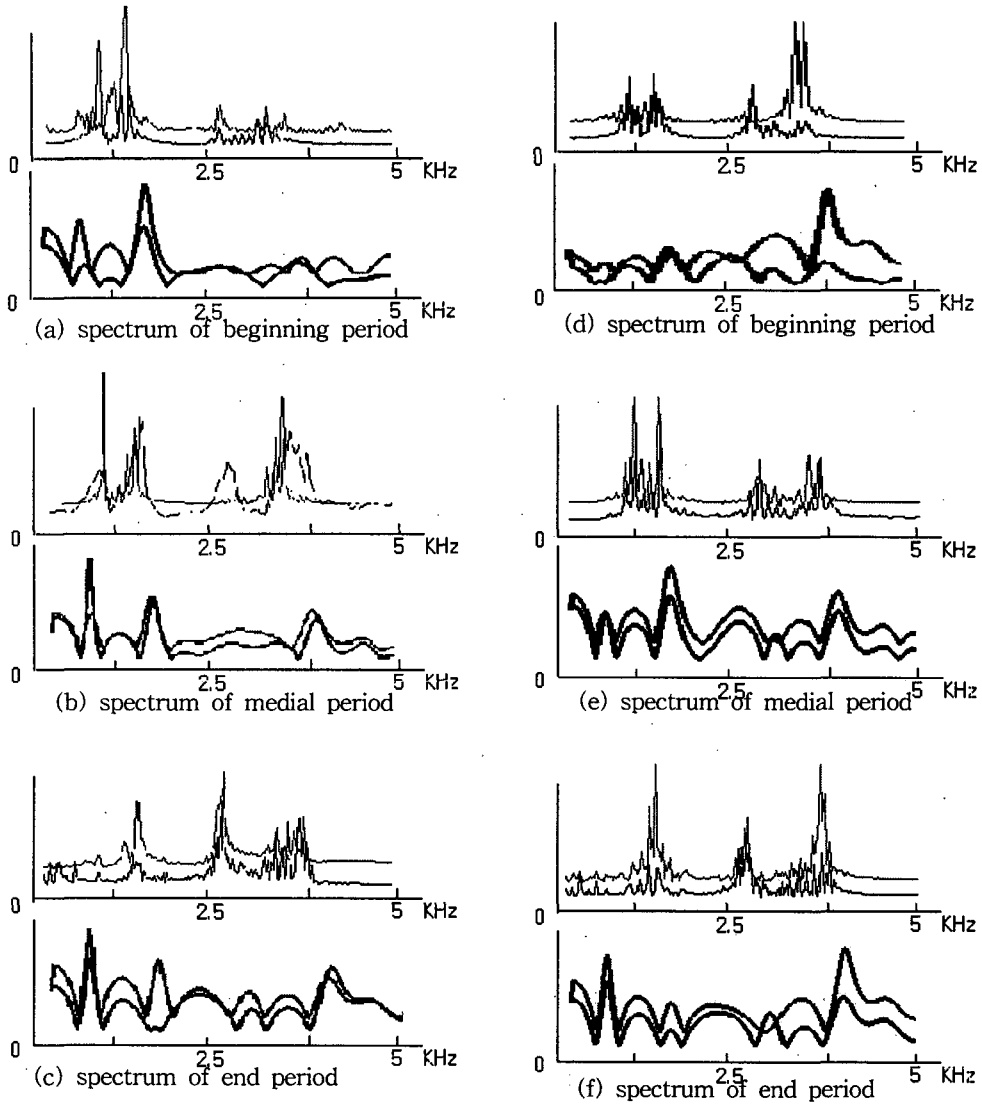


그림 6. 음성 “간”과 “난”의 스펙트럼 비교분석

그림 6에서 (a), (b), (c)는 각각 음성 “간”의 초성, 중성, 종성의 스펙트럼 특성을 나타내고, (d), (e), (f)는 음성 “난”의 초·중·종성의 구간별 스펙트럼 분석결과이며, (a)에서 (f)까지 결과에서 각각 첫 번째 결과는 “각”의 음성신호에서 초성으로 분할된 데이터에 FFT를 적용하여 스펙트럼을 분석한 결과이고 각 두번째 결과는 동일데이터에서 LPC계수를 추출한 후 LPC에 의한 스펙트럼을 분석한 결과이다.

그림 6의 결과로부터 음성 “간”의 경우 초성은 약 1.5[kHz]에 스펙트럼이 집중되어 있고 중성은 스펙트럼이 약 0.8[kHz]와 3.4[kHz]에 많이 위치하며 중성은 초성에서 종성으로 변화되는 과정으로 초성과 종성의 주파수 성분을 같이 포함하고 있음을 알 수 있다.

음성 “난”의 경우 초성은 약 2.8[kHz]에 스펙트럼이 집중되어있고 중성은 스펙트럼이 약 0.8[kHz]와 3.4[kHz]에 많이 위치하며 중성은 초성과 종성의 주파수 성분을 같이 포함하고 있음을 알 수 있다. 또한 두 음성 모두 동일 발성구간을 2구간으로 등분할한 특성분석에서는 2분할 구간의 주파수 특성이 유사함을 알 수 있었고, 두 음성의 초·중·종성의 비교 분석에서는 D/A변환기에 의한 청취 실험에서 초성의 경우 “ㄱ”과 “ㄴ”의 자음 발성의 차이가 있음을 확인할 수 있었고, 스펙트럼 비교분석에서도 “ㄱ”의 주파수 특성은 “ㄴ”의 주파수 특성에 비하여 낮은 주파수를 갖고 있음을 분석할 수 있었으며, 중성과 종성의 비교 분석에서는 두 중·종성의 청취실험에서는 깨끗한 “안”의 음성은 아니지만 중반으로 갈수록 “안”의 발성에 가깝게 발성됨을 알 수 있었고, 주파수 특징 비교분석에서도 중·종성의 주파수 특성이 매우 유사함을 확인하였다.

### 3.2 초·중·종성의 특징 분석

본 실험에서는 제안한 방법에 의하여 음성신호를 발성상 초·중·종성으로 분할후 각 구간별 발성길이, 단구간 자기상관함수에 의한 극점 평가 및 구간별 주파수 특성을 분석하였으며 그 결과는 다음과 같다.

#### 1) 음성별 초·중·종성의 발성길이 분석

천이구간의 특징분석 중에 하나가 유·무성음의 발성길이의 평가로서 본 실험에서는 음성신호를 초·중·종성으로 분할한 후 각 구간별 길이분석에 의하여 초성과 종성의 무성음의 천이특성을 분석하였다. 초성의 경우 자음에 의한 천이 특성을 고찰하였고 중성은 모음만의 특성과 종성은 모음과 자음의 혼합된 특성을 분석하였고 대표적인 음성의 결과는 표 1과 같다. 표 1에서 중성은 청취 실험으로부터 발성자의 발성이 긴 경우는 길게 평가되고 발성이 짧은 경우는 짧게 됨을 확인할 수 있었다. 초성과 종성의 초성의 발성길이 분석실험결과는 표 2와 같으며 초성의 평가는 초성이 같은 대상음성신호의 초성 길이의 평균값이고 종성은 자음이 같은 음성신호에서 평가한 종성길이의 평균값으로서 표 2로부터 알 수 있듯이 초성은 “ㄹ”의 발성길이가 짧게됨을 알 수 있고 그 외의 초성 음성들간의 길이의 차가 크지 않음을 알 수 있었다. 또한 종성의 길이분석에서는 초성의 경우와 달리 “ㄱ”이 짧게 평가되었고 “ㄹ”, “ㄴ”, “ㅇ”의 발성이 길게 됨을 확인할 수 있었다.

표 1. 음성 “각”부터 “냥”까지 발성길이 특성 (단위: 10 ms)

분 류	각	간	갈	감	갑	강	낙	난	날	남	납	냥
초 성	5	9	7	8	4	8	8	7	6	7	6	7
중 성	13	12	16	10	12	12	16	14	14	13	13	16
종 성	9	10	12	12	12	12	4	8	14	9	6	7

표 2. 음성데이터의 평균 발성길이 특성 (단위: 10ms)

초 성	각-강	낙-냥	녹-농	락-랑	릭-링	작-장	직-깡
길 이	6.8	6.8	6.8	7.3	10	8.2	7.8
중 성	“ㄱ”	“ㄴ”	“ㄷ”	“ㄹ”	“ㄴ”	“ㅇ”	
길 이	5.8	11.9	13.1	12.3	9.4	12.3	

## 2) 단구간 자기상관함수의 극점 평가에 의한 주파수 특성 분석

본 실험에서는 음성의 단구간 자기상관함수의 극점평가에 의한 음성신호의 주파수 특성을 분석하였고 음성의 동적 주파수 특성을 분석하기 위하여 음성신호를 초반부와 후반부로 등분할하여 2구간별 특징을 분석하였으며 그 결과는 그림 7과 같다. 그림 7에서 X축 Y축 Z축은 표 3과 같으며 그림 7로부터 대부분의 음성은 초반부보다 후반부의 극점 평가 값이 크게 평가됨을 알 수 있고 음성 “녹”-“농”은 다른 음성들에 비하여 비교적 극점 평가 값이 낮으며 음성 “직”-“깡”은 다른 음성보다 크게 평가됨을 알 수 있었다.

초성의 극점 평가는 “ㄴ”이 초성인 겨우 약간 작은 값으로 평가되었으나 그 외에는 커다란 차이가 없었으며 종성에 “ㄴ” 있는 음성의 경우 비교적 낮은 값으로 평가되었고 “ㄷ”을 갖는 음성은 다른 음성에 비하여 비교적 높은 값으로 평가됨을 알 수 있었다. 특히 모음이 “ㅣ”인 음성에서는 초반부에서부터 후반부까지 모두 다른 음성에 비하여 높게 평가되었음을 확인할 수 있었고 따라서 극점평가에 커다란 영향을 미치는 요인은 무성음보다 유성음임을 확인할 수 있었다.

표 3. 그림 7의 X, Y, Z축의 값

분 류	X 축		Y 축	Z 축							
	1-6	7-12		B	C	D	E	F	G	H	
음 성	초반부	후반부	극점 평가수	각	낙	녹	락	릭	작	직	
	종성이 “ㄱ”~“ㅇ”										

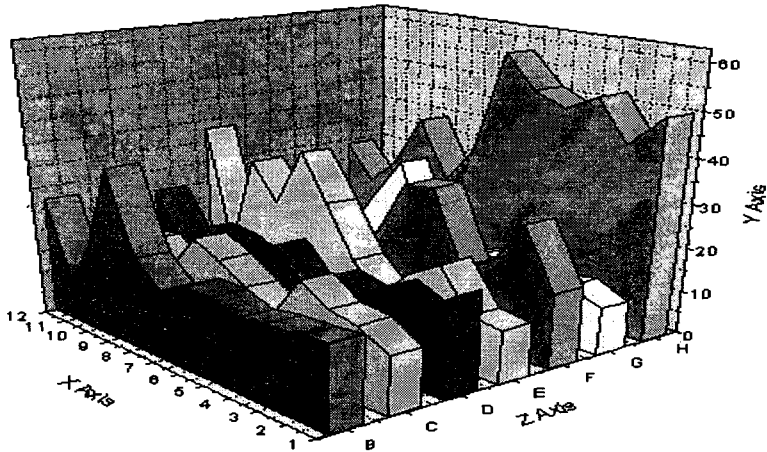


그림 7. 음성데이터들의 극점 평가 결과

본 논문에서 음성데이터에 대한 주파수 특성 분석은 모든 음성데이터를 초·중·중성으로 분할한 후 각각에 대한 제 1포먼트와 제 2포먼트를 분석하였으며 초성과 중성에 대한 결과는 표 4와 표 5에 나타내었다.

초·중·중성의 주파수 특징 분석에서는 자음의 특성 때문에 초성과 중성의 스펙트럼이 중성의 모음에 의한 스펙트럼보다 비교적 높게 평가됨을 알 수 있었으며, 중성에 모음이 “ㅏ”인 경우 대부분의 음성이 0.5 kHz와 1.8 kHz에 많이 분포됨을 알 수 있고, 모음이 “ㅓ”인 경우는 0.5 kHz와 3.5 kHz 부근에 집중됨을 확인하였다. 또한 모음이 “ㅜ”인 음성의 경우 0.5 kHz와 2.2 kHz에 많이 집중됨을 알 수 있었으며 모음이 ‘ㅣ’인 음성에서는 초·중·중성 모두 2.0 kHz 이상의 주파수 특성을 갖음을 분석할 수 있었다.

그림 8은 음성데이터 “낙”-“낭”의 초성에서 추정된 스펙트럼 분석 결과로서 X축은 주파수 Y축은 주파수의 크기, Z축은 “낙”부터 “낭”까지 음성을 나타낸다. 그림 8로부터 알 수 있듯이 초성이 모두 “ㄴ”인 자음을 갖는 음성이기 때문에 모든 음성이 제 2포먼트는 고주파 특성인 3 kHz에 스펙트럼을 갖고 있으며 제 1포먼트인 저주파수에도 유사한 특성을 갖고 있음을 확인할 수 있었으며, 모든 초성의 스펙트럼 분석결과 동일 자음의 경우 유사한 특성을 갖고 있음을 확인할 수 있었다.

본 논문에서 중성의 스펙트럼 분석은 중성이 동일 자음의 경우 스펙트럼이 유사한가를 평가하기 위한 비교 분석을 하였으며 그 결과는 그림 9와 같다.

그림 9에서 X, Y, Z축의 값은 표 5와 같으며 그림으로부터 알 수 있듯이 저주파 성분 즉 X축의 1-6까지인  $F_1$ 은 대부분이 1 kHz 미만인 특성을 갖음을 알 수 있었고, 고주파 성분인 X축의 7-12까지인  $F_2$ 는 2.5 kHz의 특성을 갖음을 분석하였다. 또한 음성 “기”의 경우 그림 7의 극점 평가값으로부터 알 수 있었듯이 다른 음성에 비하여 스펙트럼 특성이 높음을 알 수 있고 음성 “작”의 경우는 반대로 스펙트럼 특성이 낮음을 알 수 있다.



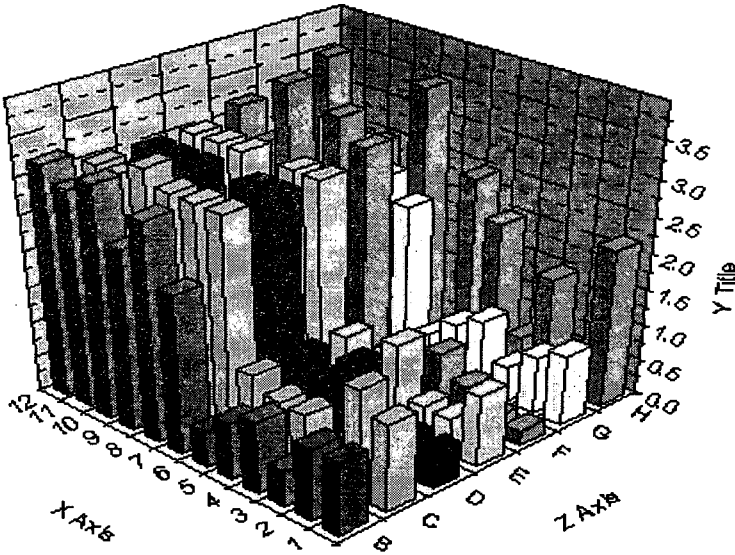


그림 9. 음성신호의 3분할에 의한 중성의 스펙트럼 비교분석

#### 4. 결 론

본 논문에서는 단구간 에너지 평가와 단구간 자기상관함수의 극점평가에 의한 발성상의 초성, 중성, 종성을 적용적으로 분할하는 방법을 제안하였다. 제안한 방법을 자음, 모음, 자음으로 이루어진 음성신호에 적용하여 발음상 초·중·종성으로 분할하고 분할구간별로 특징 분석을 수행한 결과 다음과 같은 결론을 얻었다.

첫째, 제안한 방법을 적용하여 음성신호를 발음상 초·중·종성으로 자동분할 시킨 후 분할된 구간별 청취와 스펙트럼 분석으로부터 제안한 분할방법이 음성신호를 적용적으로 분할할 수 있음을 확인하였다.

둘째, 분할된 음성신호의 초·중·종성의 구간별 스펙트럼 분석실험에서 음성의 초성에는 자음성분에 의한 스펙트럼 특성이, 중성에서는 모음성분에 의한 특성과 종성은 모음과 자음성분이 혼합된 스펙트럼 성분이 존재한다.

셋째, 초성이 동일한 자음을 갖는 음성의 경우 스펙트럼 특성이 유사함을 알 수 있었고, 초성과 중성이 다르고 종성이 동일한 자음을 갖는 음성의 경우 중성인 모음의 영향에 의하여 약간의 차이가 있으나 유사한 주파수 특성을 갖는다.

넷째, 단구간 자기상관함수의 극점 평가값이 큰 경우 음성신호의 주파수 특성이 4 kHz 부근의 높은 특성을 갖고 있음을 알 수 있고, 극점 평가값 낮으면 음성신호도 상대적으로 낮은 주파수 특성을 갖고 있음을 알 수 있었으며, 극점 평가에는 자음보다 모음성분에 영향이 크게 미침을 알 수 있었다.

본 논문은 단구간 자기상관함수의 극점평가와 단구간 에너지 평가에 의한 발성상의 초



성, 중성, 종성을 적응적으로 분할하는 방법을 제안하였고 발성상의 초·중·종성 구간별 여러 특징분석을 수행하였다. 향후 제안한 방법을 이용하여 차동음소 분류 및 대응량 어휘연구에 보다 많은 연구가 이루어져야 할 것이다.

### 참 고 문 헌

- [1] Mukund Padmanabhan and Lalit R. and David Nahamoo and Michael A. 1998. "Speaker Clustering and Transformation for Speaker Adaptation in Speech Recognition System." *IEEE Trans. on Speech and Audio Processing* 5(1), 71-77.
- [2] V. Digalakis and L. Neumeyer. 1995. "Speaker adaptation using combined Transformation and Bayesian methods." in *Proc. ICASSP*, Detroit, MI, 680-683.
- [3] T. Kosaka and S. Sagayama. 1994. "Tree-structured speaker clustering for fast speaker adaptation." in *Proc. ICASSP*, Adelaide, Australia, 245-248.
- [4] G. Zavagliogkos and R. Schwartz and J. Makhoul, 1995. "Batch, incremental and instantaneous adaptation techniques for speech recognition." in *Proc. ICASSP*, Detroit, MI, 676-679.
- [5] Rathinavelu Chengalvarayan and Li Deng. 1998. " Use of Generalized Dynamic Feature Parameters for Speech Recognition." *IEEE Trans. on Speech and Audio Processing* 5(3), 232-242.
- [6] Yaxin Zhang and Roberto Togneri and Michael Alder. 1997. "Phoneme-Based Vector Quantization in a Discrete HMM Speech Recognizer." *IEEE Trans. on Speech and Audio Processing* 5(1), 71-77.
- [7] Biing-Hwang Jung and Wu Chou and Chin-Hui Lee. 1997. "Minimum Classification Error Rate Methods for Speech Recognition", *IEEE Trans. on Speech and Audio Processing* 5(3), 257-265.
- [8] L. Bahl and P. de Souza and P. Gopalakrishnam and M. Picheny. 1993. "Context-dependent vector quantization for continuous speech recognition." in *Proc. ICASSP*, Minneapolis, MN, 632-635.
- [9] 김창석 외 4인. 1998. "변곡점 및 단구간 에너지평가에 의한 음성의 천이구간 특징분석." *음성과학* 3, 156-166.
- [10] 성종모 외 1인. 1997. "자동 음성분할 및 레이블링 시스템의 구현." *한국음향학회지* 16(5), 50-59.
- [11] T. Applebaum and B. Hanson. 1991. "Regression features for recognition of speech in quiet and in noise." *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol.2, 985-988.
- [12] S. Furui. 1986. "Speaker independent isolated word recognition using dynamic features of speech spectrum." *IEEE Trans. on Acoust., Speech Signal Processing* vol. ASSP-34, 52-59.
- [13] C. Lee and L. Rabiner and J. Wilpon. 1990. "Acoustic modeling for large vocabulary speech recognition." *Comput. Speech Lang.*, vol 4, 127-165.
- [14] K. Lee and H. Hon. 1990. "Context-dependent phonetic hidden Markov models for

- continuous speech recognition." *IEEE Trans. on Acoust., Speech Signal Processing* vol. 38, pp. 590-609.
- [15] 김창석 외 3인. 1994. "음성신호의 최적특징을 적응적으로 추출하는 방법에 관한 연구." *한국통신학회논문지*19(2), 373-380.
- [16] 김창석외 3인. 1995. "최적의 C-LPC계수 추출 알고리즘의 제안." *한국통신학회논문지* 20(1), 209-220.
- [17] Y. T. Ting, D. G. Childers. 1990. "Speech analysis using the weighted recursive least squares algorithm with a variable forgetting factor." *Proc. ICASSP*, 389-392.
- [18] Y. S. Cho, S. B. Kim, and E. J. Power. 1991. "Time-varying spectral estimation using AR models with variable forgetting factors." *IEEE Trans. Signal Processing*, vol. SP-39, 1422-1426.
- [19] 최홍섭. 1997. "음성천이구간에서의 성도 파라미터 시변추정에 관한 연구." *한국음향학회논문지*, vol.16(2), 101-106.
- [20] 김창석 외 4인. 1998. "변곡점 및 단구간 에너지평가에 의한 천이구간 특징 분석." *한국음성과학학회논문지*, vol. 3, 156-166.
- [21] J. R. Deller, Jr. et al. 1993. *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, 266~343.

접수일자 : '99. 2. 10.

게재결정 : '99. 3. 19.

▲ 장 승 관

경기도 시흥시 정왕동 3가 101  
한국산업기술대학교 전기전자공학과(우: 429-450)  
Tel : (0345) 4968-262 (O), Fax : (0345) 4968-279  
e-mail : skjang@kpu.ac.kr

▲ 최 성 연

경기도 시흥시 정왕동 3가 101  
한국산업기술대학교 전기전자공학과(우: 429-450)  
Tel : (0345) 4968-266 (O), Fax : (0345) 4968-279  
e-mail : sychoi@kpu.ac.kr

▲ 김 창 석

경기도 용인시 남동 산38-2  
명지대학교 전자정보통신공학부(우: 449-728)  
Tel : (0335) 37-2300 (O), Fax : (0335) 30-6374  
e-mail : kimcs@wh.myongji.ac.kr