# 음성 신호의 음소 단위 구분화에 관한 연구

# A Study on the Segmentation of Speech Signal into Phonemic Units

이의천*, 이강성*, 김순협*

(Yeui Cheon Lee, Gang Sung Lee, Soon Hyon Kim)

## 요약

본 연구에서는 음성신호의 음소 단위 구분화 방법을 제안한다. 제안된 구분화 시스템은 화자 독립적이고, 음성신호에 대한 사전 정보 없이도 음소 단위로 구분화를 수행할 수 있는 특성을 갖는다.

구분화 처리는 입력 음성신호를 먼저 순수 유성음 구간과 순수 유성음이 아닌 구간으로 분리 시킨 후, 각각의 구간에 대해 세분화된 음소 단위로 분리시키는 2단계 구분화 알고리즘을 적용하였고, 이때 사용된 파라미터는 유성음 검출 파라미터, 영차 LPC 켑스트럼 계수의 시간변화 파라미터, ZCR 파라미터이다.

본 연구에서 제안한 구분화 알고리즘의 유용성을 입증하기 위해 사용한 대상어는 고립단어와 연속음성으로 구성된 어휘로서 전체 어휘중에 포함된 507개 음소에 대한 구분화율은 91.7% 이다.

## ABSTRACT

This paper suggests a segmentation method of speech signal into phonemic units. The suggested segmentation system is speaker-independent and performed without anyprior information of speech signal.

In segmentation process, we first divide input speech signal into purevoiced region and not pure voiced speech regions. After then we apply the second algorithm which segments each region into the detailed phonemic units by using the voiced detection parameters, i.e., the time variation of 0th LPC cepstrum coefficient parameter and the ZCR parameter.

Types of speech, used to prove the availability of segmentation algorithm suggested in this paper, are the vocabulary composed of isolated words and continuous words. According to the experiments, the successful segmentation rate for 507 phonemic units involved in the total vocabulary is 91.7%

## I. 서 론

본 연구는 음성 인식에서 가장 필요로 하는 음소를 인식 단위로 하는 대용량 단어 인식 시스템이나 연속 음성 인식 시스템에서 음성을 음소로 분리하는데 기본이 되는 음소 분리 알고리즘을 구현하고 그 성능을 측정하는데 목적이 있다.

연속 음성이나 대어휘 단어 음성 인식을 위해서 단어 단위로 인식을 한다면 그 표준 패턴을 위한 기억 용량 및 계산 시간을 상당히 많이 필요로 할 것이다. 그러나 부단어 단위인 음소 또는 음절등으로

---

*광운대학교 전자계산기공학과

경계를 구분할 수 있다면 보다 인식시스템의 성능이 나아질 것이다.

지금까지 연구된 자동 음성 구분화 방법에는 음소 또는 음절 패턴을 미리 추출하여 저장하고 있다가 입력된 음성의 스펙트럼을 동적 계획법으로 비교하여 경계를 결정하는 템플레이트 매칭 방법[1][2]과 주파수 대역 에너지와 시간 변화 파라미터들을 이용하여 음소 경계 및 대분류를 수행하는 스펙트럼 천이 측정 방법[3][4], 그리고 여러개의 파라미터를 이용하여 음운의 경계를 검출하는 대략적인 분류(broad classification) 방법[5]등이 있다.

기존의 제안된 구분화 방법에 있어서, 템플레이트 매칭 방법은 구분화가 되면서 그 결과로 자동적으로 레이블링 된다는 장점이 있으나 추출 가능한 대상의 음소 또는 음절의 패턴을 미리 가지고 있어야 한다. 또한 음소 단위를 인식의 기본단위로 정한 경우 한음소가 다양하게 변화되는 모든 음소를 추출하지 않으면 신뢰할 만한 결과를 얻기 어렵다. 더구나 이 방법은 사람마다 음성의 패턴이 다르므로 화자 독립적으로 적용하기 힘들어 화자가 바뀜에 따라 새로 작성해야 하는 불편함이 있다.

스펙트럼 천이 측정 방법은 음성신호에 대한 자료를 미리 알고 있지 않아도 구분화를 수행하는 알고리즘으로 제안되었는데 음성신호가 약간만 변해도 실제의 음소와 비교하여 볼 때 구분화가 지나치게 많이 되어 추가의 보완 처리가 필요한 단점을 갖는다.

대략적인 분류 방법은 많은 파형과 스펙트럼 변화 파라미터를 통해 구분화 및 대분류를 수행하는 방법으로 파라미터의 수가 너무 많고 각각에 대해서 임계치를 구하는 작업이 쉽지 않으며, 또한 유성음 구간에서 음소 분리를 제대로 수행하지 않고 있어 연속적인 모음이나 모음과 비음 또는 유음사이를 잘 구분하지 못하는 단점이 있다.

이외에도 다수의 구분화 알고리즘[6][7][8]이 나와 있으나 유성음 사이의 구분에 관한 내용은 미흡하게 다루어져 있다.

본 연구에서 제안된 구분화 시스템은 화자 독립적이고, 음성에 대한 사전 정보없이 음소 단위로 구분화를 수행할 수 있다. 음성신호를 먼저 유성음 검출 파라미터에 의해서 순수 유성음 구간과 순수 유성음이 아닌 구간(유성자음, 무성음, 묵음구간)으로 나누고, 2분된 각각의 구간에 대해서 다른 파라미터와 적절한 알고리즘을 적용하여 더욱 자세한 음소 단위와 구분화를 수행한다. 입력된 음성신호로부터 음성의 특징계수(LPC)를 추출하고 이로부터 음소를 분리하는데 필요한 파라미터를 계산한 후, 음소 분리 알고리즘을 적용하여 음소 단위로 경계를 구분한다.

## Ⅱ. 음성신호의 분석

음성신호의 전처리 과정에서 고주파 성분의 잡음 및 신호 중첩 효과를 제거하기 위해 3.5kHz의 저역여파기를 사용한 다음, 8kHz 샘플링과 12비트로 A/D 변환을 수행하였다. LPC 분석은 16msec를 하나의 프레임으로 8msec씩 이동하면서 10차로 분석되었다. Preemphasis에서는 전달 함수 $H(z)=1-az^{-1}(a=0.95)$인 1차 디지탈 시스템으로 처리하여 고주파 영역의 에너지를 증폭시킨다. 프레임 처리는 128 샘플(16msec)을 하나의 프레임으로 하여 특징 분석을 행한다. 창 함수로는 전형적인 smoothing window인 Hamming 창 함수를 사용하였다.

## Ⅲ. 구분화 파라미터 추출

### 3.1 유성음 검출 파라미터

전처리 과정을 거친 음성파형은 16ms을 한 프레임으로 8msec씩 이동하면서 10차의 LPC 계수를 구한 다음 FFT를 취해 LPC 스펙트럼을 구한다. LPC 스펙트럼의 식은 다음과 같다.

$$X(k) = \sum_{n=0}^{N-1} x(n)\, W_N^{kn}, \quad k = 0, 1, 2, \ldots, N-1 \tag{1}$$

여기서

$$W_N^{kn} = e^{-j2\pi/N} \quad (N = 128) \tag{2}$$

$x(0)=1,\ x(1)=a_1,\ x(2)=a_2, \ldots, x(10)=a_{10},$

$x(11)=0, \ldots,\ x(N-1)=0$

LPC 대수 스펙트럼의 형태로 표시된 스펙트럼 [illegible]

$$V_i = \frac{1}{5}\sum_{k=1}^{5} X_i(k) \qquad (3)$$

가 순수 유성음 검출 파라미터이고, 식(3)에서 구하는 주파수 대역 양단의 주파수 $\omega_l$, $\omega_s$은 각각 62.5Hz와 312.5Hz에 해당한다.

### 3.2 영차 LPC 켑스트럼 계수

청각의 특성을 고려하기 위해 스펙트럼의 저역 부분을 강조하는, 영차 LPC 켑스트럼 계수를 이용한다. i번째 프레임의 영차 LPC 켑스트럼 계수 $C_i$는

$$C_i = x_i[0] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_i(\Omega)\ W(\Omega)\ d\Omega \qquad (4)$$

으로 표시된다. 영차 LPC 켑스트럼 계수 $C_i$는 직선 주파수 눈금상에 표시된 LPC 대수 스펙트럼에 큰 가중치를 붙인 평균치다[10].

### 3.3 영차 LPC 켑스트럼 시간 변화 파라미터

i번째 분석 프레임을 중심으로 하는 영차 LPC 켑스트럼 계수의 시계열 $C_{i+n}(|n| \le M)$에 대하여 직선식

$$x = A_i n + B \qquad (5)$$

를 적용해 회귀계수 $A_i$에 의해서 영차 켑스트럼 계수의 시간 변화의 양을 표현한다.

직선 적용 식에서 시계열 구간 양단의 절단 영향을 작게 하기 위해 평가 함수인 가중된 2승 평균 오차

$$\varepsilon = \frac{1}{2M+1}\sum_{n=-M}^{M} W_n(A_i n + B - C_{i+n})^2 \qquad (6)$$

을 이용한다. 여기서 $W_n$은 $|n| > M$에서 0이 되는 [illegible] 된 2승 평균 오차 ε를 최소로 하는 조건에서

$$A_i = K_M \sum_{n=-M}^{M} W_n\ n\ C_{i+n}\ ,$$

$$\left( K_M = \left(\sum_{n=-M}^{M} W_n\ n^2\right)^{-1} \right) \qquad (7$$

로 주어진다.

### 3.4 영교차율

스펙트럼에서 에너지가 집중되는 주파수를 찾는데 유용한 특징 파라미터로 사용되는 영교차율[11]은 성량에 적게 의존하며 음성 발생시 성대에 유성 스펙트럼이 감소되므로 순수 유성음의 에너지는 주로 3 KHz 이하에 집중된다. 그림 2에 /사랑스러운/의 음성신호와 세그멘테이션 파라미터 vi, ai, zi의 파형을 나타내었다.

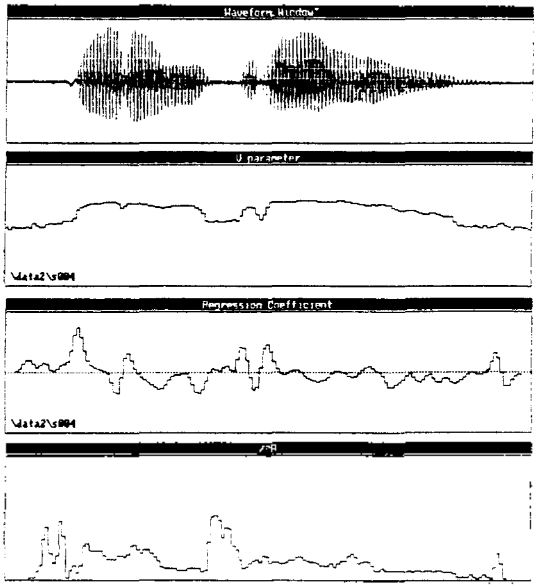

그림 1. '사랑스러운'의 음성 신호와 파라미터 파형
Fig 1. Speech signal of / sarangseureowoon / and waveform of parameter

## Ⅳ. 구분화 알고리즘의 구현

### 4.1 구분화 알고리즘의 개요

유성 신호를 먼저 순수 유성음 구간과 비 순수 유성음 구간으로 구분한 뒤, 2분된 각각의 구간에 대해서 더욱 자세한 음소 단위의 경계들로 분리한다. 구분화 알고리즘의 흐름도를 그림 2에 나타내었다.

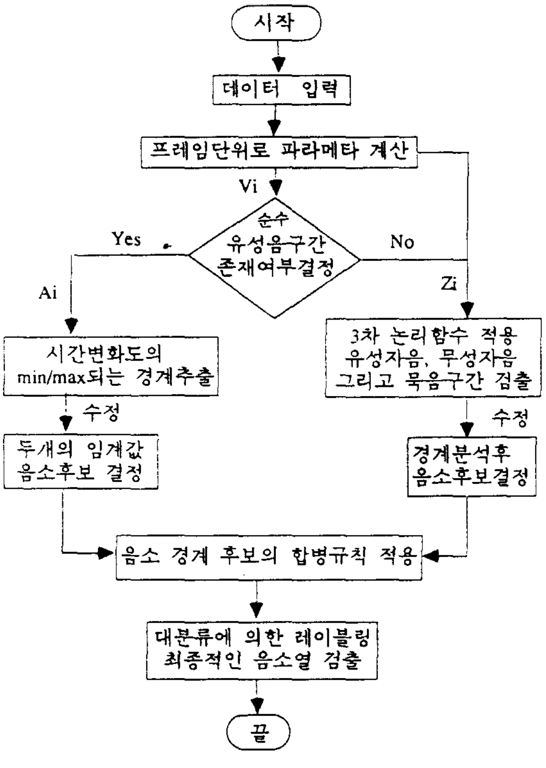

그림 2. 구분화 시스템의 구성도
Fig 2. Flow chart of the segmentation algorithm

### 4.2 순수 유성음의 분리

유성음 검출 파라미터에서 두개의 임계값($T_{VL}$=57, $T_{VH}$=62)을 가고 Rabiner와 Sambur[11]의 끝점 검출 알고리즘 방식을 적용하여 순수 유성음을 분리하였다.

### 4.3 순수 유성음의 구분화

순수 유성 구간에서 음소경계후보의 대부분은 영차 켑스트럼 시간 변화 파라미터 $a_i$가 극대치 또는 극소치를 취하는 분석 프레임에 의해 주어진다. 일반적으로 모음과 자음의 경계로는 극소치, 계속되는 자음의 경계에는 극대치를 취하나 예비실험에 의하면 그들 2개중 현저한 차이가 있는 경우에는 극치로 음소후보경계가 될 수 있다.

우선 영차 켑스트럼 시간 변화 파라미터 $a_i$의 극치를 표시하는 함수 $A_i$를

```
if (( $a_i > 0$ and $a_{i-1} < a_i \ge a_{i+1}$ ) or
   ( $a_i < 0$ and $a_{i-1} > a_i \le a_{i+1}$ ))
   then $A_i = a_i$ , else $A_i = 0$                    (8)
```

로 표시하고, 이것은 다음식

```
if ( $|A_i| < T_{A1} / 2$ )
   then $A_i = 0$
else if ( $A_i \ge T_{A1} / 2$ and $A_i < T_{A1}$
   and $i-5 \le l < i$, $|A_i - A_l| < T_{A2}$)
   then $A_i = 0$
else if ( $A_i < T_{A1} / 2$ and $A_i > -T_{A1}$
   and $i-7 \le l < i$, $|A_l - A_i| < T_{A2} * .7$)
   then $A_i = 0$
 ( $T_{A1} = 1.5$ ,   $T_{A2} = 3.0$ )
```

에 의해서 수정된다. 수정된 값 $A_i$는 파라미터 $a_i$의 극치 함수이다.

### 4.4 비 순수 유성음의 구분화

비 순수 유성음을 유음구간과 무음구간으로 구분하고, 유음구간에서 다시 평가하여 유성자음구간과 무성자음구간으로 분류한다. 이를 위해 다음의 3차 논리함수를 적용한다.

$$Z_i^T = U(Z_i - T_{ZL}) + U(Z_i - T_{ZH}) \quad (T_{ZL}=4,\ T_{ZH}=13) \qquad (10)$$

여기서 U(a) 함수는 a가 양수일 때 1, 그렇지 않을 때 0을 넘겨주는 함수이다. 이 $Z_i$값이 0이면 무음, 1이면 유성음, 2이면 무성음일 가능성이 높은 프레임 후보가 되며 이것은 다음 식으로 구성된다.

$$\text{if } ( Z_{i-1}^T + Z_i^T + Z_{i+1}^T <= 1) \quad \text{then } Z_i^T = 0 \tag{11}$$

### 4.5 대분류에 의한 레이블링과 음소 합병

본 연구에서 제안한 구분화 시스템에서는 표 1에 나타낸 바와 같이 음소 단위를 4개의 범위로 대분류하여 세그멘트의 레이블링을 수행한다.

제안된 음소 분리 알고리즘에 의해서 세그멘트된 분석 프레임 번호 i가 음소 경계후보가 되지만 그들의 경계 후보가 서로 접근하고 있는 경우에는 표 2에 나타낸 것과 같은 합병 규칙을 적용한다. 구분화 처리 결과를 그림 3에 나타내었다.

표 1. 음소 단위의 대 분류표
Table 1. Global classification of phonemic units

| 기호 | 내 용 | 해당되는 음소 단위 |
|---|---|---|
| V | 유성유음소 | 단모음, 이중모음, 유음, 경과음, 비음 |
| B | 유성자음음소 | 유성폐쇄음 |
| C | 무성자음음소 | 무성폐쇄음, 마찰음, 파찰음 |
| S | 묵 음 음 소 | 묵음 |

표 2. 병합규칙
Table 2. Merge Rule

| 구 분 | 내 용 | 결 과 |
|---|---|---|
| 유성음 합병 | V+V(7frame 미만)+무성음 | --> V+무성음 |
| 무성음 합병 | S, B, C+S, B, C | --> S,B,C로 합병 |
| | B(6frame 미만)+C(3frame 이상) | --> B 삭제 |
| | B(6frame 미만)+C(3frame 미만) | --> C 삭제 |
| | V 또는 C+B(3frame 미만) | --> B 삭제 |
| | B(3frame 미만)+V 또는 C | --> B 삭제 |
| | C(3frame 미만) | --> C 삭제 |
| | C(6frame 이상)+B(4frame 미만) | --> B 삭제 |

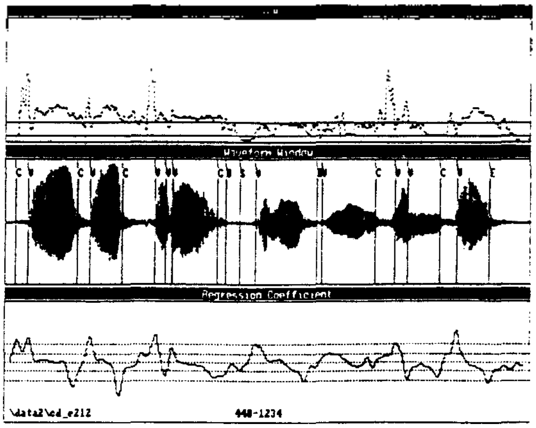

그림 3. / 448-1234 / 의 처리결과
Fig 3. The result of segmentation process

## V. 실험 결과 및 고찰

### 5.1 실험 환경 및 음성 데이터의 구성

시스템의 성능을 평가하기 위해 사용된 음성 데이터는 임의의 남성 화자 3명을 선택하여 방음 장치가 되어 있지 않고 컴퓨터 잡음이 있는 보통의 환경에서 다이나믹 마이크를 통해 저장되었다. 입력된 음성 파형의 10차 LPC 계수로부터 LPC 스펙트럼과 영차 LPC 켑스트럼 계수를 구해 이를 바탕으로 세그멘테이션 파라미터를 추출하였다.

본 연구의 실험 대상 어휘로는 고립단어와 연속 숫자음을 선정하였으며 그 내용은 표 3과 같다. 3종류의 자료를 각각 다른화자 3명이 1번 발음하였다. 전체 단어에 대응하는 음소단위의 세그먼트 갯수는 고립단어인 경우 44개이고, 연속 숫자음인 경우는 463개로 총 507개이다.

표 3. 실험 대상 어휘
Table 3. Experiment vocabulary

| 단 독 어1 | 단 독 어2 |
|---|---|
| 굳소리, 아이들, 부드러운<br>분위기, 사랑스러운 | 서울, 부산, 인천, 대구,<br>대전, 광주 |
| 연속숫자음 | |
| 512-0257, 630-1349, 734-6780<br>826-8318, 904-0371, 910-2388<br>843-4616, 729-5522, 607-7641<br>358-8736, 270-9483, 396-0011 | 408-6281, 689-6542, 209-1921<br>147-3324, 986-5066, 569-1775<br>795-9785, 448-1234, 153-0599 |

### 5.2. 구분화 결과 및 평가

제안된 구분화 시스템에 대한 실험 결과의 정확한 평가를 위해 기준 템플레이트의 생성이 필요하다. 이를 위해 실제로 D/A로 변환되어 들어온 음성 신호에서 세그먼트를 추출하기 위해 직접 헤드폰으로 청취하며 손으로 음성 파형에 대응하는 음소단위의 경계를 구분할 수 있는 세그멘테이션 유틸리티를 이용하였다.

구분화 시스템에서 얻어진 세그먼트의 경계가 손으로 구분된 세그먼트에 의해서 얻어진 결과와 일치하는지를 평가한다. 음소와 음소 사이에는 언제나 뒤 음소의 안정 구간으로 가기 위한 과도 부분이 존재하는데, 이러한 과도 구간안에서 검출된 경계는 프레임 이동을 허용한다. 이것은 화자에 따른 변화성이나 단어 환경에 기인한 음운 현상이나 조음 현상을 보완한 것이다. 앞에서 기술한 연속음성안에 포함되어 있는 507개의 음소를 구분화하여 손으로 구분된 경계와 일치하는가를 확인한 결과 구분화율은 91.7%였다. 구분화가 실패한 경우는 음소 경계의 탈락과 추가로 구분될 수 있는데, 이중 탈락은 25개, 추가는 17개로 탈락이 가장 큰 에러 발생요인이 되고 있다.

## I. 결 론

음성신호의 음소단위 구분화 방법을 제시하고, 제안된 시스템이 음소단위로 분리하는 것을 실험을 통해 확인하였다. 남성 화자 3명이 발성한 고립단어와 연속음성에 대한 구분화 실험결과, 음소단위의 탈락이 25개, 추가가 17개로 전체 에러율은 약 8.3%이다.

본 연구에서 제안된 구분화 시스템의 주요 특징은 다음과 같다.

(1) 모든 임계값들은 화자 독립적으로 사용할 수 있도록 선정하였다.

(2) 음성신호에 대한 사전 지식이 없더라도 구분화가 가능하다.

(2) 음성 특성을 효과적으로 나타내는 파라미터를 얻기 위해 저주파수 대역 부분에서 파워 스펙트럼 에너지를 검출하였으며, 시간 변화 파라미터를 이용하여 모음구간에서 구분화를 하였다.

## 참 고 문 헌

1. T. svendson and F.K. soong, "On the Automatic Segmentation of Speech Signals", Proc. ICASSP-87, pp.77-80, Dallas, 1987.
2. 한국과학기술원, "Template Matching과 Vowel Segmentation을 적용한 한국어 음소분리 Algorithm", 최종보고서 CRL-P-8902, pp.31-40, 1989. 1.
3. F.K. Soong, Rosenberg, A.E., "On the Use of Instaneous and Transitional spectral Information in Speaker Recognition", ICASSP-86, Vol.2, pp.17.5.1-17.5.4, Tokyo, 1986.
4. Regine Andre-Obrecht, "A New Statistical Approach for Automatic Segmentation of Continuous Speech Signals", IEEE, Trans. on acoustic, Speech, Signal Processing, Vol.36, No.1, Jan 1988.
5. Ronald A. Cole and Lily Hou, "Segmentation and Broad Classification of Continuous Speech", ICASSP-88, s10.12, 1988.
6. Chorkim Chan, "Voiced / Unvoiced Segmentation", ICASSP 86 TOKYO, pp.42.12.1-42.12.4, 1986.
7. S. Roucos and M. O. Dunham, "A Stochastic Segment Model for Phoneme-Based Continuous Speech Recognition", Proc. ICASSP 87, pp.3.3.1-3.3.4, BBN Lab
8. Aage Bendiksen and Kenneth Steiglitz, "Neural Networks for Voiced / Unvoiced Speech Classification", IEEE, Trans. on acoustic, Speech, Signal Processing, 1990.
9. Seiich NAKAGAWA, Mitsunori SAKAMOTO, "Evalution of FFT Cepstrum and LPC Cepstrum for speech Speaker Recognition", 일본전자학회논문집, Vol. J66-A No.12, pp.1199-1206, 1983.
10. Satoshi IMAI and Chieko FURUICHI, "Segmentation of Continuous Speech Phonemic Units", 電子情報通信學會論文誌. '89 / 1. Vol.J 72-D-11, NO.1
11. L.R.Rabiner and M.R.Sambur, "An Algorithm for Determining the End Points of Isolated Utterances", Bell Syst. Tech. J., Vol. 54, No.2, pp297-315, Feb. 1975.

**▲이 의 천(정회원)** 1962년 6월 20일생

1986년 2월 : 광운대학교 전자계산기공학과 졸업(공학사)

1991년 2월 : 광운대학교 대학원 전자계산기공학과 졸업(공학석사)

1991년 2월～현재 : 금성정보통신(주) 안양연구소 연구원

**▲이 상 신(정회원)** : 제 10권 2호 참조

**▲[illegible](정회원)** : 제 10권 2호 참조