

프랙탈 차원을 이용한 단음절 음성의 자·모음 분리

Consonant/Vowel Segmentation in Monosyllabic Speech Data Using the Fractal Dimension

최철영*, 김형순*, 김재호*, 손경식*

(Chul Young Choi*, Hyung Soon Kim*, Jae Ho Kim*, Kyung Sik Son*)

요 약

본 논문에서는 음성신호의 프랙탈 차원을 이용하여 한국어 CV(Consonant-Vowel) 단음절에서 자음과 모음을 분리하는 실험을 하였다. 프랙탈 차원은 Minkowski-Bouligand 차원을 사용하였으며, 형태학적 커버링(morphological covering) 방법을 이용하여 구하였다. 프랙탈 차원의 음성분리에 있어서의 유용성을 조사하기 위하여 프랙탈 차원과 단구간 에너지 각각을 이용한 음성분리 실험과 에너지와 프랙탈 차원을 같이 이용한 음성분리 실험을 하여 그 결과들을 비교하였다. 실험 결과 에너지의 기울기를 사용한 경우는 88.0%의 바른 분리 결과를 보였고, 프랙탈 차원의 기울기를 사용한 경우는 그 보다 더 나은 93.6%의 바른 분리 결과를 보였으며, 에너지의 기울기와 프랙탈 차원의 기울기의 곱을 사용한 경우는 96.1%로 가장 높은 바른 분리결과를 나타냈다. 이를 통해 프랙탈 차원이 음성신호의 분리에 있어서 하나의 유용한 파라메타가 될 수 있음을 확인하였다.

ABSTRACT

In this paper, we performed a class of experiments on segmenting consonant and vowel from Korean consonant-vowel(CV) monosyllable data, using the fractal dimension of the speech signals. We chose the Minkowski-Bouligand dimension as the fractal dimension, and computed it using the morphological covering method. In order to examine the usefulness of the fractal dimension in speech segmentation, we carried out speech segmentation experiments using the fractal dimension alone, using the short-time energy alone, and using both the fractal dimension and the short-time energy, and compared the results. From the experiments, segmentation accuracy of 96.1% was achieved for the case with using the multiplication of the slope of the fractal dimension and that of the energy, while the segmentation accuracies for the cases with using the slope of either the fractal dimension or energy alone were slightly lower (93.6%) or much lower (88.0%) than the above case, respectively. These results indicate that the fractal dimension can be used as a good parameter for speech segmentation.

I. 서 론

*부산대학교 전자공학과
Dept. of Electronics Engineering Pusan National University

접수일자: 1993년 3월 2일

프랙탈 기하학은 아주 불규칙하고 복잡하게 보이

는 자연현상 내에서도 그속의 어떤 법칙과 규칙성을 발견하여 그러한 자연현상을 해석하기 위한 학문분야 중의 하나로서, 1970년대 중반 Mandelbrot^[1]에 의해 제창되고 이후 Barnsley^[2]등 다른 여러 사람에게 의해 발전되어 왔다. 프랙탈은 부분의 모습이 전체와 닮은 모습을 가지는 형상으로서 구름, 산, 나무, 해안선 등이 대표적인 예이며, 그 외에도 주식시세 변동, 인구증가 변동, 음성 신호 및 일부의 잡음 신호 등의 시계열(time series) 데이터들이 프랙탈로 해석될 수 있다. 그리고, 프랙탈의 닮은 정도를 나타내는 프랙탈 차원은 프랙탈을 해석하는데 있어서 매우 중요한 역할을 한다. 닮은 정도가 큰 프랙탈일수록 높은 프랙탈 차원을 가지며 보다 거칠고 불규칙하게 보인다.

근래에 들어 음성신호처리 분야에서 음성도의 프랙탈 특성을 이용하려는 연구가 점차적으로 이루어지고 있다.^{[3][4]} 음성의 발생 과정에는 공기 흐름의 비선형적인 역학성으로 인한 어떤 혼란이 존재한다.^[3] 이러한 특성이 음성신호에 반영되어 음성신호를 여러 배율(magnification)의 크기에서 볼 때 서로 유사한 모습이 나타난다. 따라서, 음성신호는 프랙탈로 해석될 수 있으며 프랙탈 차원을 새로운 파라메타로서 음성분리(speech segmentation)를 비롯한 음성신호처리 분야에 이용할 수 있다.

이를 바탕으로 본 논문에서는 음성신호의 프랙탈 차원을 이용하여 초성 자음과 모음으로 구성된 CV (Consonant-Vowel) 단음절에서 자음과 모음을 분리하는 실험을 수행하였다. 일반적으로 음성신호에 있어서의 자음과 모음은 서로 다른 닮은 정도와 불규칙성을 가지고 있으며 그로 인하여 서로 다른 프랙탈 차원을 가진다. 그러므로 프랙탈 차원은 자음과 모음의 경계에서 큰 변화를 가지는데 이를 이용하여 프랙탈 차원 곡선의 기울기를 자음과 모음의 분리 실험에 적용하였다. 그리고 자음과 모음에서의 에너지의 특성의 차이를 이용하여 에너지 곡선의 기울기를 이용한 자음과 모음의 분리실험을 함께 수행하였으며, 또한 음성신호의 프랙탈 차원과 에너지 특성을 함께 사용한 음성분리 방법을 제안하였다. 실험 결과 자음 및 모음의 분리에 있어서 프랙탈 차원의 유용성을 확인하였다. 본 논문에서는 음성신호의 프랙탈 차원을 구하는 방법으로서 다른 프랙탈 차원보다 구현이 간단하고 강인한 특성을 지니는 것으로 알려진 Minkowski-Bouligand 차원을 사용하고 이를 형태학적 커버링(Morphological Covering)방법을 이용하여

구하였다.^[5]

본 논문의 구성은 다음과 같다. 서론에 이어 2장에서는 음성신호의 프랙탈 차원을 계산하는 방법을 설명하고, 3장에서는 프랙탈 차원을 이용한 자음과 모음의 분리 방법에 대해 설명한다. 그리고 4장에서는 분리 실험결과를 검토하여 프랙탈 차원의 음성분리에 있어서의 유용성을 조사하고, 마지막으로 5장에서 결론을 맺는다.

II. 음성신호의 프랙탈 차원 계산

음성신호의 프랙탈 차원을 구하기에 앞서, 본 장에서는 먼저 프랙탈 차원의 일반적인 개념과 계산이 간단하면서도 강인한 특성을 갖는 것으로 알려진 Minkowski-Bouligand 차원에 대해 논의한다. 그 다음으로 이산신호 형태의 음성신호의 프랙탈 차원을 구하기 위한 방법으로서 Maragos와 Sun이 제안한 형태학적 커버링 방법에 의해 Minkowski-Bouligand 차원을 계산하는 과정을 살펴본다.^[5]

1. 프랙탈 차원의 개념 및 Minkowski-Bouligand 차원 평면상의 한 프랙탈 곡선 X에 대한 프랙탈 차원의 개념은 다음과 같이 설명될 수 있다. 프랙탈은 부분이 전체와 닮은 모습을 가지고 있는 것이므로 X를 확대하여 보았을 때 확대하기 전의 X와 닮은 모습을 지니고 있을 것이다. 이때 X의 길이를 구하고자 한다. 길이가 ϵ 인 측정자를 가지고 곡선 X를 따라 가면서 놓아갈 때, X를 전부 다 덮을 수 있는 측정자의 갯수를 $N(\epsilon)$ 개라고 하자. 그러면 이 측정자에 의한 X의 길이 $L(\epsilon)$ 은 $N(\epsilon) \times \epsilon$ 로 주어진다. 이때 측정자의 길이 ϵ 을 점차로 작게 하여 나가보자(이는 X를 확대하여 보는 것과 같다). 그러면 X의 아주 미세한 부분까지 측정됨으로 인하여 ϵ 이 감소하는 비율보다 $N(\epsilon)$ 이 증가하는 비율이 더 크게 되어 측정되는 X의 길이는 점점 늘어나게 된다. 이때 프랙탈 곡선 X의 길이 $L(\epsilon)$ 은 ϵ 이 영으로 수렴함에 따라 다음 수식으로 근사화 될 수 있다.

$$L(\epsilon) \approx (\text{상수}) \times \epsilon^{1-D}, \quad \epsilon \rightarrow 0. \quad (1)$$

여기서 상수 D를 X의 프랙탈 차원이라 부르며, D가 크면 클수록 ϵ 의 감소에 대한 $L(\epsilon)$ 의 증가 비율이 커진다. 일반적으로 잘 알려져 있는 프랙탈 차원들은

모두 (1)식에 기초를 두고 있으며 (1)식에서 길이를 어떤 방법으로 구하느냐에 따라 서로 조금씩 다른 특징을 가지고 있다. 프랙탈 차원은 Hausdorff-Besicovitch 차원, Minkowski-Bouligand 차원, Box Counting 차원, Entropy 차원 등의 여러 가지가 있으나, 그 중에서 Minkowski-Bouligand 차원이 다른 방법들에 비해 구현이 간단하고 강인한 특성을 가지는 것으로 알려져 있으며^[5] 이하에 Minkowski-Bouligand 차원에 대해 보다 상세하게 설명한다.

Minkowski-Bouligand 차원의 경우, 앞서 설명한 프랙탈 곡선 X의 길이를 측정하기 위해 X의 Minkowski 덮개(cover)를 사용한다. X의 Minkowski 덮개는 X를 반경 ϵ 의 원으로서 형태학(morphology)적으로 팽창(dilation)시킨 것과 같다. 그리고 Minkowski 덮개 면적을 $A(\epsilon)$ 라고 하고, X의 길이는 $\lim_{\epsilon \rightarrow 0} L(\epsilon)$ 로 주어진다. 여기서 $L(\epsilon)$ 는 반경 ϵ 인 원을 이용한 Minkowski 덮개를 사용했을때의 X의 길이의 측정치로서 $L(\epsilon)$ 는 $L(\epsilon) = A(\epsilon)/2\epsilon$ 라는 관계가 성립한다. 이때 (1)식의 개념으로 부터 Minkowski-Bouligand 차원 D_M 은 다음과 같이 정의된다.

$$D_M = \lim_{\epsilon \rightarrow 0} \left(\frac{\log [A(\epsilon)/\epsilon^2]}{\log 1/\epsilon} \right) \quad (2)$$

2. 이산신호의 형태학적 커버링 방법

형태학적 커버링 방법은 Minkowski-Bouligand 차원 D_M 을 구하는 과정에서 원을 이용한 Minkowski 덮개 대신에 다른 임의의 형태를 가지는 평면내 집합을 이용한 형태학적 덮개(morphological cover)를 사용한 것으로 연속신호 뿐만 아니라 이산신호에 대해서도 효율적으로 적용될 수 있다.

$f[n]$, $n=0, 1, \dots, N$ 을 유한한 길이의 이산신호라고 하자. 이때 $(n, f[n])$ 은 Cartesian 좌표의 (x, y) 에 해당한다. 그리고 볼록(convex)하고, X, Y축에 대해 대칭적인 평면내의 이산집합 B에 대해 $\epsilon B = \{ \epsilon b : b \in B \}$ 으로 주어진다. 여기서 ϵ 는 1, 2, ... 의 정수이다. 이때 형태학적 팽창 연산자인 \oplus 을 사용한 $f[n]$ 의 형태학적 덮개 $C_B[\epsilon]$ 는 다음과 같이 정의된다.

$$C_B[\epsilon] \equiv f[n] \oplus \epsilon B \quad (3)$$

여기서 만약 B가 원이라면 $C_B[\epsilon]$ 는 앞서 설명한 Minkowski 덮개와 같다.

다음으로 $C_B[\epsilon]$ 의 포락선을 이용하여 $C_B[\epsilon]$ 의 면적인 $A[\epsilon]$ 를 구한다. $g[n]$ 을 B의 상위 포락선이라고 하고 $g_\epsilon[n]$ 을 ϵB 의 상위 포락선이라 할때 $C_B[\epsilon]$ 의 상위 포락선은

$$f \oplus g_\epsilon = f \oplus g^{\oplus \epsilon} = ((f \oplus g) \oplus g \dots) \oplus g \quad (4)$$

이며 하위 포락선은

$$f \ominus g_\epsilon = f \ominus g^{\ominus \epsilon} = ((f \ominus g) \ominus g \dots) \ominus g \quad (5)$$

— ϵ 번 —

이다. 그리고 일반적으로 B가 크면 형태학적 덮개가 $f[n]$ 의 파형의 변화 특성을 잘 나타내지 못하게 되므로, B는 (x, y) 이산좌표상에서 3×3 크기의 점들로 구성된 집합의 부분집합을 갖도록 권장된다.^[5] 이때 식 (4)와 (5)는 구체적으로 다음과 같이 표현될 수 있다.

$$f \oplus g[n] = \max_{-1 \leq i \leq 1} \{ f[n+i] + g[i] \}, \quad \epsilon = 1$$

$$f \oplus g^{\oplus \epsilon}[n] = (f \oplus g^{\oplus (\epsilon-1)}) \oplus g[n], \quad \epsilon \geq 2. \quad (6)$$

$$f \ominus g[n] = \min_{-1 \leq i \leq 1} \{ f[n+i] - g[i] \}, \quad \epsilon = 1$$

$$f \ominus g^{\ominus \epsilon}[n] = (f \ominus g^{\ominus (\epsilon-1)}) \ominus g[n], \quad \epsilon \geq 2 \quad (7)$$

이들 상위 포락선과 하위 포락선을 이용하여 $C_B[\epsilon]$ 의 면적 $A[\epsilon]$ 은 다음과 같이 구해질 수 있다.

$$A[\epsilon] = \sum_{n=0}^N ((f \oplus g^{\oplus \epsilon}) - (f \ominus g^{\ominus \epsilon}))[n] \quad (8)$$

그리고 마지막으로 (8)에서 구한 $A[\epsilon]$ 를 식 (2)에 적용함으로써 D_M 을 구한다. 이상의 방법으로 프랙탈 차원 D_M 을 구하는 것을 형태학적 커버링 방법이라고 한다.

본 논문에서는 식 (6), (7)에서 $g[n]=0$ ($n=-1, 0, 1$)인 수평선분 B를 사용했는데, 이 경우 $f[n]$ 의 어떠한 이동이나 크기 변화에 대해서도 $f[n]$ 의 프랙탈 차원은 변하지 않는다는 장점을 가진다.^[5] 즉, 상수 a 및 b에 대해 $f'[n] = a \cdot f[n - n_0] + b$ 이라고 할

때 $D_M(f) = D_M(f')$ 이다.

이러한 장점으로 인해 본 논문에서 음성신호의 프랙탈 차원을 구하는데 있어서 형태학적 커닝 방법에 의한 Minkowski-Bouligand 차원을 사용하였다. 참고적으로 ϵ 이 10, 30일때 $g[n]=0$ ($n=-1, 0, 1$)을 사용하여 음성신호의 상위 포락선과 하위 포락선을 구한 예를 그림 1에 나타내었다. 일반적으로 프랙탈 차원을 구하는 방법으로 잘 알려진 Box Counting 차원은 위의 $f[n]$ 과 교차하는 사각형 수와 $f'[n]$ 과 교차하는 사각형 수가 서로 달라 이로 인하여 프랙탈 차원이 달라지는 문제점이 있어 음성신호의 프랙탈 차원을 구하는데 적합하지 않다.^{[3][5]}

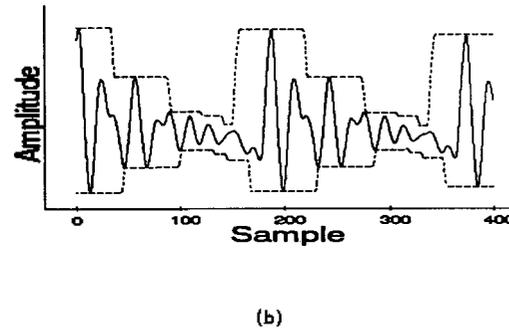
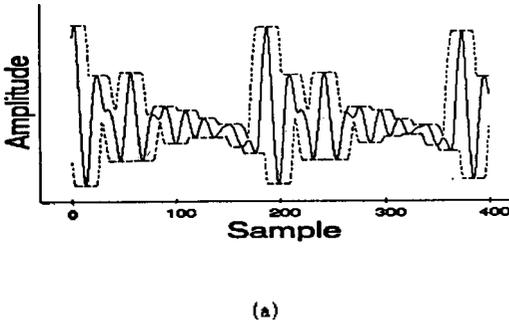


그림 1. 음성 신호 '아' (실선)에 대한 포락선(점선)의 예.
(a) $\epsilon = 10$, (b) $\epsilon = 30$.
Fig 1. Example of envelopes (dashed line) for the speech signal
(a) $\epsilon = 10$, (b) $\epsilon = 30$.

그리고 프랙탈 차원 D_M 을 구하는 과정에서 마지막으로 한가지 고려 해야될 사항이 있는데, 이는 위에서 (8)식을 (2)식에 적용할 때 ϵ 을 무한히 0으로 보낼 수 없다는 것이다. 그러므로 실제로는 어떤 임의의 ϵ

범위에서 $\log(A[\epsilon]/\epsilon^2)$ 대 $\log(1/\epsilon)$ 의 그래프를 그리고, 최소 자승법을 사용하여 그래프를 직선으로 맞추는 다음 그 직선의 기울기를 프랙탈 차원의 근사적인 값으로서 사용한다. 그런데 음성신호와 같은 실 세계 (real world)의 프랙탈은 수학적으로 생성된 Koch 곡선^[6]과 같은 이상적인 프랙탈 곡선과는 달리 모든 크기의 배율에 걸쳐 같은 크기의 프랙탈 차원을 가지지는 않는다. 즉 모든 크기의 배율에 걸쳐 동일하게 닮은 구조를 가지지 않는다는 것이다.^{[3][5]} 그러므로 어떤 크기의 배율에서 보느냐에 따라 프랙탈 차원의 측정치가 달라지는데 이것은 어떤 범위의 ϵ 값을 사용할 것인가 하는 문제를 발생시킨다.

ϵ 의 범위에 대해서 논의할때 두 가지 변수가 존재하는데 첫째는 어떤 ϵ 의 값에서 시작하는가 이며, 둘째는 ϵ 범위의 크기를 얼마로 잡는가 하는 것이다. 이 두가지 문제는 임의의 ϵ 범위를 $\{\epsilon_0, \epsilon_0 + 1, \epsilon_0 + 2, \dots, \epsilon_0 + m - 1\}$ 이라고 놓았을 때 어떤 ϵ_0 과 m 의 값을 선택할 것인가 하는 문제와 같다. ϵ_0 의 경우 ϵ_0 의 값이 크면 형태학적 덮개를 사용하는 과정에서 지나친 smoothing이 이루어져서 신호파형의 변화특성을 잘 살리지 못하는 경향이 있게 되므로, ϵ_0 의 값을 작게 하는 것이 좋다. 그리고 m 값에 대해서는 이전의 많은 연구자들이 경험적으로 결정해왔다. 본 논문에서

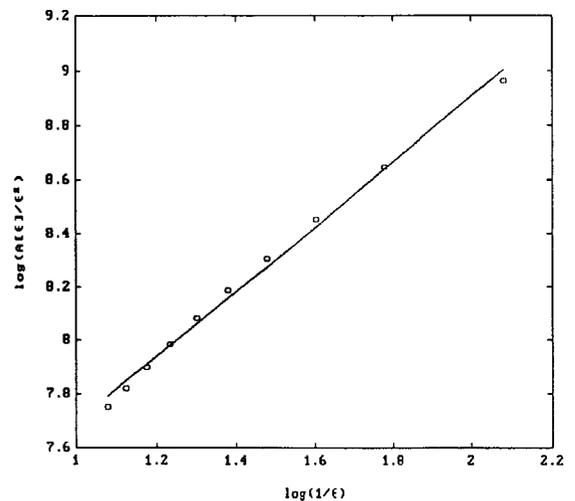


그림 2. 직선 맞춤에 의해 프랙탈 차원을 구하는 예 ($\epsilon_0 = 1$, $m = 10$ 인 경우).
Fig 2. Example of calculation of the fractal dimension by line fitting. (In case of $\epsilon_0 = 1$ and $m = 10$)

취할 때 프랙탈 차원의 값에서 1을 뺀 후 로그를 취하였다. 이는 프랙탈 차원이 낮은 부분에서의 차이를 더 강조하는 역할을 하며 비음(口, ㄴ)이 선행자음으로 올 때에 자음과 모음의 모습이 상대적으로 비슷하여 프랙탈 차원의 차이가 잘 안 나는 경우가 있어서 이 경우의 변별력을 키우기 위한 것이다. 이러한 경우의 예가 그림 3에 나타나 있다. 그림 3의 (a)에서 세로줄로 표현된 세 선들은 수작업에 의해 분리한 음성 구간과 자음과 모음의 경계위치를 나타낸다.

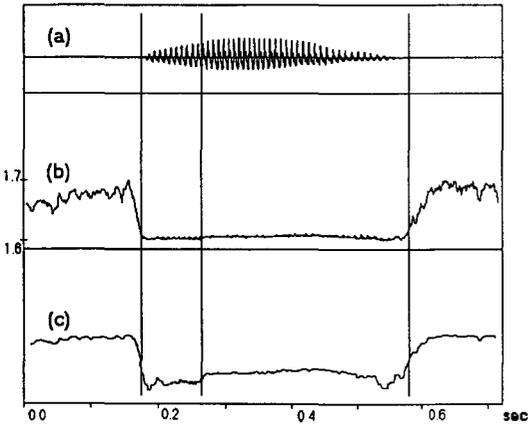


그림 3. 음성신호 '무'에 대한 프랙탈 차원.
 (a) 음성신호의 파형.
 (b) LFD의 곡선.
 (c) $\log(\text{LFD}-1)$ 의 곡선.

Fig 3. Fractal dimension of the speech signal 'mu'.
 (a) Waveform of speech signal.
 (b) Curve of LFD.
 (c) Curve of $\log(\text{LFD}-1)$.

(ii) 두 가지의 프랙탈 차원[LFD와 $\log(\text{LFD}-1)$] 곡선과 로그 에너지 곡선을 5샘플과 7샘플의 두 번의 메디안 필터(median filter)를 통과시켜 smoothing 한다.

(iii) 앞서 설명한 음성·목음 분리과정에 의해 결정된 음성의 시작부분의 위치를 N1이라 하고 끝부분을 N2라 하자. 그리고 N1에서 5ms내에는 모음이 존재 하지 않는다고 가정하여 $N1+5\text{ms}$ 인 위치를 M1이라고 놓는다. 다음으로 음성 구간에서 가장 높은 로그 에너지를 가지는 위치 M2를

찾는다. 위치 M2는 모음영역으로 볼 수 있기 때문에 자음과 모음의 경계는 M1과 M2 사이에 있을 것이다. 이때 preemphasis 하지 않은 음성 데이터를 사용한 이유가 여기 있는데 preemphasis 한 음성 데이터를 사용하여 M2를 구하면 preemphasis 과정의 고주파 성분 강조에 의해 M2가 자음구간에 있을 수 있기 때문이다.

(iv) 두 가지의 프랙탈 차원과 로그 에너지의 기울기를 구하여 M1과 M2 사이에서 기울기의 절대값 중 큰 순서대로 세 개를 찾아 그 위치를 자음과 모음의 경계 후보 위치로서 사용하였다. 이때 두 가지의 프랙탈 차원과 로그 에너지의 15샘플을 한 프레임으로 하여 각 프레임에서의 기울기를 선행회기분석방법으로 구하였다. 여기서 로그 에너지나 두 가지의 프랙탈 차원곡선을 $s[n]$ 이라고 할때 $s[n]$ 의 기울기 $D[n]$ 은 다음 식으로 주어진다.^[9]

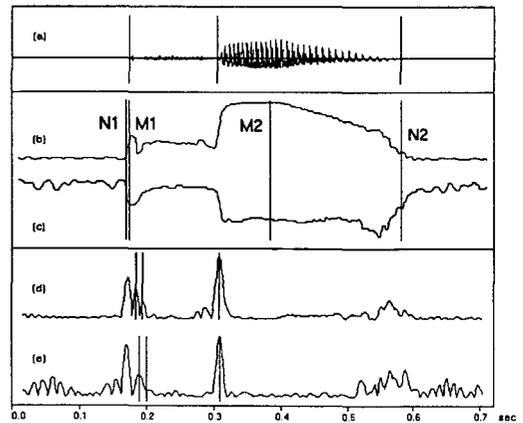


그림 4. 제안된 자·모음분리 알고리즘을 설명하기 위한 예.
 (a) 음성신호 '쿠'의 파형.
 (b) $\log(\text{Energy})$ 의 곡선.
 (c) $\log(\text{LFD}-1)$ 의 곡선.
 (d) $\log(\text{Energy})$ 의 기울기 곡선.
 (e) $\log(\text{LFD}-1)$ 의 기울기 곡선.

Fig 4. An example representing the proposed consonant/vowel segmentation algorithm.
 (a) Waveform of the speech signal 'ku'.
 (b) Curve of $\log(\text{Energy})$.
 (c) Curve of $\log(\text{LFD}-1)$.
 (d) Slope curve of $\log(\text{Energy})$.
 (e) Slope curve of $\log(\text{LFD}-1)$.

$$D[n] = \frac{\sum_{k=-7}^7 (k \cdot s[n+k])}{\sum_{k=-7}^7 k^2} \quad (9)$$

그림 4는 로그 에너지와 로그 프랙탈 차원을 사용한 경우에 대해 지금까지 설명한 알고리즘을 나타내고 있다. 그림 4의 (a)에서는 수작업에 의해 분리한 음성구간과 자음과 모음의 경계위치를 세 개의 수직선으로 나타내었으며 (b)와 (c)에서는 N1, N2, M1, M2의 위치를 보여준다. 그리고 (d)와 (e)에서는 M1과 M2사이의 세 개의 경계 후보 위치를 세 개의 수직선으로 표시했다. 그림에서 기울기는 절대값으로 표시하였으며 앞으로 나올 그림들에서도 기울기는 모두 절대값으로 표시하였다.

IV. 실험 및 결과의 검토

본 논문에서는 프랙탈 차원의 기울기를 사용한 경우, 로그 프랙탈 차원의 기울기를 사용한 경우, 로그 에너지 기울기를 사용한 경우, 그리고 로그 에너지 기울기와 로그 프랙탈 차원의 기울기 곱을 사용한 경우의 네 가지 경우에 대해서 자음과 모음의 분리 실험을 수행하였다. 세 사람의 음성 데이터를 모두 합하여 세 개의 자음과 모음의 경계 후보중에서 수작업에 의해 분리한 자음과 모음의 경계위치와 가장 가까운 후보를 선택한 경우의 분리 실험 결과를 그림 5에 나타내었다. 여기서 위의 네 가지 경우에 대한 분리 성능을 측정하는 기준으로 자동분리에 의한 자음·모음 경계와 수작업에 의한 경계 사이의 오차가 ±10ms 이내에 있는 음성 데이터의 갯수로서 정의하였는데 각각의 경우의 분리 성능을 각 후보에 대해 누계적으로 표 1에 나타내었다. 이때 후보 1, 2, 3은 기울기가 큰 순서이며, 각 후보에 대한 분리 성능은 그 후보를 사용하였을 때 다른 두 후보보다 수작업에 의해 분리한 경계 위치와 가장 가깝게 된 경우의 음성 데이터 갯수를 432개의 전체 데이터 갯수에 대한 퍼센트로서 나타내었다. 표에서 보는 바와 같이 로그 에너지 기울기를 사용한 경우 보다는 프랙탈 차원 기울기를 사용한 경우가, 프랙탈 차원 기울기를 사용한 경우 보다는 로그 프랙탈 차원의 기울기를 사용한 경우가 보다 우수한 분리 성능을 나타내었고, 특히 로그 에너지 기울기와 로그 프랙탈 차원 기울기의 곱을 사용한

경우에는 앞의 모든 경우에 비해 가장 좋은 분리 성능을 나타냈다. 예를 들어 세 개의 후보를 모두 사용하였을 때, 로그 에너지의 기울기를 사용한 경우는 88.0%, 프랙탈 차원의 기울기를 사용한 경우는 91.5%, 로그 프랙탈 차원의 기울기를 사용한 경우는 93.6%로 나타났으며 로그 에너지 기울기와 로그 프랙탈 차원의 기울기의 곱을 사용한 경우는 96.1%로서 가장 좋은 분리 성능을 나타내었다. 그리고 각 사람에 대해서도 두 곡선의 기울기의 곱을 사용한 경우 가장 좋은 분리 성능을 나타내는가 확인하기 위해 각 사람에 대해서 앞의 네 가지 경우에 대한 분리 실험을 하였다. 그 결과 역시 두 기울기의 곱을 사용한 결과가

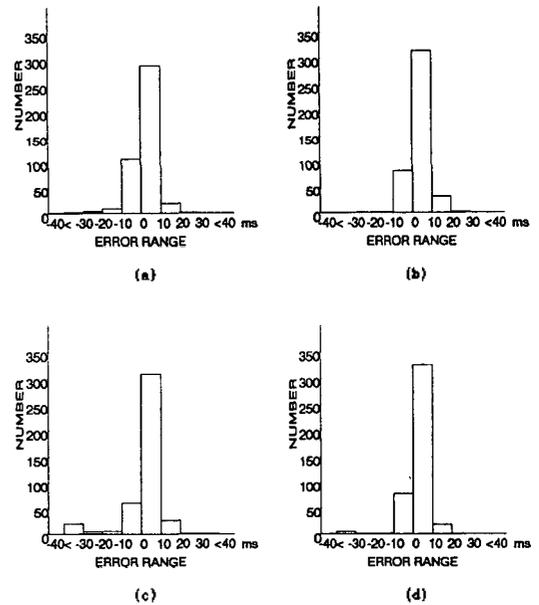


그림 5. 세 사람의 432개의 음성 데이터에 대한 자음과 모음의 분리 결과.

- (a) LFD의 기울기만을 사용한 경우.
- (b) $\log(\text{LFD}-1)$ 의 기울기만을 사용한 경우.
- (c) $\log(\text{Energy})$ 의 기울기만을 사용한 경우
- (d) $\log(\text{LFD}-1)$ 와 $\log(\text{Energy})$ 의 기울기의 곱을 사용한 경우.

Fig 5. Consonant/vowel segmentation results based on 432 speech data from three speakers.

- (a) In case of using the LFD slope alone.
- (b) In case of using the $\log(\text{LFD}-1)$ slope alone.
- (c) In case of using the $\log(\text{Energy})$ slope alone.
- (d) In case of using the multiplication of $\log(\text{Energy})$ slope and $\log(\text{LFD}-1)$ slope.

표 1. 제안된 방법에 의한 자·모음 분리 성능($\pm 10\text{ms}$ 이내의 오차).Table 1. Consonant/vowel segmentation performance based on the proposed method (considering errors within $\pm 10\text{ms}$).

분리 파라메타		사 람	HHI	PCK	RJH	전 제
LFD 기울기	후보 1		75.7%	66.7%	60.4%	67.6%
	후보 1, 2		90.3%	80.6%	73.6%	81.5%
	후보 1, 2, 3		93.8%	88.9%	91.7%	91.5%
log(LFD-1) 기울기	후보 1		84.0%	69.4%	74.3%	76.0%
	후보 1, 2		93.7%	82.6%	86.1%	87.6%
	후보 1, 2, 3		96.5%	91.6%	92.4%	93.6%
log(E[n]) 기울기	후보 1		86.1%	78.5%	61.1%	75.2%
	후보 1, 2		90.3%	86.1%	72.9%	83.1%
	후보 1, 2, 3		94.5%	89.6%	79.9%	88.0%
log(LFD-1) 기울기 × log(E[n]) 기울기	후보 1		86.8%	73.6%	74.3%	78.2%
	후보 1, 2		94.4%	88.2%	88.9%	90.5%
	후보 1, 2, 3		97.9%	95.8%	94.4%	96.1%

가장 좋은 분리 결과를 보였으며 그 결과를 그림 6에 나타내었다.

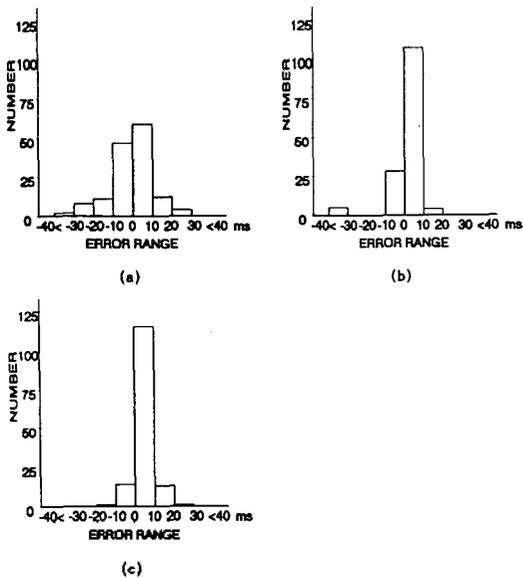


그림 6. 로그 에너지 기울기와 로그 프랙탈 차원 기울기의 곱을 사용한 세 사람 각각에 대한 분리 결과. (a) 화자 HHI. (b) 화자 PCK. (c) 화자 RJH.

Fig 6. Segmentation results for each of three speakers using the multiplication of log energy slope and slope of log fractal dimension.

(a) Speaker HHI. (b) Speaker PCK.
(c) Speaker RJH.

상기 결과로부터 에너지와 프랙탈 차원이 상호 보완적인 역할을 하여 보다 우수한 분리 결과를 얻게 되는 것을 볼 수 있었다. 이 사실을 실제 음성분리의 예를 들어 살펴 보도록 한다. 그림 7은 로그 에너지의 기울기와 로그 프랙탈 차원의 기울기를 각각 사용한 경우로 로그 프랙탈 차원의 기울기를 사용한 경우가 더 좋은 분리 결과를 보이는 예를 보여준다. 이 경우 로그 에너지의 기울기로는 자음과 모음의 경계를 거의 찾을 수 없으며, 이러한 경우는 선행자음이 비음(L, R)일 때 많이 나타난다. 그림 8은 반대인 예로서 로그 프랙탈 차원이 경계 근처에서 긴 시간에 걸쳐 완만히 변하여 세 개의 후보로서는 경계를 찾지 못한 경우이다. 여기서 네 번째 후보를 사용하면 경계를 찾을 수 있었지만 이와 같은 경우는 매우 드물어서 후보를 더 사용하는 것은 효율적이지 못한 것으로 판단되었다. 그림 9와 10에서는 그림 7과 8의 음성 데이터에 대해 로그 에너지 기울기와 로그 프랙탈 차원의 기울기의 곱을 사용한 경우를 보이고 있으며 모두 바른 분리 결과를 보이고 있다. 물론 로그 에너지 기울기와 로그 프랙탈 차원의 기울기의 곱을 사용함으로써 결과가 더 나빠지는 경우도 없는 것은 아니며, 이러한 경우를 그림 11과 12에 나타내었다. 그림 11에서와 같이 로그 프랙탈 차원의 기울기만을 사용하면 바른 분리를 할 수 있으나, 그림 12에서 두 기울기의 곱을 사용한 경우 오히려 다른 부분에서의 기울기가 강조되어 잘못된 분리 결과를 초래했다. 이러한 문제는 앞으로 알고리즘의 개선과정에서 보완되어야 할

것이다. 그러나, 그림 7, 8 및 12 등에 나타난 결과들은 일부의 예외에 속한 것으로서, 그림 5 및 표 1에서

보는 바와 같이 대부분의 경우는 바른 분리 결과를 나타내었다.

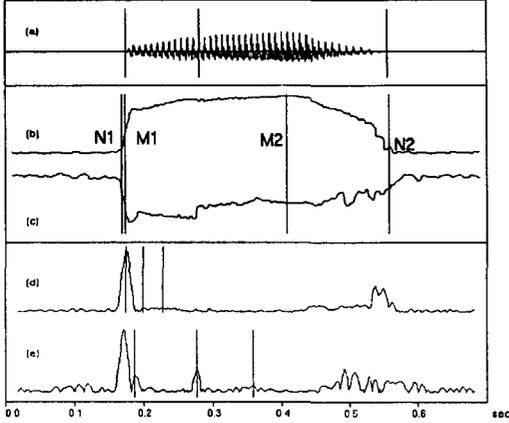


그림 7. 로그 에너지 기울기만을 사용하는 경우의 잘못된 분리된 예.

- (a) 음성신호 '네'의 파형.
- (b) $\log(\text{Energy})$ 의 곡선.
- (c) $\log(\text{LFD}-1)$ 의 곡선.
- (d) $\log(\text{Energy})$ 의 기울기 곡선.
- (e) $\log(\text{LFD}-1)$ 의 기울기 곡선.

Fig 7. Example of wrong segmentation in case of using only the log energy slope.

- (a) Waveform of the speech signal 'ne'.
- (b) Curve of $\log(\text{Energy})$.
- (c) Curve of $\log(\text{LFD}-1)$.
- (d) Slope curve of $\log(\text{Energy})$.
- (e) Slope curve of $\log(\text{LFD}-1)$.

- (a) 음성신호 '시'의 파형.
- (b) $\log(\text{Energy})$ 의 곡선.
- (c) $\log(\text{LFD}-1)$ 의 곡선.
- (d) $\log(\text{Energy})$ 와 기울기 곡선.
- (e) $\log(\text{LFD}-1)$ 의 기울기 곡선.

Fig 8. Example of wrong segmentation in case of using only the slope of log fractal dimension.

- (a) Waveform of the speech signal 'si'.
- (b) Curve of $\log(\text{Energy})$.
- (c) Curve of $\log(\text{LFD}-1)$.
- (d) Slope curve of $\log(\text{Energy})$.
- (e) Slope curve of $\log(\text{LFD}-1)$.

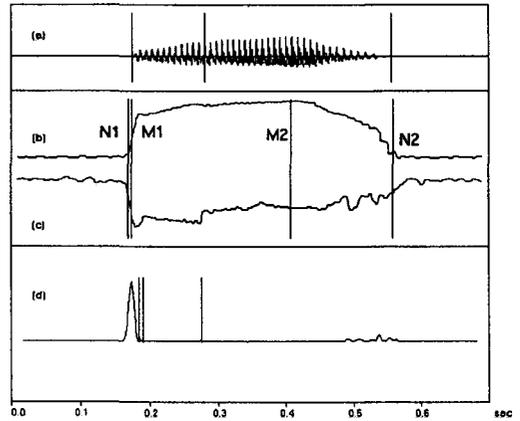


그림 9. 그림 7의 경우에 로그 에너지 기울기와 로그 프랙탈 차원 기울기의 곱을 사용한 예.

- (a) 음성신호 '네'의 파형.
- (b) $\log(\text{Energy})$ 의 곡선.
- (c) $\log(\text{LFD}-1)$ 의 곡선.
- (d) $\log(\text{Energy})$ 기울기와 $\log(\text{LFD}-1)$ 기울기의 곱의 곡선.

Fig 9. Example using the multiplication of log energy slope and slope of log fractal dimension in case of Fig 7.

- (a) Waveform of the speech signal 'ne'.
- (b) Curve of $\log(\text{Energy})$.
- (c) Curve of $\log(\text{LFD}-1)$.
- (d) Curve of multiplication of $\log(\text{Energy})$ slope and $\log(\text{LFD}-1)$ slope.

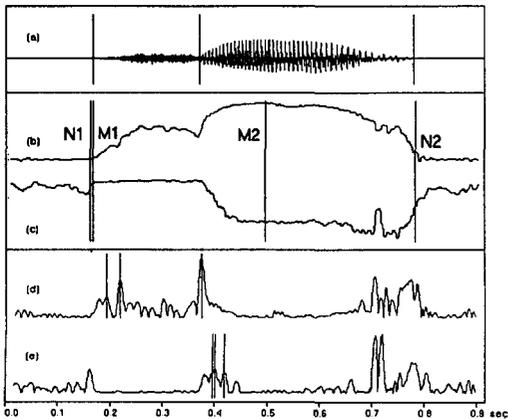


그림 8. 로그 프랙탈 차원 기울기만을 사용한 경우의 잘못된 분리된 예.

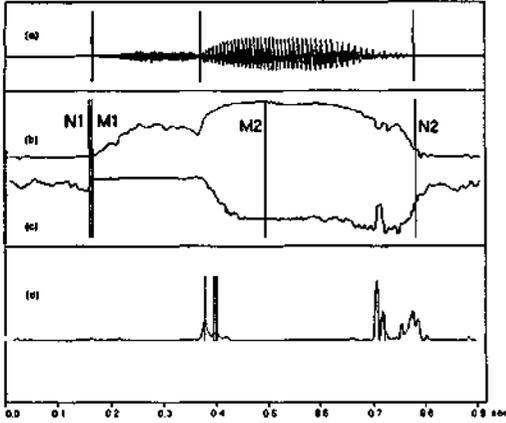


그림 10. 그림 8의 경우에 로그 에너지 기울기와 로그 프랙탈 차원 기울기의 곱을 사용한 예.

- (a) 음성신호 '시'의 파형.
- (b) $\log(\text{Energy})$ 의 곡선.
- (c) $\log(\text{LFD}-1)$ 의 곡선.
- (d) $\log(\text{Energy})$ 기울기와 $\log(\text{LFD}-1)$ 기울기의 곱의 곡선.

Fig 10. Example using the multiplication of log energy slope and slope of log fractal dimension in case of Fig 8.

- (a) Waveform of the speech signal 'si'.
- (b) Curve of $\log(\text{Energy})$.
- (c) Curve of $\log(\text{LFD}-1)$.
- (d) Curve of multiplication of $\log(\text{Energy})$ slope and $\log(\text{LFD}-1)$ slope.

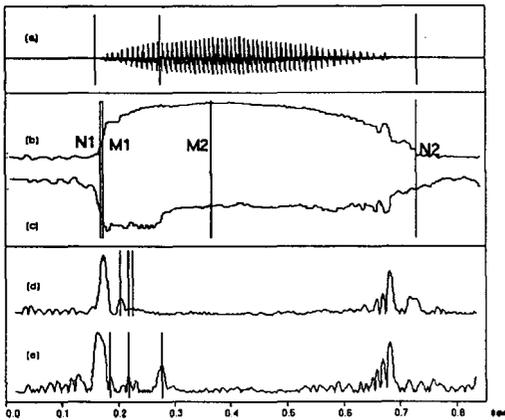


그림 11. 로그 에너지 기울기와 로그 프랙탈 차원 기울기를 각각 사용한 예.

- (a) 음성신호 '매'의 파형.
- (b) $\log(\text{Energy})$ 의 곡선.

- (c) $\log(\text{LFD}-1)$ 의 곡선.
- (d) $\log(\text{Energy})$ 의 기울기 곡선.
- (e) $\log(\text{LFD}-1)$ 의 기울기 곡선.

Fig 11. Example using log energy slope and slope of log fractal dimension, respectively.

- (a) Waveform of speech signal 'mae'.
- (b) Curve of $\log(\text{Energy})$.
- (c) Curve of $\log(\text{LFD}-1)$.
- (d) Slope curve of $\log(\text{Energy})$.
- (e) Slope curve of $\log(\text{LFD}-1)$.

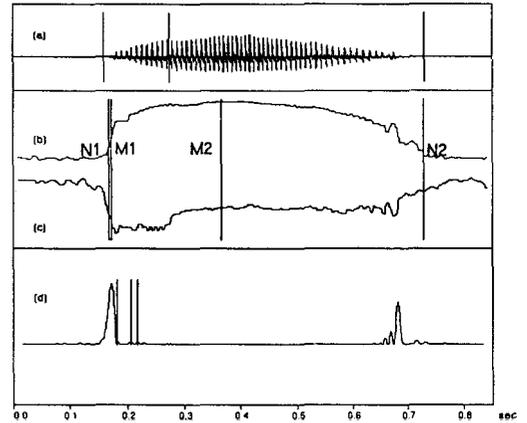


그림 12. 그림 11의 경우에 두 기울기 곱을 사용하여 잘못 분리된 예.

- (a) 음성신호 '매'의 파형.
- (b) $\log(\text{Energy})$ 의 곡선.
- (c) $\log(\text{LFD}-1)$ 의 곡선.
- (d) $\log(\text{Energy})$ 기울기와 $\log(\text{LFD}-1)$ 기울기의 곱의 곡선.

Fig 12. Example of wrong segmentation using the multiplication of two slope in case of Fig 11.

- (a) Waveform of the speech signal 'mae'.
- (b) Curve of $\log(\text{Energy})$.
- (c) Curve of $\log(\text{LFD}-1)$.
- (d) Curve of multiplication of $\log(\text{Energy})$ slope and $\log(\text{LFD}-1)$ slope.

V. 결 론

본 논문에서는 형태학적 커버링 방법으로 구한 Minkowski-bouligand 차원을 프랙탈 차원으로 사용하여 음성 신호의 분리 실험을 수행하였다. 이를 위하여 한국어 CV 단음절 음성신호의 프랙탈 차원과 에너지

