음성파형의 비대칭율을 이용한 음소의 전이구간 검출

On Detecting the Transition Regions of Phonemes by Using the Asymmetrical Rate of Speech Waveforms

배 명 진* 이 을 재* 안 수 길**

(Myung Jin Bae, Eul Jae Lee, Souguil Ann)

요 약

연속음 인식을 위해서는 음성신호의 음성학적 경계를 검출하는 분할과정이 필요하며, 본 논문에서는 음성검출의 일환으로 음소간을 결정하기 위한 파라미터로 한 프레임 내의 비대칭율을 제안하였다. 제안된 비대칭율은 한 프레임에서 음성신호의 변화율을 대별하며, 인접 프레임의 비대칭율과 비교하면 현재의 프레임이 정상상태 혹은 전이영역에 있는지를 판별할 수 있게 해 준다.

ABSTRACT

To recognize continued speech, it is necessary to segment the connected acoustic signal into phonetic units. In this paper, as a parameter to detect transition regions in continued speech, we propose a new asymmetrical rate. The suggested rate represents a change rate of magnitude of speech signals. As comparing this rate with other rate in adjacent frame, the state of the frame can be distinguished between steady state and transient state.

I. 서 론

[illegible]

고유 음성신호는 음성학적인 부류의 나열, 음절, 음소 등의 다외로 분할하여야 한다. 연속음을 이러한 단위

[illegible] 부 무슨하면 분석적인는 분석의 한계를 초월 [illegible] 있고, 음성의 [illegible] 있다.

언어학적 음절의 설의는 "두개의 sonority 최소 [illegible] 최대값을 갖는 일련의 음성적요소"[1] 내지 수 있다. 여기서 sonority는 "청각의 라에 이는

*오지대학교 전자공학과
**서울대학교 제어계측공학과

에너지의 양으로 각 음성의 분더 소리 바탕에 의해 결정되며 소리의 길이, 높이, 세기에서 오는 크기와는 무관하다"라고 정의되어 있다[7,8]. 즉, 음절의 경계는 신호의 크기에 따라 잘 정의된다. 그러데 음운의 특성에 따른 음성신호의 크기와 강세에 따른 크기 변화를 자동적으로 구분해 내기는 거의 불가능하다.

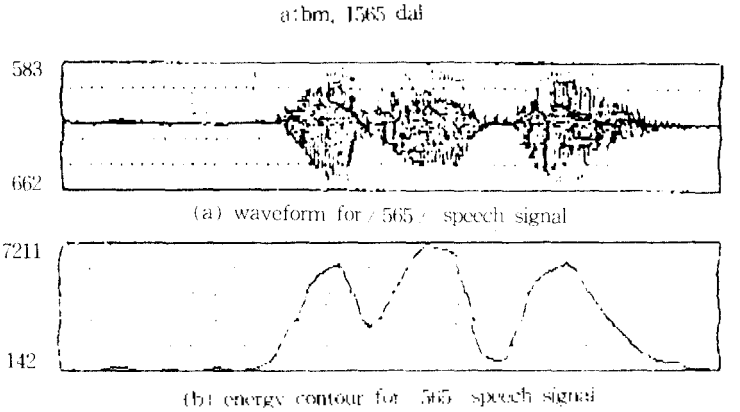
음성신호의 분할에 관한 연구는 Zue[1] 연구 이후부터 활발히 전개되어 왔다. Zue의 방법은 음성을 먼저 일차적인 분할을 행한 후 여러가지 특정 파라미터를 조합한 복잡한 판정과정을 통하여 일차적 분할결과를 수정하며, 동시에 음소의 분류를 수행한다. 여기에서 사용한 특징 파래미터들은 선형 예측 계수, 예측오차, 스펙트럼 에너지, 피치, 포만트정보 등이 적용되었다[1].

음성신호의 전이구간을 검출하는 연구는 특징파래미터를 추출한 영역에 따라 크게 시간영역법, 스펙트럼영역법, 혼성영역법으로 나눌 수 있다. 시간영역법은 시간영역에서 계산의 간편성을 취할 수 있으며, VOT(voice onset time)의 연속성이나 진폭의 증감을 이용하는 방법들이 제안되어 졌다[4,5,9]. 스펙트럼영역법은 음성신호의 음소의 변화에 따른 포먼트의 전이특성이나 주파수성분별 에너지비를 이용하는 등이 제안되어져 있다[2,4]. 또한 혼성영역법은 변환영역에서의 특징파래미터들을 이용하는 것으로[6,10], LPC 계수의 전이특성, LPC에러의 변화특성 등을 이용하고 있다.

시간영역법에서 파래미터의 검출은 비교적 쉬우나 그 변화정도를 정확히 파악하기 위한 결정논리가 상대적으로 어렵다[13]. 반면 스펙트럼영역법이나 혼성법은 비교적 정확하지만 계산의 정밀도나 변환차수의 영향을 받게 되고, 전처리과정으로 보거에는 계산량의 부담이 시간영역법에 비해 큰 편이다. 따라서 우리는 시간영역법에서 음소의 전이구간 검출용 파래미터가 갖는 결정논리의 복삽성과 부정확성에 대해 알아보고 이러한 문제점을 제거할 수 있는 새로운 파래미터를 제안하고자 한다.

Ⅱ. 음성신호의 전이구간

음성신호는 생성모델에 따라 유성음, 무성음, 혼합음, 묵음으로 구분지을 수 있다. 유성음은 준주기성과 성도의 공명으로 큰 에너지를 가지며, 무성음은 순색잡음의 낮은 에너지를 갖게 된다. 혼합음은 무성음에서 유성음으로 또는 유성음에서 무성음으로 연결되는 혼합영역이며 유·무성음의 성질이 동시에 나타나게 된다. 연속음이나 연결음에서는 이 음들이 시간에 따라 변화하게 되며, 이것은 프레임당 평균진폭의 변화형태로 그림1과 같이 나타나게 된다. 그림 1은 '오유오'라는 연결단어를 24세의 남성화자가 발성한 것이며 평균진폭의 변화도 (contour)가 음소나 음절의 변화를 잘 나타내고 있음을 알 수 있다



(a) waveform for '565' speech signal

(b) energy contour for '565' speech signal

그림 1. 연결음 '오육오'에 대한 평균진폭의 변화도
Absolute magnitude contour for connected speech 'oryukoh'

평균진폭의 변화도를 이용하여 음절의 전이구간을 분류하려고 하면 우선 평균진폭을 계산해야 하며, 이 때 그 프레임에 적용된 윈도우의 영향을 받게 된다. 윈도우의 영향으로는 윈도우의 길이와 형태에 따른 영향을 고려할 수 있다. 윈도우내의 음성 성분 주파수가 윈도우 길이의 역수에 정배수일 때가 윈도우 길이의 영향을 가장 적게 받게 된다. 그러므로 이 경우에는 윈도우의 길이가 피치에 정수배로 선택되는 것이 가장 바람직하게 된다. 그렇지만 사전에 피치를 정확히 구해야 하고 또한 윈도우의 길이가 가변적이어야 하기 때문에 윈도우 길이를 피치에 일치시키지 않고 윈도우길이 영향을 최소화시키려는 연구도 많이 제안되고 있다[9,10].

윈도우의 길이를 음성의 피치에 정수배로 하이도 정수배가 아닌 음성성분들이 또한 일부 존재하기 때문에 윈도우형태를 잘 선정할 필요가 있다. 윈도우의 형은 통과 및 차단대역의 비에 따라 방형, 삼각형, 해밍, 블랙맨 등이 있으며, 차단특성이 우수한 윈도우 함수일수록 계산과정이 복잡하게 된다[11].

윈도우의 길이에 따른 평균진폭값은 그림 2와 같이 크게 달라진다. 윈도우의 길이가 음소의 변화특성에 비해 길게되면 스므딩효과에 의해 평균진폭은 음소의 변화특성을 잘 나타낼 수 없게 된다. 반면 음소의 변화특성에 비해 윈도우의 길이를 너무 짧게 하면 평균진폭의 변화도에는 국부봉우리가 많이 나타나서 평균진폭의 변화도는 이용한 음소의 변화를 구하기 어렵게 된다.

연속음에 대해 윈도우를 잘 선정하여 평균진폭을 구하여도 그 변화특성을 수치적으로 잘 나타내는 결정논리가 필요하다. 결정논리 적용시에는 크게 두 가지의 어려움이 따른다. 첫째로는 아무리 윈도우를 잘 적용하여도 윈도우내의 음성성분이 복잡 다양하여 평균진폭의 변화도에 국부봉우리들이 나타나게 된다. 이러한 국부봉우리와 실제 음소의 전이를 나타내는 음소봉우리를 분리해야만 하는 어려움이 있다. 두번째로는 음소봉우리의 형태가 여러가지라는 점이다. 예를 들어 유성음에 이은 비음연결, 무성음과 비음 또는 유성음의 연결 등에서는 뚜렷한 봉우리

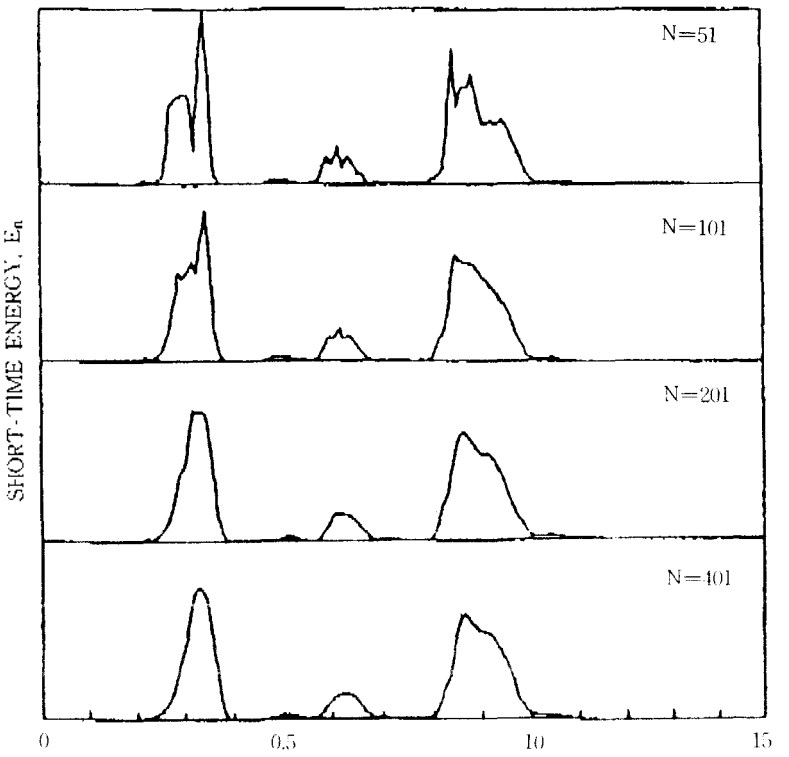


그림 2. 방형윈도우의 길이에 따른 평균진폭의 변화도
Average magnitude contours according to the length of the rectangular window.

의 형태를 찾기 힘들다.

Ⅲ. 한 프레임구간에서 파형의 비대칭율

음성신호의 특징들은 음성파형에 비해 서서히 변화하기 때문에 프레임 단위로 분석하는 것이 보통이다. 현재의 프레임이 전이구간이나 정상상태구간에 속하는지를 판정하는 방법은 현재의 프레임을 반분하였을 때 평균진폭의 비를 측정하여 판정할 수도 있다. 그 평균진폭의 비 MR(fr)은 음성신호를 s(n)이라 할 때 다음과 같이 나타낼 수 있다.

$$MR(fr)=\frac{\sum_{k=N/2}^{N-1} |s(n-k)|}{\sum_{k=0}^{N/2-1} |s(n-k)|} \quad \cdots\cdots\cdots(1)$$

여기서 변수 n은 이 프레임이 시작되는 첫 시컨스의 위치이며, N은 프레임의 길이를 나타낸다. 이 평균진폭비는 프레임길이를 1/2로 했을 때의 인근 프레임의 평균진폭비를 나타내기 때문에 제2장에서 인접한 윈도우의 영향을 역시 받게 된다.

윈도우의 영향에 무관하게 현재의 프레임이 어떤 상태에 존재하는지를 측정하는 새로운 방법으로는 다음과 같이 비대칭율 (asymmetrical rate, ASR)을 정의해서 사용할 수 있다.

$$ASR(n)=\frac{\sum_{k=1}^{P/2} |s(n-k)-s(n+k)|}{\sum_{k=P/2}^{P/2} |s(n-k)|} \quad \cdots\cdots\cdots(2)$$

여기서 n은 비대칭율을 측정하려는 중심샘플의 위치이며, P는 비대칭율을 구하려는 구간이다. 예를 들어 한 프레임의 샘플수를 N=256으로, 비대칭율을 측정하는 구간을 P=200으로 할 때, 비대칭율을 측정할 수 있는 범위 n은 n=100에서 n=155까지가 된다.

식(2)로 정의한 ASR값은 n-샘플 위치를 중심으로 좌 우파형의 대칭성을 나타내는 규준화된 비대칭율값이 된다. 이 비대칭율값이 영에 근접하면 이 샘플의 위치 n을 중심으로 좌 우파형의 대칭이 이루어졌음을 나타낸다. 따라서 주어진 프레임내에서 최소의 비대칭율을 구했을 때 이 왜율값이 영에 근접하면 이 프레임은 정상상태에 있다고 볼 수 있다. 반면, 현재 프레임내에서 최소의 비대칭율을 구했을 때, 그 왜율값이 1에 근접하면 이 프레임은 전이구간에 놓여있게 된다.

Ⅳ. 비대칭율에 의한 전이구간의 분류

23세의 여성화자가 말음한 고립단어 /삼/의 음성신호에 대해 비대칭율을 구한 것을 그림 3에 나타내었다. 여기서 그림3(a)는 음성파형을 나타내며, 그림 3(b)에는 평균진폭의 변화도를 나타내었다. 평균진폭은 각 프레임을 256샘플 단위로 하고 128 프레임씩 겹쳐서 하여 구한 것이다. 이 때 최소의 비대칭율 변화도를 그림 3(c)에 나타내었다. 이 변화도는 각 프레임내에서 최소의 비대칭율을 다음과 같이 구한 것이다.

$$ASR(fr)=MIN\{ASR(100),\ ASR(101)\cdots,\ ASR(155)\} \quad \cdots\cdots\cdots(3)$$

여기서 MIN{ }는 함수로 주어진 변수영역에서 최소값을 선택하는 함수이고, fr은 현재 프레임의 위치를 나타낸다.

그림 3(c)에 제시한 최소 비대칭율의 변화도를 살펴보면, 음소가 시작되는 영역에서는 비대칭율이 상대적으로 크다는 것을 알 수 있다. 또한 비대칭율의 변화도의 기울기를 비교할 때 음소의 전이구간 프레임 구간에서의 비대칭율값은 안정 프레임의 왜율값에 비해 훨씬 큰 기울기를 가짐을 알 수 있다. 따라서 왜율의 변화도의 기울기를 이용하면

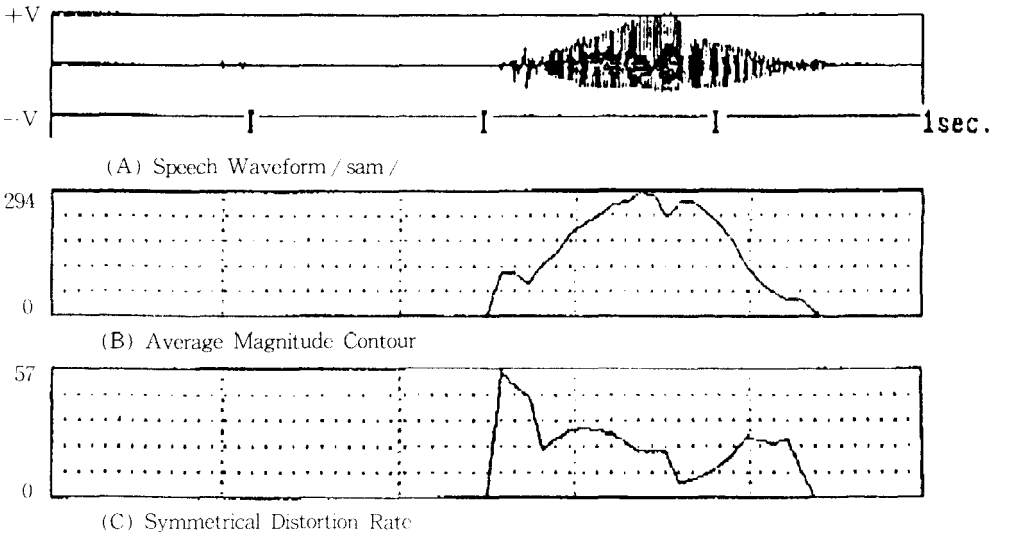


그림 3. 음성신호 / 삼 / 에 대한 최소 비대칭율의 변화도.
Minimum asymmetrical rate for speech signal 'sam'.

에서는 정상상태에 있음을 짐작할 수 있다.

또한 파형의 구조가 단순한 비음 / ㅁ / 에 비해 파형의 구조가 복잡한 / 아 / 음의 프레임에서 비대칭율은 상대적으로 높아지는 특징이 있다. 그리고 음소의 변화가 빠른 자음인 것이나 끝나는 프레임구간에서는 모음구간에 비해 비대칭율의 봉우리를 구성하는 상사가 높다라는 것을 알 수 있다.

이상을 정리해 보면 음소의 전이구간을 선택하기 위한 실험논리를 만들 수 있다. 비대칭율이 음성신호파형의 전반적인 변화도를 나타내기 때문에 각 프레임 마다 비대칭율을 구하여 그 변화도가 심한 곳을 이루면 자음의 프레임이 전이구간이 된다. 반면 곡률을 이루지 못 하는 곳은 정상상태에 있게 된다.

V. 실험 및 결과

[illegible] digital 변환기를 인터페이스 시켰다. 화자는 남성 3명과 여성화자 1명을 녹음 대상으로 선정하여 발성케하고 8KHz로 표본화하여 다음과 같이 처리하였다.

발성 1) 24세 남성화자 :
"인수네 고추는 천재소년을 좋아한다."
발성 2) 28세 남자화자 :
"서울대 전자공학과 음성신호처리 연구팀이다."
발성 3) 25세 여성화자 :
"감사합니다."

각 음성자료에 대해 한 프레임의 길이를 256샘플(= 32msec)로 하여 128샘플 단위로 겹치게 처리하였다. 각 프레임에 대해 비대칭율을 계산한 구간은 200샘플(=25 msec.) 단위로 하였고, 100샘플 뒤쪽에서 출발하여 155샘플 위치까지 56개의 비대칭율을 측정하였다. 이것이 최소의 비대칭율을 구하여 이 프레임의 대표적인 비대칭율로 사용하였다.

비대칭율을 구하는 범위는 유성음의 심한 가능한 25msec 피치범위의 1/2구간까지 적용해야만 하지 [illegible] 12.5msec 이내에 존재하며, 일반적으로 성인 남자 역시 이 범위에 대해 짧은 피치범위에 존재하고 있다. 따라서 계산의 편의를 위해 비대칭율은 56샘플 (=7 msec)까지만 구하여 적용하였다.

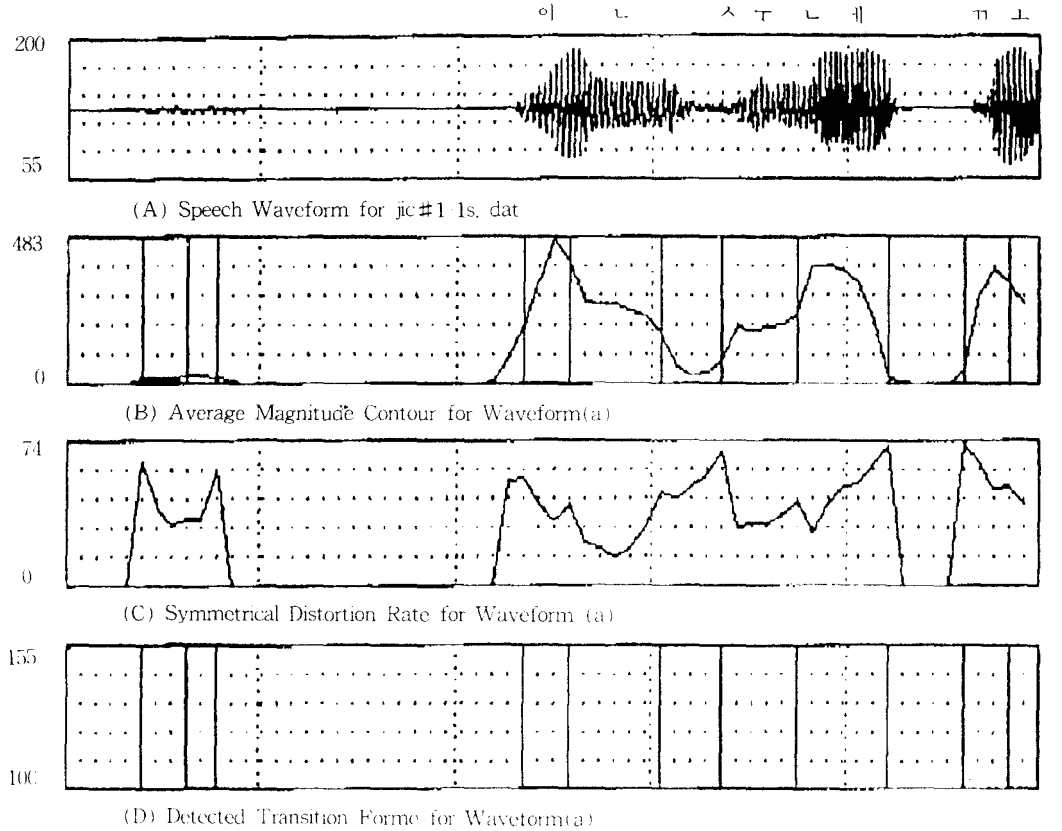
그림 4는 발성 1)에 대한 처리결과를, 그림 5에는 발성 2)에 대한 처리결과를, 그리고 그림 6)은 발성 3)에 대해 나타내었다. 각 그림에서 음성시료의 파형을 그림 (a)에, 그에 따른 평균진폭의 변화도를 그림 (b)에 나타내었으며 이것을 통해 음절분할의 개략적인 평가기준으로 삼을 수 있다. 100샘플에서 부터 155샘플까지 200샘플(≒25 msec) 단위로 구한 비대칭율을 계산하고, 이들 중에서 최소치를 그 프레임의 대표적인 비대칭율 값으로 하여 그림(c)에 나타내었다.

또한 그림(c)의 최소 비대칭율의 변화도에서 봉우리구간을 음소의 전이구간으로 결정하여 그림 (d)에 나타내었다. 그림(d)와 평균진폭의 변화도에서 변화특성을 비교하기 쉽도록 그림 (b)에도 찾은 전이구간을 동시에 종선으로 나타내었다. 여기서 연속음성의 변화특성은 비대칭율의 변화도로 잘 대별할 수 있고 앞에서 고찰한 봉우리와 골의 특성이 음소의 전이상태와 정상상태를 잘 나타내고 있음을 알 수 있다.

특히 그림 4의 두번째 초반과 중반그림이나, 그림 5의 두번째 초반그림 등에서는 비음구간인데 평균진폭의 변화도로는 구별하기 힘든 음소의 전이구간도 비음화된 한 블럭구간으로 잘 나타내고 있다. 그림 4의 세번째 초반그림은 유성음들끼리 별 변화없이 연속된 음성인 경우인데 비대칭율의 변화도는 이를 정확히 분류해주고 있다.

VI. 결 론

연속음 인식을 위해서는 음성신호의 분할과정이 필요하다. 음절단위의 분할이 잘 이루어지면 음성분석시나 인식시에 고립단어의 분석과 인식에 적용했던 많은 기법들을 쉽게 적용할 수 있게 된다. 지금까지 전이구간 검출법들이 많이 제안되어 왔지만 평균진폭의 변화도에서 전이구간을 검출하는 것이 쉽고 우수한 편이다. 그렇지만 적용과정에서 윈도우의



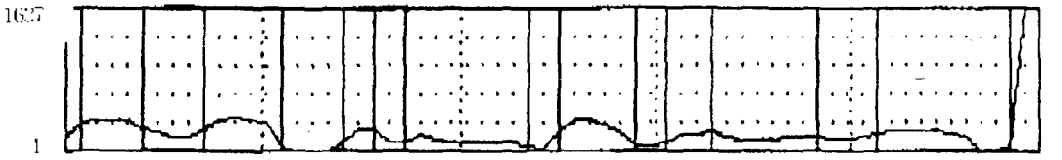
(A) Speech Waveform for jic#1 1s. dat

(B) Average Magnitude Contour for Waveform(a)

(C) Symmetrical Distortion Rate for Waveform (a)

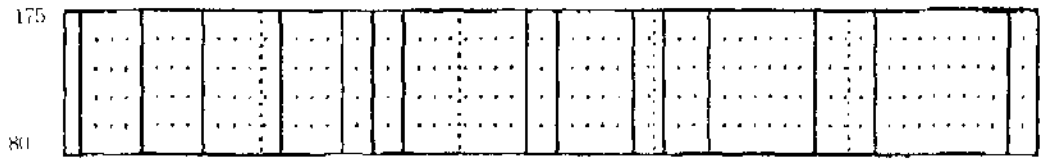
(D) Detected Transition Forme for Waveform(a)

(A) Speech Waveform for jic#1 2s. dat



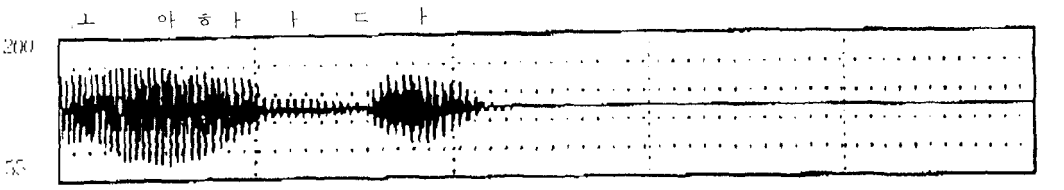
(B) Average Magnitude Contour for Waveform(a)

(C) Symmetrical Distortion Rate for Waveform (a)

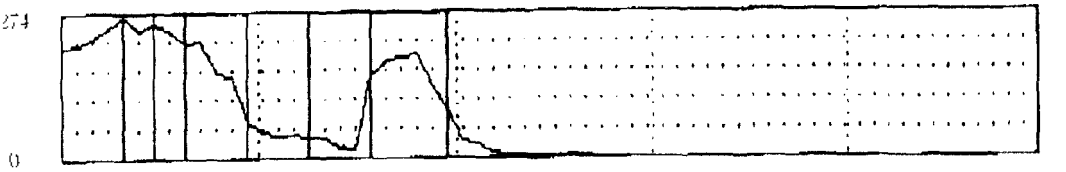


(D) Detected Transition Forme for Waveform(a)

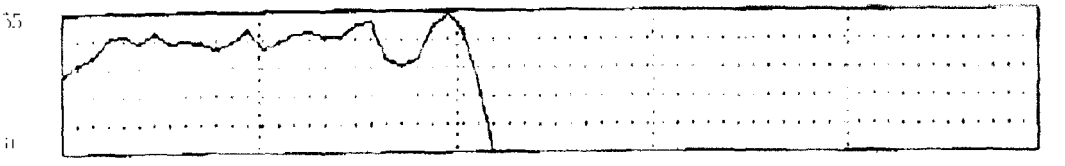
jic#1 3s. dat.



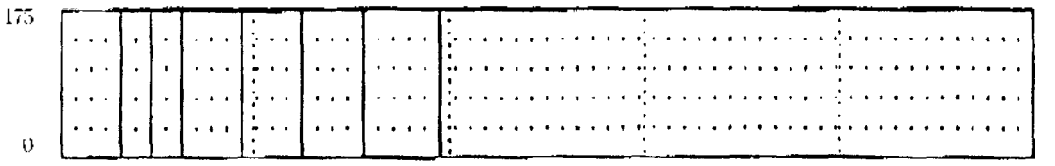
(A) Speech Waveform for jic#1-3s. dat



(B) Average Magnitude Contour for Waveform(a)



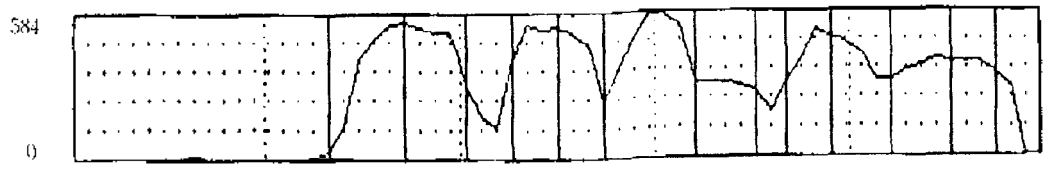
(C) Symmetrical Distortion Rate for Waveform (a)



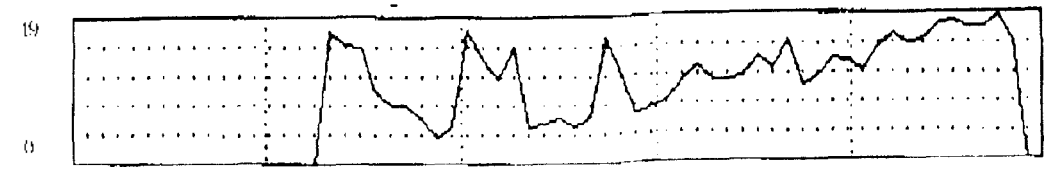
(D) Detected Transition Forme for Waveform(a)

그림 4. /인수내 고마운 천재소년을 좋아한다/ 음성에 대한 처리결과.
Results for speech / Insoonae Komanun Chunjae Sonyunwl Joahanda /.

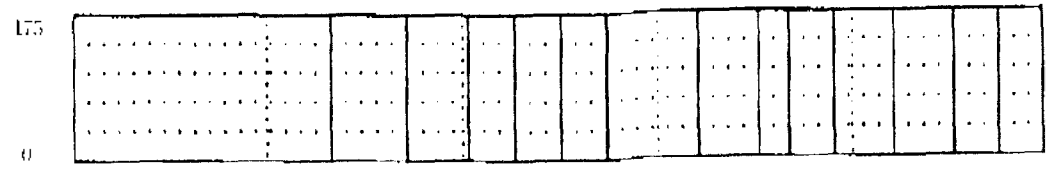
(A) Speech Waveform for b:ch#3 1s. dat



(B) Average Magnitude Contour for Waveform(a)

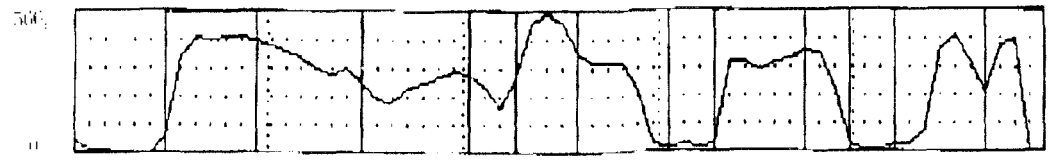


(C) Symmetrical Distortion Rate for Waveform (a)

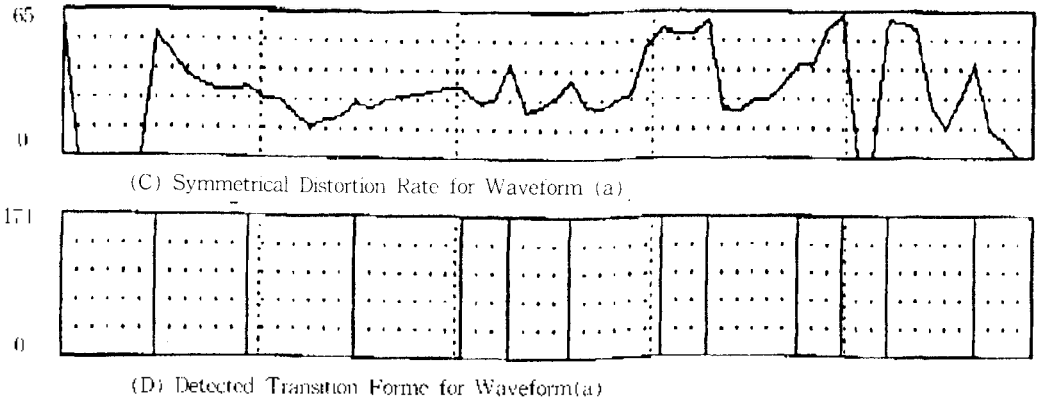


(D) Detected Transition Forme for Waveform(a)

(A) Speech Waveform for b:chs#3 2s. dat

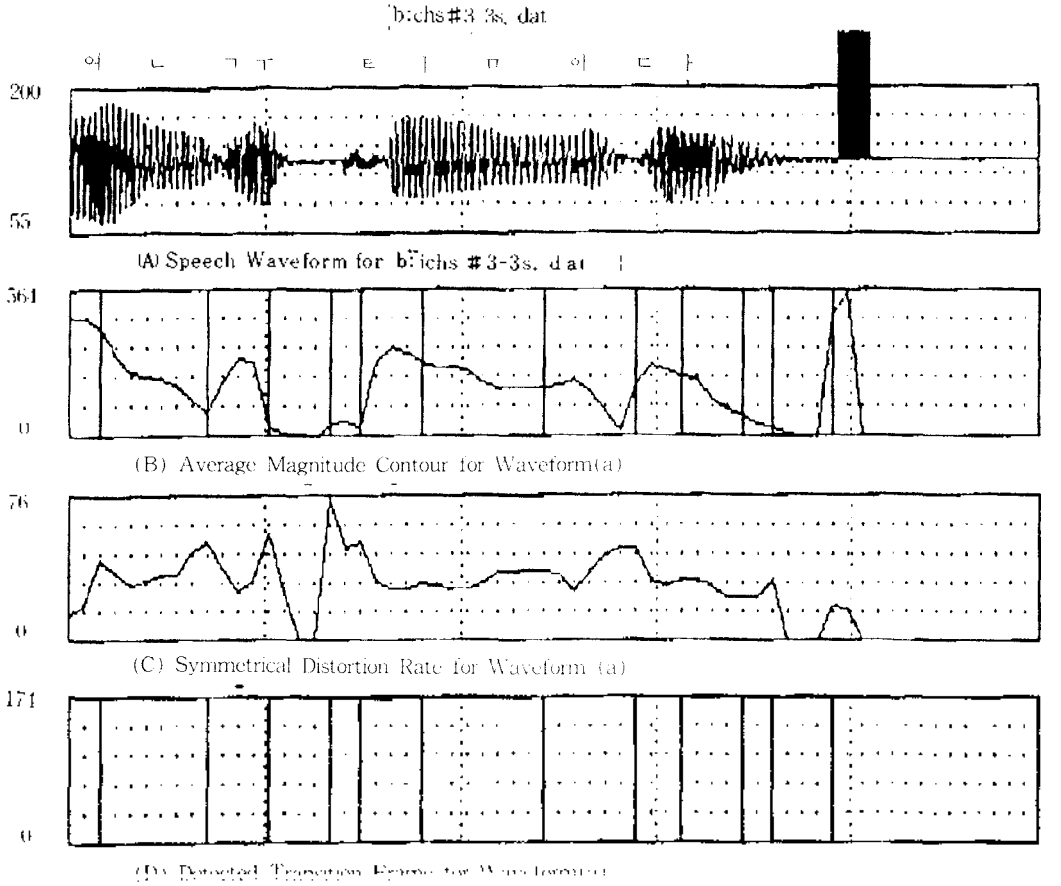


(B) Average Magnitude Contour for Waveform(a)



(C) Symmetrical Distortion Rate for Waveform (a)

(D) Detected Transition Forme for Waveform(a)



(A) Speech Waveform for b:ichs #3-3s. dat

(B) Average Magnitude Contour for Waveform(a)

(C) Symmetrical Distortion Rate for Waveform (a)

(D) Detected Transition Forme for Waveform(a)

Fig 5.

Results for Speech Sculdae Immakomyeknkwa wmsungstnhochun Yungutmada

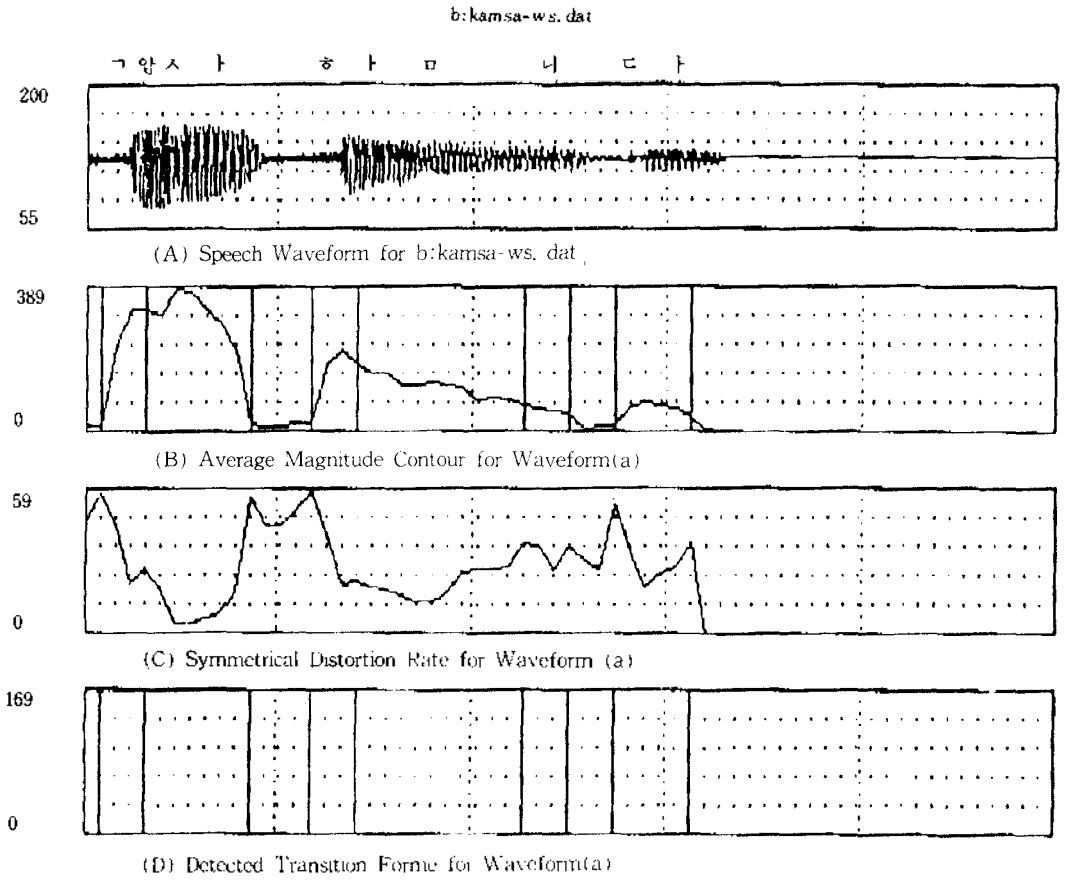


그림 6. '감사합니다' 음성에 대한 처리 결과.
Results for speech 'Kamsahamnida'.

영향을 많이 받고, 또한 전이구간 검출에 대한 성능 측정이 복잡해진다.

따라서 본 논문에서는 음소의 전이구간 검출시에 평균진폭이 갖는 제반 문제점들을 제거하기 위해서 프레임내의 파형이 이루는 비대칭율 피레이더를 평균진폭 피레이터 대신에 제안하였다. 제안된 비대칭율을 이용하면 음소의 전이구간을 간단한 비교규칙에 의해 쉽게 추정할 수 있고, 또한 전이구간의 검출이나 비대칭율값에 의해 유성음구간의 성질도 근사적으로 파악할 수 있다.

參 考 文 獻

1. C. J. Weinstein, S. S. McCandless, L. F. Mondshein, and V. W. Zue, "A System for Acoustic-Phonetic Analysis of Continuous Speech", IEEE Trans. on ASSP, vol. ASSP 23, No. 1, pp. 54~67, Feb. 1975.

2. W. F. Gimong, and J. J. Zaiorre, "Measuring phoneme Boundaries Four ways", J. Acoust. Soc. Am. vol. 68, No. 2, pp. 531~539, Aug. 1980.

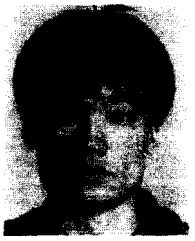
3. 고수정 외 2인, "A Segmentation Algorithm of the Connected word Speech by Statistical Method", 대한전자공학회지, vol. 26, No. 1, pp. 151~162, Apr. 1989.

4. R. Mori, P. Laface, and E. Piccolo, "Automatic Detection and Description of Syllabic Features in Continuous Speech", IEEE Trans. on ASSP, vol. ASSP 24, No. 2, pp. 365~372, Oct. 1976.

5. L. R. Rabiner, and M. R. Sambur, "Some Preliminary Experiments in the Recognition of Connected Digits", IEEE Trans. on ASSP, vol. ASSP 24, No. 2, pp. 170~182, Apr. 1976.

6. R. Mori, and P. Laface, "Use offuzzy Algorithms for phonetic and Phonemic Labeling of Continuous Speech", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. PAM-2, No. 2, pp. 136~148, Mar., 1980.
7. P. Merelstein, "Automatic Segmentation of Speech into Syllabic Units", J. Acoust. Soc. Am., vol. 58, No. 4, pp. 365~379, Oct., 1975.
8. 서울, 국어 음운학, 탑 문화사, 1985.
9. M. BAE, J. RHEEM, and S. ANN, "A Study on the Energy Extraction using G-Peak from the Speech Production Model", KIEE, vol. 21, No. 3, pp. 381~386, May 1987.
10. M. BAE and S. ANN, "On Improving the Effects of Varying the Window Length on Speech Energy Computation", J., Acoust., Soc., Korea, vol. 9, No. 2, pp. 34~41, April 1990.
11. S. D. Stearns and R. A. David, Singal Processing Algorithm, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1987.
12. 최정여, 이인섭, 배명진, 안수길, "음성신호의 전이구간 검출", 대한전자공학회 하계학술발표 논문집, vol. 12, No. 1, pp. 629~631, 1989년 7월.
13. L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Inc., 1978.

▲이 율 재



1963년 8월 1일생
1989년 2월 : 호서대학교 전자공학과 졸업
1989년 3월~현재 : 호서대학교 대학원 전자공학과 석사과정

▲배명진 9권 1호 참고
▲안수길 9권 1호 참고