규준화된 AMDF 이용한 음성파형의 안정상태 구간검출

(On Detecting the Steady State Segments of Speech Waveform by using the Normalized AMDF)

배 명 진** 김 울 제* 안 수 길***

(Myungjin Bae, Ulje Kim, Souguil Ann)

요 약

연속음 인식을 위해서는 음성신호의 음성학적 경계를 결정짓는 분할과정이 필요하다. 본 논문에서는 음성신호의 전이구간을 결정하기 위한 파라미터로 한 프레임내의 규준화된 AMDF을 제안하였다. 제안된 규준화된 AMDF은 그 프레임에서 음성진폭의 변화율을 대별하며, 인근 프레임의 규준화된 AMDF와 비교하면 현재의 프레임이 정상상태 혹은 전이영역에 있는지를 구별할 수 있게 해준다.

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 the transition regions in continued speech, we propose a new noramlized AMDF. The suggested parameter represents a change rate of magnitude of speech signals. As comparing this value with the adjactent frame s value the state of the frames can be distinguished as a level between the steady state and transient state.

I. 서 론

연속된 음성의 인식을 위해서는 전기신호로 표시된 음성신호를 음성학적 단위인 단어, 음절, 음소 등의 단위로 분할하여야 한다. 연속음을 이러한 단위로 분류하면 분석시에 분석의 반복을 줄일 수 있고, 음성인식 과정에서 고립단어의 인식기법을 연속음인식에 쉽게 연장시켜 적용할 수 있게 된다.

음성신호의 전이구간을 검출하는 연구는 특성파라미터를 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[3-5, 9]. 스펙트럼영역법은 음성신호의 음소의 변화에 따른 포먼트의 전이 특성이나 주파수성분별 에너지비를 이용하는 것 등이 제안되어져 있다[2, 4]. 또한 혼성(hybrid)영역법은 변환영역에서의 특징 파라미터를

*호서대학교 전자공학과 대학원
**호서대학교 전자공학과 조교수
***서울대학교 전자공학과 교수

을 이용하는 것으로[3,6], LPC계수의 전이특성, LPC 에러의 변화특성, cepstrum 등을 이용하고 있다.

시간영역법에서는 파라미터의 검출은 비교적 쉬우나, 그 변화정도를 정확히 파악하기 위한 결정논리가 상대적으로 어렵다[13]. 반면 스펙트럼영역법이나 혼성법은 비교적 정확하지만 계산의 성밀도나 변환 차수의 영향을 받게되고, 전처리과정으로 보기에는 계산량의 부담이 시간영역법에 비해 큰편이다.

따라서 우리는 시간영역법 전이구간 검출용 파라미터들 중에서 음성에너지 파라미터가 갖는 부정확성과 결정논리의 복잡성에 대해 알아보고, 이러한 문제점을 제거할 수 있는 새로운 파라미터를 제안하고자 한다.

II. 음성신호의 전이구간

음성신호는 생성모델에 따라 유성음, 무성음, 혼합음, 묵음으로 구분지을 수 있다. 유성음은 준주기성과 성도의 공명으로 큰 에너지를 가지며, 무성음은 준색잡음의 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 그림 1과 같이 나타나게 된다. 그림 1은 / 오육오 / 라는 연결단어를 24세의 남성화자가 발성한 것이며 평균진폭의 변화도(contour)가 음소나 음절의 변화를 잘 나타내고 있음을 알 수 있다.

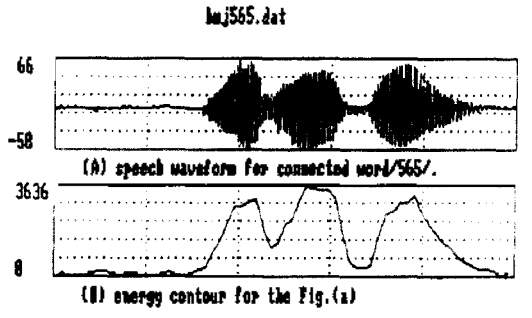


그림 1. 연결음 / 오육오 / 에 따른 평균진폭의 변화도
Fig. 1. Absolute magnitude contour for connected speech / ouyukou /

평균진폭의 변화도를 이용하여 음절의 전이구간을 분류하려면 우선 평균진폭을 계산해야 하는데, 이때 그 프레임에 적용된 윈도우의 영향을 받게 된다. 윈도우의 영향으로는 윈도우의 길이와 형태에 따른 영향을 고려할 수 있다. 윈도우내의 음성 성분 주파수가 윈도우 길이의 역수에 성배수일 때가 윈도우길이 영향을 가장 적게 받게 된다. 유성음일 경우에는 윈도우와 길이가 피치의 정수배로 선택되는 것이 가장 바람직하게 된다. 그렇지만 사전에 피치를 정확히 구해야 하고 또한 윈도우의 길이가 가변적이어야 하기 때문에 윈도우길이를 피치에 일치시키지 않고 윈도우길이 영향을 최소화시키려는 연구도 많이 제안되고 있다[9-10].

윈도우의 길이를 음성의 피치에 정수배로 하여도 음성신호 성분에는 정수배가 아닌 성분들이 또한 존재할 수 있기 때문에 윈도우형태를 잘 선정할 필요가 있다. 윈도우의 형태는 통과 및 차단대역의 비에 따라 방형, 삼각형, 해밍, 블랙맨 등이 있으며, 차단특성이 우수한 윈도우 함수일수록 계산과정이 복잡하게 된다[11].

윈도우의 길이에 따른 평균진폭의 변화는 그림 2와 같이 크게 달라진다. 윈도우의 길이가 음소의 변화특성에 비해 길게되면 스므딩되어 평균진폭의 변화도는 음소변화의 특성을 잘 나타낼 수 없게 된다. 반면 음소 변화특성에 비해 윈도우의 길이를 너무 짧게하면 평균진폭의 변화도에는 국부봉우리가 많이 나타나서 정확한 음소변화 특성을 구하기 어렵게 된다.

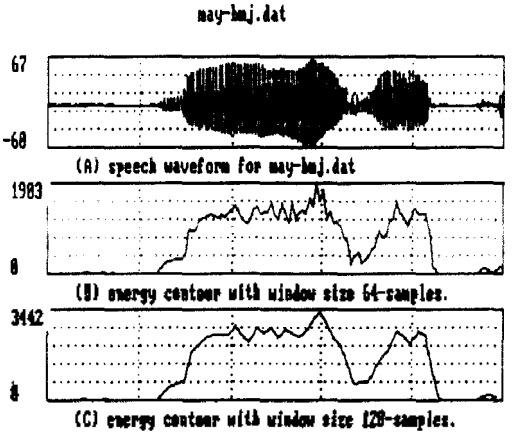


그림 2. 방형윈도우의 길이에 따른 평균진폭의 변화도.
Fig. 2. Average magnitude contours according to the length of the rectangular window.

연속음에 대해 윈도우를 잘 선정하여 평균진폭을 구하여도 그 변화특성을 수치적으로 잘 나타내는 결정논리가 필요하다. 결정논리 적용시에는 크게 두가지의 어려움이 따른다. 첫째로는 아무리 윈도우를 잘 적용하여도 윈도우내의 음성성분이 복잡하고 다양하여 평균진폭의 변화도에 국부봉우리들이 존재할 수 있다. 이러한 국부봉우리와 실제 음소전이를 나타내는 음소봉우리를 분리해야만 하는 어려움이 있다. 두번째로는 음소봉우리의형태가 여러 가지라는 점이다. 예를 들어 유성음에 이은 비음연결, 무성음과 비음 또는 유성음의 연결 등에서는 표준봉우리의 형태를 찾기 힘들다.

Ⅲ. 한 프레임구간에서 파형의 규준화된 AMDF

음소변화는 음성파형에 비해 서서히 변화하기 때문에 프레임 단위로 분석하는것이 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.

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

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

윈도우의 영향에 무관하게 현재의 프레임이 어떤 상태에 존재하는지를 측정하는 새로운 방법으로는 다음과 같이 규준화된 AMDF(normalized average magnitude difference function, NAMDF)를 정의해서 사용할 수 있다.

$$NAMDF(d)=\frac{\sum_{n=1}^{N-1} | s(n)-s(n-d) |}{\sum_{n=1}^{N-1} | s(n) |+| s(n-d) |} \quad \cdots\cdots(2)$$

여기서 d는 규준화된 AMDF를 측정하려는 지연인자이며, N는 AMDF를 구하려는 윈도우 구간이다. 지연인자를 점차 증가시키면서 이 AMDF를 구해보면, 지연인자가 프레임내의 음성피치의 정수배가 될 때 마다 NAMDF는 거의 영이되고 곱셈이 필요치 않기 때문에 자기-상관함수 대신에 주기성을 강조하는데 오랫동안 적용되어 왔다.

또한 식(2)의 AMDF 값은 지연인자 d간격을 갖는 N개 샘플간의 진폭에 대한 평균 차이값이 되기 때문에 음성파형의 d-구간사이의 유사도를 나타내는 규준화된 거리값으로 적용할 수도 있다. 식 2의 규준화된 AMDF는 d-간격의 두 음성파형 블럭에 대해, 평균진폭 차이값을 나타내지만 음성신호가 갖는 피치주기의 변화는 배제하지 않았다. 따라서 두 음성파형 블럭에 대한 피치주기의 영향을 제거하려면 시간을 고려하는 지연인자 d-의 값이 음성피치에 일치하였을 때의 규준화된 AMDF값을 두 파형블럭의 유사도값으로 사용할 수 있게 된다.

d-의 값이 음성피치와 일치하였을 때 규준화된 AMDF 값이 가장 적게 되므로 각 프레임내에서 d-를 변화시키면서 규준화된 AMDF의 최소값을 다음과 같이 구할 수 있다.

$$NNAMDF(fr)=MIN\{NAMDF(100), NAMDF(101), \cdots, NAMDF(199)\} \quad (3)$$

여기서 MIN{ · } 함수는 주어진 변수영역에서 최소치를 선택하는 함수이고, fr은 현재 프레임의 위치를 나타낸다. 지연인자를 100샘플(8KHz 표본율에서 12.5ms) 부터 구한 이유는 AMDF 값은 지연인자가

0일 때와 음성피치의 정수배일 때 최소치가 되기 때문이다. 따라서 규준화된 AMDF는 지연인자 d-를 실존하는 음성피치(2.5ms에서 25ms까지)의 최장 길이의 1/2에서 부터 증가시켰을 때만 두 파형블럭의 유사도를 나타내는 거리값이 된다.

이 거리값이 영에 근접하면 d-간격을 유지하는 두 음성파형 N-개 블럭간에는 유사성이 최대가 된다. 따라서 두 음성파형의 블럭이 놓여 있는 위도우구간은 정상상태가 이루어지고 있음을 나타낸다. 현재 프레임내에서 규준화된 AMDF을 구했을 때, 그 거리값이 1에 근접하면 이 프레임은 전이구간에 놓여있게 된다.

Ⅳ. 규준화된 AMDF에 의한 안정상태 구간의 결정

23세의 여성화자가 발음한 고립단어 /삼/의 음성신호에 대해 규준화된 AMDF을 구한 것을 그림 3에 나타내었다. 여기서 그림 3(a)는 음성파형을 나타내며, 그림 3(b)에는 평균진폭의 변화도를 나타내었다. 평균진폭은 각 프레임을 256샘플 단위로 하고 128프레임씩 겹치게하여 구한 것이다. 이때 규준화된 AMDF의 최소치에 대한 변화도를 그림 3(c)에 나타내었다.

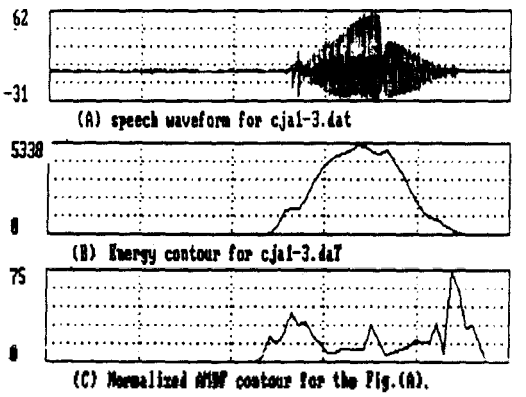


그림 3. 음성신호 /삼/에 대한 규준화된 AMDF 최소값의 변화도.

Fig 3. Minimum NAMDF Contour for speech /SAM/

그림 3(c)에 제시한 규준화된 AMDF의 변화도를 살펴보면, 음소가 시작되는 영역에서는 규준화된 AMDF가 상대적으로 크다는 것을 알 수 있다. 또한 평균진폭의 변화도와 비교할 때 음소의 전이가 주어지는 프레임 구간에서의 규준화된 AMDF 값은 인근 프레임의 값에 비해 높아지며 정점을 이루게 됨을 알 수 있다. 반면 이 변화도에서 골을 이루는 위치에서는 음성파형이 정상상태에 있음을 알 수 있다.

또한 파형의 구조가 단순한 비음 /ㅁ/에 비해 파형의 구조가 복잡한 /아/음의 프레임에서 규준화된 AMDF는 상대적으로 높아지는 특징이 있다. 그리고 음소의 변화가 빠른 자음연결이나 끝나는 프레임구간에서는 모음구간에 비해 규준화된 AMDF의 봉우리 폭과 경사가 급하다는 것을 알 수 있다.

이상을 정리해 음소의 전이구간을 선택하기 위한 결정논리를 만들 수 있다. 규준화된 AMDF가 음성신호 파형의 전반적인 변화도를 나타내기 때문에 각 프레임 마다 규준화된 AMDF을 구하여 그 변화도가 골을 이루면 지금의 프레임이 안정상태 구간이 된다. 반면 봉우리를 이루면 지금의 프레임이 안정상태 구간이 된다. 반면 봉우리를 이루면 이 구간은 인근 프레임에 비해 가장 큰 전이구간을 이루게 된다.

Ⅴ. 실험 및 결과

시뮬레이션을 위해 IBM PC/AT 시스템에 마이크로폰의 입력이 가능하도록 12-비트 analog to digital 변환기를 인터페이스 시켰다. 화자는 남성화자 2명과 여성화자 1명을 통해 다음의 연속음을 발성케 하고 8KHz로 표본화 하면서 저장시켰다.

발성1) 24세 남성화자 : "인수네 꼬마가 천재소년을 좋아한다."

발성2) 28세 남성화자 : "서울대 전자공학과 음성신호처리 연구팀이다."

발성3) 25세 여성화자 : "감사합니다."

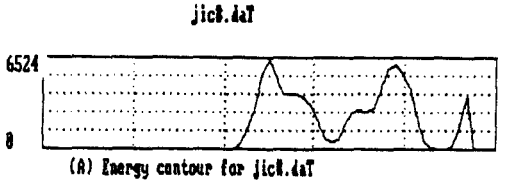
각 음성시료에 대해 한 프레임의 길이를 300샘플(=37.5msec.)로 하여 200샘플씩 겹치게 처리하였다. 각 프레임에 대해 규준화된 AMDF을 계산한

지연인자 d-의 구간은 100샘플(=12.5msec.)에서 199샘플(=25msec)까지 100개의 규준화된 AMDF을 측정하였다. 여기서 최소값(식 3의 NNAMDF(fr))을 구하여 이 프레임의 규준화된 AMDF 대표값으로 사용하였다.

그림 4, 5, 6에는 발성1), 발성2), 발성3)에 대한 처리결과를 각각 나타내었다. 각 그림에서 음성시료의 파형에 따른 평균진폭의 변화도를 그림(a)에 나타내었으며 이것을 통해 유소변화의 개략적인 평가기준으로 삼을 수 있다. 100샘플에서 부터199샘플까지 음성파형의 한 블럭을 100샘플(=12.5 msec) 단위 마다 규준화된 AMDF를 계산하고, 이들 중에서 최소치를 그 프레임의 대표값으로 하여 그림(b)에 나타내었다.

또한 그림(b)의 규준화된 AMDF의 변화도에서골을 이루는 구간을 음소의 안정상태 구간으로 결정하여 각 그림(c)에 나타내었다. 그림(c)와 평균진폭의 변화도에서 변화특성을 비교하기 쉽도록 찾아진 안정상태 구간의 에너지값들에 의해 선형적으로 인터폴레이션한 것을 그림(d)에 나타내었다. 여기서 연속음성의 진폭변화 특성은 규준화된 AMDF로 찾은 안정구간들만에 의해서도 잘 대별하고 있다.

그림(a)와 (d)를 비교해 보면, 제안한 방법으로 음성파형의 안정상태 구간들을 찾아 이 구간에서만 구체적인 분석을 행함으로써 분석의 복잡성이나 데이타량을 줄일 수 있으며, 음성인식시에도 결정논리를 간단히 할 수 있게 된다. 특히 그림 4의 두번째 초반과 중반 그림이나, 그림 5의 두번째 초반 그림 등에서는 비음구간인데 평균진폭의 변화도로는 구별하기 힘든 음소의 전이구간도 비음화된 한 블록구간으로 잘 나타내고 있다. 그림 4의 세번째 초반 그림은 유성음들 끼리 별 변화없이 연속된 음성인 경우인데 규준화된 AMDF의 변화도는 이를 정확히 분류해 주고 있다.



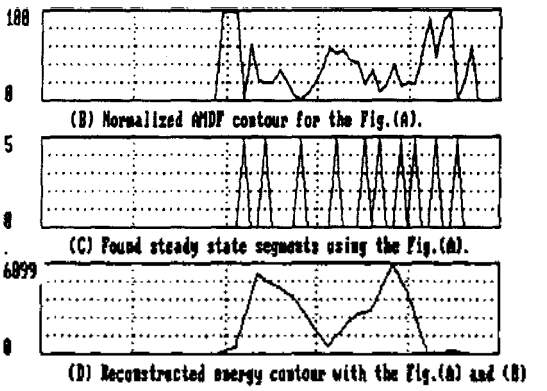


그림 4-1. / 인수네 꼬마가 천재 소년을 좋아 한다. / 음성에 대한 처리 결과
Fig. 4-1. Results for speech / insoonae komaga chunjea sonyunwl joahanda. /

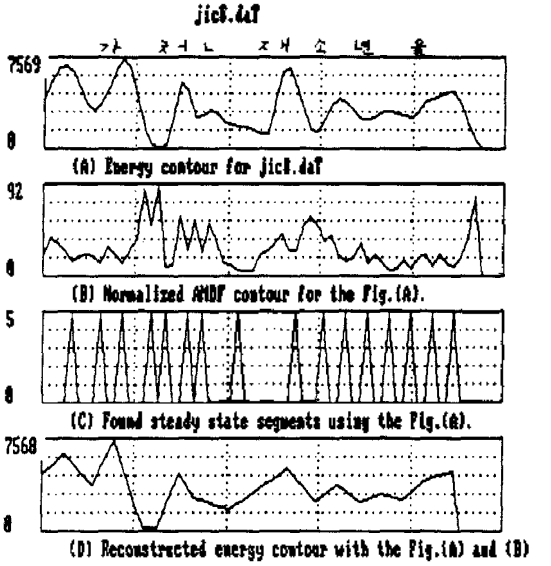
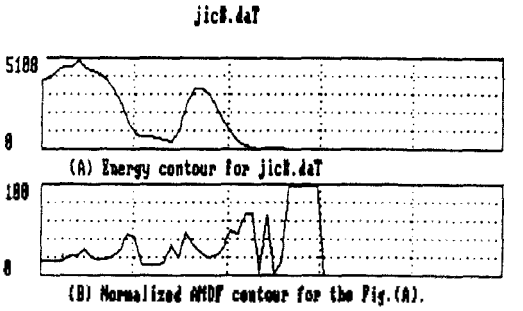


그림 4-2. / 인수네 꼬마가 천재 소년을 좋아 한다. / 음성에 대한 처리 결과
Fig. 4-2. Results for speech / insoonae komaga chunjea sonyunwl joahanda. /



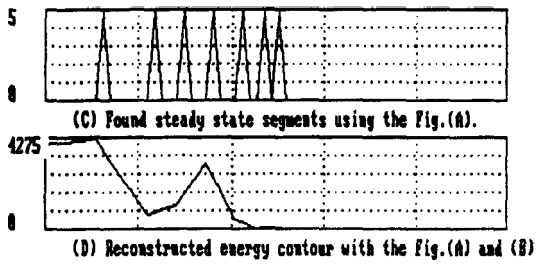


그림 4-3. / 인수내 꼬마가 천재 소년을 좋아 한다. / 음성에 대한 처리 결과
Fig. 4-3. Results for speech / insoonae komaga chunjea sonyunwl joahanda. /

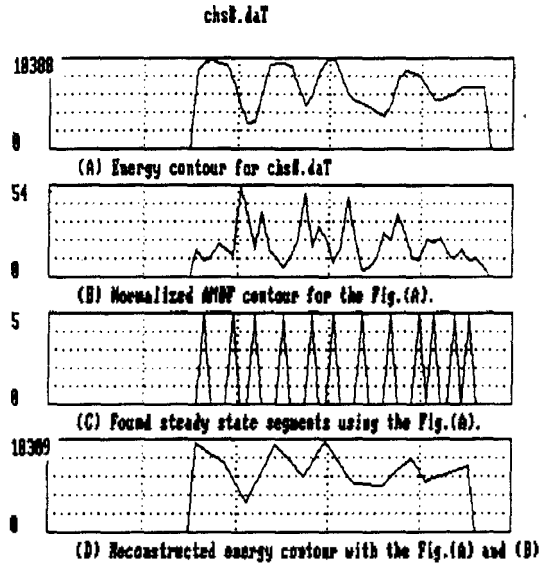
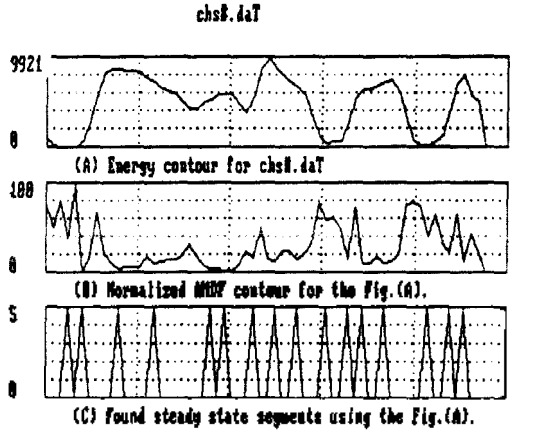


그림 5-1. / 서울대학교 음성신호처리 연구팀 입니다. / 음성에 대한 처리결과
Fig. 5-1. Result for speech / souldae junjakonghakwa wmsungsinhochuri yungutimida / .



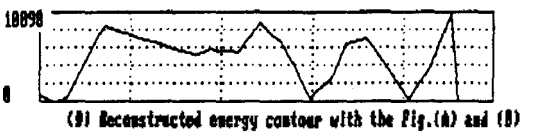


그림 5-3. / 서울대학교 음성신호처리 연구팀 입니다. / 음성에 대한 처리 결과
Fig. 5-3. Result for speech / souldae junjakonghakwa wmsungsinhochuri yungutimida /

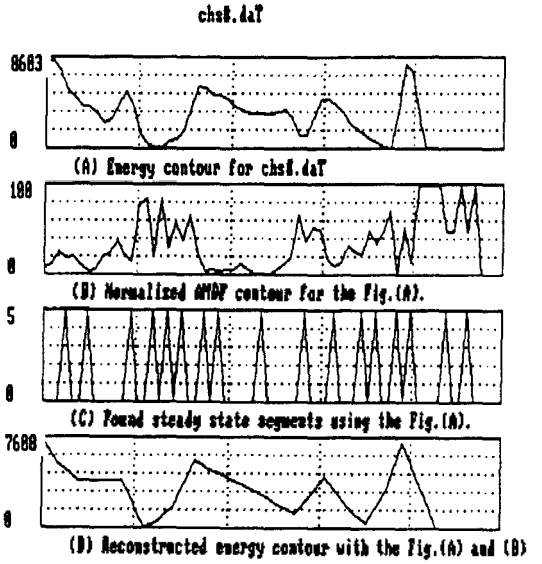


그림 5-4. / 서울대학교 음성신호처리 연구팀 입니다. / 음성에 대한 처리 결과
Fig. 5-4. Result for speech / souldae junjakonghakwa wmsungsinhochuri yungutimida /

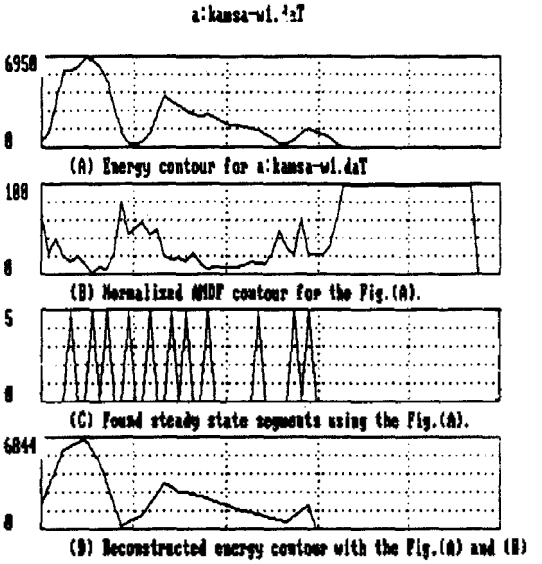


그림 6. / 감사합니다 / 음성에 대한 처리 결과.
Fig. 6. Results for speech / Kamsahamnida /

Ⅵ. 결론

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

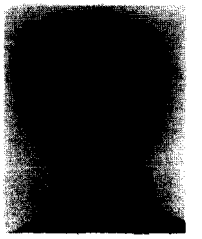
따라서 본 논문에서는 시간영역에서 음소의 전이구간 검출시에 평균진폭이 갖는 제반 문제점들을 제거하기 위해 프레임내의 파형이 이루는 규준화된 AMDF 파라미터를 평균진폭 파라미터 대신에 제안하였다. 제안된 파라미터를 파형의 안정구간 검출에 적용하면, 간단한 비교논리에 의해 쉽게 그 구간을 찾을 수 있다. 또한 전이구간의 간격이나 규준화된 AMDF값에 의해 유성음구간의 성질도 근사적으로 파악할 수도 있다.

참 고 문 헌

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. Ganong, and R.J. Zatorre, "Measuring Phoneme Boundaries Four Ways", J. Acoust. Soc. Am, Vol. 68, No. 2, pp. 431-439, Aug. 1980.
3. 김수중 외 2인, "A Segmentation Algorithm of the Connected Word Speech by Statistical Method", 대한전자공학회지, Vol. 26, No. 4, pp. 151-162, Apr., 1989.
4. R. Mori, P. Laface, and E. Piccoo, "Automatic Detection and Descripton of Syllabic Features in Continuous Speech", IEEE Trans. On ASSP, Vol. ASSP 24, No. 2, pp. 880-883, 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 of Fuzzy Algorithms for Phonetic and Phonemic Labeling of Continuous Speech", IEEE Tracs. on Pattern Analysis and Machine Intelligence, Vol.PAM-2, pp. 436-448, Mar., 1980.
7. P. Mermelstein, "Automatic Segmentation of Speech into Syllabic Units", J. Acoust. Soc. Am., Vol.58, No. 4, pp. 365-379, Oct., 1975.
8. 허 웅, 국어 음운학, 샘 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.24, 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, Signal Processing Algorithm, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1987.
12. 최정아, 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.

▲배 명 진 현 호서대학교 조교수
9권 1호 참조

▲김 을 재(학생회원) 1965년 7월 13일생



1984년 : 호서대학교 전자공학과 입학
1991년 : 호서대학교 전자공학과 졸업.
현재 : 호서대학교 전자공학과 대학원 재학.

▲안 수 길 현 서울대학교 교수
9권 1호 참조