음성 에너지계산에서 창함수-길이 변화영향의 개선에 관한 연구

On Improving the Effects of Varying the Window Length on Speech Energy Computation

배 명 진* 안 수 길**

(Myungjin Bae, Souguil Ann)

요 약

음성신호의 전처리과정에서 에너지 파라미터는 음소의 변화특성을 나타내기 때문에 많이 사용하고 있다. 그렇지만 추출과정에서 창함수를 적용하기 때문에 창함수길이에 따른 영향을 받게된다. 본논문에서는 창함수길이에 따른 영향을 측정하고 그 영향을 최소화시키는 에너지추출법을 새로이 제안하였다. 이방법으로 추출된 에너지변화도는 창함수길이의 영향을 제거시켰기 때문에 음소의 변화특성을 잘나타낸다. 또한 계산시간은 샘플당 한번의 뺄셈과 덧셈, 그리고 두번의 비교연산만 있으면 된다.

ABSTRACT

The energy parameter is widely used in pre-processing of speech signals,because it represent the phoneme characteristics of well But, the energy parameter is affected by the window length during the extracting. Thus, in this paper, the window length effects are studied in detail, and we proposed a new energy extraction algorithm that reduces the length effects. The energy contou rs with this algorithm are well representing for the characteristics of speech phon emes. And the computations to implement the algorithm are only required one subtraction, one addition, and two comparison operation per speech sample.

I. 서 론

사회가 고도로 산업화되고 모든 설비가 자동화되는 지금에는 그 설비를 운용하고 제어하기 위한 인간－기계의 통신이 중요하며, 그 통신수단으로는 음성메시지가 가장 바람직하기 때문에 음성신호처리에 관한 연구가 오랫동안 계속되어 오고 있다. 이러한 음성신호처리는 복잡하기 때문에 보통 그 과정을 전처리과정과 후처리과정으로 나누어 처리하고 있

* 호서대학교 전자공학과
** 서울대학교 전자공학과

다. 전처리과정에서는 본격적인 처리를 하기 전에 음성구간을 찾거나 음성신호를 그룹별로 분류하는 등의 사전작업을 하게 된다.

전처리과정에 적용되는 파라미터로는 음성에너지, 영교차율, 자기상관계수 등이 이용되고 있다[1]. 특히 음성신호의 시간별 에너지변화는 검출과정이 간단하고, 그 값의 의미는 유성음의 음소변화를 파악하거나 무잡음 음성(clean speech)의 구간검출(end point detection)등에 아주 유용하다.

단시간 음성에너지를 계산하려면 창함수(window function)가 적용되어야 하며, 이 때문에 창함수의 길이(duration)는 구해진 에너지의 정확성에 큰 영향을 주게된다. 창함수의 길이가 음성신호의 피치 보다 짧으면 에너지변화에 국부봉우리가 많이 나타난다. 반면 창함수길이가 피치 보다 길게 되면 스므딩현상에 의해 음성신호가 갖는 음소의 변화특성이 잘 나타나지 않게 된다. 가장 바람직한 방법으로는 창함수길이를 피치에 일치시키는 것이나 이것은 피치를 찾는 작업이 선행되어야 함으로 피치검출 에러의 영향을 받거나 계산과정의 복잡성이 부가될 수 있다. 이 때문에 기존의 음성 에너지검출법들은 창함수길이에 따른 영향을 감수하면서 음성에너지에서 얻을 수 있는 개략적인 성질만을 이용하고 있다.

음성에너지 검출시에 창함수길이의 영향이 어느 정도인지를 간단한 계산으로 알 수 있다면 이의 영향을 개선하거나 이용할 수 있게 된다. 본 연구에서는 우선 창함수길이의 변화가 에너지추출시에 어떤 영향을 나타내는 가에 대해 알아보고 그 영향을 개선하는 개선된 음성에너지 추출법을 창안하게 된다. 그런 다음에는 개선된 음성에너지 추출법의 계산시간 감축에 관한 알고리즘을 제안하게 된다.

Ⅱ. 음성신호의 에너지 추출

음성신호는 발성모델의 여기음에 따라 유성음, 무성음, 혼합음 그리고 묵음으로 분리되며, 유성음은 성대의 진동에 의해 성도가 여기되어 발성되기 때문에 그 파형은 성대진동의 기본주기인 피치와 성도 공명에 의한 큰 에너지성분을 갖게 된다. 무성음은 성대의 진동을 수반하지 않으면서 단지 성도의 협착점을 스치고 지나가기 때문에 그 파형은 에너지가 낮고, 공기의 교란에 의한 불규칙한 색잡음의 형태가 된다. 한편 혼합음(mixed excitation)의 파형은 유성음과 무성음이 함께 존재하여 복합파형을 띄게 된다.

유무성음 그리고 혼합음이 갖는 에너지 레벨이 서로 다르기 때문에 에너지의 변화형태가 음소의 변화를 근사적으로 나타내게 된다. 또한 SNR이 우수한 무잡음 음성일 경우에는 에너지의 형태만으로 음성신호구간을 구별해낼 수 있다. 음성신호 s(n)의 단시간 에너지 E(n)은 다음과 같이 나타낼 수 있다.

$$E(n)=\sum_{k=-\infty}^{\infty} s(n)s(n)w(n-k) \quad (1)$$

여기서 w(n)은 창함수로 시간의 구간 N-동안의 함수이다. 창함수에는 그림 1-1과 같이 방형창 (rectangular window), 삼각형창 (triangular window), 해밍창 (Hamming window)등이 있으며, 방형창 함수식은 다음과 같다[3].

$$w(n)=1,\ 0\leq n\leq N-1$$
$$=0,\ \text{otherwise} \quad (2)$$

식 1의 에너지관계식은 음성신호의 자승이 적용되어 값의 범위가 광범위하고 계산과정에서 곱셈이 적용됨으로 다음과 같은 평균진폭 (average magnitude) 값을 대신 사용하기도 한다.

$$M(n)=\sum_{k=-\infty}^{\infty} |s(k)|\ w(n-k) \quad (3)$$

음성에너지 계산시에 필연적으로 나타나는 문제점중 하나는 창함수길이 N에 따른 영향이다. A(n)=s(n)s(n)이라 할 때, 식 2의 방형창을 사용하여 에너지 식을 다시 쓰면,

$$E(n)=\sum_{k=-\infty}^{\infty} A(k)\ w(n-k)$$
$$=A(n)*w(n) \quad (4)$$

이 되고, 이에 대한 주파수응답은 다음과 같다.

$$E(e^{j\Omega T})=A(e^{j\Omega T})\ W(e^{j\Omega T})$$

$$=A(e^{j\Omega T})\ \frac{\sin(\Omega NT/2)}{\sin(\Omega T/2)}\ e^{j\Omega T(N-1)/2} \qquad (5)$$

따라서 에너지의 계산은 f_s를 샘플링주파수라 할 때 차단주파수가 $F=F_s/N$인 저역통과필터에 자승된 음성신호 A(n)을 통과시킨 것과 같다. 저역통과필터가 이상적이라면 차단주파수 이상의 성분들에 의한 영향이 완전히 제거되지만, 방형창의 경우에는 식 5에 주어진 것처럼 주파수응답이 sinc 함수형 저역통과필터이기 때문에 차단주파수 이상의 성분이 영향을 주는 앨리어징 (aliasing) 현상이 나타나게 된다.

앨리어징의 영향을 최소화하기 위한 기존의 방법은 두 가지가 있다. 첫째는 창함수를 선택할 때 통과대역과 차단대역이 뚜렷한 창함수를 선택하는 방법으로서, 적용이 쉬워 주로 많이 사용하는 방법이다. 이들중 방형창, 해밍창, 그리고 블랙맨창의 주파수특성을 나타내면 그림 1-2와 같고, 창함수에 따른 통과대역과 차단대역의 비를 근사적으로 나타내면, (1) 방형창 : 약 10 dB, (2) 삼각형창 : 약 30 dB, (3) 해밍창 : 약 40 dB, (4) 블랙맨창 : 약 60 dB 등이 된다[9]. 차단특성이 좋은 창함수일 수록 그 계산과정이 복잡하기 때문에 보통은 해밍창함수를 많이 사용하고 있다[10]. 한편 창함수의 차단특성이 우수해도

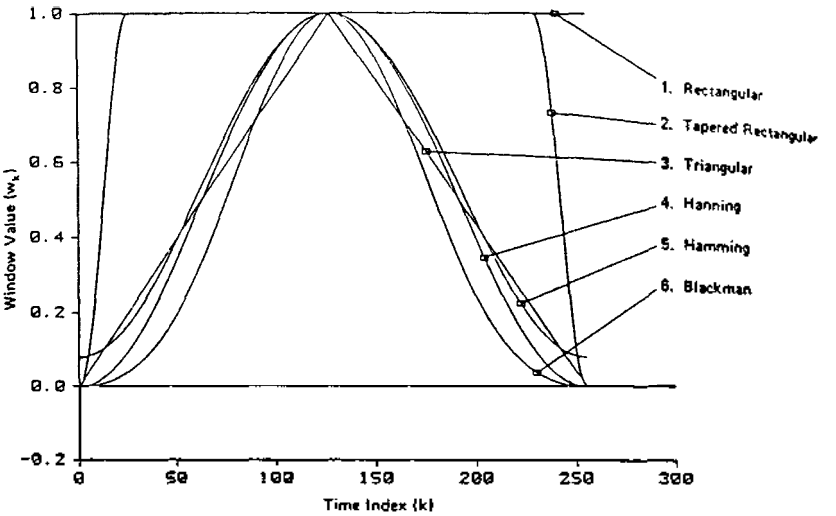

Fig. 1-1 Examples of six windows with length N=256.

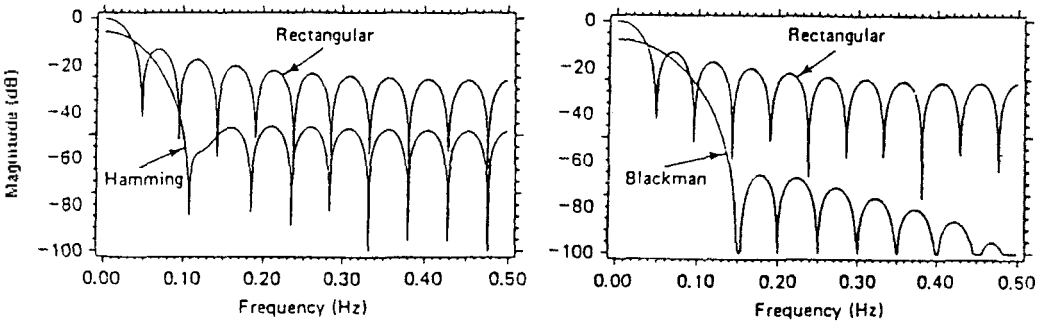

Fig. 1-2 Frequency domain characteristics of Rectangular, Hamming, and Blackman windows.

그림 1 여섯가지 창함수의 예시

창함수길이가 유성음의 피치와 다르면 차단대역에 들어있는 성분들이 창함수구간마다 다르므로 에너지 변화도를 잘 나타내지 못하고 그 에너지차들인 국부 봉우리를 많이 갖게 된다.

앨리어징효과를 줄이는 두번째 방법은 창함수를 그대로 두고 음성신호의 기본주파수에 창함수 길이 N-을 일치시키는 방법이다. 음성신호에서 에너지를 주로 갖는 것이 유성음이다. 유성음은 성대의 기본주기인 피치로 성도를 자극하여 발성되기 때문에 그 스펙트럼은 피치의 역수인 기본주파수의 하모닉스에 성도특성인 포먼트들의 주파수특성이 곱해진 형태가 된다. 따라서 기본주파수와 하모닉스들에만 주파수성분이 몰려있게 된다. 창함수길이를 피치에 일치시키면 유성음의 기본주파수 하모닉스 마다 창함수 하모닉스들의 골이 곱해져서, 차단주파수 이상의 성분들을 약화시키기 때문에 앨리어징효과를 경감시킬 수 있게 된다.

그렇지만 이방법은 창함수를 적용하기 전에 음성신호의 피치검출이 선행되야 한다. 지금까지 제안된 음성신호의 피치검출법은 음소의 천이영역이나 주변환경에 따라 많은 에러를 일으킬 수 있으며, 그 처리과정도 또한 복잡하다[4]. 그리고 창함수길이를 피치에 맞추면 음성신호의 피치는 변화되기 때문에 창함수길이를 일정하게 할 수 없다.

Ⅲ. 앨리어징의 영향

음성신호에서 특히 에너지를 지배하는 유성음은 같은 음소가 일정시간 유지되어 stationary 하다고 생각할 수 있다. 이 구간 동안에는 주파수에 따른 스펙트럼 성분비가 일정하게 되어 앨리어징현상에 의한 영향이 일정하게 나타날 것이다. 앨리어징에 의해 영향을 주는 에너지값을 d(n)이라 하면 n-번째 샘플부터 창함수를 적용할 때 얻어지는 에너지값 Ed(n)은,

$$Ed(n)=E(n)+d(n) \tag{6}$$

이 된다. d(n)은 창함수의 차단주파수 이상 성분의 에너지를 나타내며, n에 따라 위상이 달라져 에너지 Ed(n)에 영향을 주는 값이 다르다. 여기서 창함수의 Gibbs 현상에 의한 에너지의 변동은 창함수선택을 적절히하여 앨리어징 효과에 비해 무시된다고 가정한다.

창함수의 스펙트럼은 그림 1-2와 같이 차단주파수 간격마다 국부봉우리를 가지며, 이 봉우리와 곱해지는 신호의 성분주파수들이 앨리어징효과를 주로 일으키게 된다. 따라서 앨리어징효과는 창함수길이의 역수인 차단주파수의 정수배 간격의 주파수 성분들에 의해 시간 n-에 따라 위상차의 영향을 받으면서 차단주파수 안의 에너지값에 영향을 주게된다. 앨리어징에 의한 영향을 측정하기 위해, 방형 창함수를 음성샘플 단위로 적용시키면서 그 에너지 Ed(.)을 구하면 다음과 같다.

$$\begin{aligned}
Ed(n) &= E(n)+d(n)\\
Ed(n-1) &= E(n)+d(n-1)\\
Ed(n-2) &= E(n)+d(n-2)\\
&\cdots\cdots\\
&\cdots\cdots\\
Ed(n-N+1) &= E(n)+d(n-N+1)
\end{aligned} \tag{7}$$

여기서 E(n)은 창함수가 통과대역안에서 갖는 에너지값이므로 시간지연에 무관한 에너지값이 되며, d(.)은 변수값에 따라 +-로 영향을 줄 수 있는 앨리어징영향에 따른 에너지 편차값이다.

Ⅳ. 창함수길이 영향의 제거

이제 식 7에 따른 에너지변화값 Ed(.)가 얻어질 때 여기서 앨리어징의 영향이 창함수길이 N에 따라 어떻게 나타나는지를 구할 수 있다. 일정길이의 창함수 구간내에서 검출된 에너지 Ed(n)들의 최대 및 최소값을 구하면,

$$\begin{aligned}
Emax(n) &= Max\{Ed(n), Ed(n-1), \cdots, Ed(n-N+1)\}\\
Emin(n) &= Min\{Ed(n), Ed(n-1), \cdots, Ed(n-N+1)\}
\end{aligned} \tag{8}$$

이 된다. 여기서 Emax{.}와 Emin{.} 함수는 주어진 변수에서 최대값과 최소값을 각각 의미하는 함수라 정의한다. 앨리어징에 의한 에너지변화분 d(.)이 mean을 중심으로 +-로 나타날 확률이 같다면, 이 때 원하는 에너지값 E(.)은

$$E(n)=[Ed(n)+Ed(n-1)+,\cdots,+Ed(n-N+1)]/N \quad (9)$$

또는 근사적으로,

$$E(n)=[Emax+Emin]/2 \quad (10)$$

가 된다. 그렇지만 에너지변화분의 mean이 존재하려면 창함수 길이가 확률적으로 다루어지도록 상당히 길어야 한다.

이것은 창함수 길이를 실제 발성가능한 최저피치인 N=20샘플(=2.5msec)로 하여, 발성가능한 최고피치인 200샘플 (=25msec) 동안에 식 10의 Emax와 Emin의 평균을 구하면 된다. 이상의 가정이 옳은 것인가를 보기 위해 그림 2에 23세의 여성화자가 발성한 숫자음 / 삼 / 에 대한 파형과 에너지변화도를 구하였다. 먼저 음성샘플을 창함수길이 N=20에 통과시킨 후에, 식 8의 최대와 최소값을 200샘플동안 10개씩 구하여 각각을 평균한 것을 Emax와 Emin 그림으로 나타내었다. 에너지변화도에서 앨리어징에 의한 영향은 Emax(.)과 Emin의 변화곡선 사이값이

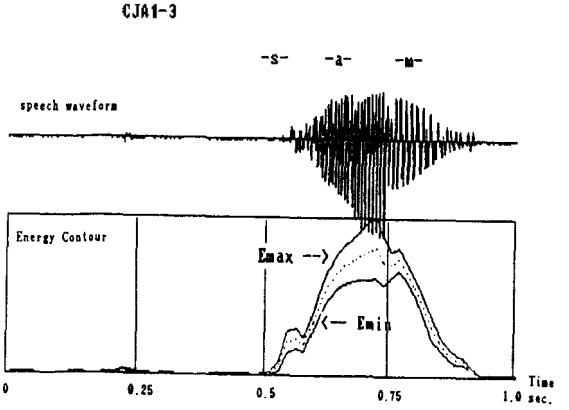

그림 2 음성신호 / 삼 / 의 에너지에 대한 앨리어징 효과
Aliasing effect for energy contour of speech / sam /

다. 앨리어징이 확률적인 mean 값을 갖게 된다는 것은 Emax(.)와 Emin(.)의 변화곡선이 닮은꼴로 변화됨을 보면 알 수 있다.

창함수길이 N-에서 식9를 계산하려면 Ed(.)를 평균진폭값으로 구하여도, 샘플당 20(=N)번의 덧셈과 최대와 최소값을 구하는데 2번의 비교연산이 소요된다. 이정도의 계산량은 전처리과정에서 많은 연산량이 아니지만 실용화하려면 계산량을 더 감소시킬 필요가 있다. Ed(n-1)에 관한 식을 다시 쓰면,

$$\begin{aligned} Ed(n-1) &= |s(n-)|+|s(n-2)| +..+ |s(n-N+1)|+|s(n-N)| \\ &= |s(n)| + |s(n-1)| +..+ |s(n-N+1) |- |s(n)| + |s(n-N)| \\ &= Ed(n)- |s(n)| + |s(n-N)| \end{aligned} \quad (11)$$

이 되어 방금전의 값 Ed(n)을 안다면 recursive하게 Ed(n-1)을 계산할 수 있다. 따라서 평균진폭 계산에는 음성샘플당 덧셈과 뺄셈 각 한번씩만 필요하게 된다.

V. 실험 및 결과

식 10의 시뮬레이션을 위해 IBM PC / AT를 사용하여 여기에 마이크입력이 가능하도록 12-비트 analog to digital converter를 인터페이스시켰다. 화자는 남성화자 2명과 여성화자 1명을 통해 다음 음성을 발성케하고 8KHz로 표본화하면서 저장시켰다.

발성1) 23세 남성화자 :
"인수네 꼬마는 천재소년을 좋아한다."

발성2) 32세 남성화자 :
"예수님은 천지창조의 교훈을 말씀하셨다."

발성3) 25세 여성화자 :
"감사합니다."

각 음성자료에 대해 방형창함수의 길이 N-을 32, 64, 그리고 128로 하고 앨리어징효과를 고려치 않고 평균진폭의 변화도를 나타낸 것이 각 결과 그림의 (c)에 나타내었다. 각 발성화자의 중심피치값은 발성 1=57, 발성 2=60, 그리고 발성 3=32샘플 정도를 차지하였다. 그림 3에서 그림 5까지는 결과그림이며 그림(c)를 보면 창함수 길이가 각 발성자의 중심피치 보다 짧은(예를 들어 남성화자의 경우 N=32) 경우에는 앨리어징현상 때문에 그림(a) 파형의 변화를 잘 나타내지 못하고 많은 국부봉우리가 발생하게 된다. 반면, 창함수길이가 피치보다 길게 (예를 들어 N=128)되면 창함수길이 내에서 앨리어징효과 보다 음성에너지 변화에 의한 스므딩현상이 더 크게 되어, 음소의 변화를 잘 나타내지 못하게 된다.

결과 그림의 (b)도는 앞에서 제안했던 앨리어징의 영향이 제거된 평균진폭의 변화도를 결과로 제시하였다. 결국 음성피치의 변화에 무관하게 음소의 변동을 파형의 진폭변화의 형태로 잘 대별하고 있음을 알 수 있다.

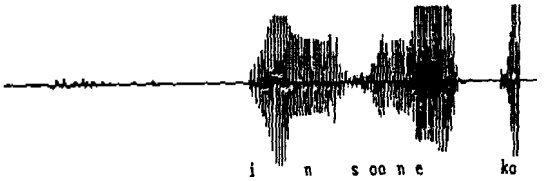

Fig. 3-1a speech waveform for / insoone koma / .

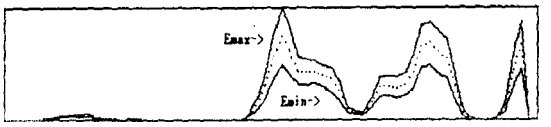

Fig. 3-1b the proposed energy contour.

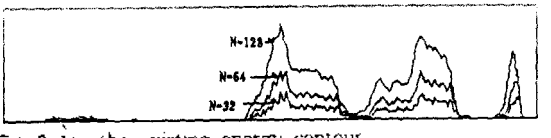

Fig. 3-1c the existing energy contour.

Fig. 3-2a speech waveform for / manun chunjesonyu-

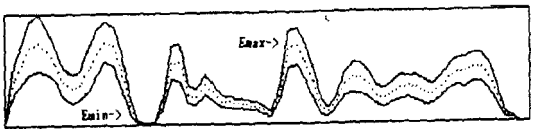

Fig. 3-2b the proposed energy contour.

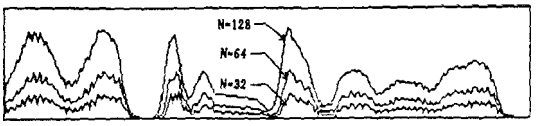

Fig. 3-2c the existing energy contours.

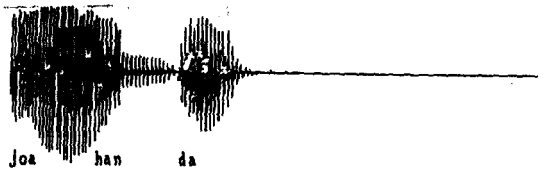

Fig. 3-3a speech waveform for / joahanda / .

Fig. 3-3b the proposed energy contour.

Fig. 3-3c the existing energy contours.

그림 3 "인수내 꼬마는 천재소년을 좋아한다" 음성에 대한 처리결과
Results for speech / Insoonae Komanun Chunjae Sonyunwl joahanda / .

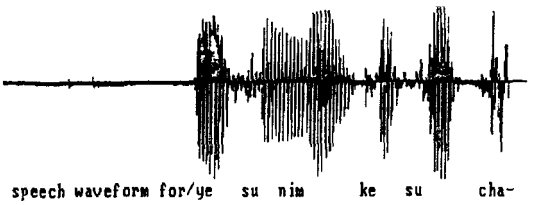

Fig. 4-1a speech waveform for / ye su min ke su cha-

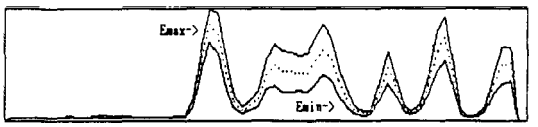

Fig. 4-1b the proposed energy contour

Fig. 4-1c the existing energy contour

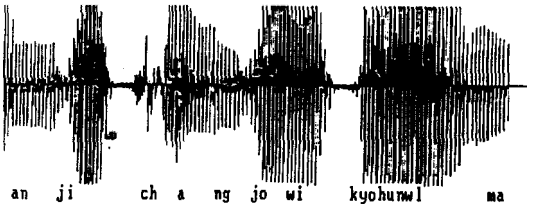

Fig. 4-2a an ji ch a ng jo wi kyohunwl ma

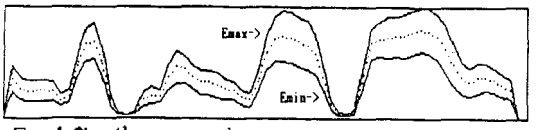

Fig. 4-2b the proposed energy contour

Fig. 4-2c the existing energy contours

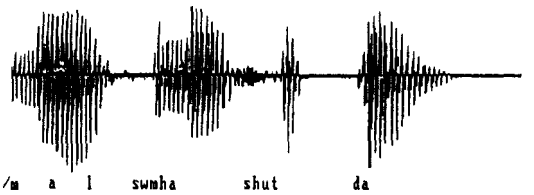

Fig. 4-3a / m a l swmha shut da

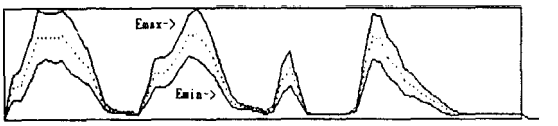

Fig. 4-3b the proposed energy contour

Fig. 4-3c the existing energy contour

그림 4 "예수님은 천지창조의 교훈을 말씀하셨다." 음성에 대한 처리결과.
Results for speech / Yesoonimeun Chungichangjowi Kyohwunwl malswmhasyetda.

Fig. 5-1a speech waveform / KAMSAHAMNIDA /.

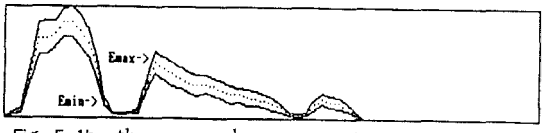

Fig. 5-1b the porposed energy contour.

Fig. 5-1c the existing energy contour.

그림 5 "감사합니다." 음성에 대한 처리결과.
Results for speech / Kamsahamnida /.

Ⅵ. 결 론

음성신호처리는 보통 전처리과정과 후처리과정으로 나누어 처리하고 있다. 전처리과정에서 주로 이용

하는 음성에너지 변화도는 분석시에 적용한 창함수 길이의 영향을 많이 받는다. 이것은 창함수가 저역통과 필터적인 측면에서 앨리어징효과에 의한 것이다. 앨리어징효과를 경감시키기 위한 한 방법으로는 차단특성이 민감한 창함수를 사용하면 된다. 이러한 방법은 계산과정이 복잡하여 전처리과정으로는 바람직하지 못하다. 다른 한 방법으로는 창함수길이를 음성신호의 피치에 일치시키면 된다. 그렇지만 이것은 창함수길이가 음성신호 마다 가변되어야 하고, 또한 에너지검출 전에 음성의 피치를 구해야하는 어려움이 따르게 된다.

따라서 본 논문에서는 앨리어징에 의한 영향을 에너지계산시에 파악하여 그 영향을 최소화하는 새로운 에너지검출법을 제안하였다. 이 방법은 창함수의 길이를 피치에 꼭 일치시킬 필요가 없으며, 창함수길이에 거의 영향을 받지 않는 장점이 있다. 또한 그 계산과정에서 음성샘플당 덧셈과 뺄셈이 한번씩, 그리고 비교논리를 두번씩만 실행하면 된다.

參 考 文 獻

1. L. R. Rabiner & R. W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978.
2. E. O. Brigham, The Fast Fourier Transform, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1974.
3. S. D. Stearns & R. A. David, Signal Processing Algorithms, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1988.
4. P. E. Papamichalis, Practical Speech Processing, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.
5. S. Chandra and W. C. Lin, "Linear Prediction with a Variable Analysis Frame Size, IEEE Trans., Acoust., Speech, Signal Processing, Vol. ASSP-25, No. 4, August 1977.
6. M. BAE, I. CHUNG, and S. ANN, "The Extraction of Nasal Sound by using G-Peak in Continued Speech", KIEE, Vol. 24, No. 2. pp. 92~97, March 1987.
7. 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.

▲배명진(정회원) 9권 1호 참조
▲안수길(정회원) 9권 1호 참조