2000년도 한국음향학회 학술발표대회 논문집 제19권 제1(s)호

시간축 스케일링 윈도우를 이용한 스펙트럼 누설 감소

이 희 원, 나 덕 수, 배 명 진
숭실대학교 정보통신공학과
156-743 서울시 동작구 상도5동 1-1
jx1004@hanmail.net

Spectrum Leakage Reduction using Time Scaling Window

HeeWon LEE, DuckSu NA, MyungJin BAE
Dept. of Information and Telecommunication Engr., Soongsil University
1-1 Sangdo-5Dong, Dongjak-Ku, Seoul 156-743, KOREA
jx1004@hanmail.net

요 약

음성 신호는 시간에 따라 변하지만 일정 구간에서는 특성이 변하지 않는다고 가정하여 윈도우를 취해 단구간 분석을 한다. 이 때 윈도우의 적용은 필수적이다. 하지만 단구간 분석을 위해서 사용되는 윈도우에 의해 생기는 누설에너지 때문에 음성신호의 스펙트럼 정보가 왜곡되어버린다.

본 논문에서는 스펙트럼 분석 시 발생되는 누설에너지를 최소화하는 방법을 제안하고자 한다. 음성신호에 고정된 크기의 rectangular window를 취한 후 처음 샘플과 차이가 가장 작은 샘플을 프레임 크기의 3/4인 지점에서부터 검색하여 최소인 부분까지 시간축 스케일링을 한 후 기존의 윈도우 크기과 같은 크기로 만든다. interpolation과 decimation을 조합하여 시간축 스케일링을 한다. 이렇게 윈도우가 적용된 신호를 처리 후 위의 역과정을 수행한다.

제안한 윈도우의 SNRseg는 rectangular window보다 평균 7.88dB 낮고, kaiser window보다 평균 1.65dB 높았다. 또한 제안한 윈도우의 SD는 rectangular window보다 평균 1.73dB^2 낮았다.

1. 서 론

음성신호는 시간에 따라 변한다. 음성을 느리게 발성하면 성도(vocal tract)의 모양과 여기(excitation) 유형이 200ms 동안 변하지 않는다. 그러나 대부분 음소의 지속시간이 약 80ms 정도이기 때문에 더 빨리 변한다. 조음기관과 F0의 변화는 이전 피치 주기와는 다른 피치주기를 생성한다. 그럼에도 불구하고 음성 분석은 보통 신호의 성질이 시간에 따라 비교적 천천히 변한다고 가정한다. 이러한 가정으로 인해 프레임 단위로 단구간 분석이 가능하게 된다[10]. 단구간 분석을 위해서는 윈도우 적용이 필요하다. 그러나 윈도우 적용시 누설현상이 발생하여 스펙트럼을 왜곡시키게 된다. 창함수의 종류와 길이에 따른 스펙트럼 왜곡을 줄이기 위한 많은 연구가 진행되었지만 원래의 스펙트럼에 근사 시키는 것은 몇 가지 어려운 문제를 내포하고 있다[1]-[3].

본 논문에서는 고정된 크기의 윈도우를 적용한 후 스펙트럼 누설이 최소인 지점을 찾아 나머지 부분은 무시하고 누설이 최소인 지점까지의 샘플들을 기존의 윈도우 크기에 맞게 시간축 스케일링을 하는 새로운 방법을 제안하였다.

먼저 2장에서는 스펙트럼 누설현상에 대해, 3장에서는 시간축 스케일링을 이용한 윈도우, 4장에서는 제안한 방법에 대한 실험 및 결과에 대해 설명하고, 5장에서 결론을 맺는다.

2. 스펙트럼 누설현상

윈도우 함수 구간 NT내에 정수배의 주기를 갖지 않는 신호에 대한 스펙트럼은 원래의 신호주파수 근처에 대칭으로 분산되어 나타나는 현상을 leakage현상이라고 한다. 이 현상은 Discrete Fourier Transform이 구간 NT에서 모든 성분이 주기적이라는 가정하에 Fourier Series 전개를 한 것이기 때문이다[4][5].

주기성을 갖는 유성음에 대해 일정위치에서 유한구간을 windowing하여 구한 스펙트럼은 윈도우를 고려하지 않은 유성음의 스펙트럼에 윈도우의 스펙트럼이 convolution된 결과가 된다.

$s(\cdot)$, $f(\cdot)$, $w(\cdot)$을 각각 음성신호 및 윈도우 함수가 적용된 음성신호 그리고 윈도우 함수라 할 때 시간영역에서는

$$f(n) = s(n)w(n-k) \tag{2.1}$$

으로 표시되고 주파수영역에서는

$$F(e^{j\Omega T}) = S(e^{j\Omega T}) * W(e^{j\Omega T}) \tag{2.2}$$

으로 표현할 수 있다. 이상적인 윈도우 함수라면 주파수 영역에서 convolution 되더라도 원래의 음성 스펙트럼에 근사 되도록 나타날 것이다. rectangular window일 경우 좁은 대역폭의 main lobe에 의해 하모닉스의 첨예도가 커지나 큰 side lobe에 의해 누설현상이 두드러져 스펙트럼의 왜곡이 발생한다. 이때 윈도우 함수의 길이가 피치의 정수 배로 되면 누설현상이 최소화되어 스펙트럼의 첨예도가 더욱 커지고 원래의 스펙트럼에 근사 되어 나타나게 된다. 그렇지만 윈도우의 길이를 피치와 동기시키기 위해서는 피치를 구하는 일이 선행되어야 하는 어려움이 따르게 된다. 따라서, 윈도우 함수의 길이에 의한 영향을 줄이려는 연구보다는 윈도우의 모서리에 의한 누설현상을 줄일 수 있는 윈도우를 개발하려고 노력했다. 지금까지 연구 개발된 윈도우 함수로는 rectangular, triangle, hanning, hamming, kaiser, blackman 윈도우 등이 있으며 차단특성이 좋은 윈도우 함수일수록 계산이 복잡하여 주로 hamming window를 많이 사용하고 있다[6].

윈도우의 모서리에 의한 영향은 좋은 윈도우를 선택한다고 완전히 제거되지는 않는다. 창함수의 적용위치에 따라 신호의 불연속이 크고 작게 나타날 수 있다. 따라서, 원래의 스펙트럼에 근접시키기 위해서는 창함수의 적용위치와 크기를 조절할 필요가 있다.

3. 시간축 스케일링을 이용한 윈도우

시간축 스케일링을 위해 decimation과 interpolation을 사용하였다.

먼저 decimation에 대해 알아보자. 샘플링 율을 factor M로 줄이고 싶을 때 $x_a(t)$의 샘플을 주기가 $T' = MT$이 되도록 한다. 즉,

$$y(n) = x_a(nT') = x_a(nTM) \tag{3.1}$$

이것을 다시 표현하면 다음과 같다.

$$y(n) = x(Mn), \quad -\infty < n < \infty \tag{3.2}$$

즉, $y(n)$은 M개의 샘플 중 하나의 샘플만 주기적으로 남기면 얻어진다. 샘플링 이론에서 $1/T = 2F_N$이면 $y(n)$은 원래 아날로그 신호를 표현하기에 적당하다는 것을 주의해야 한다. $x(n)$과 $y(n)$의 Fourier transform은 다음과 같은 관계가 있다.

$$Y(e^{j\Omega T}) = \frac{1}{M}\sum_{k=0}^{M-1} X(e^{\frac{j(\Omega T - 2\pi k)}{M}}) \tag{3.3}$$

식 (3.3)에서 $X(e^{j\Omega T})$의 상들 사이에 오버랩(overlap)을 없애기 위해 $1/T = 2F_N$를 만족해야 한다. 만약 샘플링율이 나이퀴스트율보다 낮으면($1/T < 2F_N$) aliasing 이 발생한다. 이것을 막기 위해서 차단 주파수가 π/T인 디지털 저역통과 필터를 통과 시킨다. 일반적인 decimation 시스템을 그림 3-1.에 나타내었다.

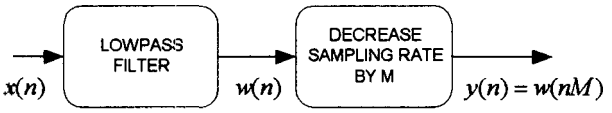

그림 3-1. Decimation의 블럭도

다음으로 interpolation에 대해 알아보자. 아날로그 파형의 다음과 같은 $x(n) = x_a(nT)$ 샘플을 가지고 있다고 하자. 정수 factor L로 샘플링율을 높히려면 $T' = T/L$의 주기를 가지고 새로운 시퀀스를 만들어야 한다. 즉, 식 (3.4)와 같이 표현된다.

$$y(n) = x_a(nT') = x_a(nT/L) \tag{3.4}$$

다음과 같은 시퀀스를 생각해 보자.

$$\begin{aligned} v(n) &= x\left(\frac{n}{L}\right) \quad n = 0, \pm L, \pm 2L, \cdots \\ &= 0 \qquad otherwise \end{aligned} \tag{3.5}$$

$v(n)$의 Fourier transform은 다음과 같다.

$$\begin{aligned} V(e^{j\Omega T}) &= X(e^{j\Omega TL}) \\ &= X(e^{j\Omega T}) \end{aligned} \tag{3.6}$$

$V(e^{j\Omega T})$은 $2\pi/T = 2\pi/(LT')$의 주기를 갖는다. $v(n)$에서 $y(n) = x_a(nT')$인 시퀀스를 얻기 위해 식(3.7)을 확실하게 해야한다.

$$Y(e^{j\Omega T}) = \frac{1}{T} X_a(j\Omega) \quad -\frac{\pi}{T} \le \Omega \le \frac{\pi}{T} \tag{3.7}$$

식(3.8)과 같이 가정하면 $V_a(j\Omega)$에서 $X_a(j\Omega)$의 이미지

를 제거해야 한다.

$$X(e^{j\Omega T}) = \frac{1}{T} X_a(j\Omega) \quad -\frac{\pi}{T} \le \Omega \le \frac{\pi}{T} \quad (3.8)$$

이렇게 하기 위해서 $-\frac{\pi}{T} \le \Omega \le \frac{\pi}{T}$ 범위의 모든 주파수 성분을 저역통과필터를 사용하여 뽑아낸다. 그림 3-2.는 일반적인 interpolation 시스템의 블록도이다[2].

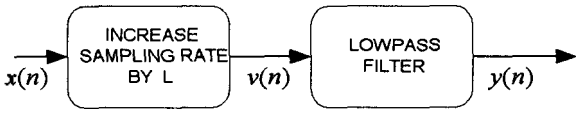

그림 3-2. Interpolation의 블럭도

4. 실험 및 결과

제안한 방법을 실험하기 위해서 먼저 IBM PC(233 MHz)에 마이크 입력이 가능한 A/D 변환기를 인터페이스 하였다. 음성시료는 남자와 여자가 연구실 환경(30dB의 SNR)에서 발성한 음성을 8kHz로 표본화하고 16bit로 양자화하여 사용하였다. 발성한 문장은 다음과 같다.

발성1) "인수네 꼬마는 천재소년을 좋아한다."
발성2) "예수님께서 천지창조의 교훈을 말씀하셨다."
발성3) 핸드폰 음성 사서함에 녹음하는 대화체 음성
발성4) 일기예보 아나운서 음성 시료

그림 4-1.은 제안한 방법에 대한 순서도이다. 먼저 단구간 분석을 위해 고정된 크기(256 샘플)의 rectangular window를 취한다. 스펙트럼 누설을 최소화하기 위해 프레임의 첫 번째 샘플과 오차가 최소가 되는 샘플을 프레임 크기의 3/4되는 지점부터 프레임의 끝까지 검색한다. 여기서 얻어지는 인덱스에서 interpolation factor L로 나누어 소숫점을 무시하고 decimation factor M을 구하였다. 여기서 계산 시간을 단축하기 위해 interpolation factor는 16으로 고정했다. 소숫점을 무시해서 생기는 오차를 보정하기 위해 M과 L을 곱한 값을 인덱스로 재조정하였다. 이 과정을 통하여 윈도우 크기는 고정되고 스펙트럼 누설을 최소화하는 윈도우를 적용한다. 그 후 원하는 처리를 한다. 본 논문에서는 제안한 방법을 평가하기 위해 제안한 윈도우를 적용한 음성신호를 LPC 분석 후 합성하였다. 처리가 끝나면 전단에서 decimation factor로 사용되었던 M으로 interpolation을 수행하고, interpolation factor로 사용되었던 L로 decimation을 수행한다. 결과적으로 원래 신호와 같은 샘플링율을 갖는 신호가 생성된다.

제안한 방법을 평가하기 위해 원 신호와 rectangular window를 적용한 신호, kaiser window를 적용한 신호, 제안한 윈도우를 적용한 신호의 LPC를 구한 후 LPC 계수를 8비트로 코딩하였다. 그것을 이용하여 음성을 합성 한 후 SNRseg과 Spectrum Distortion(SD)을 계산하였다.

먼저 SNRseg을 계산하기 위해 한 프레임 안에서 다음과 같이 원 신호 $s(n)$와 합성음 $\hat{s}(n)$의 차를 계산한다.

$$e(n) = s(n) - \hat{s}(n) \quad \text{식(4.1)}$$

그 후 각 프레임에 대해 원 신호의 에너지와 오차 신호의 에너지를 계산하여 SNR(m)을 구한 후 마지막으로 다음 식을 이용하여 SNRseg를 계산한다.

$$SNRseg = \frac{1}{M} \sum_{m=1}^{M} SNR(m) \quad \text{식(4.2)}$$

각 발성에 대한 결과를 표 4.1에 나타내었다. 제안한 윈도우의 SNRseg는 rectangular window보다 평균 7.88dB 낮고, kaiser window보다 평균 1.65dB 높은 것을 볼 수 있다.

표 4-1. SNRseg 비교(dB)

방법 / 발성	rectangular	kaiser	제안한 window
발성 1	38.52	28.26	30.20
발성 2	40.69	30.21	31.89
발성 3	41.95	35.27	37.22
발성 4	40.87	32.11	33.15

kaiser window를 적용한 신호의 스펙트럼 $A(n)$과 retangular와 제안한 윈도우를 적용한 스펙트럼 $\hat{A}(n)$을 계산한 후, 한 프레임에 대한 sd는 식(4.3)에 의해서 구한다.

$$sd = \frac{1}{N/2} \sum_{k=1}^{N/2} \left[10 \log_{10} \frac{|\hat{A}(k)|^2}{|A(k)|^2} \right]^2 \quad \text{식(4.3)}$$

전체적인 SD를 구하기 위해 식(4.4)를 사용한다.

$$SD = \frac{1}{M} \sum_{n=1}^{M} sd(n) \quad \text{식(4.4)}$$

각 발성에 대한 SD를 표 4-2.에 나타내었다. 제안한 윈도우의 SD는 rectangular window보다 평균 1.73dB^2 낮음을 볼 수 있다.

표 4-2. SD(Spectrum Distortion) 비교(dB^2)

방법 / 발성	rectangular	제안한 window
발성 1	2.8750	1.0021
발성 2	3.4571	1.8532
발성 3	3.0282	1.3874
발성 4	2.9987	1.1865

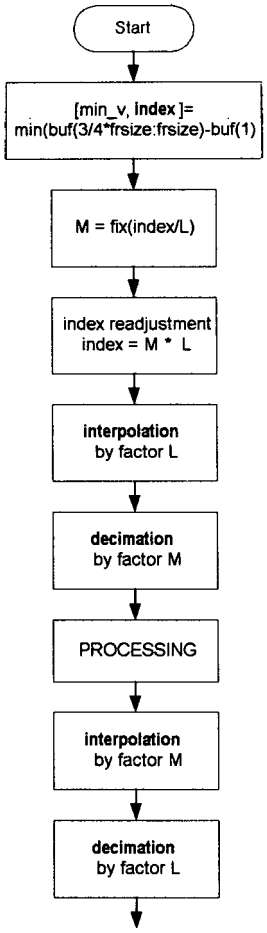

그림 4-1. 시간축 스케일링 윈도우를 이용한 처리 방법

5. 결 론

음성 신호는 시간에 따라 변하지만 일정 구간에서는 특성이 변하지 않는다고 가정하여 윈도우를 취해 단구간 분석을 한다. 이 때 윈도우의 적용은 필수적이다. 하지만 단구간 분석을 위해서 사용되는 윈도우에 의해 생기는 누설에너지 때문에 음성신호의 스펙트럼 정보가 왜곡되어버린다.

먼저 단구간 분석을 위해 고정된 크기(256 샘플)의 rectangular window를 취한다. 스펙트럼 누설을 최소화하기 위해 프레임의 첫 번째 샘플과 오차가 최소가 되는 샘플을 프레임 크기의 3/4되는 지점부터 프레임의 끝까지 검색한다. 기존의 윈도우 크기로 시간축 스케링을 하기 위해 interpolation과 decimation을 수행한다. 시간축 스케일링된 신호를 가지고 원하는 처리를 한 후 다시 interpolation과 decimation을 수행하여 역 스케일링을 한다. 제안한 방법을 평가하기 위해 rectangular, kaiser, 제안한 윈도우를 사용하여 LPC 합성 후 SNRseg와 SD를 계산한 결과 제안한 윈도우의 SNRseg는 rectangular window보다 평균 7.88dB 낮고, kaiser window보다 평균 1.65dB 높았다. 또한 제안한 윈도우의 SD는 rectangular window보다 평균 1.73dB^2 낮았다.

6. 참 고 문 헌

[1] Panos E.Papamichalis, Practical Approaches to Speech Coding, Prentice-hall inc., Englewood cliffs, N.J., 1987.

[2] L.R. Rabiner and R.W. Schafer, "Digital processing of Speech Signals Englewood Cliffs", New Jersey : Prentice-Hall, 1978.

[3] Thomas W. Parsons, Voice and Speech Processing, Mcgraw-Hill, 1987.

[4] Nasir Ahmed, T.Natarajan, Discrete-time Signal and System, Prentice-Hall, 1983.

[5] R.W.Schafer and L.R.Rabiner,"Digital representation of Speech Signal", Proc.IEEE, vol.63, pp.662-667, Apr. 1975.

[6] P.D.Welch, "The use of the Fast Fourier Transform for the Estimation of Power Spectra", IEEE Trans. Audio and Electro Acoust., vol. pp.70-73, Au-15, 1967.

[7] 강동규, 이을재, 배명진, 안수길, "스펙트럼 누설현상을 이용한 중심피치 검출", 대한전자공학회 하기 학술발표회 논문집, 제 13권, 제1호, pp.665-668, 1990.

[8] 배명진, 안수길, "음성 에너지 계산에서 창함수-길이 개선에 관한 연구", 한국음향학회, 제9권, 2호, pp.34-41, 1990년 4월.

[9] 강동규, 박찬수, 배명진, 안수길, "스펙트럼 비교에 의한 음성신호의 창함수 적용싯점 검출", 음성통신 및 신호처리 Workshop 논문집, pp.88-92, 1990년 8월.

[10] Speech Communication, IEEE PRESS, 2000.