2000년도 한국음향학회 학술발표대회 논문집 제19권 제1(s)호

Shifted Window를 이용한 음성신호의 분석에 관한 연구

강 은 영, 민 소 연, 배 명 진

숭실대학교 정보통신공학과
156-743 서울시 동작구 상도동 1-1

On a Study of Analysis Using Shifted Window in the Speech Signal

EunYoung kang, SoYeon Min, MyungJin Bae
Dept. of Information and telecommunication Engr., Soongsil University
1-1 Sangdo-5Dong, Dongjak-Ku, Seoul 156-743, KOREA
cjjung@ifcom.soongsil.ac.kr

요 약

음성신호처리에서 스펙트럼 분석은 매우 중요하다. 하지만 스펙트럼 분석을 위해서 사용되는 윈도우에 의해 생기는 누설에너지 때문에 음성신호의 스펙트럼 정보가 왜곡된다.

본 논문에서는 스펙트럼 분석 시 발생되는 창함수 사용에 의해 생기는 누설에너지를 최소화하기 위한 새로운 창함수를 제안하고자 한다. 그 형태는 전체 창함수크기의 반을 방형창으로 나머지 반을 해밍창으로 하고 창의 처음 부분은 ±20표본에서 영점을 찾아주는 것이다. 이 창함수의 특징은 신호분석에 있어서 왜곡은 크지만 그 형태에 있어서 가장 이상적인 방형창함수의 장점과 side lobe가 작아 비교적 왜곡이 적은 해밍창함수의 장점을 취한것이라 하겠다.

실제 음성 신호에의 적용에 있어서 방형창과 해밍창의 적용비는 신호의 종류 및 용도에 따라 달리 할수 있다. 제안한 창함수는 해밍창함수 보다는 좁은 main lobe 특성으로 음성신호의 단구간 스펙트럼 분석시 음성의 빠른 변화특성을 적절히 보여줄 수 있고 방형창보다는 side lobe의 영향을 줄일수 있다.

1. 서 론

음성신호처리에서 음성신호를 주파수영역에서 처리 할 수 있으므로 시간영역에서 처리가 어려웠던 많은 문제들이 해결되었다. 그러나, FFT 수행시 창함수 적용 이 불가피하고 이로 인해 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]

음성신호에 창함수를 적용할 때 창함수의 길이가 음성신호의 피치보다 짧으면 음성신호가 가지고 있는 특성을 전체적으로 반영 할 수 없고 길게 되면 스무딩현상에 의해 음성신호가 갖는 음소의 변화 특성이 잘 나타나지 않게 된다[4].

본 논문에서는 창함수의 종류와 길이에 따른 문제점들을 해결하는 방안으로 창함수의 형태를 변형시켜 여러 창함수들의 단점을 보완하고 창의 시작 부분에서 영점을 찾아줌으로써 불연속 왜곡을 최소화하려 한다.

2. 스펙트럼 누설현상

창함수 구간 내에 정수 배의 주기를 갖지 않는 신호에 대한 스펙트럼은 원래의 신호주파수 근처에 대칭으로 분산되어 나타나는 현상을 leakage 현상이라고 한다. 이 현상은 Discrete Fourier Transform이 창함수 구간에서 모든 성분이 주기적이라는 가정 하에 Fourier Series 전개를 한 것이기 때문이다[4,5].

음성신호에서 주기성을 갖는 유성음에 대해 일정위치에서 유한구간 windowing하여 구한 스펙트럼은 창함수를 고려하지 않은 유성음의 스펙트럼에 창함수의 스펙트럼 convolution된 결과가 된다. $s(\cdot)$, $f(\cdot)$, $w(\cdot)$을 각각 음성신호 및 창함수가 적용된 음성신호 그리고 창함수라 할 때 시간영역에서는

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

으로 표시되고 주파수영역에서는

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

으로 표현할 수 있다. 이상적인 창함수라면 주파수 영역에서 convolution 되더라도 원래의 음성 스펙트럼에 근사 되도록 나타날 것이다. 방형 창함수일 경우 좁은 대역폭의 main lobe에 의해 하모닉스의 첨예도가 커지나 큰 side lobe에 의해 누설현상이 두드러져 스펙트럼의 왜곡이 발생한다. 이때 창함수의 길이가 피치의 정수 배로 되면 누설현상이 최소화되어 스펙트럼의 첨예도가 더욱 커지고 원래의 스펙트럼에 근사 되어 나타나게 된다. 그렇지만 창함수의 길이를 피치와 동기시키기 위해서는 피치를 구하는 일이 선행되어야 하는 어려움이 따른다. 따라서, 창함수의 길이에 의한 영향을 줄이려는 연구보다는 창함수의 모서리에 의한 누설현상을 줄일 수 있는 창함수를 개발하려고 노력했다. 지금까지 연구 개발된 창함수로는 방형창, 삼각형창, 해닝창, 해밍창, 카이저창, 블랙맨창 등이 있으며 차단특성이 좋은 창함수일수록 계산이 복잡하여 주로 해밍 창함수를 많이 사용하고 있다[6].

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

3. 기존의 창함수의 종류 및 특징

기존의 창함수로는 방형창, 삼각형창, 해닝창, 해밍창, 카이저창, 블랙맨창 등이 있으며 차단특성이 좋은 창함수일수록 계산이 복잡하다.

이상적인 창의 주파수응답특성은 매우 좁은 main lobe를 갖고 side lobe들이 없는 것이다.

하지만 현실적으로는 불가능하기 때문에 어떤 종류의 신호를 어떤 목적으로 분석하느냐에 따라 적절한 창함수를 선택하여 사용하게 된다.

기존의 창함수들은 다음과 같이 정의되어진다.

방형창 :

$$w(n) = \begin{cases} 1 & ; \ 0 \le n \le N-1 \\ 0 & ; \ otherwise \end{cases}$$

삼각형창 :

$$w(n) = \begin{cases} \dfrac{2n}{N-1} & ; \ 0 \le n \le \dfrac{N-1}{2} \\ 2 - \dfrac{2n}{N-1} & ; \ \dfrac{N-1}{2} \le n \le N-1 \\ 0 & ; \ otherwise \end{cases}$$

해밍창 :

$$w(n) = \begin{cases} 0.54 - 0.46\cos(2\pi \dfrac{n}{N-1}) & ; \ 0 \le n \le N-1 \\ 0 & ; \ otherwise \end{cases}$$

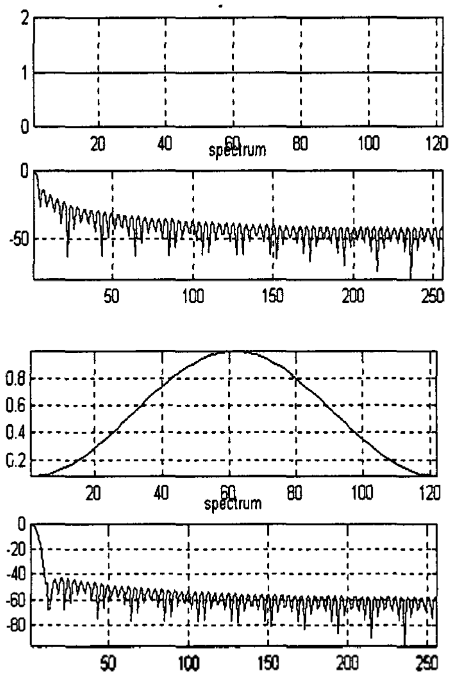
해닝창 :

$$w(n) = \begin{cases} 0.5 - 0.5\cos(2\pi \dfrac{n}{N-1}) & ; \ 0 \le n \le N-1 \\ 0 & ; \ otherwise \end{cases}$$

블랙맨창 :

$$w(n) = \begin{cases} 0.42 - 0.5\cos(2\pi \dfrac{n}{N-1}) + 0.08\cos(2\pi \dfrac{2n}{N-1}) \\ \qquad ; \ 0 \le n \le N-1 \\ 0 \qquad ; \ otherwise \end{cases}$$

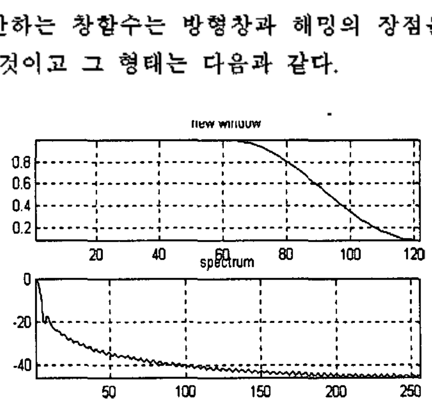
음성 신호 분석에서 자주 쓰이는 것은 해밍창이다. 가장 이상적인 방형창은 side lobe의 영향 때문에 신호의 스펙트럼 왜곡이 심하다. 반면 해밍창은 스펙트럼 왜곡의 원인인 불연속점을 완화시켜줌으로써 side lobe의 영향을 많이 줄인다. 하지만 main lobe폭이 넓어서 음성신호의 주파수 분해능이 작아 음성의 빠른 변화특성을 분석하기가 어렵다. 방형창과 해밍창의 주파수 응답특성이 아래와 같다.[4]



그림<3-1> 기존의창함수들과 스펙트럼

4. 제안하는 창함수의 형태

제안하는 창함수는 방형창과 해밍의 장점을 취하는 것이고 그 형태는 다음과 같다.



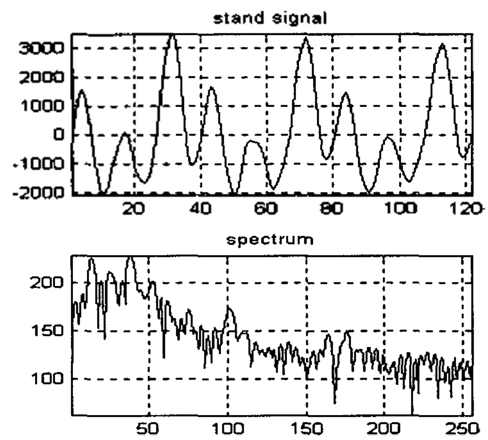
그림<4-1> 제안하는 창함수와 스펙트럼

5. 실험 및 결과

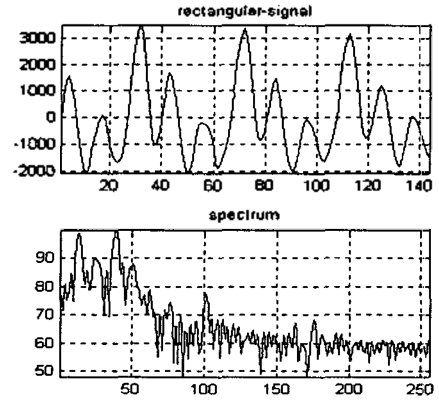
제안한 방법을 실험하기 위해서 먼저 IBM PC(233 MHz)에 마이크 입력이 가능한 A/D 변환기를 인터페이스 하였다. 음성시료는 여성이 연구실 환경(30dB의 SNR)에서 발성한 음성을 8kHz로 표본화하고 16bit로 양자화 하여 사용하였다. 발성한 문장은 다음과 같다.

발성1) "음성 신호 처리"
발성2) "창공을 나는 새"
발성3) "천지창조의 교훈"

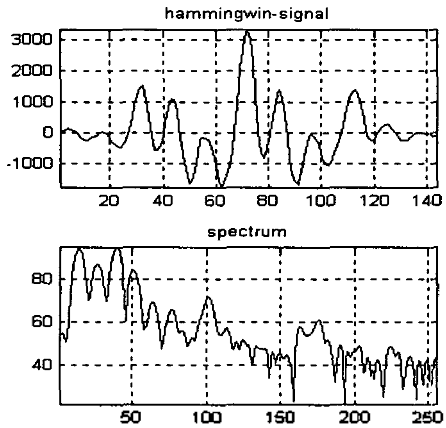
아래 그림들은 제안한 창함수를 실제 음성신호에 적용하고 기존의 창함수에 적용되었을 때와 비교한 것이다. 그림<5-1>은 음성신호의 정수배 피치에 대한 스펙트럼이고 이를 비교기준으로 삼았다. 각 창함수를 적용한 음성신호는 정수배 피치이상으로 하여 스펙트럼의 왜곡을 유도하였다. 그림<5-2>를 보면 방형창을 적용한 음성신호의 스펙트럼 특성이 가장 첨예한 특성을 보이고 있다는 것을 알 수 있다. 그러나 큰 side lobe 때문에 왜곡이 크다. 그림<5-3>은 해밍을 적용한 음성신호와 스펙트럼이고 왜곡이 적지만 주파수 분해능이 너무 떨어져 음성의 빠른 특성을 제대로 분석하기 어렵다. 그림<5-4>에서처럼 본 논문에서 제안한 창함수를 적용했을때는 왜곡을 줄이면서 음성신호의 주파수 분해능을 높이므로써 스펙트럼의 첨예한 특성을 볼수 있었다.



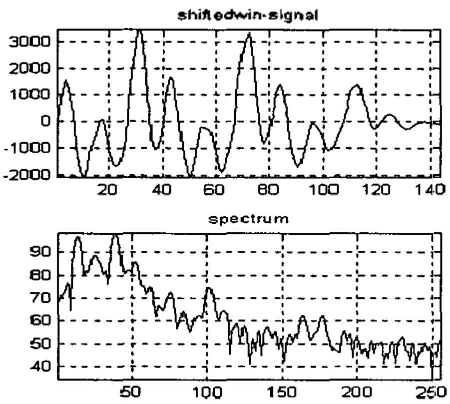
그림<5-1> 정수배피치 음성신호와 스펙트럼



그림<5-2> 방형창을 적용한 신호와 스펙트럼



그림<5-3> 해밍창을 적용한 신호와 스펙트럼



그림<5-4> 제안창을 적용한 신호와 스펙트럼

6. 결 론

음성신호를 주파수영역으로 처리할 때 창함수 적용이 불가피 하고 이로 인해 누설현상이 발생하여 스펙트럼 정보가 왜곡된다. 이러한 왜곡을 줄이기 위해서는 창함수의 길이가 유성음에서 나타나는 피치주기의 정수배가 되어야 한다. 하지만 음성의 피치 주기는 계속 변하여 찾기가 그리 쉽지는 않다. 분석하고자 하는것에 따라 적절한 창함수를 쓰는 것이 중요하다. 따라서 본 논문에서는 가장 이상적으로 음성의 특징을 반영해주는 방형창 함수와 왜곡이 적은 해밍창 함수의 장점을 갖는 창함수를 제안하였다.

창함수 전체 길이중 앞쪽 반은 방형창으로 또 뒤쪽 반은 해밍창으로 하여 그 모양은 그림 <4-1>과 같다. 실제 음성에 적용해 본 결과 해밍보다는 첨예한 신호 특성을 얻을 수 있었고 방형창보다는 적은 왜곡특성을 보였다.

7. 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] A. M. Kondoz, *Digital Speech*, 1994

[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, 이올재, 배명진, 안수길, "스펙트럼 누설현상을 이용한 중심피치 검출", 대한전자공학회 하기학술발표회 논문집, 제 13권, 제1호, pp.665-668, 1990.

[8] 배명진, 안수길, "음성 에너지 계산에서 창함수-길이 개선에 관한 연구", 한국음향학회, 제 9권, 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.