

다해상도 신호해석을 이용한 음성개선 방식 비교

한미경, 식종원, 배건성
경북대학교 전자·전기 공학부

Comparison of Speech Enhancement Methods Using Multiresolutional Signal Analysis

Mi Kyung Han, Jong Won Seok, Keun Sung Bae
School of Electronic and Electrical Engineering, Kyungpook National University

요 약

본 논문에서는 최근들어 널리 연구되고 있는 다해상도 신호해석 방법인 웨이브렛 변환, 웨이브렛 패킷, 그리고 코사인 패킷 알고리즘을 음성개선에 이용하여 각각의 성능을 비교하였으며, 또한 이를 기존의 스펙트럼 차감법의 성능과 비교 분석 하였다. 성능비교의 척도로는 SNR과 켈스트랄 거리를 이용하였다.

실험결과 SNR면에서는 코사인 패킷이 가장 좋은 결과를 보였다. 그리고 켈스트랄 거리의 경우 코사인 패킷과 웨이브렛 패킷이 훨씬 나은 결과를 보였으며 주관적인 청취결과 역시 코사인 패킷이 가장 좋은 결과를 보였고, 기존의 스펙트럼 차감법은 musical noise의 영향으로 인해 상대적으로 다른 방식에 비해 합성음의 음질이 많이 떨어짐을 확인할 수 있었다.

1. 서 론

음성통신 분야에 사용되는 기본적인 기술에 관련된 연구의 대부분은 주로 주변잡음이 없거나 무시할 수 있을 정도의 양질의 음성을 대상으로 연구가 진행되어 왔다. 이러한 잡음환경의 영향을 고려하지 않은 음성신호 처리 시스템의 성능은 실제 환경에서 급격히 감소하게 된다. 따라서 음성신호에 포함된 잡음을 제거하는 음성개선(speech enhancement) 기술에 대한 연구 및 개발이 꾸준히 이루어지고 있다[1-3].

음성개선을 위해 제안된 여러 방법중 가장 널리 사용되고 있는 알고리즘의 하나가 스펙트럼 차감법(spectral

substraction)[1] 이다. 이 방식은 단구간의 스펙트럼 추정(short-time spectrum estimation)에 기초를 둔 방법으로, 추정된 잡음의 스펙트럼을 입력 신호에서 빼주어서 잡음이 제거된 스펙트럼에서 음성을 재합성하는 방법으로서, 미리 잡음이나 원하는 음성의 통계적 특성이 잘 알려져 있을 때 적합한 음성개선 방법으로 알려져 있다.

본 논문에서는 단구간 푸리에 변환의 단점을 극복하고자 널리 연구되고 있는 다해상도 신호해석 방법인 웨이브렛 변환, 웨이브렛 패킷, 그리고 코사인 패킷 알고리즘[4-7]을 음성개선에 이용하여 각각의 성능을 비교하였으며, 또한 이를 기존의 스펙트럼 차감법의 성능과 비교 분석 하였다.

2. 스펙트럼 차감법을 이용한 음성개선

스펙트럼 차감법은 주변잡음에 의해 손상된 음성 스펙트럼에서 잡음 스펙트럼의 크기 성분만을 제거하는 방법이다. 이는 주변잡음이 음성에 산술적으로 더해진다는 가정과 음성을 인지하는 청각의 특성은 음성의 주파수 성분별 위상 정보보다는 크기 정보에 더 많이 영향을 받는다는 결과에 기초한다.

스펙트럼 차감법은 배경잡음의 스펙트럼 형태를 미리 알고 있거나, 잡음의 스펙트럼을 추정하기에 충분한 목음구간이 주어져야 한다. 배경잡음은 최소한 부분적으로 안정(stationary)한 특성을 가져야 하며, 통계적 특성이 서서히 변하는 환경에서는 음성이 존재하는 구간과 잡음만이 존재하는 구간을 검출할 수 있는 방법이 필요하다. 하지만 이 방법은 실제잡음과 그 평균값과의 오차로

인하여 추정된 음성과 원음성 사이에 스펙트럼 에러가 생겨 잡음저리과정에서 musical noise가 발생하며, 경우에 따라서는 잡음을 제거한 음성이 잡음음성보다 음질이 떨어지는 결과를 초래할 수 있는 단점을 가지고 있다. 아래의 그림 1은 전체적인 스펙트럼 차감법의 과정을 보여주고 있다.

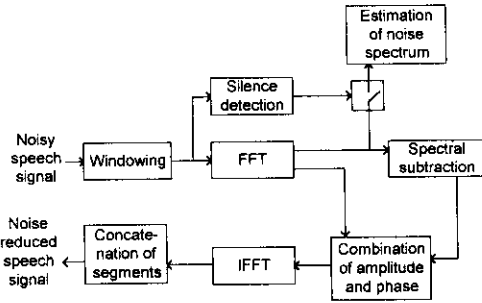


그림 1. 스펙트럼 차감법의 블록도

3. 다해상도 신호해석을 이용한 음성개선

3.1 웨이브렛 변환

웨이브렛 이론은 응용수학에서 처음 소개된 후 최근 컴퓨터 비전 분야에서 연구되어 온 다중해상도 표현과 연관성이 있음이 밝혀졌으며 이산 웨이브렛 변환 이론은 이산신호의 subband 분해방법과도 연관성이 존재한다[4.5]. 연속 웨이브렛 변환은 다음과 같이 정의된다.

$$CWT(t, a) = \int f(t) \varphi^*_{a,\tau} dt \quad (1)$$

$$\varphi_{a,\tau} = \frac{1}{\sqrt{a}} \varphi\left(\frac{t-\tau}{a}\right) \quad (2)$$

식 (2)에서 $\varphi(t)$ 는 원형(prototype) 웨이브렛이며 $\varphi_{a,\tau}$ 는 $\varphi(t)$ 를 이동(shift)과 확장(dilation)하여 구해진다. 즉, 짧은 지속시간을 갖는 고주파신호에 대해서는 짧은 창함수를 사용하고 긴 지속시간을 갖는 저주파신호에 대해서는 긴 창함수를 이용하는 결과가 되어 주파수 영역에 따른 다중해상도를 갖게 된다. 식 (1)의 연속 웨이브렛 변환은 시간과 scale factor가 연속인 값을 갖기 때문에 이를 실제 이용 가능한 이산적인 형태인 dyadic 웨이브렛 변환으로 표시하면 다음과 같다.

$$d_{j,k} = \frac{1}{\sqrt{2^j}} \int f(t) \varphi^*\left(\frac{t}{2^j} - kT\right) dt \quad (3)$$

$$\varphi_{j,k}(t) = 2^{-\frac{j}{2}} \varphi(2^{-j}t - kT) \quad (4)$$

웨이브렛 변환시 각 scale은 각각 바로 위의 scale에서 factor 2로 decimation하여 구해지게 되므로 각 scale의 샘플수는 상위 scale의 반이 되는 것이다. 신호를 웨이브렛 변환을 이용하여 분할하기 위해서 Tree 형태의 필터뱅크를 구성한다. 입력신호가 저역통과 필터와 고역통과 필터를 거치게 되면 한 번의 웨이브렛 변환이 이루어지며, 이러한 과정을 반복적으로 수행하여 웨이브렛 변환된 신호를 얻을 수 있다. 그림 2는 이와같은 과정을 거친 신호들 중에 고역통과 필터를 거친 신호의 주파수 해상도를 나타내고 있고 이것이 바로 웨이브렛 변환의 주파수 해상도이다.

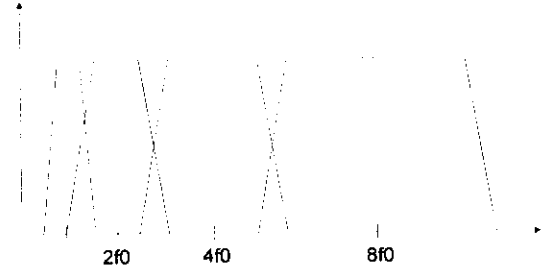


그림 2. Wavelet 변환의 주파수 분포

3.2 웨이브렛 패킷과 코사인 패킷

웨이브렛 패킷은 신호해석에 있어서 더 넓은 범위의 요구를 만족시킬 수 있는 웨이브렛 변환의 일반화이다. 즉, 웨이브렛 패킷은 신호의 다차원 분해를 위한 일반성을 제시하고 있으며 가능한 모든 서브밴드 트리 구성을 가능하게 한다. 웨이브렛 변환이 단순히 저대역부분만 계속 분해해 나가는 것과는 달리, 웨이브렛 패킷은 전 영역에 걸쳐 대역분할이 이루어지게 된다[6.7].

웨이브렛 패킷이 가지는 가장 큰 장점은 주어진 신호에 대하여 최상의 기저(best basis)를 찾는 것이 가능하다는 것이다. 최적의 기저벡터를 찾아내기 위하여 전체 트리구조에서 각 노드에 cost 함수를 적용하여 비교 값을 구한 다음 트리의 제일 밑에서부터 그 값들을 비교하여 가지치기(pruning)를 수행하게 된다. 그리고 이때 사용되는 cost 함수는 일반적으로 식 (5)와 같이 정의되는 엔트로피 함수가 사용되게 된다[6]. 그림 3은 분해된 웨이브렛 패킷 트리를 나타내고 있다.

$$H(x) = - \sum_k \frac{|x_k|}{\|x\|} \log \frac{|x_k|}{\|x\|} \quad (5)$$

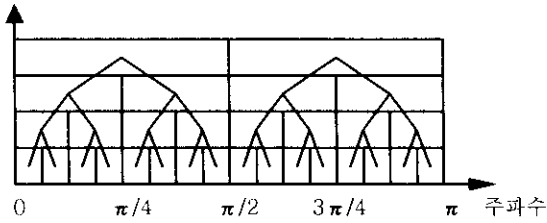


그림 3. 웨이블릿 패킷 트리

웨이블릿 변환이나 웨이블릿 패킷이 lowpass와 highpass 필터링되어 대역제한된 신호를 기저벡터로 사용하는 것과는 달리 코사인 패킷은 일정 길이의 신호를 길이가 다른 여러개의 부분으로 나누어서 각각을 DCT(discrete cosine transform)하여 이를 기저벡터로 사용하게 된다. 즉, 코사인 패킷의 기저벡터는 다양한 길이를 가지는 time segment들이 되는 것이다. 그림 4는 코사인 패킷 트리를 보여주고 있다. 그림 3과 그림 4에서도 확인할 수 있듯이 웨이블릿 패킷은 대역제한된 신호가 기저벡터가 되고 코사인 패킷은 시간축을 길이가 다르게 분리하여 이를 기저벡터로 사용하게 된다.

코사인 패킷 역시 여러개의 기저벡터의 집합들 중에서 최적의 기저벡터를 선정할 수 있으며, 이때 사용하는 cost 함수는 웨이블릿 패킷과 마찬가지로 식 (5)를 사용하게 된다.

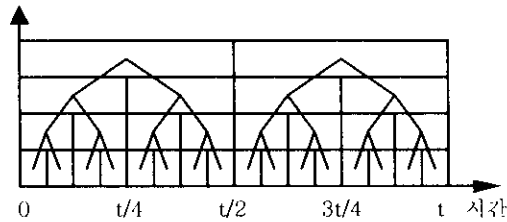


그림 4. 코사인 패킷 트리

3.3 다해상도 신호해석을 이용한 음성개선

앞에서 언급한 다해상도 신호해석 방법인 웨이블릿 변환, 웨이블릿 패킷, 그리고 코사인 패킷을 이용한 음성개선 방식은 기존의 스펙트럼 차감법과 방법에 있어서 매우 유사하다.

우선 웨이블릿을 이용한 방식은, 잡음신호를 프레임별로 웨이블릿 변환하여 여러개의 밴드로 분해하여 각 밴드별로 잡음을 제거하게 된다. 잡음만이 존재하는 초기 몇 프레임의 목음구간에서 추정된 각 밴드의 잡음을 추정하여 이를 각 밴드별로 빼주어서 잡음음성에 포함된 잡음성분을 제거하게 된다.

웨이블릿 패킷은 우선 웨이블릿 패킷 트리를 구성한 다음 식 (5)에서 정의된 cost 함수를 이용하여 최상의 기저벡터 집합을 구성하게 된다. 그러다음 목음구간에서 추정된 잡음성분을 최상의 기저벡터들에서 빼주어서 잡음을 제거한다.

코사인 패킷의 잡음제거 방식은 앞에서도 언급하였듯이 기저벡터 구성시 필터링 대신에 DCT를 이용하는 것 외에는 웨이블릿 패킷과 동일하다.

4. 실험결과 및 검토

실험에 사용된 음성은 10 kHz 샘플링되고 16 bits 로 양자화된 음성신호를 사용하였다. 잡음음성을 위해서는 원음성에 additive white Gaussian noise를 첨가하여 0dB에서 15dB까지의 잡음신호를 만들어 실험을 수행하였다. 분석 프레임은 256 샘플이고, 합성시에는 128 샘플단위로 overlap-add를 수행하여 합성하였으며, 사용된 합성 window는 triangular window를 사용하였다.

웨이블릿 변환과 웨이블릿 패킷을 구성하기 위해서 Daubechies 8 tap 필터를 사용하였고, DCT TYPE-IV를 사용하여 코사인 패킷을 구성하였다. 각 방식의 성능을 비교하기 위한 척도로는 SNR과 켈스트랄 거리를 이용하였다. 표 1과 2는 각 방식에 따른 SNR과 켈스트랄 거리를 보여주고 있다.

표 1. 각 방식에 따른 SNR 비교

방식* SNR	A	B	C	D
15 dB	16.15	16.98	16.93	15.43
10 dB	13.11	13.60	13.36	12.20
5 dB	10.09	10.10	9.85	9.00
0 dB	7.12	6.50	6.39	6.02

표 2. 각 방식에 따른 켈스트랄 거리 비교

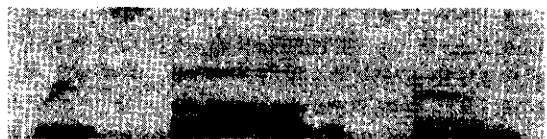
방식* SNR	A	B	C	D
15 dB	0.2699	0.1770	0.1887	0.2387
10 dB	0.2982	0.2382	0.2380	0.2925
5 dB	0.3690	0.2994	0.3558	0.3500
0 dB	0.4490	0.3761	0.4471	0.4185

* A : 스펙트럼 차감법 B : 코사인 패킷
C : 웨이블릿 패킷 D : 웨이블릿 변환

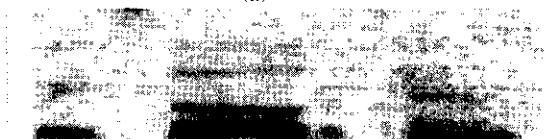
표 1에서도 확인할 수 있듯이 SNR의 경우 각 방식에 따라 큰 차이는 보이지 않았지만 코사인 패킷이 가장 높은 SNR을 나타내었고 웨이블릿 변환을 이용한 방식이 가장 낮은 SNR을 보였다. 하지만 각 방식별로 SNR

의 차이가 평균 1dB 미만이어서 뚜렷한 차이를 보여주지는 못하였다. 켈스트랄 거리의 경우 SNR과는 달리 각 방식별로 뚜렷한 차이가 나타내었다. 코사인 패킷의 경우 SNR에서와 마찬가지로 가장 좋은 성능을 보였고 다른 방식에 비해 켈스트랄 거리가 많이 줄어들음을 확인할 수 있었다. 그리고 웨이브렛 변환을 이용한 방식의 경우 비록 SNR에서는 가장 낮은 수치를 보였지만 켈스트랄 거리에서는 기존의 스펙트럼 차감법보다 더 나은 결과를 보였다.

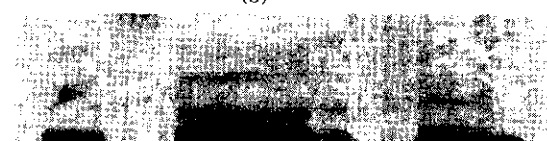
전체적으로 보면 코사인 패킷이 가장 좋은 성능을 보였고 웨이브렛 패킷의 경우 코사인 패킷보다는 성능이 떨어지지만 기존의 스펙트럼 차감법보다는 좋은 성능을 보였다. 주관적인 청취결과 역시 코사인 패킷이 가장 좋은 결과를 보였고, 기존의 스펙트럼 차감법은 필연적인 musical noise의 영향으로 인해 상대적으로 다른 방식에 비해 합성음의 음질이 많이 떨어졌다. 그림 5는 SNR 10dB에서 각 방식을 사용하여 나온 합성음의 스펙트로그램을 보여주고 있다.



(a)



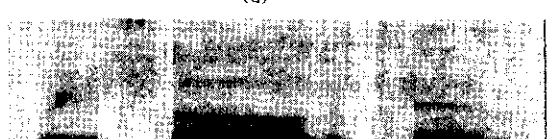
(b)



(c)



(d)



(e)

(a)잡음음성 (b)스펙트럼 차감법 (c)코사인 패킷 방식 (d)웨이브렛 패킷 방식 (e)웨이브렛 변환 방식
그림 5. 각 방식에 따른 합성음의 스펙트로그램

5. 결론

본 연구에서는 최근들어 단구간 푸리에 변환의 단점을 극복하고자 널리 연구되고 있는 다해상도 신호해석 방법인 웨이브렛 변환, 웨이브렛 패킷, 그리고 코사인 패킷 알고리즘을 음성개선에 이용하여 각각의 성능을 비교하였으며, 또한 이를 기존의 스펙트럼 차감법의 성능과 비교 분석 하였다. 성능비교의 척도로는 SNR과 켈스트랄 거리를 이용하였다.

실험결과 SNR면에서는 코사인 패킷이 가장 좋은 결과를 보였으나, 각 방식에 따라 대략 평균 1dB이내의 차이를 보여 큰 차이를 나타내지는 않았다. 하지만 켈스트랄 거리의 경우 코사인 패킷과 웨이브렛 패킷이 훨씬 나은 결과를 보였으며 주관적인 청취결과 역시 코사인 패킷이 가장 좋은 결과를 보였고, 기존의 스펙트럼 차감법은 musical noise의 영향으로 인해 상대적으로 다른 방식에 비해 합성음의 음질이 많이 떨어짐을 확인할 수 있었다.

본 연구는 한국과학기술재단의 핵심전문연구비(과제번호 : 971-0917-103-2) 지원으로 수행되었으며, 지원에 감사 드립니다.

참고문헌

- [1] S. F. Boll, "Suppression of acoustical noise in speech using spectral subtraction," IEEE Trans. on ASSP, vol. 27, no. 2, 1979.
- [2] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," Proc. IEEE, vol. 67, no. 12, 1979.
- [3] J. Hansen and M. Clements, "Constrained iterative speech enhancement with to speech recognition," IEEE Trans. on ASSP, vol. 39, no. 4, 1989.
- [4] I. Daubechies, Ten Lectures on Wavelets, SIAM, 1992.
- [5] O. Rioul and M. Vetterli, "Wavelet and Signal Processing", IEEE Signal Processing Magazine, pp. 14-38, Oct. 1991
- [6] R. R. Coifman, Y. Meyer, and V. Wickerhauser, "Entropy based algorithms for best basis selection," IEEE Trans. on Infomation Theory, vol. 32, 1992.
- [7] V. Wickerhauser, "Adapted wavelet analysis from theory to software," IEEE Press, 1994.