

음성/잡음 차등 주파수차감법에 의한 잡음처리 및 기존 주파수차감법과의 성능 비교

황규연, *이경준 **정제창

한양대학교, *한양대학교, **한양대학교

bascara@naver.com, *kjlee888@naver.com, **jjeong@hanyang.ac.kr

A Speech Enhancement Using Speech/Noise-dominant Frequency Subtraction and Comparing with Normal Frequency Subtraction

Hwang Kyu-Yeon *Lee Kyung-Jun **Jeong Je-Chang

Hanyang University *Hanyang University **Hanyang University

요 약

본 논문에서는 기존에 쓰이던 주파수차감법과 다른 새로운 방법을 제안한다. 본 논문에서 다루는 방법은, 특정한 주파수의 대역에서 음성과 잡음의 우세도를 결정하고, 인간의 청각기와 관련된 마스크 성질을 기반으로 하여 주파수 차감법을 이용해 제거한다. 이에 대하여 다양한 성능 평가를 하였고, 기존의 일반적인 주파수차감법과 비교하여 보다 효과적으로 잡음처리를 할 수 있음을 알 수 있다.

1. 서론

다양한 음성처리 시스템의 계속된 성장으로 인하여 채널에서의 고성능 잡음처리 시스템의 필요성이 증가되고 있는 추세이다. 음성에 잡음이 추가되는 경우, 음성인식기에서의 인식률이 떨어지며 음성부호기로의 코딩 효율성을 낮춘다. 일반적으로, 잡음은 실제의 음성과 처리된 음성 사이에서 많은 차이를 발생시키게 되며 또한 듣는 사람에게 피로감을 준다. 이러한 점을 보완하기 위해 잡음을 줄이려는 시도로 다양한 방법이 있었다. 주파수차감법[1-4], 소프트 디시전(soft-decision) 필터링 방법[5], 평균제곱오차의 추정(MMSE estimation) 방법[6,7], 음성모델에 기반 잡음처리방법[8-10], 인간 청각 특성 이용한 잡음처리방법 등이 있었다[11-14]. 이러한 방법들은 통계적인 정보가 필요하고, 목음이 있는 구간을 주로 이용하여 잡음을 추정한다. 하지만 목음의 구간에 대한 정확한 검출은 매우 어렵고, 또한 잘못된 검출은 오히려 음성을 더욱 왜곡하는 결과를 낳게 된다. 과거에도 목음의 구간 검출 없이 음성 등 잡음을 줄이려는 시도가 있었지만[15, 16], 만족스런 결과를 내지 못했다.

본 논문에서는, 첫째로 새로운 음성/잡음 차등 주파수차감법 처리 시스템에 대해 간략히 설명한다. 영어로 문장을 제안하고, 이에 자동차 소음, 탱크잡음, F16 잡음, car 잡음, 백색잡음, 핑크잡음을 이용해 테스트를 한다. 새로운 주파수 차감법을 제안하며 이 성능을 기존 주파수차감법과 비교하며 얼마나 차이가 나는지 확인한다.

2. 본론

2.1 음성/잡음 차등 잡음처리 시스템

본 논문에서는, 어떤 프레임 특정 대역에 포함되는 주파수의 성분에 대한 크기의 합과, 그러한 특정 대역의 과거 주파수에 대한 성분 크기의 합들과 비교한 후, 그 프레임 특정 대역에서 음성인 성분이 우세한지 (음성우세), 잡음인 성분이 우세한지 (잡음우세) 결정을 하고 마스크 성질(인간 청각기) 기반하여, 음성의 왜곡은 최소로 하고 잡음을 최대한 제거한다. 과거 프레임들의 특정 대역의 합들을 오름차순 정렬하여, 그 모양에 기반하여 대역에 대해 두 가지의 종류로 분류한다. 그리고 종류에 따른 다른 기준을 적용함으로써 오름차순 정렬 합과 비교하여 음성우세인지 잡음우세인지 결정을 한다. 이러한 결정과 마스크 성질(인간 청각기)에 기초하여, 각각 음성우세/잡음우세 대역에서 잡음을 주파수 차감법을 활용하여 제거한다[22].

본 방법은 3 단계로 이루어졌다. 첫째, 음성 $y[n]$ 을 $w[n]$ 으로 썬다. 그 후 신호를 DFT 계수로써 변환한다. 임계 밴드 계수 크기를 합하고, 이를 오름차순으로 정렬한다. 둘째, 근사함수로써 모양에 따라 두 종류로 분류한다. 종류에 따라 다른 기준을 적용하여 음성우세, 잡음우세를 결정 한다. 셋째, 이를 마스크 성질에 기반해 각각의 잡음을 적절히 주파수 차감법을 이용해 제거한다. 그림 1 은 전체 시스템을 보여준다[22]. 그림 2, 3 은 정렬 및 종류 분류를 보여준다. 종류 1 과 종류 2 의 구분은 음성성분이 잡음성분과 명확히 구별되는지 여부로 판단한다.

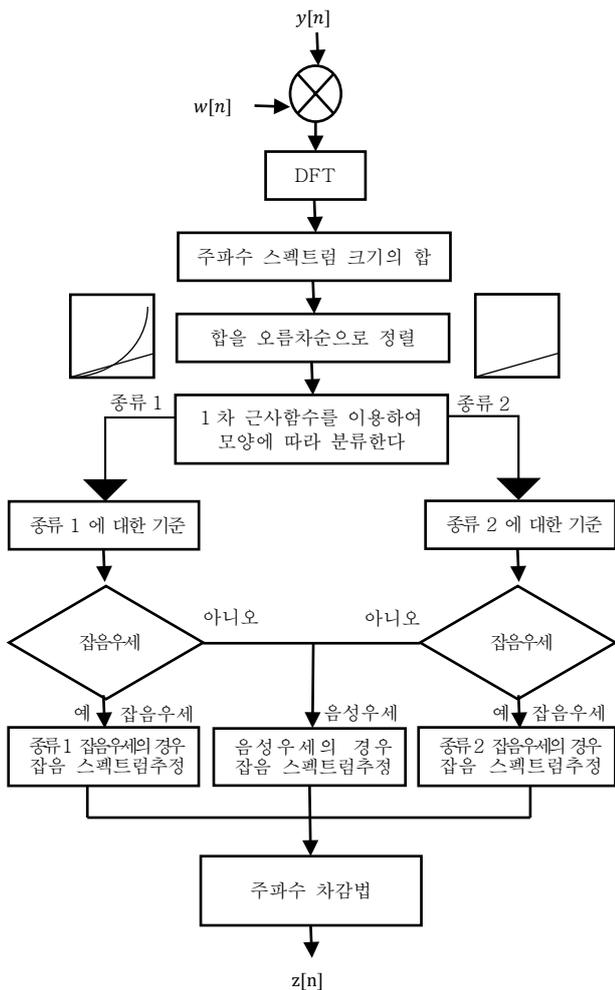


그림 1. 전체 시스템

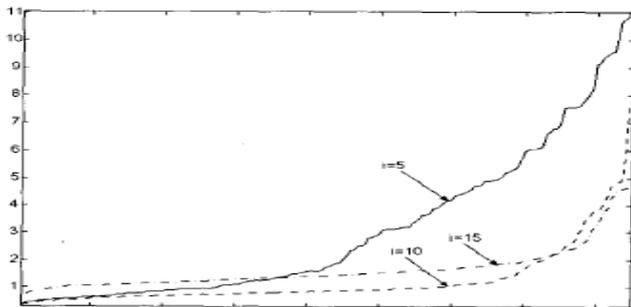


그림 2. i 번째 임계밴드에 대한 오름차순 정렬된 수열의 그래프 ($i=5, 10, 15$) [22]

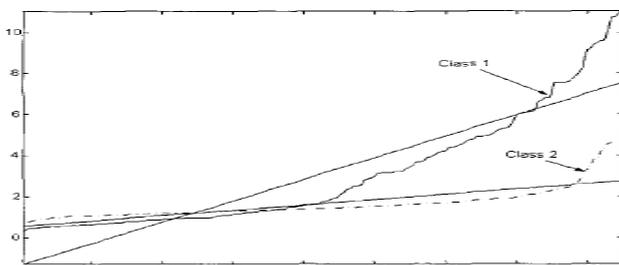


그림 3. 종류 1(Class 1), 종류(Class 2)로의 분류 [22]

2.2 음성/잡음 차등 주파수차감법과 기존 주파수차감법의 성능 비교

제안된 방법의 성능을 보여주기 위해 우리는 다양한 잡음상황에서 평가를 해보았다. 잡음처리에 대해 실험할 문장은 다음 두 문장이다. “She had your dark suit in greasy wash water all year”와 Scholastic aptitude is judged by standardized tests” – 각각 남성, 여성에 의해서 읽힌 문장이다.

Noisex-92 데이터베이스에서 6 개의 다른 잡음을 뽑아내어 평가하는데 사용하였다. 자동차잡음, F16 잡음, 백색잡음, 핑크잡음, 탱크잡음, 혼선잡음이다.

스펙트로그램은 시각적으로 알 수 있는 측정 도구로서 가로축은 시간, 세로축은 주파수를 나타낸다. 스펙트로그램은 시각적으로 음성의 상태를 쉽게 알 수 있다는 특징이 있다. 또한 세그멘탈 SNR의 방법도 사용하였다.

2.2.1 자동차 잡음에서 음성/잡음 차등 주파수차감법과 기존 주파수차감법의 성능 비교

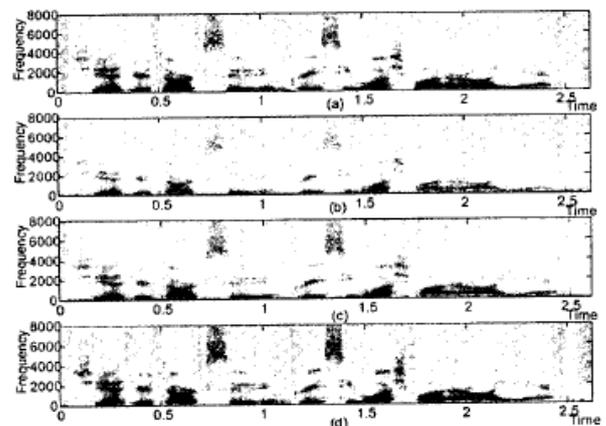


그림 4. (a)깨끗한 음성(“She had your dark suit in greasy wash water all year” 남자 음성), (b)자동차잡음에 의해 오염된 음성(SNR=-5dB), (c)목음구간에서의 잡음추정을 이용한 주파수차감법에 의해 잡음처리 된 음성 (d)제안된 방법에 의해 잡음처리 된 음성의 스펙트로그램 [22]

그림 4(a), (b), (c), (d)는 각각 깨끗한 음성(“She had your dark suit in greasy wash water all year” 남자의 음성), 자동차잡음에 의해 오염된 음성(SNR=-5dB), 목음구간에서의 잡음추정을 이용한 주파수차감법에 의해 잡음처리 된 음성, 그리고 제안된 방법에 의해 잡음처리 된 음성의 스펙트로그램을 나타낸다.

그림 4(b)는 자동차잡음에 의해 오염된 음성이다. 자동차잡음의 특성은 저주파(100Hz 부근)에 대부분의 잡음성분이 몰려 있는 특성이 있다. 그 외의 부분은 백색잡음처럼 넓게 고르게 퍼져있다.

그림 4(c)는 목음구간을 이용한 잡음처리를 통해 나오는

스펙트로그램인데, 깨끗한 음성의 스펙트로그램을 나타내는 그림 4(a)와 비교해서 그리 선명한 모습을 보여 주지 못한다. 특히 저주파 부근의 잡음 성분은 거의 제거되지 않은 결과를 확인할 수 있다.

그림 4(d)를 보면 그림 4(c)에 비해 선명하고 저주파 부근에 잡음이 거의 사라졌음을 관찰할 수 있다.

2.2.2 탱크 잡음에서 음성/잡음 차등 주파수차감법과 기존 주파수차감법의 성능 비교

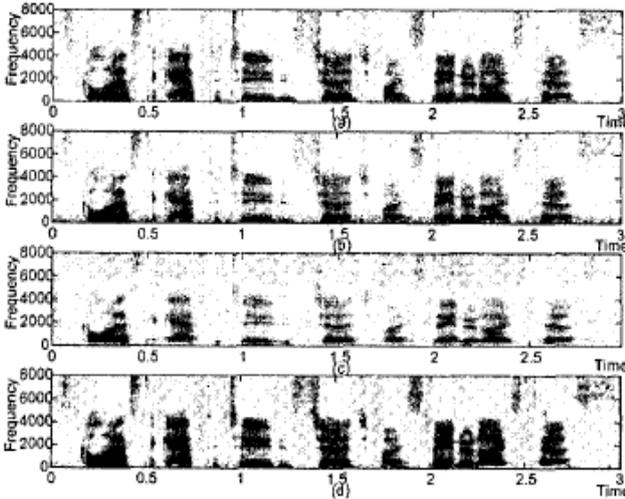


그림 5. (a)깨끗한 음성(“Scholastic aptitude is judged by standardized tests” 여자 음성), (b)탱크잡음에 의해 오염된 음성(SNR=10dB), (c)목음구간에서의 잡음추정을 이용한 주파수차감법에 의해 잡음처리 된 음성, (d)제안된 방법에 의해 잡음처리 된 음성의 스펙트로그램 [22]

그림 5 는 각각 (a)깨끗한 음성(“Scholastic aptitude is judged by standardized tests” 여자의 음성), (b)탱크잡음에 의해 오염된 음성(SNR=10dB), (c)목음구간에서의 잡음추정을 이용한 주파수차감법에 의해 잡음처리 된 음성 그리고 (d)제안된 방법에 의해 잡음처리 된 음성의 스펙트로그램을 나타낸다.

그림 5(b)를 보면 탱크잡음은 저주파 (1kHz 아래부분) 부근에 강하게 나타나고 그 외의 부분은 저주파보다는 약한 형태로 나타난다.

그림 5(c)를 보면 저주파부근에 잡음성분은 어느 정도 제거하였으나 그 외의 부분은 제대로 처리하지 못 하였음을 볼 수 있다. 이는 목음구간만을 통해 잡음을 추정하므로 변화하는 잡음에 빠르게 대처하지 못하기 때문이다.

그림 5(d)를 보면 탱크잡음의 저주파성분 (1kHz 아래부분)뿐만 아니라 그 외의 구역도 매우 깨끗하게 잡음이 제거되어 있고 음성성분도 잘 보존되어 있음을 알 수 있다.

2.2.3 여러가지 잡음에서 음성/잡음 차등 주파수차감법과 기존 주파수차감법의 성능 비교

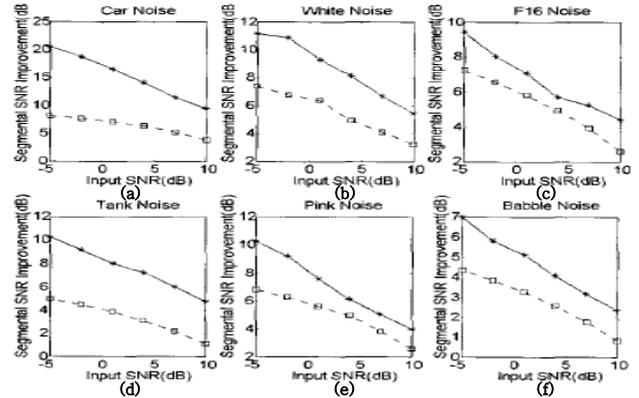


그림 6. 입력 SNR 대 향상된 segmental SNR 의 그래프 (a) 자동차 잡음, (b)F16 잡음, (c)백색 잡음, (d)핑크 잡음, (e)탱크 잡음, (f)혼선잡음 (실선 : 제안된 방법, 점선 : 목음구간에서의 잡음추정을 이용한 주파수차감법) [22]

그림 6 은 입력 SNR 대 향상된 세그멘탈(segmental) SNR 의 그래프를 나타내었다. - (a)자동차 잡음, (b)F16 잡음, (c)백색 잡음, (d)핑크 잡음, (e)탱크 잡음, (f)혼선잡음에 대해서 제안된 방법과 목음구간추정을 이용한 주파수차감법을 각각 비교하여 나타내었다.

그림 6 을 보면 다양한 잡음상황에서도 기존의 방법보다 전체적으로 향상된 성능을 보인다. 특히, 자동차잡음에 대해선 기존의 방법에 비해 탁월 한 성능을 보이고 있음을 알 수 있다.

3. 결론

본 논문에서 우리는 잡음에 대한 사전정보나 목음구간에서의 잡음추정을 필요로 하지 않는 새로운 잡음처리기법을 소개하였다. 또한, 그 성능을 비교하여 음성/잡음 우세를 이용한 잡음처리가 기존 방법에 비해 매우 효과적이라는 사실을 알 수 있었다.

제안된 방법은 다양한 환경에서 성능평가가 이루어졌다. (자동차 잡음, F16 잡음, 백색 잡음, 핑크 잡음, 탱크 잡음, 혼선잡음) 그리고 시각적 측정 척도인 스펙트로그램 뿐만 아니라 세그멘탈 신호대 잡음비를 통해, 음성왜곡은 줄이면서 효과적으로 잡음을 줄일 수 있음을 알 수 있다.

감사의 글

이 논문은 2015 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2015R1A2A2A01006004).

참고문헌

- [1] S. F. Boll, "Suppression of acoustic noise speech using spectral subtraction," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-27, pp.113-120, Apr. 1979.
- [2] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*. vol. 67, pp. 1586-1604, Dec. 1979.
- [3] M. Erouti and R. Chwartz, "Enhancement of speech corrupted by acoustic noise," *Proc. IEEE ICASPP*, Washington, DC, pp. 208-211, Apr. 1979.
- [4] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor(NSS), hidden Markov models and projection, for robust recognition in cars," *Speech Commun.*, vol. 11, pp. 215-228, June, 1992.
- [5] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, pp. 137-145, Apr., 1980.
- [6] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, pp. 1109-1121, Dec., 1984.
- [7] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean Square Error Log-Spectral Amplitude Estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 443-445, Apr., 1985
- [8] Y. Ephraim, "A Bayesian Estimation Approach for Speech Enhancement Using Hidden Markov Models," *IEEE Trans on Signal Processing*, vol. 40, pp.725-735, Apr. 1992.
- [9] Y. Ephraim, "Statistical-Model-Based Speech Enhancement System," *IEEE proceeding*, vol. 80. pp.1526-1555, Oct., 1992.
- [10] H. Sameti, H. Sheikhzadeh, and L. Deng, "HMMBased Strategies for Enhancement of Speech Signals Embedded in Nonstationary noise," *IEEE trans. on speech and audio processing*, vol. 6, pp. 445-455, Sep., 1998.
- [11] D. Tsoukalas, M. Paraskevas, and J. Mourjopoulos, "Speech enhancement using psycho-acoustic criteria," *proc. IEEE ICASSP*, pp. 359-361, Apr., 1993.
- [12] T. Usagawa, M. Iwata, and M. Iwata, "Speech parameter extraction in noisy environment using a masking model," *proc. IEEE ICASSP*, pp. 81-84, Apr., 1994.
- [13] S. Nandkumar and J. H. L. Hansen, "Dual-channel iterative speech enhancement with constrains on an auditory-based spectrum," *IEEE tran. on speech and audio processing*, vol 7, pp. 22-34, Jan. 1995.
- [14] N. Virag, "Single Channel Speech Enhancement Based on Masking Property of the Human Auditory System," *IEEE Trans. On speech and audio processing*, vol. 7, pp. 126-137, Mar., 1999.
- [15] V. Stahl, A. Fischer, and R.Bippus, "Quantile Based Noise Estimation for Spectral subtraction and Wiener Filtering" in ICASSP, 2000.
- [16] H. G. Hirsch and C. Ehrlicher, "Noise Estimation Techniques for Robust Speech Recognition," in *Proc. ICASSP*, pp. 153-157, 1995.
- [17] D. Kincaid and W. Cheney, *Numerical Analysis*, Brooks/Cole Publishing Company, 1996.
- [18] S. Quakenbush, T. Barnwell, and M. Clements, *Objective Measures of Speech Quality*, Prentice-Hall, 1988.
- [19] T. Painter and A. Spanias, "A Review of Algorithms for Perceptual Coding of Digital Audio Signals," *Digital Signal Processing Proceedings*, vol. 1, pp 179-208, 1997.
- [20] B. Scharf, *Critical Bands, in foundation of modern auditory theory*, New York : Academic Press, 1970.
- [21] T. Painter and A. Spanias, "Perceptual Coding of Digital Audio," *Proc. IEEE*, vol. 88, PP. 451-513, 2000.
- [22] 윤석현, 유창동 "시간-주파수 영역에서의 음성/잡음 우세결정에 의한 새로운 잡음처리," *한국음향학회지*, vol. 20, no.3, pp.48-55, 2001