

# 결정적/확률적 요소로의 음성 분해와 심리음향 모델 기반 잡음 제거 기법

\*조석환, 유창동

한국과학기술원 전자전산학부 전기및전자공학전공  
e-mail : antiland00@kaist.ac.kr, cdyoo@ee.kaist.ac.kr

## Speech Enhancement with Decomposition into Deterministic and Stochastic components and Psychoacoustic Model

\*Seokhwan Jo, Chang D. Yoo

School of Electrical Engineering and Computer Science  
Division of Electrical Engineering  
Korea Advanced Institute of Science and Technology

### Abstract

A speech enhancement algorithm based on both a decomposition of speech into deterministic and stochastic components and a psychoacoustic model is proposed. Noisy speech is decomposed into deterministic and stochastic components, and then each component is enhanced preserving its individual characteristics. A psychoacoustic model is taken into account when enhancing the stochastic component. Simulation results show that the proposed algorithm performs better than some of the more popular algorithms.

### I. 서론

음성에 섞인 잡음은 음성통신과 음성인식 등 음성을 사용하는 시스템의 성능 저하를 일으키는 요인이 될 뿐만 아니라 듣는 사람에게 피로를 일으키는 요인이 되므로 음성에 섞인 잡음을 제거해야 할 필요가 있다. 그로 인해, 음성에 섞인 잡음을 제거하는 기법은 오래 전부터 연구되어 왔다.

잡음 제거 알고리즘은 잡음을 제거하는데 치중하게

되면 원래 음성의 왜곡이 필연적으로 발생하기 때문에 잡음을 최대한 줄이면서 음성의 왜곡을 최소화하는 것이 중요하다. 이 논문에서는 새로운 잡음 제거 알고리즘에 대해 서술한다.

### II. 본론

음성을 살펴보면 조화 변조 사인함수 형태 - 결정적 성분 - 와 랜덤 잡음 형태 - 확률적 성분 - 가 동시에 존재하는 것을 볼 수 있다[1]. 이런 음성의 특성을 반영하여 잡음 제거 알고리즘을 설계하여 잡음을 최대한으로 줄이고 음성의 왜곡을 최소화하도록 한다.

이 논문에서 제시하는 잡음 제거 알고리즘은 먼저 잡음이 섞인 음성을 결정적 성분과 확률적 성분으로 나누는 뒤, 잡음이 확률적 성질을 가지고 있다고 가정하여 결정적 성분에서는 간단한 수정을 통해 원 음성의 결정적 성분을 추정하고 확률적 성분에서 대부분의 잡음 제거가 이루어져 원 음성의 확률적 성분을 추정한다. 결정적 성분을 추출하기 위해서는 음성의 음조(pitch)와 조화 진폭(harmonic amplitude)을 추정하는데 잡음이 섞이더라도 음성의 음조는 크게 변화하지 않는다는 것이 밝혀져 있어[1] 원 음성의 결정적 성분을 추정하기 위해서는 조화 진폭의 수정을 통해 쉽게

추정한다. 잡음이 섞인 음성에서 앞에서 추출한 원 음성의 결정적 성분을 제외한 부분을 확률적 성분이라고 하고 여기서 위너 필터를 통과, 잡음을 제거하고 원 음성의 확률적 성분을 추출하여 최종적으로 이 두 신호의 합으로 원 음성을 복원할 수 있다는 것이 이 논문에서 제시하는 잡음 제거 방법이다. 이 때 잡음이 첨가된 음성의 확률적 성분에 잡음이 들어있다고 가정하였고 대부분 음성의 결정적 성분보다 확률적 성분에 에너지가 작기 때문에 원 음성의 확률적 성분을 추정하는 것이 어렵다. 이를 위해 이 논문에서는 심리음향 모델, 그 중에서도 마스킹 문턱값을 사용하는데 이 마스킹 문턱값 아래의 신호는 사람이 들을 수 없기 때문에 잡음 제거를 많이 하지 않고 마스킹 문턱값 위의 신호에서 잡음 제거를 많이 한다[2]. 즉, 마스킹 문턱값을 사용하여 잡음 제거에 쓰이는 위너 필터의 계수를 조정하면 원 음성의 확률적 성분을 잘 복원할 수 있게 된다.

### III. 실험 결과

이 논문에서 서술한 알고리즘과 지금까지의 대표적인 잡음 제거 알고리즘들의 성능을 비교하였다.

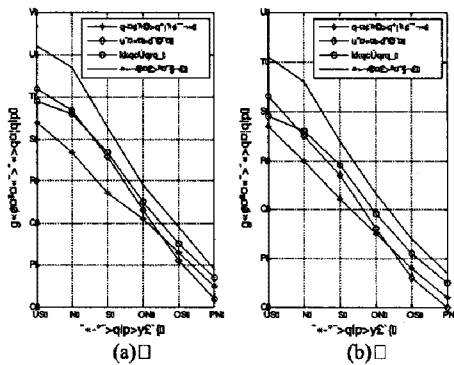


그림 1 Segmental SNR 비교

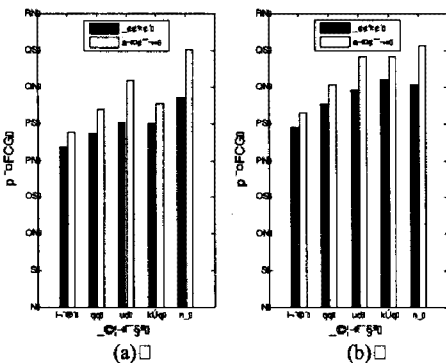


그림 2 잡음제거 후 음성인식 결과

실험결과를 살펴보면 비교대상이 되는 주파수 차감법 (spectral subtraction) [3], 위너 필터링 (Wiener filter)[4], 최소평균오차 주파수크기성분 추정기 (minimum mean square error short time spectral amplitude estimator)[4]보다 성능이 좋은 것을 알 수 있다.

### IV. 결론 및 향후 연구 방향

과거의 잡음 제거 알고리즘들이 음성과 잡음에 대한 간단한 모델과 가정을 한 반면 이 논문에서는 음성의 특성을 반영하여 잡음을 제거하였다. 음성의 특성을 반영하였을 뿐만 아니라 심리 음향 모델을 사용하여 과거의 잡음 제거 알고리즘들과 비교해 잡음을 최대한 줄이면서 음성의 왜곡을 최소화하는 목적에 더 가깝도록 만들었기 때문에 음질 측면에서, 그리고 사람이 듣기에도 나은 성능을 나타내었다.

앞으로는 음성의 특성을 지금보다 더 적극적으로 반영할 수 있는 잡음제거 알고리즘을 개발하려고 한다.

### 참고문헌

[1] C.D. Yoo J. Hardwick and J.S. Lim, "Speech enhancement using the dual excitation model," *Proc. IEEE Int. Conf. Acoustic, Speech and Signal Processing*, vol. 2, pp. 367 - -370, 1993.

[2] N. Virag, "Single channel speech enhancement based on masking properties of human auditory system," *IEEE Transactions on Speech and Audio Processing*, vol. 7, pp. 126 - -137, 1999.

[3] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, April 1979.

[4] Y. Ephraim, "Speech enhancement using a minimum meansquare error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, September December 1984.