

지능로봇에 적합한 잡음 환경에서의 원거리 음성인식 전처리 시스템

권세도*, 정홍**
포항공과대학교 전자전기공학과

Remote speech recognition preprocessing system for intelligent robot in noisy environment

Sedo Gwon*, Hong Jeong**
Department of Electronics and Electrical Engineering
POSTECH
E-mail : *ralrara@postech.ac.kr, **hjeong@postech.ac.kr

Abstract

This paper describes a pre-processing methodology which can apply to remote speech recognition system of service robot in noisy environment. By combining beamforming and blind source separation, we can overcome the weakness of beamforming (reverberation) and blind source separation (distributed noise, permutation ambiguity). As this method is designed to be implemented with hardware, we can achieve real-time execution with FPGA by using systolic array architecture.

접한 발화 환경)에서만 어느 정도의 성과를 나타낼 뿐, 정작 일상적인 환경(일상적 잡음 환경에서 마이크가 로봇에 장착된 원거리 음성인식)에 있을 때에는 그 인식이 현저히 떨어지게 된다.

이에 본 논문에서는 음성 인식의 전처리 단계에서 Beamforming 기법과 Blind Source Separation 기법을 이용해 잡음 환경에서의 원거리 음성인식 시스템의 성능을 향상시킬 수 있는 방법에 대해 기술한다.

I. 서론

최근 지능로봇에 관한 관심이 집중되고 있고 실제로 두 발로 걸을 수 있는 휴머노이드가 속속 등장하면서 어느 때 보다 로봇에 대한 기대가 높아지고 있다. 하드웨어와 제어 기술의 발전으로 인해 로봇의 몸체는 점점 사람과 가까워지고 있는데 반해, 인공지능 분야에서의 성과는 아직 미진하여 진정한 의미의 '인간을 닮은 로봇' 휴머노이드에는 미치지 못하는 수준이다. 특히 인간과 로봇의 인터페이스에 있어 가장 자연스러운 형태인 음성을 통한 의사소통 능력에 대해서는 아직 제한된 환경(주위 잡음이 없고, 유무선 마이크를 통한 근

II. 본론

마이크 어레이를 이용해 Adaptive beamforming 과 BSS 를 혼합하고자 하는 노력들은 꾸준히 있어 왔다. [1] 하지만 지금까지의 접근들은 대부분 frequency domain 에서 이루어져, 실시간으로의 구현이 어려웠다. 이에 본 논문에서는 Beamforming 과 BSS 를 time domain 상에서 연결하여 Beamforming 의 약점인 reverberation 문제와 BSS 의 약점인 distributed noise 에 대한 처리와 permutation ambiguity 를 상호 보완하려고 하였다. Time domain 에서의 BSS 는 그림 2 와 같이 본 연구실의 연구 결과[3]를 활용했으며, 그림 1 의 adaptive beamforming[2] 을 통해 얻어진 enhanced signal 과 noise signal 을 BSS 의 input 으로 넣어주는 구조를 가지고 있다. 기존의 BSS

알고리즘이 Source 의 개수와 마이크의 개수가 같아야하는 제약 조건이 있었고, 또한 계산량의 증가로 인해 2 개 이상의 마이크를 사용하기 어려웠던 점을 앞 단에 beamforming 을 넣어줌으로써 극복할 수 있도록 했다. 또한 BSS 의 문제점이었던 distributed noise 를 beamforming 을 통해 point source 로 바꿀 수 있으므로[4] 성능의 향상을 가져올 수 있다.

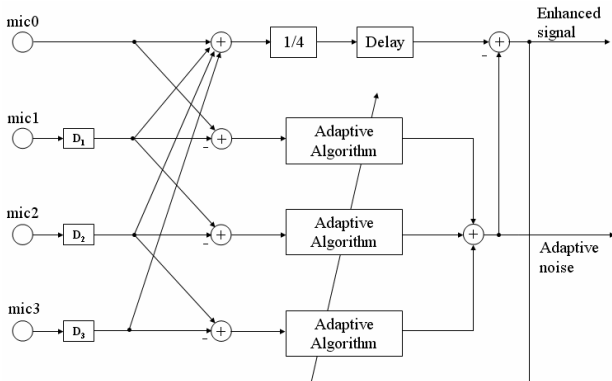


그림 1. Block diagram of 4-Channel Griffiths-Jim adaptive beamformer

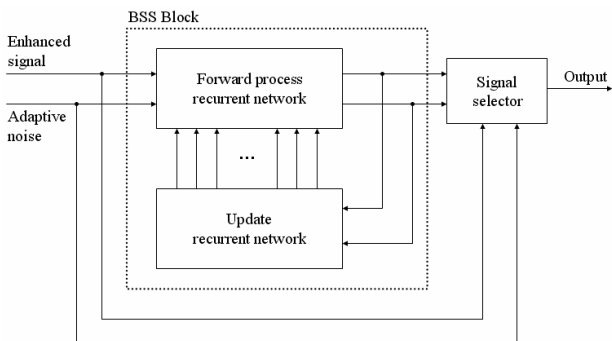


그림 2. Block diagram of 2-Channel BSS using systolic array

위의 그림 1 은 4 개의 마이크 어레이에서 입력이 들어와 맨 위의 mic0 쪽에서는 4 개의 입력을 delay-sum 방식을 이용해 향상된 output 을 구하고, 아래쪽 3 개의 마이크에서는 들어오는 signal 은 이웃한 마이크 간에 delay-subtraction 방식을 이용해 그림 3 과 같이 원하는 신호를 제외한 노이즈 신호를 얻어내게 된다. Beamforming 을 통해 얻어진 enhanced signal 에도 distributed noise 는 섞여 있으므로, 이를 noise signal 과 BSS 를 수행함으로써 distributed noise 의 감쇠 효과를 얻을 수 있다.[4] 또한 일반적인 BSS 에서 생기는 permutation ambiguity 를 Beamforming 의 결과로 얻은 signal 들로 해결할 수 있다.

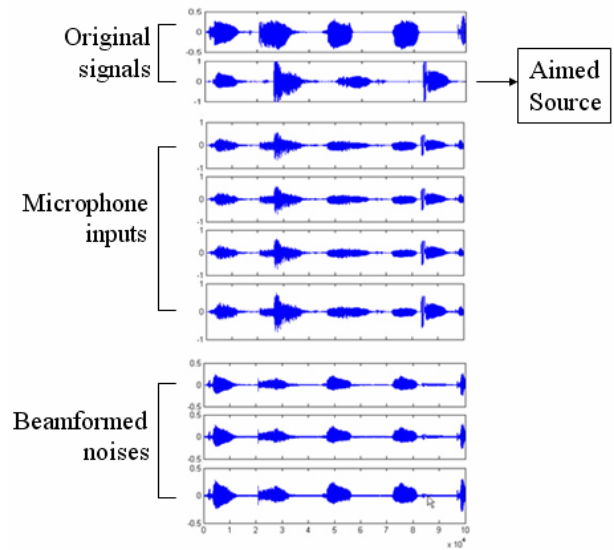


그림 3. Noise signal from beamforming

IV. 결론 및 향후 연구 방향

본 논문에서는 Adaptive Beamforming 과 BSS 를 이용하여 잡음 환경에 강인한 음성인식 시스템을 만들기 위한 전처리 시스템을 제안했다. 또한 이를 real time 실현을 위해 time domain 에서 구현하여, 실제 음성 인식에 활용할 수 있도록 하였다. 추후 speaker 가 이동하는 경우나 위치가 확실치 않은 경우에 대해서도 시스템이 동작하도록 하도록 시스템을 개선할 계획이다.

참고문헌

- [1] Qiongfeng Pan and Tyseer Aboulnasr, "Combined spatial/beamforming and time/frequency processing for blind source separation", EUSIPCO 2005
- [2] Tony. P. W. Price, David. M. Howard, Alwyn. V. Lewis, Andy. M. Tyrrell, "Adaptive Microphone Array Beamforming for Teleconference Using VHDL and Parallel Architectures", Parallel and Distributed Processing, 1999. PDP '99. Proceedings of the Seventh Euromicro Workshop on 3-5 Feb. 1999 Page(s):13 - 18
- [3] H. Jeong and Y. Kim, "Parallel feedback network architecture for blind source separation", ELECTRONICS LETTERS 19th August 2004 Vol.40 No.17.
- [4] Erik Visser, Manabu Otsuka, Te-Won Lee "A spatio-temporal speech enhancement scheme for robust speech recognition in noisy environment", Speech Communications, 2003