원거리 음성인식 시스템의 잡음 제거 기법에 대한 연구

*우성민, 이상훈, 정홍
포항공과대학교 전자전기공학과
e-mail : *innosm@postech.ac.kr, god1710@postech.ac.kr hjeong@postech.ac.kr*

Noise removal algorithm for intelligent service robots in the high noise level environment

*Sung-Min Woo, Sang-Hoon Lee, Hong Jeong
Electrical and Electronic Engineering
Pohang University of Science and Technology

Abstract

Successful speech recognition in noisy environments for intelligent robots depends on the performance of preprocessing elements employed. We propose an architecture that effectively combines adaptive beamforming (ABF) and blind source separation (BSS) algorithms in the spatial domain to avoid permutation ambiguity and heavy computational complexity. We evaluated the structure and assessed its performance with a DSP module. The experimental results of speech recognition test shows that the proposed combined system guarantees high speech recognition rate in the noisy environment and better performance than the ABF and BSS system.

I. 서론

최근 두발로 걷는 휴머노이드 로봇이 속속 등장함에 따라 지능형 로봇에 대한 관심도 증가하고 있다. 하드웨어와 제어기술의 발전으로 로봇의 움직임은 사람과 점점 닮아가고 있지만 인간과 로봇의 의사소통을 위한 가장 기본적인 음성인식기술은 상대적으로 발전이 더디다. 현재, 소음이 거의 없는 환경에서만 음성인식 성공률이 보장되고 원거리 발화의 경우에도 성능이 저하된다. 본 논문에서는 음성인식성능을 향상시키기 위하여 특정방향의 소리만을 감지하여 집중하는 인간의 청각능력을 벤치마킹하였다. 이것을 각각 ABF와 BSS 기법으로 표현하여 조합하는 알고리즘을 소개한다.

II. 본론

주위의 장애물에 의해 음성신호가 반사되는 환경에서 ABF기법의 단점인 cross-talk 문제와 BSS기법의 permutation ambiguity를 해결하기 위해 두 기법을 조합하는 알고리즘을 제안한다.

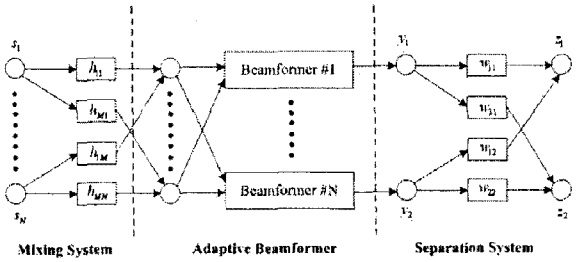

그림 1. 제안된 전처리 잡음제거 구조

Adaptive beamformer를 위해 적은 수의 filter tap을 가지고도 좋은 성능을 보이는 Adaptive Generalized Sidelobe Canceller(AGSC)를 사용했다[1]. 4개의 마이크를 통해 각각의 입력신호를 받아 물리적인 거리에 따른 delay를 적용한 뒤 이것을 합치고 blocking filter와 cancelling필터를 적용한 이전 신호와 더해 출력신호를 얻는다. 다른 하나의 출력에서는 잡음과 유사한 신호를 만들어낸다. blocking filter와 cancelling filter의 weight는 information maximization learning rule에 의해 업데이트 된다. 제안된 시스템의 두 번째 단계는 신호분리시스템으로써 time-domain convolutive BSS를 이용하였다[2].

$$y(t) = x(t) + \sum_{p=0}^{L} W_p(t) y(t-p),$$

$$= [I - W_0]^{-1} \{x(t) + \sum_{p=1}^{L} W_p y(t-p)\}.$$

Time-domain convolutive mixture model은 위의 식으로 표현된다. Vector x와 y는 각각 입력과 출력이며 W는 weight vector, L은 filter의 tap수이다. 이 식의 해를 구하기 위해서는 W_0의 inverse를 계산해야하므로 상당히 복잡하다. 하지만 $p=0$ 일 때를 무시하면, 즉 하나의 출력이 sampling time에서 다른 하나의 출력에 영향을 미치지 않는다고 가정하면 위의 식은 다음과 같이 간략화 된다.

$$y(t) = x(t) + \sum_{p=1}^{L} W_p(t) y(t-p),$$

$$= x(t) + \sum_{p=1}^{L} W_p y(t-p).$$

이 과정으로 DSP implementation을 할 경우, 계산량은 상당히 줄어들고 출력에는 거의 변화가 없는 것을 확인하였다.

앞에서 기술된 두 기법을 조합하여 전체의 시스템을 구현하였다. 이 구조는 두 개의 stage로 구성되어있으며 AGSC에서 특정한 방향의 신호만을 받아들여 BSS에서 그것을 잡음과 원하는 음성신호로 분리한다.

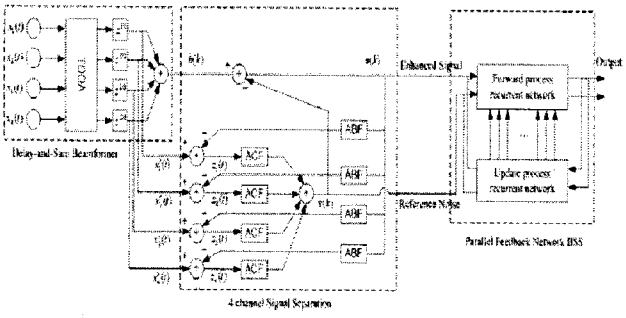

그림 2. Combined Structure of ABF and BSS

Ⅲ. 성능검증 및 비교

제안된 알고리즘을 TI의 TMS320C6713를 통해 구현하였으며 출력을 HTK를 이용하여 3음절이상의 단어 인식실험을 하였다. AGSC는 64개의 tap을 사용하였고 본 연구실에서 개발된 parallel feelback network BSS[3]를 사용했다. 잡음이 없는 환경에서는 ABF, BSS와 비슷한 인식률을 보였지만 주위잡음이 커질수록 다음과 같이 인식률에 있어서 차이를 보였다.

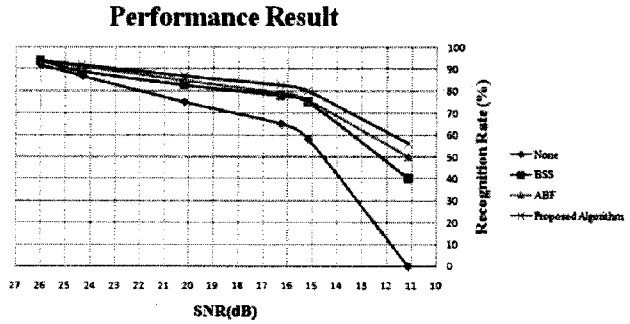

그림 3. 음성 인식률

Ⅳ. 결론 및 향후 연구 방향

본 논문에서 제안한 combined architecture는 지능형 서비스 로봇의 음성인식을 위한 잡음제거에 효과적임을 보였다. 두 개의 ABF와 BSS 알고리즘을 조합하여 cross-talk, permutation 문제를 극복했다. 향후 향상된 음성인식성능을 위해서 사람의 음성신호만을 선택적으로 받아들이는 연구가 필요하겠다.

참고문헌

[1] C.Choi, D.Kong, J,Kim : Speech enhancement and recognition using circular microphone array for service robits, In Proceedings IEEE/RSJ International Conference on IRS, 2003, pp.3516-3521

[2] K. Torkkola : Blind separation of convolved sources based on information maximazation, Proc. IEEE Wkshp. Neural Networks Signal Processing, Kyoto, Japan, Sept. 1996, pp. 423-432.

[3] H. Jeong, Y. Kim : Parallel feedback network architecture for blind source separation, ELECTRONICS LETTERS, 19th, August 2004, vol. 40, No. 17.