

Hybrid Window 알고리즘을 이용한 마이크로폰 어레이 시스템의 성능 개선

An Enhancement of Microphone Array System Using Hybrid Window Algorithm

이학주*, 김기만**, 이원철***, 차일환*, 윤대회*, 이충용*

* 연세대학교 전기·컴퓨터공학과, ** 해양대학교 전파공학과

*** 숭실대학교 정보통신전자공학부

Hak-Ju Lee*, Ki-Man Kim**, Won-Cheol Lee***, Il-Whan Cha*, Dae-Hee Youn*,
Chungyong Lee*

* Dept. of Electrical and Computer Engineering, Yonsei University

** Dept. of Radio Science and Engineering, Korea Maritime University

*** School of Electronic Engineering, Soongsil University

E-mail) midus@cyclon.yonsei.ac.kr

요약문

본 연구에서는 화자의 음성신호를 이용하여 추출된 공간정보를 통해 화자의 위치를 실시간으로 추적하는 시스템을 제안하고 실시간 구현하였다. 기존의 대표적인 화자 위치 추출 알고리즘인 CPSP(Cross Power Spectrum Phase)는 실내환경에서 심각하게 일어나는 반향신호에 취약한 단점을 갖고 있으므로 구현된 시스템에서는 위치 추적 성능 개선을 위하여 반향신호에 강한 hybrid window 알고리즘을 제안하여 적용하였다. Hybrid window 알고리즘은 실내 환경에 적합한 hybrid window를 설계하여 수신된 음성신호에 적용함으로써 반향신호에 의한 상호 상관관계를 줄이고 직접 경로에

의한 신호들의 상관관계를 높임으로써 보다 정확한 시간 지연 추정을 가능하게 한다. 제안된 시스템의 성능 분석을 위해 DSP를 이용해 실시간 구현된 하드웨어를 이용해 기존의 CPSP 알고리즘과 제안된 hybrid window를 적용한 시스템을 실제 환경에서의 실험하였고 제안한 알고리즘을 적용한 시스템이 22% 이상 성공적으로 화자의 위치를 추적하였다.

1. 서론

최근 유·무선 네트워크의 발전으로 다자간 원격회의에 대한 관심이 높아지고 있다. 이러한 원격회의 시스템에서 원하는 화자의 위치를 추정함으로써 원하지 않는 화자의 영향을 줄이고 원하는 화자의 음성만을 선택적으로 취득하는 기술은 보다 현실감있고 생동감있는 회의를 가능케 하는 중요한 기술이 된다. 본 논문에서는 화자의 위치를 실시간으로 추적, 기준점으로부터 화자 위치를 좌표화하는 기존 시스템의 성능을 개선시켜 향

본 연구는 한국과학재단의 99특정기초사업 지원으로 이루어졌습니다.

과제번호 : 1999-1-302-010-2

후 반향 및 잡음신호 제거 및 빔형성기법을 이용한 고품질 음성신호 취득에 적용 가능한 시스템 개발을 그 목적으로 한다.

화자의 위치를 추정해내는 방법으로 활발히 연구되고 있는 분야로 지연 시간(TDOA: Time Difference Of Arrival) 추정 알고리즘[1]을 들 수 있다. 이는 수신신호 간의 도달 시간 차이를 추정함으로써 음원의 공간정보를 추출하는 방법인데, 이러한 추정기법은 공간 내 주변 잡음 및 벽면 등에 반사되어 입사하는 반향신호와 같은 추정 성능 저하 요인을 갖고 있다. 최근에는 이러한 요인들에 강인한 지연 시간 추정 기법들이 연구되고 있지만[2,3], 주로 주변 잡음만을 고려한 성능 향상 알고리즘에 대한 연구가 대부분을 차지하고 있다.

본 논문의 구성은 다음과 같다. 2장에서는 위치 추정 알고리즘으로 MCPSP(Modified Cross Power Spectrum Phase) 함수를 이용한 지연 시간 추정 기법을 소개하고 있으며, 3장에서는 제안된 성능 개선 방안인 hybrid windowing 기법을 제시한다. 4장에서는 실제 환경에서의 실험을 위한 하드웨어 제작 및 실험 결과를 다루며 마지막으로 5장에서 결론을 맺는다.

2. 화자 위치 추정 알고리즘

2개의 마이크로폰에 수신되는 신호를 $x_1(t), x_2(t)$ 라 할 때, 수신 신호는 다음과 같이 나타난다.

$$\begin{aligned} x_1(t) &= s(t - \tau_1) + n_1(t) \\ x_2(t) &= s(t - \tau_2) + n_2(t) \end{aligned} \quad (1)$$

식(1)에서 $s(t)$ 는 화자의 음성 신호, τ_i 는 음원으로부터 i 번째 마이크로폰까지의 도달 시간, $n_i(t)$ 는 i 번째 마이크로폰에 수신되는 잡음 신호 및 반향신호등의 음향 잡음을 의미한다. 이 때, $n_1(t), n_2(t)$ 는 i.i.d.이며 음원 신호 $s(t)$ 와 서로 상관관계가 없다고 가정한다.

두 마이크로폰간에 도달하는 입사신호의 상호상관관계를 구하면 마이크로폰간의 도달시간차이인 $\tau_1 - \tau_2$ 에서 최대값을 갖게 된다. 이를 TDOA(Time Difference Of Arrival)라 하며 신호원의 공간정보를 담고 있는 중요한 성분이 된다.

$$R_{x_1 x_2}(\tau) = \mathcal{F}^{-1}(X_1(w) \cdot X_2^*(w)) \quad (2)$$

이 때, $X_i(w) = \mathcal{F}(x_i(t))$, \mathcal{F} 는 Fourier 변환, \mathcal{F}^{-1} 는 Fourier 역변환을 의미한다.

그러나, 사전에 신호에 대한 아무런 통계적 특성을 갖고 있지 못한 경우, 단순한 상호상관관계보다는 식(3)과

같이 정규화된 상관관계를 사용함으로써 잡음의 영향을 줄일 수 있다. 이렇게 정규화된 상관관계를 나타내는 것으로 다음과 같은 MCPSP(Modified Cross Power Spectrum Phase) 함수를 정의한다[4].

$$MCPSP_{12}(\tau) = \mathcal{F}^{-1} \left(\frac{X_1(w) \cdot X_2^*(w)}{(|X_1(w)| |X_2(w)|)^p} \right) \quad (3)$$

이때, p 는 정규화되는 정도를 나타낸 상수로 실내의 반향 정도에 따라 조정되며 보통 0.75 정도의 값을 사용한다.

이렇게 구한 MCPSP 함수의 최대값 인덱스로부터 TDOA를 찾아낸다.

$$TDOA_{12} = \max_{\tau} MCPSP_{12}(\tau) \quad (4)$$

위와 같이 추정한 TDOA에는 신호원의 공간정보를 갖고 있으므로 충분한 쌍으로 구성된 마이크로폰 구조를 설정하고 이로부터 각기 다른 TDOA를 추정, 신호원의 위치를 추적하게 된다.

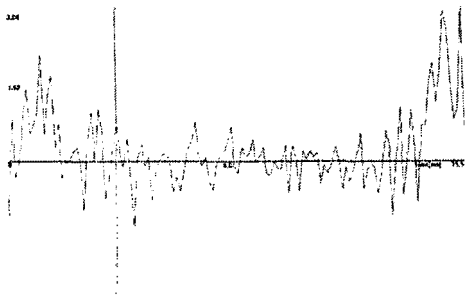
3. 제안된 성능 향상 알고리즘

도달 시간 차이(TDOA)를 이용한 신호원의 위치 추정 알고리즘은 공간내의 반향신호에 취약한 단점을 가지고 있다. 따라서 별도의 반향 제거 기법을 필요로 하는데, 이를 위하여 아래와 같은 hybrid window 알고리즘을 제안하였다.

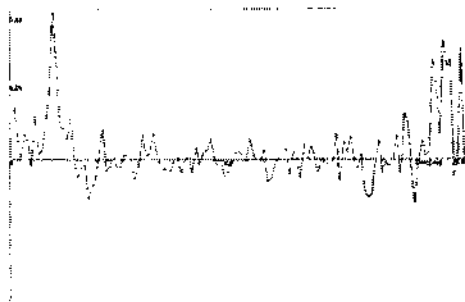
일반적으로 반향신호는 직접경로에 의한 신호에 비하여 긴 도달 시간을 가지고 있다. 따라서 단조감소하는 형태의 Window를 수신신호에 곱함으로써 반향신호에 의한 MCPSP 왜곡을 최소화할 수 있다. 이 때, 직접경로 신호에 의한 상관관계를 최대화하기 위하여 일정구간만큼 지속구간을 둔다. 지속구간의 폭과 단조감소폭은 각각의 마이크로폰의 배치 및 구조에 따라 최적의 값으로 결정한다. 이를 일반화시키면 다음과 같다.

$$w_i(n) = \begin{cases} e^{a_{i1}(n-N_{i1})} & n \leq N_{i1} \\ 1 & N_{i1} \leq n \leq N_{i2} \\ e^{-a_{i2}(n-N_{i2})} & n \geq N_{i2} \end{cases} \quad (5)$$

이때, $w_i(n)$ 은 i 번째 마이크로폰에 적용하는 윈도우 함수이고, N_{i1}, N_{i2} 은 i 번째 마이크로폰에 수신되는 직접경로 신호의 도달 시간 범위를, a_{i1}, a_{i2} 는 i 번째 마이크로폰 윈도우에 적용되는 감소폭을 나타낸다. 제안된 알고리즘에 의한 모의 실험결과는 그림 1과 같다.



(a) 기존 알고리즘



(b) 제안된 알고리즘

그림 1. Hybrid Window의 성능 분석

위의 그림은 시간축에서 MCPSP함수를 나타낸 것이며 실제 TDOA값은 가로축의 왼쪽부분에서 나타난다. 제안된 알고리즘이 적용되지 않은 기존 알고리즘에서는 (a)에서 보는 바와 같이 반향의 영향으로 인하여 잘못된 공간정보인 오른쪽 부분에서 최대값을 갖고 있음에 비하여 제안된 알고리즘을 적용한 (b)에서는 정확한 정보인 왼쪽 부분에서 최대값을 갖고 있음을 알 수 있다.

4. 하드웨어 제작 및 실험 결과

실제 환경에서 실험을 위하여 아래와 같은 시스템을 구성하였다.

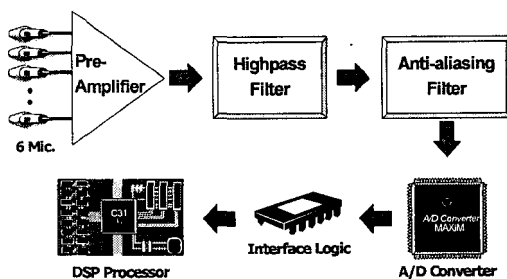


그림 2. 하드웨어 구성도

6개의 단방향성 콘덴서 마이크로폰으로부터 음성 신호를 취득, 각각의 수신신호를 3000배 내지 5000배 정도로 증폭한다. 이후 음성대역만을 선택하기 위하여 고역 통과 필터 및 저역통과 필터를 거쳐 333Hz~3.33kHz대역의 신호만을 추출하고 다채널 12-bit A/D Converter를 사용하여 10kHz 표본화된 신호로 바뀌어 DSP 보드로 전송된다.

프로세서로는 25 MIPS를 갖는 Texas Instruments사의 TMS320C31 부동 소수점 DSP를 이용하였고, 내부 메모리 2Kword에 외부 메모리 128Kword를 추가하였다. 최대 연산량 점유율은 81%정도이며 블록 프로세싱을 거치기 때문에 빠른 DSP를 요구하지 않는 장점을 갖고 있다.

아래 그림은 구현된 알고리즘의 블록 다이어그램이다.

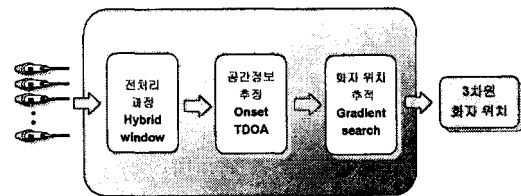


그림 3. 구현된 알고리즘의 블록 다이어그램

6개의 마이크로폰으로부터 들어오는 수신신호는 충분한 음성구간을 취득하기 위한 버퍼링과정을 거치게 되고 성능 개선을 위한 전처리 과정을 거쳐 이 중에서 가장 효율적인 5개의 마이크로폰쌍을 형성, 식(3)과 같은 MCPSP함수를 구성하게 된다.

위와 같이 구현된 하드웨어로 다음과 같은 실제 환경에서 실험을 거쳐 제안한 알고리즘의 성능 개선을 평가하였다. 이 때 시스템의 해상도를 결정짓는 중요한 요소가 되는 마이크로폰의 배치 및 간격은 아래 그림과 같다. 화자의 위치는 그림에서와 같이 마이크로폰의 중심에서부터 수직으로 2m, 3m되는 지점과 수평으로 1m간격으로 각각 5군데씩, 10군데로 고정하였다.

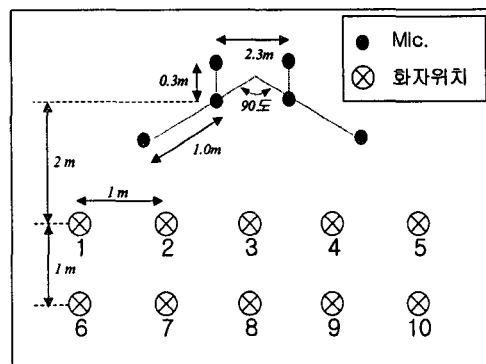


그림 4. 실험 환경

실험결과와 그림 5,6과 같이 허용오차 50cm 범위로 화자 위치를 추정하는 확률로 나타내었다. 이러한 오차는 비교적 정확하게 추정하였다고 여길 수 있는데, 제안된 알고리즘을 적용하지 않은 기존 알고리즘의 경우에는 그림 5와 같으며 평균적으로 56%의 정확도를 갖는다. 그러나 제안된 알고리즘을 적용한 경우에는 그림 6과 같이 평균 78% 정확도를 나타내 제안된 알고리즘이 22% 정도 성능 향상을 가져오고 있음을 알 수 있다.

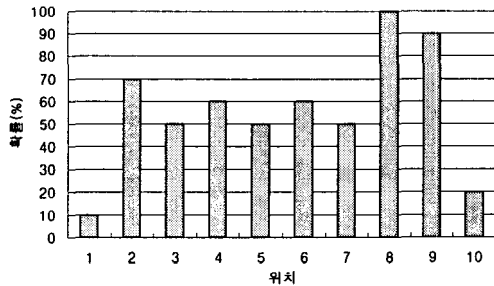


그림 5. 기존 알고리즘의 실험 결과

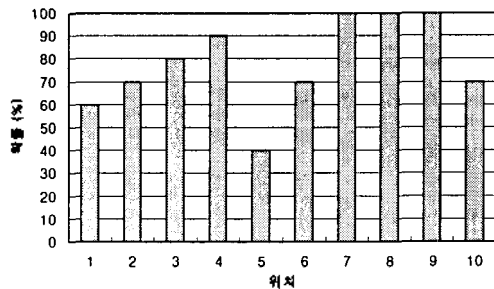


그림 6. 제안된 알고리즘의 실험 결과

위의 결과에서 특히 1,5,6,10번 위치와 같이 벽과 가까운 위치에서는 다른 지점에 비해 두드러진 성능 저하를 보이는데, 이러한 원인으로서는 콘크리트벽, 또는 유리창에 의한 공간 반향 신호를 들 수 있다. 즉 화자의 음성이 직접 경로에 의해 수신되는 신호보다 벽 또는 유리창에 반사하여 수신하는 모든 반향신호의 합이 오히려 더 커지고 있음을 의미한다.

5. 결론

본 논문에서는 기존의 위치 추정 알고리즘으로부터 반향의 영향에 강인한 전처리 과정을 거침으로써 전체적인 시스템의 성능을 향상시키는 알고리즘 개발 및 실제 환경에서의 실험을 위한 하드웨어를 제작하여 통합된 형태의 시스템 개발을 다루었다.

도달 시간 지연 차이를 이용한 기존의 위치 추적 알고리즘은 우수한 분해능을 가짐에도 불구하고 반향에

민감한 단점을 가지고 있는데, 이를 보완하기 위하여 본 논문에서는 각각의 마이크로폰에 수신되는 음성신호에 다른 형태의 윈도우를 적용하여 반향에 의한 상관관계를 최소화시키는 hybrid windowing 기법을 제안하였다. 또한 실제 환경에서 실험을 위해 Texas Instrument사의 TMS320C31 DSP 프로세서를 이용하여 위치 추적 시스템을 실시간으로 구현하였고, 실험 결과 제안된 알고리즘은 기존의 알고리즘에 비하여 위치 추적 정확도면에서 22%이상 우수한 성능을 갖고 있음을 확인하였다. 한편, 시스템의 성능 저하를 가져오는 주된 요인으로 여전히 벽 또는 유리창에 의한 반향을 들 수 있음을 확인하였다.

따라서 향후 잡음 뿐만 아니라 반향 신호에 강인한 지연 시간 추정 알고리즘의 연구 및 반향 제거 알고리즘 등의 연구가 병행하여 진행된다면 더욱 높은 성능의 위치 추정이 가능할 것으로 기대된다.

참 고 문 헌

- [1] Chales H. Knapp, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, no. 4, pp.320-327, Aug. 1976.
- [2] Y. T. Li and A. L. Kurkjian, "Arrival Time Estimation Using Iterative Signal Reconstruction from the Phase of the Cross-Spectrum," *Proc. of the Computer-Aided Seismic Analysis and Discrimination*, pp.87-91, Aug. 1981.
- [3] Renbiao Wu, "Super Resolution Time Delay Estimation via MODE- WRELAX," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 35, no. 1, Jan. 1999.
- [4] Daniel V. Rabinin, *Optimum Sensor Placement for Microphone Arrays*, New Brunswick, May 1998.
- [5] Richard James Renomeron, *Spatially Selective Sound Capture for Teleconferencing Systems*, New Brunswick, Oct. 1997.