

반사 공간에서 고음질의 소리를 얻기위한 Matched Filter Array (MFA) 처리 기법

노 용 주* 윤 종 략*

* : 부경대학교 정보통신공학과

Matched Filter Array Processing for High-Quality Sound Capture

Ro, Yong-Ju* Yoon, Jong-Rak*

* : Pukyong National University Dept. of telematics engineering

E-mail : royju@woongbi.pknu.ac.kr jryoon@dolphin.pknu.ac.kr

Abstract

강당, 회의실, 강의실 등과 같은 닫혀진 공간에서 포착된 소리의 질은 반향음과 간섭 잡음 등에 의해 영향을 받는다. 마이크로폰에 수신되는 신호는 직접전달과 벽면에 의한 반사파들이 더해지므로 해서 실제 발생음을 명확히 얻기가 어렵다. 수신측 마이크로폰에서 반사음의 영향을 제거하면 실제 발생음에 가까운 양질의 소리를 얻을 수 있을 것이다.

잡음과 반향음의 영향이 큰 음향 공간에서 고음질의 소리를 얻기 위한 방법으로 마이크로폰 배열의 병렬 신호 처리 기법이 있다.

본 연구에서 제시된 마이크로폰 배열의 병렬 신호처리 기법은 공간적 음량 선택성을 제공하기 위해 각 마이크로폰 센서들의 matched filter 처리와 병렬 처리 기법을 결합한다. 이 기법은 다중경로 왜곡(반향)과 간섭 잡음을 제거하는 수단을 제공한다.

1. 서론

Matched filter(정합 필터) 동작은 SNR 향상과 채널 왜곡의 감소를 위한 최적의 신호처리 기법이다. 밀폐된 실내 공간의 신호전송에 있어서, 공간은 하나의 필터로 생각할 수 있다. 정합 필터는 공간의 임펄스 응답의 시역전 함수이다. 화상회의 환경에서, 마이크로폰 시스템에 수신된 음성신호는 다중경로 왜곡(반향)과 간섭 음향잡음에 의하여 음질이 떨어진다. 수신된 낮은 음질의 음성신호로부터 다중경로 왜곡과 간섭 음향잡음을 제거하여 고음질의 원음과 같은 신호를 얻기위한 방법으로 마이크로폰 배열 병렬 신호처리 기법이 제안되어 있다. 이 기술은 적당한 반향환경의 강당과 화상회의실 등에서 고음질의 소리를 얻기 위한 효과적 방법을 제공한다.

배열 시스템은 공간적으로 위치한 마이크로폰으로 구

성되어 있고 각 마이크로폰의 출력은 정합필터 처리되어 병렬 처리 기법에 의해 더하여 진다. Matched Filter Array(MFA) 처리기법은 간섭 잡음의 주요 원인이 발생음의 반사된 에너지인 경우의 높은 반향 환경에서 더욱 효과적이다. 정합필터는 음향 공간의 임펄스 응답의 시역전 함수와 주파수 응답의 복소 공액으로 정의되는데 밀폐된 음향공간에서 반향에너지를 제거한다.

본 논문에서는 정합필터 계수를 구현하여 반향에너지를 제거하는 MFA 처리 기법의 이론적인 내용을 살펴보고 반향환경의 모델링에 의한 수치 모의실험과 실제 강의실에서 실시된 실험결과에 대해 살펴보도록 한다.

2. Matched Filter Arrays(MFA) 이론

정합필터는 잡음환경하에서 신호수신에 최적의 기술로 알려져 있다.

밀폐된 공간에서 음원 $s(t)$ 가 수신 마이크로폰에 의해 수신된 신호는 다음식으로 나타내어 진다.

$$r(t) = \int s(\tau)h(t-\tau)d\tau = s(t)*h(t) \quad (1)$$

여기서 $h(t)$ 는 밀폐된 음향공간모델의 전송함수이다.

정합필터는 각 마이크로폰의 임펄스 응답의 시역전함수 $h(-t)$ 로 정의된다.

따라서, 마이크로폰에서 정합 필터된 신호 $y(t)$ 는 다음과 같이 얻을 수 있다.

$$y(t) = r(t)*h(-t) = s(t)*h(t)*h(-t) \quad (2)$$

N개의 MFA의 출력은 각 정합필터 출력의 합으로써 정의된다.

$$\begin{aligned} y_o(t) &= \sum_{i=0}^N y_i(t) \\ &= \sum_{i=0}^N s(t)*h_i(t)*h_i(-t) \\ &= s(t)*\sum_{i=0}^N h_i(t)*h_i(-t) \end{aligned} \quad (3)$$

$y_i(t)$: 각 마이크로폰에서 정합필터된 신호
 $h_i(t)$: 각 마이크로폰의 임펄스 응답
 이상의 MFA 이론을 이산신호식으로 표기하면,
 음원 $s(t) \rightarrow S(z) = a_0 + a_1z + a_2z^2 + \dots$
 모델의 전송함수

$$h(t) \rightarrow H(z) = h_0 + h_1z + h_2z^2 + \dots$$

마이크로폰에서 수신된 신호

$$r(t) \rightarrow R(z) = r_0 + r_1z + r_2z^2 + \dots$$

따라서,

$$R(z) = S(z)H(z) \quad (4)$$

$$Y(z) = R(z)H(z^{-1}) = S(z)H(z)H(z^{-1}) \quad (5)$$

여기서 정합필터 $H(z^{-1})$ 는 $H(z)$ 의 공역복소수이다. 그러므로 N개의 MFA 출력은 다음식과 같다.

$$Y_o(z) = S(z) \sum_{i=0}^N H_i(z)H_i(z^{-1}) \quad (6)$$

MFA의 가장 큰 장점은 수신된 신호로부터 반향효과를 제거하는 것이다. MFA에 대한 SNR은 반사의 수에 독립적이다.

만약 음향 환경에서 K개의 반사가 있다면, 단일 임펄스 여기에 대한 정합 필터의 출력은 K^2 개의 임펄스의 합으로 나타난다. 여기서 비왜곡된 신호를 형성하는 K는 중심에서 나타나고 반향 혹은 부가잡음 신호를 형성하는 나머지 $K(K-1)$ 는 여러 시간에 분포한다.

N-채널 MFA에서, NK^2 의 NK 는 시간축 중심에서 나타나고 반향잡음으로 남은 $NK(K-1)$ 는 여러 시간축에 분포한다. 이러한 개념을 Fig. 2-1에서 설명한다. 여기서 $NK(K-1)$ 만큼의 반향 신호의 적은 수가 동일한 시간 지수에서 나타난다고 가정한다.

따라서 MFA 처리시의 SNR은 다음과 같다.

$$SNR = \frac{(NK)^2}{NK(K-1)} = \frac{NK}{K-1} \quad (7)$$

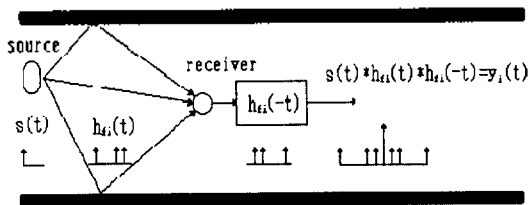


Fig. 2-1 Effect of matched filtering in an enclosure.

그러나 실제 공간에서 임펄스 응답을 구하는 것은 쉬운 일이 아니다. 실제 음향공간 임펄스 응답 $h(t)$ 대신 유사 임펄스 응답 $e(t)$ 를 사용하여 정합 필터링을 수행한다. 실제 임펄스 응답과 유사 임펄스 응답의 비가 1이 되면 정확한 정합필터 계수를 구할 수 있다. 이 $e(t)$ 는 실제 음향공간의 기하학적인 정보로부터 실험을 통하여 구할 수 있다.

유사 임펄스 응답

$$e(t) \rightarrow E(z) = e_0 + e_1z + e_2z^2 + \dots$$

따라서, 정합 필터는 $E(z)$ 의 공역복소수이다.

$$E(z^{-1}) = (e_0 + e_1z^{-1} + \dots + e_{-M}z^{-M})z^M \quad (8)$$

여기서, M은 필터의 유한길이 (만약, 필터의 길이가 길면 과도한 계산시간으로 인하여 실시간 구현이 어렵다), 그리고 z^M 은 Causal한 필터를 생성하기 위해 삽입된 M개의 고정 시간 지연을 의미한다.

유사 임펄스 응답에 의한 각 마이크로폰의 정합필터된 신호는 다음과 같다.

$$M(z) = R(z)E(z^{-1}) \quad (9)$$

그러므로 N개의 MFA의 출력은 다음식과 같다.

$$M_o(z) = \sum_{i=0}^N R_i(z)E_i(z^{-1}) \quad (10)$$

3. 수치모의실험

식 (2-10)를 이용한 MFA 처리에 의한 음질의 향상율을 나타내기 위하여 마이크로폰으로 수신된 신호와 MFA 처리된 신호의 SNR의 차를 계산하여 이를 SNR 향상값으로 정의하였다. SNR 계산을 간단히 하기 위하여, MFA 처리된 혹은 처리되지 않은 출력신호를 다음과 같이 정의한다.

$$r(t) = \alpha s(t-d) + n(t) \quad (11)$$

여기서 $s(t)$ 는 원 음성신호이고, d 는 시간 지연값, α 는 전송에 따른 감쇠 상수 그리고 $n(t)$ 는 간섭신호, 배경잡음, 반사등에 의한 부가잡음이다.

따라서 SNR은 다음과같이 계산되어 진다.

먼저, 감쇠상수의 역수는 다음과 같다.

$$\frac{1}{\alpha} = \sqrt{\frac{\sum_{t=0}^N s^2(t)}{\sum_{t=0}^N r^2(t)}} \quad (12)$$

N : $s(t)$ 의 길이, N_r : $r(t)$ 의 길이
 그러므로 원 음성신호에 대한 비례 출력신호

$$r_s(t) = \frac{1}{\alpha} r(t+d) \text{ 이다.}$$

따라서, 부가잡음은 $n(t) = r_s(t) - s(t)$ 이다.

결과적으로 SNR은 다음식과 같다.

$$SNR = 10 \log_{10} \frac{\sum_{t=0}^N s^2(t)}{\sum_{t=0}^N n^2(t)} \quad (13)$$

수치모의실험은 Ray 모드 계산식으로 밀폐된 공간에서의 신호전송 모델을 구성하여 수신신호를 예측하였고 예측된 수신신호로 정합필터링을 수행하였다. 이때에 신호전송시에 벽에 의한 흡수와 모서리 부분에서의 회절현상에 의한 효과는 무시하고, 신호감쇠는 거리에 반비례한다고 가정하였다.

수치모의실험의 음향공간 모델은 가로 7m, 세로 6m,

높이 3m의 정방형 구조이며, 각각 (0, 3.05, 1.5), (0, 3.25, 1.5), (1.05, 0, 1.6), (1.25, 0, 1.6) m 위치의 4개의 마이크로폰 배열을 가정하였다. 그리고 음원은 KBS 시험관 음성자료의 발음예문 「바람과 햇님」 중 남성이 발음한 “한 나그네가 따뜻한 외투를 입고 걸어 왔습니다.”(지속시간:3.5sec, 샘플링주파수:16Hz)를 사용하였다.

수치모의실험은 (1)각 정합된 음원위치에 대한 SNR 향상값 계산과 (2)정합된 음원위치와 비정합된 음원위치에서의 SNR 향상값 비교로 수행하였다. 각각에 대하여 실험 음원간의 위치는 2m와 0.3m로 하였다.

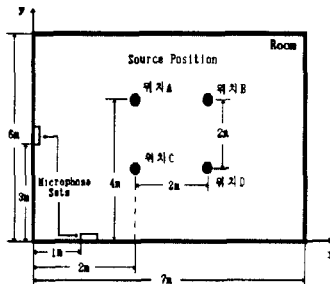


Fig. 3-1 The trial source position and microphone sets position in computer simulation

Fig. 3-1은 위치A (2.0, 4.0, 1.4), 위치B (4.0, 4.0, 1.4), 위치C (2.0, 2.0, 1.4), 위치D (4.0, 2.0, 1.4)의 4개의 실험대상 음원위치와 마이크로폰 배열 위치를 보여주고 있는데 음원간의 거리는 2m로 넓게 한 경우이다. 그리고 Fig. 3-2는 1m 이하의 음원위치 변화에 대한 MFA 처리의 민감도를 실험하기위해 음원간의 거리를 0.3m로 좁게한 경우의 16개의 실험대상 음원위치를 나타내고 있다. 그리고 수치모의 실험에서 정합필터의 길이는 65 msec로 일정하게 하였다.

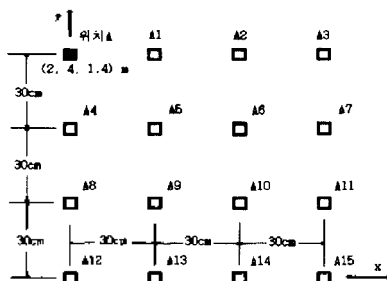


Fig. 3-2 Source positions in the computer simulation for testing the sensitivity of the MFA to changes in source position of less than 1 meter.

<표 1>은 각 반향시간에 따른 Fig. 3-1의 위치에서의 평균 SNR 향상값을 보여주고 있다.(정합된 음원위치에서의 모의실험 결과) <표 1>의 결과에서 이론상으로는 반향시간이 클수록 SNR 향상값이 크게 나타나야 하지만 반향시간 70msec 이상인 경우의 SNR 향상값은 감소하였다. 이것은 정합필터 길이가 65msec

로 제한되었고 채널 수가 4개로 제한된 결과로 판단된다. <표 2>는 Fig. 3-2에서 정합 음원위치 'A'와 비정합 음원위치들의 SNR 향상값의 관계를 보여주고 있다. 표에서 각 음원들이 정합위치에서 멀어질수록 SNR 향상값은 급격하게 떨어짐을 알수 있다. 이것은 MFA 처리가 공간적 신호 선택성을 가짐을 보여준다.

반향 시간 (msec)	평균 SNR 향상값 (dB)
30	3.51
50	4.42
70	4.61
100	4.48

<표 1> MFA 처리된 음성의 평균 SNR 향상값

음원위치(거리)	반향시간(msec)		
	30	70	100
A (0.00m)	2.62	3.61	3.60
A1 (0.30m)	-1.51	-1.80	-1.80
A2 (0.42m)	-4.05	-3.71	-3.63
A3 (0.60m)	-2.51	-2.39	-2.38
A5 (0.67m)	-4.72	-4.80	-4.72
A6 (0.85m)	-3.28	-3.08	-2.99
A7 (0.90m)	0.00	-0.30	-0.30
A10 (0.95m)	-2.70	-2.86	-2.81
A11 (1.08m)	-3.23	-2.61	-2.56
A15 (1.27m)	-1.88	-1.91	-1.91

<표 2> 정합위치와 비정합위치의 SNR 향상값 (dB)

4. 실험 및 결과

실제 실험은 밀폐된 강의실에서 행하였다. 실험 강의실의 크기는 가로 13.89m, 세로 7.31m, 높이 2.93m이며, 각각 (0.13, 3.05, 1.5), (0.13, 3.15, 1.5), (1.05, 0.13, 1.6), (1.15, 0.13, 1.6)m 위치의 4개의 마이크로폰 배열을 사용하였다.

실제 강의실에서의 실험은 수치모의실험과 같이 (1) 각 정합된 음원위치에 대한 SNR 향상값 계산과 (2)정합된 음원위치와 비정합된 음원위치에서의 SNR 향상값 비교를 행하였다. Fig. 4-1은 실제 강의실 실험에서의 음원위치와 마이크로폰 배열의 위치를 보여주고 있다.(음원의 높이는 0.93m로 일정) 정합필터의 길이는 65 msec로 일정하게 하였고, 데이터는 16 kHz의 샘플링 주파수로 샘플링하였다.

<표 2>은 각 음원위치에 따른 SNR 향상값을 보여주고 있다. 모든 위치에 대한 평균 SNR 향상값은 대략 1dB 정도로 수치모의실험에서 보다 많이 낮았다. 이것은 실제 강의실에서는 정합필터의 길이(65 msec) 이상의 시간에서 많은 반사가 일어나며, 강의실의 크기가 크고 배경잡음이 많은 음향환경이기 때문으로 설명되어 질 수 있다.

<표 3>는 정합위치와 비정합 음원위치와의 SNR 향상값의 차이를 보여준다. 그림에서 정합위치 'C1'와

가장 큰 SNR 향상값의 차이는 3.4m 떨어진 위치 'C3'에서 대략 0.8dB 정도로 수치모의실험에서와 같은 공간적 신호 선택성을 나타내지는 않았다. 이상과 같이 강의실 실험에서 좋은 결과를 얻지 못한 것은 스피커와 마이크론의 지향성과 주파수대역 등의 실험장치 특성과 주변잡음 등의 환경적 요인으로 인해 정확한 정합필터를 구현하지 못한 결과로 판단된다.

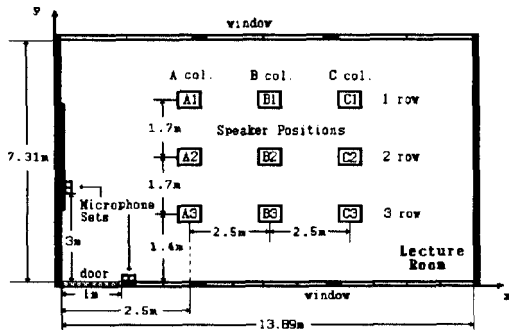


Fig. 4-1 The trial source position and microphone sets position in a lecture room

	A 열	B 열	C 열
1 위치	0.95	0.88	1.27
2 위치	0.81	0.94	1.12
3 위치	1.06	0.83	0.84

<표 2> 실제 강의실에서의 각 위치에 정합된 4-채널 MFA에 대한 평균 SNR 향상값 (dB)

	A 열	B 열	C 열
1 위치	0.74	0.51	<1.27>
2 위치	0.82	0.81	1.02
3 위치	0.79	0.57	0.50

<표 3> C열 1위치에 정합된 4-채널 MFA에 대한 각 위치에서의 평균 SNR 향상값 (dB)

5. 결론

수치모의 실험의 반향 음향공간 모델에서 MFA 처리의 SNR 향상은 반향시간이 30msec부터 100msec까지 대략 4dB로 일정하였다. 따라서 MFA 처리기법은 음향공간에서 다중경로 왜곡의 제거에 효과적으로 사용될 수 있다. 그러나 실제 강의실에서는 음질이 거의 향상되지 않았다. 이것은 유사 임펄스응답을 정확히 구하지 못한 것으로 판단된다. 지금 현재, 실제 음향공간에서 유사 임펄스응답을 정확히 임기위한 방법의 연구가 진행되고 있다.

이 MFA 처리 기법이 화상회의 통신시스템 등에 응용되기 위해서는 앞으로 더 많은 연구가 진행되어야 한다. 먼저, 실제 화상회의실에서 좁은 정합영역을 얻기 위해 전체 회의실에 대한 많은 정합필터 계수를 구할 필요가 있다. 이것은 실시간 구현에 있어 큰 걸림돌로 작용하므로 요구되는 계수를 정확히 계산하고 저장하는

효과적인 방법의 개발이 필요하다. 또한 주파수, 필터 길이, 마이크론 배열 위치 등과 관계되는 MFA 동작의 더 정밀한 수학적 모델이 필요하다. 그리고 MFA 처리 기법은 공간적인 음원위치 정보가 정확히 구해져야 할 필요가 있다. 따라서 정확한 음원 위치 추적알고리즘의 연구도 뒤따라야 할 것이다.

참고문헌

- [1] R.J. Renomeron, D.V. Rabinkin, J.C. French and J.L. Flanagan, "Small-scale matched filter array processing for spatially selective sound capture", 134th meeting of ASA, December 5, 1997.
- [2] H.F. Silverman, W.R. Patterson, J.L. Flanagan, and D.V. Rabinkin, "A digital processing system for source location and sound capture by large microphone arrays", ICASSP-97, Vol. I, pp 251-254, 1997
- [3] D. Rabinkin, R. Renomeron, A. Dahl, J. French, M. Bianchi, and J. Flanagan. "A DSP implementation of source location using microphone arrays", SPIE Annual Meeting, 1996
- [4] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A closed-form method for finding source locations from microphone-array time-delay estimates", ICASSP-95, Vol. 5, pp 3019-3022
- [5] C. Sydow, "Broadband beamforming for a microphone array", JASA, 96(2), pp 845-849, 1994
- [6] C. S. Clay and Saimu Li, "Time domain signal transmission and source location in a waveguide : Matched filter and deconvolution experiments", JASA, 83(4), pp 1377-1383, 1988.
- [7] S. Li and C. S. Clay, "Optimum time domain signal transmission and source location in a waveguide : experiments in an ideal wedge waveguide", JASA 82(4), pp 1409-1417, 1987.
- [8] C. S. Clay, "Optimum time domain signal transmission and source location in a waveguide", JASA, 81(3), pp 660-664, 1987.
- [9] J. L. Flanagan, J. D. Johnson, R. Zahn, and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms", JASA, 78(5), pp 1508-1518, 1985
- [10] E. C. Shang, "Passive harmonic source ranging in waveguides by using mode filter", JASA, 78(1), pp 172-175, 1985.
- [11] N. Aoshima, "Computer-generated pulse signal applied for sound measurement", JASA, 69(5), pp 1484-1488, 1981