

AMDF를 이용한 화자위치 추정

송도훈*, 차경환**, 김천덕***

*부경대학교 대학원 전기공학과, **동서대학교 전자기계공학부, ***부경대학교 전기공학과

Speech source estimation using AMDF

Do-Hoon Song*, Kyung-Hwan Cha**, Chun-Duck Kim***

* Dept. of Electrical Eng. Graduate School Pukyong National Univ.

** Faculty of Electronics and Mechanical Eng. Dongseo Univ.

*** Dept. of Electrical Eng. Pukyong National Univ.

요 약

본 연구에서는 원격 화상회의 시스템 등에서 Camera를 자동적으로 제어하기 위해 화자의 음성신호를 4개의 마이크로폰 배열(Microphone Array)로 수음하여 그 신호에 의해 화자의 위치를 추정한다. 마이크로폰으로 수음한 음성신호의 TDE(Time Delay Estimation)를 계산할 때 그 연산량을 감소시키기 위해 AMDF 알고리즘을 사용한다. 각 마이크로폰 출력신호에 대해 AMDF 알고리즘으로 시간지연을 구하고 DOA(Direction of Arrival)를 계산한다. 그리고 다시 공간 기하계산을 통해 공간내 화자의 위치를 추정한다.

시험 신호로써 음성신호 '아' 음을 사용한 수직 시뮬레이션과 반사음이 존재하는 일반 강의실에서 아나운서가 발생하는 음을 사용하여 AMDF 알고리즘을 이용한 화자위치 추정의 정확도를 조사하였다.

1. 서 론

최근 통신기술과 고속모뎀 개발에 의해 기존의 CATV망은 멀티미디어 통신망으로 사용 가능해 지고 시간적 경제적으로 큰 장점을 가지는 원격 화상회의 시스템이 실용화되고 있다.[1]

원격 화상회의에서 회의 참가자들이 친밀감을 느끼기 위해서는 상황에 맞는 Camera 제어가 필요하다. 회의 참가자들에 의한 수동적 Camera 제어나 외부 부착물을 착용해야 하는 기존의 적외선 기법은 매우 번거롭다. 따라서 화자의 음성신호를 이용하여 능동적으로 Camera를 제어하는 연구가 진행되고 있다.[2]

화자의 위치에 따른 Camera 제어를 위해 마이크로폰 배열의 각 마이크로폰간 지연시간 추정을 하고 다시 수평각, 수직각, 공간 3각 계산을 통해 음원의 3차원 위치를 추정할 수 있다. 이동하는 피사체에 대한 자동초점 기법 및 배경과 피사체 경계의 흐림 방지 기법이 실용화 단계에 있으므로 음원까지 거리계산을 자동초점 기능으로 대체할 수 있을 것으로 판단된다.[2][3]

TDE를 위해 화자의 음성신호를 수음하여 상호상관(Cross-Correlation) 함수나 Cross Power Spectrum을 적용하면 송신, FFT, IFFT연산을 해야 하므로 연산량이 크게 증가하고, 하드웨어로 구현할 때 송신에 대한 고속 연산을 위해 DSP(Digital Signal Processor)나 Signal Hold를 위한 많은 양의 레지스터(Register)를 필요로 한다. 본 연구에서는 마이크로 프로세서에 의한 하드웨어 구현을 위해 제안된 AMDF(Average Magnitude Difference Function) 알고리즘을 이용한다.[4][5]

II. 마이크로폰 배열과 DOA 추정

Fig. 1.은 4개의 마이크로폰 2차원 배열을 나타낸 것이다. 실음장에서는 회의실 벽면의 재질에 따라 다양한 경로의 반사음이 존재한다. 배열을 구성하는 각 마이크로폰 특성으로 인한 신호 왜곡은 무시하고 외부잡음을 고려한다. 마이크로폰 배열의 각 마이크로폰에 입력되는 신호를 표현하면 식 (1)과 같다.

$$P_i(n) = \alpha S(n-q_i) + h_i(n)*s(n) + n_i(n)$$

$$P_j(n) = \beta S(n-q_j) + h_j(n)*s(n) + n_j(n) \quad (1)$$

여기서 $S(n)$ 는 평면파로 가정한 음원신호, α 와 β 는 음원과 마이크로폰간의 거리에 따른 감쇄 상수이다. 그리고 $h_i(n)*s(n)$, $h_j(n)*s(n)$ 는 다중 경로를 통한 반사음, $n_i(n)$, $n_j(n)$ 는 음원신호와 상관성이 없는 외부잡음을 나타낸다. Fig. 1.에서 (Mic_1, Mic_2) (Mic_3, Mic_4) (Mic_1, Mic_3) (Mic_2, Mic_3) 간의 TDE는 다음 식(2)에 의해 계산된다.

$$AMDF(q) = \sum_{n=0}^{N-1} |P_i(n) - P_j(n+q)|$$

$$q_{ij} = \arg \min AMDF(q) \quad (2)$$

TDE를 위한 Cross-Correlation 알고리즘과 AMDF 알고리즘의 연산량을 비교하면 Table 1.과 같다.

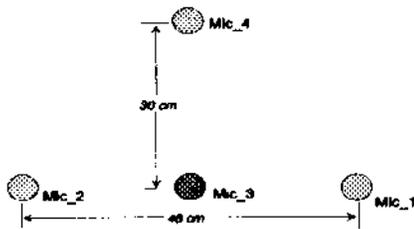


Fig. 1. 4개의 마이크로폰 2차원 배열

Table 1. 두 알고리즘의 연산량 비교

	Cross-correlation 알고리즘		AMDF 알고리즘	
	실수	복소수	실수	복소수
승산	$(2N+1)N$	$4(2N+1)N$	0	0
가산	$(2N+1)(N+1)$	$2(2N+1)(N+1)$	$(2N+1)(N+1)$	$(2N+1)(N+1)$
감산	0	0	$(2N+1)N$	$2(2N+1)N$
비교	승산을 감산으로 대체			

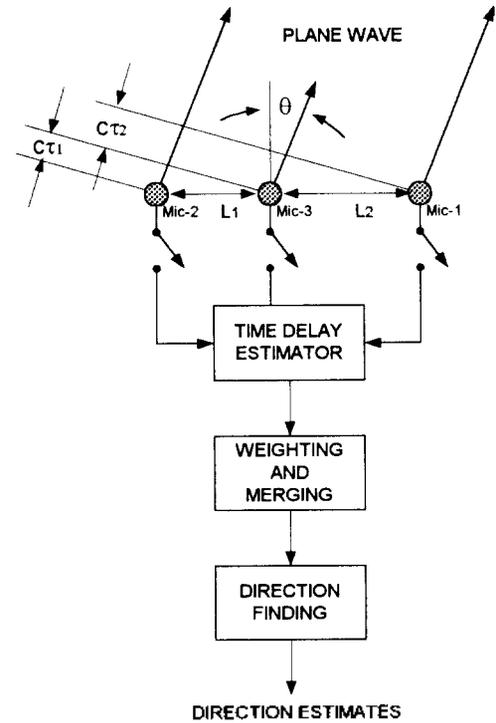


Fig. 2. 음원의 DOA 추정

Fig. 2.는 AMDF 알고리즘을 이용하여 두 마이크로폰 사이의 상대적인 TDE를 계산하고 Cosine 법칙을 이용하여 음원의 DOA를 추정한다. 마이크로폰 사이의 시간지연 q 는 식(3)에 의해 계산된다.

$$q_1 = \frac{R_l - R}{c} = \frac{1}{c} \{ [L_1^2 + R^2 + 2L_1R\sin\beta]^{1/2} - R \}$$

$$q_2 = \frac{R - R_{III}}{c} = \frac{1}{c} \{ R - [L_2^2 + R^2 - 2L_2R\sin\beta]^{1/2} \} \quad (3)$$

여기서 R_l , R , R_{III} 는 음원과 Mic_2, Mic_3, Mic_1과의 거리이고 C 는 공기중 음속이다. 음원을 평면파로 가정한다면 식(3)은 Taylor 급수 전개에 의해 식(4)와 같다.

$$q_1 = \frac{1}{c} (L_1 \sin\beta + L_1^2 \cos^2\beta / 2R)$$

$$q_2 = \frac{1}{c} (L_2 \sin\beta - L_2^2 \cos^2\beta / 2R) \quad (4)$$

식(4)에서 $\cos^2\beta/2R$ 을 연립하여 풀면 다음 식 (5),(6)과 같다.

$$\sin \beta = \frac{c}{2L} \left[\frac{L_2}{L_1} q_1 + \frac{L_1}{L_2} q_2 \right] \quad (5)$$

$$\beta = \sin^{-1} \frac{c}{2L} \left[\frac{L_2}{L_1} q_1 + \frac{L_1}{L_2} q_2 \right] \quad (6)$$

본 연구에서는 식(6)의 L_1, L_2 가 0.23[m]로 잡고, AMDF 알고리즘에 의해 지연시간을 구하면 식(6)에 의해 DOA를 추정할 수 있다.

III. 단음절 음성신호에 의한 수치 시블레이션

AMDF 알고리즘을 이용하여 TDE를 위해 단음절 음성신호 '아' 음에 대한 시블레이션을 하였다. 수치 시블레이션 조건은 다음과 같다.

두 마이크로폰간 거리는 0.46[m]로 하고 샘플링 주파수는 12kHz, 샘플 데이터수 N는 2048개이고 각 데이터는 16bit A/D변환기에 의해 획득된다. 음원은 마이크로폰 배열면의 법선 방향을 기준으로 +62.5°에 위치한다고 가정한다.

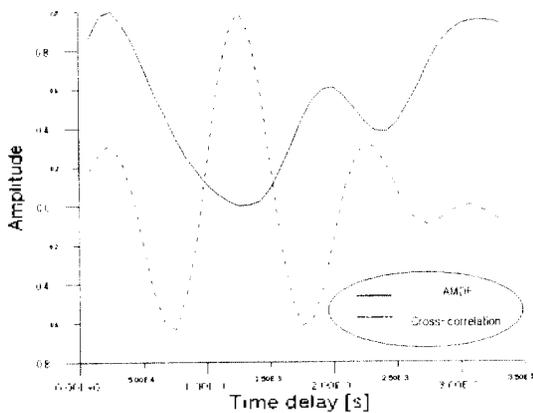


Fig. 3. '아' 음에 대한 AMDF 와 Cross-correlation Algorithm의 TDE

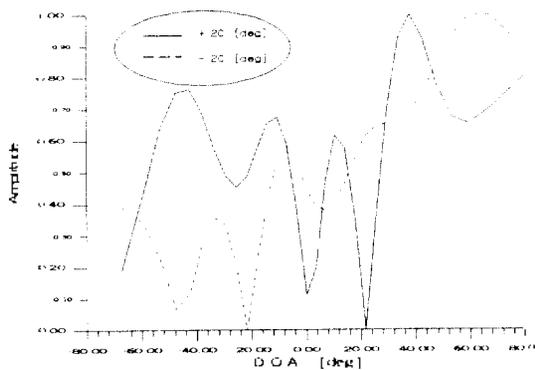


Fig. 4. AMDF에 의한 양방향 DOA 추정

Fig. 3.은 상호상관함수 알고리즘과 AMDF 알고리즘에 대한 TDE를 비교한 그림으로 1.25ms 지점에서 Peak가 나타난다. Fig. 4.는 음원이 +20.2°와 -20.2°에 존재할 때 AMDF 알고리즘에 의한 DOA 추정 결과로써 양방향 모두 DOA를 추정하고 있다. 수치 시블레이션 결과 TDE, DOA 추정에서 AMDF 알고리즘의 유용함을 알 수 있었다.

IV. 일반 강의실에서의 실험

회상회의 공간과 유사한 조건으로 테이블과 벽면에 의한 반사가 존재하는 일반 강의실(462×870×250[cm])에서 실험을 하였다. 실험시 압소음은 57[dB]이다. Fig. 5.는 음원과 마이크로폰 사이의 거리를 1m로 했을 때 White Noise에 대해 Cross Spectrum법으로 측정된 강의실내 임펄스 응답이다.

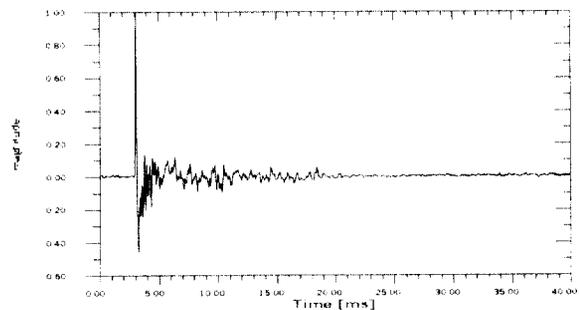


Fig. 5. 일반 강의실에서 음원과 Mic.1사이의 임펄스 응답

화자 위치 추정 실험에 사용된 측정 마이크로폰은 부지향성인 B&K4130이고 프리앰프는 B&K2642이다. Fig. 6.은 음성 신호를 획득하는 과정과 TDE를 계산하는 능력도이다. 먼저 마이크로폰 배열로 수음한 음성신호를 12kHz로 샘플링 하여 16CH DAT(SONY PC216AX)에 녹음한 후 전용 인터페이스(PCIF-5)를 이용하여 퍼스널 컴퓨터(PentiumII)에 저장하고 AMDF 알고리즘을 이용하여 TDE를 계산한다.

Fig. 7.은 DOA 추정을 위해 +20°방향에 SNR이 20[dB]인 Radio News(여자 아나운서)를 음원으로 사용하였다. Fig. 4.의 수치 시블레이션은 두 마이크로폰에 수음되는 음성신호의 음압이 같으므로 AMDF의 최소치가 0이 된다. 그러나 실제 음장에서는 음원과 마이크로폰사이의 거리에 따른 감쇄(α, β) 때문에 두 마이크로폰에 수음되는 음성신호의 음압이 달라지므로 AMDF의 최소치가 0이 되지 않는다.

AMDF 알고리즘은 TDE를 위해 최소치까지의 지연시간을 계산하므로 그래프 최소값이 0이 아니더라도 TDE를 구할 수 있다.

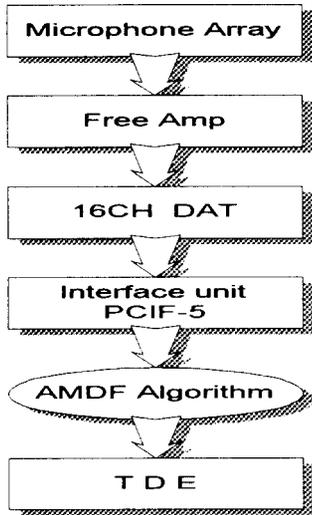


Fig. 6. 음성 신호처리 블록도

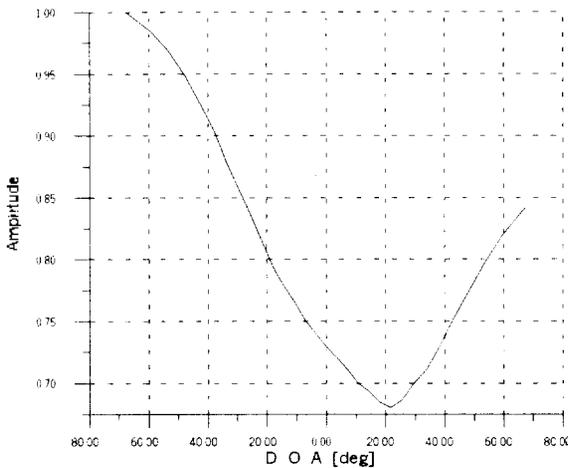


Fig. 7. 실음장에서 Radio News가 음원일 때 DOA 추정

V. 결론

화상위치 추정을 위한 TDE를 구하기 위해 제안된 AMDF 알고리즘은 Cross-Correlation 알고리즘에 비해 승산 대신 간산을 수행하므로 마이크로 프로세서에 의한 하드웨어 구현이 용이하다. 단음절 음성신호 '아' 음에 대한 수치 시뮬레이션과 일반 상의실 실험을 통해 정확한 지연시간 추정을 확인하였다. 하지만 프레임 선택에 따라 TDE 정확도가 약간 차이를 보인다.

MAC등을 응용한 TDE 분해능 향상과 Camera와 마이크로폰 배열면의 상대적 위치를 고려한 실제 Camera 제어시스템 구성에 대해 연구되어야 할 것이다.

<참 고 문 헌>

- [1] 송분규, 이상설, "무선 CATV망의 원격 관리 및 제어 시스템의 개발", 한국통신학회논문집 Vol 23, No 4 p1134-1147, 1998.
- [2] Hong Wang and Peter Chu, "Voice source localization for automatic camera pointing system in videoconferencing", In Proceeding of the ICASSP IEEE, p187-190, 1997.
- [3] 장성익, 이용만, 정기욱, 김원찬, 이충웅, "자동추적 조절 기법 및 하드웨어의 구현", 한국통신학회 논문집 Vol 20, No 2 p502-517, 1995.
- [4] 송도훈, 차경환, 김천덕, "음원위치 추정 시스템에서 속도향상을 위한 AMDF의 적용", 한국음향학회 영남지회 학술발표회 논문집 Vol 5, p64-67, 1998.
- [5] 백승관, "음원위치 추정을 위한 마이크로폰 설계", 부경대학교 석사학위논문, 1998.
- [6] L.R.Rabiner and R.W.Schafer, Digital processing of speech signals, Englewood Cliffs, Prentice Hall, New Jersey, 1978.
- [7] G.clifford carter, "Coherence and Time Delay Estimation", proceedings of the IEEE, Vol. 75, no. 2, p236-255, Feb. 1987.
- [8] C.H.Knapp and G.C.Carter, "The Generalized Correlation Method for Estimation of Time delay", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol 24. ASSP, August 1976.