1999년도 한국음향학회 학술발표대회 논문집 제18권 제2(s)호

Onset 알고리듬을 이용한 시 지연(TDOA) 추정 알고리듬의 개선에 관한 연구

김선영*, 박규식*
*상명대학교 컴퓨터·정보·통신학부
충남 천안시 안서동 산 98-20
전화) 0417-550-5350, Fax) 0417-550-5370
E-mail : sykim@smuc.sangmyung.ac.kr
kspark@smuc.sangmyung.ac.kr

요 약

본 연구는 2개의 마이크로폰을 이용한 지연 시간(TDOA) 추정에 대한 연구로서 기존의 Cross-Correlation 알고리듬과 Roth Processor 알고리듬에 Onset 알고리듬을 적용하여 기존 알고리듬의 성능 개선에 대한 연구를 수행하였다. 작은 사무실이나 회의실내 주변 잡음과 음향반향이 동시에 존재한다는 가정 하에 Onset 알고리듬은 마이크로폰의 수신신호에 포함되어 있는 반향 신호를 효율적으로 마스킹하여 반향 현상을 제거함으로서 정확한 시간 지연 추정을 가능하게 한다. Onset을 적용한 Cross-Correlation 과 Roth processor 알고리듬의 우수성을 입증하기 위하여 7개의 SNR 환경에서 총 420번의 컴퓨터 모의 실험을 수행하였으며 실험 결과 SNR이 5dB 이상에서는 Onset 알고리듬을 적용한 알고리듬이 Onset을 적용하지 않은 알고리듬에 비해 우수함을 입증할 수 있었다.

I. 서 론

최근의 마이크로폰 어레이 기술은 음원 추적, 음성 인식, 음성습득, 원격회의 시스템, 비디오 감시 시스템 등 다양한 관련 응용 분야의 기술 발전에 원동력이 되어왔다. 이러한 마이크로폰 어레이 관련 응용 기술들은 공통적으로 다음과 같은 2가지 근본 과정을 밟게 되는데 첫째 2개 혹은 그 이상의 마이크로폰에 수신되는 음원들간의 신호 지연 시간을 추정하고, 둘째 추정된 지연 시간을 이용하여 음원의 방향각과 마이크로폰 배치 기하학을 이용하여 음원의 위치를 추적한다. 마지막으로는 이러한 음원의 위치 정보를 각 응용분야에 적절히 사용하여 기존의 기술들을 향상 발전시키는 것이다. 예를 들면, 음성인식 분야에서는 음원의 위치로부터 고 품질의 음성신호를 받아들여 인식률을 향상시킬 수 있는 방법에 대한 연구가 활발하고[1], 원격회의 시스템 등에서는 사람에 의한 수동적인 비디오 카메라 조절이 아니라 음원의 위치 정보를 이용하여 자동적으로 화자를 비추어 줄 수 있는 음향-화상 융합시스템에 대한 연구가 진행중이다.[2]

마이크로폰 어레이에 수신되는 음원 신호들간의 지연 시간(TDOA: Time Delay of Arrival)추정은 크게 다음과 같이 2가지의 추정 성능 저하 요소를 갖고 있다. 첫째는 회의실이나 사무실 등의 비교적 작은 공간 내에 컴퓨터, 환풍기, 에어콘, 소형 냉장고등으로 인한 비교적 500Hz 이하의 저주파 대역에 퍼져 있는 잡음들로서 정확한 시간 지연 추정을 어렵게 한다. 둘째는 음향 반향 신호인데 이는 폐쇄된 공간 내에서 음원 신호가 벽면을 반사하여 지연되어 마이크로폰에 유입됨으로서 직선 경로의 음원 신호를 방해한다. 따라서 이러한 주변 잡음과 음향 반향 같은 저해 요소들을 효율적으로 제거 혹은 완화하여 정확한 시간 지연 추정이 가능한 알고리듬에 대한 연구가 필수적이라 할 수 있다.

최근까지의 지연 시간(TDOA) 추정 알고리듬의 동향은 폐쇄 공간 내 주변 잡음만을 고려하여 잡음에 강인한 알고리듬에 대한 연구가 대부분을 차지하고 있으며 주변 잡음과 함께 반향 신호의 억제 혹은 완화에 대한 연구는 상대적으로 빈약하다고 할 수 있다. 주변 잡음을 고려한 관련 논문들로는 크게 cross-correlation 함수를 이용하는 방법[3] 과 일반화된 cross-correlation 함수[4][5]를 이용하는 방법으로 다양한 알고리듬들이 개발되어 있다. 반면 주변 잡음 및 반향신호를 동시에 고려한 것으로는 Steve[2] 와 Huang[6]이 제안한 알고리듬이 있다.

본 연구는 음원추적을 위한 지연 시간(TDOA) 추정에 대한 연구로서 작은 사무실이나 회의실내

주변 잡음과 음향반향이 동시에 존재한다는 가정하에 기존의 cross-correlation 알고리듬과 Roth Processor 알고리듬에 Steve의 Onset[6]알고리듬을 적용하여 성능 개선에 대한 연구를 목적으로 한다.

II. 음향 반향 마스킹을 위한 Onset 알고리듬

Onset 신호는 마이크로폰에 수신된 음성 및 음향 신호에 반향 신호가 포함되어 있을 경우 반향신호가 도착하기 전의 원 음성 신호 혹은 반향 신호를 마스킹한 결과 신호만을 일컫는다. 반향이 섞여 있는 마이크로폰 수신신호에서 Onset 신호를 만드는 방법은 수신신호의 첫 번째 silent 구간을 추적하여 그 구간의 신호를 취하는 방법[6] 그리고 수신신호의 시간에 따른 크기 변화를 추적하여 수신신호로부터 반향신호만을 마스킹하는 방법[2]이 알려져 있다. 이중에서도 반향 마스킹 방법은 수신신호의 에너지 변화를 추적하여 반향신호를 검출함으로서 비교적 정확한 결과를 보이고 있으며 본 논문에서도 Steve가 제안한 마스킹 알고리듬을 사용하도록 한다.

Steve가 제안한 반향 마스킹 Onset 알고리듬을 간략히 소개하면 다음과 같다. 먼저 마이크로폰에 수신된 신호로부터 엔벨롭 과정을 거쳐 peak rectifier에 의해 각 마이크로폰 입력신호의 크기 변화 모양을 결정한다. 엔벨롭 과정은 다음과 같은 수식으로 표현된다.

$$ENV(n) = \max(\beta \times ENV(n-1), |x(n)|) \quad (1)$$

수식 (1)에서 $x(n)$은 반향신호를 포함한 2개의 마이크로폰 수신 신호, β 는 엔벨롭 계수를 그리고, $ENV(n)$ 은 엔벨롭 과정후의 엔벨롭 신호를 나타낸다. 이러한 엔벨롭 과정은 큰 폭으로 증가하는 입력신호에 대해 빠르게 증가하는 엔벨롭 신호를 만들어 낸 후 그 뒤에 따르게 되는 반향신호를 마스킹하기 위하여 천천히 감소하는 엔벨롭 신호를 만들어 낸다. 반면에 엔벨롭 신호는 수신신호의 일정한 silence period 이후에 0의 값에 도달하게 된다. 그림 1(b)는 그림 1(a)의 반향 신호가 포함된 단순 sine 신호에 대한 엔벨롭 처리 과정을 보여주고 있다.

엔벨롭 처리 과정 후에는 다음과 같은 Onset 처리 과정을 거쳐

$$\text{if } (ENV(n-1) < ENV(n)) \quad \{ONSET(n) = \max(0.0, ENV(n) - ENV(n-1))\} \quad (2)$$

엔벨롭 신호로부터 증가하는 경사 신호 부분만을 추출한다. 그림 1(c)는 그림 1(b)의 신호에 수식 (2)를 적용한 최종 onset 신호를 보여주고 있다. 그림 1(a) 와 1(b)에서 보는 바와 같이 수신신호에 포함되어 있는 반향신호(2번째 sine 신호)가 잘 마스킹 되는 것을 볼 수 있다.

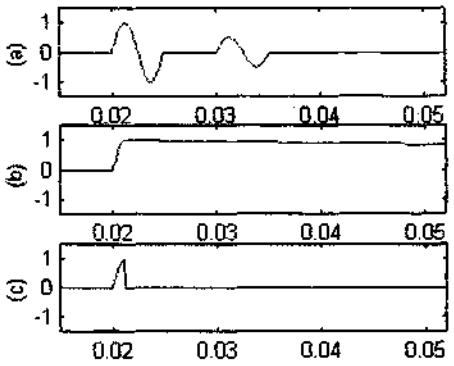

그림 1: Onset 알고리듬의 반향 마스킹 효과
(a) 반향신호를 포함한 단순 sin 신호, (b) 엔벨롭 신호, (c) Onset 신호

III. 시간 지연 추정 알고리듬

2개의 microphone에 수신된 신호로부터 음원의 지연 시간 추정은 크게 두 신호간의 상관성 관계를 이용한 cross-correlation 방법과 신호의 상관성에 가중치를 부가하는 Generalized cross-correlation (Generalized CC:GCC) 2가지로 나누어진다.

먼저 2개의 마이크로폰에 수신되는 신호를 $x_1(t)$, $x_2(t)$이라고 할 때, 수신 신호는 다음과 같이 표현된다.

$$\begin{aligned} x_1(t) &= s_1(t) + n_1(t) \\ x_2(t) &= s_1(t+D) + n_2(t) \end{aligned} \quad (3)$$

수식 (3)에서 D 는 음원신호의 도착 지연 시간, $s_1(t)$ 는 음원 신호, $n_1(t), n_2(t)$는 공간 내에 존재하는 잡음신호와 반향 신호등의 음향 잡음을 의미한다. 여기서 잡음신호 $n_1(t), n_2(t)$는 AWGN(Additive White Gaussian Noise)로 가정되어 음원 신호 $s_1(t)$ 와 비 상관 관계에 있다고 가정한다.

Cross-correlation 방식은 마이크로폰에 수신된 2개 신호의 cross-correlation 함수를 구하여 이 함수를 최대화하는 Time Delay를 계산하는 방식으로, cross-correlation은 다음과 같이 정의된다

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} S_{x_1x_2}(f)\, e^{j2\pi f\tau} d\tau \quad (4)$$

여기서 $S_{x_1x_2}(f)$는 Cross Power Spectrum을 나

타내며, 시간 지연 추정은 수식(2)의 cross-correlation 함수를 최대화하는 지연 시간 D를 추정하는 것이다. Cross-correlation 방식은 공간내의 주변잡음과 반향에 상당히 민감하게 반응하는 단점을 가지고 있다.

Generalized cross-correlation (GCC) 방법은 수식 (4)의 Cross Power Spectrum에 가중함수를 주어 기존의 cross-correlation 방법의 잡음에 대한 단점을 보완하는 것으로서 다음과 같이 정의된다.

$$R_{x_1x_2}^{(g)}(\tau) = \int_{-\infty}^{\infty} \psi_g(f) S_{x_1x_2}(f) e^{j2\pi f \tau} df \qquad (5)$$

여기서 $\psi_g(f)$는 가중함수로서 실제 시간 지연 추정은 가중 Cross Power Spectrum인 $S_{x_1x_2}^{(g)}(f) = \psi_g(f) S_{x_1x_2}(f)$의 역 푸리에 변환 값인 GCC 함수를 최대화하는 값으로 구해진다.

Roth Processor 알고리듬은 수식 (5)의 GCC 함수에서 가중 함수를 다음과 같이 입력신호 $x_1(t)$ 의 Auto-power spectrum의 절대값을 취해줌으로서 정의된다.

$$\psi_R(f) = \frac{1}{|S_{x_1x_1}(f)|}$$
$$R_{x_1x_2}^{(R)} = F^{-1}\left[\frac{S_{x_1x_2}(f)}{|S_{x_1x_1}(f)|}\right] \qquad (6)$$

이 경우 일반화된 잡음 및 음원신호의 비상관성을 가정하면 GCC는 다음과 같이 표현될 수 있다.

$$R_{x_1x_2}^{(R)}(\tau) = \delta(\tau - D) * \int_{-\infty}^{\infty} \frac{S_{s_1s_2}(f)}{(S_{s_1s_1}(f) + S_{n_1n_1}(f))} e^{j2\pi f\tau} df \qquad (7)$$

수식 (7)에서 볼 수 있듯이 Roth processor 알고리듬은 시간 지연을 추정하는 GCC 함수의 최대 피크치가 상대적으로 퍼지게 되는 단점은 있지만, 수신신호의 잡음이 많은 주파수 대역에서 잡음에 의한 spectrum을 억제하는 장점을 가지고 있다.

Ⅳ. 컴퓨터 모의 실험

본 연구에서는 마이크로폰 수신 신호에 주변 잡음과 반향 신호가 존재한다는 가정 하에 기존의 시간 지연 추정을 위한 수식(4)의 cross-correlation 방법과 수식(6)의 Roth processor 알고리듬의 성능 개선을 위해 Onset 알고리듬을 적용하고 컴퓨터 모의 실험을 통하여 성능 평가를 수행한다. 실험 환경은 우선 작은 사무실이나 회의실을 가정하여 인공적인 음원 소스 신호를 발생하고, 이러한 음원 신호에 AWGN 잡음신호와 Eyring의 image model 감쇠 계수와 지연시간을 이용한 반향신호를 더하였다. 이때 2개의 마이크로폰 수신 신호간에 지연 시간추정을 위해 인의의 지연시간을 주어서 2개의 마이크로폰 수신신호를 만들어낸다. 구체적인 마이크로폰 수신신호 발생과 모의 실험 절차는 다음과 같다.

먼저 수식 (1)에 따라 인공적으로 음원 신호 $s_1(t)$, 잡음신호 $n_1(t)$, $n_2(t)$를 발생시키기 위해 샘플링 주파수를 16Khz로 하여 2초 분량의 3개 white gaussian sequence를 MATLAB rpiid() 함수를 이용하여 발생하였다. 이중에서 하나의 white gaussian sequence는 AR(6) 필터를 통과시켜 음성과 유사한 특징을 갖는 신호를 만들었으며 나머지 2개의 white sequence는 잡음신호로 사용하였다. 또한 알고리듬의 성능을 사무실내 잡음뿐만이 아니라 음향 반향 환경 하에서도 고려하기 위해, 음원신호의 크기를 감쇠 하여 인위적으로 지연시켜 반향 신호를 만들었다. 이때 반향신호의 감쇠정도는 Eyring의 image model을 사용하여 반사 계수를 0.5로 하였다.

위에서 만들어진 음원신호, 잡음신호, 반향신호는 서로 더해져서 2개의 마이크로폰에 수신되는 입력 신호들을 만들었으며 이때 두 개의 입력 신호간에 상대적인 시간 지연을 갖도록 인위적으로 5 sample 지연 시간을 갖도록 하였다.

수행된 컴퓨터 모의 실험은 수식(4)의 Basic Cross-Correlation(BCC) 알고리듬과, BCC에 Onset 알고리듬을 적용한 Onset Cross-Correlation(OCC), 그리고 수식 (6)의 Roth Process 알고리듬에 Onset을 적용한 Onset Roth Processor(ORP) 알고리듬에 대해 각각 SNR이 20dB, 15dB, 10dB, 5dB, 0dB, -5dB, -10dB 등 7가지 신호 대 잡음비를 고려하였으며 각 SNR에서 각각 20번의 반복실험을 하여 총 420번의 독립 실험을 반복하였다. 각각의 알고리듬 성능을 비교하기 위해서는 최종 Cross-Correlation 함수로부터 추정된 지연 시간과 PFR(Peak to Floor Ratio)을 계산하였다. PFR은 최종 Cross -Correlation 결과로부터의 최대 피크 치(지연 시간 추정)와 이 값을 제외한 나머지 값들의 합과의 비교를 통해 최대 피크치 근처에 발생하는 사이드 피크 값의 방해 여부를 측정하는 것으로서 지연 시간 추정 알고리듬의 정확성을 평가할 수 있는 중요한 척도가 되며 다음과 같이 정의된다.

$$PFR = \frac{\max_k R_{x_1x_2}^{(G)}(k)}{\sum_{j \neq k} R_{x_1x_2}^{(G)}(j)} \qquad (8)$$

그림 2는 SNR이 10dB에서 각 알고리듬의 최종 cross-correlation 함수를 보여 주고 있다. 모든 알고리듬에서 시간 지연 값에 해당하는 최대

피크치를 비교적 잘 검출하고 있으나, Onset 알고리듬을 적용하지 않은 BCC의 경우 음향반향으로 인한 사이드 피크가 심각함을 볼 수 있다.

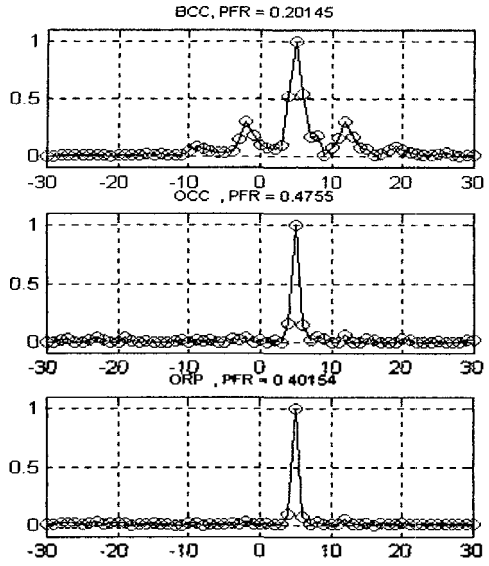

그림 2. SNR=10dB 에서의 Cross-Correlation 결과 함수 비교

그림 3은 SNR에 따른 PFR 값의 변화를 보이고 있다.

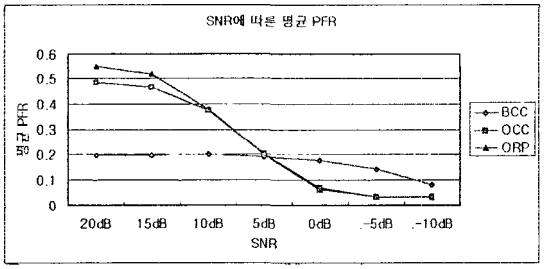

그림 3. SNR에 따른 PFR 값의 비교

그림에서 보는바와 같이 SNR이 10dB 이상인 경우에는Onset 알고리듬을 적용한 ORP 와 OCC 알고리듬 모두 BCC에 비해 우수함을 볼 수 있다. SNR이 10dB 이하인 경우에는 Onset 알고리듬을 적용한 알고리듬 보다 Onset을 적용하지 않은 BCC 알고리듬이 우수함을 볼 수 있다. 이러한 현상은 Onset 알고리듬의 반향신호 마스킹 과정으로부터 제기되는데 이는 마이크로폰 수신 신호가 Onset 과정을 거치면서 급격한 경사 신호(그림 1 참조), 즉 고주파 신호로 변환됨으로 인해서 잡음이 많은 환경 하에서는 원 음성신호의 에너지보다 잡음신호에 의한 오차를 많이 발생하기 때문이다.

V. 결 론

본 연구에서는 기존의 지연 시간(TDOA) 추정 알고리듬의 음향반향으로 인한 성능저하를 개선하기 위하여 기존의 Basic Cross-Correlation 알고리듬과 Roth Processor 알고리듬에 Onset 알고리듬을 적용하여 성능 개선에 대한 연구를 수행하였다. Onset 알고리듬은 마이크로폰의 수신 신호에 포함되어 있는 반향 신호를 효율적으로 마스킹 하여 정확한 시간 지연 추정을 가능하게 함으로서 컴퓨터 모의 실험 결과 Onset을 적용한 Cross-Correlation 과 Roth processor 알고리듬이 onset을 적용하지 않은 알고리듬에 비해 성능이 우수함을 입증할 수 있었다. 반면 비교적 잡음이 많은 환경 하에서의 Onset 알고리듬은 반향신호의 마스킹 과정에서 발생하는 저주파 대역 음원 신호 억제현상으로 인하여 Basic Cross-Correlation 보다 성능이 저하되는 현상을 보이고 있는데, 현재 이에 대한 보완 연구가 진행중이다.

참고문헌

[1] M. Omologo and P. Svaizer, " Talker localization and speech enhancement in a noisy environment using a microphone array based acquisition system", Proc. of Eurospeech, 1993

[2] Steve G., " Multimedia fusion for intelligent camera control and human-computer interaction", Ph.D Thesis, North Carolina Univ., 1997

[3] A. papoulis, Probability, Random Variables and Stochastic Process, New York, McGraw-Hill, 1965

[4] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, Aug. 1976

[5] P. R. Roth, "Effective measurements using digital signal analysis", IEEE Spectrum, vol. 8, Apr. 1971

[6] Jie Huang, Noboru Ohnishi, "A biomimetic system for localization and separation of multiple sound sources", IEEE Trans. on Inst. and Mea., vol. 44, No. 3, June 1995