

인간의 청각 시스템을 응용한 음원위치 추정에 관한 연구

A study imitating human auditory system for tracking the position of sound source

배진만*, 조선희**, 박종국***

(Jeen Man Bae and Sun Ho Cho and Chong Kuk Park)

* 경희대학교 전자공학과(전화:(031)201-2957, 팩스:(031)201-4968, E-mail : marujeen@hotmail.com)

** 경희대학교 전자공학과(전화:(031)201-2957, 팩스:(031)201-4968, E-mail : chosunho@netian.com)

*** 경희대학교 전자공학과(전화:(031)201-2957, 팩스:(031)204-4968, E-mail : ckpark@khu.ac.kr)

Abstract : To acquire an appointed speaker's clear voice signal from inspect-camera, picture-conference or hands free microphone eliminating interference noises needs to be preceded speaker's position automatically. Presumption of sound source position's basic algorithm is about measuring TDOA(Time Difference Of Arrival) from reaching same signals between two microphones. This main project uses ADF(Adaptive Delay Filter) [4] and CPS(Cross Power Spectrum) [5] which are one of the most important analysis of TDOA. From these analysis this project proposes presumption of real time sound source position and improved model NI-ADF which makes possible to presume both directions of sound source position. NI-ADF noticed that if auditory sense of humankind reaches above to some specified level in specified frequency, it will accept sound through activated nerve. NI-ADF also proposes practicable algorithm, the presumption of real time sound source position including both directions, that when microphone loads to some specified system, it will use sounds level difference from external system related to sounds of diffraction phenomenon. In accordance with the project, when existing both direction adaptation filter's algorithm measures sound source, it increases more than twice number by measuring one way. Preserving this weak point, this project proposes improved algorithm to presume real time in both directions.

I. 서론

음원의 위치를 추정하는 기본 알고리즘은 두개 이상의 마이크로폰에 수신되는 신호들을 도착시간지연(Time Difference of Arrival)추정[1]을 통하여 음파의 속도와 시간과의 관계로부터 위치를 음원의 위치를 추정할수 있다.

수신위치에 따른 신호의 시간지연 정보를 이용하는 방법은 지금까지 다양한 알고리즘이 제시되어 왔다. 일반적으로 시간지연 검출에는 크게 두 가지 방법이 있는데 상호상관함수를 이용한 방법과 적응필터를 이용하는 방법으로 분류할 수 있다. 상호상관함수를 이용하는 방법은 비교되는 두 신호 간 지연시간의 합수로 표현되는 상호상관함수에서 피크 값의 위치를 검출함으로써 시간지연 값을 추정하는 방법이며 상호상관함수를 이용하는 대부분의 방법들은 주파수영역에서 두 신호의 상호파워스펙트럼(cross power spectrum)[5]을 구한 다음 신호대 잡음비의 향상을 위해 필터처리한다.

적응필터를 이용한 방법은 두신호의 상대적 시간지연 값을 반복적 알고리즘을 이용하여 결정한다[4].이 적응지연 알고리즘은 에러제곱평균[MSE][6]을 최소화 하는 시간지연 값을 찾기 위해 gradient technique를 사용한다.

위에서 언급한 두가지 알고리즘의 단점은 기준이 되는 마이크로폰이 미리 지정되어야 하기 때문에 단방향 음원 위치추정에 한정된다는 것이다. 결국 두개의 마이크로폰으로 위치추정이 가능한 각도가 0°~90°로 제한된다. 따라서 본 논문에서는 두개의 마이크로폰으로 일반 알고리즘의 두 배의 측정범위, 즉 0°~180°의 측정범위를 갖는 알고리즘을 제안한다. 또한 적응필터 알고리즘에서 양방향 음원위치 추정을 위해 단방향 추정 보다 두 배 많은 가중치와 계산시간이 필요한 단점을 마이크로폰 간 음압레벨을 이용함으로써 기준이 되는 마이크로폰을 선 지정하여 단방향 추정과 비슷한 계산시간으로 양방향 추정이 가능한 신경발화 적응필터 알고리즘을 제안하였다.

이때 음압 레벨 차는 마이크로폰들이 어떠한 시스템에 탑재될 경우 마이크로폰 사이에 음의 회절현상과 관련하여 음압레벨차가 발생하는 것에 착안한 알고리즘이다.

본 논문에서는 인간의 형상을 가진 휴먼로봇에 적용시키기 위한 음원위치 추정알고리즘을 제안한다.

II. 시간지연 추정 기본 알고리즘

마이크로폰이 수신하는 비 상관 잡음이 포함된 동일 신호 사이의 상대적 지연시간을 구하는데 상관함수

(correlation function)[2] 를 이용 할 수 있다.

한 음원에서 발생하는 신호를 분리된 두 마이크로폰이 비상관 잡음과 함께 수신하는 과정은 다음과 같이 수학적 모델로 표현 할 수 있다.

$$\begin{aligned} r_1(t) &= a_1 s(t - \tau_1) + n_1(t) \\ r_2(t) &= a_2 s(t - \tau_2) + n_2(t) \end{aligned} \quad (1)$$

두 마이크로폰에서 받은 신호의 상호상관함수는

$$R_{r_1, r_2}(\tau) = E[r_1(t)r_2(t - \tau)] = \int_{-\infty}^{\infty} r_1(t)r_2(t - \tau)dt \quad (2)$$

이고 여기서 E[]는 기댓값(시간평균)을 나타낸다.

그러나 실제로는 무한구간에서 측정 할 수 없으며 유한관찰시간에 대해서 식(2)을 근사화 한다.

$$R_{r_1, r_2}(\tau) = \frac{1}{T - \tau} \int_0^{T - \tau} r_1(t)r_2(t - \tau)dt, \quad 0 < \tau \leq T \quad (3)$$

위에서 적분을 수행하기 전에 $r_1(t)$ 와 $r_2(t)$ 를 그림1과 같이 필터 H_1, H_2 를 각각 통과시키면, 신호 y_i 는 필터 H_i 를 통과하여 출력 y_i ($i=1,2$)이 된다.

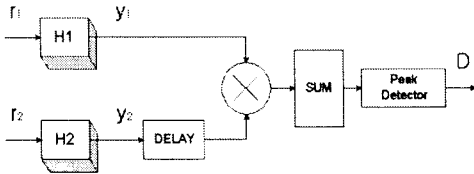


그림1. Flow chart for a peak detector based on the generalized correlation method

다음은 입력에 관한 상호상관 스펙트럼이다.

$$R_{r_1, r_2}(\tau) = \int_{-\infty}^{\infty} G_{r_1, r_2}(f)e^{j2\pi f\tau}df \quad (4)$$

여기서 필터를 통과한 y_1, y_2 의 상호스펙트럼은 다음과 같이 나타낼수 있다.

$$G_{y_1, y_2}(f) = H_1(f)H_2(f)G_{r_1, r_2}(f) \quad (5)$$

결과적으로 주파수 영역에서의 상호상관함수는 다음과 같다.

$$R_{y_1, y_2}^{(g)}(\tau) = \int_{-\infty}^{\infty} \phi_g(f)G_{r_1, r_2}(f)e^{j2\pi f\tau}df \quad (6)$$

여기서 $\phi_g(f) = H_1(f)H_2(f)$ 는 "일반화된 주파수 가중치(general frequency weighting)"라고 정의 하며 신호와 잡음에 관련된 일정 주파수 특성을 강조 하여 시간지연 측정이 용이하게 정할 수 있다.

III. CPS 알고리즘

이 알고리즘은 두개의 마이크로폰에서 수신된 신호의 cross power spectrum의 성질 중 위상에 관련된 정보만으로 주파수영역에서의 신호의 시간지연 값을 구하는 방법이다.

$$G_{r_1, r_2}(f) = \alpha_1 \alpha_2 G_{ss}(f)e^{-j2\pi f\tau} + G_{n_1, n_2}(f) \quad (7)$$

상상관방법에 의해 두신호 사이에 포함된 잡음이 서

로 비상관인 신호 일 때 위상은 다음과 같다.

$$R_{y_1, y_2}^{(g)}(\tau) = \int_{-\infty}^{\infty} \frac{G_{r_1, r_2}(f)}{|G_{r_1, r_2}(f)|} e^{j2\pi f\tau}df \quad (8)$$

$$\frac{G_{r_1, r_2}(f)}{|G_{r_1, r_2}(f)|} = e^{-j\theta(f)} = e^{-j2\pi fD} \quad (9)$$

그러므로 식 (8)은 다음과 같이 쓸수 있다.

$$R_{y_1, y_2}^{(g)}(\tau) = \delta(\tau - D) \quad (10)$$

식 (11)는 CPS를 구하는 결과식이다.

$$R_{r_1, r_2}(\tau) = F^{-1} \left\{ \frac{F\{r_1(t)\}F^*\{r_2(t)\}}{|F\{r_1(t)\}| |F\{r_2(t)\}|} \right\} \quad (11)$$

IV. ADF 알고리즘

적용필터에 의한 시간지연 검출은 두 신호 간의 차이인 추정에러신호 e_n 을 최소로 수렴해 가기 위해 필터의 특성을 매 샘플마다 반복적으로 보정하여 필터의 시간지연 변수 d 가 실제의 지연 D 에 접근하도록 하는 원리이다.

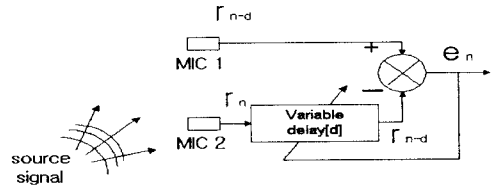


그림 2. TDOA using an adaptive filter

그림2에 대한 수식적 해석은 다음과 같다.

$$e(n) = r(n - D) - r(n - d) \quad (12)$$

이며, 여기서 얻은 에러신호에 자승을 취해 파워성분으로 변환하면 다음과 같다.

$$\begin{aligned} e^2(n) &= \{r(n - D) - r(n - d)\}^2 \\ &= r^2(n - D) - 2r(n - D)r(n - d) + r^2(n - d) \end{aligned} \quad (13)$$

양 변에 기대값을 취하면,

$$E[e^2(n)] = 2R_{rr}(0) - 2R_{rr}(D - d) \quad (14)$$

여기서 R 는 신호 r 의 자기상관함수이다. 따라서 에러자승평균 $E[e^2(n)]$ 는 입력신호의 자기상관함수에 의해 완전히 구해짐을 알 수 있다. 결국 식 (14)에서 구해진 d 가 두 신호 사이의 시간지연을 나타낸다.

다음은 최소평균자승(Least Mean Square)알고리즘으로서 가장 널리 사용되는 적응 알고리즘이다.

가장 큰 특징으로는 계산이 간결해서 알고리즘 구현이 용이하다는 점이다. 적응제어 시스템에서는 보통 유한 임펄스응답 필터 (FIR Filter)를 사용한다. FIR필터는 다음과 같은 시스템 함수를 갖는다.

$$H(z) = b_0 + b_1 z^{-1} + \dots + b_M z^{-M} = \sum_{n=0}^M b_n z^{-n} \quad (15)$$

그러므로 임펄스 응답 $h(n)$ 은 $0 \leq n \leq M$ 구간에서

수열 b_n 이므로 출력을 입력의 차분방정식 표현으로 나타내면 다음과 같다.

$$y(n) = b_0 r(n) + b_1 r(n-1) + \dots + b_M r(n-M) \quad (16)$$

$$r[n] = [r(n), r(n-1), \dots, r(n-M)]^T \quad (17)$$

여기서 필터의 계수벡터를 w 라하면 출력 $y(n)$ 는 다음과 같다.

$$y(n) = r(n)^T w = w^T r(n) \quad (18)$$

y_n 또한 순간에러 e_n 는 회망응답 d_n 와 필터 출력와의 차에 의해서 구해지므로

$$e(n) = d(n) - y(n) = d(n) - w^T r(n) \quad (19)$$

이다. 따라서 에러자승평균 $E[e^2(n)]$ 는

$$E[e^2(n)] = E[d^2(n)] + w^T E[r(n)r(n)^T] w - 2E[d_n r(n)] w \quad (20)$$

식(20)의 우변 제2항, 제3항의 $E[r(n)r(n)^T]$, $E[d_n r(n)]$ 는 각각 필터입력벡터의 자기상관행렬 R 과 입력과 회망응답과의 상호상관벡터 P 로 표현할 수 있다. 따라서

$$MSE = J = E[d^2(n)] + w^T R w - 2P^T w \quad (21)$$

여기서 다시 필터계수의 워너해를 찾기 위한 알고리즘은 다음과 같다.

$$\begin{aligned} \phi(n) &= \frac{\partial e^2(n)}{\partial w} = 2e(n) \frac{\partial e(n)}{\partial w} \\ \frac{\partial e(n)}{\partial w} &= -r(n) \end{aligned} \quad (22)$$

$$w(n+1) = w(n) + \mu(-\phi_n) = w(n) + 2\mu e(n)r(n)$$

여기서 μ 는 보정량에 대한 가중치로 수렴상수라고 하며 수렴상수에 따라서 수렴속도와 안정도가 결정된다.

μ 는 임의의 상수이지만 최적의 워너해에 접근하기 위해서는 $0 < \mu < \frac{2}{\lambda_{\max}}$ 를 만족해야 한다.

여기서 λ 는 자기상관행렬 R 의 고유치를 나타낸다. 신호의 분산정도에 따라 고유치가 비례하므로 분산정도가 높은 음성신호 같은 경우 수렴속도가 느려질 것을 알 수 있다.

IV. 양방향 추정 ADF 알고리즘

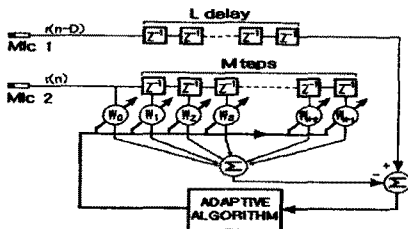


그림 3. 양방향 추정 ADF diagram

기준이 되는 마이크로폰에 인위적인 지연을 발생시킴으로써 양방향 음원위치추정이 가능하도록 하는 것으로서 인위적 지연의 결정은 다음과 같다.

$$e(n) = r(n-D-L) - r(n-d) \quad (23)$$

이때 MSE는 다음과 같이 나타낼수 있다.

$$e(n)^2 = r(n-D-L) - r(n-d)^2 \quad (24)$$

$$= r^2(n-D-L) - 2r(n-D-L)r(n-d) + r^2(n-d)$$

$$E[e(n)^2] = 2R_{rr}(0) - 2R_{rr}((D+L)-d) \quad (25)$$

MSE값이 0으로 수렴하려면

$$2R_{rr}(0) = 2R_{rr}(D+L-d) \quad (26)$$

$D+L=d$ 가 되어야 한다. 여기서 $L=M/2$

이 경우 샘플링 횟수의 증가로 계산량의 증가가 불가피하다.

V. NI-ADF 알고리즘

본 알고리즘은 두개의 서로 다른 마이크로폰으로부터 수음된 음을 인간의 신경과 같이 작용하는 병렬 다중 대역필터로 주파수 영역에 분리한 후 상관관계가 가장 높은 지연시간을 찾아내는 방법이다. 이것은 대역필터를 통과한 신호가 일반 오디오 시스템의 그래픽 이퀄라이저 같은 펄스 레벨 형태로 표현되기 때문에 상관관계를 구하는데 상당한 계산시간 단축효과를 가져올수 있다. 또한 NI-ADF 알고리즘은 양방향 음원위치 추정을 목적으로 제한한 방법으로 앞서 언급한 알고리즘들이 기준마이크로폰이 결정된 후에 시간지연을 알아내는 방법과는 달리 음압레벨과 관련한 전처리 과정을 통하여 기준마이크로폰을 자동선정 하는 방법을 취한다.

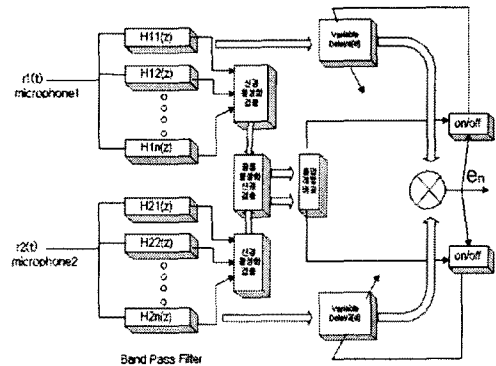


그림 4. NI-ADF Block Diagram

그림에서 보듯이 각 대역필터는 인간의 청각신경과 같이 특정주파수에 일정레벨이상에서 활성화 하여 양쪽 마이크로폰에 들어온 공통 활성화 신경만을 추출하여 두 마이크로폰 간 레벨을 비교하여 레벨이 낮은 쪽을 기준으로 하여 높은 쪽 신호를 지연시키는 방법으로 지연시간을 찾아내게 된다.

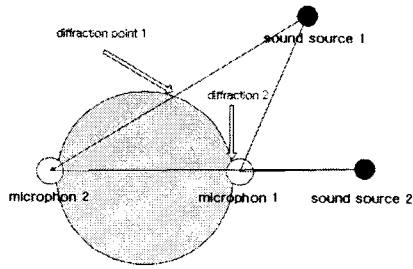


그림 5 음의 회절관련도

음압레벨 차는 그림4에서 sound source 2 에서 음이 microphone1 에는 감쇠 없이 전달되며, microphone2에서는 일정 레벨 이상의 주파수에서 감쇠하는 현상이 일어나게 된다. 로봇의 머리를 보다 긴 파장을 가진 음들은 레벨차가 거의 없으나 지름보다 짧은 파장을 가진 음은 감쇠가 일어나 음압레벨이 발생하게 된다. source1의 경우는 source2의 경우보다 더 짧은 파장이상에서 감쇠가 일어난다. 이를 이용하여 두음의 좌우방향을 먼저검출 하기 때문에 양방향 위치검출에 있어서 NI-ADF는 효과적인 알고리즘이라 할수 있다.

VI. 실험

본 연구의 실험을 위해서 인간의 머리 형상의 지름 20cm의 구에 마이크로폰을 두개를 탑재하였으며 회절 임계점은 1.6KHz로 정하여 음압레벨을 비교하였다. 즉 1.6KHz이상의 대역 필터만을 이용하여 음의 감쇠가 일어나지 않는 공통영역을 제거하여 계산량을 줄이도록 하였다. 검출한 후에 또한 44KHz로 샘플링된 신호를 이용하였기 때문에 Delay Time 은 최대 27개의 샘플링수를 delay 시키는 것으로 가능하다. 음원은 인간의 음성과 음악을 오디오 스피커를 통해 발생하였으며, 실제 인간의 귀 구조와 흡사한 형태로 마이크로폰을 삽입하여 시스템을 구성하였다. 본 논문에서 제안한 NI-ADF와 ADF 알고리즘을 비교하기 위하여 음성 신호에 대해 양방향 측정시간을 비교하였다.

음원추정 시간비교

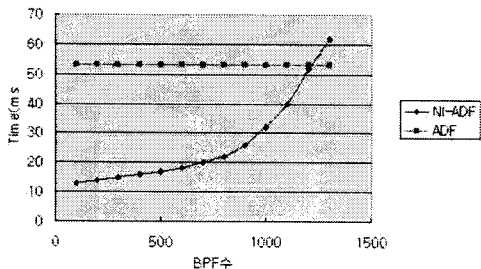


그림 6. 실험 결과

그림 5에서 보듯이 NI-ADF알고리즘에 신경수가 낮은 경우 ADF보다 빠른 처리속도를 나타내지만 그 수가 증가할수록 ADF와 근사한 처리속도를 나타내게 된

다. 본 실험에 사용된 ADF는 1024 FFT를 사용하였기 때문에 신경의 약 1000개 부분에서 비슷한 결과를 나타내고 있다. 그러나 1500개 이상의 부분에서 ADF보다 처리속도가 느려지는 것을 볼 수 있으나 이것은 정확도가 필요이상으로 증가했기 때문에 나타난 것으로 소음제어 같은 정확도가 높은 시스템에 응용될 수 있는 영역으로 사용할 수 있다. 또 이 경우 고정 소음원을 제거 하는데 효과적으로 적용될 수 있다. 그리고 본 실험의 필터는 프로그램의 처리하여 순차적인 처리를 하는 프로그램으로서 병렬 처리를 하도록 설계된 필터에 상당한 시간지연을 가지므로 이것을 하드웨어로 설계하였을 경우는 현재의 결과보다 빠른 응답속도를 보장할 것으로 기대된다.

VI. 결론

본 연구는 임의의 평면상의 음원을 2개의 마이크로폰으로 위치 추정하는 알고리즘을 좀더 발전된 형태로 빠른 응답시간을 보장하는 알고리즘을 제안하였다. 또한 실제 인간의 청각기능이 음의 시간지연과 음압레벨 차 모두를 이용하는데 착안하여 고안된 알고리즘으로 단 방향에 극한 되거나 양방향을 지향하기 위하여 단 방향에 두 배 이상의 처리시간을 필요로 하는 기존의 알고리즘에 비해 양방향 측정 시 빠른 응답을 보장할 수 있는 것을 실험으로 확인하였다. 현재 평면상의 180° 부분의 음원 위치를 추정할 수 있는 본 알고리즘을 인간의 귀바퀴 기능과 같은 전 처리 과정을 통하여 360°로 확장하고 더 나아가 공간상의 음원위치를 추정할 수 있을 것으로 기대된다.

참고문헌

- [1] Michael S. "A Closed-Form Location Estimator for Use with Room Environment Microphone Arrays", IEEE Trans On Acoustic, Speech, and Signal Processing, Vol. 5, No.1, June 1997
- [2] Michael S. "Microphone-array localization error estimation with application to sensor placement", J. Acoust. Soc. Am. 99(6), June 1996
- [3] Pi. Sheng Chang and Alan "performance of 3D Speaker Localization Using a Small Array of Microphone", Proc. IEEE Int.Symp. on Circuits and system, 1998
- [4] Don H. Johnson " array Signal Processing :concepts and Techniques", P T E Prentice Hal, englewood Cliffs, NJ 07632
- [5] Maurizio Omologo and piergiorgio Svaizer, "Acoustic source Localization in Noise and Reververant Environment Using CSP Analysis", Proc IEEE Int. Conf. Acoust., Speech, Signal Processing, Atlanta 1996, Vol. 2, pp. 921-924
- [7] 오상현, 박규식 "subban CPSP를 이용한 음원 추적 시스템에 관한 연구" 전자공학회 논문집 제 23권, 제1호, 2000.