

CPSP의 저주파 위상 복원을 이용한 화자 위치 추적 알고리즘의 성능 개선

An Enhancement of Microphone Array System Using Hybrid Window Algorithm

이학주^{*}, 김기만^{**}, 이원철^{***}, 이충용^{*}

^{*}연세대학교 전기전자공학과, ^{**}해양대학교 전파공학과

^{***}승실대학교 정보통신전자공학부

Hak-Ju Lee^{*}, Ki-Man Kim^{**}, Won-Cheol Lee^{***}, Chungyong Lee^{*}

^{*} Dept. of Electrical and Electronic 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 마이크로폰 어레이를 이용하여 화자의 음성신호로부터 화자의 위치를 추정하는 기존의 대표적인 알고리즘인 CPSP(Cross Power Spectrum Phase)로부터 보다 반향에 강인한 알고리즘인 저주파 위상 복원 알고리즘을 제안한다. CPSP 함수는 상호 상관관계(Cross Correlation)가 정규화 되어있는 형태를 갖는데, CPSP 함수의 최대값 인덱스로부터 화자의 공간정보인 TDOA(Time Difference Of Arrival)를 추출한다. 그러나 CPSP 함수를 이용한 공간정보 추정 알고리즘은 실내환경에서 심각하게 일어나는 반향신호에 대해서 취약한 단점을 갖고 있다. 본 논문에서 제안하는 저주파 위상 복원 알고리즘은 주파수 축면에서 반향신호가 CPSP 함수에 미치는 영향을 분석하여 반향으로 인하여 왜곡된

위상 성분을 복원함으로써 보다 신뢰도 있는 TDOA 추정을 가능하게 한다. 반향신호로 인한 CPSP의 위상은 저주파보다 고주파에서 심하게 왜곡되는데, 각각의 반향 신호의 도달 시간을 기하학적 분포를 갖는 확률변수로 모델링하여 이를 수학적으로 증명하였다. 또한 실제 환경에서 채집한 음성신호를 이용한 모의 실험을 통해 개선된 알고리즘의 성능 개선을 확인하였다.

1. 서론

화자 위치 추적 시스템이란 화자의 음성신호를 이용하여 이로부터 화자의 공간정보를 추정, 실시간으로 화자의 위치를 추적하는 시스템이다. 화자의 공간정보를 추정하는 대표적인 알고리즘으로 음성의 도달 시간 차이(TDOA: Time Difference of Arrival)를 이용하는 방법을 들 수 있다. 두 개의 마이크로폰으로부터 화자의 음성신호를 수신할 때, 화자의 위치로부터 각각의 마이크로폰간에 발생하는 거리 차이로 인하여 도달되는 시간 차이가 발생하게 된다. 이러한 도달 시간 차이를 TDOA

본 연구는 한국과학재단의 99특정기초사업 지원으로 이루어졌습니다.

과제번호 : 1999-1-302-010-2

라 하는데, 화자의 공간정보로 이용되는 중요한 성분이 된다. 널리 알려진 TDOA 추정 기법으로 Omologo와 Svaizer가 제안한 Cross-Power Spectrum Phase (CPSP) 알고리즘을 들 수 있으며[1], 반향 또는 잡음에 따라 정규화 정도를 조정하는 알고리즘도 사용되고 있다[2][3]. 그러나 기존의 CPSP 알고리즘은 실내에서 발생하는 반향 및 잡음에 심각한 성능저하를 보이는 단점을 갖고 있어 본 논문에서는 이러한 반향의 영향에 강인한 알고리즘 개발한다.

2. 기존의 TDOA 추정 알고리즘

그림 1과 같이 두 개의 마이크로폰에 수신되는 신호를 각각 $x_1(t)$, $x_2(t)$ 라고 하면 다음과 같다.

$$\begin{cases} x_1(t) = A_1 s(t - \tau_1) + \sum_{l=1}^L a_{1l} s(t - \mu_{1l}) \\ x_2(t) = A_2 s(t - \tau_2) + \sum_{m=1}^M a_{2m} s(t - \mu_{2m}) \end{cases} \quad (1)$$

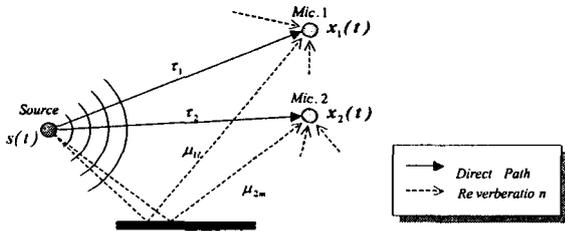


그림 1. 두 개의 마이크로폰으로 수신한 음성신호

이때, $s(t)$ 는 화자의 음성신호이며, A_i 는 화자로부터 i 번째 마이크로폰까지 전달된 정도를 의미한다. 또한 τ_i 는 직접경로에 의한 i 번째 마이크로폰까지의 도달 시간을, μ_{ij} 는 i 번째 마이크로폰에 입사하는 j 번째 반향 경로의 도달 시간을 나타낸다. 따라서 주파수 영역에서 두 수신신호의 CPSP 함수를 구하면 다음과 같다.

$$\begin{aligned} CPSP(w) &= X_1(w) X_2^*(w) \\ &= [A_1 A_2 e^{-j(\tau_1 - \tau_2)w} \\ &\quad + \sum_{l=1}^L A_2 a_{1l} e^{-j(\mu_{1l} - \tau_2)w} + \sum_{m=1}^M A_1 a_{2m} e^{-j(\tau_1 - \mu_{2m})w} \\ &\quad + \sum_{l=1}^L \sum_{m=1}^M a_{1l} a_{2m} e^{-j(\mu_{1l} - \mu_{2m})w}] |S(w)|^2 \end{aligned}$$

(2) 일반적으로 반향신호는 직접경로에 의한 도달 신호보다 작은 전력을 가지므로 각각의 반향경로에 의한 상관관계인 우변의 마지막항은 무시할 수 있다. 또한 식 (3)과 같이 CPSP함수를 정규화함으로써 $|S(w)|^2$ 의 영향을 근사적으로 제거할 수 있다. 이렇게 변형된 MCPSP(Modified CPSP) 함수는 아래와 같다[4].

$$\begin{aligned} MCPSP(w) &= \frac{X_1(w) X_2^*(w)}{|X_1(w)| |X_2(w)|} \\ &\approx [A_1 A_2 e^{-j(\tau_1 - \tau_2)w} \\ &\quad + \sum_{l=1}^L A_2 a_{1l} e^{-j(\mu_{1l} - \tau_2)w} \\ &\quad + \sum_{m=1}^M A_1 a_{2m} e^{-j(\tau_1 - \mu_{2m})w}] \quad (3) \end{aligned}$$

$$\begin{aligned} mcpssp(t) &= A_1 A_2 \delta(t - (\tau_1 - \tau_2)) \\ &\quad + \sum_{l=1}^L A_2 a_{1l} \delta(t - (\mu_{1l} - \tau_2)) \\ &\quad + \sum_{m=1}^M A_1 a_{2m} \delta(t - (\tau_1 - \mu_{2m})) \end{aligned}$$

위 식으로부터 주파수 영역에서의 MCPSP 함수는 두 마이크로폰 각각에 들어오는 직접 경로 신호에 의한 상관관계 뿐만 아니라 반향신호와 직접 경로 신호간의 상관관계들의 합으로 나타나고 있음을 알 수 있다. 따라서 여러 경로를 통한 반향신호의 영향이 더해진 형태로 MCPSP 함수에 미치기 때문에 이것이 직접 경로들에 의한 상관관계보다 커질 경우 잘못된 공간 정보를 추정하게 되는 문제점을 갖고 있다.

3. 제안한 CPSP 저주파 위상 복원

반향신호의 영향

일정 기준치 이상의 파워를 갖는 반향신호의 특성을 결정짓는 중요한 요소인 도달시간 μ_{1l} , μ_{2m} 은 톨 충격 응답으로부터 다음과 같은 지수함수 분포를 갖는 확률 변수로 모델링할 수 있다.

$$\begin{aligned} f_{M_{1l}}(\mu_{1l}) &= \lambda e^{-\lambda(\mu_{1l} - \tau_1)} \quad (\mu_{1l} > \tau_1) \\ f_{M_{2m}}(\mu_{2m}) &= \lambda e^{-\lambda(\mu_{2m} - \tau_2)} \quad (\mu_{2m} > \tau_2) \end{aligned} \quad \lambda > 0 \quad (4)$$

이때 λ 는 실내의 환경에 따른 반향 정도를 결정짓는 변수이며, 두 마이크로폰간의 간격이 작을 경우 같은 값을 적용할 수 있으며 각각의 반향경로들은 실내 환경의 영향에 의한 것이므로 서로 독립적이다. 따라서 이에 따른 반향신호의 도달시간 또한 서로 i.i.d하므로, 다음과 같이 주파수 영역에서 MCPSP함수의 기대값을 얻을 수 있다.

$E[MCPSP]$

$$\begin{aligned}
 &= A_1 A_2 e^{-\lambda(\tau_1 - \tau_2)w} \\
 &\quad + LA_2 \overline{a_1} \int_{-\infty}^{\infty} e^{-\lambda(\mu_{1i} - \tau_2)w} \times f_{M_{1i}}(\mu_{1i}) d\mu_{1i} \\
 &\quad + MA_1 \overline{a_2} \int_{-\infty}^{\infty} e^{-\lambda(\tau_1 - \mu_{2m})w} \times f_{M_{2m}}(\mu_{2m}) d\mu_{2m} \\
 &= [A_1 A_2 e^{-\lambda(\tau_1 - \tau_2)w}] \\
 &\quad + \left[\begin{aligned} &LA_2 \overline{a_1} \frac{\lambda}{\lambda + jw} e^{-\lambda(\tau_1 - \tau_2)w} \\ &+ MA_1 \overline{a_2} \frac{\lambda}{\lambda - jw} e^{-\lambda(\tau_1 - \tau_2)w} \end{aligned} \right] \\
 &= \overline{MCPSP_{direct\ path}} + \overline{MCPSP_{reverberation}} \quad (5)
 \end{aligned}$$

이 때, $\overline{a_i}$ 는 i 번째 마이크로폰에 들어오는 반향신호의 평균 크기를 의미한다.

제안된 CPSP 저주파 알고리즘

이와 같이 각 반향경로 신호들을 독립된 확률변수로 생각하고 평균값을 취하면, 직접경로 신호에 의한 상관관계 및 반향신호에 의한 영향들로 나눌 수 있음을 알 수 있다. 이 식에 의하면 $\frac{\lambda}{\lambda \pm jw}$ 와 같은 복소수 형태로 나타나는 반향신호의 영향으로 실제 공간 정보를 갖고 있는 지수함수의 위상성분이 왜곡되게 되어 잘못된 TDOA를 추정하게 된다.

이러한 MCPSP 함수의 반향성분에 의한 영향은 다음과 같다.

$\overline{MCPSP_{reverberation}}(w)$

$$\begin{aligned}
 &= LA_2 \overline{a_1} \frac{\lambda}{\lambda + jw} e^{-\lambda(\tau_1 - \tau_2)w} \\
 &\quad + MA_1 \overline{a_2} \frac{\lambda}{\lambda - jw} e^{-\lambda(\tau_1 - \tau_2)w} \quad (6)
 \end{aligned}$$

위 식에서 위상 성분만을 추출하면,

$\angle \overline{MCPSP_{reverberation}}(w)$

$$\begin{aligned}
 &= -(\tau_1 - \tau_2)w \\
 &\quad - \left[\frac{LA_2 \overline{a_1} - MA_1 \overline{a_2}}{(LA_2 \overline{a_1} + MA_1 \overline{a_2}) \lambda} \right] w \quad (7)
 \end{aligned}$$

$$= \angle TDOA_{TRUE} + \angle TDOA_{FALSE}$$

따라서, 위 식으로부터 MCPSP 함수의 반향성분에 의한 위상 왜곡 성분은 그림 2와 같이 w 가 작을 경우, 즉 저주파에서 작아지는 것을 알 수 있다. 따라서 반향신호는 MCPSP 함수의 고주파 성분에 대해 큰 영향을 끼치고 있으므로 상대적으로 왜곡이 적은 저주파 영역에서 MCPSP 함수의 위상으로 보다 신뢰도 있는 TDOA를 추정을 할 수 있다.

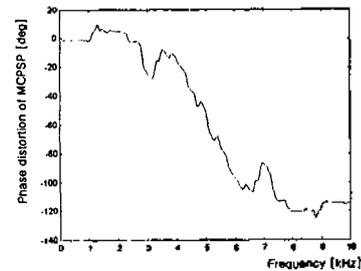


그림 2. 주파수에 따른 MCPSP 함수의 반향성분에 의한 위상 왜곡

4. 실험 결과

제안된 저주파 위상 복원 알고리즘의 성능 분석을 위해 5개의 화자 위치에서 기존 알고리즘 및 제안한 알고리즘을 이용하여 각각 20개씩의 CPSP 함수를 구성하고 TDOA 오차를 분석하는 실험을 수행하였다. 표 1은 각 실험 위치별로 평균 TDOA 추정 오차를 나타낸 것이며 그림 3은 이러한 TDOA 추정 오차의 도수분포도이다.

따라서 따라서 제안한 알고리즘을 적용했을 경우 평균적으로 15샘플 정도 TDOA 추정 오차를 줄일 수 있음을 확인할 수 있고 그림으로부터 기존 알고리즘에서는 특히 심각한 위치 추적 오차를 발생시킬 수 있는 40샘플 이상의 TDOA 오차가 20%이상 발생하고 있음에 비하여 제안한 알고리즘을 적용하였을 경우 10샘플 미만의 TDOA 오차가 70%이상이므로 위치 추적 오차에 큰 영향을 주지 않음을 알 수 있다.

표 1. 알고리즘에 따른 TDOA 추정 오차
(단위 : 샘플)

	기존 알고리즘	제안한 알고리즘
Position 1	38.4950	13.0920
Position 2	20.8670	6.6106
Position 3	14.4879	3.9370
Position 4	15.3297	8.5401
Position 5	35.4109	19.4799
Total	26.8929	11.6876

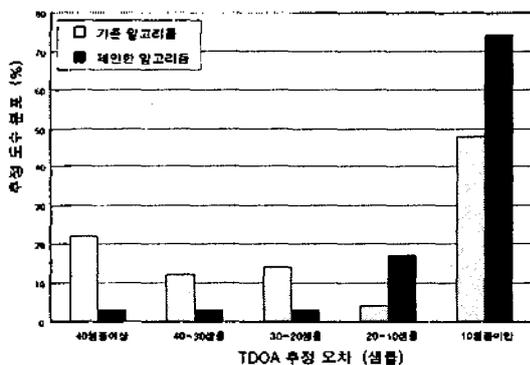


그림 3. TDOA 추정 오차의 도수 분포도

5. 결론

본 논문에서는 기존의 TDOA 추정 방법인 CPSP 알고리즘으로부터 반향의 영향에 강인한 CPSP 저주파 위상 복원 알고리즘을 제안하고 실측 데이터를 이용한 실험을 통해 그 성능을 분석하였다. 제안한 알고리즘은 반향 경로의 모델링을 통해 상대적으로 CPSP 저주파 영역에서의 반향의 영향이 줄어들고 있음을 분석하여 저주파

정보만을 이용함으로써 보다 신뢰도 있는 TDOA 추정을 가능하게 한다.

실험 결과 제안한 알고리즘을 적용한 경우 TDOA 추정 오차를 평균 15샘플 정도 줄일 수 있었으며 심각한 위치 추적 오차를 불러일으킬 수 있는 40샘플 이상의 오차 발생 확률을 크게 개선함으로써 위치 추적 성능을 향상시킬 수 있음을 확인하였다.

참고 문헌

- [1] M. Omologo and P. Svaizer, "Acoustic Event Localization using a Cross Power-Spectrum Phase based Technique," *Proc. of ICASSP*, pp. 273-276, Adelaide, Australia, Apr. 1994.
- [2] D. Rabinkin, R. Renomeron, A. Dahl, J. French, J. Flanagan, and M. Bianchi, "A DSP implementation of source location using microphone arrays," *Proc. of the SPIE*, vol. 2846, pp. 88-99, Aug. 1996.
- [3] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A practical time-delay estimator for localizing speech sources with a microphone array," *Computer, Speech, and Language*, vol. 9, no. 2, pp. 53-69, Apr. 1995.
- [4] D. Rabinkin, *Optimum Sensor Placement for Microphone Arrays*, New Brunswick, New Jersey, May 1998.