2001년도 대한전기학회/대한전자공학회 시스템및제어분야 합동 추계학술대회 논문집 (2001.11.24)

대화형 로봇의 화자 추종을 위한 sound localization

심현민, 이종실, 권오상*, 이응혁**, 홍승홍
인하대학교 전자공학과
*(주) 한울로보틱스
** 한국산업기술대학교 전자공학과

Sound localization for Teller Following of A dialog type Humanoid Robot

H.M. Shim, J.S. Lee, O.S Kwon, E.H. Lee, S.H. Hong
Dept., of Electronic engineering, Inha university
*Han-wool robotics co., Inc.
**Dept., of Electronic engineering, Korea Polytechnic university
E-mail : robot@elecage.pe.kr

Abstract - In this paper, we supposed teller following algorithm that using sound localization for developing dialog type humanoid robot. A sound localization is studied for develop the techniques of an efficient 3-D sound system based on the psychoacoustics of spatial hearing with multimedia or virtual reality. When a robot talk with human, it is necessary that robot follow human for improved human interface and adaptive noise canceling. We apply this algorithm to robot system.

1. 서 론

과학 기술의 발달로 인해 로봇은 산업 용도뿐만 아니라 가사와 교육, 유희 도구로까지 그 쓰임이 넓어졌다. 이러한 로봇들은 기계 제어와 더불어 인공지능과 감성, 시청각적인 요소들도 요구된다.

최근에는 사람과 대화가 가능하고 인간과 유사한 감성과 시청각 기능을 가지는 인간형 로봇에 대한 연구가 활발히 진행되고 있다. 그 대표적인 예로 MIT의 AI Lab.의 Kismet이나 일본 동경대와 Japan Science and Technology Corp.의 SIG를 들 수 있다[3][5]. 이들 로봇은 화자를 식별하고 간단한 대화를 할 수 있으며 감정의 표현이 가능하다.

이러한 대화형 로봇은 사람과 대화를 할 때 사람이 자리에서 위치를 이동하더라도 추종하여 마주봄으로써 사람의 기계에 대한 거부감을 줄일 수 있고 음성 인식시 화자의 위치를 잡음 제거의 중요한 단서로서 사용하게 된다. 이 때 사용할 수 있는 방법이 음상 정위(sound localization)이다.

음상 정위란 음원에서 음파가 방사될 때 그 음상의 위치를 판단하는 것으로 사람과 대화를 할 때 화자의 말소리를 단서로 화자의 방향을 판별하는 것이다[2][4]. 음원들 특정 각도에 정위시키는데 있어서 가장 중요한 단서는 수평면에 위치한 두 귀에서의 파면의 상대적인 차이이다. 음상정위에 사용되는 단서로는 두 귀간의 시간차(Inter-aural Time Difference : ITD)와 두 귀간의 레벨차(Inter-aural Level Diffrence : ILD)가 있다. 특히 ITD는 음원의 방위(azimuth)에 중요한 단서(cue)가 된다.

음상 정위에는 수평면 내에 있는 음원의 방향을 지각하는 수평면 정위(horizontal plane localization), 정중면 상의 음원의 방향을 지각하는 정중면 정위(median plane localization), 수평면과 정중면 이외의 음원의 위치를 판별하는 상반구면 정위 등이 있다[2].

본 논문에서는 이와 같은 음상 정위 단서를 이용하여 로봇에 인간의 두 귀와 같이 2개의 마이크로 음의 입력을 받아 음상 정위 단서를 이용하여 대화하고자 하는 화자의 방향을 향하는 로봇 시스템을 구현한다.

본론에서는 음상 정위 단서를 이용하여 방향을 판단하는 방법을 설명하고 구현된 시스템에 대해 설명한다. 다음으로 이를 이용한 실험의 결과 및 향후 연구 방향에 대해서 기술한다.

2. 음상 정위 단서에 의한 방위 판별

2.1 두 귀간의 시간차(Inter-aural Time Difference : ITD)

음원의 방위 판별을 위해 사용되는 가장 중요한 단서는

ITD이다. 음파는 상온에서 약 340m/s의 속도로 전달된다. 따라서 그림 1과 같이 인간의 두 귀에 음이 도달하는 데에 약간의 시간차가 발생하게 된다.

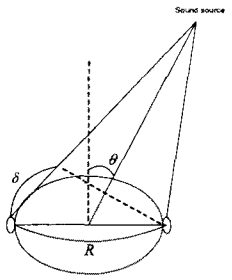

그림 1. ITD의 개념도

그림 1을 살펴보면 음원과 양쪽 귀와의 거리가 δ만큼의 차이가 나므로 정면을 기준으로 하는 음원의 방위 θ는 다음과 같이 구할 수 있다.

$$\theta \approx \sin^{-1}\left(\frac{\delta}{R}\right) \qquad (1)$$

양쪽 귀에 들리는 음원이 왜곡이 없고 시간 지연만 있다고 가정하면 그림 2에서 보는 바와 같이 시간 만큼의 위상차이가 발생한다. 여기서 음파의 속도를 C라고 하면 $\delta = \tau / C$이므로 θ를 구하는 식은 식 (2)와 같이 된다.

$$\theta \approx \sin^{-1}\left(\frac{\tau}{R \cdot C}\right) \qquad (2)$$

이를 로봇에 적용할 경우에 음파는 일정한 시간간격으로 표본화한 이산 신호로 입력을 받게 되며 한쪽 신호를 시간축에 대해 이동(shift)시키며 상호 상관(cross correlation)을 조사하여 상관도가 가장 높은 점을 찾으면 원래의 파형에서 최대 상관도를 보이는 지점과의 차이가 위상차를 나타내는 것이다. 식(3)은 위상차 τ를 구하는 계산식이다.

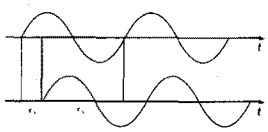

그림 2. 두 귀간의 위상 지연

$$\tau = \arg\left\{\max \sum_{n} x_l[k] \cdot x_r[k+\tau_i]\right\} \qquad (3)$$

$x_r[k]$를 τ_i만큼 쉬프트 시켜서 $x_l[k]$와 곱한 값을 n개 더한 값 중 최대가 되는 τ_i가 지연된 샘플 만큼의 차가 된다. 만약 두 음원의 위상차가 180° 이상이 될 경우에는 오히려 멀리 있는 방향의 음이 먼저 입력된 것처럼 나타날 수도 있다. 따라서 고주파일수록 파장의 간격이 좁아지므로 ITD에 의한 방향 판별이 어려워진다.

일반적으로 ITD는 1.5kHz 이상의 주파수에서는 위상편이에 따른 시간지연의 측정이 힘들어지며 따라서 ITD에 의한 음상 정위는 1.5kHz 이하의 저주파 대역의 음에 대한 판별에 사용하게 된다.

2.2 두 귀간의 레벨차(Inter-aural Level Difference : ILD)

고주파의 음상 정위에 주로 사용하는 ILD는 1.5kHz에서 6kHz 사이의 정위에 용이하며 음원의 거리 판별에 중요한 단서가 된다.

음파는 음원으로부터 멀어짐에 따라 점점 감쇄하게 되며 음원에서 보다 가까운 쪽의 귀가 반대쪽 귀보다 더 큰 레벨의 음을 입력받는다. 거리와 레벨과의 관계는 역자승 법칙을 이용할 수 있다. 즉 주어진 기준 레벨과 거리에 대해 그림 3에서 보는 것과 같이 무지향성 음원으로부터 거리가 2배 멀어지면 레벨은 6dB 감소하게 된다.

ILD에 의한 정위는 ITD를 이용한 음상 정위에 비해 정확한 판별이 힘들고 1.5kHz 이하의 저주파 영역에서는 파장의 길이가 양 귀 사이의 거리보다 길기 때문에 양쪽 귀에서 레벨의 차이를 느낄 수 없으므로 효과가 없다. 따라서 로봇에 적용할 때 ILD는 음상 정위에서 단독으로 사용하기 보다는 ITD 단서에 대한 보조적인 단서로서 사용하게 된다.

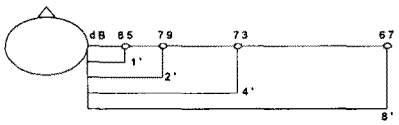

그림 3. 역자승 법칙에 의한 소리의 세기 감소

3. 시스템 구현 및 실험 방법

본 논문에서는 음상 정위의 주요 단서로 ITD를 사용하였으며 ILD는 오차를 줄이기 위해 보조적으로 사용하였다.

실험을 위해서 가로 18cm, 세로 24cm 크기의 로봇 머리를 만들고 양쪽에 각각 마이크로폰을 부착하였다. 이 로봇의 머리는 DC 서보 모터를 통해 좌·우 전방향으로 회전이 가능하도록 구성하였다.

그림 4.는 로봇 시스템의 개략적인 구성도이다. 본 시스템에서 음의 입력은 -40dB 감도의 무지향성 콘덴서 마이크로폰을 사용하였다. 양쪽 마이크로폰 사이의 거리 R

은 18cm이다.

마이크로폰 입력 신호를 PC의 사운드 카드의 라인 입력을 위해서는 마이크로폰 신호를 증폭해야 한다. 실험에서 사용한 증폭기의 증폭도는 50으로 하였다. 증폭된 신호는 전원에 의한 60Hz 잡음과 저주파 잡음, 음성 대역 이상의 불요 잡음 등이 섞이게 되므로 대역 통과 필터를 사용하여 잡음을 제거하였다.

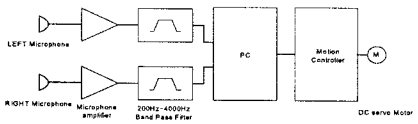

그림 4. 시스템 구성도

이렇게 입력된 신호는 PC의 사운드 카드를 통해 48000Hz의 표본화된 이산 신호로 변환된다. 음파의 속도는 상온에서 340m/s라고 할 때 하나의 샘플에서 다음 샘플까지의 시간은 0.21μs이므로 그 거리 차이는 약 0.71cm가 된다. 따라서 간략화된 계산식이 식(4)와 같다.

$$\theta = \sin^{-1}\left(\frac{0.71 \cdot \tau}{18}\right) \qquad (4)$$

계산된 방위로 로봇의 머리를 회전시키기 위해 PC에서 Motion controller를 통해 모터를 제어하게 되는데 본 실험에서 사용한 Motion controller는 (주)한울로보틱스의 HWR-DMC2로서 USB 방식으로 PC와 연결되어 PC에서 PID 제어로 회전 각도의 제어가 가능하도록 설계되었다. 로봇의 머리를 회전시키기 위한 모터는 스위스 Maxon사의 서보 모터를 사용하였으며 표 1은 서보 모터의 사양을 나타내었다.

표 1. 서보 모터 제원

Power Rating (W)	Normal Voltage (V)	No Load Speed (rpm)	Continuous Torque (mNm)	Continuous Current (mA)
20	24	6420	47.3	1350

사용된 모터는 23.04:1의 감속비를 갖는 감속기가 일체형으로 부착되어 있으며 인코더를 통해 모터의 회전 각도를 알 수 있다.

4. 실험 결과 및 고찰

본 실험의 실험 결과를 위하여 로봇으로부터 약 1m 거리에서 600Hz의 정현파의 음파를 발생시킨 후 실험을 하였다. 16비트 48000Hz로 표본화 된 신호의 세기는 0~1까지의 크기로 정규화 시킨 후 0.15 이하의 신호는 무시하였다. 이 중 2048개의 연속된 유효 신호를 추출하여 상호 상관도를 계산하여 방향을 판별하여 로봇의 머리를 회전하도록 하였다.

식(4)를 사용하여 에서 $0.71 \cdot \tau / 18 < 1$을 만족하는 최대 정수 τ는 25이며 1부터 25까지의 τ값에 따른 θ는 표 2와 같다.

표 2. 샘플링 지연에 따른 방위

τ	θ	τ	θ
1	2.26	14	33.52
2	4.52	15	36.28
3	6.80	16	39.13
4	9.08	17	42.11
5	11.37	18	45.23
6	13.69	19	48.54
7	16.03	20	52.08
8	18.39	21	55.93
9	20.79	22	60.20
10	23.23	23	65.12
11	25.71	24	71.20
12	28.25	25	80.44
13	30.85		

표 2를 보면 정중면쪽에 가까울수록 해상도가 높고 한쪽으로 치우쳐지는 경우에는 판별할 수 있는 각도가 커지는 것을 알 수 있다.

실험 결과 정면에서 약 ±60° 이내의 각도에서는 오차 10° 이내의 비교적 정확한 정위를 하였다. 그러나 한쪽 방향으로 치우쳐질 경우에는 오차가 점점 심해지기 시작하였으며 정면에서 90° 방향, 즉 왼쪽 또는 오른쪽으로 치우쳐졌을 때에는 약 15% 정도는 오히려 반대 방향으로 회전을 하는 경우도 보였다. 이는 실내가 완벽한 자유 음장 공간이 아니며 음이 머리를 지나면서 많은 왜곡이 발생하기 때문으로 보인다.

비슷한 거리에서 사람의 음성으로 로봇에게 소리를 내도 비슷한 결과를 얻을 수 있었는데 오히려 정현파 신호보다 사람의 음성에 대한 반응의 오차가 약간 더 적은 것을 확인할 수 있었다.

5. 결 론

본 연구에서는 대화형 로봇을 개발하기 위한 기본 단계로 로봇과 대화하고자 하는 화자의 목소리에 반응하여 화자를 추종하는 기법과 이를 구현한 실험의 결과를 기술하였다. 본 논문의 실험 결과를 보면 반사파가 적고 외부 잡음이 크지 않은 실내 공간에서의 음상 정위는 만족할 만한 결과를 얻었다고 볼 수 있다.

본 논문의 향후 과제로는 로봇의 머리를 구동하는 모터의 소음 및 외부 소음 등에 대한 잡음 제거에 관한 연구를 통하여 음성 인식 기능을 추가하였을 때 인식률을 높일 수 있는 연구를 진행해 나갈 것이다.

또한 박수와 같은 짧은 시간동안 신호의 에너지가 급격히 변하는 소리는 오차가 더욱 줄었는데 이러한 소리에 로봇이 민감하게 반응할 경우 오히려 화자보다 다른 쪽의 시끄러운 소리를 추종할 수가 있다. 따라서 로봇이 음원을 추종하기에 앞서 그 음이 음성인지 비음성인지를 판별할 수 있도록 하는 알고리즘에 대한 연구를 진행할 것이다.

[참 고 문 헌]

[1] C. Schauer, H.-M. Gross, " Model and Application of a Binaural 360° Sound localization System", Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on , Volume: 2 ,pp.1132-1137, 2001

[2] 강성훈, 강경옥 공저, "입체 음향", 기전 연구사, pp.40-81, 1997

[3] Kazuhiro Nakadai, Tatsuya Matsui, Hiroshi G. Okuno, and Hiroaki Kitano, "Active Audition System and Humanoid Exterior Design", Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.1453-1461

[4] William G. Gardner, " 3-D Audio Using Loudspeakers", Kluwer Academic Publishers, pp.99-115

[5] Adams, B., Breazeal, C., Brooks, R.A., Scassellati, B., IEEE Intelligent Systems, Volume: 15 Issue: 4 , pp.25-31, 2000

[6] Chun, G.D., Caudell, T.P. "A model for auditory localization in robotic systems based on the neurobiology of the inferior colliculus and analysis of HRTF data ", Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on , Volume: 2 , pp.1107-1111, 2001

[7] Mark Kahrs, Karlheinz Brandenburg, "Applications of Digital Signal Processing to Audio and Acoustics", Kluwer Academic Publishers, pp.85-97