2차원 마이크로폰 배열에 의한 능동 청각 시스템

Active Audition System based on 2-Dimensional Microphone Array

이 창 훈*, 김 용 호*

* 배재대학교 전자공학과(전화:(042)520-5702, 팩스:(042)520-5773, E-mail : naviro@pcu.ac.kr)

Abstract : This paper describes a active audition system for robot-human interface in real environment. We propose a strategy for a robust sound localization and for-talking speech recognition(60-300cm) based on 2-dimensional microphone array. We consider spatial features, the relation of position and interaural time differences, and realize speaker tracking system using fuzzy inference process based on inference rules generated by its spatial features.

Keywords : active audition system, 2-dimensional microphone array, sound localization, nonlinear compression

Ⅰ. 서 론

산업용으로 주로 쓰이던 로봇이 인간의 생활에 밀접하게 다가오면서 인간과 로봇의 활동공간의 공유가 많아지게 되었다. 이로 인하여 인간과 접하는 시간이 증대되어 인간과 친밀한 인터페이스 구현에 대한 연구가 활발히 진행되고 있으며 이와 관련하여 음성인식은 휴먼로봇을 개발하는 곳에서는 필연적으로 사용이 되고 있다.

최근 통신분야와 관련하여 음성인식과 음성합성의 기술이 급속히 발전하고 있으나, 음성인식에 있어 현재 헤드셋을 이용하거나 마이크에서 30cm정도 떨어진 거리에서 음성을 인식하는 것이 일반적이며, 그 이상의 거리에서나 잡음이 존재하는 실제 환경에서 인식률이 급격히 떨어져 이동로봇과 같은 실용시스템에 접목하는 기술이 부족한 상황이다.

그리고 인간은 대화할 때 서로 마주보며 대화하는 것이 일반적이며, 그렇지 않는 경우에는 인간 간의 대화에서도 위화감을 주기 마련이다. 로봇에게 있어서 대화 상대를 주시함으로써 음성인식률의 향상과 더불어 인간에게 친밀감을 줄 수 있는 기능에 대한 필요성이 높아지고 있다.

따라서 본 연구에서는 화자의 방향을 검지하여 추종함으로써 로봇과의 대화에 있어 위화감을 줄일 수 있으며, 또한 지향성을 증대시켜 잡음에 대해 상대적으로 신호를 강조할 수 있어 음성인식률을 증대시킬 수 있다.

그래서, 본 논문에서는 대화형 로봇에 있어 화자의 방향을 검지하여 추종하며 원거리의 음성을 인식하는 능동 청각 시스템에 관하여 논의한다. 특히, 이동로봇의 특성상 화자와의 거리가 반경 3m이내에서 전방뿐만 아니라 후방, 즉, 임의의 방위에 대하여 화자의 방향을 판별해야하며, 음성인식이 가능해야한다는 제약조건에서 3개의 마이크로폰을 2차원으로 배치하여 음성의 입력받아 비선형증폭을 하는 전처리, 지연시간차와 음원방향과의 관계에 퍼지이론을 적용하여 화자에 능동적으로 대응하는 청각 시스템을 구현한다.

Ⅱ. 능동 청각 시스템

2.1 시스템 구성

반경 15cm의 정삼각형으로 배열된 3개의 마이크(왼쪽: *l*, 오른쪽: *r*, 전방중앙: *c*)로부터 각각 입력된 신호를 이용하여 상관관계를 계산하고 마이크 위치의 특성으로부터 얻어진 정성적인 관계를 추론하여 이를 통하여 음원의 방향을 검출하여 화자를 추종하고, 음성활성검지(Voice Activity Detection)를 통하여 음성영역을 추출하여 잡음 제거와 목적음의 강조를 통하여 얻은 결과를 음성 인식부로 전달하여 음성인식을 하는 시스템을 구성한다. 시스템의 전체 구성도는 그림 1과 같다.

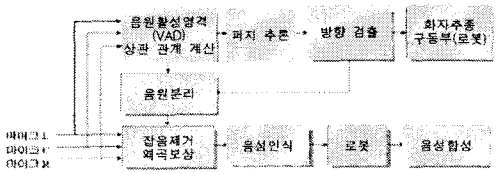

그림 10 능동 청각 시스템 구성도

본 논문에서는 능동 청각 시스템을 구현하기 위하여 구체적으로 다음 3가지 사항에 대하여 논의한다. (i)음성신호 검지 거리개선, (ii)잡음에 대한 강인성 향상, (iii)전 방위(360°)에 대해 균일한 정밀도를 갖는 방향검지.

2.2 음성신호 검지 거리개선

마이크로폰으로부터 증폭도가 일정한 증폭기를 통하

여 신호를 입력받는 일반적인 경우, 증폭도가 낮을 때는 음압이 낮으면 음원을 검출하기 힘들며 증폭도가 높을 때는 음압이 높으면 포화되게 되어 신호가 크게 왜곡된다. 따라서 음원검출의 거리를 늘리기 위한 유효한 한 방법으로 능동적으로 신호를 증폭할 수 있는 비선형 증폭기를 생각할 수 있다. 여기서는 이를 위하여 그림 2에 나타낸 비선형 압축 소자를 이용한 회로를 통하여 근거리와 원거리의 음원을 모두 방향검지를 위한 입력으로 사용할 수 있다.

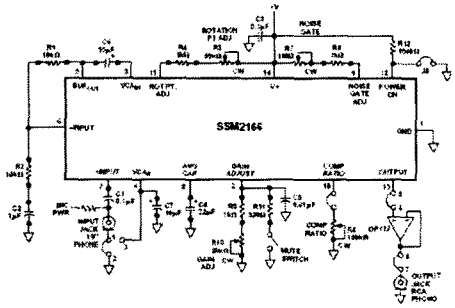

그림 2 비선형 증폭회로 예

2.4 잡음에 대한 강인성 향상

잡음에 대하여 강인성을 향상시키기 위하여 음성코딩이나 음성통신에서 일반적으로 사용되는 음성활성영역의 검출 방법을 활용할 수 있다. 차이점으로는 일반적인 목적으로 사용되는 VAD 기법과는 달리 방향검지의 정확성을 목적으로 하는 경우에는 반드시 무성음을 포함하는 음성영역을 찾아야 할 필요는 없다. 왜냐하면 무성음의 경우 음서의 일부분이지만 주기성을 갖지 않기 때문에 무성음 영역을 방향검지에 이용할 경우 잘못된 결과를 얻게 되는 일이 빈번하게 발생한다. 바꾸어 말하면, 무성음 영역의 신호에서는 잡음과 마찬가지로 방향검지에 있어 유효한 정보를 얻기 힘들다는 것을 의미한다. 이러한 이유로 본 연구에서는 접근방향을 일반적인 음성활성영역 판별을 지양하고 유성음 영역의 판별에 주안점을 둔다. 이를 위하여 단구간 에너지(short term energy), 피치(pitch), 영교차률(zero crossing rate)의 3가지 특징벡터를 사용한다. 이 중 피치에 대해서는 신호의 크기에 의존하지 않으며 유성음을 잘 찾아내는 특성을 가진 캡스트럼(cepstrum)을 이용한 검출 방법을 이용한다.

2.4 균일 정밀도를 갖는 방향 검지

방향 판별을 위하여 일반적으로 (i)두 귀사이의 시간차이(Interaural Time Differences: ITD), (ii)두 귀사이의 음압 차이(Inteaural Intensity Differences: IID), (iii) 헤드전달함수(Head-related Transfer Functions: HTF)를 사용한다. 이들 중 두 귀사이의 음압 또는 강도의 차이를 실제로 마이크로폰을 통해 이용할 때는 여러 증폭기의 증폭도를 일정하게 유지하기가 어려우며, 위치와 시간에 따른 증폭도 변화 등 유효한 오차 범위를 유지하기 위해서는 여러 가지 어려움이 따른다. 한편 두 귀사이의 시간 차이를 이용할 경우 고주파에서는 방향 판별이 어려우나 1.5kHz 근방이나 이하에서는 효율적이다. 본 연구에서는 샘플링 주파수를 높이면 판별의 정확도를 높일 수 있고, 음원이 음성이므로 비교적 간편한 두 귀사이의 시간 차이에 근거한 방향 판별 방법을 고려한다.

두개의 마이크로폰을 이용할 경우 측면 방향에 음원이 있을 경우 판별 오차가 커지며, 전방과 후방을 구별하기가 어렵다. 이러한 문제를 고려하여 모든 방위에 대하여 상당히 균일한 오차를 가지며 전방위를 판별할 수 있게 하기 위하여 본 연구에서는 2차원 평면상에 정삼각형 형태로 3개의 마이크로폰을 배치하여 사용한다.

먼저, 이해를 돕기 위하여 두개의 마이크로폰의 경우에 대하여 살펴보도록 하자.

두 마이크 사이의 거리를 l, 두 마이크를 이은 선분과 수직인 선분과 과 음원과의 각도차를 θ라 하면, 음원에서 두 마이크에 도달하는 거리의 차이 d와의 관계식은 아래와 같다.

$$\theta \approx sin^{-1}\left(\frac{d}{l}\right) \tag{1}$$

$$d = \frac{v}{fs}k \tag{2}$$

여기서, v는 음파의 속도, fs는 샘플링 주파수, k는 위상차이다. 그리고 두 마이크로 입력된 신호를 각각 $s_x(n)$, $s_y(n)$라 하면 두 신호 사이의 상호-상관관계 $R_{xy}(k)$는 다음과 같다. 이 값 중에서 최대가 되는 k가 지연된 샘플의 차이 값이 된다.

$$R_{xy}(k) = \frac{\sum_n \{s_x(n-k)s_y(n)\}}{\sqrt{\sum_n s_x(n-k)^2}\sqrt{\sum_n s_y(n)^2}} \tag{3}$$

이제 3개의 마이크로폰을 원점을 중심으로 정삼각형으로 배치하여, 각 마이크를 통해 입력된 디지털 신호를 $s_l(n)$, $s_r(n)$, $s_c(n)$라 하고, 이들 중 각 쌍의 상관관계, $R_{rl}(k)$, $R_{rc}(k)$, $R_{cl}(k)$라 하자. 이때의 값은 -1에서 1의 값을 가진다.

그러면, 위치와 시간 지연과의 관계를 살펴보자. 반경 60cm-300cm와 각도 1°~360°에 해당하는 위치와 왼쪽과 오른쪽 마이크 사이의 지연 시간과의 관계를 계산하여 정규화한 것을 그림 2에 나타낸다. 이 그림을 보면 지연 시간은 거의 각도에 의존함을 알 수 있다. 또한 마이크가 x축과 평행하게 배치되어 있으므로 그

림에서 볼 수 있듯이 수평 방향으로 조밀하게 되어있어 분해능이 떨어짐을 알 수 있다.

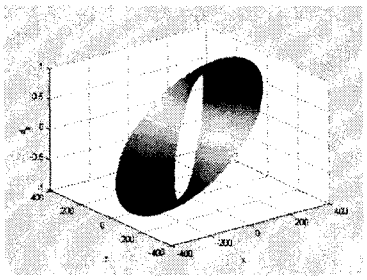

그림 3 위치와 지연과의 관계 $d_b(x, y)$

이에 대해 정삼각형으로 배치함으로써 조밀하게 된 각도 영역에 대하여 또 다른 마이크로폰 쌍에 의해 보완되어지게 된다. 이것이 전방위에 대해 분해능을 일정하게 유지할 수 있는 이유이다. 그리고 마이크로폰 사이의 간격 정보를 이용하여 최대 지연 시간을 계산하여 이를 이용함으로써 반사음이나 일부 예외 상황에 대한 영향을 배제시키는 효과를 얻을 수 있다.

그 다음은 어느 각도에서 화자가 말을 했는지의 여부를 알아내기 위하여, 앞에서 얻은 두개의 마이크 쌍에 대한 위치와 지연 시간과의 관계로부터 각도를 추론할 수 있다.

표 1 방위 계산을 위한 퍼지추론 규칙

	Mic l-r	Mic r-c	Mic c-l
30°	NM	PB	NM
60°	NB	PB	ZR
90°	NB	PM	PM
120°	NB	ZR	PB
150°	NM	NM	PB
180°	ZR	NB	PB
210°	PM	NB	PM
240°	PB	NB	ZR
270°	PB	NM	NM
300°	PB	ZR	NB
330°	PM	PM	NB
360°	ZR	PB	NB

예를 들어, 음원이 180°일 때를 생각해 보자. 마이크로폰 l과 r에 대해서는 둘을 잇는 선분과 수직이므로 지연이 없어 0이 됨을 알 수 있다. 그리고 마이크로폰 r과 c에 대해서는 음으로 큰 값을 갖게 된다. 또한 마이크로폰 c과 l에 대해서는 양으로 큰 값을 가짐을 알 수 있다. 반대로 입력 신호로부터 상호-상관관계를 계산하여 지연 시간이 위와 같은 경우에도 서로의 관계를 종합하여 추론하면 방위 값을 얻을 수 있음을 알 수 있다. 이와 같은 관계로부터 30° 간격으로 작성한 추론규칙은 표 1과 같다. 자세한 지연 시간과 각도와의 관계는 이 표의 규칙을 보면 이해할 수 있다.

III. 실험 및 고찰

본 연구에서는 두 귀 사이의 시간차를 근거로 전절에서 제안한 방법을 통해 능동청각시스템을 구현하였다. 반경 15cm로 3개의 -38dB 감도의 무지향성 콘덴서 마이크로폰을 정삼각형으로 배치하여, 입력 신호들 간의 지연 시간을 계산하여 각도와 지연 시간과의 관계로부터 얻어진 표 1의 퍼지규칙으로부터 퍼지추론을 하여 전방위에 대해 균일한 분해능으로 화자의 방위를 얻도록 구성하였다. 하드웨어 구성은 그림 4과 같다.

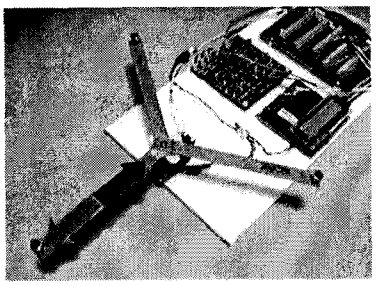

그림 4 능동청각시스템의 하드웨어구성

마이크로폰과 신호처리부와의 사이에는 60~300cm 거리에서도 화자의 방향을 검지하고 음성인식부의 입력으로 사용하여 인식 가능하도록 하기 위하여 2.2절에서 설명한 신호압축 등의 전처리회로가 구성되어 있고, A/D변환기로는 National Instrument사의 16비트, 200kHz의 카드형을 이용하였다. 실험에서는 각 채널에 대해 11kHz로 샘플링하였고, 입력신호 범위는 ±2.5V로 하였었다. 음파의 속도는 20°C에서 343.4m/s로 하였으며, 이때 최대 지연 샘플은 17이었다. 그리고 퍼지추론 입력으로는 3쌍의 마이크로폰에 대한 상호-상관관계 계수값이 최대가 되는 위상차(지연 샘플수)의 상위 5%의 값을 이용하였다.

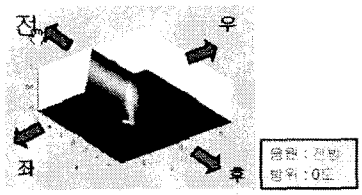

그림 5 방위 표시와 0° 음원

방위 표시와 0° 음원에 대한 실험결과를 그림 5에 나타낸다. 잡음이 없는 이상적인 경우에는 화자의 방향에 따라 매끄러운 그래프를 그리지만, 잡음이 있는 실제 상황에서는 매끄러운 결과가 나오지 않는다. 이

러한 이유에서 방향검지에 앞서 일부에 음성판별을 위한 문턱값을 정하는 알고리즘의 일부를 사용하였다.

실험환경은 그림 6에 나타낸 것과 같이 실제 가정의 거실에서 TV방송, 오디오, 청소기 잡음이 있는 상황을 설정하였다. 잡음의 정확한 레벨은 측정하지 못하였으나, 모터 잡음과 인간이 충분히 들을 수 있을 정도의 TV방송과 오디오가 동작되고 있는 환경에서 실험하였다. 여기서 오른쪽에 흰색으로 표시된 것은 1m 간격을 나타내고 있으며 음원의 거리는 1m~2.8m사이로 하여 실험하였다. 실험결과의 예를 그림 6과 7에 나타낸다. 자세한 결과는 참고문헌 [5]의 동영상을 참고하기 바란다.

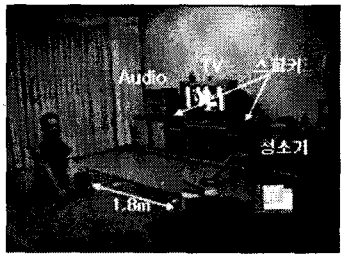

그림 6 실험 환경

화자의 방향검지의 오차 범위는 ±8°이었다. 또한 온도가 다소 차이가 나는 환경에서 실험한 결과도 양호하게 나타났다. 이처럼 온도 변화가 있는 환경에서도 오차가 커지지 않는 이유는 정성적인 규칙에 바탕으로 퍼지추론을 하기 때문인 것으로 판단된다.

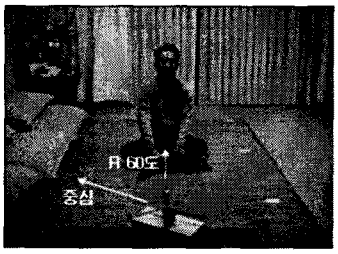

그림 7 화자 추종 결과

방향을 검지한 후 음원분리를 통해 음성인식부의 입력으로 사용하여 인식한 결과 60~300cm 거리에서 인식률이 85% 정도이었다. 인식에 있어서는 문제점이 남아 있으나 제반 환경을 고려하면 이 결과는 근거리와 원거리에 관계없이 인식한 결과로 양호한 것으로 판단되어진다. 인식률이 떨어진 이유로는 압축된 신호를 완전히 선형으로 복원하지 않아 음성에 왜곡이 생긴 것이 하나의 이유이고, 또 하나는 음성 인식 프로그램 자체의 인식률이 환경에 민감한 것이 원인으로 여겨진다.

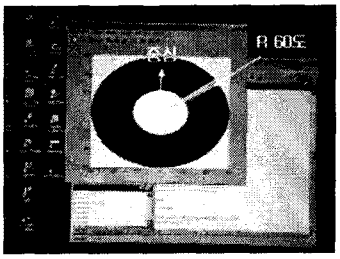

그림 8 방향검지와 음성인식 결과

Ⅳ. 결 론

본 연구에서는 로봇과 인간과의 인터페이스에 있어 능동 청각 시스템을 통하여 화자의 방향을 검지하여 로봇이 화자를 추종함으로써 인간에게 친근감을 증대시키고, 3m 이내에서 명령한 것을 로봇이 음성인식을 하여 명령을 수행하게 하는 것을 목적으로 하였다. 이 목적을 2차원으로 마이크로폰을 배치하여 비선형 신호압축기법과 유성음영역 검출기법, 화자의 위치와 신호들의 상관관계를 정성적으로 표현하여 퍼지추론을 통하여 실현하였다.

인간과 같이 두 귀에 해당하는 2개의 마이크로폰을 이용하여 방향 검지를 하는 경우에는 본문에서 언급한 바와 같이 마이크로폰을 잇는 선분에 가까운 곳에서의 음원을 검지할 경우에는 오차가 커질 수밖에 없으며, 전방위에 대한 검지도 어렵다. 그렇지 않는 경우에는 상당히 복잡한 알고리즘으로 구현하여야 하는 부담이 생긴다. 이에 반해 3개의 마이크로폰을 2차원으로 배치하여 사용하는 경우에는 비교적 간단한 방법으로 전방위에 대해 정밀도를 균일하게 유지할 수 있으며 전방위에 대해 화자의 방향을 검지할 수 있음을 알 수 있다.

참 고 문 헌

[1] Jens Blauer, Spatial Hearing : The Psychophysics of Human Sound Localization. MIT Press, 1996.

[2] C. Schauer, H.-M. Gross, "Model and Application of a Binaural 360° Sound Localization System," IEEE Conf. pp.1132-1137, 2001.

[3] N. Roman, D. Wang, G. J. Brown, "Speech segregation based on sound localization," IEEE Conf., pp.2861-2866, 2001.

[4] K. Nakadai, K. Hidai, H. G. Okuno, H. Kitano, "Real-Time Active Human Tracking by Hierarchical Integration of Audition and Vision," Proc. of IEEE-RAS Humanoid 2001, pp.91-98, Nov. 2001.

[5] http://ee.pcu.ac.kr/~chlee/ActiveAuditionSystem