

실시간 지능형 능동 청각 시스템

이창훈*(배재대학교 전자공학), 김현돈, 최종석, 김문상(한국과학기술연구원)

주제어 : 능동 청각 시스템 (Active Audition System), 실시간 (Real-time), 비선형압축 (Nonlinear Compression), 전 방위 (Omni-directional), 퍼지추론 (Fuzzy Inference)

산업용으로 주로 쓰이던 로봇이 인간의 생활에 밀접하게 다가오면서 인간과 로봇의 활동공간의 공유가 늘어가고 있다. 이로 인하여 접하는 시간이 증대되어 인간과 친밀한 인터페이스 구현에 대한 연구가 활발히 진행되고 있으며, 이와 관련하여 지능형 로봇에 있어 음성시스템은 필수적이다.

최근 통신분야와 관련하여 음성인식과 음성합성의 기술이 급속히 발전하고 있으나, 음성인식에 있어 현재 헤드셋을 이용하거나 마이크로폰에서 약 30cm정도 떨어진 거리에서 음성을 인식하는 것이 일반적이며, 그 이상의 거리에서나 잡음이 존재하는 실제 환경에서 인식률이 급격히 떨어져 이동로봇과 같은 실용시스템과의 접목하는 기술이 부족한 상황이다. 그리고 인간은 대화할 때 서로 마주보며 대화하는 것이 일반적이며, 그렇지 않는 경우에는 인간 간의 대화에서도 위화감을 주기 마련이다. 로봇에게 있어서도 대화 상대를 주시함으로써 음성인식률의 향상과 더불어 인간에게 친밀감을 줄 수 있는 기능에 대한 필요성이 높아지고 있다.

따라서 본 논문에서는 화자의 방향을 검지하여 추적하며 원거리의 음성을 인식하는 능동 청각 시스템에 관하여 논의한다. 가정내의 로봇을 대상으로 하고 있기 때문에 수가 적은(3개의) 마이크로폰으로 화자와의 거리가 반경 3m이내에서 모든 방위에 대하여 화자의 방향을 판별하고, 음성인식이 가능하게 하는 것을 목표로 한다. 이를 구현하기 위하여 1. 음성신호 검지 거리의 개선, 2. 잡음에 대한 강인성 향상, 3. 균일 정밀도를 갖는 방향 검지에 대하여 구체적인 방안을 논의한다.

마이크로폰으로부터 증폭도가 일정한 증폭기를 통하여 신호를 입력으로 하는 일반적인 경우, 증폭도가 작을 때는 음압이 낮으면 음원을 검출하기 힘들며, 증폭도가 클 때는 음압이 높으면 포화되어 신호가 크게 왜곡된다. 그래서, 음성 신호의 검지 거리를 향상시키기 위하여 능동적으로 신호를 증폭할 수 있는 비선형 압축 소자를 이용하여 이러한 문제를 해결한다. 실제 가정을 적용 대상의 환경으로 하고 있으므로 잡음에 대해 강인성이 요구된다. 음성코딩이나 음성통신에서 압축률 향상과 잡음에 대하여 강인성을 향상을 위해서 일반적으로 음성활성검출(VAD: voice activity detection) 방법을 사용하고 있는데 본 연구에도 유사한 접근 방법을 취하고 있다. 일반적인 VAD기법과는 달리 방향검지의 정확성을 목적으로 하는 경우에는 반드시 무성음을 포함하는 음성영역을 찾아야 할 필요는 없다. 왜냐하면 무성음의 경우 음성의 일부분이지만 주기성을 갖지 않기 때문에 이를 방향검지에 이용할 경우 잘 못된 결과를 얻게 되는 일이 빈번히 발생한다. 그래서, 일반적인 음성활성검출을 지양하고 3가지 음성 특징벡터 : 단구간 에너지, 피치, 영교차율을 이용하여 유성음 영역의 판별에 주안점을 두고 접근한다. 방향판별을 위하여 각 마이크로폰에 도달하는 신호의 시간 차이를 이용한다. 마이크로폰의 배치에 있어서는 신호처리분야에서 일반화된 일직선상의 배열은 앞과 뒤가 구분이 되지 않으며, 정면과 측면에서 각도의 정밀도가 일정하지 않다. 그래서, 정삼각형 형태로 마이크로폰을 배치하여 각 쌍의 위치와 시간차이의 관계를 바탕으로 퍼지추론을 통하여 모든 방위에 대하여 균일한 분해능을 가지며 실시간으로 음원의 방향을 검지한다.

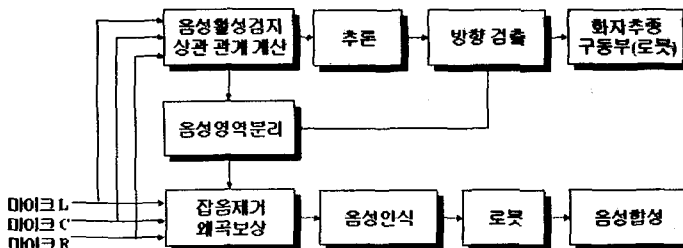


Fig. 1 Active Audition System

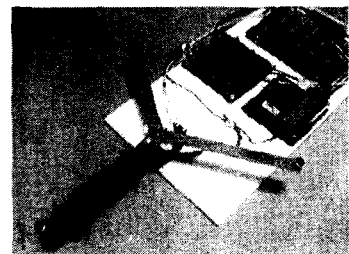


Fig. 2 Hardware Configuration