

퍼지를 이용한 인간형 능동 청각 시스템

Humanoid Active Audition System Using the Fuzzy Logic System

김 현 돈, 최 종 석, 이 창 훈, 박 귀 태, 김 문 상
(Hyun-Don Kim, Jong-Suk Choi, Chang-Hoon Lee, Gwi-Tea Park, and Mun-Sang Kim)

Abstract : We propose a humanoid active audition system which detects the direction of sound using just three microphones. Compared with previous researches, this system which has a simpler algorithm and less microphone shows better performance. Moreover, it uses the fuzzy logic system to remove the reflective sound and makes the resultant improvement of confidence. In order to verify our system's performance, we install the proposed active audition system to the home service robot, called hombot II, which has been developed at the KIST(Korea Institute of Science and Technology), and confirm its excellent performance through the experiment.

Keywords : active audition system, fuzzy logic system, speech recognition, detect the direction of sound

I. 서론

음성인식분야는 현재 여러 분야에 적용되고 있으며 성능도 많이 발전되었다. 특히 인간 친화적인 로봇을 개발하는 곳에서는 필연적으로 음성인식을 이용한 로봇과 인간과의 인터페이스 환경이 제공되었다.

따라서 신뢰성 있는 음성인식을 위해서 여러 음원에서 음성명령 신호만 분리해 내거나 각종 노이즈 제거에 관련된 기술 개발이 주요 연구목표로 떠오르고 있다. 뿐만 아니라 컴퓨터 비전 및 각종 인지센서와의 조합을 통하여 좀더 인간의 행동과 유사한 인간형 로봇이 개발되고 있는 추세이다[1]-[3].

본 연구에서는 로봇이 자신의 이름을 음성인식으로 인식 후 인간과 유사하게 음원의 방향을 검지하고 로봇의 얼굴 부분을 음원 방향으로 돌림으로써 화상인식과 연계하여 명령자의 얼굴을 인식하고 명령자와 마주보고 대화 할 수 있는 기술을 구현하고자 한다.

음원의 방향을 검지하기 위한 방법으로는 크게 입사신호의 파워로부터 얻어내는 방법(Power Estimation)과 도달시간 차이를 이용하는 방법(Time Delay Estimation)이 있다[4][5].

그 중에서 도달시간 차이를 이용하는 방법은 각 마이크로폰에 입사되는 지연시간차를 이용하는 방법이며 이는 반향신호의 영향이 적은 환경에서 우수한 성능을 나타낸다. 따라서 주로 이 방법을 사용하여 음원신호의 공간정보를 추출해 내는 연구가 개발되고 있다[3]-[5].

그러나 도달시간 차이를 이용한 방법 사용자 마이크로폰의 배열형태(array)의 배치[4][6]로 많은 수의 마이크로폰이

필요하며 계산량도 많아지게 되고 반향신호에 대한 영향이 많은 곳에서는 방향검지에 대한 신뢰도가 떨어지는 단점이 발생한다[5].

이를 보완하고자 이 논문에서는 최소한의 마이크로폰을 사용하고(3개) 처리시간이 짧은 방향검지 알고리즘을 적용하며 음원의 방향 검지시 가장 신뢰도를 떨어뜨리는 요소인 반향신호의 영향을 최소화하고 검지된 방향의 오차를 줄이기 위하여 퍼지로지직[7]-[9]을 적용하며 반향신호 필터링 역할 및 신뢰성있는 각도추론을 하도록 한다.

제안된 방향검지 시스템의 검증을 위하여 현재 한국과학기술연구원에서 개발중인 가정용 서비스로봇의 홈봇2에 탑재하여 음성인식 시스템과의 연계를 통한 구현으로 신뢰성 있는 성능을 확인하고자 한다.

II. 능동 청각 시스템

1. 능동 청각 시스템의 하드웨어 구성

그림 1은 사용된 능동 청각 시스템의 하드웨어 사진으로 제작된 전처리 보드 및 마이크가 장착된 삼각대, adventech사의 PCI-1711 A/D 보드로 구성되어 있다.

마이크로폰의 배치는 3개를 삼각형의 모양으로 배치하되 밑변의 중앙에서 각 마이크까지의 거리가 동일하도록 하였으며 각 마이크 당 비선형 증폭을 거친 신호를 11kHz의 속도로 A/D에서 샘플링 하게 된다.

2. 비선형 증폭 칩을 이용한 음성신호 검지 거리개선

특히 음원검출의 신뢰거리를 늘리기 위해서는 거리에 따른 신호의 크기에 따라서 능동적인 증폭을 할 수 있는 비선형 증폭이 필요하게 된다. 만약 증폭비가 일정하다면 동일한 크기의 음성신호라 할지라도 멀리서 발생한 음성신호는 크기가 줄게 되므로 노이즈에 상쇄되어 신호의 검출이 어려울 뿐만 아니라 검출된 신호를 적용하기에는 신호의 크기가 작으며 반면에 음성신호가 마이크로폰에 가까울수록 신호의 크기는 커지므로 증폭시 포화되어 신호가 왜곡

논문접수 : 2002. 11. 12., 채택확정 : 2003. 3. 17.

김현돈 : 고려대학교 전기공학부(reynolds@kist.re.kr)

최종석 : 한국과학기술연구원(cjs@kist.re.kr)

이창훈 : 배재대학교 정보통신공학부(naviro@pcu.ac.kr)

박귀태 : 고려대학교 전기공학부(gtpark@korea.ac.kr)

김문상 : 한국과학기술연구원(munsang@kist.re.kr)

* 본 논문은 과학기술부에서 지원하여 연구하였습니다.

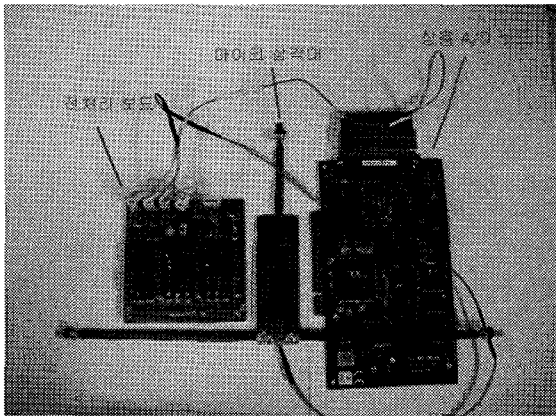


그림 1. 능동 청각 시스템의 하드웨어 실물사진.
Fig. 1. The photograph of Active Audition system.

될 가능성이 있다.

때문에 소리의 크기에 따라 증폭도를 달리하는 비선형 증폭회로를 구현함으로써 검출거리를 늘릴 수 있는 시스템이 필요하다. 이와 관련된 비선형 증폭회로를 구현시 상당히 복잡해지는 단점이 있으므로 여기서는 아날로그 디바이스에서 출시된 SSM2166 비선형 증폭칩 사용을 제안한다.

그림 2와 3에서 보듯이 SSM2166은 1:1 ~ 15:1의 압축비율을 조절할 수 있으며 증폭 한계치(Limiting threshold), 노이즈 입력 차단치(Noise Gate) 및 내부 증폭비등을 세부적으로 설정가능하며 하나의 칩에 위의 모든 기능이 집적되어 있으므로 회로가 간단해 지는 장점도 아울러 가지고 있다.

세부 설정방법을 알아보면 우선 압축비(Compression Regi on) 조절을 들 수 있다. 여기서 압축비란 그림 2에서 보듯이 comp ratio를 나타내며 1:1을 제외한 그래프에서 입력 대 출력의 증폭비가 신호의 크기에 따라서 틀린 것을 볼 수 있으며 압축비가 높을수록 상대적으로 크기가 센 신호의 증폭비율은 큰 영향이 없는 반면, 크기가 작은 신호의 입력에 대한 증폭비율이 높아지는 역할을 하는 것을 볼 수 있다. 따라서 압축비로 신호의 크기에 따라서 증폭비가 틀려지는 역할을 설정할 수 있으며 편 10번에 연결된 저항값을 조정하여 설정이 가능하다. 저항값에 따른 압축비는 표 1과 같다.

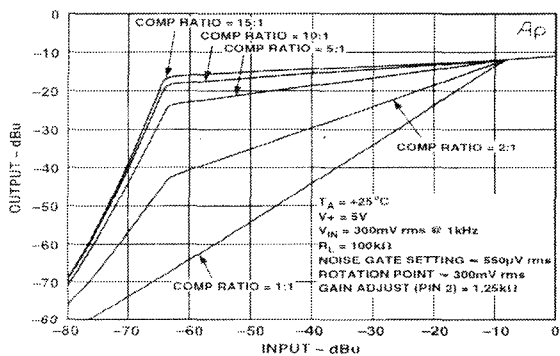


그림 2. SSM2166 입력력 신호 특성곡선.
Fig. 2. Output vs. Input Characteristics.

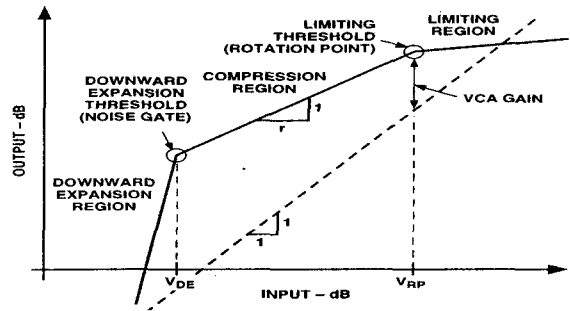


그림 3. SSM2166 일반적 입출력 특성곡선.
Fig. 3. General Input/Output Characteristics.

표 1. 저항의 변동값에 따른 압축비.

Table 1. Compression Ratio vs. Rcomp.

Compression Ratio	1:1	2:1	5:1	10:1	15:1
Rotation Point					
100mVrms	0	12.5kΩ	96kΩ	215kΩ	395kΩ
300mVrms	0	12.5kΩ	96kΩ	215kΩ	395kΩ
1Vrms	0	12.5kΩ	96kΩ	215kΩ	395kΩ

그림 4는 본 시스템에 맞도록 압축비를 1:2로 하여 저항값을 조절한 회로도를 나타낸다.

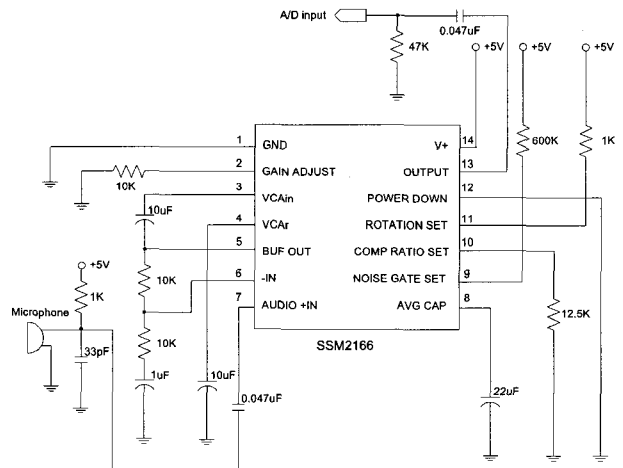


그림 4. 비선형 증폭 회로도.
Fig. 4. Circuit of Nonlinear Pre-Amp.

III. 방향검지 구현

이 논문에서 제안하는 음원의 방향을 검출하는 원리로 각 마이크에 도달하는 신호의 시간차를 이용하는 것이다. 마이크 사이의 떨어진 거리에 따라서 한 방향에서 들려오는 신호는 짧지만 시간차를 두고 각각의 마이크에 들어오게 된다. 이를 A/D의 고속 샘플링을 통하여 동시에 음성신호를 변환할 수 있으며 샘플링 구간에서 상관계수를 구하게 되면 얼마만큼의 시간차를 두고 들어왔는지 계산할 수 있다. 이로

써 각 시간차를 가지고 위치를 추적해 내는 원리이다.

그림 5에서의 같이 3개의 마이크 위치를 삼각대 중심에서 동일한 거리로 배치 시켰다면 마이크 C를 기준으로 'A와 C' 및 'B와 C' 두 개의 마이크 조합이 생기며 각 마이크의 일직선으로 신호가 들어 올 때 마이크 사이에 들어오는 신호가 최대로 지연됨을 알 수 있다.

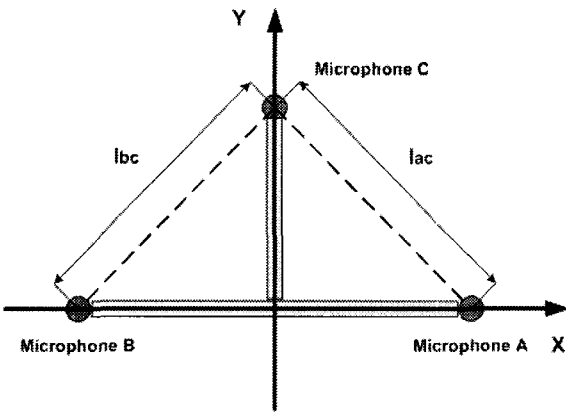


그림 5. 마이크의 위치도.
Fig. 5. Location of microphones.

이때 최대 지연되는 거리는 l_{ac} 와 l_{bc} 로 정의 할 수 있다.

$$n_{ac} = \frac{l_{ac}}{v} F_s \quad (1)$$

$$n_{bc} = \frac{l_{bc}}{v} F_s \quad (2)$$

또한 음속(voice velocity)을 v , 샘플링 주파수를 F_s 라고 놓으면 최대 지연거리에 대한 샘플링 개수를 (1), (2)와 같이 구할 수 있다. 여기서 n_{ac} 는 마이크 A와 C의 간의 최대 지연될 수 있는 샘플링 수이며 n_{bc} 는 B와 C마이크 간의 최대 지연될 수 있는 샘플링 수이다.

그러나 음원의 방향에 따라서 지연되는 순서는 틀리므로 방향에 따른 샘플링 값을 고려해 주어야 한다. 따라서 n_{ac} 와 n_{bc} 의 값은 마이크 A 와 B가 C보다 먼저 신호를 받았을 경우의 값이며 $-n_{ac}$ 와 $-n_{bc}$ 는 그 반대로 C가 A 와 B보다 먼저 신호를 받았을 경우이므로 결론적으로 샘플링 k 는 $-n_{ac} \sim n_{ac}$ 와 $-n_{bc} \sim n_{bc}$ 의 경우를 고려하여 적용하여야 한다.

$$R_{ac}(k) = \frac{\sum_{g=0}^{\infty} A(t_g - k)C(t_g)}{\sqrt{\sum_{g=0}^{\infty} A(t_g - k)^2} \sqrt{\sum_{g=0}^{\infty} C(t_g)^2}} \quad (3)$$

$$R_{bc}(k) = \frac{\sum_{g=0}^{\infty} B(t_g - k)C(t_g)}{\sqrt{\sum_{g=0}^{\infty} B(t_g - k)^2} \sqrt{\sum_{g=0}^{\infty} C(t_g)^2}} \quad (4)$$

위의 수식은 마이크 C와 A의 상관계수 (3)과 마이크 C와 B의 상관계수 (4)이다. 여기서 t_g 는 샘플링 구간에서 기준 샘플 target 값이며 위의 수식은 상관관계 적용 구간이 $g = 0$ 부터 ∞ 로 되어있으나 사실상 ∞ 의 영역을 고려해 주는

것은 불가능 하므로 적절한 샘플링 구간을 정해주어 계산 하도록 한다. 단, 샘플링 구간이 너무 작으면 이후 방향 검지시 오차의 범위가 늘어나며 너무 많아도 계산 시간도 많아지며 주변 반향신호의 영향을 받을 확률도 많아 역시 오차의 범위가 늘어남을 생각하여 실험적으로 최적의 값을 구하는 과정이 필요하다.

또한 k 는 지연된 샘플링 개수를 나타내므로 고려해 주어야 되는 범위는 이전에 언급했듯이 R_{ac} 의 경우 $-n_{ac} \sim n_{ac}$ 가 되고 R_{bc} 의 경우 $-n_{bc} \sim n_{bc}$ 이다. 따라서 실제 지연된 값 k_{ac} 을 구하기 위해서는 수식 3의 k 값에 지연 가능한 $-n_{ac} \sim n_{ac}$ 값을 모두 대입해 주면 마이크 C를 기준으로 마이크 A의 샘플링 지연값을 임의로 변경하면서 상관계수를 구할 수 있으며 이 과정을 통하여 양쪽 마이크에 들어오는 신호의 지연이 없을 때, 즉 각 마이크의 두개 신호가 가장 일치된 모양을 가질 경우 상관계수는 1과 근접한 최대값을 가지게 되며 이때의 k 값이 실제 지연된 샘플링 개수 k_{ac} 으로 볼 수 있으며 k_{bc} 의 경우도 마찬가지로 방법으로 구할 수 있다.

여기서 기존 제안된 방법과 비교해보면 실제 지연된 값을 구하기 위해서 사용되는 방법으로 주파수 영역에서 정규화된 상관계수를 구하는 방법을 주로 사용하게 된다. 이를 MCPSP(Modified Cross Power Spectrum Phase)라 하며 이때는 주파수 영역에서 지연된 값을 구해내는 방법으로 DF T(Discrete Fourier Transform) 및 IDFT(Inverse DFT)가 사용되며 이는 상당한 계산량을 필요로 하게된다[1]-[6].

또한 공간정보를 구해내기 위해서는 array 형태의 많은 마이크의 개수[4][6]가 필요하나 이 논문에서는 시간영역에서 상관계수를 구함으로써 간단하고 계산량이 적은 수식을 제안하였고 단 3개의 마이크로 음원의 방향을 찾아내는 방법을 구현하였다.

이제 마이크 사이의 지연을 모두 고려하여 R_{ac} 와 R_{bc} 의 상관계수가 계산이 되었으면 이를 이용하여 최종적으로 실제 음성신호의 방향이 어디인지 구하는 과정이 필요하다.

그림 6에서는 A와 C 마이크에서 음원이 들어오는 입사각과 실제 지연되는 샘플링 개수를 나타낸다.

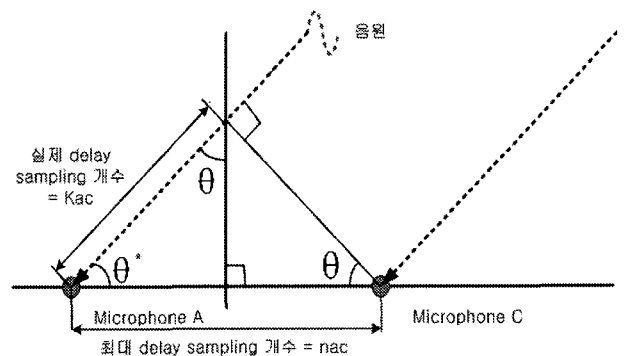


그림 6. 실제 delay sampling 값 kac 구간.
Fig. 6. kac section about number of delay sampling.

위에서 보듯이 A와 C의 마이크에서 들어오는 음원의 입사각은 수식 (5)와 (6)과 같이 나타낼 수 있다.

$$k_{ac} = n_{ac} \times \cos \theta^* \quad (5)$$

$$k_{ac} = n_{ac} \times \sin \theta \quad (6)$$

위에서 수식 (3)과 (4)를 통하여 이미 k_{ac} 및 k_{bc} 는 구해졌으므로 마이크 A 와 C의 음원의 입사각 θ 는 수식 (7)과 같이 나타낼 수 있으며 마이크 B 와 C의 경우에도 똑같은 방식으로 수식을 표현할 수 있다.

$$\theta = \sin^{-1} \frac{k_{ac}}{n_{ac}} \quad (7)$$

이제 구해진 A 와 C 또는 B 와 C 마이크사이의 음원의 입사각을 이용하여 마이크 C를 중심으로 삼각형으로 배치된 마이크 구조에서 음원의 최종 입사각을 검출해 내야한다.

그림 7을 보듯이 같은 종류의 선들은 서로 수평하며 실제 음원의 위치 (xi, yi)에서 음원이 수평으로 각 마이크에 들어온다고 가정하면 입사각도를 구하기 위해서는 θ 와 x축에서 마이크 A 와 C의 일직선이 기울어진 각도 θ_{ac} 를 더해 주면 된다.

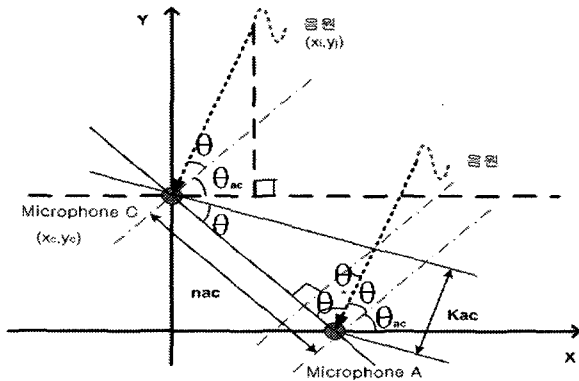


그림 7. k_{ac} 를 구하기 위한 θ 값 범위.

Fig. 7. θ section on account of k_{ac} calculation.

여기서 θ_{ac} 는 45°가 되며 x축에서 마이크 B와 C의 일직선이 기울어진 각도 θ_{bc} 는 135°가 된다.

$$\theta_{final} = \sin^{-1} \frac{k_{ac}}{n_{ac}} + \theta_{ac} \quad (8)$$

$$\theta_{final} = \sin^{-1} \frac{k_{bc}}{n_{bc}} + \theta_{bc} \quad (9)$$

위의 수식 (8)과 (9)가 최종 음원의 방향을 구하는 수식이며 이를 통해서 구해진 음원의 입사각도 θ_{final} 은 서로 일치하게 된다.

여기서 주의해 주어야 할 것이 있는데 그림 6을 예로들어 실제 k_{ac} 의 지연값을 가지는 입사각은 θ 뿐만 아니라 x축을 기준으로 맞은편에 같은 각도로 들어오는 입사각이 또 하나 존재하게 된다.

$$\theta_{ref} = 180^\circ - \theta \quad (180^\circ > \theta \geq 0^\circ) \quad (10)$$

$$\theta_{ref} = -180^\circ - \theta \quad (-180^\circ \leq \theta < 0^\circ) \quad (11)$$

수식 (10)과 (11)은 실제 들어오는 입사각과 동일한 지연 시간을 가지는 다른 하나의 각도 θ_{ref} 를 나타낸다.

따라서 실제 θ_{final} 값은 실제 입사각 θ 로 구해진 값과 θ 와 동일한 각도로 x축 맞은편으로 들어오는 입사각 θ_{ref} 로 구해진 두 가지의 값이 구해지게 되며 θ_{ref} 를 제거하기 위해서 나머지 마이크 B 와 C에 들어오는 입사각도 함께 구해주어 비교하여 서로 일치된 각도를 뽑아내면 최종 음원의 방향 θ_{final} 이 구해지게 된다. 따라서 A와 C 및 B 와 C 마이크의 입사각도를 서로 고려해 주어서 앞서 언급한 맞은편에서 들어오는 거짓각도를 제거하면 전 방위의 음원 방향을 구할 수 있게 된다.

IV. 반향신호 제거를 위한 퍼지로지 시스템

이제까지 음원의 방향검지를 위한 이론들을 살펴보았으나 일반적인 환경에서 위의 내용을 수행하면 실제와 전혀 다른 방향을 검지하는 경우가 많이 생긴다. 이유는 반향신호 때문이며 이는 주변에 벽 혹은 음원이 반사될 수 있는 조건이 많은 곳일수록 빈번히 일어 나게되어 방향검지의 신뢰도를 급격히 떨어뜨리게 된다.

이 문제점으로 인하여 각 마이크에 들어오는 음성신호의 지연 차이를 이용하기보다는 각 마이크에서 검출되는 음성신호의 크기를 처리하여 방향을 검출하기도 한다. 이 경우 반향신호는 원 신호보다 신호의 크기가 작게 되므로 반향신호에 대한 영향을 줄일 수 있으나 본 논문에서 제안한 신호의 상관관계를 이용하는 것은 반향신호도 본 신호와 파형이 일치하므로 문제가 발생한다.

이 논문에서는 음원의 방향 검지 시 반향신호에 의한 오검출을 줄이기 위해서 퍼지를 응용하였으며 3개의 검출된 각도를 이용하여 일정한 범위에 검출각도가 들어오면 음원의 방향 검지로 인식하여 결과를 출력하고 하나라도 일정한 각도를 벗어나는 신호를 검지하게 되면 반향신호로 인식하여 결과값을 출력하지 못하도록 하였다.

퍼지를 이용함으로써 인해서 크게 두가지의 이점을 얻을수 있는데 한가지는 반향신호의 제거이며 또 하나는 실제 음원의 방향 검지시 검출 방향에 따라서 8° ~ 14°의 오차가 필연적으로 발생하는데 퍼지를 사용함으로써 3개의 샘플된 각도값을 추론하여 중간값을 결정하여 각도의 분해력을 향상시킬 수 있다.

1. Fuzzifier

검출된 음원방향의 각도를 퍼지에 적용하기 위해서 아래 식과 같이 퍼지 입력값을 결정한다.

$$x_1 = \text{angle}(n-2) - \text{angle}(n-1) \quad (12)$$

$$x_2 = \text{angle}(n-1) - \text{angle}(n) \quad (13)$$

검출된 방향 3개의 값을 (12)와 (13)을 이용하여 구해진 각도의 차이 x_1 과 x_2 를 퍼지 멤버쉽 함수에 적용되게 된다.

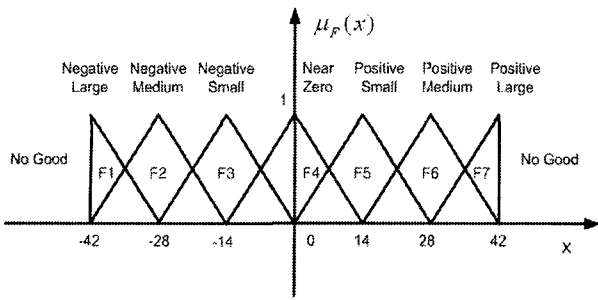


그림 8. 적용된 퍼지 멤버쉽 함수.
Fig. 8. membership function.

그림 8의 퍼지 멤버쉽 함수를 정의하면 다음과 같다.

$$\begin{aligned} \mu_{F_i}(x) &= \frac{1}{14}(-|x - b_i| + 14), \quad x \geq b_i \\ &= \frac{1}{14}(|x - b_i| + 14), \quad x < b_i \\ &= \text{No Good}, \quad x \geq 42, \quad x < -42 \end{aligned} \quad (14)$$

여기서 i 는 멤버쉽 함수의 개수로 $i=1, \dots, 7$ 이며, b_i 는 함수의 중심값으로 $b_i = -42, -28, -14, 0, 14, 28, 42$ 로 정의할 수 있다. 또한 위의 멤버쉽 함수의 x 축은 구해진 각도의 차이를 나타내며 만약 각도의 차이가 42° 이상 차이가 나게 되면 반향신호로 간주하여 신호를 무시하게 된다.

2. Rule & Inference

반향신호를 제거하기 위한 퍼지제어 규칙은 다음과 같은 조건문으로 표현 될 수 있다.

$$R^l: \text{IF } x_1 \text{ is } F_i \text{ and } x_2 \text{ is } F_j \text{ then } y_{ij} \text{ is } C_{ij} \quad (15)$$

(15)에서 $R^l(l=1, \dots, M)$ 는 M 개의 퍼지로직을 나타내며 F_i 는 x_1 및 F_j 는 x_2 의 퍼지 집합을 나타내며 C_{ij} 는 출력 y_{ij} 의 퍼지 집합을 나타낸다.

표 2. 반향신호 제거에 관한 퍼지제어 규칙표.
Table 2. Rule Table.

		x_1								
		NL	NM	NS	NZ	PS	PM	PL	NG	
x_2	NL	NG	NG	NL	NL	NM	NS	NZ	NG	
	NM	NG	NL	NL	NM	NS	NZ	PS	NG	
	NS	NL	NL	NM	NS	NZ	PS	PM	NG	
	NZ	NL	NM	NS	NZ	PS	PM	PL	NG	
	PS	NM	NS	NZ	PS	PM	PL	PL	NG	
	PM	NS	NZ	PS	PM	PL	PL	NG	NG	
	PL	NZ	PS	PM	PL	PL	NG	NG	NG	
	NG	NG	NG	NG	NG	NG	NG	NG	NG	NG

위의 규칙표에 따라 출력값이 결정되게 되며 수식 (12) 와 (13)에 의해 구해진 각도의 차이 x_1 혹은 x_2 가 42° 이상 차이되거나 x_1 과 x_2 의 전체 각도의 오차범위가 56° 이상 차이나면 위의 조건문에 의하여 NG(No Good)가 선택되어지

고 반향신호(noise)로 간주되어 출력신호는 무시되게 된다. 위의 정의한 퍼지 집합의 이름은 아래와 같다.

표 3. 퍼지집합의 이름.

Table 3. Name of membership function.

NL	Negative Large	PL	Positive Large
NM	Negative Medium	PM	Positive Medium
NS	Negative Small	PS	Positive Small
NZ	Near Zero	NG	No Good (Reject)

3. Defuzzifier

비 퍼지화는 평균 중심법(Center-of-Average method)을 사용하였다.

$$y = \frac{\sum_{i=1}^M \mu_{F_i}(x_i) y_i}{\sum_{i=1}^M \mu_{F_i}(x_i)} \quad (16)$$

(16)에서 y_i 는 i 번째 규칙 출력 소속함수의 정점이 있는 중심값이고 M 은 64개 이다.

$$\mu_{F_i}(x_i) = \min(\mu_{F_i}(x_1), \mu_{F_i}(x_2)) \quad (17)$$

여기서 각 퍼지 규칙에 속하는 소속함수는 (17)과 같이 T -norm (minimum)으로 정의된다.

$$\text{angle}(n) + y = \text{result} \quad (18)$$

최종 적용되는 출력은 (18)과 같이 최근의 검출된 각도에 퍼지가 적용된 보상값 y 를 더해줌으로서 최종 결과값이 출력되게 된다.

V. 홈봇 2 음성인식 시스템 및 적용결과

1. 로봇 적용사례

음성인식 시스템과 관련하여 현재 한국과학기술연구원(KIST) 휴먼로봇연구센터에서 개발중인 홈 서비스 로봇인 홈봇 2에 직접 적용 하였다. 데모를 수행시 시나리오는 3명의 화자가 서로 번갈아 가면서 다른 위치에서 ‘홈 나들봐’ 라는 음성명령을 주었을때 음성인식에 의하여 음성명령을 인지하고 음원의 방향으로 카메라를 회전후 화상인식을 실행하게 되어있다.

그림 9는 홈봇 2가 방향을 검지하고 음성을 인식하는 순서도를 나타낸다.

2. 방향검지 실험결과

로봇에 적용될 음성 모듈 개발시 가장 핵심적인 부분이 음원의 방향검지 부분이다. 실험 결과는 표 4 및 표 5와 같다. 실험 조건은 앞면에 언급한 알고리즘을 이용하였고 반향신호 제거를 위해 퍼지 로직을 적용 하였다. 또한 추가로 고려해 주어야 할 것은 음원의 일정 크기이상인 검출 되었을 때 방향을 검지하는 부분이 필요하다. 만약 이 부분을 고려해 주지 않는다면 극단적으로 조용한 곳에서 노이

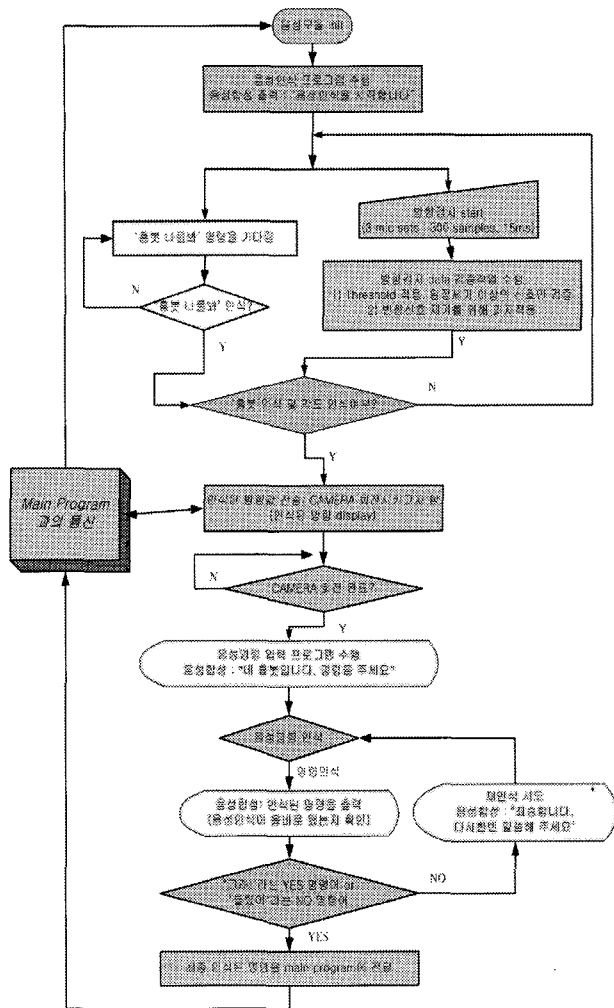


그림 9. 홈봇 2 음성인식 순서도.
Fig. 9. Flow chart of voice sequence.

즈만으로도 방향을 검출하게 된다. 따라서 여기서는 일정한 시간마다(1초당 10번 수행) 샘플링을 수행하고 신호의 크기가 일정 이상의 값이 나왔을 때 방향검지 알고리즘을 수행하게 되어있다.

표 4와 5는 0°~360° 사이의 전 방향에 대하여 45° 각도 만큼 구간을 나누어서 각 방향마다 “홈봇 나를봐” 라는 음성을 말했을 시 이에 대한 방향이 검지된 실험 결과이다. 각 방향마다 10번을 시도하였고 이에 대한 각도오차의 평균을 마지막 오차간에 적었으며 마지막 줄의 오차에 대한 평균을 평균칸 아래에 적어 놓았다.

단, 표 4에서는 반향신호의 영향으로 전혀 엉뚱한 값이 계산된 경우(표내 음영이 되어있는 곳)는 평균오차에서 제외되었다.

표 4에서와 같이 반향신호를 전혀 고려해 주지 않았을 경우 전혀 다른 각도(45° 이상 오차발생)를 검출해 내는 경우가 총 100번의 실험 결과 중 13번(표내 음영이 되어있는 곳)이 검지되었다. 이는 반향신호에 의해서 나타나는 현상으로 실제 로봇에 적용시 엉뚱한 방향을 나타내어 문제점을 발생시킨다. 또한 반향신호를 제외한 실제 데이터의 평

표 4. 퍼지 미 적용시 방향검지 실험결과.

Table 4. Result for detection of voice direction without fuzzy logic system.

	45°	90°	135°	180°	225°	270°	315°	360°	평균
1	315°	342°	137°	189°	238°	262°	315°	8°	
2	32°	90°	0°	189°	122°	267°	315°	0°	
3	32°	90°	167°	171°	223°	160°	315°	199°	
4	43°	81°	155°	181°	224°	128°	0°	348°	
5	32°	77°	135°	189°	255°	0°	314°	348°	
6	31°	90°	137°	198°	238°	258°	283°	258°	
7	32°	102°	161°	189°	225°	258°	283°	283°	
8	32°	90°	180°	180°	213°	262°	306°	348°	
9	45°	90°	155°	164°	228°	252°	270°	348°	
10	43°	77°	161°	171°	238°	270°	303°	355°	
오차	9°	5°	16°	9°	10°	9°	11°	9°	10°

표 5. 퍼지 적용시 방향검지 실험결과.

Table 5. Result for detection of voice direction using fuzzy logic system.

	45°	90°	135°	180°	225°	270°	315°	360°	평균
1	51°	95°	144°	190°	238°	264°	298°	360°	
2	36°	81°	146°	183°	214°	273°	312°	352°	
3	50°	87°	129°	189°	231°	262°	318°	358°	
4	42°	98°	149°	190°	211°	266°	328°	5°	
5	54°	87°	145°	183°	217°	280°	320°	352°	
6	45°	80°	136°	172°	215°	279°	315°	360°	
7	45°	99°	135°	184°	213°	275°	301°	9°	
8	41°	93°	136°	176°	222°	279°	312°	352°	
9	40°	84°	147°	189°	225°	275°	306°	352°	
10	51°	90°	129°	180°	213°	254°	313°	352°	
오차	5°	6°	7°	6°	9°	8°	7°	6°	7°

균오차는 10°이며 5°~16°오차 범위를 나타내는 것을 알 수 있다.

이에 반하여 표 5는 퍼지 로직을 적용한 결과이며 표 5에서 보듯이 반향신호의 영향은 없는 것을 알 수 있다. 또한 평균 오차는 7°로 퍼지 로직을 적용하지 않았을때와 비교하여 약 3°가량의 오차가 줄었으며 오차범위 또한 5°~9°로 4°의 오차범위를 보였다. 따라서 적용하지 않았을 때의 오차범위 11°와 비교해 보면 퍼지 로직을 적용 했을 때가 좀더 신뢰성 있는 방향을 검지할뿐만 아니라 동일한 방향에 대하여 좀더 일정한 결과값을 출력함을 알수있다. 이는 적용한 퍼지 로직에 의하여 이전에 얻은 3개의 각도가 퍼지로 추론 되어지고, 그 결과 구해진 보상값이 출력값에 고려가 되기 때문에 오차범위가 줄어든다.

홈봇 2 에 적용시 음원의 방향을 검지후 화상인식에 쓰였으며 평균 10° 이내의 방향오차는 화상검지에 심각한 영향을 줄 정도로 크지는 않았다.

3. 음성인식 실험결과

홈봇 2에서 음성인식은 무선마이크를 통하여 인식되었으므로 방향검지와는 직접적인 관련이 없다고 말할 수 있다.

표 6. 음성인식을 실험결과.

Table 6. Result for speech recognition.

단문	여기로와	100%	70%	80%
	나를봐	100%	90%	90%
	정지해	100%	100%	100%
인식률	평균 인식률 (92%)	100%	87%	90%
2 문장 조합	거실로 가	90%	100%	70%
	오른쪽으로 회전해	100%	80%	80%
	신문을 가져와	100%	100%	100%
	책을 집어	100%	100%	90%
	우리집을 순찰해	80%	80%	60%
	맵을 구성해	80%	90%	90%
	원격으로 제어해	70%	90%	60%
쉬어모드로 전환해	100%	100%	100%	
인식률	평균 인식률 (88%)	90%	93%	81%
4 문장 조합	테이블 위에 컵을 가져와	60%	90%	70%
	쇼파 앞에 책을 집어	40%	60%	100%
인식률	평균 인식률 (72%)	50%	75%	85%

하지만 화자의 음성명령이 무선마이크를 통해 입력되고 사람의 행동패턴과 유사하게 화자를 향해 로봇의 얼굴부분을 돌린 후 음성명령을 인식하는 것이 최종목적이라고 할 때는 음성인식과 관련된 작업도 요하게 되므로 여기서 간단히 언급하도록 한다.

홈봇 2에서 음성인식은 L&H 사의 SDK(Software Developer Kit)를 사용하였다. 이를 이용하여 visual C++ 개발 tool 을 사용하여 프로그래밍 하였다. L&H의 특징으로 단어를 조합하여 문장인식을 할 수가 있으며 데이터는 컴퓨터의 사운드카드에서 얻어 낼 수 있다.

표 6에서 보듯이 인식률은 문장의 조합이 복잡할수록 인식률이 낮아지며 인식률의 변화폭도 높은 것을 알수있다. 그러나 2문장 조합 이하의 음성인식률은 90%이상으로 실효성이 있는 것을 알 수 있다. 단지 외부의 노이즈가 많으면 인식률이 급격히 떨어지는 특징과 등록된 단어와 관련 없는 노이즈가 들어와도 음성인식 엔진이 동작하여 등록된 가장 유사한 단어를 끄집어 내는 단점이 있으며 음성인식 엔진의 특징을 고려할 때 보완하여야 될 문제는 많은 상황이다. 다시 말해서 노이즈를 분리해 내는 기술 및 원하지 않은 음원이 들어왔을 경우 거부해 내는 기술 등의 향상이 음성인식을 실생활에 적용하는데 가장 큰 전제 조건임을 알 수 있다.

음성인식을 로봇에 적용함에 있어서 한 가지 더 고려해 주어야 할 기능이 음성합성에 관한 것이다. 음성합성을 통하여 인식된 음성명령이 잘 인식 되었는지 확인 할 수 있을 뿐 아니라 음성의 인식 및 출력을 동시에 적용함으로써 로봇과 사람과의 대화형 인터페이스를 구현 할 수 있다. 여기서는 음성합성 엔진으로 국내 보이스웨어 제품을 사용하였다. 음성합성은 TTS(Text-To-Speech) 방식으로 text로 입력된 문장을 문장 분석을 통하여 운율과 리듬에 맞게 합성하

여 최대한 자연스러운 음성을 스피커를 통하여 출력하게 되어있다.

VI. 결론

본 논문에서는 이전에 제시되었던 array 형태의 마이크를 지양하고 단 3개의 마이크를 사용하여 전 방위에서 발생된 음원의 방향을 검지해 낼 수 있는 간단하고 신뢰성 있는 알고리즘 제안 및 하드웨어를 구현하였다.

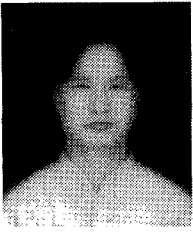
또한 음성신호의 방향검지를 위한 방법중 하나인 마이크에 도달하는 음성신호의 시간차를 이용한 방법의 단점인 반향신호의 영향을 최소화하기 위해서 퍼지 로직을 적용하였다. 때문에 반향신호의 발생이 많은 곳에서도 신뢰성 있게 방향을 검지해 낼 수 있었다.

이를 홈 서비스 로봇에 적용하여 실제 음원의 방향을 인식하여 화상인식 카메라를 돌려 명령자의 화상인식 및 음성명령을 인식하는 기능을 확인하였다.

향후 실생활에 직접 적용하기 위한 음성인식 및 방향검지를 위하여 여러 가지 음원이 섞여 들어오는 신호에서 사람의 말소리만 분리해 내고 주변 노이즈 제거에 대한 신뢰성 있는 기술 등이 연구되어야 한다.

참고문헌

- [1] K. Nakadai, T. Matsui, H. G. Okuno, H. Kitano, "Active audition system and humanoid exterior design," *IEEE/RSJ* vol. 2, pp. 1453-1461, 2000.
- [2] K. Nakadai, H. G. Okuno, H. Kitano, "Epipolar geometry based sound localization and extraction for humanoid audition", *IEEE/RSJ* vol. 3, pp. 1395-1401, 2001.
- [3] S. A. Sekmen, M. Wilkes, K. Kawamura, "An application of passive human-robot interaction: Human tracking based on attention distraction", *IEEE Trans. Sys., Man and Cybern.*, vol. 32, no. 2, pp. 248-259, 2002.
- [4] T. Nishiura, T. Yamada, S. Nakamura, K. Shikano, "Localization of multiple sound sources based on a CSP analysis with a microphone array," *IEEE/ICASSP* vol. 2, pp. 1053-1056, 2000.
- [5] 이학주, 김효원, 이규하, "마이크로폰 어레이를 이용한 화자 추적 시스템 구현," TI korea 제 4회 DSP 디자인 컨테스트 우수 논문집, 2001
- [6] H. Kamiyanagida, H. Saruwatari, K. Takeda, F. Itakura, "Direction of arrival estimation based on nonlinear microphone array", *IEEE Int. Conf. Acoustics, Speech, and Signal Process*, vol. 5, pp. 3033-3036, 2001.
- [7] J. M. Mendel, "Uncertain rule-based fuzzy logic systems," Prentice Hall
- [8] C. C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller, part I", *IEEE Trans. Sys., Man and Cybern.*, vol. 20, no. 2, pp. 404-418, 1990.
- [9] C. C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller, part II", *IEEE Trans. Sys., Man and Cybern.*, vol. 20, no. 2, pp. 419-435, 1990.



김 현 돈

1974년 12월 17일생. 1997년 고려대학교 제어계측공학과 (공학사). 2002년~현재 고려대학교 전기공학과 석사과정, KIST 학생연구원. 관심분야는 휴먼로봇, Embedded System, speech recognition system.



최 중 석

1970년 8월 18일생. 1994년 한국과학기술원 전기공학과 (공학사). 1996 한국과학기술원 전기공학과 (공학석사). 한국과학기술원 전기공학과 (공학박사). 2001년~현재 KIST 선임연구원. 관심분야는 이동로봇제어, 최적제어,

로봇추구.



이 창 훈

1965년 6월 25일생. 1989년 연세대학교 전자공학과 (공학사). 1991년 연세대학교 전자공학과 (공학석사). 1999년 일본 동경공업대학 시스템과학과 (공학박사). 2002년~현재 배재대학교 전자공학과 전임강사. 관심분야는 지능

제어, 로보틱스, 시스템의 안정도 해석과 응용, 임베디드 시스템, 디지털 워터마킹.



박 귀 태

1947년 10월 25일생. 1975년 고려대학교 전기공학과 (공학사). 1977년 고려대학교 전기공학과 (공학석사). 1981년 고려대학교 전기공학과 (공학박사). 1981년~현재 고려대학교 전기공학과 교수, IBS Korea 회장. 관심분야는

intelligent controls, image processing, mobile robotics, home automation, intelligent building system(IBS).



김 문 상

1957년 12월 4일생. 1980년 서울대 기계설계공학과 (공학사). 1982년 서울대 기계설계공학과 (공학석사). 1987년 독일 Technical Univ. of Berlin (공학박사). 1987년~현재 한국과학기술연구소 책임연구원. 관심분야는 로봇공학, 햅

틱 시스템.