



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0097296
(43) 공개일자 2011년08월31일

(51) Int. Cl.

G10L 21/06 (2006.01) G10L 21/02 (2006.01)

(21) 출원번호 10-2010-0017069

(22) 출원일자 2010년02월25일

심사청구일자 2010년02월25일

(71) 출원인

서강대학교산학협력단

서울 마포구 신수동 1-1 서강대학교

경북대학교 산학협력단

대구광역시 북구 산격동 1370 경북대학교내

(72) 발명자

박형민

서울특별시 서초구 잠원동 56-3 잠원한신아파트
5동 903호

이수연

서울특별시 광진구 중곡4동 4

이민호

대구광역시 수성구 범어4동 장원맨션 102동 1203
호

(74) 대리인

이지연

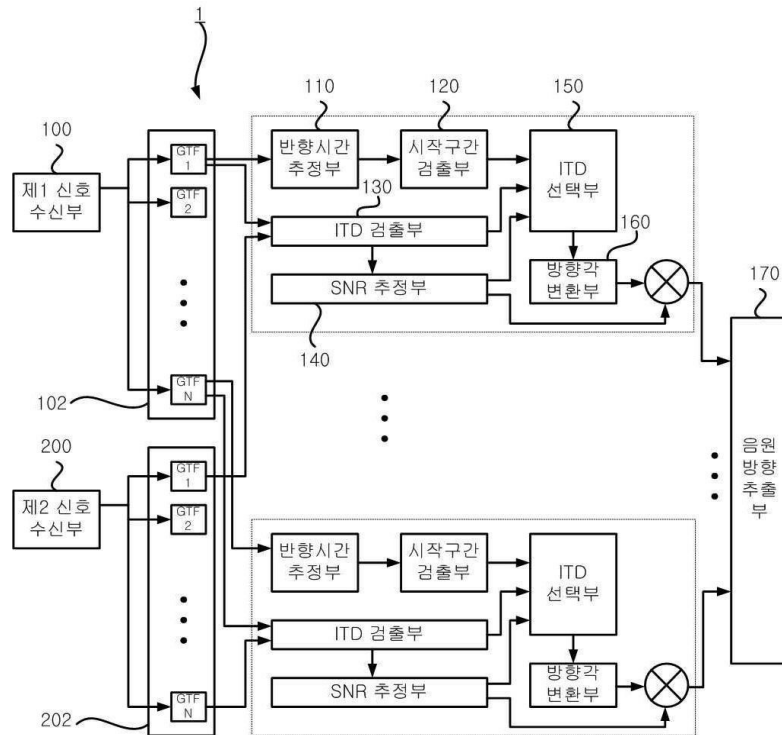
전체 청구항 수 : 총 10 항

(54) 반향 환경에서의 영교차 기반 다음원 국지화 방법 및 장치

(57) 요약

본 발명은 다음원 국지화 장치에 관한 것이다. 다음원 국지화 장치는, 외부로부터 신호를 수신하는 제1 및 제2 신호 수신부; 상기 제1 및 제2 신호 수신부로부터 제공되는 신호를 사전에 설정된 주파수 대역별로 분리하여 각 채널로 출력하는 제1 및 제2 필터 뱅크; 상기 제1 필터 뱅크의 각 채널로부터 출력된 신호의 포락선 정보를 이용하여 반향 시간을 추정하는 반향 시간 추정부; 상기 반향 시간을 이용하여 반향이 없는 시작 구간을 검출하는 시작 구간 추정부; 영교차점을 기반으로 하는 ITD를 추정하는 ITD 추정부; 상기 채널들로부터 출력된 신호들에 대한 신호 대 잡음비(SNR)를 검출하는 SNR 추정부; 시작 구간 및 SNR들을 이용하여 상기 ITD들 중 신뢰성있는 ITD들을 선택하는 ITD 선택부; 상기 ITD 선택부에 의해 선택된 ITD들을 방향각으로 변환시키는 방향각 변환부; 상기 방향각 변환부로부터 출력된 방향각에 추정된 SNR 가중치로 누적시켜 히스토그램을 생성하고, 상기 히스토그램을 이용하여 음원의 방향을 추출하는 음원 방향 추출부; 를 구비한다. 상기 반향 시간 추정부, 시작 구간 추정부, ITD 추정부, SNR 추정부, ITD 선택부 및 방향각 변환부는 제1 필터 뱅크의 모든 채널에 각각 구비되고, 모든 방향각 변환부로부터 출력된 방향각 정보들은 음원 방향 추출부로 제공되어 반향 환경에서의 음원 방향을 추출한다.

대표도



이 발명을 지원한 국가연구개발사업

과제고유번호 2009-0086194 / 200912240100

부처명 한국연구재단 / 교육과학기술부

연구관리전문기관

연구사업명 중견연구자지원사업 / 신기술 융합형 성장동력사업

연구과제명 인지 향상 기술 개발 모바일 단말기를 위한 지능형 음성대화 인터페이스 개발 / 다중감각추론 일반 모델 및

기여율

주관기관 서강대학교 산학협력단 / 경북대학교

연구기간 2009년 09월 01일 ~ 2010년 08월 31일 / 2009년 07월 01일 ~ 2010년 06월 30일

특허청구의 범위

청구항 1

외부로부터 신호를 수신하는 제1 신호 수신부;

상기 제1 신호 수신부와 일정 거리 이격된 위치에 배치되어 외부로부터 신호를 수신하는 제2 신호 수신부;

상기 제1 신호 수신부로부터 제공되는 신호를 사전에 설정된 주파수 대역별로 분리하여 각 채널로 출력하는 제1 필터 뱅크;

상기 제2 신호 수신부로부터 제공되는 신호를 사전에 설정된 주파수 대역별로 분리하여 각 채널로 출력하는 제2 필터 뱅크;

상기 제1 필터 뱅크의 각 채널에 연결되며, 상기 연결된 채널로부터 신호의 포락선 정보를 수신하고, 수신된 포락선 정보를 이용하여 반향 시간을 추정하는 반향 시간 추정부;

상기 반향 시간 추정부로부터 반향 시간을 제공받고, 반향 시간을 이용하여 반향이 없는 시작 구간을 검출하는 시작 구간 추정부;

상기 제1 및 제2 필터 뱅크의 각 채널에 연결되며, 상기 연결된 채널들로부터 수신된 신호들을 이용하여 영교차점을 기반으로 하는 ITD를 추정하는 ITD 추정부;

상기 채널들로부터 수신된 신호들에 대한 신호 대 잡음비(SNR)를 검출하는 SNR 추정부;

상기 ITD 추정부에 의해 추정된 ITD들, 상기 시작 구간 추정부에 의해 검출된 시작 구간, 상기 SNR 추정부에 의해 검출된 각 신호에 대한 SNR들을 입력받고, 시작 구간 및 SNR들을 이용하여 상기 입력된 ITD들 중 신뢰성있는 ITD들을 선택하는 ITD 선택부;

상기 ITD 선택부에 의해 선택된 ITD들을 방향각으로 변환하는 방향각 변환부;

상기 방향각 변환부로부터 출력된 방향각들을 누적시켜 히스토그램을 생성하고, 상기 히스토그램을 이용하여 음원의 방향을 추출하는 음원 방향 추출부;

를 구비하고, 상기 반향 시간 추정부, 시작 구간 추정부, ITD 추정부, SNR 추정부, ITD 선택부 및 방향각 변환부는 제1 필터 뱅크의 모든 채널에 각각 구비되고, 모든 방향각 변환부로부터 출력된 방향각 정보들은 음원 방향 추출부로 제공되어 반향 환경에서의 음원 방향을 추출하는 것을 특징으로 하는 다음원 국지화 장치.

청구항 2

제1항에 있어서, 제1 및 제2 필터 뱅크는 감마톤 필터 뱅크로 구성되는 것을 특징으로 하는 다음원 국지화 장치.

청구항 3

제1항에 있어서, 상기 반향 시간 추정부는 반향 시간을 추정하며, 반향 시간(RT60)은 신호가 수신된 후 수신된 신호의 세기가 초기의 신호보다 60dB 감소되기까지의 시간을 나타내는 것을 특징으로 하는 다음원 국지화 장치.

청구항 4

제3항에 있어서, 상기 반향 시간 추정부는 하기 수학식을 반복적으로 학습하여 신호의 포락선의 감쇄 상수 (α)의 최대 우도를 추정하며, 추정된 포락선의 감쇄 상수를 이용하여 반향 시간(RT60)을 계산하는 것을 특징으로 하는 다음원 국지화 장치.

$$\Delta a \propto \frac{\partial \ln L(y(n))}{\partial a} = \frac{M}{a} \left[-\frac{M-1}{2} + \frac{\sum_{n=0}^{M-1} n a^{-2n} y^2(n)}{\sum_{n=0}^{M-1} a^{-2n} y^2(n)} \right]$$

여기서, a 는 포락선의 감쇄 상수, M 은 표본의 개수, $y(n)$ 은 필터 뱅크 특정 채널의 n 번째 출력 신호임.

청구항 5

제1항에 있어서, 시작 구간 추정부는 반향 시간을 이용하여 반향 신호를 추정하고, 추정된 반향 신호 ($E_i(n)$)와 원 입력 신호($env_i(n)$)가 하기의 수학적식을 만족하는 구간을 시작 구간으로 설정하는 것을 특징으로 하는 다음원 국지화 장치.

$$\frac{env_i(n)}{E_i(n)} \geq \rho_{th}$$

여기서, ρ_{th} 는 원 입력 신호와 추정된 반향 신호의 비율의 한계값임.

청구항 6

제1항에 있어서, 상기 방향각 변환부는 선택된 각 ITD들을 방향각으로 변환시킨 후, 각 ITD에 해당하는 신호의 SNR를 가중치로 하여 방향각들을 가중시켜 가중된 방향각을 출력하는 것을 특징으로 하는 다음원 국지화 장치.

청구항 7

서로 일정 거리 이격된 위치에 배치된 2개의 신호 수신부들로부터 입력된 신호들을 주파수 대역별로 분리하여 채널별로 출력하는 2개의 필터 뱅크로부터 출력되는 신호들을 이용하여 음원의 방향을 추정하는 방법에 있어서,

- 상기 필터 뱅크의 채널로부터 수신된 신호의 포락선 정보를 이용하여 반향 시간(RT60)을 추정하는 단계;
- 상기 추정된 반향 시간을 이용하여 반향이 없는 시작 구간을 검출하는 단계;
- 상기 필터 뱅크의 채널들로부터 출력된 신호들을 이용하여 영교차점을 기반으로 하는 ITD를 추정하는 단계;
- 상기 필터 뱅크의 채널들로부터 출력된 신호들에 대한 신호 대 잡음비(SNR)를 검출하는 단계;
- 상기 시작 구간 및 SNR들을 이용하여 상기 검출된 ITD들 중 신뢰성있는 ITD들을 선택하는 단계;
- 상기 선택된 ITD들을 방향각으로 변환시키는 단계;
- 상기 (a) 내지 (f) 단계는 필터 뱅크의 모든 채널에 대하여 반복적으로 수행하여, 모든 채널에 대하여 방향각들을 획득하는 단계;
- 모든 채널에 대하여 획득한 방향각들을 누적시켜 히스토그램을 생성하고, 상기 히스토그램을 이용하여 음원의 방향을 추출하는 단계;

를 구비하여, 반향 환경에서의 음원 방향을 추출하는 것을 특징으로 하는 다음원 국지화 방법.

청구항 8

제7항에 있어서, 상기 (a) 단계의 반향 시간(RT60)은 신호의 세기가 초기의 신호보다 60dB 감소되기 까지의 경과 시간을 나타내는 것을 특징으로 하는 다음원 국지화 방법.

청구항 9

제7항에 있어서, 상기 (a) 단계는 하기 수학적식을 반복적으로 학습하여 신호의 포락선의 감쇄 상수(α)의 최대 우도를 추정하며, 추정된 포락선의 감쇄 상수를 이용하여 반향 시간(RT60)을 계산하는 것을 특징으로 하는 다음원 국지화 방법.

$$\Delta a \propto \frac{\partial \ln L(y(n))}{\partial a} = \frac{M}{a} \left[-\frac{M-1}{2} + \frac{\sum_{n=0}^{M-1} n a^{-2n} y^2(n)}{\sum_{n=0}^{M-1} a^{-2n} y^2(n)} \right]$$

여기서, α 는 포락선의 감쇄 상수, M 은 표본의 개수, $y(n)$ 은 필터 뱅크 특정 채널의 n번째 출력 신호임.

청구항 10

제7항에 있어서, 상기 (b) 단계는 반향 시간을 이용하여 반향 신호를 추정하고, 추정된 반향 신호($E_i(n)$)와 원 입력 신호($env_i(n)$)가 하기의 수학적식을 만족하는 구간을 시작 구간으로 설정하는 것을 특징으로 하는 다음원 국지화 방법.

$$\frac{env_i(n)}{E_i(n)} \geq \rho_{th}$$

여기서, ρ_{th} 는 원 입력 신호와 추정된 반향 신호의 비율의 한계값임.

명세서

기술분야

[0001] 본 발명은 반향환경에서의 영교차 기반 다음원 국지화 방법 및 장치에 관한 것으로서, 더욱 구체적으로는 영교차점을 기반으로 한 두 신호간 시간차(Interaural Time Difference; 이하 'ITD'라 한다)를 이용한 종래의 국지화 방법에 반향이 존재하지 않는 음원의 시작 구간을 검출하는 방법을 적용함으로써, 정확한 음원의 방향을 찾아낼 수 있는 반향환경에서의 영교차 기반 다음원 국지화 방법 및 장치에 관한 것이다.

배경기술

[0002] 음원 위치 국지화는 여러 다른 소리들과 잡음이 혼재하는 환경에서 핸드즈프리(hands-free) 시스템 사용 또는 음성 회의(audio conference)시 매우 중요한 역할을 한다. 인간은 잡음이 많이 존재하는 환경에서도 여러 음원의 위치를 추적하는 데 뛰어난 능력을 보이고 있다. 특히 두 귀에서 감지하는 신호의 시간차(Interaural Time Difference, ITD)와 신호의 세기차(Interaural Intensity Difference, IID)가 중요한 요소이다. 신호의 시간차(ITD)는 배경 잡음에 취약하고 신호의 세기차(IID)는 반향환경에서 취약하다고 알려져 있다. 최근에 각 주파수 대역에서 신호들의 영교차점(zero-crossing, ZC)을 비교하여 신호의 시간차(ITD)를 계산하는 방법이 제안되었다. 이 방법은 적은 계산량으로 잡음 환경 하에서 원하는 소스의 방향을 추적하는데 뛰어난 성능을 보여주고

있다. 비록 이 방법은 확산잡음 환경 하에서는 좋은 성능을 보이지만 실제 환경에 적용하기 위해서는 반향을 고려해야 된다.

[0003] 반향이 존재하는 곳에서의 신호는 음원으로부터 마이크까지 직선 경로를 통해 입력되는 신호뿐 아니라 주위 벽에 반사되어 나중에 들어오는 반향 신호들이 존재하므로, 반향 신호에 의해 잘못된 음원 위치 정보를 갖게 되는 신호의 시간차 값이 발생된다.

[0004] 한편, 반향 환경에 대한 일반적인 채널 필터는 직선 경로(direct path) 성분, 초기 반사(early reflection) 성분 그리고 지연 반사(late reflection) 성분 이렇게 세 가지 성분으로 구분할 수 있다. 직선 경로는 음원으로부터 마이크 사이의 가장 짧은 경로로서, 직선 경로를 통과하여 마이크에 도달한 신호가 다른 경로를 통과한 신호보다 빨리 도착하게 되고 이 신호만이 음원의 방향에 상응하는 신호의 시간차(ITD)를 얻는데 도움이 된다. 음원 신호의 시작 구간(onset)은 마이크에 입력되는 신호 중 직선경로를 통해서 마이크에 도달한 신호가 지배적이므로 상대적으로 반향에 영향을 받지 않는 구간이다. 따라서 음원의 시작 구간 검출은 반향에 영향을 받지 않는 신호의 시간차를 구함으로써 음원의 위치를 추적하는 데 중요한 역할을 한다. 음원의 시작 구간 검출을 이용한 음원 위치 추적 연구에는 두가지 방법이 효과적이었다. 하나는 영교차점에 기반하여 입력 신호의 포락선(envelope)에서 에너지가 갑작스럽게 증가하는 부분을 음원의 시작구간이라 가정하고 검출하는 방법이며, 다른 하나는 선작효과의 반향 회피 모델을 이용한 방법이다.

[0005] 전자의 방법에서는, 입력 신호의 평활(smoothing)된 포락선의 현재 값과 Δn -샘플 이전 값을 비교하여 그 비율이 특정의 한계 값 이상일 경우 그 부분을 음원의 시작구간으로 설정한다. 이 간단한 알고리즘은 적은 계산량으로도 음원 신호의 시작 구간을 쉽게 구할 수 있지만 그 과정에서 변수 값들과 입력 신호에 매우 민감하다.

[0006] 후자의 방법에서는, 반향 회피(echo avoidance) 모델에 기반하여 반향이 없는 음원의 시작 구간을 검출한다. 실제 환경에서 우리는 일반적으로 반향에 영향을 받은 신호를 인지하게 된다. 따라서 인지된 신호의 대부분은 반향의 영향을 받은 시간차를 갖게 된다. 원하는 음원의 방향은 반향의 영향을 받지 않은 구간, 즉 묵음 구간 뒤에 갑자기 음성이 시작되는 부분에서만 얻어진다. 반향에 영향을 받지 않는 음원의 시작 구간은 원 신호와 반향 신호의 비율이 큰 부분을 선택하여 검출해 낼 수 있다. 여기서 반향은 정형화된 임펄스 응답(impulse response)을 이용하여 이전 음성으로부터 추정된다. 이 방법은 음원 신호의 크기에 상관없으며 잡음에 민감하지 않는 특성이 있다. 전자의 방법과 비교할 때 이 방법이 훨씬 더 변수들에 강인하지만 여전히 반향환경은 이 알고리즘의 성능을 저하시키는 문제점을 갖고 있다.

[0007] 따라서, 본 발명에서는 반향이 큰 환경에서도 잘 적용할 수 있는 음원 위치 추적 방법에 대해 제안하고자 한다.

발명의 내용

해결하려는 과제

[0008] 전술한 문제점을 해결하기 위한 본 발명은 반향이 적은 환경뿐만 아니라 반향이 많은 환경에서도 잘 적용될 수 있는 다음원 국지화 방법 및 장치를 제공하는 것이다.

과제의 해결 수단

[0009] 전술한 기술적 과제를 달성하기 위한 본 발명의 특징에 따른 다음원 국지화 장치는, 외부로부터 신호를 수신하는 제1 신호 수신부; 상기 제1 신호 수신부와 일정 거리 이격된 위치에 배치되어 외부로부터 신호를 수신하는 제2 신호 수신부; 상기 제1 신호 수신부로부터 제공되는 신호를 사전에 설정된 주파수 대역별로 분리하여 각 채널로 출력하는 제1 필터 뱅크; 상기 제2 신호 수신부로부터 제공되는 신호를 사전에 설정된 주파수 대역별로 분리하여 각 채널로 출력하는 제2 필터 뱅크; 상기 제1 필터 뱅크의 각 채널에 연결되며, 상기 연결된 채널로부터 신호의 포락선 정보를 수신하고, 수신된 포락선 정보를 이용하여 반향 시간을 추정하는 반향 시간 추정부; 상기 반향 시간 추정부로부터 반향 시간을 제공받고, 반향 시간을 이용하여 반향이 없는 시작 구간을 검출하는 시작 구간 추정부; 상기 제1 및 제2 필터 뱅크의 각 채널에 연결되며, 상기 연결된 채널들로부터 수신된 신호들을 이용하여 영교차점을 기반으로 하는 ITD를 추정하는 ITD 추정부; 상기 채널들로부터 수신된 신호들에 대한 신호대 잡음비(SNR)를 검출하는 SNR 추정부; 상기 ITD 추정부에 의해 추정된 ITD들, 상기 시작 구간 추정부에 의해 검출된 시작 구간, 상기 SNR 추정부에 의해 검출된 각 신호에 대한 SNR들을 입력받고, 시작 구간 및 SNR들을 이용하여 상기 입력된 ITD들 중 신뢰성있는 ITD들을 선택하는 ITD 선택부; 상기 ITD 선택부에 의해 선택된 ITD들을 방향각으로 변환하는 방향각 변환부; 상기 방향각 변환부로부터 출력된 방향각들을 누적시켜 히스토그램을 생성하고, 상기 히스토그램을 이용하여 음원의 방향을 추출하는 음원 방향 추출부; 를 구비하고, 상기 반향 시

간 추정부, 시작 구간 추정부, ITD 추정부, SNR 추정부, ITD 선택부 및 방향각 변환부는 제1 필터 뱅크의 모든 채널에 각각 구비되고, 모든 방향각 변환부로부터 출력된 방향각 정보들은 음원 방향 추출부로 제공되어 반향 환경에서의 음원 방향을 추출한다.

[0010] 전술한 특징에 따른 다음원 국지화 장치에 있어서, 제1 및 제2 필터 뱅크는 감마톤 필터 뱅크로 구성되는 것이 바람직하다.

[0011] 전술한 특징에 따른 다음원 국지화 장치에 있어서, 상기 반향 시간 추정부는 반향 시간을 추정하며, 반향 시간(RT60)은 신호가 수신된 후 수신된 신호의 세기가 초기의 신호보다 60dB 감소되기 까지의 시간을 나타내며, 상기 반향 시간 추정부는 하기 수학적식을 반복적으로 학습하여 신호의 포락선의 감쇄 상수(α)의 최대 우도를 추정하며, 추정된 포락선의 감쇄 상수를 이용하여 반향 시간(RT60)을 계산하는 것이 바람직하다.

$$\Delta a \propto \frac{\partial \ln L(y(n))}{\partial a} = \frac{M}{a} \left[-\frac{M-1}{2} + \frac{\sum_{n=0}^{M-1} n a^{-2n} y^2(n)}{\sum_{n=0}^{M-1} a^{-2n} y^2(n)} \right]$$

[0012]

[0013] (여기서, α 는 포락선의 감쇄 상수, M 은 표본의 개수, $y(n)$ 은 필터 뱅크 특정 채널의 n 번째 출력 신호임).

[0014] 전술한 특징에 따른 다음원 국지화 장치에 있어서, 상기 시작 구간 추정부는 반향 시간을 이용하여 반향 신호를 추정하고, 추정된 반향 신호($E_i(n)$)와 원 입력 신호($env_i(n)$)가 하기의 수학적식을 만족하는 구간을 시작 구간으로 설정하는 것이 바람직하다.

$$\frac{env_i(n)}{E_i(n)} \geq \rho_{th}$$

[0015]

[0016] (여기서, ρ_{th} 는 원 입력 신호와 추정된 반향 신호의 비율의 한계값임).

[0017] 전술한 특징에 따른 다음원 국지화 장치에 있어서, 상기 방향각 변환부는 선택된 각 ITD들을 방향각으로 변환시키는 것이 바람직하며, 음원 방향 추출부는 상기 방향각 변환부로부터 출력된 방향각을 각 ITD에 대해 추정된 SNR 가중치로 누적시켜 히스토그램을 생성하고, 상기 히스토그램을 이용하여 음원의 방향을 추출하는 것이 바람직하다.

[0018] 본 발명의 다른 특징에 따른 다음원 국지화 방법은, 마이크로프로세서나 컴퓨터, 또는 신호처리장치 등에 의해 수행되며, 서로 일정 거리 이격된 위치에 배치된 2개의 신호 수신부들로부터 입력된 신호들을 주파수 대역별로 분리하여 채널별로 출력하는 2개의 필터 뱅크로부터 입력되는 신호들을 이용하여 음원의 방향을 추정하는 방법에 있어서, (a) 상기 필터 뱅크의 채널로부터 수신된 신호의 포락선 정보를 이용하여 반향 시간(RT60)을 추정하는 단계; (b) 상기 추정된 반향 시간을 이용하여 반향이 없는 시작 구간을 검출하는 단계; (c) 상기 필터 뱅크의 채널들로부터 수신된 신호들을 이용하여 영교차점을 기반으로 하는 ITD를 추정하는 단계; (d) 상기 필터 뱅크의 채널들로부터 수신된 신호들에 대한 신호 대 잡음비(SNR)를 검출하는 단계; (e) 상기 시작 구간 및 SNR들을 이용하여 상기 검출된 ITD들 중 신뢰성있는 ITD들을 선택하는 단계; (f) 상기 선택된 ITD들을 방향각으로 변환시키는 단계; (g) 상기 (a) 내지 (f) 단계는 필터 뱅크의 모든 채널에 대하여 반복적으로 수행하여, 모든 채널에 대하여 방향각들을 획득하는 단계; (h) 모든 채널에 대하여 획득한 방향각들을 누적시켜 히스토그램을 생성하고, 상기 히스토그램을 이용하여 음원의 방향을 추출하는 단계;를 구비하여, 반향 환경에서의 음원 방향을 추출한다.

[0019] 전술한 특징에 따른 다음원 국지화 방법에 있어서, 상기 (a) 단계의 반향 시간(RT60)은 신호가 수신된 후 수신

된 신호의 세기가 초기의 신호보다 60dB 감소되기 까지의 시간을 나타낸다.

[0020] 전술한 특징에 따른 다음원 국지화 방법에 있어서, 상기 (a) 단계는 하기 수학적식을 반복적으로 학습하여 신호의 포락선의 감쇄 상수(α)의 최대 우도를 추정하며, 추정된 포락선의 감쇄 상수를 이용하여 반향 시간(RT60)을 계산하는 것이 바람직하다.

$$\Delta a \propto \frac{\partial \ln L(y(n))}{\partial a} = \frac{M}{a} \left[-\frac{M-1}{2} + \frac{\sum_{n=0}^{M-1} n a^{-2n} y^2(n)}{\sum_{n=0}^{M-1} a^{-2n} y^2(n)} \right]$$

[0021]

[0022] (여기서, α 는 포락선의 감쇄 상수, M 은 표본의 개수, $y(n)$ 은 필터 뱅크 특정 채널의 n번째 출력 신호임).

[0023] 전술한 특징에 따른 다음원 국지화 방법에 있어서, 상기 (b) 단계는 반향 시간을 이용하여 반향 신호를 추정하고, 추정된 반향 신호($E_i(n)$)와 원 입력 신호($env_i(n)$)가 하기의 수학적식을 만족하는 구간을 시작 구간으로 설정하는 것이 바람직하다.

$$\frac{env_i(n)}{E_i(n)} \geq \rho_{th}$$

[0024]

[0025] (여기서, ρ_{th} 는 원 입력 신호와 추정된 반향 신호의 비율의 한계값임).

발명의 효과

[0026] 본 발명에 따른 다음원 국지화 방법 및 장치의 성능을 평가하기 위하여 실제 반향 환경에서 녹음한 음성을 이용하였다. 다양한 반향 환경을 고려하기 위하여 일반 사무실과 반향이 심한 공간, 즉 두 곳에서 녹음한 음성을 사용하였다. 또한 기존의 다른 방법과도 비교해 보았다. 도 6은 본 발명에 따른 다음원 국지화 방법의 성능을 평가하기 위하여 녹음이 진행된 공간에서의 마이크와 화자의 배치 상태를 나타낸 그림으로서, (a)는 일반 사무실을 나타내며 (b)는 반향이 큰 방을 나타낸 그림이다. 도 6과 같은 방에서 표본 주파수(sampling frequency) 16kHz로 두 개의 마이크를 이용하여 두 사람이 동시에 다른 각도에서 약 20초간 발생한다.

[0027] 두 방의 높이는 모두 3m 이고 마이크의 높이는 1.5m이다. 2개의 음원의 위치를 찾기 위한 실험을 위하여 2명의 화자가 두 개의 마이크 중앙으로부터 0.3m 떨어져서 각각 0° 와 -45° 에서 발생하게 된다. 반향의 정도는 마이크와 화자의 위치, 거리에 따라 달라지게 되는데, 거리가 멀어질수록 그리고 각도가 커져서 벽에 붙을수록 화자에서 마이크로 입력되는 신호의 반사되는 양이 많아져 반향이 심해진다. 또한 회의실이나 강당 같이 울림이 있는 곳은 반향의 정도가 더욱 심해지게 된다. 따라서 반향의 정도에 따라 기존 방법과 본 발명에 따른 방법의 성능의 차이를 보이기 위하여, 각도와 마이크로부터 화자의 거리는 고정된 채 반향이 적은 방과 반향이 심한 방 두 곳에서 음성을 녹음한 뒤 성능을 비교하여 보았다. 표 1은 본 발명에 따른 방법과 기존 방법의 성능을 평가할 때 사용된 변수 값들이다.

표 1

Parameter	α_{fe}	τ_{fe}	ρ_{th}	M
Value	0.5	8ms	3	800

[0028]

[0029]

도 7은 반향이 적은 일반 방에서 0° 그리고 -45° 위치에서 각각 화자가 발생했을 때의 음원 국지화 결과를 나타낸 음원 위치 히스토그램으로서, (a)는 본 발명에 따른 방법으로 수행한 결과이며, (b)는 종래의 에너지 기반의 음원 시작 구간 검출을 적용한 방법으로 수행한 결과이며, (c)는 종래의 반향이 없는 음원 시작 구간 검출을 적용한 방법으로 수행한 결과이다. 세 방법 모두 반향이 적은 곳에서는 어느 정도 정확한 결과를 나타내고 있다.

[0030]

도 8은 반향이 큰 환경에서 0° 그리고 -45° 위치에서 각각 화자가 발생했을 때의 음원 국지화 결과를 나타낸 음원 위치 히스토그램으로서, (a)는 본 발명에 따른 방법으로 수행한 결과이며, (b)는 종래의 에너지 기반의 음원 시작 구간 검출을 적용한 방법으로 수행한 결과이며, (c)는 종래의 반향이 없는 음원 시작 구간 검출을 적용한 방법으로 수행한 결과이다. 도 8을 통해 반향이 심한 환경에서는 음원 국지화가 좀 더 어려워졌음을 쉽게 확인할 수 있다. 도 8의 (b)의 결과는 -45°의 정보가 정확하지 않음을 확인할 수 있고 (c)도 역시 -45° 부근에서는 히스토그램이 뭉개져서 정확한 피크를 찾기 힘들다. 하지만, 본 발명에 따른 (a)는 두 개의 방향 정보의 침투값이 정확히 나타나고 있음을 확인할 수 있다.

[0031]

마지막으로 본 발명이 좀 더 다양한 환경에서도 적용 가능한지 확인하기 위하여 여러 환경에 대한 데이터를 실험적으로 생성하여 실험하였다. 실험에서 마이크로 입력되는 반향신호들은 음원 신호와 음원에서 마이크까지의 음향 반향에 대한 채널 필터의 컨벌루션 결과의 혼합으로 생성하였다. 반향 시간(RT60)이 0.1초부터 1.0초까지 10가지의 경우에 대해 TIMIT 데이터베이스에서 16명의 화자로부터 발생된 각 10개의 문장을 사용하였고 이 때 화자는 한문장을 2번씩 발생하게 된다. 또한 마이크 음원의 방향은 0°, 30°, 60°로 바뀌가며 실험하였다. 또한 이렇게 생성된 데이터를 좀 더 실제 환경에 가깝게 하기 위하여 생성된 데이터에 신호 대 잡음비 30dB로 확산 잡음을 더해 주었다. 도 9는 생성된 총 9600개의 음원에 대해 음원 국지화의 성공률을 허용오차 3°로 하여 나타낸 결과를 나타낸 그래프이다. 도 9를 통해 제안한 방법을 사용했을 경우 음원 국지화의 성공률이 종래의 방법들에 비해 향상됨을 확인할 수 있다.

[0032]

전술한 바와 같이, 본 발명에 따른 방법은 실제 환경에서 녹음한 자료로 실험한 결과 종래의 방법에 비해 매우 효과적으로 음원의 방향을 찾게 됨을 확인할 수 있다. 또한 다양한 반향 환경과 화자의 위치에서도, 본 발명에 따른 방법의 성능 향상을 보기 위하여 여러 조건에 따라 실험을 반복한 결과 본 발명에 따른 방법이 더 좋은 성능을 가짐을 확인할 수 있다.

도면의 간단한 설명

[0033]

도 1은 본 발명의 바람직한 실시예에 따른 다음원 국지화 장치를 전체적으로 도시한 블록도이다.

도 2는 입력 신호와 추정된 반향 신호와의 관계를 도시한 그래프이다. 도 2의 (a)는 입력 신호들을 표시한 것이며, 도 2의 (b)는 입력 신호들에 대한 반향 신호들을 추정한 신호를 표시한 것이다.

도 3은 수학적 10에 따라 이전의 모든 시간에 의한 전체의 반향 신호($E_i(n)$)를 추정하는 피드백 알고리즘을 도시한 것이다.

도 4는 본 발명의 바람직한 실시예에 따른 다음원 국지화 장치에 있어서, 방향 시간 추정부 및 시작 구간 검출부에 의해 음원 신호의 시작 구간 검출 과정을 순차적으로 도시한 그래프이다.

도 5는 필터뱅크 출력 신호의 포락선에서의 영교차점의 검출에 기반하여 두 신호의 시간차를 도시한 그래프이다.

도 6은 본 발명에 따른 다음원 국지화 방법의 성능을 평가하기 위하여 녹음이 진행된 공간에서의 마이크와 화자

의 배치 상태를 나타낸 그림으로서, (a)는 일반 사무실을 나타내며 (b)는 반향이 큰 방을 나타낸 그림이다.

도 7은 반향이 적은 일반 방에서 0° 그리고 -45° 위치에서 각각 화자가 발생했을 때의 음원 국지화 결과를 나타낸 음원 위치 히스토그램으로서, (a)는 본 발명에 따른 방법으로 수행한 결과이며, (b)는 종래의 에너지 기반의 음원 시작 구간 검출을 적용한 방법으로 수행한 결과이며, (c)는 종래의 반향이 없는 음원 시작 구간 검출을 적용한 방법으로 수행한 결과이다.

도 8은 반향이 큰 환경에서 0° 그리고 -45° 위치에서 각각 화자가 발생했을 때의 음원 국지화 결과를 나타낸 음원 위치 히스토그램으로서, (a)는 본 발명에 따른 방법으로 수행한 결과이며, (b)는 종래의 에너지 기반의 음원 시작 구간 검출을 적용한 방법으로 수행한 결과이며, (c)는 종래의 반향이 없는 음원 시작 구간 검출을 적용한 방법으로 수행한 결과이다.

도 9는 생성된 총 9600개의 음원에 대해 음원 국지화의 성공률을 허용오차 3° 로 하여 나타낸 결과를 나타낸 그래프이다.

발명을 실시하기 위한 구체적인 내용

- [0034] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예에 따른 다음원 국지화 장치 및 방법에 대하여 구체적으로 설명한다.
- [0035] 도 1은 본 발명의 바람직한 실시예에 따른 다음원 국지화 장치를 전체적으로 도시한 블록도이다. 도 1을 참조하면, 본 실시예에 따른 다음원 국지화 장치(1)는 외부의 음원으로부터 발생된 신호를 수신하는 제1 및 제2 신호 수신부(100, 200), 감마톤 필터 뱅크로 구성되는 제1 및 제2 필터뱅크(102, 202), 제1 필터 뱅크의 각 채널로부터 출력되는 신호의 포락선(envelope)을 이용하여 반향시간을 추정하는 반향시간 추정부(110), 추정된 반향시간을 이용하여 반향이 없는 시작구간을 검출하는 시작구간 검출부(120), 제1 및 제2 필터 뱅크의 채널로부터 출력되는 신호의 파형을 이용하여 영교차점을 기반으로 한 ITD를 추정하는 ITD 추정부(130), SNR을 추정하는 SNR 추정부(140), 추정된 시작 구간 및 SNR을 이용하여 신뢰성있는 ITD를 선택하는 ITD 선택부(150), 선택된 ITD를 방향각으로 변환하는 방향각 변환부(160), 각 채널별로 얻어진 방향각을 각 ITD에 대해 추정된 SNR 가중치로 가중시킨 후 누적하여 히스토그램을 생성하고 히스토그램을 이용하여 음원의 방향을 추출하는 음원 방향 추출부(170)를 구비한다. 본 발명에 따른 다음원 국지화 장치(1)의 반향 시간 추정부, 시작 구간 검출부, ITD 추정부, SNR 추정부, ITD 선택부 및 방향각 변환부는 각 주파수 채널별로 구비된다.
- [0036] 이하, 전술한 각 구성요소들의 구조 및 동작에 대하여 구체적으로 설명한다.
- [0037] 제1 및 제2 신호 수신부(100, 200)는 마이크와 같이 외부로부터 입력되는 신호를 검출하는 센서들로 구성되며, 외부로부터 입력되는 신호를 수신하여 제1 및 제2 필터뱅크로 제공한다.
- [0038] 제1 및 제2 필터뱅크(102, 202)는 인간의 달팽이관과 같은 주파수 특성을 갖는 감마톤 필터뱅크(Gammatone Filterbank)로 구성되며, 감마톤 필터 뱅크는 중심 주파수가 100Hz부터 4kHz까지 ERB(Equivalent Rectangular Bandwidth) 스케일에서 선형적으로 증가하는 25개의 채널로 구성된다. 제1 및 제2 필터뱅크(102, 202)는 제1 및 제2 신호 수신부와 각각 연결되어, 제1 및 제2 신호 수신부로부터 각각 입력되는 신호들을 사전에 설정된 각 주파수 대역별로 분리하여 해당 채널로 출력한다. 제1 필터뱅크는 각 반향시간추정부로 해당 주파수 대역의 신호에 대한 포락선 신호를 출력하며 각 ITD 검출부로 해당 주파수 대역의 신호를 출력한다. 제2 필터뱅크는 각 ITD 검출부로 해당 주파수 대역의 신호를 출력한다.
- [0039] 상기 반향시간추정부(110)는 RT60으로 표현되는 반향 시간을 추정하여 시작구간 검출부로 제공한다. 이하, 본 발명에 따른 반향시간 추정부의 반향시간 추정과정을 구체적으로 설명한다.
- [0040] 반향 시간(reverberation time)은 청공간(auditory space)의 특성을 나타내는 중요한 척도이다. 방안의 반향(room reverberation)의 정도는 음성의 인지나 음원의 국지화에 많은 영향을 끼친다. 반향이 커질수록 음성의 인식률이 감소하게 된다. 따라서 보청기(hearing aids)나 핸드프리(hands-free) 기기 등 다양한 기기에 반향 추정이 적용되면 주변 환경에 가장 적합한 신호 처리 방법을 선택할 수 있도록 하여 청자가 좀 더 쉽게 상대방의 음성을 인지할 수 있게 된다. 반향의 정도를 나타내기 위해 가장 많이 사용되는 척도는 반향 시간(reverberation time, RT60)으로서, RT60은 소리가 난 뒤 그 소리의 정도가 초반보다 60dB 감소되기까지의 시간이다. 이를 수학적으로 모델링하면 시정수(τ)로 지수적으로 감소하는 신호의 모양으로 가정할 수 있다. 반향 시간과 시정수 사이의 관계는 $RT60 = 6.91\tau$ 로 나타낼 수 있다. 이와 같이 반향의 정도인 반향 시간을 추정하여 여러 기기들에 적용하기 위해서는 주위 환경인 방의 구조 또는 입력 신호들의 특성에 대한 사전 정보 없이 마이

크로 입력된 신호만을 가지고 반향 시간을 추정해 내야 한다. 따라서, 마이크에 입력된 신호로부터 감쇄 시정수를 우도의 최대화(maximum-likelihood) 과정을 통해 추적해가는 방식을 통해 반향 시간을 추정해낸다.

[0041] 일반적인 방의 임펄스(room impulse) 응답은 직선 경로 성분을 포함하는 부분, 직선 경로 성분을 포함하는 부분에 뒤이어 나오는 초기 지연 반사 성분들, 및 지수적으로 감소하는 반향 꼬리(reverberant tail) 부분으로 나눌 수 있다. 정확한 음원의 위치를 알기 위해서 필요했던 부분이 채널 필터 앞부분의 직선 경로를 통과한 부분이었다면, 반향의 정도를 나타내는 반향 시간을 추정하기 위해서는 마지막 반향 꼬리 부분이 필요하다. 직선 경로 부분과 초기 지연 반사 부분과는 달리 마지막의 반향 꼬리 부분은 포락선이 감소하는 형태의 캐리어 신호를 포함하고 있다. 따라서, 반향 꼬리 부분은 시정수(τ)에 의해 포락선이 지수적으로 감소하는 비상관 확률 과정(uncorrelated random process) 모형으로 간주한다.

[0042] 따라서 반향의 감쇄 모델은 정규 분포 $N(0, \sigma)$ 인 독립적이고 동일한 분포(independent and identically distributed)를 따는 확률 변수 $x(n)$, $n > 0$ 과 시정수(τ_d) 를 가지는 포락선의 감쇄상수 $a = \exp(-1/\tau_d)$ 의 곱으로 나타낼 수 있다. 그러므로 반향 환경에서 관찰되는 신호를 수학적 1과 같이 나타낼 수 있다.

수학적 1

[0043]
$$y(n) = a^n x(n)$$

[0044] 이 신호의 확률 분포 함수(probability density function) 는 $N(0, a^n \sigma)$ 로 나타낼 수 있으며 감쇄 시정수(τ_d)를 추정하기 위하여 수학적 1에 대한 우도(likelihood) 함수를 구한다. 이 우도함수의 변수는 a 와 σ 로 주어진다. 또한 추정 연산을 할 때 M개의 표본 단위로 연산이 수행된다.

[0045] 우도 함수가 주어졌을 경우 변수를 추정하기 위해서 우도 함수 최대화(maximum likelihood) 과정이 수행된다. 그러기 위해서는 우도 함수에 로그를 취해 주어야 한다. 로그 우도 함수는 수학적 2와 같이 나타난다.

수학적 2

[0046]
$$L(y(n)) = -\frac{M(M-1)}{2} \ln(a) - \frac{M}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{M-1} a^{-2n} y^2(n).$$

[0047] 수학적 2 의 편미분은 수학적 3 및 4와 같이 나타난다.

수학식 3

$$\frac{\partial \ln L(y(n))}{\partial a} = -\frac{M(M-1)}{2a} + \frac{1}{a\sigma^2} \sum_{n=0}^{M-1} na^{-2n} y^2(n)$$

[0048]

수학식 4

$$\frac{\partial \ln L(y(n))}{\partial \sigma} = -\frac{M}{\sigma} + \frac{1}{\sigma^3} \sum_{n=0}^{M-1} a^{-2n} y^2(n)$$

[0049]

[0050] 수학식 3 및 4 의 편미분을 0으로 하여 a 와 σ 의 최대 우도 값을 추정해 낼 수 있으며, 이는 수학식 5 및 6과 같이 표현된다.

수학식 5

$$0 = -\frac{M(M-1)}{2a} + \frac{1}{a\sigma^2} \sum_{n=0}^{M-1} na^{-2n} y^2(n),$$

[0051]

수학식 6

$$\sigma^2 = \frac{1}{M} \sum_{n=0}^{M-1} a^{-2n} y^2(n)$$

[0052]

[0053] 수학식 6의 a 는 최대 경사법을 통하여 구할 수 있다. a 를 추정하기 위하여 수학식 6을 수학식 3에 대입하여 정리하면 수학식 7을 얻을 수 있다.

수학식 7

$$\Delta a \propto \frac{\partial \ln L(y(n))}{\partial a} = \frac{M}{a} \left[-\frac{M-1}{2} + \frac{\sum_{n=0}^{M-1} na^{-2n} y^2(n)}{\sum_{n=0}^{M-1} a^{-2n} y^2(n)} \right]$$

[0054]

[0055] 수학식 7을 반복적으로 학습하여 α 의 최대 우도를 추정한다. 포락선의 감쇄 상수 $\alpha = \exp(-1/\tau_d)$

이므로, 전술한 과정을 통해 추정된 α 를 이용하여 반향 시간을 추정하기 위한 감쇄 시정수(τ_d)를 구한다.

[0056] 반향 시간(RT60)은 6.91τ 이므로, 전술한 과정에서 구한 감쇄 시정수를 이용하여 반향 시간을 추정한다. 본 발명에 따른 다음원 국지화 장치는 전술한 반향 시간 추정부를 각 채널별로 구비하여, 각 채널별로 반향 환경에 맞는 반향 신호를 추정하게 된다.

[0057] 시작구간 검출부(120)는 반향시간 추정부로부터 반향시간을 입력받고, 이를 이용하여 반향이 없는 시작 구간을 검출한다. 이하, 본 발명에 따른 시작구간 검출부가 반향이 없는 시작 구간을 검출하는 과정을 구체적으로 설명한다.

[0058] 어떠한 소리가 반향 환경에 존재할 때 마이크에 도달하는 신호들은 직선 경로를 통과한 성분들과 그 뒤를 따르는 지연 반사 성분을 통과한 성분들로 이루어진다. 지연 반사 성분들은 음원의 위치와 주변 환경에서의 반향 정도 등에 따라 다르게 나타나게 된다. 즉 음원의 위치와 음원과 센서 사이의 거리 등의 조합에 따라 다양한 특성을 나타낸다. 하지만, 인간이 실세계에서 접하는 음향 신호는 보통 대략적으로 시간에 따라 급격하게(exponential) 크기가 감소하는 특성을 보인다. 두 마이크에 도달하는 신호의 차이를 이용하여 음원의 위치를 추정할 때, 반향이 없는 음원의 시작 구간을 제외하고는 정확한 음원의 위치를 추정해 낼 수 없다. 모든 지연 반사 성분들은 직선 경로를 통한 성분들보다 센서에 도달하기까지의 시간 지연이 있으며 크기도 감쇄하게 된다.

이러한 특성을 바탕으로 우리는 반향에 관한 전형적인 임펄스 패턴($h_p(n)$)을 수학식 8과 같이 세울 수 있다.

수학식 8

$$h_p(n) = \begin{cases} 0, & 0 \leq n \leq \tau_{fe} \\ \alpha_{fe} e^{-\frac{n-\tau_{fe}}{\tau_d}}, & n \geq \tau_{fe} \end{cases}$$

[0059]

[0060] α_{fe} 와 τ_{fe} 는 첫 반향(echo)의 감쇄와 지연 시간을 나타내며 τ_d 는 반향의 감쇄 정도를 나타내는 시

정수이다. 필터뱅크를 통과한 출력 신호 중 i 번째 채널신호의 포락선을 $env_i(n)$ 로 나타내며 시간 n_p 에서의 값에 의해 나타낼 수 있는 반향을 수학식 9와 같이 나타낼 수 있다.

수학식 9

$$e_i(n, n_p) = env_i(n_p) h_p(n - n_p)$$

[0061]

[0062] 도 2는 입력 신호와 추정된 반향 신호와의 관계를 도시한 그래프로서, (a)는 입력 신호들을 표시한 것이며, 도 2의 (b)는 입력 신호들에 대하여 추정된 반향 신호들을 표시한 것이다. 도 2의 (b)에 도시된 바와 같이, 각 입력 신호에 대한 반향 신호들 중 가장 큰 세기를 갖는 신호가 실제 반향 신호가 된다. 따라서, 이전의 모든 시간에 의한 전체의 반향 신호는 수학식 10과 같이 구해질 수 있다.

수학식 10

$$E_i(n) = \text{Max}\{e_i(n, n_p), 0 \leq n_p < \infty\}$$

[0063]

[0064] 여기서, Max 함수는 입력되는 값의 최대값을 취하는 것을 의미한다. 반향의 추정에서 반향 신호의 감쇄는 지수함수 형태로 나타나게 되므로 포락선의 감쇄 상수(α)는 수학식 11과 같이 상수로 나타낼 수 있다.

수학식 11

$$\alpha = \frac{E_i(n+1)}{E_i(n)} = e^{-\frac{1}{\tau_d}}$$

[0065]

[0066] 한편, 포락선의 감쇄 상수(α)는 반향 시간 추정부에 의해 추정된 포락선의 감쇄 상수와 동일한 값을 의미한다.

[0067] 도 3은 수학식 9와 10에 따라 이전의 모든 시간에 의한 전체의 반향 신호($E_i(n)$)를 추정하는 피드백 알고리즘을 도시한 것이다.

[0068] 반향 신호가 추정되면, 추정된 반향 신호($E_i(n)$)를 원 입력 신호($env_i(n)$)와 비교하여 상대적인 반향의 크기를 찾고 이를 통해 반향이 없는 음원의 시작 구간을 수학식 12와 같이 검출하게 된다.

수학식 12

$$\frac{env_i(n)}{E_i(n)} \geq \rho_{th}$$

[0069]

[0070] 여기서, ρ_{th} 는 원 입력 신호와 추정된 반향 신호의 비율의 한계값이다. 즉, 음원의 시작 구간은 원 입력 신호와 추정된 반향 신호의 비율이 사전에 설정된 한계값(ρ_{th})보다 크거나 같은 구간으로 추정된다.

[0071] 반향을 추정하여 반향이 존재하지 않는 음원의 시작 구간을 검출하기까지의 전체적인 실험 결과들을 도 4에 도시하였다. 도 4는 본 발명의 바람직한 실시예에 따른 다음원 국지화 장치에 있어서, 시작 구간 검출부에 의해 반향이 없는 음원의 시작 구간 검출 과정의 결과들을 순차적으로 도시한 그래프이다.

[0072] 도 4의 첫번째 그림이 필터뱅크를 통과한 한 채널 신호의 포락선을 나타낸 것이며, 두 번째 그림은 수학식 10에 의해 추정된 반향 신호를 나타낸 것이며, 세 번째 그림이 첫 번째 그림의 신호와 두 번째 신호의 비를 나타낸 그림이며, 마지막 그림은 두 신호의 비를 통해 반향이 존재하지 않는 음원의 시작 부분을 검출해 낸 결과를 나타낸 것이다. 입력 신호와 추정한 반향 신호와의 비율을 통하여 반향이 없는 음원의 시작 구간을 검출하는 방법을 이용한 음원 위치 국지화 방법은 도 3에서와 같이 피드백 알고리즘을 이용하여 적은 계산량으로 구현할 수

있다.

[0073] 본 발명에 따른 시작 구간 검출부는 전술한 반향 시간 추정부에 의해 추정된 반향 감쇄 시정수를 사용하여 반향이 없는 음원의 시작구간(Onset)을 검출함으로써, 다양한 반향 환경에 대처할 수 있도록 한다. 하지만, 종래의 방법들은 반향을 추정할 경우 정형화된 임펄스 응답 모형을 사용하였으며, 이때 반향의 정도를 나타내는 감쇄 시정수는 모든 환경에 대해 고정되어져 있기 때문에, 반향의 정도가 커지게 되거나 주변 환경이 매우 달라질 경우 성능이 저하되는 문제점을 안고 있었다.

[0074] ITD 추정부(130)는 제1 및 제2 필터뱅크의 동일한 주파수 채널들로부터 출력되는 신호들의 영교차점을 검출하고, 이에 기반하여 두 신호의 시간차를 계산하여 출력한다.

[0075] 영교차점은 동일한 주파수 채널에 대하여 제1 및 제2 필터뱅크를 통과한 출력신호들로부터 도 5와 같이 검출된다. 도 5는 필터뱅크 출력 신호의 영교차점의 검출에 기반하여 두 신호의 시간차를 도시한 그래프이다. 도 5에 도시된 바와 같이, 제1 및 제2 필터뱅크를 통과한 신호들의 정확한 영교차점은 영점을 지나기 직전과 직후의 샘플값을 선형적으로 이어서 영점과 교차되는 시간을 찾는 것이다. 영교차점에 기반한 신호의 시간차(ITD)를 계산하기 위해서 제1 신호 수신부의 입력신호의 영교차점을 제2 신호 수신부의 입력 신호의 영교차점들과 비교하여, 가장 근접한 영교차점 사이의 시간차(ITD)를 계산한다. 이 시간차(ITD)는 음향 신호가 제1 및 제2 신호 수신부에 도달하는데 걸리는 최대 시간차보다 작으면 유효하다고 판단하며, 최대 시간차보다 큰 시간차를 가지는 모든 정보는 무시한다. 위에 언급한 바와 같이 제1 및 제2 신호 수신부의 이격 거리를 신호처리가 필요한 최대 주파수까지 공간 엘리머싱을 피할 수 있을 만큼 가깝게 설정한다. 이 때 신호의 시간차의 최대값은 입력 신호의 최대 주파수의 반주기 이하가 되므로 제일 작은 영교차점의 시간차와 음원의 방향에 상응하는 신호의 시간차(ITD)값이 된다.

[0076] SNR 추정부(140)는 각 신호에 대한 신호 대 잡음비(SNR)를 추정하여 출력한다. 이하, SNR 추정부(140)가 각 신호에 대한 SNR을 추정하는 과정을 설명한다.

[0077] 특정 음원이 마이크에 도달했을 경우의 시간차는 거의 일정한 값을 나타내나, 음원이 없는 구간 즉 확산 잡음만 존재하는 구간에서의 시간차는 제각각으로 나타난다. 따라서 일정 구간의 신호 간 시간차들에 대한 분산을 이용하여 해당 신호에서 신호 대 잡음비(SNR)를 추정하여 정확한 음원의 방향을 나타내는 신호의 시간차만을 선택한다. SNR은 수학식 13에 의하여 구할 수 있다.

수학식 13

$$\widehat{SNR}_i(j) = \begin{cases} 10 \log_{10} \frac{1}{w_i^2 S_i^2(j)}, & \text{if } w_i^2 S_i^2(j) < 1, \\ 0, & \text{otherwise.} \end{cases}$$

[0078]

[0079] w_i 는 감마톤 필터뱅크의 i 번째 채널의 중심 주파수를 나타내고, $S_i^2(j)$ 는 j번째 신호의 시간차를 기준으로 한 분산값을 나타낸다. 큰 분산값을 갖는 ITD는 신호 대 잡음비(SNR)가 낮기 때문에, 확산잡음으로부터 영향을 많이 받은 것으로 볼 수 있다. 따라서, 신호 대 잡음비가 크게 추정된 신호의 시간차만을 주로 음원방향 추정에 사용하게 된다.

[0080] ITD 선택부(150)는 시작 구간 및 SNR을 이용하여 ITD 추정부에 의해 추정된 ITD들 중에서 신뢰성있는 ITD들을 선택하여 출력한다. ITD 선택부(150)는 시작 구간 추정부로부터 반향이 없는 시작 구간이 입력되고, ITD 추정부로부터 ITD 값들이 입력되며, SNR 추정부로부터 SNR이 입력된다. ITD선택부(150)는 ITD 추정부로부터 입력된 ITD값들중에서 반향이 없는 시작 구간의 신호의 ITD 및 SNR이 크게 추정된 신호의 ITD만을 선택하여 신뢰성있는 ITD만을 출력한다.

[0081] 방향각 변환부(160)는 ITD 선택부에 의해 선택된 신뢰성있는 ITD들을 음원의 방향에 해당하는 방향각으로 변환시킨다.

[0082] 음원 방향 추출부(170)는 각 채널별로 마련된 방향각 변환부들로부터 출력된 방향각을 이용하여 히스토그램을

생성하고, 상기 생성된 히스토그램을 이용하여 음원의 방향을 추출한다. 이 때, 신호 대 잡음비가 크게 추정된 신호의 시간차에 대한 반영률을 상대적으로 높이기 위하여, 각 시간차에 상응하는 방향각에 추정된 신호 대 잡음비를 가중치로하여 히스토그램을 생성한다. 이렇게 구성된 히스토그램으로부터 얻은 첨두(peak) 값들이 음원의 방향에 해당한다.

산업상 이용가능성

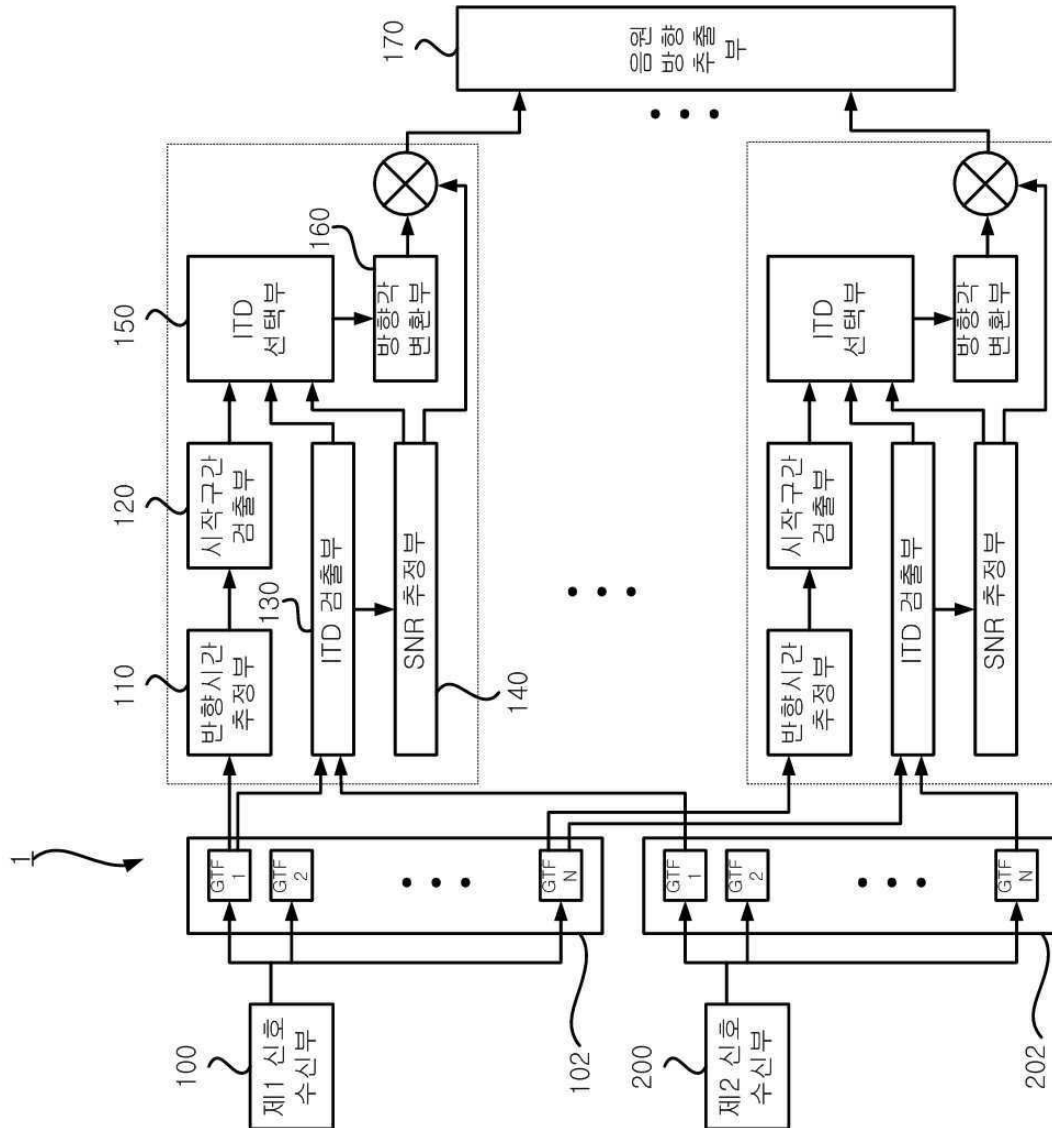
[0083] 본 발명에 따른 다음원 국지화 방법 및 장치는 여러 소리들과 잡음이 혼재하는 환경에서 사용되는 핸즈프리 시스템이나 음성 회의 시스템에 있어서, 음원 위치 국지화하는 방법 등에 사용될 수 있다.

부호의 설명

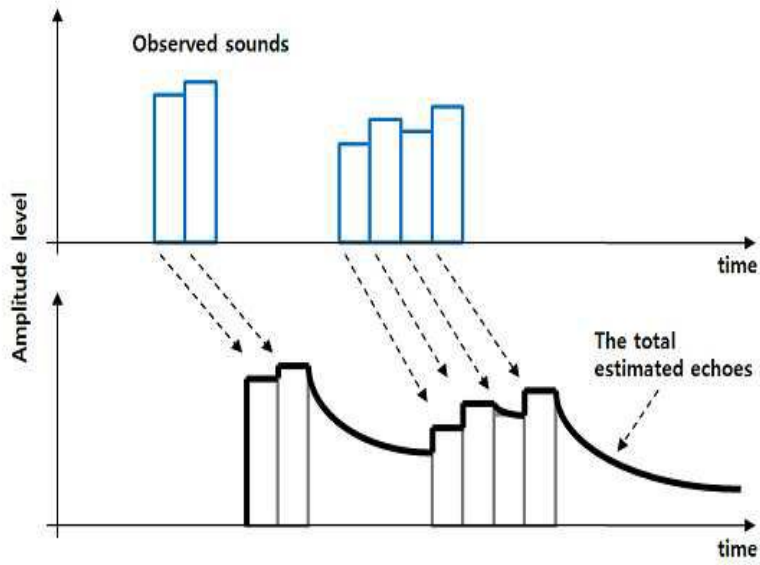
[0084] 1 : 다음원 국지화 장치
 100 : 제1 신호 수신부
 200 : 제2 신호 수신부
 102 : 제1 필터뱅크
 202 : 제2 필터뱅크
 110 : 반향시간 추정부
 120 : 시작구간 검출부
 130 : ITD 추정부
 140 : SNR 추정부
 150 : ITD 선택부
 160 : 방향각 변환부
 170 : 음원 방향 추출부

도면

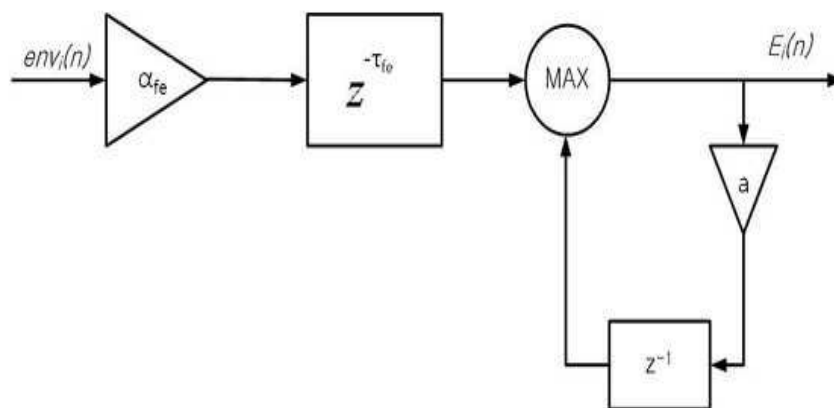
도면1



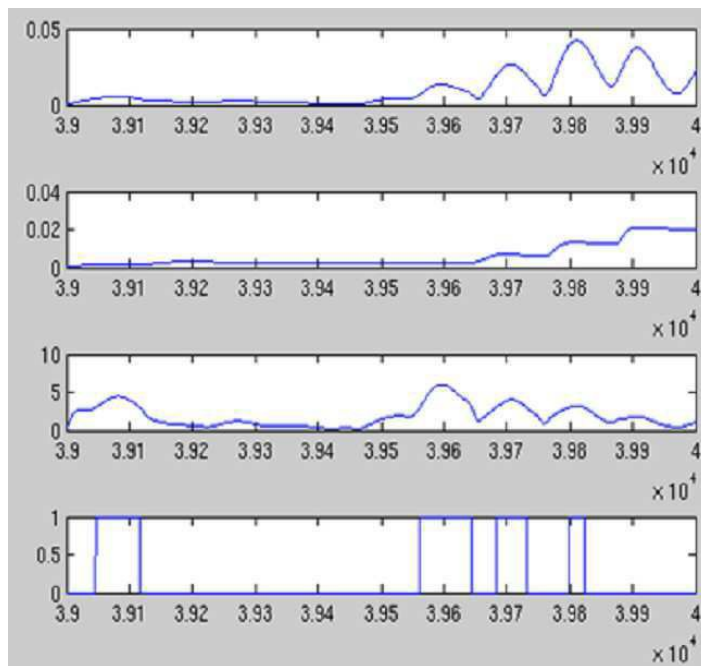
도면2



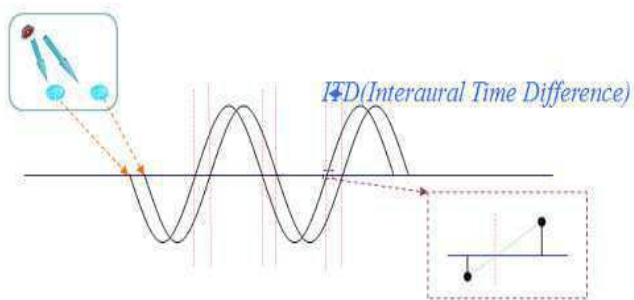
도면3



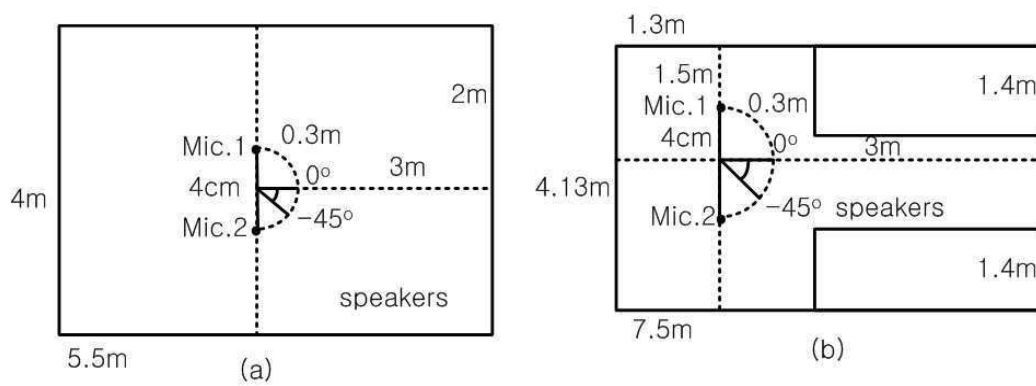
도면4



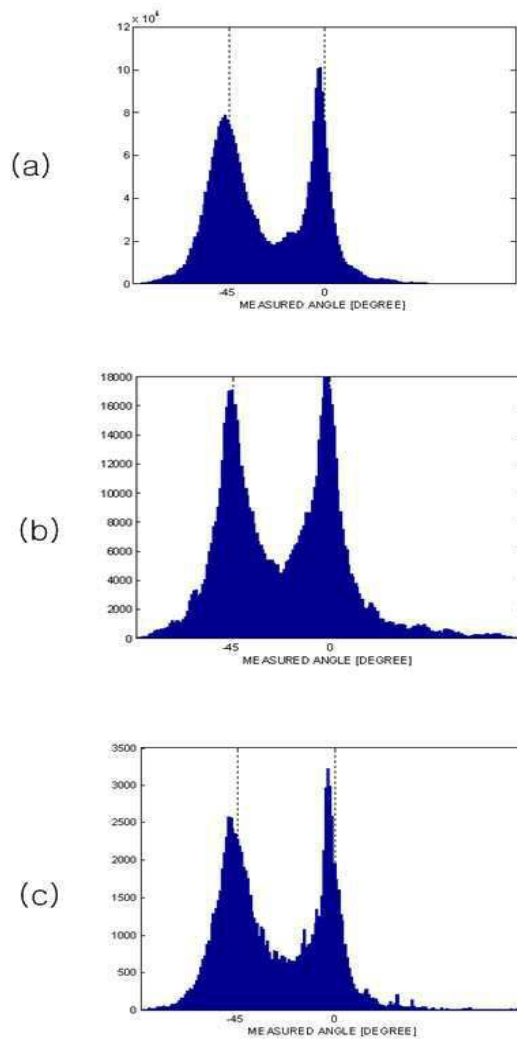
도면5



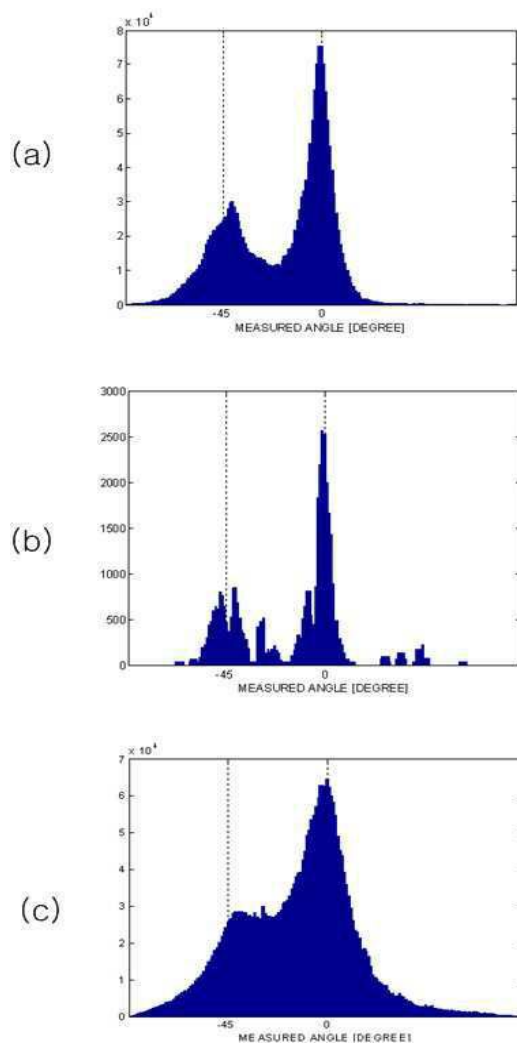
도면6



도면7



도면8



도면9

