(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2009-0127709
(43) 공개일자 2009년12월14일

(51) Int. Cl.
G10L 21/02 (2006.01) *G10L 15/20* (2006.01)
(21) 출원번호 10-2008-0053810
(22) 출원일자 2008년06월09일
심사청구일자 없음

(71) 출원인
삼성전자주식회사
경기도 수원시 영통구 매탄동 416
고려대학교 산학협력단
서울 성북구 안암동5가1 고려대학교 내
(72) 발명자
김현수
경기도 용인시 수지구 상현동 만현마을 두산위브 아파트 803동 501호
고한석
서울특별시 용산구 서빙고동 신동아아파트 16-703
(뒷면에 계속)
(74) 대리인
박상수

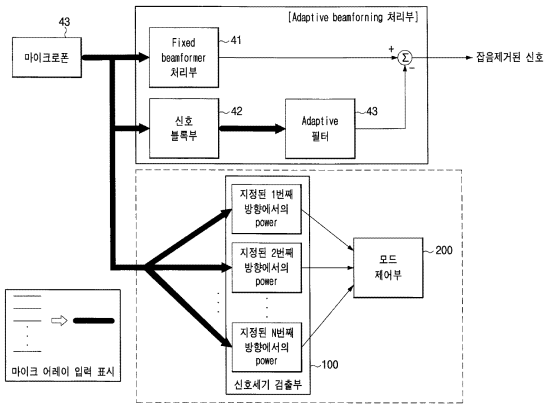
전체 청구항 수 : 총 4 항

(54) 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응모드 제어 장치 및 방법

(57) 요 약

본 발명에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치는 Adaptive beamformer 처리부로 제공하는 적어도 하나 이상의 마이크로폰을 통해 입력된 어레이 입력 신호를 입력받으면, 각각의 지정된 방향의 신호 세기를 검색하여 최대값을 갖는 신호 세기를 검출하는 신호 세기 검출부; 및 상기 신호 세기 검출부를 통해 검출된 최대값을 갖는 신호 세기와 임계값을 비교하여 최대값을 갖는 신호 세기가 임계값을 초과할 경우 GSC의 adaptive 모드를 수행하지 않도록 제어하는 adaptive 모드 제어부를 포함하는 것으로 적응 빔포밍이 성능을 보장하기 위한 조건 중에 하나이나, 자기 상관도가 높은 음원의 잡음을 제거할 경우 적응 필터의 적응을 적용하지 않아야 하나 이를 제어할 수 없는 종래의 문제점을 해결하는 이점이 있다.

대 표 도



(72) 발명자

배정훈

서울특별시 중구 신당6동 58-44

이택진

서울특별시 관악구 신림4동 472-33

특허청구의 범위

청구항 1

적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치에 있어서,

Adaptive beamformer 처리부로 제공하는 적어도 하나 이상의 마이크로폰을 통해 입력된 어레이 입력 신호를 입력받으면, 각각의 지정된 방향의 신호 세기를 검색하여 최대값을 갖는 신호 세기를 검출하는 신호 세기 검출부; 및

상기 신호 세기 검출부를 통해 검출된 최대값을 갖는 신호 세기와 임계값을 비교하여 최대값을 갖는 신호 세기가 임계값을 초과할 경우 GSC의 adaptive 모드를 수행하지 않도록 제어하는 adaptive 모드 제어부를 포함하는 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치.

청구항 2

제 1항에 있어서,

상기 신호 세기 검출부는,

어레이의 각 마이크로폰별로 입력된 잡음 섞인 음성을 임의의 길이의 hanning window를 씌어 프레임으로 구분하는 윈도우 처리부;

주파수 분석을 위해 각 마이크로폰 및 구분된 프레임 별로 DFT를 수행하는 DFT 처리부;

Array를 구성하는 각 마이크 별로 pair를 이루어 탐지하려고 하는 방향으로 빔(beam)을 steering한 후 Cross-power spectrum 추정하는 상관관계 계산부;

Cross-power spectrum을 정규화 시켜주기 위한 phase-transform weight를 구하는 weight 추정부; 및

방향이 검색되면 해당 방향으로부터 들어오는 소리의 세기를 측정하는 신호 세기 측정부를 포함하는 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치.

청구항 3

적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 방법에 있어서,

고정 beamformer와 신호 블록부로 제공하는 적어도 하나 이상의 마이크로폰을 통해 입력된 어레이 입력 신호를 입력받으면, 각각의 지정된 방향의 신호 세기를 검색하여 최대값을 갖는 신호 세기를 검출하는 단계;

상기 신호 세기 단계에서 검출된 최대값을 갖는 신호 세기와 임계값을 비교하는 단계; 및

상기 판단 단계에서 최대값을 갖는 신호 세기가 임계값을 초과하면, GSC의 adaptive 모드를 수행하지 않도록 제어하는 단계를 포함하는 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 방법.

청구항 4

제 3항에 있어서,

상기 최대값을 갖는 신호 세기를 검출하는 단계는,

어레이의 각 마이크로폰별로 입력된 잡음 섞인 음성을 임의의 길이의 hanning window를 씌어 프레임으로 구분하는 단계;

주파수 분석을 위해 각 마이크로폰 및 구분된 프레임 별로 DFT를 수행하는 단계;

Array를 구성하는 각 마이크 별로 pair를 이루어 탐지하려고 하는 방향으로 빔(beam)을 steering한 후 Cross-power spectrum 추정하는 단계;

Cross-power spectrum을 정규화 시켜주기 위한 phase-transform weight를 구하는 단계; 및

방향이 검색되면 해당 방향으로부터 들어오는 소리의 세기를 측정하는 단계를 포함하는 적응 빔포밍을 위한 사

용자 방향의 소리 검출 기반의 적응 모드 제어 방법.

명 세 서

발명의 상세한 설명

기 술 분 야

<1> 본 발명은 잡음 제거를 위한 적응 빔포밍(adaptive beamforming)에서 핵심 분야인 adaptive 모드 제어에 관한 것이다.

<2> 적응 빔포밍은 사용자의 음성이 나오는 방향에 acoustic beam을 쏘아서 음성 이외의 소리를 억압하는 기술이다.

<3> 적응 빔포밍이 성능을 보장하기 위한 조건 중에 하나가 사용자의 발성이 있을 때에는 적응 빔포밍에 쓰이는 적응 필터(adaptive filter)의 적응(adaption)을 멈추어야 한다는 것이며, 이것을 결정하는 것이 adaptive 모드 제어이다.

배 경 기 술

<4> 종래의 기술들은 어레이의 각각의 마이크로 입력되는 신호의 상관관계(correlation)를 이용하는 제 1 방법과, 목적(target) 신호와 참조(reference) 신호와의 에너지의 비를 이용하는 제 2 방법이 있다.

<5> 종래 마이크로폰 어레이를 이용한 잡음 제거 시스템은 적어도 하나 이상의 마이크로폰, 마이크로폰에 각각 연결된 단구간 분석부, 에코 제거부, 방향성 잡음 제거와, 정면소리 유/무를 기반으로 필터 weight update on/off하는 adaptive beamforming 처리부, 마이크간의 신호의 상관관계를 이용하여 정면소리를 정면소리 검출부, 잔여잡음 제거와 정면소리 유/무를 기반으로 잔여 잡음을 제어하는 Post-filtering부 및 Overlap and add 처리부를 포함한다.

<6> 이러한, 종래 마이크로폰 어레이를 이용한 잡음 제거 시스템 및 방법은 방향성 잡음의 위치가 바뀌거나 큰 에너지를 가지는 돌발성 잡음이 발생하게 되면 GSC의 adaptive filter는 adaptation을 제대로 하지 못한다. 왜냐하면, 잡음의 변동성을 tracking하지 못하기 때문이다.

<7> 또한 잡음원이 사람의 목소리와 같이 자기상관도가 높은 경우에도 adaptive filter의 adaptation 성능도 하락한다. 이로 인해 잔여잡음이 존재하게 되는 문제점이 생긴다.

<8> 여기서 상관관계를 이용하는 제 1 방법은 제거하려는 방향의 잡음이 음악이라던가 TV 서리 등 자기상관도가 높은 colored noise의 형태일수록 성능 저하를 일으키게 키기 때문에 실제 환경에서는 사용할 수 없는 문제점이 있었다.

<9> 제 2 방법은 신호 대 잡음비가 낮아질수록 성능 저하를 일으키므로 이것 또한 실제 환경에는 적합하지 않는 문제점이 있었다.

발명의 내용

해결 하고자하는 과제

<10> 따라서 본 발명은 상기와 같은 종래의 문제점을 해결하기 위한 것으로서, 본 발명의 목적은 적응 빔포밍에서의 핵심 부분인 adaptive 모드 control 부분의 성능을 향상시켜 적응 빔포밍을 이용한 잡음제거 기술의 성능을 향상시키는 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치 및 방법을 제공하는 데 있다.

<11> 본 발명은 Hi(k, l)를 추정하여 Yi(k, l)를 제거하고 adaptive beamforming을 이용하여 Ni(k, l)을 제거함으로써 사용자의 음성 Si(k, l)을 복원하는데 있다.

과제 해결수단

<12> 상기한 목적을 달성하기 위한 본 발명에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치의 일 측면에 따르면, Adaptive beamformer 처리부로 제공하는 적어도 하나 이상의 마이크로폰을 통해 입력된 어레이 입력 신호를 입력받으면, 각각의 지정된 방향의 신호 세기를 검색하여 최대값을 갖는 신호 세기

를 검출하는 신호 세기 검출부; 및 상기 신호 세기 검출부를 통해 검출된 최대값을 갖는 신호 세기와 임계값을 비교하여 최대값을 갖는 신호 세기가 임계값을 초과할 경우 GSC의 adaptive 모드를 수행하지 않도록 제어하는 adaptive 모드 제어부를 포함한다.

<13> 상기 신호 세기 검출부는 어레이의 각 마이크로폰별로 입력된 잡음 섞인 음성을 임의의 길이의 hanning window를 씌어 프레임으로 구분하는 윈도우 처리부; 주파수 분석을 위해 각 마이크로폰 및 구분된 프레임 별로 DFT를 수행하는 DFT 처리부; Array를 구성하는 각 마이크 별로 pair를 이루어 탐지하려고 하는 방향으로 빔(beam)을 steering한 후 Cross-power spectrum 추정하는 상관관계 계산부; Cross-power spectrum을 정규화 시켜주기 위한 phase-transform weight를 구하는 weight 추정부; 및 방향이 검색되면 해당 방향으로부터 들어오는 소리의 세기를 측정하는 신호 세기 측정부를 포함할 수 있다.

<14> 본 발명에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 방법의 일 측면에 따르면, 고정 beamformer와 신호 블록부로 제공하는 적어도 하나 이상의 마이크로폰을 통해 입력된 어레이 입력 신호를 입력받으면, 각각의 지정된 방향의 신호 세기를 검색하여 최대값을 갖는 신호 세기를 검출하는 단계; 상기 신호 세기 단계에서 검출된 최대값을 갖는 신호 세기와 임계값을 비교하는 단계; 및 상기 판단 단계에서 최대값을 갖는 신호 세기가 임계값을 초과하면, GSC의 adaptive 모드를 수행하지 않도록 제어하는 단계를 포함한다.

<15> 상기 최대값을 갖는 신호 세기를 검출하는 단계는, 어레이의 각 마이크로폰별로 입력된 잡음 섞인 음성을 임의의 길이의 hanning window를 씌어 프레임으로 구분하는 단계; 주파수 분석을 위해 각 마이크로폰 및 구분된 프레임 별로 DFT를 수행하는 단계; Array를 구성하는 각 마이크 별로 pair를 이루어 탐지하려고 하는 방향으로 빔(beam)을 steering한 후 Cross-power spectrum 추정하는 단계; Cross-power spectrum을 정규화 시켜주기 위한 phase-transform weight를 구하는 단계; 및 방향이 검색되면 해당 방향으로부터 들어오는 소리의 세기를 측정하는 단계를 포함할 수 있다.

효 과

<16> 상술한 바와 같이 본 발명에 의한 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치 및 방법에 의하면, 적응 빔포밍이 성능을 보장하기 위한 조건 중에 하나이나, 자기 상관도가 높은 음원의 잡음을 제거할 경우 적응 필터의 적응을 적용하지 않아야 하나 이를 제어할 수 없는 종래의 문제점을 해결하는 뛰어난 효과가 있다.

발명의 실시를 위한 구체적인 내용

<17> 이하, 본 발명에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치 및 방법에 대한 바람직한 실시 예에 대하여 첨부한 도면을 참조하여 상세하게 살펴보기로 한다. 이 때, 아래에서 설명하는 시스템 구성은 본 발명의 설명을 위해서 인용한 시스템으로써 아래 시스템으로 본 발명을 한정하지 않음을 이 분야의 통상의 지식을 가진 자라면 이해해야할 것이다.

<18> 도 1은 종래 마이크로폰 어레이를 이용한 방향성 잡음 제거 시스템을 나타낸 도면으로써, 적어도 하나 이상의 마이크로폰(10), 마이크로폰(10)에 각각 연결된 단구간 분석부(20), 에코 제거부(30), 방향성 잡음 제거와, 정면소리 유/무를 기반으로 필터 weight update on/off하는 adaptive beamforming 처리부(40), 마이크간의 신호의 상관관계를 이용하여 정면소리를 정면소리 검출부(50), 잔여잡음 제거와 정면소리 유/무를 기반으로 잔여 잡음을 제어하는 Post-filtering부(60) 및 Overlap and add 처리부(70)를 포함한다.

<19> 한편, 하기에서의 기호 및 약어는 다음의 [표 1]과 같다.

<20> [표 1]

구분	기호	의미	기호	의미
공통	k	discrete 주파수 인덱스	N	잡음
	m	discrete time 인덱스	Φ_{AB}	A와 B의 cross-power spectrum
	l	frame 인덱스	μ	forgetting factor
	i	마이크 인덱스	$\hat{}$	추정치, 예) $\hat{S}$는 추정된 음성
	*	conjugate	w	window function
	Z	입력신호	SNR	signal-to-noise ratio
	Y	에코	SER	signal-to-echo ratio
	H	에코 path transfer function	DFT	discrete Fourier transform
	X	far-end 신호	FFT	fast Fourier transform
	S	음성	LMS	least mean square
에코 제거	Z^{aec}	에코가 제거된 신호	P^{far}	far-end 신호의 단구간 power
	η	double-talk검출 measure		
Adaptive beamforming	Z^{fb}	fixed beamformer output	E	error 신호
	Z^{sb}	signal blocking output	P^{gsc}	reference잡음의 power spectrum
	Z^{gsc}	adaptive beamformer output	A	spatial path transfer function
정면소리 detection	P^{srp}	정면 방향 소리의 power		
Post-filtering	ξ	a priori SNR	λ_S	음성 power-spectrum
	γ	a posteriori SNR	λ_N	잡음 power-spectrum

<21>

<22> 적어도 하나 이상의 마이크로폰(10)으로 이루어져 있으나, 4개의 마이크로폰(10)으로 어레이가 이루어져 있음을 예를 들어 설명하면, 각각의 마이크로폰(10)으로 입력되는 신호는 [수학식 1]과 같이 나타낼 수 있다.

<23> [수학식 1]

<24> $$Z_i(k,l) = Y_i(k,l) + N_i(k,l), i = 1...4$$

<25> 여기서, Z는 입력신호, Y는 에코, S는 음성, N은 잡음이며, i는 마이크 인덱스, k는 discrete 주파수 인덱스, l은 frame 인덱스이다.

<26> 한편, 각 마이크로폰(10)으로 들어오는 에코 $Y_i(k, l)$는 각각의 에코 path $H_i(k)$를 거쳐 4개의 마이크로폰(10)으로 입력되며, 각각의 마이크폰으로 입력되는 에코 신호는 [수학식 2]와 같이 나타낼 수 있다.

<27> [수학식 2]

<28> $$Y_i(k,l) = H_i(k)X(k,l), i = 1...4$$

<29> 여기서, Y는 에코, H는 에코 path transfer function, X는 far-end 신호, i는 마이크 인덱스, k는 discrete 주파수 인덱스, l은 frame 인덱스이다.

<30> 이때, X(k, l), N(k, l), S(k, l)는 [수학식 1], [수학식 2]에서 서로 관계있다고 가정한다.

<31> 상기 각 마이크로폰(10)을 통해 입력된 음성들은 각각의 단구간 분석부(20)를 통해 주파수 영역을 분석한다.

<32> 예를 들어 설명하면, 하나의 프레임은 2.56ms에 해당하며 이동구간은 128ms이다. 따라서 16Khz 샘플링에서 256ms는 4,096개로 샘플링된다.

<33> Hanning window를 사용할 경우 [수학식 3]과 같이 나타낼 수 있다.

<34> 그 이유는 에코 path impulse response를 모델링하기 위해서이다.

<35> 만약, 에코 path impulse response의 길이가 프레임 크기의 절반인 128ms 보다 길면, 에코 path는 제대로 추정되어지지 않으며 결과적으로 기술의 성능 하락을 일으키게 된다. 그 이유는 사용되는 필터들이 모두 주파수 영역에서 이루어지며 결국 시간 영역에서 본다면 circular convolution이 되기 때문이다.

<36> [수학식 3]

$$w(m) = 0.5\left(1 - \cos\left(2\pi \frac{m}{M}\right)\right), 0 \le m < M$$

<37>

<38> 여기서, w는 window function, M은 프레임을 구성하는 샘플의 개수, m은 discrete time 인덱스이다.

<39> 즉, 이동구간의 샘플의 개수를 T라고 가정하면, I 번째 프레임에서 입력신호와 far-end 신호의 주파수 영역 신호는 [수학식 3]의 윈도우와 DEF를 이용하여 다음의 [수학식 4]와 [수학식 5]와 같이 나타낸다.

<40> [수학식 4]

$$Z_i(k,l) = \sum_{m=0}^{M-1} w(m) z_i(l(M-T)+m) e^{-j\frac{2\pi}{M}mk}, 0 \le k < M, i = 1,\ldots,4$$

<41>

<42> 여기서, Z는 입력신호, i는 마이크 인덱스, k는 discrete 주파수 인덱스, l은 frame 인덱스, w는 window function, M은 프레임을 구성하는 샘플의 개수, m은 discrete time 인덱스이다.

<43> [수학식 5]

$$X(k,l) = \sum_{m=0}^{M-1} w(m) x(l(M-T)+m) e^{-j\frac{2\pi}{M}mk}, 0 \le k < M$$

<44>

<45> 여기서, X는 far-end, k는 discrete 주파수 인덱스, l은 frame 인덱스, w는 window function, M은 프레임을 구성하는 샘플의 개수, m은 discrete time 인덱스이다.

<46> 이후, DFT를 하기 위해 real-FFT를 이용하며, 소스코드는 ETSI 표준 특징 추출 프로그램을 이용한다.

<47> 여기서, M=4,096이며, FFT의 order는 M과 같다.

<48> 즉, [수학식 4]와 [수학식 5]의 신호를 가지고 에코와 잡음을 제거하고 복원된 사용자 음성 신호를 $\hat{S}(k,l)$라 가정하면, 이 신호는 inverse real-FFT를 거쳐 다시 [수학식 6]에서와 같은 시간 영역 신호로 복원된다.

<49> [수학식 6]

$$\hat{s}(l(M-T)+m) = \sum_{k=0}^{M-1} \hat{S}(k,l) e^{j\frac{2\pi}{M}mk}, 0 \le m < M$$

<50>

<51> 여기서, 추정된 음성, S는 음성, k는 discrete 주파수 인덱스, l은 frame 인덱스, M은 프레임을 구성하는 샘플의 개수, m은 discrete time 인덱스이다.

<52> 이렇게 복원된 신호는 window를 씌운 형태로 나타나게 되며 각 프레임의 복원된 신호는 이동구간만큼 중첩되어 합하여진다. 즉, I 번째 그리고 I+1 번째 프레임의 복원된 신호를 이용하여 T개의 샘플이 복원되며 [수학식 7]과 같이 나타낼 수 있다.

<53> [수학식 7]

$$\hat{s}(m) = \hat{s}(l(M-T)+m+T) + \hat{s}((l+1)(M-T)+m), 0 \le m < T$$

<54>

<55> 여기서, $\hat{S}$는 추정된 음성, S는 음성, k는 discrete 주파수 인덱스, l은 frame 인덱스, M은 프레임을 구성하는 샘플의 개수, m은 discrete time 인덱스이다.

<56> 상기와 같인 방법을 이용하여 도 2a 내지 도 2e에 도시된 바와 같이 신호를 중첩 구간만큼만 합하면, 그 구간에서의 신호 값들은 원 신호화할 수 있게 된다.

<57> 여기서, 도 2a는 원본 신호이며, 도 2b는 window이며, 도 2c는 첫 번째 프래임 신호이며, 도 2d는 두 번째 프레임 신호이며, 도 2e는 복원된 신호이다.

<58> 이러한 형태로 입력된 신호를 프레임 단위로 처리하여 복원하게 된다.

<59> 한편, 상기에서와 같이 에코가 제거된 신호는 adaptive beamforming 처리부(40)를 통해 방향성 잡음이

제거한다.

<60> 상기 adaptive beamforming 처리부(40)는 generalized sidelobe canceller(이하 "GSC"라 함)를 이용한다. 여기서 이용되는 GSC는 도 3에 도시된 바와 같이 고정 beamformer(41)와 신호 블록부(42), Adaptive 필터(43), 및 interference 제거부(44)와 같이 4 부분으로 이루어진다.

<61> 고정 beamformer(41)는 어레이를 사용자 방향(정면)에 steering한다. 즉, 정면에서 발성되는 소리이므로 마이크간 사용자의 음성 신호에 대해서 지연(delay)이 없으므로 [수학식 8]에서와 같이 그냥 평균을 취한다.

<62> [수학식 8]

<63>
$$Z^{fb}(k,l) = \frac{1}{4}\sum_{i=1}^{4} Z_i^{aec}(k,l)$$

<64> 여기서, Z^{fb}는 fixed beamformer output, k는 discrete 주파수 인덱스, l은 frame 인덱스, Z^{aec}는 에코가 제거된 신호, i는 마이크 인덱스이다.

<65> 신호 블록부(42)는 정면 방향을 main-lobe, 나머지 방향을 side-lobe라 한다. side-lobe 잡음을 구하는 과정이며 [수학식 9]를 통해 정면 소리는 제거되고 잡음만 얻는다.

<66> [수학식 9]

<67>
$$\begin{bmatrix} Z_1^{sb}(k,l) \\ Z_2^{sb}(k,l) \\ Z_3^{sb}(k,l) \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} Z_1^{aec}(k,l) \\ Z_2^{aec}(k,l) \\ Z_3^{aec}(k,l) \\ Z_4^{aec}(k,l) \end{bmatrix}$$

<68> 여기서, Z^{sb}는 signal blocking output, Z^{aec}는 에코가 제거된 신호, k는 discrete 주파수 인덱스, l은 frame 인덱스이다.

<69> 이후, Adaptive 필터(43)는 side-lobe에서 발생한 잡음은 A(k, l)이라는 spatial path transfer function을 거쳐서 어레이로 입력된다는 가정을 한다.

<70> 그러면, adaptive하게 A(k, l)를 추정하고 [수학식 9]에서 구하여진 Zsb를 이용하여 방향성 잡음을 제거하게 된다.

<71> 이후, Far-end 신호가 스피커에 어레이로 도착하는 path를 추정하여 에코를 제거하는 것과 마찬가지이다. 단, 각각의 마이크의 특성이 다르므로 [수학식 9]의 결과에는 사용자 음성이 약간은 남아 있다.

<72> 따라서 사용자 음성이 존재하는 경우에는 adaptation을 하지 않는다.

<73> 여기서, 정면소리 검출을 통해 adaptation 수행 여부를 결정한다.

<74> Adaptation 방식은 [수학식 10], [수학식 11], [수학식 12]를 통해 Complex LMS를 응용하여 주파수 영역 normalized LMS를 구현한다.

<75> [수학식 10]

<76>
$$\hat{A}_i(k,l+1) = \hat{A}_i(k,l) + (1-\mu)\frac{\xi(k,l)Z_i^*(k,l)}{P^{gsc}(k,l)}$$

<77> 여기서, A는 spatial path transfer function, ^는 추정치이며, x는 a priori SNR, k는 discrete 주파수 인덱스, l은 frame 인덱스, m은 forgetting factor, Z는 입력신호, *는 conjugate, i는 마이크 인덱스, P^{gsc}는 far-end 신호의 단구간 power이다.

<78> [수학식 11]

<79>
$$P^{gsc}(k,l) = \mu P^{gsc}(k,l-1) + (1-\mu)\sum_{i=1}^{3} |Z_i^{sb}(k,l)|^2$$

<80> 여기서, P^{gsc}는 far-end 신호의 단구간 power, k는 discrete 주파수 인덱스, l은 frame 인덱스, m은 forgetting factor, Z^{sb}는 signal blocking output, i는 마이크 인덱스이다.

<81> [수학식 12]

$$E(k,l) = Z^{fb}(k,l) - \sum_{i=1}^{3} \hat{A}_i(k,l) Z_i^{sb}(k,l)$$

<82>

<83> 여기서, E는 error 신호, Z^{fb}는 fixed beamformer output, k는 discrete 주파수 인덱스, l은 frame 인덱스, A는 spatial path transfer function, ^는 추정치이며, x는 a priori SNR, Z^{sb}는 signal blocking output이다.

<84> 이후, [수학식 13]에서와 같이 Interference 제거한다.

<85> [수학식 13]

$$Z^{gsc}(k,l) = \xi(k,l) = Z^{fb}(k,l) - \sum_{i=1}^{3} \hat{A}_i(k,l) Z_i^{sb}(k,l)$$

<86>

<87> 한편, 정면소리 검출은 Steered response power phase transform(SRP-PHAT)를 이용하여 정면방향에서 들어오는 소리의 세기(power)를 구한다. 이후, 에코제거를 거친 각 마이크로폰(10) 별 신호를 이용하여 다음 [수학식 14]를 이용하여 구한다.

<88> [수학식 14]

$$P^{srp}(l) = \frac{1}{M/2-1} \sum_{i=1}^{4} \sum_{j=i+1}^{4} \sum_{k=1}^{M/2-1} \frac{\Phi_{Z^{aec}_i Z^{aec}_j}(k,l)}{|\Phi_{Z^{aec}_i Z^{aec}_j}(k,l)|}$$

<89>

$$\Phi_{Z^{aec}_i Z^{aec}_j}(k,l) = (1-\mu)\Phi_{Z^{aec}_i Z^{aec}_j}(k,l-1) + \mu Z_i^{aec}(k,l) Z_j^{aec}(k,l)^*$$

<90>

<91> 여기서, P^{srp}는 정면 방향 소리의 power, F_{AB}는 A와 B의 cross-power spectrum, Z^{aec}는 에코가 제거된 신호, k는 discrete 주파수 인덱스, l은 frame 인덱스이며, $P^{srp}(l)$ 값은 0~6 값을 가진다.

<92> 이 값과 미리 정하여진 임계값과 비교하여 정면에서 소리가 존재하는지의 여부는 [수학식 15]을 통해 알아낸다.

<93> [수학식 15]

$$Flag^{srp}(l) = \begin{cases} 1, if\ P^{srp}(l) < TH^{srp} \\ 0, elsewhere \end{cases}$$

<94>

<95> 여기서, TH^{srp} 값은 1로 두었으며 환경에 따라 값을 바꿔줄 필요가 있다.

<96> 환경이라 함은 본 기술이 사용되어지는 공간의 reverberant의 정도를 말한다.

<97> SRP-PHAT 값은 magnitude로 normalization되기 때문에 조그만 소리라도 정면에서 발생해도 값이 커지는 경우가 있다.

<98> 따라서 좀더 안정적으로 정면소리를 검출하기 위해서 GSC의 output log power를 구하여 그 값이 미리 정해진 임계값과 비교하여 [수학식 16]과 함께 정면소리를 검출한다.

<99> [수학식 16]

$$Flag^{out}(l) = \begin{cases} 1, if\ P^{out}(l) < TH^{out} \\ 0, elsewhere \end{cases}$$

<100>

$$P^{out}(l) = \log\left(\frac{1}{M/2-1} \sum_{k=1}^{M/2-1} |Z^{gsc}(k,l)|^2 \right)$$

<101>

<102> 여기서, Z^{gsc}는 adaptive beamformer output, P^{out}는 출력 power이다.

<103> TH^{out}은 [수학식 16]과 같이 정의하였으며 이 값은 환경에 따라 값을 바꿔줄 필요가 있다.

<104> 환경이라 함은 본 기술이 사용되어진 경우 어레이된 마이크로폰과 화자의 거리를 말한다.

<105> [수학식 17]

<106>
$$Flag^{usr}(l) = \begin{cases} 1, if\ Flag^{out}(l) = 1\ and\ Flag^{out}(l) = 1 \\ 0, elsewhere \end{cases}$$

<107> 그리고 돌발잡음 또는 일반적으로 reverberant 환경에서는 beamforming의 성능이 하락하고 잔여잡음이 발생함에 따라 상기와 같은 상황에서 발생된 잔여잡음을 더욱 감쇄시키기 위해 부가적인 post-filter를 이용한다. 여기서, post-filter는 GSC를 거친 신호에 적용된다.

<108> 따라서 post-filter는 minimum mean square estimation of log-spectral amplitude(MMSE-LSA)를 기반으로 한다.

<109> [수학식 18]

<110>
$$G_{lsa}(k,l) = \frac{\xi(k,l)}{1+\xi(k,l)} \exp\left(\frac{1}{2}\int_{\upsilon(k,l)}^{\infty} \frac{e^{-\tau}}{\tau} d\tau\right)$$

<111> 여기서, x는 a priori SNR, k는 discrete 주파수 인덱스, l은 frame 인덱스이다.

<112> [수학식 19]

<113>
$$\xi(l,k) \equiv \frac{\lambda_S(l,k)}{\lambda_N(l,k)}, \gamma(l,k) \equiv \frac{|Z^{gsc}(l,k)|^2}{\lambda_N(l,k)}, \upsilon(l,k) \equiv \frac{\gamma(l,k)\xi(l,k)}{1+\xi(l,k)}$$

<114> 여기서, x는 a priori SNR, k는 discrete 주파수 인덱스, l은 frame 인덱스, l_s는 음성power-spectrum, l_N는 잡음 power-spectrum, g는 a posteriori SNR, m는 forgetting factor이다.

<115> [수학식 19], [수학식 20]에서 lN(l, k)는 [수학식 21]과 같이 추정된다.

<116> [수학식 20]

<117>
$$\hat{\lambda}_N(k,l) = \begin{cases} \mu\hat{\lambda}_N(k,l-1) + (1-\mu)|Z^{gsc}(k,l)|^2, if\ Flag^{usr}(l) = 0 \\ \hat{\lambda}_N(k,l-1), elsewhere \end{cases}$$

<118> 여기서, l_N는 잡음 power-spectrum, k는 discrete 주파수 인덱스, l은 frame 인덱스, m는 forgetting factor, Z^{gsc}는 adaptive beamformer output이다.

<119> 또한 lS(l, k) 자체는 추정이 어려우며 대신 x(k, l) 자체를 [수학식 21]과 같이 추정한다.

<120> [수학식 21]

<121>
$$\hat{\xi}(k,l) = (1-\mu)G_{lsa}^2(k,l-1)\gamma(k,l-1) + \mu\max\{\gamma(k,l)-1, 0\}$$

<122> 여기서, x는 a priori SNR, k는 discrete 주파수 인덱스, l은 frame 인덱스, g는 a posteriori SNR, m은 forgetting factor이다.

<123> 이렇게 계산된 G_{lsa}(k, l)와 최종적인 gain을 구하고, GSC를 통해 나온 신호에 적용하여 에코와 잡음이 제거된 을 구한다.

<124> 이는 [수학식 22]

<125> $$G(k,l)=\begin{cases}0.0001, if\ Flag^{usr}(l)=0\ and\ \gamma(k,l)>2\\ G_{ba}(k,l), elsewhere\end{cases}$$

<126> $\hat{S}(k,l)=G(k,l)Z^{fm}(k,l)$ 과 같다.

<127> S는 음성, ^는 추정치, k는 discrete 주파수 인덱스, l은 frame 인덱스이다.

<128> [수학식 22]를 보면 돌발잡음이 발생할 경우 0.0001이라는 작은 값으로 G(k, l)가 정해진다.

<129> 여기서 돌발잡음이란 정면소리가 검출되지 않았음에도 불구하고 a posteriori SNR g(k, l)값이 큰 경우를 말한다. 즉 사용자와 다른 각도에서 큰 소리가 발생했다는 것을 뜻한다.

<130> 도 3은 본 발명에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치의 구성을 나타낸 도면으로서, 본 발명에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치는 신호 세기 검출부(100), 및 adaptive 모드 제어부(200)를 포함한다.

<131> 신호 세기 검출부(100)는 고정 beamformer(41)와 신호 블록부(42), Adaptive 필터(43)로 구성된 adaptive beamforming 처리부(40)로 제공하는 적어도 하나 이상의 마이크로폰(10)을 통해 입력된 어레이 입력 신호를 입력받으면, 각각의 지정된 방향의 신호 세기를 검색하여 최대값을 갖는 신호 세기를 검출하며, 도 4에 도시된 바와 같이 신호 세기 검출부(100)는 윈도우 처리부(110), DFT 처리부(120), 상관관계 계산부(130), weight 추정부(140) 및 신호 세기 측정부(150)를 포함한다.

<132> 신호 세기 검출부(100)의 윈도우 처리부(110)는 어레이의 각 마이크로폰(10)별로 입력된 잡음 섞인 음성을 임의의 길이의 hanning window를 씌어 프레임으로 구분한다.

<133> 그리고 신호 세기 검출부(100)의 DFT 처리부(120)는 주파수 분석을 위해 각 마이크로폰(10) 및 구분된 프레임별로 DFT를 수행한다.

<134> 또한 신호 세기 검출부(100)의 상관관계 계산부(130)는 Array를 구성하는 각 마이크 별로 쌍(pair)을 이루어 탐지하려고 하는 방향으로 빔(beam)을 steering한 후 Cross-power spectrum 추정한다.

<135> 그리고 신호 세기 검출부(100)의 weight 추정부(140)는 Cross-power spectrum을 정규화 시켜주기 위한 phase-transform weight를 구한다.

<136> 또한 신호 세기 검출부(100)의 신호 세기 측정부(150)는 방향이 검색되면 해당 방향으로부터 들어오는 소리의 세기를 측정한다.

<137> 그리고 adaptive 모드 제어부(200)는 신호 세기 검출부(100)를 통해 검출된 최대값을 갖는 신호 세기와 임계값을 비교하여 최대값을 갖는 신호 세기가 임계값을 초과할 경우 GSC의 adaptive 모드를 수행하지 않도록 제어한다.

<138> 상기한 각 구성들에 대한 일반적인 기능 및 각각의 상세한 동작에 대하여는 그 설명을 생략하고, 본 발명에 상응하는 동작 위주로 그 동작들을 설명하기로 한다.

<139> 먼저, 적어도 하나 이상의 마이크로폰(10), 각각의 단구간 분석부(20)와 에코 제거부(30)를 통해 입력된 어레이 입력 신호에 대하여 고정 beamformer(41)와 신호 블록부(42), Adaptive 필터(43)로 구성된 adaptive beamforming 처리부(40)를 통해 SGC를 수행한다.

<140> 이때, adaptive beamforming 처리부(40)로 입력되는 어리이 입력 신호를 신호 세기 검출부(100)도 입력받는다.

<141> 그러면, 신호 세기 검출부(100)의 윈도우 처리부(110)는 어레이의 각 마이크로폰(10)별로 입력된 잡음 섞인 음성을 임의의 길이의 hanning window를 씌어 프레임으로 구분한다.

<142> 이어서, 신호 세기 검출부(100)의 DFT 처리부(120)는 주파수 분석을 위해 각 마이크로폰(10) 및 구분된 프레임별로 DFT를 수행한다.

<143> 이후, 신호 세기 검출부(100)의 상관관계 계산부(130)는 어레이를 구성하는 각 마이크 별로 쌍(pair)을 이루어 탐지하려고 하는 방향으로 빔(beam)을 steering한 후 Cross-power spectrum 추정한다.

<144> 한편, 신호 세기 검출부(100)의 weight 추정부(140)는 Cross-power spectrum을 정규화 시켜주기 위한 phase-transform weight를 구한다.

<145> 이후, 신호 세기 검출부(100)의 신호 세기 측정부(150)는 방향이 검색되면 해당 방향으로부터 들어오는 소리의 세기를 측정한다.

<146> 상기에서와 같이 각각의 신호 세기 측정부(150)를 통해 각 방향에서의 신호 세기가 측정되면, adaptive 모드 제어부(200)는 각 신호 세기 검출부(100)를 통해 검출된 최대값을 갖는 신호 세기와 기 설정된 임계값을 비교하여 최대값을 갖는 신호 세기가 임계값을 초과할 경우 GSC의 adaptive 모드를 수행하지 않도록 adaptive beamforming 처리부(40)를 제어한다.

<147> 만약, 최대값을 갖는 신호 세기가 임계값을 초과하지 않을 경우, 종래 기술에서와 같이 GSC의 adaptive 모드를 수행한다.

<148> 그러면, 상기와 같은 구성을 가지는 본 발명에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 방법에 대해 도 5를 참조하여 설명하기로 한다.

<149> 먼저, adaptive beamforming 처리부(40)로 제공하는 어레이 입력 신호를 입력받으면, 각각의 지정된 방향의 신호 세기를 검색하여 최대값을 갖는 신호 세기를 검출한다(S1).

<150> 이하 하기에서는 최대값을 갖는 신호 세기를 검출하는 단계(S1)의 세부 과정을 도 6을 참조하여 설명하기로 한다.

<151> 먼저, 어레이의 각 마이크로폰(10)별로 입력된 잡음 섞인 음성을 임의의 길이의 hanning window를 씌어 프레임으로 구분한다(S11).

<152> 이어서, 주파수 분석을 위해 각 마이크로폰(10) 및 구분된 프레임 별로 DFT를 수행한다(S12).

<153> 이후, Array를 구성하는 각 마이크 별로 pair를 이루어 탐지하려고 하는 방향으로 빔(beam)을 steering한 후 Cross-power spectrum 추정한다(S13).

<154> 또한 Cross-power spectrum을 정규화 시켜주기 위한 phase-transform weight를 구한다(S14).

<155> 이후, 방향이 검색되면 해당 방향으로부터 들어오는 소리의 세기를 측정한다(S15).

<156> 이어서, 검출된 최대값을 갖는 신호 세기가 임계값을 초과하는지의 여부를 판단한다(S2).

<157> 만약, 판단 단계(S2)에서 최대값을 갖는 신호 세기가 임계값을 초과하면(YES), adaptive beamforming 처리부(40)를 통해 GSC의 adaptive 모드가 수행되지 않도록 제어한다(S3).

<158> 한편, 판단 단계(S2)에서 최대값을 갖는 신호 세기가 임계값을 초과하지 않면(NO), adaptive beamforming 처리부(40)를 통해 GSC의 adaptive 모드가 수행되도록 제어한다(S4).

<159> 이상에서 본 발명은 기재된 구체적인 실시 예에 대해서만 상세히 설명하였지만 본 발명의 기술 사상 범위 내에서 다양한 변형 및 수정이 가능함은 당업자에게 있어서 명백한 것이며, 이러한 변형 및 수정이 첨부된 특허청구범위에 속함은 당연한 것이다.

도면의 간단한 설명

<160> 도 1은 종래 적응 빔포밍을 위한 사용자 방향의 소리 검출 시스템을 나타낸 기능블록도.

<161> 도 2a 내지 도 2e는 도 1에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 시스템에서 각 구간에서의 신호를 나태는 도면.

<162> 도 3은 본 발명에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치의 구성을 나타낸 기능블록도.

<163> 도 4는 도 3에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 장치에서 신호세기 검출부의 세부 구성을 나타낸 기능블록도.

<164> 도 5는 본 발명에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 방법을 나타낸 순서도,

<165> 도 6은 도 5에 따른 적응 빔포밍을 위한 사용자 방향의 소리 검출 기반의 적응 모드 제어 방법에서 신호 세기를 검출하는 단계의 세부 구성을 나타낸 순서도이다.

<166> <도면의 주요 부분에 대한 부호의 설명>

<167> 10 : 적어도 하나 이상의 마이크로폰 20 : 단구간 분석부

<168> 30 : 에코 제거부 40 : adaptive beamforming 처리부

<169> 50 : 정면소리 검출부 60 : Post-filtering부

<170> 70 : Overlap and add 처리부

<171> 100 : 신호 세기 검출부 110 : 윈도우 처리부

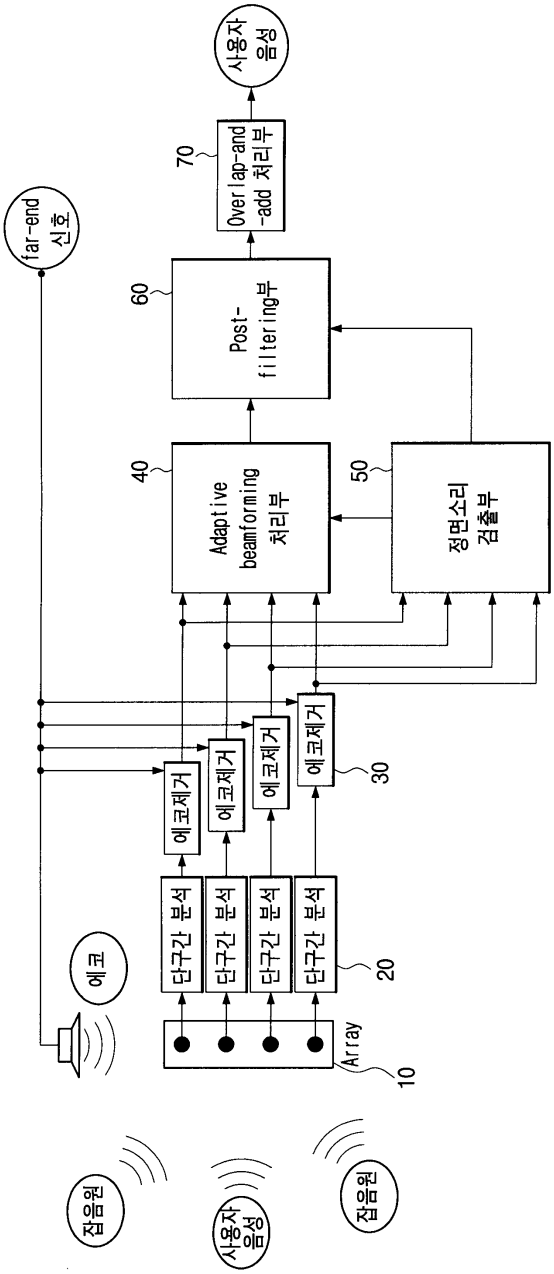
<172> 120 : DFT 처리부 130 : 상관관계 계산부

<173> 140 : weight 추정부 150 : 신호 세기 측정부

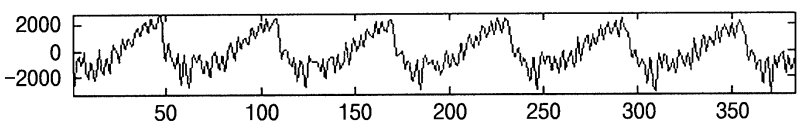
<174> 200 : adaptive 모드 제어부

도면

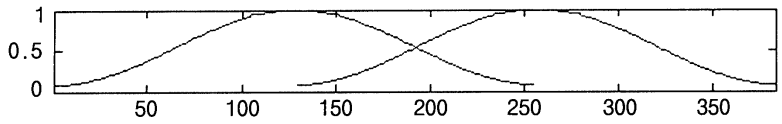
도면1


사용자 음성
70
Overlap-and -add 처리부
far-end 신호
60
Post-filtering부
40
Adaptive beamforming 처리부
50
정면소리 검출부
에코제거
에코제거
에코제거
에코제거
30
단구간 분석
단구간 분석
단구간 분석
단구간 분석
20
에코
Array
10
잡음원
사용자 음성
잡음원

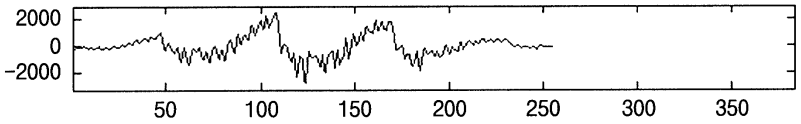
도면2a


2000
0
-2000
50
100
150
200
250
300
350

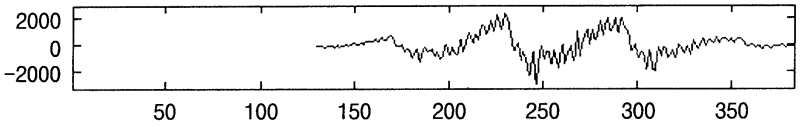
도면2b



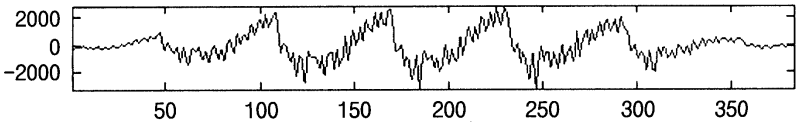
도면2c



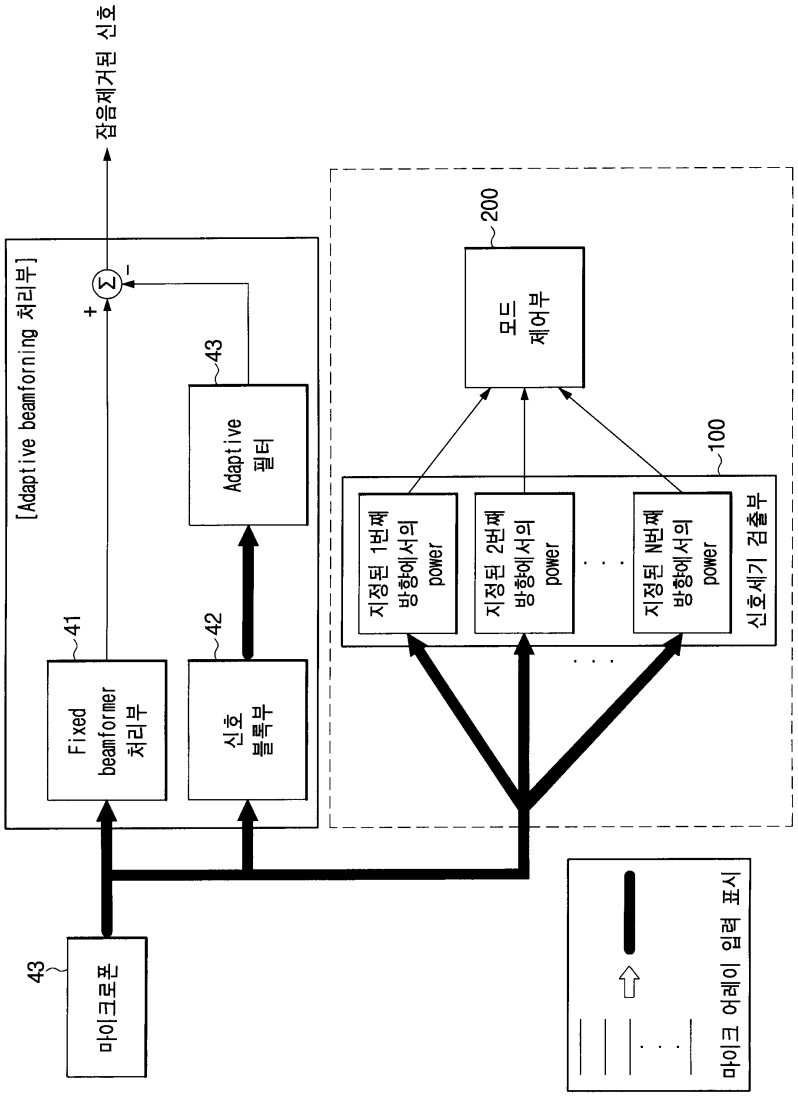
도면2d



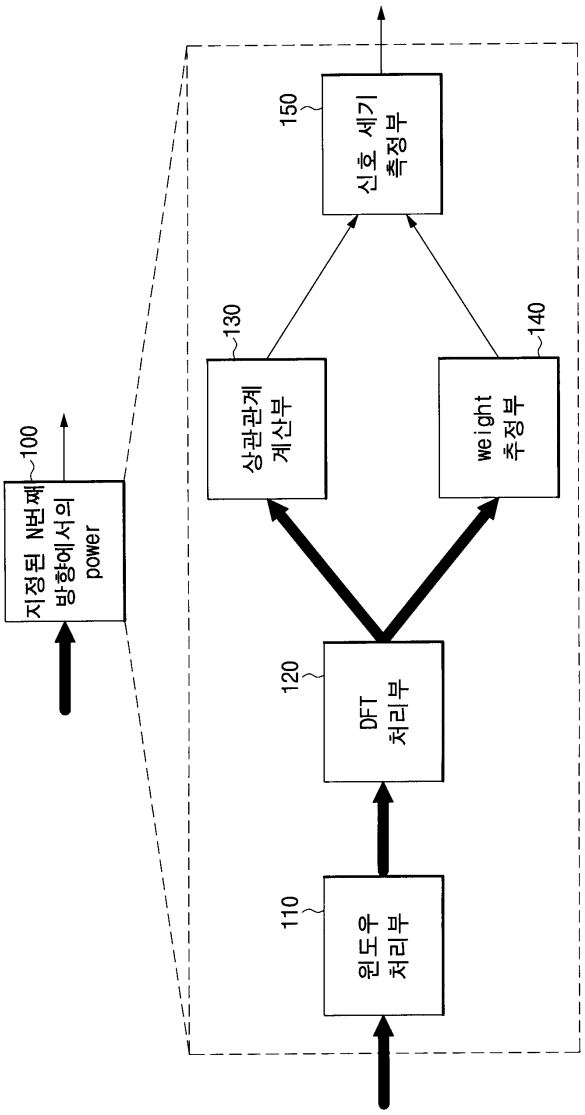
도면2e



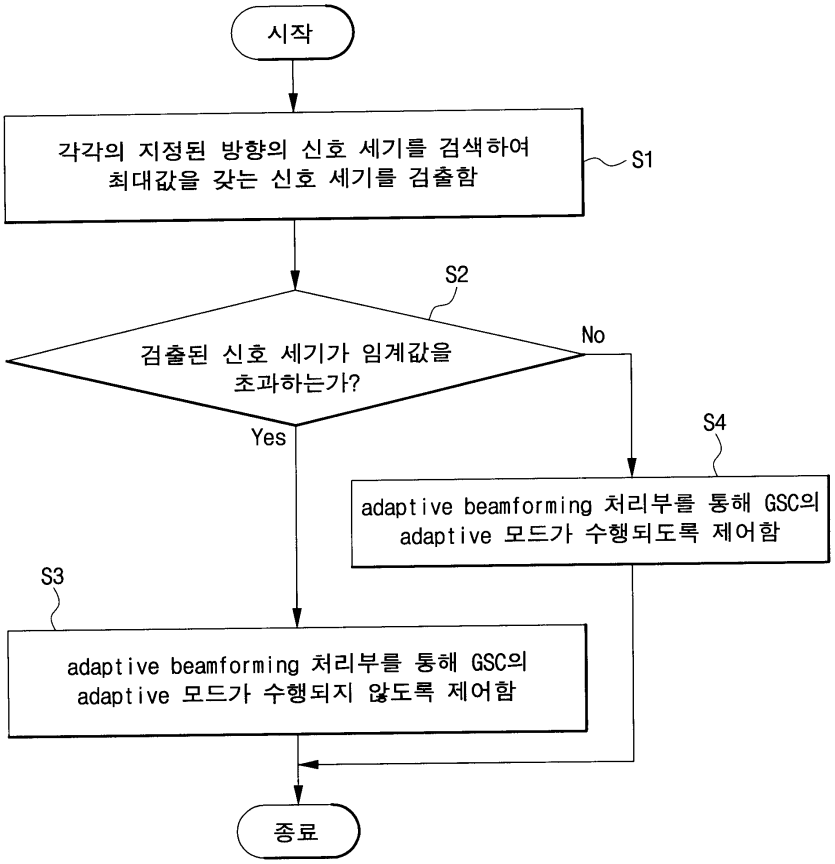
도면3


잡음제거된 신호
[Adaptive beamforming 처리부]
Σ
+
-
43
Adaptive 필터
41
42
Fixed beamformer 처리부
신호 블록부
43
마이크로폰
200
모드 제어부
100
지정된 1번째 방향에서의 power
지정된 2번째 방향에서의 power
지정된 N번째 방향에서의 power
신호세기 검출부
마이크 어레이 입력 표시

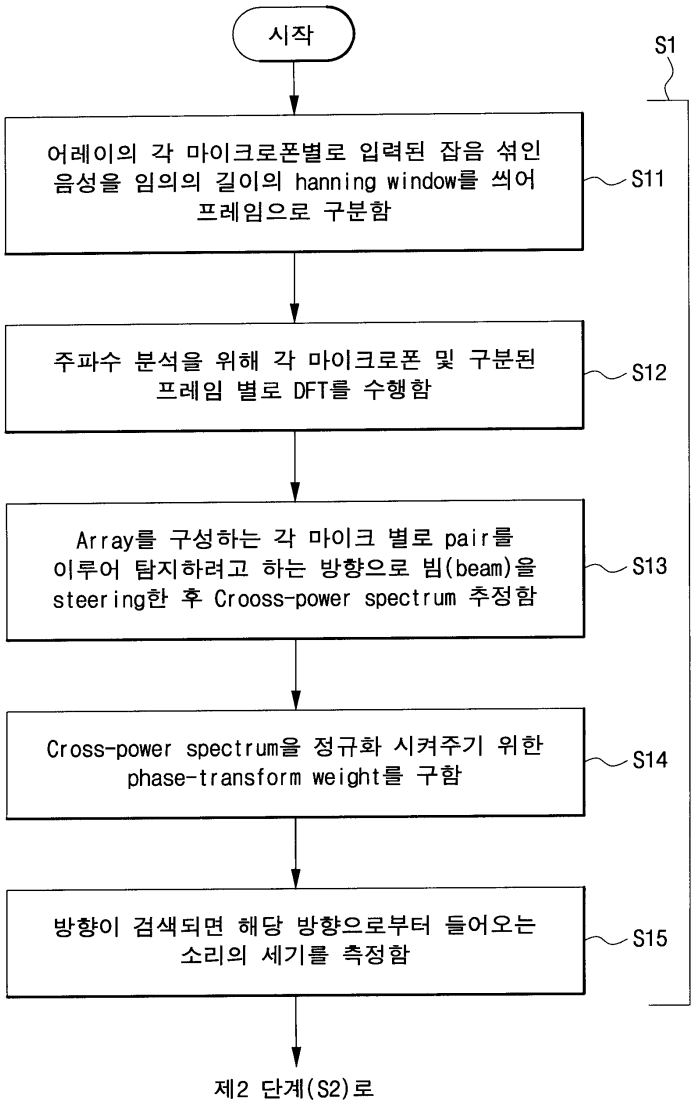
도면4


100
지정된 N번째 방향에서의 power
110
윈도우 처리부
120
DFT 처리부
130
상관관계 계산부
140
weight 추정부
150
신호 세기 측정부

도면5


시작
각각의 지정된 방향의 신호 세기를 검색하여 최대값을 갖는 신호 세기를 검출함
S1
S2
검출된 신호 세기가 임계값을 초과하는가?
No
Yes
S4
adaptive beamforming 처리부를 통해 GSC의 adaptive 모드가 수행되도록 제어함
S3
adaptive beamforming 처리부를 통해 GSC의 adaptive 모드가 수행되지 않도록 제어함
종료

도면6


시작
S1
어레이의 각 마이크로폰별로 입력된 잡음 섞인 음성을 임의의 길이의 hanning window를 씌어 프레임으로 구분함
S11
주파수 분석을 위해 각 마이크로폰 및 구분된 프레임 별로 DFT를 수행함
S12
Array를 구성하는 각 마이크 별로 pair를 이루어 탐지하려고 하는 방향으로 빔(beam)을 steering한 후 Crooss-power spectrum 추정함
S13
Cross-power spectrum을 정규화 시켜주기 위한 phase-transform weight를 구함
S14
방향이 검색되면 해당 방향으로부터 들어오는 소리의 세기를 측정함
S15
제2 단계(S2)로