



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0073852
(43) 공개일자 2019년06월27일

(51) 국제특허분류(Int. Cl.)
G10L 21/0208 (2013.01) G10L 21/0216 (2013.01)
(52) CPC특허분류
G10L 21/0208 (2013.01)
G10L 2021/02166 (2013.01)
(21) 출원번호 10-2017-0175057
(22) 출원일자 2017년12월19일
심사청구일자 2017년12월19일

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
박형민
서울특별시 강남구 삼성로 212 3동 1002호 (대치동, 은마아파트)
이준민
서울특별시 마포구 노고산동 109-59번지 502호
(74) 대리인
이지연

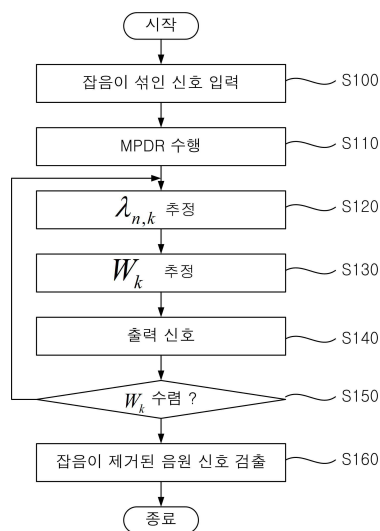
전체 청구항 수 : 총 6 항

(54) 발명의 명칭 **우도 최대화를 이용한 빔포밍 방법**

(57) 요약

본 발명은 우도 최대화를 이용한 빔포밍 방법에 관한 것이다. 상기 빔포밍 방법은, (a) 시간 프레임 n 및 주파수 k 에서의 잡음이 섞인 입력 신호($X_{n,k}$)를 수신하는 단계; (b) 입력 신호에서 잡음을 제거한 목표 신호($Y_{n,k}$)가 시간-주파수 영역에서 평균이 0인 complex generalized Gaussian 분포 또는 complex gamma distribution 을 따른다고 가정하여, 상기 목표 신호($Y_{n,k}$)에 대한 확률 밀도 함수를 결정하는 단계; (c) 상기 확률 밀도 함수에 대한 로그 우도를 최대화시키도록 상기 목표 신호의 분산($\lambda_{n,k}$)을 추정하는 단계; (d) 상기 확률 밀도 함수에 대한 로그 우도를 최대화시키도록 상기 비용함수를 최대화하는 필터(w_k)를 추정하는 단계; (e) 상기 필터(w_k)가 수렴될 때까지 상기 (c) 단계와 (d) 단계를 반복적으로 추정하여, 최종적으로 필터(w_k)를 얻는 단계; 를 구비한다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호 10063424

부처명 산업통상자원부

연구관리전문기관 한국산업기술평가관리원

연구사업명 로봇산업핵심기술개발사업

연구과제명 실내용 음성대화 로봇을 위한 원거리 음성인식 기술 및 멀티 태스크 대화처리 기술 개발

기 여 율 1/1

주관기관 한국과학기술원

연구기간 2016.05.01 ~ 2020.04.30

명세서

청구범위

청구항 1

- (a) 시간 프레임 n 및 주파수 k 에서의 잡음이 섞인 입력 신호($X_{n,k}$)를 수신하는 단계;
- (b) 입력 신호에서 잡음을 제거한 목표 신호($Y_{n,k}$)가 시간-주파수 영역에서 평균이 0인 complex generalized Gaussian 분포 또는 complex Gamma distribution을 따른다고 가정하여, 상기 목표 신호($Y_{n,k}$)에 대한 확률 밀도

함수($p_{\text{CGG}}(Y_{n,k})$)를 결정하는 단계;

- (c) 상기 확률 밀도 함수에 대한 로그 우도를 최대화시키도록 상기 목표 신호의 분산($\lambda_{n,k}$)을 추정하는 단계;
- (d) 상기 확률 밀도 함수에 대한 로그 우도를 최대화시키도록 상기 비용함수를 최대화하는 필터(w_k)를 추정하는 단계;
- (e) 상기 필터(w_k)가 수렴될 때까지 상기 (c) 단계와 (d) 단계를 반복적으로 추정하여, 최종적으로 필터(w_k)를 얻는 단계;

를 구비하는 것을 특징으로 하는 우도 최대화를 이용한 빔포밍 방법.

청구항 2

제1항에 있어서, 상기 필터(w_k)는 아래의 수학식에 따라 입력 신호($X_{n,k}$)로부터 목표 신호($Y_{n,k}$)를 얻기 위한 필터인 것을 특징으로 하는 우도 최대화를 이용한 빔포밍 방법.

$$Y_{n,k} = w_k^H X_{n,k}$$

청구항 3

제1항에 있어서, 상기 (b) 단계에서의 확률 밀도 함수($p_{\text{CGG}}(Y_{n,k})$)는 아래의 수학식으로 나타내며,

$$p_{\text{CGG}}(Y_{n,k}) = \frac{\alpha}{\lambda_{n,k}} \exp \left\{ - \left[\frac{Y_{n,k} Y_{n,k}^*}{\lambda_{n,k}} \right]^c \right\}.$$

여기서, α 는 $p_{\text{CGG}}(Y_{n,k})$ 가 확률밀도함수가 되도록 하는 상수이며, 양수 c 는 가우시안 분포의 Shape 파라미터로서, $c = 1$ 인 경우 complex Gaussian 분포를 나타내고 $c < 1$ 인 경우 가우시안 분포보다 sparse한 분포를 나타내고 $c > 1$ 인 경우 가우시안 분포보다 덜 sparse한 분포를 나타내는 것을 특징으로 하는 우도 최대화를 이용한 빔포밍 방법.

청구항 4

제1항에 있어서, 상기 (b) 단계에서의 확률 밀도 함수($p_{CGG}(Y_{n,k})$)는 아래의 수학식으로 나타내며,

$$p_{CGG}(Y_{n,k}) = \frac{1}{\Gamma(\alpha)\beta^\alpha} Y_{n,k}^{\alpha-1} e^{-Y_{n,k}/\beta}$$

여기서, α 는 $|\alpha| > 0$ 인 복소수 상수이며, β 는 양의 실수인 것을 특징으로 하는 우도 최대화를 이용한 빔포밍 방법.

청구항 5

제1항에 있어서, 상기 (c) 단계 및 (d) 단계에서 $\mathbf{w}_k^H \mathbf{a}_k = 1$ (여기서, $\mathbf{a}_k$ 는 방향 벡터임)의 제한 조건을 만족하는 로그 우도를 최대화시켜, 상기 필터와 분산을 추정하는 것을 특징으로 하는 우도 최대화를 이용한 빔포밍 방법.

청구항 6

제1항에 있어서, 상기 분산($\lambda_{n,k}$)은 $Y_{n,k} Y_{n,k}^*$ 를 포함한 주변 프레임들의 평균값을 사용하고, 초기화를

위하여 필요시 MVDR 빔포밍 또는 MPDR 빔포밍으로 얻어진 $Y_{n,k} Y_{n,k}^*$ 를 이용하는 것을 특징으로 하는 우도 최대화를 이용한 빔포밍 방법.

발명의 설명

기술분야

[0001] 본 발명은 우도 최대화를 이용한 빔포밍 방법에 관한 것으로서, 더욱 구체적으로는 잡음에 의해 왜곡이 발생하는 음향 신호들에 대하여 다중 마이크로폰을 이용한 빔포밍 출력 신호를 complex generalized Gaussian 분포 또는 complex Gamma distribution 을 따르는 것으로 가정하고 확률밀도함수의 로그 우도가 최대화되는 필터를 추정하는 빔포밍 방법에 관한 것이다.

배경기술

[0002] 일반적인 실내 환경에서 녹음된 음향 신호는 잡음에 의해서 왜곡이 발생하여 음질이 저하되거나 음성인식의 성능이 떨어지게 된다. 이러한 문제를 극복하기 위해 다중 마이크로폰을 이용한 빔포밍 출력 신호를 최소화 시키는 방향으로 필터를 추정하는 기법을 사용하여 잡음을 제거해오고 있다.

[0003] 이러한 필터 추정 기법 중 하나가 Minimum Variance Distortionless Response (MVDR) 빔포밍 알고리즘이다. 이하, MVDR 빔포밍 알고리즘에 대하여 개략적으로 설명한다.

[0004] 단구간 푸리에 변환(Short-time Fourier Transform)을 통해 시간-주파수 영역에서 M개의 마이크로폰으로 취득된 입력 신호(X)는 다음과 같이 표현할 수 있다.

수학식 1

$$\mathbf{X}_{n,k} = \mathbf{H}_k^1 S_{n,k} + \sum_{i=2}^N \mathbf{H}_k^i N_{n,k}^i$$

[0005]

[0006] 여기서 $X_{n,k}$, $S_{n,k}$, $N_{n,k}$, H_k 는 각각 시간프레임 n , 주파수 k 에서 마이크로폰 입력 신호로 구성된 M 차원 벡터, 음원 신호, 잡음 신호, 그리고 각 음성과 잡음 신호별 적용되는 공간필터로 이루어진 $M \times N$ 차원 행렬을 나타낸다. H_k^i 는 음원으로부터 마이크로폰까지의 공간 필터를 나타낸다.

[0007] 음질 또는 음성인식 성능이 떨어지는데 주된 영향을 미치는 잡음 성분을 제거하기 위해서 입력 신호에서 잡음을 제거한 목표 신호($Y_{n,k}$)는 다음과 같이 나타낼 수 있다.

수학식 2

$$Y_{n,k} = \mathbf{w}_k^H \mathbf{X}_{n,k}$$

[0008]

[0009] 빔포밍의 목적은 공간 필터를 통과하여 들어온 잡음 신호($N_{n,k}$)가 필터링된 출력의 파워

$\mathbf{w}_k^H \mathbf{X}_{n,k} (\mathbf{w}_k^H \mathbf{X}_{n,k})^H$ 를 최소화하며 수학식 3과 같이 목표 음원의 방향의 이득을 1로 고정하는 알고리즘이다.

수학식 3

$$\mathbf{w}_k^H \mathbf{a}_k = 1$$

[0010]

[0011] 여기서, w_k 는 비용함수를 최대화하는 필터이며, a_k 는 방향 벡터이며, 위첨자 H 는 행렬의 hermitian을 나타낸다.

[0012] L_k , d , c , F_s 를 각각 총 주파수 bin의 개수, 음원과 마이크 사이의 거리, 소리의 속도, 샘플링 주파수라 하면 방향 벡터(steering vector ; a_k)는 수학식 4와 같이 나타낼 수 있다. 상기 L_k 는 이산푸리에변환(discrete Fourier transform)의 DFT 포인트수를 나타낸다.

수학식 4

$$\mathbf{a}_k = \exp(-2\pi \frac{(k-1)}{L_k} \frac{d}{c} F_s)$$

[0013]

[0014] Lagrange multiplier를 이용하여 조건을 만족하는 비용함수는 다음과 같다.

수학식 5

$$J_k = \mathbf{w}_k^H \sum_{n=1}^{L_n} \mathbf{N}_{n,k} \mathbf{N}_{n,k}^H \mathbf{w}_k + \alpha (\mathbf{w}_k^H \mathbf{a} - 1) = \mathbf{w}_k^H \mathbf{R}_k^N \mathbf{w}_k + \alpha (\mathbf{w}_k^H \mathbf{a}_k - 1)$$

[0015]

[0016] 이 때, 비용함수를 최대화 하는 필터($\mathbf{w}_k$)는 아래와 같다. 여기서, $\mathbf{R}_k$ 는 프레임 n , 주파수 k 에서의 입력신호 $X_{n,k}$ 에 대한 공분산 행렬을 나타내며, $\mathbf{R}_k^N$ 은 입력신호의 잡음성분인 $N_{n,k}$ 에 대한 공분산 행렬을 나타낸다.

수학식 6

$$\mathbf{w}_k = \frac{\mathbf{R}_k^{N^{-1}} \mathbf{a}_k}{\mathbf{a}_k^H \mathbf{R}_k^{N^{-1}} \mathbf{a}_k}$$

[0017]

[0018] 전술한 MVDR 알고리즘은 목표 음성이 없고 잡음만 존재할 때는 잡음 신호($N_{n,k}$)를 추정할 수 있지만, 목표 음성이 존재하는 구간에서는 잡음 신호를 정확히 추정하기 힘들다.

[0019] 본 발명에서는 기존의 빔포밍 출력 신호를 최소화 시키는 필터를 추정하는 대신 빔포밍 출력 신호를 complex generalized Gaussian 분포를 따른다고 가정하고 이 확률밀도함수의 로그 우도가 최대화되는 필터를 추정하는 알고리즘을 제안한다.

선행기술문헌

특허문헌

[0020] (특허문헌 0001) 한국공개특허공보 제 10-2013-0068251호

발명의 내용

해결하려는 과제

[0021] 전술한 문제점을 해결하기 위한 본 발명의 목적은 빔포밍 출력 신호를 complex generalized Gaussian 분포를 따른다고 가정하고 이 확률밀도함수의 로그 우도가 최대화되는 필터를 추정하는 빔포밍 방법을 제공하는 것이다.

과제의 해결 수단

[0022] 전술한 기술적 과제를 달성하기 위한 본 발명의 특징에 따른 우도 최대화를 이용한 빔포밍 방법은, (a) 시간 프레임 n 및 주파수 k 에서의 잡음이 섞인 입력 신호($X_{n,k}$)를 수신하는 단계; (b) 입력 신호에서 잡음을 제거한 목표 신호($Y_{n,k}$)가 시간-주파수 영역에서 평균이 0인 complex generalized Gaussian 분포 또는 complex Gamma

distribution을 따른다고 가정하여, 상기 목표 신호($Y_{n,k}$)에 대한 확률 밀도 함수($p_{\text{CGG}}(Y_{n,k})$)를 결정하는 단계; (c) 상기 확률 밀도 함수에 대한 로그 우도를 최대화시키도록 상기 목표 신호의 분산($\lambda_{n,k}$)을 추정하는 단계; (d) 상기 확률 밀도 함수에 대한 로그 우도를 최대화시키도록 상기 비용함수를 최대화하는 필터($\mathbf{w}_k$)를 추정하는 단계; (e) 상기 필터($\mathbf{w}_k$)가 수렴될 때까지 상기 (c) 단계와 (d) 단계를 반복적으로 추정하여, 최종적으로 필터($\mathbf{w}_k$)를 얻는 단계; 를 구비한다.

[0023] 전술한 특징에 따른 우도 최대화를 이용한 빔포밍 방법에 있어서, 상기 필터(w_k)는 아래의 수학식에 따라 입력 신호($X_{n,k}$)로부터 목표 신호($Y_{n,k}$)를 얻기 위한 필터인 것이 바람직하다.

$$Y_{n,k} = w_k^H X_{n,k}$$

[0024]

[0025] 전술한 특징에 따른 우도 최대화를 이용한 빔포밍 방법에 있어서, 상기 (b) 단계에서의 확률 밀도 함수

($p_{CGG}(Y_{n,k})$)는 아래의 수학식으로 나타내며,

$$p_{CGG}(Y_{n,k}) = \frac{\alpha}{\lambda_{n,k}} \exp \left\{ - \left[\frac{Y_{n,k} Y_{n,k}^*}{\lambda_{n,k}} \right]^c \right\}.$$

[0026]

[0027] 여기서, α 는 $p_{CGG}(Y_{n,k})$ 가 확률밀도함수가 되도록 하는 상수이며, 양수 c 는 가우시안 분포의 Shape 파라미터로서, $c = 1$ 인 경우 complex Gaussian 분포를 나타내고 $c < 1$ 인 경우 가우시안 분포보다 sparse한 분포를 나타내고 $c > 1$ 인 경우 가우시안 분포보다 덜 sparse한 분포를 나타낸다.

[0028] 전술한 특징에 따른 우도 최대화를 이용한 빔포밍 방법에 있어서, 상기 (b) 단계에서의 확률 밀도 함수

($p_{CGG}(Y_{n,k})$)는 아래의 수학식으로 나타내며,

$$p_{CG}(Y_{n,k}) = \frac{1}{\Gamma(\alpha)\beta^\alpha} Y_{n,k}^{\alpha-1} e^{-Y_{n,k}/\beta}$$

[0029]

[0030] 여기서, α 는 $|a| > 0$ 인 복소수 상수이며, β 는 양의 실수이다.

[0031] 전술한 특징에 따른 우도 최대화를 이용한 빔포밍 방법에 있어서, 상기 (c) 단계 및 (d) 단계에서

$$w_k^H a_k = 1$$

(여기서, a_k 는 방향 벡터임)의 제한 조건을 만족하는 로그 우도를 최대화시켜, 상기 필터와 분산을 추정하는 것이 바람직하다.

[0032] 전술한 특징에 따른 우도 최대화를 이용한 빔포밍 방법에 있어서, 상기 분산($\lambda_{n,k}$)은 $Y_{n,k} Y_{n,k}^*$ 를 포함한 주변 프레임들의 평균값을 사용하고, 초기화를 위하여 필요시 MVDR 빔포밍 또는 MPDR 빔포밍으로 얻어진

$Y_{n,k} Y_{n,k}^*$ 를 이용하는 것이 바람직하다.

발명의 효과

[0033] 본 발명에 따른 빔포밍 방법은 목표 신호인 잡음이 제거된 음성 신호가 complex generalized Gaussian 분포를 따른다고 가정하여 해당 분포의 로그 우도가 최대화되도록 필터와 목표 신호에 대한 분산을 추정한다.

[0034] 따라서, 본 발명에 따른 빔포밍 방법은 종래의 다른 빔포밍 방법들에 비하여 단어 오 인식률이 낮다.

도면의 간단한 설명

[0035] 도 1은 본 발명의 바람직한 실시예에 따른 우도 최대화를 이용한 빔포밍 방법을 도시한 흐름도이다.

도 2는 본 발명의 바람직한 실시예에 따른 우도 최대화를 이용한 빔포밍 방법 및 종래의 방법들을 이용하여 단어 오인식율을 측정한 결과표를 도시한 도표이다.

발명을 실시하기 위한 구체적인 내용

[0036] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예에 따른 우도 최대화를 이용한 빔포밍 방법에 대하여 구체적으로 설명한다.

[0037] 도 1은 본 발명의 바람직한 실시예에 따른 우도 최대화를 이용한 빔포밍 방법을 도시한 흐름도이다.

[0038] 전술한 MVDR 알고리즘은 목표 음성이 없고 잡음만 존재할 때는 잡음 신호를 추정할 수 있지만, 목표 음성이 존재하는 구간에서는 잡음 신호를 정확히 추정하기 힘들다. 따라서, 도 1을 참조하면, 본 발명에 따른 빔포밍 방법은 잡음이 섞인 신호가 입력되면(단계 100), Minimum Power Distortionless Response (MPDR) 빔포밍 알고리즘을 이용하여 잡음 신호의 출력 파워를 최소화 하는 대신 입력 신호의 출력의 파워를 최소화하도록 비용함수(J_k)를 수학식 7과 같이 설정한다(단계 110).

수학식 7

$$J_k = \mathbf{w}_k^H \sum_{n=1}^{L_n} \mathbf{X}_{n,k} \mathbf{X}_{n,k}^H \mathbf{w}_k + \alpha (\mathbf{w}_k^H \mathbf{a} - 1) = \mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k + \alpha (\mathbf{w}_k^H \mathbf{a}_k - 1)$$

[0039]

[0040] 여기서 L_n 은 전체 프레임의 개수를 나타낸다.

[0041] MVDR과 같은 방법으로 비용함수를 최소화 하는 필터($\mathbf{w}_k$)는 수학식 8과 같다.

수학식 8

$$\mathbf{w}_k = \frac{\mathbf{R}_k^{-1} \mathbf{a}_k}{\mathbf{a}_k^H \mathbf{R}_k^{-1} \mathbf{a}_k}$$

[0042]

[0043] 만약, $\mathbf{R}_k$ 가 singular matrix에 가까우면 역행렬을 제대로 추정할 수 없으므로 행렬의 diagonal 부분에 작은 값을 더해 주어 최종적으로 추정된 필터는 수학식 9와 같다.

수학식 9

$$\mathbf{w}_k = \frac{(\mathbf{R}_k + \delta \mathbf{I})^{-1} \mathbf{a}_k}{\mathbf{a}_k^H (\mathbf{R}_k + \delta \mathbf{I})^{-1} \mathbf{a}_k}$$

[0044]

[0045] 여기서 δ 는 공분산이 singular matrix 가 되지 않도록 하는 작은 상수값이며, $\mathbf{R}_k$ 는 프레임 n , 주파수 k 에서의 입력 신호 $\mathbf{X}_{n,k}$ 의 공분산행렬을 나타낸다.

[0046] 전술한 MPDR 빔포밍 알고리즘은 단구간 푸리에 변환을 통해 시간-주파수 영역에서 빔포밍 출력 신호의 파워를 최소화하도록 필터를 추정하였지만, 잡음 신호의 공분산과 입력 신호의 공분산의 차이로 인하여 잡음이 잘 제거되지 않는다.

[0047] 따라서, 본 발명에 따른 빔포밍 방법은 Complex Gaussian 분포를 이용한 최대우도 기반 빔포밍 알고리즘을 이용

하여 잡음이 제거된 음성 신호가 Gaussian 분포를 따른다고 가정하여 해당 분포의 로그 우도가 최대화 되도록 잡음을 제거한다. 이 때, $Y_{n,k}$ 가 시간-주파수 영역에서 평균이 0인 complex generalized Gaussian 분포를 따른다고 가정하면 $Y_{n,k}$ 의 확률밀도함수는 수학적 10과 같이 나타낼 수 있다.

수학적 10

$$p_{\text{CGG}}(Y_{n,k}) = \frac{\alpha}{\lambda_{n,k}} \exp \left\{ - \left[\frac{Y_{n,k} Y_{n,k}^*}{\lambda_{n,k}} \right]^c \right\}.$$

[0048]

[0049] 여기서 α 는 $p_{\text{CGG}}(Y_{n,k})$ 가 확률밀도함수가 되도록 하는 상수이며, 양수 c 는 가우시안 분포의 Shape 파라미터로서, $c = 1$ 인 경우 complex Gaussian 분포를 나타내며 $c < 1$ 인 경우 Gaussian 분포보다 sparse한 분포를 나타내며, $c > 1$ 인 경우 Gaussian 분포보다 덜 sparse한 분포를 나타낸다. $\lambda_{n,k}$ 는 $Y_{n,k}$ 의 분산을 나타낸다.

[0050] 한편, 수학적 11과 같은 complex Gamma distribution 과 같이 sparse한 분포를 나타낼 수 있는 다양한 분포 함수에도 적용 가능하다.

수학적 11

$$p_{\text{CG}}(Y_{n,k}) = \frac{1}{\Gamma(\alpha)\beta^\alpha} Y_{n,k}^{\alpha-1} e^{-Y_{n,k}/\beta}$$

[0051]

[0052] 여기서, α 는 $|\alpha| > 0$ 인 복소수 상수이며, β 는 양의 실수이다.

[0053] 최대 우도 기법으로 필터 w_k 와 $\lambda_{n,k}$ 를 추정하기 위해 $\mathbf{w}_k^H \mathbf{a}_k = 1$ 의 제한조건을 만족하는 로그우도를 최대화한다. 로그 우도 함수는 수학적 12로 나타낼 수 있다.

수학적 12

$$\begin{aligned} l(\mathbf{w}_k, \lambda_{n,k}) &= \sum_{n=1}^{L_n} \left\{ \log \alpha - \log(\lambda_{n,k}) - \left[\frac{Y_{n,k} Y_{n,k}^*}{\lambda_{n,k}} \right]^c \right\} + \alpha (\mathbf{w}_k^H \mathbf{a}_k - 1) \\ &= \sum_{n=1}^{L_n} \left\{ \log \alpha - \log(\lambda_{n,k}) - \left[\frac{\mathbf{w}_k^H \mathbf{X}_{n,k} \mathbf{X}_{n,k}^H \mathbf{w}_k}{\lambda_{n,k}} \right]^c \right\} + \alpha (\mathbf{w}_k^H \mathbf{a}_k - 1) \end{aligned}$$

[0054]

[0055] 로그우도를 최대화하는 분산 $\lambda_{n,k}$ 는 수학적 13 및 수학적 14와 같이 구할 수 있다(단계 120).

수학식 13

$$\frac{\partial l(\mathbf{w}_k, \lambda_{n,k})}{\partial \lambda_{n,k}} = -\frac{1}{\lambda_{n,k}} + c \frac{[Y_{n,k} Y_{n,k}^*]^c}{\lambda_{n,k}^{c+1}} = 0.$$

[0056]

수학식 14

$$\lambda_{n,k} = c^{1/c} Y_{n,k} Y_{n,k}^*$$

[0057]

[0058] 추정하는 $\lambda_{n,k}$ 의 신뢰성을 향상시켜주기 위하여, 수학식 15와 같이, $Y_{n,k} Y_{n,k}^*$ 를 포함한 주변 프레임들의 평균을 $\lambda_{n,k}$ 값으로 사용한다.

수학식 15

$$\lambda_{n,k} = \frac{1}{2\tau+1} \sum_{l=n-\tau}^{n+\tau} Y_{l,k} Y_{l,k}^*$$

[0059]

[0060] 처음에는 $Y_{n,k}$ 의 값을 알 수 없으므로 기존의 입력 신호의 출력 파워를 최대화 시키는 필터 $\mathbf{w}_k$ 를 이용하여 $\lambda_{n,k}$ 를 초기화한다.

[0061] 비슷한 방법으로 로그 우도를 최대화 하는 필터 $\mathbf{w}_k$ 는 수학식 16 및 수학식 17과 같이 구할 수 있다(단계 130).

수학식 16

$$\frac{\partial l(\mathbf{w}_k, \lambda_{n,k})}{\partial \mathbf{w}_k^*} = -c \sum_{n=1}^{L_n} \left[\frac{Y_{n,k} Y_{n,k}^*}{\lambda_{n,k}} \right]^{c-1} \left(\frac{\mathbf{X}_{n,k} \mathbf{X}_{n,k}^H}{\lambda_{n,k}} \right) \mathbf{w}_k + \alpha \mathbf{a}_k = \mathbf{P}_k \mathbf{w}_k + \alpha \mathbf{a}_k = 0$$

[0062]

수학식 17

$$\mathbf{w}_k = \frac{(\mathbf{P}_k + \delta \mathbf{I})^{-1} \mathbf{a}_k}{\mathbf{a}_k^H (\mathbf{P}_k + \delta \mathbf{I})^{-1} \mathbf{a}_k}$$

[0063]

$$P_k = -c \sum_{n=1}^{L_n} \left[\frac{Y_{n,k} Y_{n,k}^*}{\lambda_{n,k}} \right]^{c-1} \left(\frac{X_{n,k} X_{n,k}^H}{\lambda_{n,k}} \right)$$

[0064] 여기서, 이다.

[0065] P_k 는 complex generalized Gaussian 분포 기반의 우도로부터 계산된 가중치가 적용된 공분산 행렬이다.

[0066] 위의 $\lambda_{n,k}$ 와 w_k 를 반복적으로 추정하여 최종 필터 w_k 를 얻고(단계150), 최종 필터를 이용하여 잡음이 제거된 음원 신호를 검출한다(단계 160).

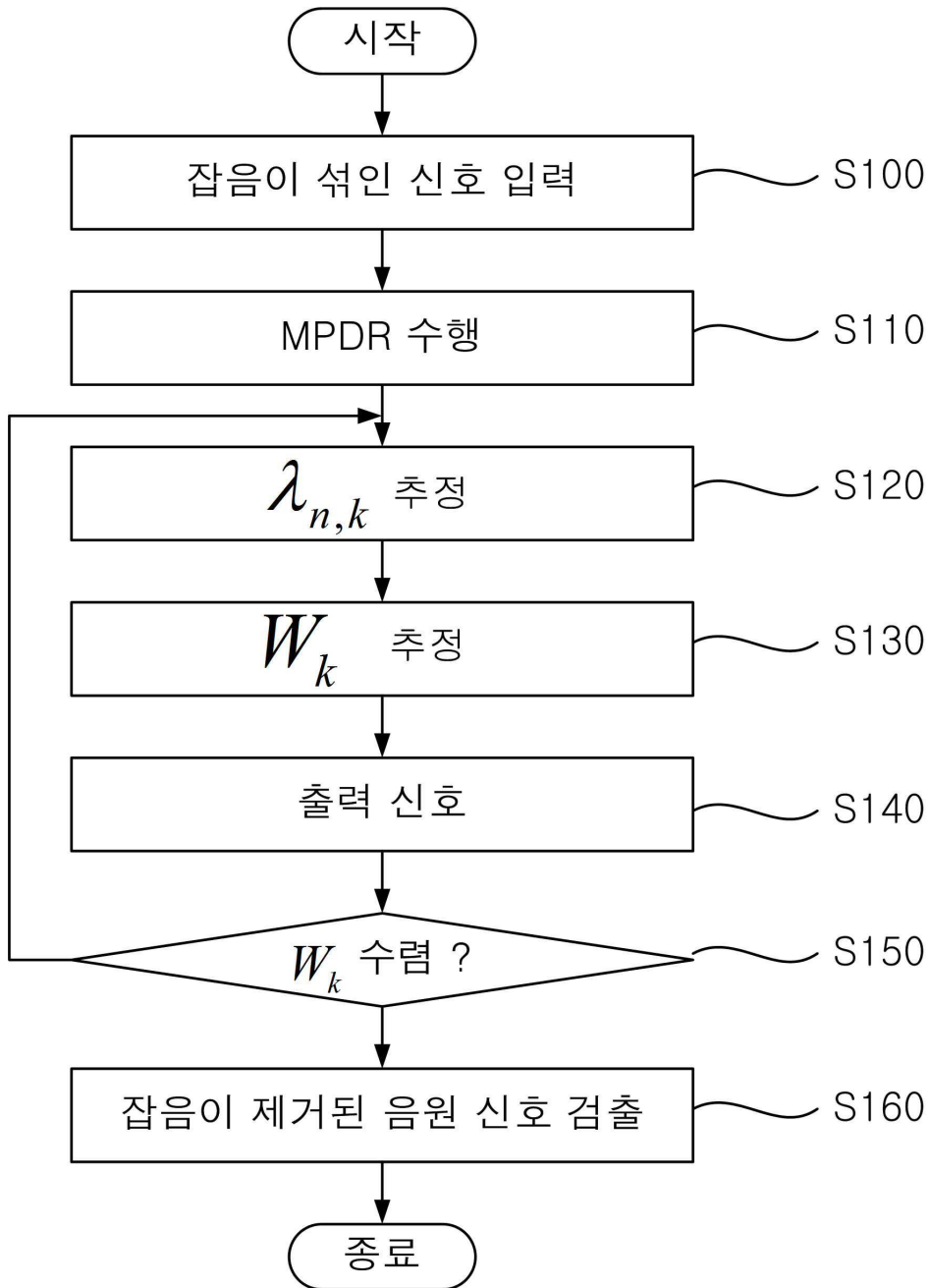
[0067] 이하, 전술한 구성을 갖는 본 발명에 따른 빔포밍 방법에 대한 성능 평가를 위하여, Kaldi 음성 인식 툴을 이용하여 단어 오 인식률을 측정하였다. 음성 인식기는 Reverb Challenge 2014에 사용된 recipe를 이용하여 DNN학습을 하였고, MC-WSJ-AV를 실환경에서 재녹음하여 Signal-to-Noise (SNR) 비율별로 잡음과 혼합하였다. MVDR 알고리즘은 잡음의 추정을 위해 묵음 구간의 프레임에 의하여 공분산 추정을 하였고, 최대 우도 기반 빔포밍 알고리즘에 사용된 반복횟수는 7, β 값은 2이다.

[0068] 도 2는 본 발명의 바람직한 실시예에 따른 우도 최대화를 이용한 빔포밍 방법 및 종래의 방법들을 이용하여 단어 오 인식율을 측정한 결과표를 도시한 도표이다. 도 2를 참조하면, 종래의 MVDR 빔포밍 및 MPDR 빔포밍에 비하여 본 발명에 따라 제안된 빔포밍 방법이 잡음이 없는 경우에는 비슷한 수준을 달성하면서도 그 외 잡음이 있는 경우에는 단어 오 인식률이 낮음을 파악할 수 있다. 따라서, 본 발명에 따른 빔포밍 방법이 잡음 환경에서 더욱 강인한 전처리 방법으로 사용될 수 있음을 알 수 있다.

[0069] 이상에서 본 발명에 대하여 그 바람직한 실시예를 중심으로 설명하였으나, 이는 단지 예시일 뿐 본 발명을 한정하는 것이 아니며, 본 발명이 속하는 분야의 통상의 지식을 가진 자라면 본 발명의 본질적인 특성을 벗어나지 않는 범위에서 이상에 예시되지 않은 여러 가지의 변형과 응용이 가능함을 알 수 있을 것이다. 그리고, 이러한 변형과 응용에 관계된 차이점들은 첨부된 청구 범위에서 규정하는 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.

도면

도면1



도면2

단위 : 단어 오 인식률(%)

SNR(dB)	Baseline	MVDR	MPDR	Proposed
0	83.85	50.44	51.91	40.09
5	72.91	36.18	37.79	29.86
10	58.23	26.03	27.01	22.41
15	42.50	20.11	20.40	18.56
20	30.49	17.25	17.14	16.76
∞	19.71	15.95	15.83	16.29