



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0126318
(43) 공개일자 2013년11월20일

(51) 국제특허분류(Int. Cl.)

G10L 21/02 (2006.01)

(21) 출원번호 10-2012-0050356

(22) 출원일자 2012년05월11일

심사청구일자 없음

(71) 출원인

엘지전자 주식회사

서울특별시 영등포구 여의대로 128 (여의도동)

(72) 발명자

전혜정

서울특별시 서초구 우면동 16 LG전자 전자기술원

박형민

서울특별시 서초구 잠원동 한신아파트 5동 903호

(뒷면에 계속)

(74) 대리인

서교준

전체 청구항 수 : 총 10 항

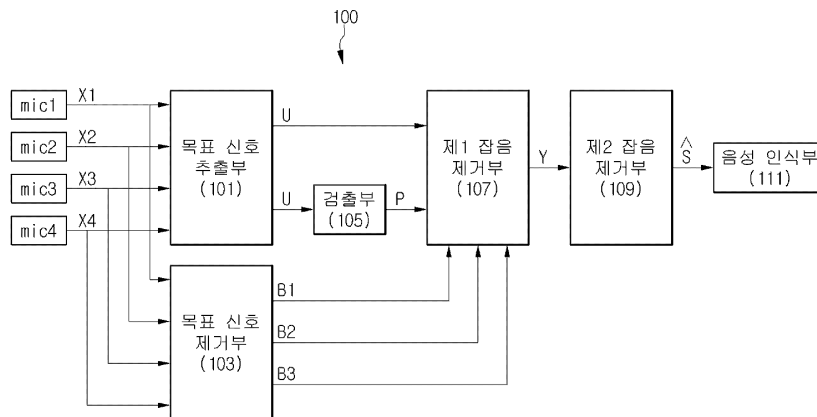
(54) 발명의 명칭 잡음 제거 장치 및 방법

(57) 요약

본 발명은 잡음 제거 장치 및 방법에 관한 것으로, 특히 음성 인식을 위한 잡음 제거 장치 및 방법에 관한 것이다.

본 발명은 입력 신호의 잡음을 제거하는 장치로서, 제1 분리 벡터를 이용하여 상기 입력 신호로부터 제1 목표 신호를 추출하는 목표 신호 추출부, 제2 분리 벡터를 이용하여 상기 입력 신호로부터 제1 잡음 신호를 추출하는 목표 신호 제거부, 상기 제1 목표 신호의 음성 구간 정보를 추출하는 검출부 및 상기 음성 구간 정보를 이용하여 상기 제1 잡음 신호로부터 가중치를 연산하고, 상기 가중치를 이용하여 상기 제1 목표 신호로부터 잡음을 제거하는 제1 잡음 제거부를 포함한다.

대표도



(72) 발명자

이철민

서울특별시 서초구 우면동 16 LG전자 전자기술원

이홍규

서울특별시 서초구 우면동 16 LG전자 전자기술원

특허청구의 범위

청구항 1

입력 신호의 잡음을 제거하는 장치에 있어서,

제1 분리 벡터를 이용하여 상기 입력 신호로부터 제1 목표 신호를 추출하는 목표 신호 추출부;

제2 분리 벡터를 이용하여 상기 입력 신호로부터 제1 잡음 신호를 추출하는 목표 신호 제거부;

상기 제1 목표 신호의 음성 구간 정보를 추출하는 검출부; 및

상기 음성 구간 정보를 이용하여 상기 제1 잡음 신호로부터 가중치를 연산하고, 상기 가중치를 이용하여 상기 제1 목표 신호로부터 잡음을 제거하는 제1 잡음 제거부를 포함하는

잡음 제거 장치.

청구항 2

제1 항에 있어서,

상기 목표 신호 추출부는

독립 벡터 분석 알고리즘을 이용하여 상기 입력 신호로부터 상기 제1 분리 벡터를 결정하는

잡음 제거 장치.

청구항 3

제1 항에 있어서,

상기 목표 신호 제거부는

독립 벡터 분석 알고리즘을 이용하여 상기 입력 신호로부터 상기 제2 분리 벡터를 결정하는

잡음 제거 장치.

청구항 4

제1 항에 있어서,

상기 검출부는

최소 값 제어 재귀 평균 알고리즘을 이용하여 상기 음성 구간 정보를 추출하는

잡음 제거 장치.

청구항 5

제1 항에 있어서,

상기 제1 잡음 제거부는

상기 음성 구간 정보를 이용하여 상기 제1 잡음 신호를 공분산 연산하고, 상기 제1 잡음 신호의 공분산 연산 결과를 이용하여 상기 가중치를 연산하는

잡음 제거 장치.

청구항 6

입력 신호의 잡음을 제거하는 방법에 있어서,

제1 분리 벡터를 이용하여 상기 입력 신호로부터 제1 목표 신호를 추출하는 단계;

제2 분리 벡터를 이용하여 상기 입력 신호로부터 제1 잡음 신호를 추출하는 단계;

상기 제1 목표 신호의 음성 구간 정보를 추출하는 단계;
 상기 음성 구간 정보를 이용하여 상기 제1 잡음 신호로부터 가중치를 연산하는 단계; 및
 상기 가중치를 이용하여 상기 제1 목표 신호로부터 잡음을 제거하는 단계를 포함하는
 잡음 제거 방법.

청구항 7

제6 항에 있어서,
 상기 제1 목표 신호를 추출하는 단계는
 독립 벡터 분석 알고리즘을 이용하여 상기 입력 신호로부터 상기 제1 분리 벡터를 결정하는 단계를 포함하는
 잡음 제거 방법.

청구항 8

제6 항에 있어서,
 상기 제1 잡음 신호를 추출하는 단계는
 독립 벡터 분석 알고리즘을 이용하여 상기 입력 신호로부터 상기 제2 분리 벡터를 결정하는 단계를 포함하는
 잡음 제거 방법.

청구항 9

제6 항에 있어서,
 상기 음성 구간 정보를 추출하는 단계는
 최소 값 제어 재귀 평균 알고리즘을 이용하여 상기 음성 구간 정보를 추출하는 단계를 포함하는
 잡음 제거 방법.

청구항 10

제6 항에 있어서,
 상기 가중치 연산 단계는
 상기 음성 구간 정보를 이용하여 상기 제1 잡음 신호를 공분산 연산하는 단계; 및
 상기 공분산 연산 결과를 이용하여 상기 가중치를 연산하는 단계를 포함하는
 잡음 제거 방법.

명세서

기술분야

[0001] 본 발명은 잡음 제거 장치 및 방법에 관한 것으로, 특히 음성 인식을 위한 잡음 제거 장치 및 방법에 관한 것이다.

배경기술

[0002] 음성 인식 등을 목적으로 음성 신호가 입력되는 장치는 음성 신호에 복수의 잡음 신호가 혼합되어 입력되는 경우, 혼합 신호로부터 원하는 음성 신호를 추출함으로써 음성 인식 기능을 수행한다.

[0003] 예를 들면, 음성 인식을 통하여 IPTV의 동작을 제어하는 경우, 음악 소리 또는 IPTV 자체의 소리 등 다양한 종류의 잡음원이 존재하는 잡음 환경에서, 잡음과 사용자의 음성을 분리하고 음성 신호를 추출하여 음성 신호의 음질을 높임으로써, 잡음이 없는 환경과 유사한 수준의 음성 인식 성능을 보장할 수 있다.

[0004] 이에 따라, 신호 분리 기술로서 빔포밍(beamforming) 기술이 사용된다. 대표적인 빔포밍 기술은 입력된 혼합 신호 중에서 특정 방향으로 입력된 신호를 다른 방향으로 입력된 신호보다 강화하는 방법 및 강화된 특정 방향 신호만을 제거하는 방법을 함께 사용할 수 있다. 이어서, 특정 방향 신호가 제거된 신호를 이용하여 특정 방향 신호의 잡음을 제거함으로써 신호 분리 기능을 수행할 수 있다.

[0005] 다만, 종래의 빔포밍 기술에 의하면, 고정된 음원으로부터 발생된 음성 신호가 입력되는 경우에 한하여 적절한 신호 분리 동작을 수행할 수 있다. 따라서, 음원이 이동하는 경우, 음원의 위치 정보를 얻기 위하여 음원 추적 장치 또는 음원 추적 방법이 추가적으로 요구된다.

발명의 내용

해결하려는 과제

[0006] 본 발명의 실시 예는 음원이 이동하는 경우에도 효과적으로 잡음을 제거하는 잡음 제거 장치를 제공한다.

과제의 해결 수단

[0007] 본 발명의 일 실시 예에 따른 입력 신호의 잡음을 제거하는 장치는 제1 분리 벡터를 이용하여 상기 입력 신호로부터 제1 목표 신호를 추출하는 목표 신호 추출부, 제2 분리 벡터를 이용하여 상기 입력 신호로부터 제1 잡음 신호를 추출하는 목표 신호 제거부, 상기 제1 목표 신호의 음성 구간 정보를 추출하는 검출부 및 상기 음성 구간 정보를 이용하여 상기 제1 잡음 신호로부터 가중치를 연산하고, 상기 가중치를 이용하여 상기 제1 목표 신호로부터 잡음을 제거하는 제1 잡음 제거부를 포함한다.

[0008] 본 발명의 일 실시 예에 따른 잡음 제거 방법은 제1 분리 벡터를 이용하여 상기 입력 신호로부터 제1 목표 신호를 추출하는 단계, 제2 분리 벡터를 이용하여 상기 입력 신호로부터 제1 잡음 신호를 추출하는 단계, 상기 제1 목표 신호의 음성 구간 정보를 추출하는 단계, 상기 음성 구간 정보를 이용하여 상기 제1 잡음 신호로부터 가중치를 연산하는 단계 및 상기 가중치를 이용하여 상기 제1 목표 신호로부터 잡음을 제거하는 단계를 포함한다.

발명의 효과

[0009] 본 발명의 실시 예에 따른 잡음 제거 장치는 음원이 이동하는 경우, 추가적인 음원 추적 장치 없이도 효과적으로 잡음을 제거한다.

도면의 간단한 설명

[0010] 도 1은 본 발명의 일 실시 예와 관련된 잡음 제거 장치의 블록 다이어그램(block diagram)이다.

도 2는 본 발명의 일 실시 예와 관련된 잡음 제거 방법의 흐름도이다.

도 3 내지 도 5는 본 발명의 일 실시 예와 관련된 잡음 제거 장치의 효과를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0011] 이하, 본 발명과 관련된 잡음 제거 장치에 대하여 도면을 참조하여 보다 상세하게 설명한다. 이하의 설명에서 사용되는 구성요소에 대한 접미사 "모듈" 및 "부"는 명세서 작성의 용이함만이 고려되어 부여되거나 혼용되는 것으로서, 그 자체로 서로 구별되는 의미 또는 역할을 갖는 것은 아니다.

[0012] 다음은 도 1을 참조하여 본 발명의 일 실시 예에 따른 잡음 제거 장치의 구조를 설명한다.

[0013] 도 1은 본 발명의 일 실시 예와 관련된 잡음 제거 장치의 블록 다이어그램(block diagram)이다.

[0014] 본 발명의 일 실시 예에 따른 잡음 제거 장치는 복수의 마이크로폰(mic1, mic2, mic3, mic4), 목표 신호 추출부(101), 목표 신호 제거부(103), 검출부(105), 제1 잡음 제거부(107)를 포함할 수 있다. 추가적으로, 제2 잡음 제거부(109)를 포함하여 보다 깨끗한 음성 신호를 추출할 수 있다. 이어서, 음성 인식부(111)는 제1 잡음 제거부(107) 또는 제2 잡음 제거부(109)로부터 추출된 음성 신호를 인식할 수 있다.

[0015] 복수의 마이크로폰(mic1, mic2, mic3, mic4)은 혼합 신호(미도시됨)를 수신한다. 혼합 신호는 복수의 음원으로부터 생성된 신호를 포함할 수 있다. 혼합 신호는 예컨대, IPTV를 제어하기 위한 음성 신호, 전자 기기 사용시 발생

하는 소리와 같은 일상 생활에서 발생할 수 있는 모든 신호를 포함할 수 있다.

[0016] 기기 음향 제거부(미도시됨)는 혼합 신호에서 기기 음향을 제거한다. 예컨대, IPTV의 마이크론이 혼합 신호를 수신할 때, 혼합 신호는 IPTV 자체에서 발생하는 소리를 포함할 수 있다. 기기 음향 제거부(미도시됨)가 혼합 신호로부터 IPTV 자체에서 발생하는 소리를 제거함으로써, 혼합 신호 중 음향 크기가 가장 큰 잡음을 제거할 수 있고, 따라서 음성 인식의 성능을 향상시킬 수 있다. 결과적으로, 입력 신호(X1,X2,X3,X4)는 혼합 신호에서 기기 음향이 제거된 신호를 포함할 수 있다.

[0017] 목표 신호 추출부(101)는 복수의 입력 신호(X1,X2,X3,X4)를 입력받아, 제1 목표 신호(U)를 출력한다. 이때, 목표 신호 추출부(101)는 복수의 입력 신호(X1,X2,X3,X4)를 입력받아, 그 중 음성 인식의 대상이 되는 하나의 음성 신호를 분리하는 제1 분리 벡터(w1)를 결정하고, 제1 분리 벡터(w1)를 이용하여 제1 목표 신호(U)를 추출할 수 있다. 목표 신호 추출부(101)는 독립 벡터 분석(IVA, Independent Vector Analysis) 알고리즘을 이용하여 제1 분리 벡터(w1)를 결정할 수 있다. 독립 벡터 분석 알고리즘은 독립 성분 분석(Independent component analysis)을 주파수 영역으로 확장한 개념이다.

[0018] 수학식 1은 독립 성분 분석 문제를 일반적인 모델로 표현한 것이다.

수학식 1

$$x = As$$

$$y = \hat{s} = Wx = WAs \approx A^{-1} As$$

[0019]

[0020] x는 입력 벡터, A는 혼합 행렬, s는 음원 벡터, y는 출력 벡터, W는 A의 의사 역행렬(Pseudo-inverse Matrix)을 의미한다. 특정 음원으로부터 발생한 신호를 추출하기 위하여 의사 역행렬(W)를 학습하는 것이 독립 성분 분석 문제가 된다. 의사 역행렬(W) 학습시, 비용 함수가 이용될 수 있다. 예컨대, 비용 함수가 최소화될수록, 음원 벡터(s)와 최대한 비슷한 출력 벡터(y)가 추출될 수 있다.

[0021] 독립 벡터 분석 알고리즘은 주파수 영역으로 확장된 독립 성분 분석 알고리즘이므로, 독립 벡터 분석 문제는 제1 분리 벡터(w1)를 결정하는 것이다.

[0022] 목표 신호 추출부(101)는 제1 분리 벡터(w1)를 결정하기 위한 전처리로서, 입력 벡터(x)를 백색화(whitening) 연산할 수 있다. 백색화 연산을 실행함으로써 복수의 입력 신호(X1,X2,X3,X4)들의 상관 관계(correlation)를 제거할 수 있다. 수학식 5는 입력 벡터(x)의 백색화 연산 과정을 표현한 것이다.

수학식 5

$$z = Vx = D^{-1/2} E^T x$$

[0023]

[0024] 복수의 입력 신호(X1,X2,X3,X4)의 공분산(covariance) 행렬에서, 공분산 행렬의 고유 값들의 대각 행렬을 D라

$$D^{-\frac{1}{2}} = \text{diag}\left(\frac{1}{\sqrt{\lambda_i}}\right), i=1,2,\dots,p$$

하고, 공분산 행렬의 고유 벡터들의 모음 행렬을 E라 한다. 이때,

의미하고, E^T 는 E 행렬의 전치 행렬을 의미한다.

[0025] 수학식 11은 수학식 5의 백색화된 입력 벡터(z)를 이용한 비용 함수(J_g)를 나타낸 것이다.

수학식 11

$$J_g(w_1^k) = E\{g(\sum_k |(w_1^k)^H z^k|^2)\}$$

[0026]

[0027]

k는 주파수 인덱스, E는 기대 값(expectation), g는 비선형 함수로서 예컨대, 첨도(kurtosis) 기반일 때 $g(y)=y^4$ 로 근사화될 수 있는 비선형 함수, H은 헤르미티안 행렬(Hermitian Matrix)을 의미한다. 수학식 11은 독립 벡터 분석에서 벡터 상관도(vector correlation), 벡터 상호 정보(vector mutual information)에 따른 비용 함수를 나타낸 것이다.

[0028]

수학식 12는 수학식 11의 비용 함수를 미분한 결과로서, 제1 분리 벡터(w1)의 최대경사법 알고리즘에 포함되는 학습 규칙을 의미한다.

수학식 12

$$\begin{aligned} \triangle w_1 &\propto -E\{z^k [(w_1^k)^H z^k]^* g(\sum_k |(w_1^k)^H z^k|^2)\} \\ w_1 &\leftarrow w_1 / \|w_1\| \end{aligned}$$

[0029]

[0030]

*는 공액 행렬(Conjugate Matrix)을 의미한다.

[0031]

수학식 13은 수학식 12의 1차 미분, 2차 미분을 이용한 결과로서, 뉴턴 알고리즘에 포함되는 고속 독립 벡터 분석 학습 규칙을 의미한다.

수학식 13

$$\begin{aligned} w_1^k &\leftarrow E\{z^k ((w_1^k)^H z^k)^* g(\sum_k |(w_1^k)^H z^k|^2)\} - E\{g(\sum_k |(w_1^k)^H z^k|^2) + |(w_1^k)^H z^k|^2 g'(\sum_k |(w_1^k)^H z^k|^2)\} w_1^k \\ w_1 &\leftarrow w_1 / \|w_1\| \end{aligned}$$

[0032]

[0033]

$(w_1^k)^H z^k$ 는 제1 목표 신호(U)를 의미한다. 목표 신호 추출부(101)는 수학식 13으로부터 제1 분리 벡터(w1)를 결정할 수 있고, 제1 분리 벡터(w1)를 이용하여 복수의 입력 신호(X1,X2,X3,X4)로부터 제1 목표 신호(U)를 추출할 수 있다.

[0034]

목표 신호 제거부(103)는 복수의 입력 신호(X1,X2,X3,X4)를 입력받아, 복수의 제1 잡음 신호(B1,B2,B3)를 출력한다. 즉, 목표 신호 제거부(103)는 목표 신호 추출부(101)와 반대로, 복수의 입력 신호(X1,X2,X3,X4)를 입력받아, 그 중 하나의 음성 신호를 제외한 나머지 음향 신호를 분리하는 제2 분리 벡터(w2)를 결정하고, 제2 분리 벡터(w2)를 이용하여 제1 잡음 신호(B1,B2,B3)를 추출할 수 있다. 목표 신호 제거부(103)는 하나의 출력 신호를 제거하기 위하여 목표 신호 추출부(101)가 이용한 독립 벡터 분석 알고리즘을 이용할 수 있다. 다만, 목표 신호 제거부(103)는 복수의 입력 신호를 분리한 후, 그 중에서 하나의 출력 신호를 제거해야 하기 때문에 수학식 12에서 표현된 학습 규칙과 반대 부호를 가지는 학습 규칙을 이용한다. 따라서, 목표 신호 제거부(103)가 이용하는 최대경사법 알고리즘에 포함되는 학습 규칙은 수학식 14와 같다.

수학식 14

$$\triangle w_2 \propto E\{z^k [(w_2^k)^H z^k]^* g(\sum_k |(w_2^k)^H z^k|^2)\}$$

$$w_2 \leftarrow w_2 / \|w_2\|$$

[0035]

[0036] z 는 입력 벡터(x)의 백색화 연산 결과, $(w_2^k)^H z^k$ 는 제1 잡음 신호($B1, B2, B3$)를 의미한다. 목표 신호 제거부(103)는 수학식 14를 이용하여 제2 분리 벡터(w_2)를 결정할 수 있다. 이어서, 목표 신호 제거부(103)는 제2 분리 벡터(w_2)를 이용하여 복수의 입력 신호($X1, X2, X3, X4$)로부터 제1 잡음 신호($B1, B2, B3$)를 추출할 수 있다.

[0037] 목표 신호 제거부(103)는 복수의 잡음으로부터 목표 신호를 제외하기 위하여 예컨대, 복수의 입력 신호 중 두 개의 입력 신호를 선택하고, 선택된 두 개의 입력 신호 중에서 목표 신호를 제외하는 방법, 복수의 입력 신호 전체에서 목표 신호를 직접 제외하는 방법 등을 이용할 수 있다.

[0038] 이와 같이, 본 발명의 일 실시 예에 따라 목표 신호 추출부(101) 및 목표 신호 제거부(103)가 독립 벡터 분석 알고리즘을 이용함으로써, 음원이 이동하는 경우에도 효과적으로 음성을 인식할 수 있다.

[0039] 검출부(105)는 제1 목표 신호(U)의 각 프레임마다 음성 신호가 존재하는 확률 정보(P)를 출력한다. 목표 신호 추출부(101)가 출력한 제1 목표 신호(U)는 단일 음성 신호뿐만 아니라, 그 외의 잡음도 포함하고 있다. 검출부(105)는 제1 목표 신호(U)에 포함된 잡음을 제거하기 위한 전처리로서, 제1 목표 신호(U)의 각각의 프레임에 음성 신호 존재 여부를 검출하여 검출 결과를 확률 정보(P)로 출력할 수 있다. 검출부(105)는 제1 목표 신호(U)의 음성 존재 확률 정보(P)를 출력하기 위하여 최소 값 제어 재귀 평균(MCRA, Minima Controlled Recursive Averaging) 알고리즘을 이용할 수 있다.

[0040] 먼저, 제1 목표 신호(U)는 단구간 푸리에 변환(short-time Fourier Transform)을 통해 주파수 영역에서 표시될 수 있다. 주파수 영역의 제1 목표 신호($U(k, l)$)에서, k 는 주파수 인덱스, l 은 프레임 인덱스를 의미한다. 프레임 인덱스(l)는 신호 전체를 복수의 단 구간으로 나누어 분석하는 경우, 신호를 나누는 각각의 단 구간을 표시한 것이다. 이때, 단 구간 푸리에 변환은 목표 신호 추출부(101), 목표 신호 제거부(103) 단계에서 실행될 수 있다.

[0041] 수학식 23은 주파수 영역의 제1 목표 신호($U(k, l)$)의 에너지 스펙트럼($P(k, l)$)을 추출하는 연산을 나타내는 식이다.

수학식 23

$$P(k, l) = \alpha_s P(k, l-1) + (1 - \alpha_s) |U(k, l)|^2$$

[0042]

[0043] α_s 는 0보다 크고 1보다 작은 값을 가지는 스무딩 매개 변수를 의미한다. 검출부(105)는 수학식 23을 이용하여 주파수 영역의 제1 목표 신호($U(k, l)$)의 스무딩 연산 결과인 에너지 스펙트럼($P(k, l)$)을 출력할 수 있다. 수학식 23은 과거의 에너지 스펙트럼($P(k, l-1)$)을 입력받아 현재의 에너지 스펙트럼($P(k, l)$)을 출력하는 재귀 평균 연산을 나타낸다.

[0044] 수학식 24는 수학식 23의 에너지 스펙트럼($P(k, l)$)을 이용하여 각각의 주파수마다 에너지 스펙트럼의 최소값을 결정하는 연산을 나타내는 식이다.

수학식 24

$$P_{\min}(k, l) = \min \{ P_{\min}(k, l-1), P(k, l) \}$$

$$P_{\text{tmp}}(k, l) = \min \{ P_{\text{tmp}}(k, l-1), P(k, l) \}$$

[0045]

[0046]

$P_{\text{tmp}}(k, l)$ 은 에너지 스펙트럼의 최소 값($P_{\min}(k, l)$)을 결정하기 위한 에너지 스펙트럼의 임시 값을 의미한다. 검출부(105)는 현재 에너지 스펙트럼 성분($P(k, l)$)과 과거 에너지 스펙트럼 성분의 최소 값($P_{\min}(k, l-1)$)을 비교한 결과를 이용하여 각각의 주파수에서 에너지 스펙트럼의 최소 값($P_{\min}(k, l)$)을 결정할 수 있다.

[0047]

수학식 25는 수학식 23 및 수학식 24의 결과를 이용하여 현재 에너지 스펙트럼의 최소 값($P_{\min}(k, l)$) 및 임시 값($P_{\text{tmp}}(k, l)$)을 갱신하는 연산을 나타내는 식이다.

수학식 25

$$P_{\min}(k, l) = \min \{ P_{\text{tmp}}(k, l-1), P(k, l) \}$$

$$P_{\text{tmp}}(k, l) = P(k, l),$$

[0048]

[0049]

검출부(105)는 수학식 25의 연산을 이용하여 현재 에너지 스펙트럼의 최소 값($P_{\min}(k, l)$) 및 임시 값($P_{\text{tmp}}(k, l)$)을 갱신할 수 있다.

[0050]

수학식 26은 수학식 25의 결과를 이용하여, 에너지 스펙트럼의 최소 값($P_{\min}(k, l)$)과 현재 에너지 스펙트럼 성분($P(k, l)$)의 비율($S_r(k, l)$)을 연산하는 식이다.

수학식 26

$$S_r(k, l) = P(k, l) / P_{\min}(k, l)$$

[0051]

[0052]

검출부(105)는 수학식 26의 연산을 이용하여 각 주파수에서 에너지 스펙트럼의 최소 값($P_{\min}(k, l)$)과 현재 에너지 스펙트럼 성분($P(k, l)$)의 비율($S_r(k, l)$)을 구할 수 있다.

[0053]

수학식 27은 수학식 26 및 음성 존재 확률 인덱스($I(k, l)$)를 이용하여 음성 존재 확률($\hat{p}'(k, l)$)을 연산하는 식이다.

수학식 27

$$\hat{p}'_r(k, l) = \alpha_s \hat{p}'(k, l-1) + (1 - \alpha_s) I(k, l)$$

[0054]

[0055]

검출부(105)는 수학식 27의 연산을 이용하여 음성 존재 확률($\hat{p}'(k, l)$)을 출력할 수 있다. 이때, 음성 존재

확률 인덱스($I(k,1)$)는 수학식 26의 결과인 $S_r(k,1)$ 이 특정 문턱 값(δ)보다 작은 경우 0이 되고, $S_r(k,1)$ 이 특정 문턱 값(δ)보다 큰 경우 1이 된다. 이와 같이, 검출부(105)는 수학식 23 내지 수학식 27에 표현된 연산을 포함하는 최소값 제어 재귀 평균 알고리즘을 이용하여, 제1 목표 신호(U)의 각 프레임이 음성이 존재하는 구간인지 또는 음성이 존재하지 않는 구간인지 여부를 출력할 수 있다.

[0056] 제1 잡음 제거부(107)는 제1 잡음 신호($B1, B2, B3$)를 이용하여 제1 목표 신호(U)의 잔여 잡음 신호를 추가적으로 제거한 결과인 제2 목표 신호(Y)를 추출한다. 이때, 제1 목표 신호(U)에 포함된 제2 목표 신호(Y)가 왜곡 없이 출력될수록, 잔여 잡음 신호가 최소화되어 출력될수록, 보다 효과적인 음성 인식이 이루어질 수 있다. 한편, 제1 잡음 신호($B1, B2, B3$)에는 제거되지 못한 잔여 음성 신호가 존재한다. 이때, 제1 잡음 신호($B1, B2, B3$)에 포함된 잔여 음성 신호가 최소화될수록 제2 목표 신호(Y)가 왜곡 없이 출력될 수 있다. 제1 잡음 제거부(107)는 제1 목표 신호(U)의 잔여 잡음 신호를 제거하기 위하여, 위너 필터(Wiener Filter)를 이용할 수 있다.

[0057] 먼저, 제1 잡음 제거부(107)는 수학식 28에 나타난 바와 같이, 제1 잡음 신호($B1, B2, B3$)의 벡터(b)를 잔여 음성 신호가 존재하는 구간 및 잔여 음성 신호가 존재하지 않는 구간으로 구분하여 정의할 수 있다.

수학식 28

잔여 음성 신호가 존재하는 구간 : $b^k = s^k + v^k$

잔여 음성 신호가 존재하지 않는 구간 : $b^k = v^k$

[0058]

[0059] b 는 제1 잡음 신호 벡터, s 는 제1 잡음 신호에 포함된 잔여 음성 신호 벡터, v 는 제1 잡음 신호에 포함된 제2 잡음 신호 벡터를 의미한다.

[0060] 한편, 제1 잡음 제거부(107)는 수학식 29에 나타난 연산을 통해 제2 목표 신호(y)를 추출할 수 있다. 수학식 29는 제1 목표 신호(u), 제1 잡음 신호(b), 위너 필터의 가중치 파라미터(w)를 이용하여 제2 목표 신호(y)를 연산한 식을 나타낸다.

수학식 29

$$y^k = u^k - (w^k)^H b^k$$

$$J_k = \frac{1}{\mu} E\{|(w^k)^H s^k|^2\} + E\{|u^k - (w^k)^H v^k|^2\}$$

[0061]

[0062] u 는 제1 목표 신호, b 는 제1 잡음 신호 벡터, w 는 위너 필터의 가중치 파라미터, y 는 제2 목표 신호, J_k 는 비용 함수를 의미한다. 비용 함수(J_k)는 제1 잡음 신호(b)에 포함된 잔여 음성 신호 벡터를 최소화하기 위한 계수인 μ 를 포함할 수 있다.

[0063] 한편, 비용 함수(J_k)가 최소가 되어야, 제2 목표 신호(y)가 최적화될 수 있다. 다시 말하면, 비용 함수(J_k)가 최소일 때, 제2 목표 신호(y)에서 음성이 존재하지 않는 부분이 최소가 될 수 있다. 비용 함수(J_k)를 최소화하기 위한 위너 필터의 가중치 파라미터(w)는 수학식 30과 같다.

수학식 30

$$w^k = \left(\frac{1}{\mu} E\{s^k(s^k)^H\} + E\{v^k(v^k)^H\} \right)^{-1} E\{(u^k)^* v^k\}$$

[0064]

[0065]

이때, 위너 필터의 가중치 파라미터(w)를 연산하기 위한 수학식 30에 수학식 31의 결과를 대입할 수 있다.

수학식 31

$$\text{잔여 음성 신호가 존재하는 구간} : \hat{E}\{s^k(s^k)^H\} = E\{b^k(b^k)^H\} - \hat{E}\{v^k(v^k)^H\}$$

$$\text{잔여 음성 신호가 존재하지 않는 구간} : \hat{E}\{v^k(v^k)^H\} = E\{b^k(b^k)^H\}$$

[0066]

[0067]

수학식 31은 수학식 28의 정의를 이용한 결과로서, 잔여 음성 신호가 존재하는 구간의 제1 잡음 신호의 공분산과 잔여 음성 신호가 존재하지 않는 구간의 제1 잡음 신호의 공분산을 의미한다. 이와 같이, 잔여 음성 신호의 존재 여부에 따른 공분산 연산 결과를 위너 필터의 가중치 파라미터(w) 연산에 적용함으로써, 음성 부존재 구간에서 제1 목표 신호의 에너지를 최소화할 수 있다. 즉, 음성 부존재 구간에서 제1 목표 신호의 에너지가 최소화됨으로써, 제1 목표 신호의 잔여 잡음 신호가 최소화될 수 있다.

[0068]

제2 잡음 제거부(109)는 제2 목표 신호(Y)를 수신한 후, 제2 목표 신호(Y)에 남아 있는 잡음을 추가로 제거하여

인식 대상인 음성 신호와 가장 비슷한 제3 목표 신호($\hat{S}$)를 출력한다. 제2 잡음 제거부(109)는 제3 목표 신

호($\hat{S}$)를 출력하기 위하여 제2 목표 신호(Y)에 최소값 제어 재귀 평균 알고리즘을 적용하여 음성 존재 확률을 출력할 수 있다. 또한, 제2 잡음 제거부(109)는 제2 목표 신호(Y)에 최소 평균 제곱 오차(Minimum-mean square error) 알고리즘을 적용하여 제2 목표 신호(Y)의 잡음을 제거할 수 있다. 또한, 제2 잡음 제거부(109)는 제2 목표 신호(Y)에 끝점 검출(end-point detection) 알고리즘을 적용하여 제2 목표 신호(Y)에 포함된 음성 신

호 즉, 제3 목표 신호($\hat{S}$)의 시작점 및 종료점을 검출할 수 있다.

[0069]

음성 인식부(111)는 제2 잡음 제거부(109)에서 출력된 제3 목표 신호($\hat{S}$)를 수신하여, 사용자가 IPTV 등을

제어하기 위하여 전송한 음성 신호를 인식할 수 있다. 예컨대, 음성 인식부(111)는 제3 목표 신호($\hat{S}$)에 HMM(Hidden Markov Model) 알고리즘을 적용함으로써, 복수의 마이크로폰(mic1, mic2, mic3, mic4)이 수신한 혼합 신호 중 IPTV 제어를 위한 음성 신호를 인식할 수 있다.

[0070]

다음은 도 2를 참조하여 본 발명의 일 실시 예에 따른 잡음 제거 방법을 설명한다.

[0071]

도 2는 본 발명의 일 실시 예와 관련된 잡음 제거 방법의 흐름도이다.

[0072]

복수의 마이크로폰(mic1, mic2, mic3, mic4)은 혼합 신호를 수신한다(S201). 음성 인식을 이용한 IPTV 작동 제어를 위하여, IPTV는 음성 신호를 수신하는 마이크로폰을 포함한다. IPTV는 효과적인 음성 인식을 위하여 적어도 하나 이상의 마이크로폰을 포함할 수 있다. 혼합 신호는 IPTV 제어를 위한 음성 신호뿐만 아니라, 일상 생활에서 발생하는 모든 종류의 잡음 신호를 포함할 수 있다.

[0073]

기기 음향 제거부(미도시됨)는 혼합 신호에서 기기 음향을 제거하여 입력 신호(X1, X2, X3, X4)를 생성한다(S203).

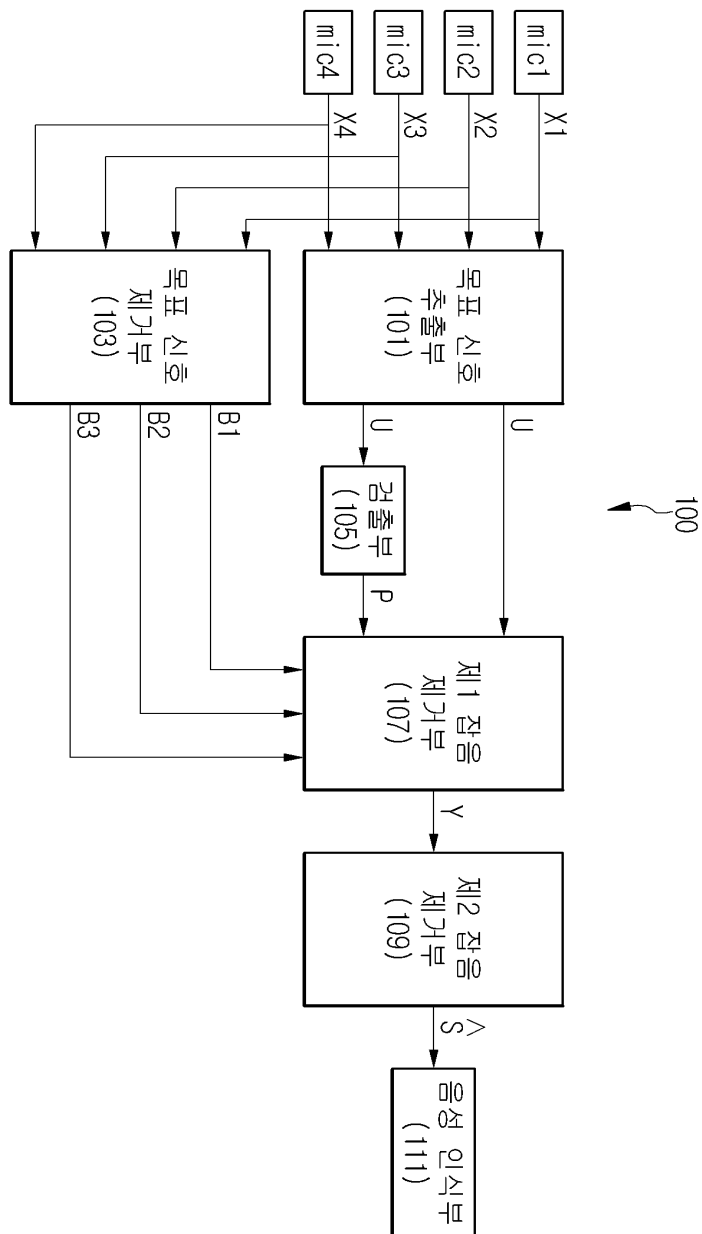
예컨대, IPTV 작동 제어를 위한 음성 인식의 경우, 기기 음향은 IPTV 작동시 발생하는 소리를 의미한다.

- [0074] 목표 신호 추출부(101)는 복수의 입력 신호(X_1, X_2, X_3, X_4)를 수신하여, 입력 신호(X_1, X_2, X_3, X_4)를 백색화하는 연산을 실행한다(S205). 수학식 5를 참조하면, 목표 신호 추출부(101)는 백색화 연산을 통하여 복수의 입력 신호(X_1, X_2, X_3, X_4)들의 상관 관계를 제거함으로써, 제1 목표 신호(U)를 보다 효과적으로 복수의 입력 신호(X_1, X_2, X_3, X_4)로부터 분리할 수 있다. 제1 목표 신호(U)는 예컨대, IPTV 작동 제어를 위하여 인식되어야 하는 음성 신호를 포함할 수 있다.
- [0075] 이어서, 목표 신호 추출부(101)는 독립 벡터 분석 학습 규칙을 이용하여 백색화된 입력 신호로부터 제1 분리 벡터(w_1)를 결정한다(S207). 독립 벡터 분석 학습 규칙은 독립 성분 분석 문제를 주파수 영역으로 확장한 개념으로써, 목표 신호 추출부(101)는 수학식 11의 벡터 상관도 및 벡터 상호 정보에 따른 비용 함수, 수학식 12의 최대경사법 알고리즘, 수학식 13의 뉴턴 알고리즘을 이용하여, 제1 분리 벡터(w_1)를 결정할 수 있다.
- [0076] 이어서, 목표 신호 추출부(101)는 제1 분리 벡터(w_1)를 이용하여 입력 신호(X_1, X_2, X_3, X_4)로부터 제1 목표 신호(U)를 추출한다(S209).
- [0077] 목표 신호 제거부(103)는 복수의 입력 신호(X_1, X_2, X_3, X_4)를 수신하여, 입력 신호(X_1, X_2, X_3, X_4)를 백색화하는 연산을 실행한다(S205).
- [0078] 이어서, 목표 신호 제거부(103)는 독립 벡터 분석 학습 규칙을 이용하여 백색화된 입력 신호로부터 제2 분리 벡터(w_2)를 결정한다(S211). 독립 벡터 분석 학습 규칙은 도면 부호 S207 단계에서 설명한 바와 동일하므로, 이하 생략한다.
- [0079] 이어서, 목표 신호 제거부(103)는 제2 분리 벡터(w_2)를 이용하여 입력 신호(X_1, X_2, X_3, X_4)로부터 제1 잡음 신호(B_1, B_2, B_3)를 추출한다(S213). 목표 신호 제거부(103)는 목표 신호 추출부(101)와 반대로 예컨대, IPTV 작동 제어를 위하여 인식되어야 하는 음성 신호를 제외한 나머지 잡음을 추출함으로써, 목표 신호를 제거할 수 있다. 따라서, 목표 신호 제거부(103)는 수학식 11, 목표 신호 추출부(101)에서 이용하는 수학식 12의 최대경사법 알고리즘과 부호만 반대인 수학식 14, 수학식 13을 이용하여, 제2 분리 벡터(w_2)를 결정할 수 있다.
- [0080] 검출부(105)는 최소 값 제어 재귀 평균 알고리즘을 이용하여 제1 목표 신호(U)로부터 제1 목표 신호(U)의 음성 존재 확률 정보(P)를 출력한다(S215). 검출부(105)는 수학식 23의 에너지 스펙트럼($P(k, l)$) 추출 연산, 수학식 24의 최소 에너지 스펙트럼($P_{min}(k, l)$) 결정 연산, 수학식 25의 에너지 스펙트럼 갱신 연산, 수학식 26의 현재 에너지 스펙트럼 성분($P(k, l)$)에 대한 최소 에너지 스펙트럼($P_{min}(k, l)$)의 비율 연산, 수학식 27의 음성 존재 확률($\hat{p}'(k, l)$) 연산을 이용하여, 제1 목표 신호(U)의 음성 존재 확률 정보(P)를 출력할 수 있다.
- [0081] 제1 잡음 제거부(107)는 제1 잡음 신호 벡터(b)를 수학식 28과 같이 잔여 음성 신호의 존재 또는 부존재 구간으로 구분하고, 수학식 31을 이용하여 잔여 음성 신호가 존재하는 구간에서 제1 잡음 신호 벡터(b)의 공분산을 계산하고(S217), 잔여 음성 신호가 존재하지 않는 구간에서 제1 잡음 신호 벡터(b)의 공분산을 계산한다(S219).
- [0082] 제1 잡음 제거부(107)는 제1 잡음 신호 벡터(b)의 공분산 연산 결과를 이용하여 음성 부존재 구간에서 제1 목표 신호(u)의 에너지가 최소가 되는 가중치 파라미터(w)를 결정한다(S221). 제1 잡음 제거부(107)는 가중치 파라미터(w)를 연산하기 위하여 수학식 30을 이용할 수 있다.
- [0083] 제1 잡음 제거부(107)는 제1 목표 신호(u)로부터 제1 잡음 신호(b)에 가중치 파라미터(w)를 연산한 신호를 제거한 결과, 제2 목표 신호(y)를 생성한다(S223). 제1 잡음 제거부(107)는 제2 목표 신호(y)를 생성하기 위하여 수학식 29를 이용할 수 있다. 이때, 제1 목표 신호(u)에서 제1 잡음 신호(b)에 가중치 파라미터(w)를 연산한 신호가 제거되면, 제1 목표 신호(u)에 포함된 음성 신호의 왜곡은 최소화되고, 제1 목표 신호(u)에 포함된 잡음은 최소화될 수 있으므로, 음성 인식부(111)에 보다 깨끗한 음성 신호를 전송할 수 있다.
- [0084] 참고로, 제2 목표 신호(y)에 포함된 잡음을 추가적으로 제거하기 위하여, 본 발명의 일 실시 예에 따른 잡음 제거 장치에 제2 잡음 제거부(109)가 부가될 수 있다. 제2 잡음 제거부(109)는 최소 값 제어 재귀 평균 알고리즘 및 최소 평균 제곱 오차 알고리즘을 이용하여 제2 목표 신호(y)의 잡음을 제거한 제3 목표 신호($\hat{S}$)를 생성한다(S225).

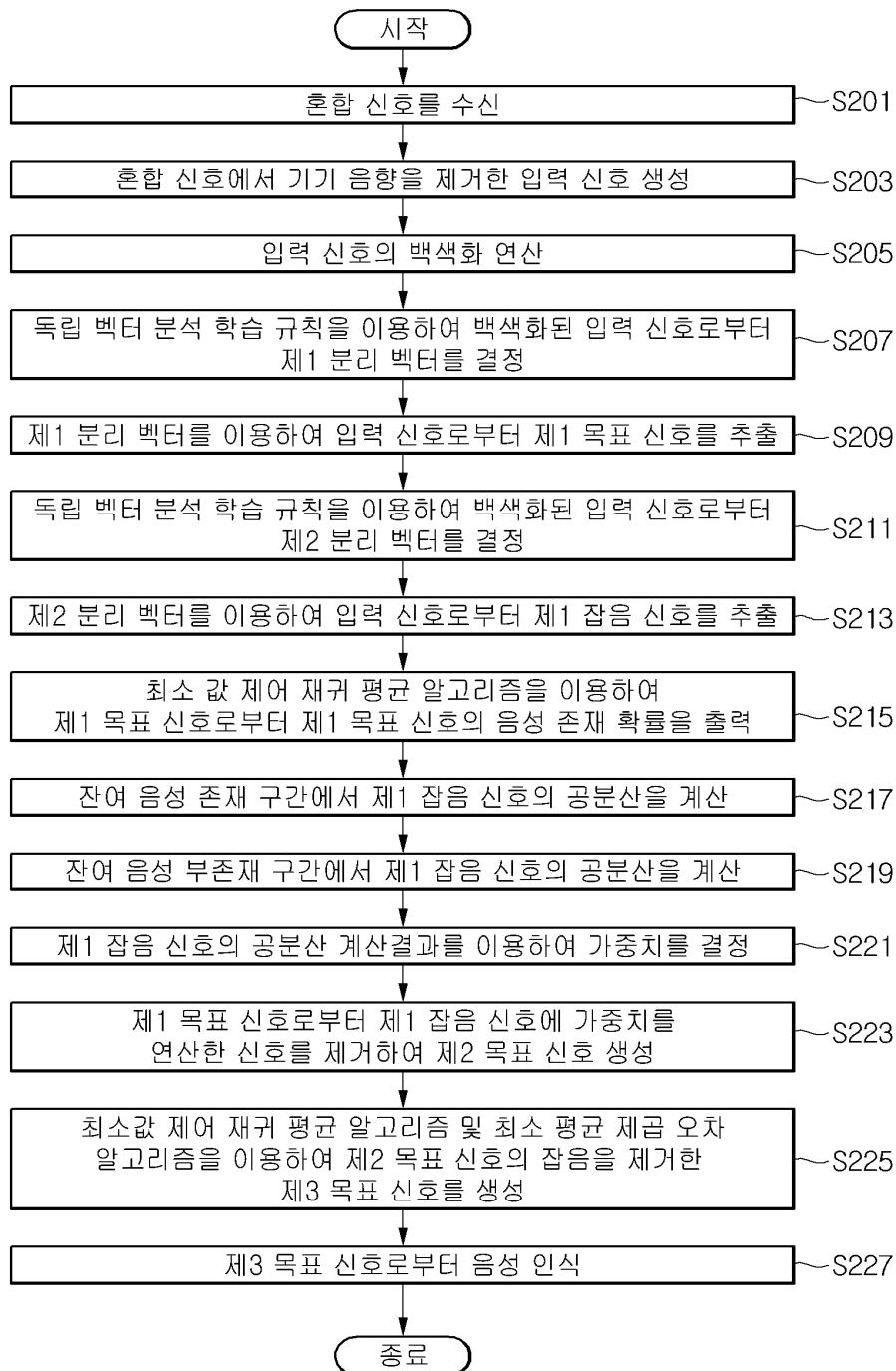
- [0085] 음성 인식부(111)는 제2 목표 신호(y) 또는 제3 목표 신호($\hat{S}$)를 수신한 후, 음성 신호를 인식한다(S227).
- [0086] 다음은 도 3을 참조하여 본 발명의 일 실시 예에 따른 잡음 제거 장치의 구조를 설명한다.
- [0087] 도 3 내지 도 5는 본 발명의 일 실시 예와 관련된 잡음 제거 장치의 효과를 설명하기 위한 도면이다.
- [0088] 도 3을 참조하면, 본 발명의 일 실시 예에 따른 잡음 제거 장치의 효과를 설명하기 위하여 복수의 마이크로폰(301)이 다양한 음향 신호를 수신하는 경우를 시뮬레이션 환경으로 설정한다. 시뮬레이션 환경에서, 2 개의 IPTV 스피커(303)가 IPTV의 복수의 마이크로폰(301)으로부터 0.3m 떨어진 위치에 180° 방향으로 각각 놓여 있고, 복수의 마이크로폰(301)으로부터 1.5m 떨어진 위치에 정면으로 음성 신호의 음원(S)이 있다. 다양한 종류의 잡음원(N)들이 복수의 마이크로폰(301)으로부터 1.5m 떨어진 위치에 30° 간격으로 반원을 그리며 놓여 있다. 다양한 종류의 잡음원(N)들은 예컨대, 자동차 엔진 소리, 음악 소리, 음성 데이터베이스 예컨대, TIMIT 데이터베이스에 포함된 남녀 음성 신호를 포함할 수 있다.
- [0089] 본 발명의 일 실시 예에 따른 잡음 제거 장치는 위와 같은 시뮬레이션 환경에서 복수의 마이크로폰(301)이 수신한 혼합 신호로부터 음성 신호를 분리하여 추출한 후 인식한다. 이때, 기기 음향 제거부는 독립 성분 분석 알고리즘을 이용하여 IPTV 스피커(303)에서 발생하는 소리를 제거하고, 음성 인식부(111)는 HMM 알고리즘을 이용하여 음성을 인식한다.
- [0090] 도 4 및 도 5를 참조하면, 시뮬레이션 적용 결과, 사용자(S)의 음성 신호 대 IPTV 스피커(303)에서 발생하는 기기 음향 간의 신호 대 잡음비(SNR, Signal to Noise Ratio)가 5dB일 때, 사용자(S)의 음성 신호 대 다양한 종류의 잡음원(N)에서 발생하는 잡음의 신호 대 잡음비(NOISE-SNR)에 따른 단어 인식률을 비교할 수 있다. 이때, 음성 신호 대 잡음비(NOISE-SNR)가 0(zero)인 경우는, 음성 신호의 크기와 잡음의 크기가 동일한 경우를 의미하고, 음성 신호 대 잡음비(NOISE-SNR)가 커질수록 음성 신호의 크기가 잡음의 크기보다 커진다.
- [0091] 마이크로폰(mic) 입력은 복수의 마이크로폰(mic1, mic2, mic3, mic4)이 수신한 혼합 신호, 기기 음향 제거 출력은 혼합 신호에서 기기 음향이 제거된 신호, 제1 목표 신호(U)는 목표 신호 추출부(101)가 출력한 신호, 제2 목표 신호(Y)는 제 잡음 제거부(107)가 출력한 신호, 제3 목표 신호($\hat{S}$)는 제2 잡음 제거부(109)가 출력한 신호를 의미한다.
- [0092] 시뮬레이션 결과, 본 발명의 일 실시 예에 따른 잡음 제거 방법의 각 단계별 출력을 비교하면, 본 발명의 일 실시 예에 따른 잡음 제거 방법이 음성 인식에 효과적이며 특히, 음성 인식 환경이 나쁜 경우 예컨대, 음성 신호 대 잡음비(NOISE-SNR)가 0(zero)인 경우에도 본 발명의 일 실시 예에 따른 잡음 제거 방법을 적용하면 음성 인식이 점점 좋아진다
- [0093] 본 발명의 일 실시 예에 의하면, 전술한 방법은 프로그램이 기록된 매체에 프로세서가 읽을 수 있는 코드로서 구현하는 것이 가능하다.
- [0094] 상기와 같이 설명된 잡음 제거 장치는 상기 설명된 실시 예들의 구성과 방법이 한정되게 적용될 수 있는 것이 아니라, 상기 실시 예들은 다양한 변형이 이루어질 수 있도록 각 실시 예들의 전부 또는 일부가 선택적으로 조합되어 구성될 수 있다.

도면

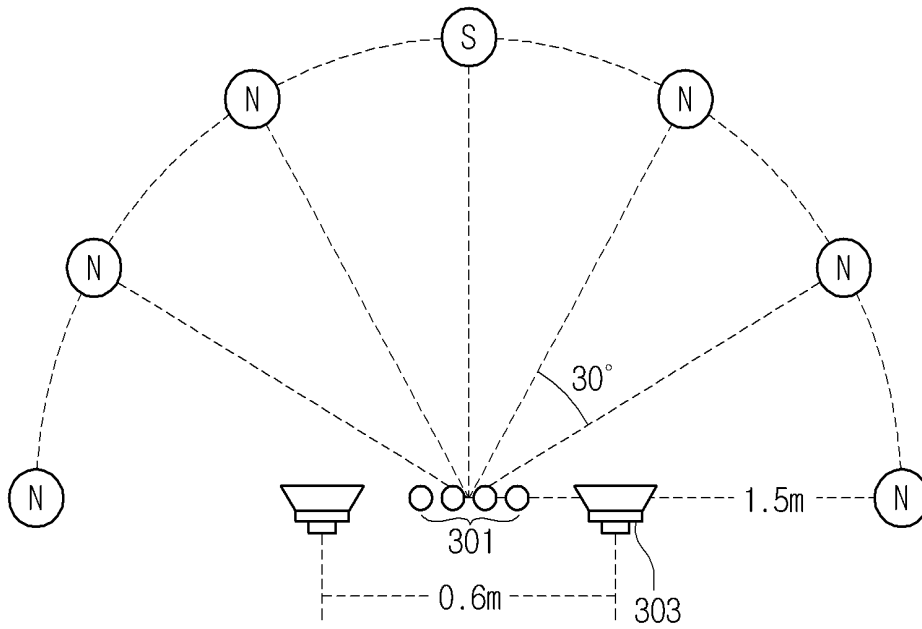
도면1



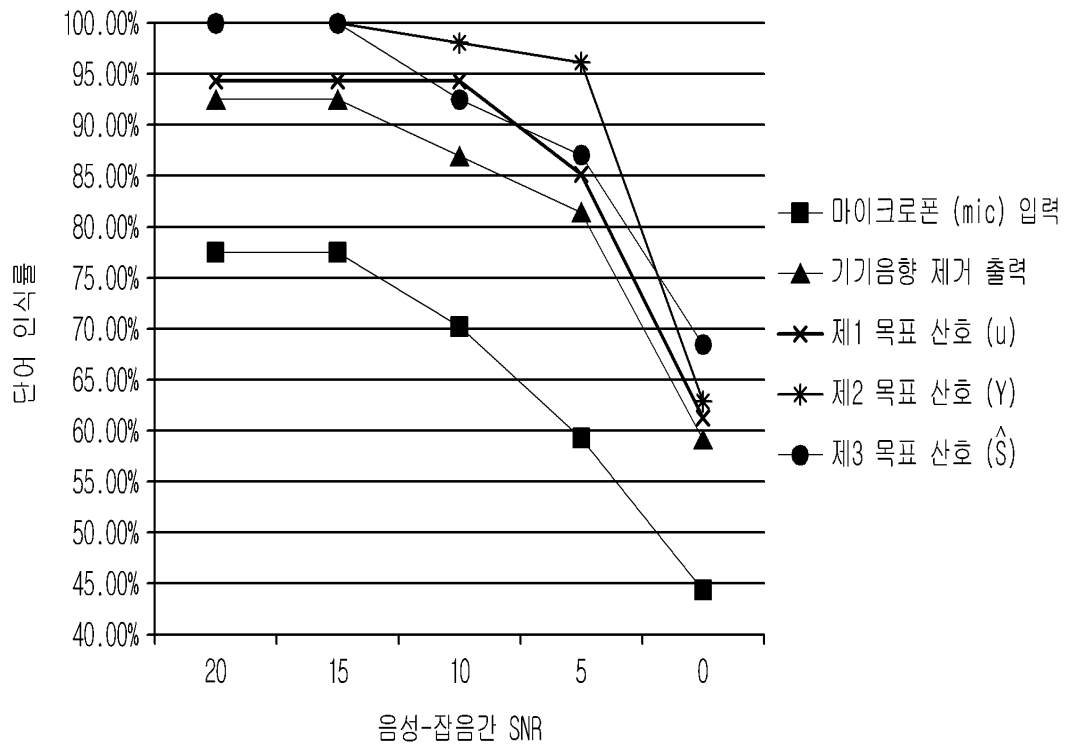
도면2



도면3



도면4



도면5

음성-잡음간 SNR	20	15	10	5	0
마이크로폰(mic) 입력	77.78%	77.78%	70.37%	59.26%	44.44%
기기 음향 제거 출력	92.59%	92.59%	87.04%	81.48%	59.26%
제1 목표 신호(U)	94.44%	94.44%	94.44%	85.19%	61.11%
제2 목표 신호(Y)	100.00%	100.00%	98.15%	96.30%	62.96%
제3 목표 신호($\hat{S}$)	100.00%	100.00%	92.59%	87.04%	68.52%