



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0130875
(43) 공개일자 2022년09월27일

(51) 국제특허분류(Int. Cl.)
G10L 21/0208 (2013.01) H04R 3/02 (2006.01)
(52) CPC특허분류
G10L 21/0208 (2013.01)
H04R 3/02 (2013.01)
(21) 출원번호 10-2021-0035663
(22) 출원일자 2021년03월19일
심사청구일자 없음

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
박형민
서울특별시 강남구 삼성로 212, 3동 1002호 (대치동, 은마아파트)
신의협
서울특별시 영등포구 당산로49길 11-4, 906호(당산동6가, 유담 유블레스2차)
(74) 대리인
이준영

전체 청구항 수 : 총 18 항

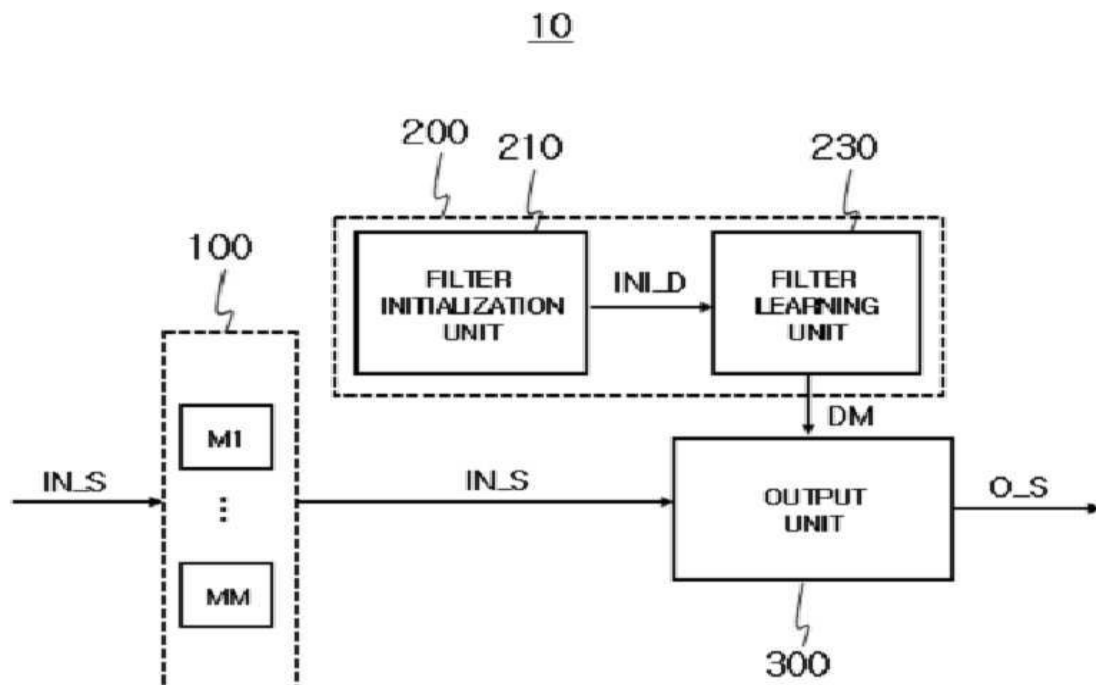
(54) 발명의 명칭 타겟 음성신호 추출 방법 및 타겟 음성신호 추출 장치

(57) 요약

본 발명의 실시예에 따른 타겟 음성신호 추출 방법에서는, 복수의 마이크들이 외부 입력신호를 수신할 수 있다. 필터 초기화부가 채널 수에 상응하는 항등행렬에서 제1 채널을 출력채널로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널로 설정하여 디믹싱 매트릭스를 초기화할

(뒷면에 계속)

대표도 - 도1



수 있다. 필터 학습부가 출력채널을 통해서 출력되는 출력신호 및 더미채널을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스 및 방향벡터에 대한 일정한 추가조건을 만족시키는 비용함수를 설정하고, 비용함수에 기초하여 디믹싱 매트릭스를 업데이트할 수 있다. 출력부가 디믹싱 매트릭스 및 외부 입력신호에 기초하여 타겟 소스에 상응하는 출력신호를 제공할 수 있다.

본 발명에 따른 타겟 음성신호 추출 방법에서는, 출력채널을 통해서 출력되는 출력신호 및 더미채널을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스 및 방향벡터에 대한 일정한 추가조건을 만족시키는 비용함수를 설정하고, 비용함수에 기초하여 디믹싱 매트릭스를 업데이트함으로써 타겟 음원신호에 대한 무왜곡 조건을 걸어 음원신호 왜곡문제를 해결할 수 있다.

이 발명을 지원한 국가연구개발사업

과제고유번호	1711102796
과제번호	2019-0-01376-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	혁신성장동력프로젝트(R&D)(과기정통부)
연구과제명	다중 화자간 대화 음성인식 기술개발
기 여 율	1/1
과제수행기관명	서강대학교산학협력단
연구기간	2019.04.01 ~ 2021.12.31

명세서

청구범위

청구항 1

복수의 마이크들이 외부 입력신호를 수신하는 단계;

필터 초기화부가 채널 수에 상응하는 항등행렬에서 제1 채널을 출력채널로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널로 설정하여 디믹싱 매트릭스를 초기화하는 단계;

필터 학습부가 상기 출력채널을 통해서 출력되는 출력신호 및 상기 더미채널을 통해서 출력되는 상기 더미신호 간의 독립성이 최대화되고, 상기 디믹싱 매트릭스 및 방향벡터에 대한 일정한 추가조건을 만족시키는 비용함수를 설정하고, 상기 비용함수에 기초하여 상기 디믹싱 매트릭스를 업데이트하는 단계;

출력부가 상기 디믹싱 매트릭스 및 상기 외부 입력신호에 기초하여 상기 타겟 소스에 상응하는 상기 출력신호를 제공하는 단계를 포함하는 타겟 음성신호 추출 방법.

청구항 2

제1항에 있어서,

상기 추가조건은,

상기 디믹싱 매트릭스의 N 번째 행 벡터(N은 자연수)와 상기 방향벡터의 곱이 채널에 따라 일정한 값을 갖는 것을 특징으로 하는 타겟 음성신호 추출 방법.

청구항 3

제2항에 있어서,

상기 제1 채널에 대해서 상기 디믹싱 매트릭스의 첫번째 행 벡터와 상기 방향벡터의 곱은 1이고,

상기 더미채널에 대해서 상기 디믹싱 매트릭스의 N 번째 행 벡터(N은 1보다 큰 자연수)와 상기 방향벡터의 곱은 0인 것을 특징으로 하는 타겟 음성신호 추출 방법.

청구항 4

제1항에 있어서,

상기 필터 학습부는 주요화최소화(Majorization-Minimization) 기법에 기반한 어그질러리 함수를 상기 비용함수로 설정하고, 상기 비용함수에 기초하여 상기 디믹싱 매트릭스를 업데이트하는 것을 특징으로 하는 타겟 음성신호 추출 방법.

청구항 5

제4항에 있어서,

상기 어그질러리 변수(Auxiliary variable)에 적용되는 블록 사이즈는 1인 것을 특징으로 하는 타겟 음성신호 추출 방법.

청구항 6

제5항에 있어서,

상기 타겟 소스에 상응하는 상기 출력신호는 온라인으로 제공하는 것을 특징으로 하는 타겟 음성신호 추출 방법.

청구항 7

외부 입력신호를 수신하는 복수의 마이크들;

채널 수에 상응하는 항등행렬에서 제1 채널을 출력채널로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널로 설정하여 디믹싱 매트릭스를 초기화하는 필터 초기화부;

상기 출력채널을 통해서 출력되는 출력신호 및 상기 더미채널을 통해서 출력되는 상기 더미신호 간의 독립성이 최대화되고, 상기 디믹싱 매트릭스 및 상기 방향벡터에 대한 일정한 추가조건을 만족시키는 비용함수를 설정하고, 상기 비용함수에 기초하여 상기 디믹싱 매트릭스를 업데이트하는 필터 학습부; 및

상기 디믹싱 매트릭스 및 상기 외부 입력신호에 기초하여 상기 타겟 소스에 상응하는 상기 출력신호를 제공하는 출력부를 포함하는 타겟 음성신호 추출 장치.

청구항 8

제7항에 있어서,

상기 제1 채널에 대해서 상기 디믹싱 매트릭스의 첫번째 행 벡터와 상기 방향벡터의 곱은 1이고,

상기 더미채널에 대해서 상기 디믹싱 매트릭스의 N 번째 행 벡터(N은 1보다 큰 자연수)와 상기 방향벡터의 곱은 0인 것을 특징으로 하는 타겟 음성신호 추출 장치.

청구항 9

제7항에 있어서,

상기 필터 학습부는 상기 디믹싱 매트릭스 및 상기 방향벡터에 대한 일정한 추가조건을 만족시키며, 주요화최소화(Majorization-Minimization) 기법에 기반한 어그질러리 함수를 비용함수로 설정하고, 상기 비용함수에 기초하여 상기 디믹싱 매트릭스를 업데이트하는 것을 특징으로 하는 타겟 음성신호 추출 장치.

청구항 10

제9항에 있어서,

상기 출력신호에 따라 계산된 분산에 기초하여 상기 디믹싱 매트릭스를 업데이트하는 것을 특징으로 하는 타겟 음성신호 추출 장치.

청구항 11

제10항에 있어서,

상기 분산은 현재 시간에 상응하는 현재 프레임을 기준으로 상기 현재 시간 이전의 이전 프레임들 및 상기 현재 시간 이후의 이후 프레임들에 기초하여 계산되는 것을 특징으로 하는 타겟 음성신호 추출 장치.

청구항 12

제11항에 있어서,

상기 분산은 현재 시간을 기준으로 현재 시간 이전의 이전 분산에 일정한 가중치를 곱하여 계산되는 것을 특징으로 하는 타겟 음성신호 추출 장치.

청구항 13

제1항에 있어서,

상기 방향벡터는 상기 입력신호의 파워 및 상기 입력신호 중 타겟소스의 파워 사이의 비율을 나타내는 값에 해당하는 타겟 마스크에 기초하여 산출되는 것을 특징으로 하는 타겟 음성신호 추출 방법.

청구항 14

제13항에 있어서,

상기 방향벡터는 온라인 업데이트 방식으로 실시간 업데이트 되는 것을 특징으로 하는 타겟 음성신호 추출

방법.

청구항 15

제4항에 있어서,

상기 어그질러리 변수는 상기 출력신호를 제공하는 출력모델에 따라서 결정되는 가중치 함수에 기초하여 결정되는 것을 특징으로 하는 타겟 음성신호 추출 방법.

청구항 16

복수의 마이크들이 외부 입력신호를 수신하는 단계;

필터 초기화부가 채널 수에 상응하는 항등행렬에서 제1 채널을 출력채널로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널로 설정하여 초기화된 방향벡터를 뺀 값으로 디믹싱 매트릭스를 초기화하는 단계;

필터 학습부가 상기 출력채널을 통해서 출력되는 출력신호에 기초하여 분산을 계산하고, 상기 계산된 분산에 따라 가중치 함수를 산출하고, 상기 출력신호의 스케일을 조정하여 스케일 조정된 출력신호에 따라 방향벡터를 계산하고, 상기 분산 및 상기 방향벡터에 기초하여 상기 디믹싱 매트릭스를 업데이트하는 단계;

출력부가 상기 디믹싱 매트릭스 및 상기 외부 입력신호에 기초하여 상기 타겟 소스에 상응하는 상기 출력신호를 제공하는 단계를 포함하는 타겟 음성신호 추출 방법.

청구항 17

제16항에 있어서,

상기 분산은 온라인 업데이트 방식으로 업데이트되는 것을 특징으로 하는 타겟 음성신호 추출 방법.

청구항 18

외부 입력신호를 수신하는 복수의 마이크들;

채널 수에 상응하는 항등행렬에서 제1 채널을 출력채널로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널로 설정하여 초기화된 방향벡터를 뺀 값으로 디믹싱 매트릭스를 초기화하는 필터 초기화부;

상기 출력채널을 통해서 출력되는 출력신호에 기초하여 분산을 계산하고, 상기 계산된 분산에 따라 가중치 함수를 산출하고, 상기 출력신호의 스케일을 조정하여 스케일 조정된 출력신호에 따라 방향벡터를 계산하고, 상기 분산 및 상기 방향벡터에 기초하여 상기 디믹싱 매트릭스를 업데이트하는 필터 학습부;

상기 디믹싱 매트릭스 및 상기 외부 입력신호에 기초하여 상기 타겟 소스에 상응하는 상기 출력신호를 제공하는 출력부를 포함하는 타겟 음성신호 추출 장치.

발명의 설명

기술 분야

[0001] 본 발명은 타겟 음성신호 추출 방법 및 타겟 음성신호 추출 장치에 관한 것이다.

배경 기술

[0002] 마이크를 통해서 입력되는 소리 입력신호는 음성인식에 필요한 타겟 음성뿐만 아니라 음성인식에 방해가 되는 노이즈들이 포함될 수 있다. 소리 입력신호에서 노이즈를 제거하고, 원하는 타겟 음성만을 추출하여 음성인식의 성능을 높이기 위한 다양한 연구가 진행되고 있다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) (한국등록특허) 제10-1133308호 (등록일자, 2012.3.28)

발명의 내용

해결하려는 과제

[0004] 본 발명이 이루고자 하는 기술적 과제는 출력채널을 통해서 출력되는 출력신호 및 더미채널을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스 및 방향벡터에 대한 일정한 추가조건을 만족시키는 비용함수를 설정하고, 비용함수에 기초하여 디믹싱 매트릭스를 업데이트함으로써 타겟 음원신호에 대한 무왜곡 조건을 걸어 음원신호 왜곡문제를 해결할 수 있는 타겟 음성신호 추출 방법을 제공하는 것이다.

[0005] 또한, 본 발명이 이루고자 하는 기술적 과제는 출력채널을 통해서 출력되는 출력신호 및 더미채널을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스 및 방향벡터에 대한 일정한 추가조건을 만족시키며, 주요화최소화(Majorization-Minimization) 기법에 기반한 어그질러리 함수를 비용함수로 설정하고, 비용함수에 기초하여 디믹싱 매트릭스를 업데이트함으로써 타겟 음원신호에 대한 무왜곡 조건을 걸어 음원신호 왜곡문제를 해결할 수 있을 뿐만 아니라, 온라인 프로세싱이 가능할 수 있는 타겟 음성신호 추출 방법을 제공하는 것이다.

과제의 해결 수단

[0006] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 타겟 음성신호 추출 방법에서는, 복수의 마이크들이 외부 입력신호를 수신할 수 있다. 필터 초기화부가 채널 수에 상응하는 항등행렬에서 제1 채널을 출력채널로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널로 설정하여 디믹싱 매트릭스를 초기화할 수 있다. 필터 학습부가 상기 출력채널을 통해서 출력되는 출력신호 및 상기 더미채널을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 상기 디믹싱 매트릭스 및 상기 방향벡터에 대한 일정한 추가조건을 만족시키는 비용함수를 설정하고, 상기 비용함수에 기초하여 상기 디믹싱 매트릭스를 업데이트할 수 있다. 출력부가 상기 디믹싱 매트릭스 및 상기 외부 입력신호에 기초하여 상기 타겟 소스에 상응하는 상기 출력신호를 제공할 수 있다.

[0007] 일 실시예에 있어서, 상기 추가조건은 상기 디믹싱 매트릭스의 N 번째 행 벡터(N 은 자연수)와 상기 방향벡터의 곱이 채널에 따라 일정한 값을 갖을 수 있다.

[0008] 일 실시예에 있어서, 상기 제1 채널에 대해서 상기 디믹싱 매트릭스의 첫번째 행 벡터와 상기 방향벡터의 곱은 1일 수 있다.

[0009] 일 실시예에 있어서, 상기 더미채널에 대해서 상기 디믹싱 매트릭스의 N 번째 행 벡터(N 은 1보다 큰 자연수)와 상기 방향벡터의 곱은 0일 수 있다.

[0010] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 타겟 음성신호 추출 방법에서는, 복수의 마이크들이 외부 입력신호를 수신할 수 있다. 필터 초기화부가 채널 수에 상응하는 항등행렬에서 제1 채널을 출력채널로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널로 설정하여 디믹싱 매트릭스를 초기화할 수 있다. 필터 학습부가 상기 출력채널을 통해서 출력되는 출력신호 및 상기 더미채널을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 상기 디믹싱 매트릭스 및 상기 방향벡터에 대한 일정한 추가조건을 만족시키며, 주요화최소화(Majorization-Minimization) 기법에 기반한 어그질러리 함수를 비용함수로 설정하고, 상기 비용함수에 기초하여 상기 디믹싱 매트릭스를 업데이트할 수 있다. 출력부가 상기 디믹싱 매트릭스 및 상기 외부 입력신호에 기초하여 상기 타겟 소스에 상응하는 상기 출력신호를 제공할 수 있다.

[0012] 일 실시예에 있어서, 상기 타겟 소스에 상응하는 상기 출력신호는 온라인으로 제공될 수 있다.

[0013] 일 실시예에 있어서, 상기 추가조건은 상기 디믹싱 매트릭스의 N 번째 행 벡터(N 은 자연수)와 상기 방향벡터의 곱이 채널에 따라 일정한 값을 갖는 것일 수 있다.

[0014] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 타겟 음성신호 추출 장치는 복수의 마이크들, 필터 초

기화부, 필터 학습부 및 출력부를 포함할 수 있다. 복수의 마이크들은 외부 입력신호를 수신할 수 있다. 필터 초기화부는 채널 수에 상응하는 항등행렬에서 제1 채널을 출력채널로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널로 설정하여 디믹싱 매트릭스를 초기화할 수 있다. 필터 학습부는 상기 출력채널을 통해서 출력되는 출력신호 및 상기 더미채널을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 상기 디믹싱 매트릭스 및 상기 방향벡터에 대한 일정한 추가조건을 만족시키는 비용함수를 설정하고, 상기 비용함수에 기초하여 상기 디믹싱 매트릭스를 업데이트할 수 있다. 출력부는 상기 디믹싱 매트릭스 및 상기 외부 입력신호에 기초하여 상기 타겟 소스에 상응하는 상기 출력신호를 제공할 수 있다.

[0015] 일 실시예에 있어서, 상기 제1 채널에 대해서 상기 디믹싱 매트릭스의 첫번째 행 벡터와 상기 방향벡터의 곱은 1이고, 상기 더미채널에 대해서 상기 디믹싱 매트릭스의 N 번째 행 벡터(N은 1보다 큰 자연수)와 상기 방향벡터의 곱은 0일 수 있다.

[0016] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 타겟 음성신호 추출 장치는 복수의 마이크들, 필터 초기화부, 필터 학습부 및 출력부를 포함할 수 있다. 복수의 마이크들은 외부 입력신호를 수신할 수 있다. 필터 초기화부는 채널 수에 상응하는 항등행렬에서 제1 채널을 출력채널로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널로 설정하여 디믹싱 매트릭스를 초기화할 수 있다. 필터 학습부는 상기 출력채널을 통해서 출력되는 출력신호 및 상기 더미채널을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 상기 디믹싱 매트릭스 및 상기 방향벡터에 대한 일정한 추가조건을 만족시키며, 주요화최소화(Majorization-Minimization) 기법에 기반한 어그질러리 함수를 비용함수로 설정하고, 상기 비용함수에 기초하여 상기 디믹싱 매트릭스를 업데이트할 수 있다. 출력부는 상기 디믹싱 매트릭스 및 상기 외부 입력신호에 기초하여 상기 타겟 소스에 상응하는 상기 출력신호를 제공할 수 있다.

[0017] 위에서 언급된 본 발명의 기술적 과제 외에도, 본 발명의 다른 특징 및 이점들이 이하에서 기술되거나, 그러한 기술 및 설명으로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

발명의 효과

[0018] 이상과 같은 본 발명에 따르면 다음과 같은 효과가 있다.

[0019] 본 발명에 따른 타겟 음성신호 추출 방법에서는, 출력채널을 통해서 출력되는 출력신호 및 더미채널을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스 및 방향벡터에 대한 일정한 추가조건을 만족시키는 비용함수를 설정하고, 비용함수에 기초하여 디믹싱 매트릭스를 업데이트함으로써 타겟 음원신호에 대한 무왜곡 조건을 걸어 음원신호 왜곡문제를 해결할 수 있다.

[0020] 또한, 본 발명에 따른 타겟 음성신호 추출 방법에서는, 출력채널을 통해서 출력되는 출력신호 및 더미채널을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스 및 방향벡터에 대한 일정한 추가조건을 만족시키며, 주요화최소화(Majorization-Minimization) 기법에 기반한 어그질러리 함수를 비용함수로 설정하고, 비용함수에 기초하여 디믹싱 매트릭스를 업데이트함으로써 타겟 음원신호에 대한 무왜곡 조건을 걸어 음원신호 왜곡문제를 해결할 수 있을 뿐만 아니라, 온라인 프로세싱이 가능할 수 있다.

[0021] 이 밖에도, 본 발명의 실시 예들을 통해 본 발명의 또 다른 특징 및 이점들이 새롭게 파악될 수도 있을 것이다.

도면의 간단한 설명

[0022] 도 1은 본 발명의 실시예들에 따른 타겟 음성신호 추출 장치를 나타내는 도면이다.

도 2는 본 발명의 실시예들에 따른 타겟 음성신호 추출 방법을 나타내는 순서도이다.

도 3은 도 1의 타겟 음성신호를 분리하기 위한 일 예를 설명하기 위한 도면이다.

도 4는 도 1의 타겟 음성신호 추출 장치에서 사용되는 외부 입력신호의 일 예를 나타내는 도면이다.

도 5는 도 2의 타겟 음성신호 추출 방법에서 사용되는 디믹싱 매트릭스를 초기화하는 과정을 설명하기 위한 도면이다.

도 6은 도 2의 타겟 음성신호 추출 방법에서 사용되는 방향벡터를 설명하기 위한 도면이다.

도 7은 본 발명의 실시예들에 따른 타겟 음성신호 추출 방법을 나타내는 순서도이다.

도 8은 본 발명의 일 실시예에 따른 출력신호에 따라 계산되는 분산에 기초하여 디믹싱 매트릭스를 업데이트하는 과정을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0023] 본 명세서에서 각 도면의 구성 요소들에 참조번호를 부가함에 있어서 동일한 구성 요소들에 한해서는 비록 다른 도면상에 표시되더라도 가능한한 동일한 번호를 가지도록 하고 있음에 유의하여야 한다.
- [0024] 한편, 본 명세서에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.
- [0025] 단수의 표현은 문맥상 명백하게 다르게 정의하지 않는 한, 복수의 표현을 포함하는 것으로 이해되어야 하는 것으로, 이들 용어들에 의해 권리범위가 한정되어서는 아니 된다.
- [0026] "포함하다" 또는 "가지다" 등의 용어는 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0027] 이하, 첨부되는 도면을 참고하여 상기 문제점을 해결하기 위해 고안된 본 발명의 바람직한 실시예들에 대해 상세히 설명한다.
- [0028] 도 1은 본 발명의 실시예들에 따른 타겟 음성신호 추출 장치를 나타내는 도면이고, 도 2는 본 발명의 실시예들에 따른 타겟 음성신호 추출 방법을 나타내는 순서도이고, 도 3은 도 1의 타겟 음성신호를 추출하기 위한 일 예를 설명하기 위한 도면이고, 도 4는 도 1의 타겟 음성신호 추출 장치에서 사용되는 외부 입력신호의 일 예를 나타내는 도면이고, 도 5는 도 2의 타겟 음성신호 추출 방법에서 사용되는 디믹싱 매트릭스를 초기화하는 과정을 설명하기 위한 도면이고, 도 6은 도 2의 타겟 음성신호 추출 방법에서 사용되는 방향벡터를 설명하기 위한 도면이다.
- [0029] 도 1 내지 6을 참조하면, 본 발명의 실시예에 따른 타겟 음성신호 추출 장치(10)는 복수의 마이크들(100), 필터부(200) 및 출력부(300)를 포함할 수 있다. 필터부(200)는 필터 초기화부(210) 및 필터 학습부(230)를 포함할 수 있다. 본 발명에 따른 타겟 음성신호 추출 방법에서는, 복수의 마이크들(100)이 외부 입력신호(IN_S)를 수신할 수 있다(S100). 예를 들어, 외부의 음원 소스는 제1 음원(S1) 내지 제N 음원(SN)을 포함할 수 있다. 제1 음원(S1) 내지 제N 음원(SN)으로부터 전달되는 음원신호들은 제1 마이크(M1) 내지 제N 마이크(MN)로 입력될 수 있다. 제1 마이크(M1) 내지 제N 마이크(MN)로 입력되는 신호는 외부 입력신호(IN_S)일 수 있다. 외부 입력신호(IN_S)는 시간에 따른 주파수 성분에 대한 정보가 포함될 수 있다. 이 경우, 제1 방향(D1)을 따라서 시간(T)이 표시될 수 있고, 제2 방향(D2)을 따라 주파수(F)가 표시될 수 있다. 외부 입력신호(IN_S)는 $\mathbf{x}(f, \tau)$ 로 표현될 수 있고, $\mathbf{x}(f, \tau)$ 는 아래 [수학식 1]과 같이 나타낼 수 있다.
- [0030] [수학식 1]
- [0031]
$$\mathbf{x}(f, \tau) = \mathbf{A}(f)\mathbf{s}(f, \tau)$$
- [0032] 여기서, $\mathbf{A}(f)$ 는 믹싱 매트릭스일 수 있고, $\mathbf{s}(f, \tau)$ 는 음원신호일 수 있다.
- [0033] 필터 초기화부(210)가 채널 수에 상응하는 항등행렬에서 제1 채널(CH1)을 출력채널(OC)로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널(DC)로 설정하여 디믹싱 매트릭스(DM)를 초기화(INI_D)할 수 있다. 예를 들어, 제1 음원(S1)은 타겟 소스일 수 있다. 제1 음원(S1)이 타겟 소스인 경우, 제1 음원(S1)에 상응하는 채널은 제1 채널(CH1)일 수 있다. 더미채널(DC)은 제2 채널 내지 제N 채널을 포함할 수 있다. 더미채널(DC)은 항등행렬의 제2 채널 내지 제N 채널에서 방향벡터(DV)를 뺀 값으로 초기화하여 더미채널에서 타겟 소스가 제거될 수 있다.
- [0034] 필터 학습부(230)가 출력채널(OC)을 통해서 출력되는 출력신호(O_S) 및 더미채널(DC)을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스(DM) 및 방향벡터(DV)에 대한 일정한 추가조건을 만족시키는 비용함수(Q)를 설정하고, 비용함수(Q)에 기초하여 디믹싱 매트릭스(DM)를 업데이트할 수 있다.
- [0035] 예를 들어, 비용함수(Q)는 아래의 [수학식 2]와 같이 표현될 수 있다.

[0036] [수학식 2]

$$Q(\mathbf{W}(f), \mathbf{V}(f)) = \frac{1}{2} \sum_{n=1}^N \mathbf{w}_n^H(f) \mathbf{V}_n(f) \mathbf{w}_n(f) - \log|\det \mathbf{W}(f)| + \alpha(f)(\mathbf{w}_n^H(f) \mathbf{h}(f) - \beta_n)$$

$$\frac{\partial Q(\mathbf{W}(f), \mathbf{V}(f))}{\partial \mathbf{w}_n^H(f)} = \mathbf{V}_n(f) \mathbf{w}_n(f) - \frac{\partial}{\partial \mathbf{w}_n^H(f)} \log|\det \mathbf{W}(f)| + \alpha(f) \mathbf{h}(f) = 0$$

[0037]

[0038] 여기서, $\mathbf{w}_n(f)$ 는 디믹싱 매트릭스, $\mathbf{V}_n(f)$ 는 어그질러리 변수(Auxiliary variable), $\alpha(f)$ 는 라그랑제 멀티플라이어(Lagrange multiplier), $\mathbf{h}(f)$ 는 방향벡터, $\beta_n = \begin{cases} 1 & (n = 1) \\ 0 & (n \neq 1) \end{cases}$, n은 채널 수일 수 있다.

[0039] 여기서, 이러한 비용함수(Q)는 아래와 같이 독립성을 최대화하기 위한 기본적인 비용함수에 디믹싱 매트릭스(DM) 및 방향벡터(DV)에 대한 일정한 추가조건을 만족시키는 [수학식 2-1]의 upper bound로서 설정될 수 있다.

[0040] [수학식 2-1]

$$J(\mathbf{W}(f)) = \sum_{n=1}^N E[-\log q(Y_m(f, \tau))] - \log|\det \mathbf{W}(f)| + \sum_{n=1}^N \alpha(f)(\mathbf{w}_n^H(f) \mathbf{h}(f) - \beta_n)$$

[0041]

[0042] 여기서, $q(Y_m(f, \tau))$ 는 후술할 출력 모델일 수 있다. 이로부터 아래의 [수학식 2-2]와 같은 부등식을 사용하여 $J(\mathbf{W}(f)) \leq Q(\mathbf{W}(f), \mathbf{V}(f))$ 의 관계가 될 수 있도록 설정할 수 있다.

[0043] [수학식 2-2]

$$G(\mathbf{y}) \leq \frac{G'_R(r_0)}{2r_0} \|\mathbf{y}\|_2^2 + \left(G_R(r_0) - \frac{r_0 G'_R(r_0)}{2} \right)$$

[0044]

[0045] 여기서 r_0 은 스칼라, $\mathbf{y}$ 는 벡터에 해당하며 등식은 $r_0 = \|\mathbf{y}\|_2^2$ 일 때 성립한다. 이러한 부등식을 적용하여 설정된 비용함수(Q)의 어그질러리 변수는 아래의 [수학식 2-3]과 같이 표현될 수 있다.

[0046] [수학식 2-3]

$$\mathbf{V}_m(f) = E[\phi_m(f, \tau) \mathbf{x}(f, \tau) \mathbf{x}^H(f, \tau)]$$

[0047]

[0048] 여기서, $\phi_m(f, \tau)$ 는 출력 모델에 따라서 결정되는 가중치 함수(weighting function)이다. 이 함수는 위의 [수학식 2-2]의 $\frac{G'_R(r_0)}{2r_0}$ 값으로 설정된다.

[0049] 우선, 함수 $\phi_m(f, \tau)$ 의 설정 및 분산의 추정을 위해서 필요한 출력 신호의 통계적 모델링에 대한 내용이 선행될 수 있다. 이 경우, 출력신호는 아래의 [수학식 2-4]과 같이 채널별로 표현될 수 있다.

[0050] [수학식 2-4]

$$\mathbf{y}(f, \tau) = \begin{bmatrix} Y_1(f, \tau) \\ \vdots \\ Y_M(f, \tau) \end{bmatrix}$$

[0051]

[0052] 이러한 출력신호는 아래의 [수학식 2-5]와 같이 모델링될 수 있다.

[0053] [수학식 2-5]

$$q(Y_m(f, \tau)) \propto \frac{1}{\lambda_m(f, \tau)} \exp \left\{ - \left(\frac{|Y_m(f, \tau)|^2}{\lambda_m(f, \tau)} \right)^\zeta \right\}$$

[0054]

$$\lambda_m(f, \tau)$$

[0055] 여기서, $\lambda_m(f, \tau)$ 는 분산, ζ 는 셰이프 파라미터(shape parameter)를 나타낼 수 있다. 이로부터 아래의 [수학식 2-6]과 같이 우도 최대화 기반으로 분산을 추정할 수 있다.

[0056] [수학식 2-6]

$$\lambda_m(f, \tau) = \zeta^{\frac{1}{\zeta}} |Y_m(f, \tau)|^2$$

[0057]

[0058] 또한, 분산은 현재 시간에 상응하는 현재 프레임을 기준으로 현재 시간 이전의 이전 프레임들 및 현재 시간 이후의 이후 프레임들에 기초하여 아래의 [수학식 2-7]와 같이 계산될 수 있다.

[0059] [수학식 2-7]

$$\lambda_m(f, \tau) = \frac{\zeta^{\frac{1}{\zeta}}}{2\tau_0 + 1} \sum_{t=\tau-\tau_0}^{\tau+\tau_0} |Y_m(f, t)|^2$$

[0060]

$$\tau_0$$

[0061] 여기서 τ_0 는 인접 프레임의 개수일 수 있다.

$$\phi_m(f, \tau)$$

[0062] 한편, 출력 모델에 따라서 $\phi_m(f, \tau)$ 는 아래의 [수학식 2-8]와 같이 계산될 수 있다.

[0063] [수학식 2-8]

$$\phi_m(f, \tau) = \zeta \frac{|Y_m(f, \tau)|^{2\zeta-2}}{\lambda_m^\zeta(f, \tau)}.$$

[0064]

$$\phi_0$$

[0065] 이 경우, 안정적인 성능을 위하여 위의 가중치 함수는 아래의 [수학식 2-8]과 같이 미리 설정된 값 ϕ_0 에 의해서 min함수를 통해서 클리핑될 수 있다.

[0066] [수학식 2-8]

$$\phi_m(f, \tau) = \min \left(\zeta \frac{|Y_m(f, \tau)|^{2\zeta-2}}{\lambda_m^\zeta(f, \tau)}, \phi_0 \right)$$

[0067]

[0068] 또한, 출력신호는 아래의 [수학식 2-9] 와 같이 좀 더 클릭(clique)를 사용한 확장된 모델을 사용할 수 있다.

[0069] [수학식 2-9]

$$q(Y_m(1, \tau), \dots, Y_m(F, \tau)) \propto \frac{1}{\prod_{c=1}^C \tilde{\lambda}_m^{N_c}(c, \tau)} \exp \left\{ - \left(\sum_{c=1}^C \left(\frac{\sum_{f \in \Omega_c} |Y_m(f, \tau)|^2}{\tilde{\lambda}_m(c, \tau)} \right)^\gamma \right)^\beta \right\}$$

[0070]

$$c \quad C \quad \Omega_c$$

[0071] 여기서 이때, c 와 C 는 각각 clique 인덱스와 clique의 개수, 그리고 해당 clique에 속하는 주파수 bin의

$$\tilde{\lambda}_m(c, \tau)$$

$$\beta \quad \gamma$$

집합이다. 또한 $\tilde{\lambda}_m(c, \tau)$ 는 m번째 출력의 c번째 clique의 분산일 수 있다. β 및 γ 는 shape parameter이다. 여기서 clique구조의 설계와 shape parameter 및 분산의 설정에 따라서 다양한 종류의 pdf를 나타낼 수 있다. 이때 분산이 시간 또는 clique에 따라서 변하는 variance라고 가정하면, 우도 최대화 기반으로 아래와 같이 추정

할 수 있다.

[수학식 2-10]

$$\tilde{\lambda}_m(c, \tau) = \frac{1}{C} \psi_m(\tau) \kappa_m(c, \tau) \sum_{f \in \Omega_c} |Y_m(f, \tau)|^2$$

$$\psi_m(\tau) = \beta \left[\sum_{c=1}^C \left(\frac{\sum_{f \in \Omega_c} |Y_m(f, \tau)|^2}{\tilde{\lambda}_m(c, \tau)} \right)^\gamma \right]^{\beta-1}$$

$$\kappa_m(c, \tau) = \gamma \left(\frac{\sum_{f \in \Omega_c} |Y_m(f, \tau)|^2}{\tilde{\lambda}_m(c, \tau)} \right)^{\gamma-1}$$

또한, clique기반의 방법의 경우 비용함수(Q)의 어그질러리 변수는 아래의 [수학식 2-12]과 같이 주어진다.

[수학식 2-12]

$$\mathbf{V}_m(f) = E \left[\beta \rho_m^{2\beta-2}(\tau) \sum_{c=1}^C \frac{(\sum_{f \in \Omega_c} |Y_m(f, \tau)|^2)^{\gamma-1}}{\tilde{\lambda}_m^\gamma(c, \tau)} \mathbf{x}(f, \tau) \mathbf{x}^H(f, \tau) \right]$$

$$\rho_m(\tau) = \sqrt{\sum_{c=1}^C \left(\frac{\sum_{f \in \Omega_c} |Y_m(f, \tau)|^2}{\tilde{\lambda}_m(c, \tau)} \right)^\gamma}$$

또한, 방향벡터(DV)는 $\boldsymbol{\gamma}_f = [\Gamma_f^1, \dots, \Gamma_f^{M-1}]^T$, $\Gamma_k = \exp\{j\omega_k d \sin \theta_{\text{target}}/c\}$ 와 같이 초기화될 수 있다. 혹은 방

향이 주어지지 않을 경우 $\theta = 0$ 으로 가정하여 $h(f)$ 는 1로 초기화될 수 있으며, 방향벡터는 아래와 같은 방법으로 계속 업데이트 될 수 있다.

우선, [수학식 1]의 입력신호에서 음원신호는 아래의 [수학식 1-1]과 같이 표현될 수 있다.

[수학식 1-1]

$$\mathbf{s}(f, \tau) = \begin{bmatrix} S(f, \tau) \\ \mathbf{n}(f, \tau) \end{bmatrix}$$

여기서, $S(f, \tau)$ 은 타겟소스, $\mathbf{n}(f, \tau)$ 는 노이즈를 나타낼 수 있다. 타겟 소스와 노이즈가 혼합된 입력 신호 $\mathbf{x}(f, \tau)$

의 파워에서 타겟 소스 파워의 비율을 나타내는 값에 해당하는 타겟 마스크는 아래의 [수학식 1-2]와 같이 표현될 수 있다.

[수학식 1-2]

$$\mathcal{M}(f, \tau) = \frac{|S(f, \tau)|^2}{|S(f, \tau)|^2 + \|\mathbf{n}(f, \tau)\|_2^2}$$

여기서, 타겟 소스 및 노이즈는 후술하는 출력채널, 더미채널 및 [수학식 6]으로부터 계산되는 출력신호를 통해서 아래의 [수학식 1-3]과 같이 추정될 수 있다.

[0088] [수학식 1-3]

$$diag(\mathbf{W}^{-1}(f))\mathbf{y}(f, \tau) = \begin{bmatrix} \hat{S}(f, \tau) \\ \hat{\mathbf{n}}(f, \tau) \end{bmatrix}$$

[0089]

$$diag(\mathbf{W}^{-1}(f)) \quad \mathbf{W}^{-1}(f)$$

[0090] 는 출력 신호의 scale를 조정하는 값으로 디믹싱 매트릭스의 역행렬 을 계산하여 이의

$$diag(\cdot)$$

대각선 성분만을 남기고, 그 외의 성분은 0으로 반환 하는 대각행렬을 사용할 수 있다.

[0091] 따라서, 이를 통해서 타겟 마스크를 아래의 [수학식 1-4]와 같이 추정할 수 있다.

[0092] [수학식 1-4]

$$\hat{\mathcal{M}}(f, \tau) = \frac{|\hat{S}(f, \tau)|^2}{|\hat{S}(f, \tau)|^2 + \|\hat{\mathbf{n}}(f, \tau)\|_2^2}$$

[0093]

[0094] 여기서, 타겟 소스에 대한 공분산 행렬은 아래의 [수학식 1-5]와 같이 추정될 수 있다.

[0095] [수학식 1-5]

$$\hat{\mathbf{R}}_s(f) = \frac{1}{\sum_{\tau} \hat{\mathcal{M}}(f, \tau)} \sum_{\tau} \hat{\mathcal{M}}(f, \tau) \mathbf{x}(f, \tau) \mathbf{x}^H(f, \tau)$$

[0096]

[0097] 따라서, 위의 행렬을 Eigen Value Decomposition을 통해서 Principal Eigen vector로부터 방향벡터를 추정할 수 있다.

$$\mathbf{h}^H(f; \tau)$$

[0098] 또한, 온라인 추정방법이 추가될 수 있다. 즉, 이는 을 실시간으로 업데이트하는 것에 대한 것일 수 있다.

[0099] [수학식 1-6]

$$diag(\mathbf{A}(f; \tau))\mathbf{y}(f, \tau) = \begin{bmatrix} \hat{S}(f, \tau) \\ \hat{\mathbf{n}}(f, \tau) \end{bmatrix}$$

[0100]

$$diag(\mathbf{W}^{-1}(f))$$

[0101] 여기서, 는 출력 신호의 scale를 조정하는 값으로 계속 업데이트가 되는 디믹싱 매트릭스의 역

$$\mathbf{A}(f)$$

행렬 을 바로 사용할 수 있다. 따라서 이를 통해서 아래의 [수학식 1-7]과 같이 타겟 마스크를 추정할 수 있다.

[0102] [수학식 1-7]

$$\hat{\mathcal{M}}(f, \tau) = \frac{|\hat{S}(f, \tau)|^2}{|\hat{S}(f, \tau)|^2 + \|\hat{\mathbf{n}}(f, \tau)\|_2^2}$$

[0103]

[0104] 이로부터 타겟 소스에 대한 공분산 행렬을 [수학식 1-8]과 같이 moving average를 통해 타겟에 대한 공분산 행렬을 실시간으로 업데이트할 수 있다.

[0105] [수학식 1-8]

$$\hat{\mathbf{R}}_s(f; \tau) = \eta \hat{\mathbf{R}}_s(f; \tau - 1) + (1 - \eta) \hat{\mathcal{M}}(f, \tau) \mathbf{x}(f, \tau) \mathbf{x}^H(f, \tau)$$

[0106]

$$\eta$$

[0107] 여기서 는 smoothing factor를 나타낼 수 있다. 이 행렬을 Eigen Value Decomposition을 통해서 Principal

Eigen vector로부터 방향벡터를 추정할 수 있다.

[0108] 일 실시예에 있어서, 추가조건은 디믹싱 매트릭스(DM)의 N 번째 행 벡터(N은 자연수)와 방향벡터(DV)의 곱이 채널에 따라 일정한 값을 갖을 수 있다. 예를 들어, 제1 채널(CH1)에 대해서 디믹싱 매트릭스(DM)의 첫번째 행 벡터와 방향벡터(DV)의 곱은 1일 수 있고, 더미채널(DC)에 대해서 디믹싱 매트릭스(DM)의 N 번째 행 벡터(N은 1보다 큰 자연수)와 방향벡터(DV)의 곱은 0일 수 있다. 위와 같은 내용은 아래의 [수학식 3]과 같이 표현될 수 있다.

[0109] [수학식 3]

$$w_n^H(f)h(f) = \beta_n, \quad \beta_n = \begin{cases} 1 & (n = 1) \\ 0 & (n \neq 1) \end{cases}$$

[0110]

[0111] 이 경우, 제1 채널(CH1)(출력채널)에 대한 디믹싱 매트릭스(DM)는 [수학식 4]와 같이 표현될 수 있고, 더미채널(DC)에 대한 디믹싱 매트릭스(DM)는 [수학식 5]와 같이 표현될 수 있다.

[0112] [수학식 4]

$$w_1(f) \leftarrow \frac{(W(f)V_1(f))^{-1}e_1}{h^H(f)(W(f)V_1(f))^{-1}e_1}, \quad n=1(\text{출력채널})$$

[0113]

[0114] [수학식 5]

$$w_n(f) = (W(f)V_n(f))^{-1}e_n,$$

$$w_n(f) \leftarrow w_n(f) / \sqrt{w_n^H(f)V_n^H(f)w_n(f)}, \quad n>1(\text{더미채널})$$

[0115]

[0116] 그리고 위의 [수학식 3]으로부터 아래의 [수학식 3-1]과 같은 관계가 성립할 수 있다.

[0117] [수학식 3-1]

$$W(f)h(f) = e_1$$

[0118]

[0119] 따라서, $A^{-1}(f) = W(f)$ 이므로 아래의 [수학식 3-2]과 같은 관계가 성립할 수 있다.

[0120] [수학식 3-2]

$$A(f)e_1 = h(f)$$

[0121]

[0122] 이 경우, [수학식 4]는 아래의 [수학식 4-1]과 같이 정리될 수 있다.

[0123] [수학식 4-1]

$$w_1(f) \leftarrow \frac{(W(f)V_1(f))^{-1}e_1}{h^H(f)(W(f)V_1(f))^{-1}e_1} = \frac{V_1^{-1}(f)h(f)}{h^H(f)V_1^{-1}(f)h(f)}$$

[0124]

[0125] 또한, 더미 채널에 대한 디믹싱 매트릭스(DM)를 나타내는 [수학식 5]의 경우에는 라그랑제 멀티플라이어가 반영되지 않은 것으로서, [수학식 5]에 라그랑제 멀티플라이어를 반영하면, 아래의 [수학식 5-1]과 같이 변경할 수 있다.

[0126] [수학식 5-1]

$$w_m(f) \leftarrow (W(f)V_m(f))^{-1}e_m - \frac{h^H(f)(W(f)V_m(f))^{-1}e_m}{h^H(f)(W(f)V_m(f))^{-1}e_1} (W(f)V_m(f))^{-1}e_1$$

[0127]

[0128] 따라서, 이는 아래의 [수학식 5-2]과 같이 정리될 수 있다.

[0129] [수학식 5-2]

$$\mathbf{u}_m(f) = (\mathbf{W}(f)\mathbf{V}_m(f))^{-1}\mathbf{e}_m$$

$$\hat{\mathbf{u}}_m(f) = \frac{\mathbf{V}_m^{-1}(f)\mathbf{h}(f)}{\mathbf{h}^H(f)\mathbf{V}_m^{-1}(f)\mathbf{h}(f)}$$

[0130]

[0131] 출력부(300)가 디믹싱 매트릭스(DM) 및 외부 입력신호(IN_S)에 기초하여 타겟 소스에 상응하는 출력신호(O_S)를 제공할 수 있다. 출력신호(O_S)는 $\mathbf{y}(f, \tau)$ 로 표현될 수 있고, $\mathbf{y}(f, \tau)$ 는 아래 [수학식 6]와 같이 나타낼 수 있다.

[0132] [수학식 6]

[0133]

$$\mathbf{y}(f, \tau) = \mathbf{W}(f)\mathbf{x}(f, \tau)$$

[0134] 본 발명에 따른 타겟 음성신호 추출 방법에서는, 출력채널(OC)을 통해서 출력되는 출력신호(O_S) 및 더미채널(DC)을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스(DM) 및 방향벡터(DV)에 대한 일정한 추가조건을 만족시키는 비용함수(Q)를 설정하고, 비용함수(Q)에 기초하여 디믹싱 매트릭스(DM)를 업데이트함으로써 타겟 음원신호에 대한 무왜곡 조건을 걸어 음원신호 왜곡문제를 해결할 수 있다.

[0135] 도 7은 본 발명의 실시예들에 따른 타겟 음성신호 추출 방법을 나타내는 순서도이다.

[0136] 도 1 내지 7을 참조하면, 본 발명의 실시예에 따른 타겟 음성신호 추출 장치(10)는 복수의 마이크들(100), 필터 초기화부(210), 필터 학습부(230) 및 출력부(300)를 포함할 수 있다. 본 발명에 따른 타겟 음성신호 추출 방법에서는, 복수의 마이크들(100)이 외부 입력신호(IN_S)를 수신할 수 있다. 필터 초기화부(210)가 채널 수에 상응하는 항등행렬에서 제1 채널(CH1)을 출력채널(OC)로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널(DC)로 설정하여 디믹싱 매트릭스(DM)를 초기화(INI_D)할 수 있다. 필터 학습부(230)가 출력채널(OC)을 통해서 출력되는 출력신호(O_S) 및 더미채널(DC)을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스(DM) 및 방향벡터(DV)에 대한 일정한 추가조건을 만족시키며, 주요화최소화(Majorization-Minimization) 기법에 기반한 어그질러리 함수를 비용함수로 설정하고, 비용함수(Q)에 기초하여 디믹싱 매트릭스(DM)를 업데이트할 수 있다. 출력부(300)가 디믹싱 매트릭스(DM) 및 외부 입력신호(IN_S)에 기초하여 타겟 소스에 상응하는 출력신호(O_S)를 제공할 수 있다.

[0137] 예를 들어, 어그질러리 변수는 비용함수(Q)에 포함될 수 있다. 온라인 업데이트를 위해 시간에 대한 변수 τ 가 추가된 어그질러리 변수는 아래의 [수학식 7]과 같이 온라인 업데이트 방식으로 표현될 수 있다.

[0138] [수학식 7]

$$\mathbf{V}_m(f; \tau) = \eta \mathbf{V}_m(f; \tau - 1) + (1 - \eta) \phi_m(f, \tau) \mathbf{x}(f, \tau) \mathbf{x}^H(f, \tau)$$

[0140] 여기서, η 는 smoothing factor일 수 있고, 실제로는 이에 대한 역행렬을 사용하기 때문에 이를 $\mathbf{U}_m(f; \tau) = \mathbf{V}_m^{-1}(f; \tau)$

로 정의하여 바로 업데이트를 아래와 같이 진행할 수 있다. 또한,

$$(1 - \eta) \phi_m(f, \tau) = p_m(f, \tau)$$

라고 할 때 아래의 [수학식 7-1]와 같이 주어질 수 있다.

[0141] [수학식 7-1]

$$\mathbf{U}_m(f; \tau) = \frac{1}{\eta} \mathbf{U}_m(f; \tau - 1) + \frac{p_m(f, \tau) \mathbf{U}_m(f; \tau - 1) \mathbf{x}(f, \tau) \mathbf{x}^H(f, \tau) \mathbf{U}_m^H(f; \tau - 1)}{\eta^2 + \eta p_m(f, \tau) \mathbf{x}^H(f, \tau) \mathbf{U}_m(f; \tau - 1) \mathbf{x}(f, \tau)}$$

[0142]

[0143] 이 경우, woodbury matrix inversion을 통해서 매 프레임마다 디믹싱 매트릭스(DM)의 인버전 $\mathbf{W}^{-1}(f; \tau)$ 및

어그질러리 변수의 인버전 $\mathbf{V}_n^{-1}(f; \tau)$ 를 업데이트 함으로써 타겟 소스에 상응하는 출력신호(O_S)는 온라인으로 제공될 수 있다. 이 경우, 디믹싱 매트릭스(DM)의 인버전은 [수학식 9]와 같이 표현될 수 있다.

[수학식 8]

$$\mathbf{A}(f; \tau) = \mathbf{A}(f; \tau) - \frac{\mathbf{A}(f; \tau) \mathbf{e}_n \Delta \mathbf{w}_n^H(f; \tau) \mathbf{A}(f; \tau)}{1 + \Delta \mathbf{w}_n^H(f; \tau) \mathbf{A}(f; \tau) \mathbf{e}_n}$$

여기서, $\mathbf{A}(f; \tau) = \mathbf{W}^{-1}(f; \tau)$

따라서 이를 통해서 출력 weight를 아래의 [수학식 9]과 같이 온라인으로 업데이트할 수 있다.

[수학식 9]

타겟 채널 weight

$$\mathbf{w}_1(f; \tau) \leftarrow \frac{\mathbf{U}_1(f; \tau) \mathbf{h}^H(f; \tau)}{\mathbf{h}^H(f; \tau) \mathbf{U}_1(f; \tau) \mathbf{h}^H(f; \tau)}$$

노이즈 채널 weight

$$\mathbf{u}_m(f; \tau) = \mathbf{U}_m(f; \tau) \mathbf{A}(f; \tau) \mathbf{e}_m$$

$$\hat{\mathbf{u}}_m(f; \tau) = \frac{\mathbf{U}_m(f; \tau) \mathbf{h}^H(f; \tau)}{\mathbf{h}^H(f; \tau) \mathbf{U}_m(f; \tau) \mathbf{h}^H(f; \tau)}$$

$$\mathbf{w}_m(f; \tau) \leftarrow \mathbf{u}_m(f; \tau) - (\mathbf{h}^H(f; \tau) \mathbf{u}_m(f; \tau)) \hat{\mathbf{u}}_m(f; \tau)$$

또한, [수학식 2-7]에 해당하는 분산은 현재 시간을 기준으로 현재 시간 이전의 이전 분산에 일정한 가중치를 곱하여 아래의 [수학식 10]과 같이 계산될 수 있다.

[수학식 10]

$$\lambda_m(f, \tau) = \mu \lambda_m(f, \tau) + (1 - \mu) \zeta^{\frac{1}{\gamma}} |Y_m(f, \tau)|^2$$

여기서 $\lambda_m(f, \tau)$ 는 분산, μ 는 가중치, $Y_m(f, \tau)$ 는 출력신호, τ 는 시간을 나타낸다.

그리고, [수학식 2-8]에 해당하는 클릭 분산은 아래와 같이 재귀적으로 계산될 수 있다.

[수학식 11]

$$\tilde{\lambda}_m(c, \tau) = \frac{1}{C} \psi_m(\tau; \tau - 1) \kappa_m(c, \tau; \tau - 1) \sum_{f \in \Omega_c} |Y_m(f, \tau)|^2$$

$$\psi_m(\tau; \tau - 1) = \beta \left[\sum_{c=1}^C \left(\frac{\sum_{f \in \Omega_c} |Y_m(f, \tau)|^2}{\tilde{\lambda}_m(c, \tau - 1)} \right)^\gamma \right]^{\beta-1}$$

$$\kappa_m(c, \tau; \tau - 1) = \gamma \left(\frac{\sum_{f \in \Omega_c} |Y_m(f, \tau)|^2}{\tilde{\lambda}_m(c, \tau - 1)} \right)^{\gamma-1}$$

본 발명에 따른 타겟 음성신호 추출 방법에서는, 출력채널(OC)을 통해서 출력되는 출력신호(O_S) 및 더미채널(DC)을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스(DM) 및 방향벡터(DV)에 대한 일정한 추가조건을 만족시킴, 주요화최소화(Majorization-Minimization) 기법에 기반한 어그질러리 함수를 비용함수로 설정하고, 비용함수(Q)에 기초하여 디믹싱 매트릭스(DM)를 업데이트함으로써 타겟 음원신호에 대한 무왜

곡 조건을 걸어 음원신호 왜곡문제를 해결할 수 있을 뿐만 아니라, 온라인 프로세싱이 가능할 수 있다.

[0160] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 타겟 음성신호 추출 장치(10)는 복수의 마이크들(100), 필터 초기화부(210), 필터 학습부(230) 및 출력부(300)를 포함할 수 있다. 복수의 마이크들(100)은 외부 입력신호(IN_S)를 수신할 수 있다. 필터 초기화부(210)는 채널 수에 상응하는 항등행렬에서 제1 채널(CH1)을 출력채널(OC)로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널(DC)로 설정하여 디믹싱 매트릭스(DM)를 초기화(INI_D)할 수 있다. 필터 학습부(230)는 출력채널(OC)을 통해서 출력되는 출력신호(O_S) 및 더미채널(DC)을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스(DM) 및 방향벡터(DV)에 대한 일정한 추가조건을 만족시키는 비용함수(Q)를 설정하고, 비용함수(Q)에 기초하여 디믹싱 매트릭스(DM)를 업데이트할 수 있다. 출력부(300)는 디믹싱 매트릭스(DM) 및 외부 입력신호(IN_S)에 기초하여 타겟 소스에 상응하는 출력신호(O_S)를 제공할 수 있다. 일 실시예에 있어서, 제1 채널(CH1)에 대해서 디믹싱 매트릭스(DM)의 첫번째 행 벡터와 방향벡터(DV)의 곱은 1이고, 더미채널(DC)에 대해서 디믹싱 매트릭스(DM)의 N 번째 행 벡터(N은 1보다 큰 자연수)와 방향벡터(DV)의 곱은 0일 수 있다.

[0161] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 타겟 음성신호 추출 장치(10)는 복수의 마이크들(100), 필터 초기화부(210), 필터 학습부(230) 및 출력부(300)를 포함할 수 있다. 복수의 마이크들(100)은 외부 입력신호(IN_S)를 수신할 수 있다. 필터 초기화부(210)는 채널 수에 상응하는 항등행렬에서 제1 채널(CH1)을 출력채널(OC)로 설정하고, 상기 제1 채널을 제외한 나머지 채널들에서 타겟 음원신호를 제거한 더미신호를 제공하는 더미채널(DC)로 설정하여 디믹싱 매트릭스(DM)를 초기화(INI_D)할 수 있다. 필터 학습부(230)는 출력채널(OC)을 통해서 출력되는 출력신호(O_S) 및 더미채널(DC)을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스(DM) 및 방향벡터(DV)에 대한 일정한 추가조건을 만족시키며, 주요화최소화(Majorization-Minimization) 기법에 기반한 어그질러리 함수를 비용함수로 설정하고, 비용함수(Q)에 기초하여 디믹싱 매트릭스(DM)를 업데이트할 수 있다. 출력부(300)는 디믹싱 매트릭스(DM) 및 외부 입력신호(IN_S)에 기초하여 타겟 소스에 상응하는 출력신호(O_S)를 제공할 수 있다.

[0162] 본 발명에 따른 타겟 음성신호 추출 장치(10)에서는, 출력채널(OC)을 통해서 출력되는 출력신호(O_S) 및 더미채널(DC)을 통해서 출력되는 더미신호 간의 독립성이 최대화되고, 디믹싱 매트릭스(DM) 및 방향벡터(DV)에 대한 일정한 추가조건을 만족시키며, 주요화최소화(Majorization-Minimization) 기법에 기반한 어그질러리 함수를 비용함수로 설정하고, 비용함수(Q)에 기초하여 디믹싱 매트릭스(DM)를 업데이트함으로써 타겟 음원신호에 대한 무왜곡 조건을 걸어 음원신호 왜곡문제를 해결할 수 있을 뿐만 아니라, 온라인 프로세싱이 가능할 수 있다.

[0163] 도 8은 본 발명의 일 실시예에 따른 출력신호에 따라 계산되는 분산에 기초하여 디믹싱 매트릭스를 업데이트하는 과정을 설명하기 위한 도면이다.

[0164] 도 1 내지 8을 참조하면, 디믹싱 매트릭스(DM)는 초기화될 수 있다(S120). 본 발명에 따른 타겟 음성신호 추출 장치(10)는 디믹싱 매트릭스(DM)에 기초하여 출력신호(O_S)를 산출할 수 있다(S220). 출력신호에 기초하여 분산을 계산하고, 출력신호 및 분산에 기초하여 가중치 함수를 계산할 수 있다(S321, S322). 출력신호의 스케일을 조정하여 스케일 조정된 출력신호에 기초하여 방향벡터를 계산할 수 있다(S323, S324). 출력신호(O_S)에 따라 계산된 분산 및 방향벡터에 기초하여 디믹싱 매트릭스(DM)를 업데이트할 수 있다(S430). 이후, 디믹싱 매트릭스(DM)에 기초하여 출력신호(O_S)가 산출될 수 있고, 수렴여부를 판단할 수 있다(S530, S630).

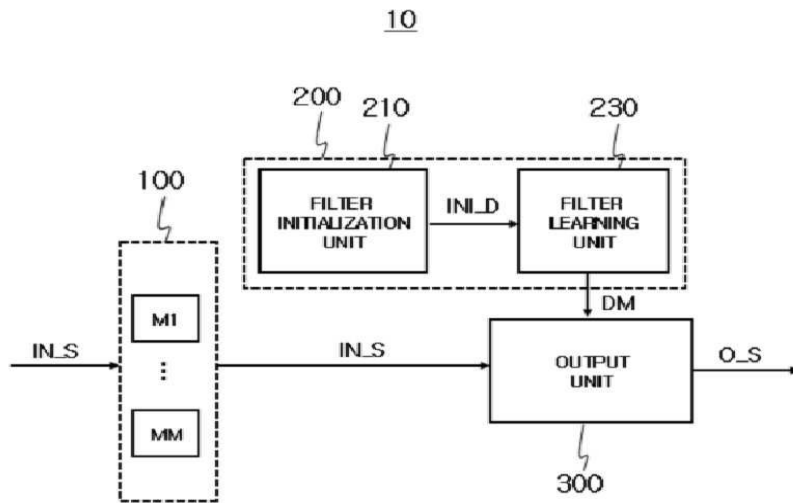
[0165] 위에서 언급된 본 발명의 기술적 과제 외에도, 본 발명의 다른 특징 및 이점들이 이하에서 기술되거나, 그러한 기술 및 설명으로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

부호의 설명

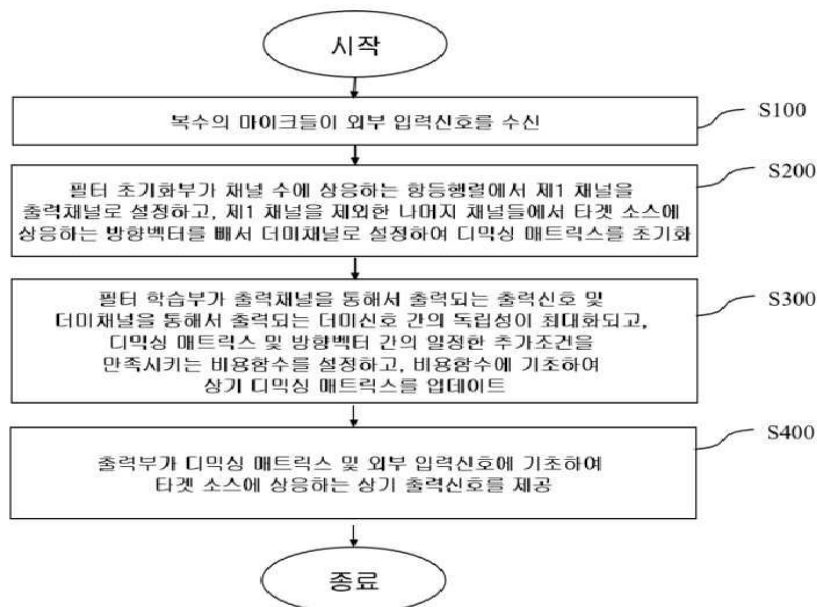
[0166] 10: 타겟 음성신호 추출 장치 100: 복수의 마이크들
200: 필터부 300: 출력부
210: 필터 초기화부 230: 필터 학습부

도면

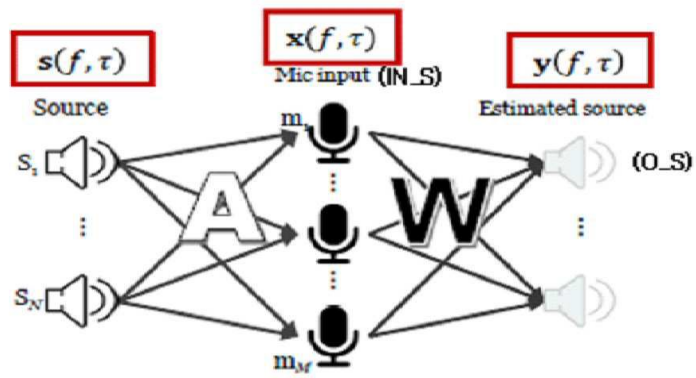
도면1



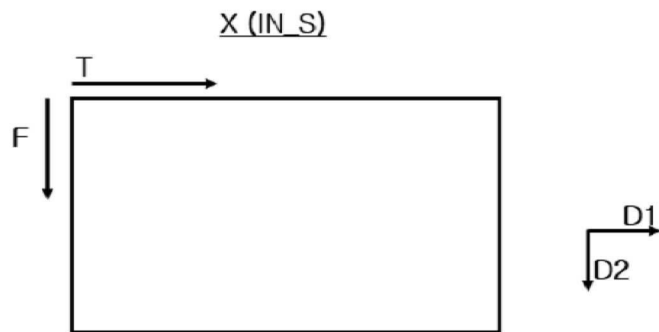
도면2



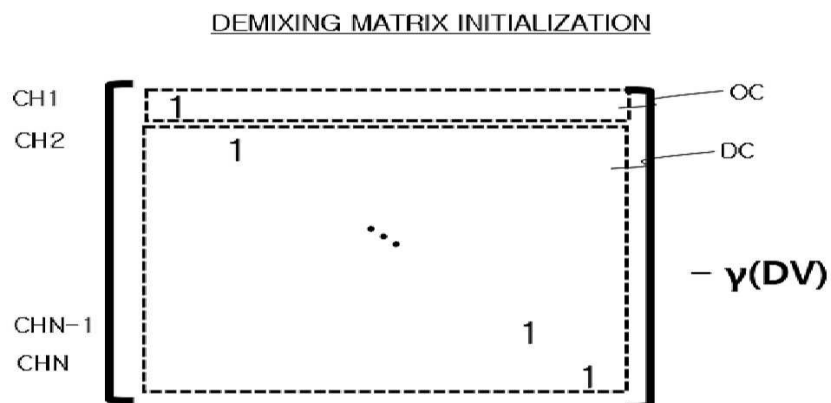
도면3



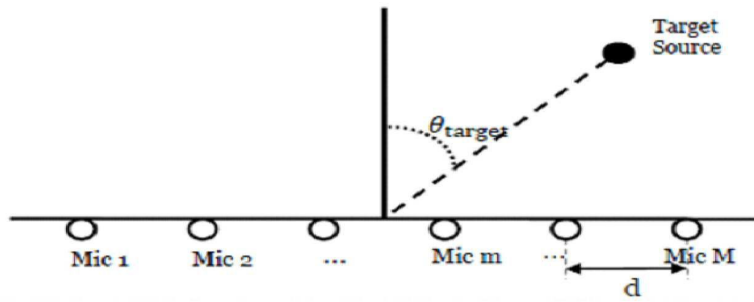
도면4



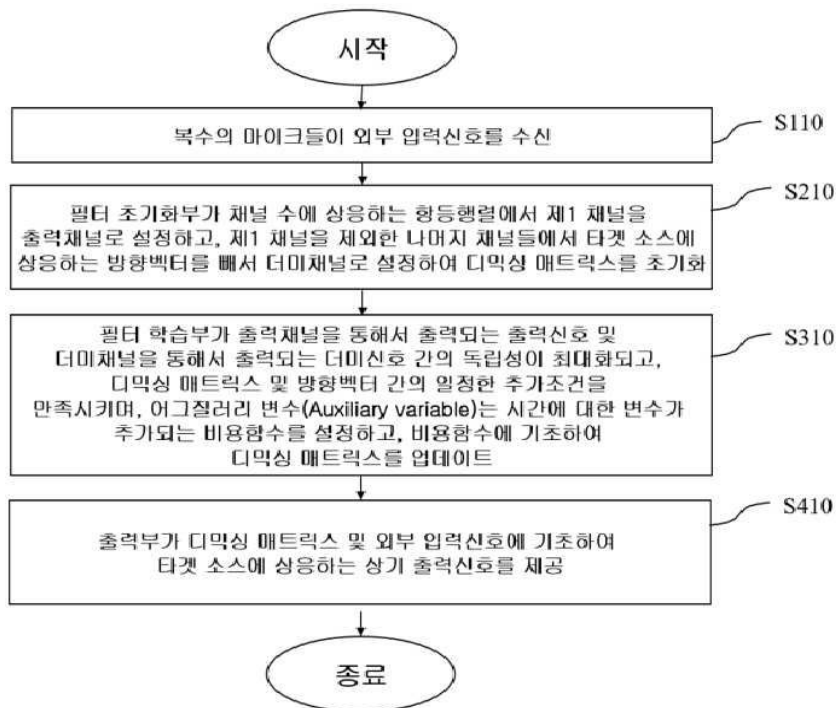
도면5



도면6



도면7



도면8

