



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0010356
(43) 공개일자 2010년02월01일

(51) Int. Cl.

G10L 11/00 (2010.01)

(21) 출원번호 10-2008-0071287

(22) 출원일자 2008년07월22일

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

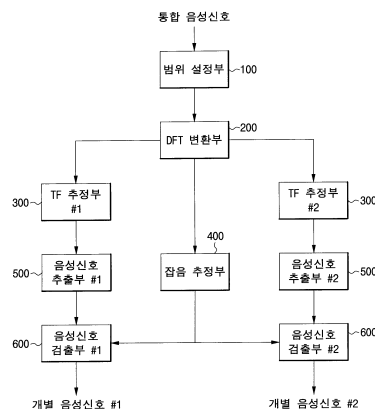
전체 청구항 수 : 총 18 항

(54) 빔포밍 기술을 이용한 음원 분리 방법 및 시스템

(57) 요약

본 발명에 따른 빔포밍 기술을 이용한 음원 분리 방법 및 시스템은 마이크 어레이를 통해 적어도 하나 이상의 음성이 포함된 통합 음성신호가 입력되면 그 통합 음성신호에 대하여 기 설정된 방향으로 hanning 윈도우를 씌어 프레임으로 구분하는 windowing 처리부; 프레임별로 구분된 각각의 개별 음성신호를 주파수축으로 변환하는 DFT 변환부; 주파수 축으로 변환된 프레임에 대한 임펄스 응답을 구하여 개별 음성신호의 트랜스퍼 평선(Transfer Function)을 추정하는 적어도 하나 이상의 TF 추정부; 통합 음성신호에서 개별 음성신호의 트랜스퍼 평선들을 제거하여 잡음 신호를 추정하는 잡음 추정부; 개별 음성신호들 중 추출하고자 하는 개별 음성신호를 제외한 나머지 개별 음성신호를 통합 음성신호에서 제거하는 적어도 하나 이상의 음성신호 추출부; 및 검출하고자 하는 개별 음성신호의 트랜스퍼 평선에서 잡음 추정부를 통해 제공되는 잡음 부분을 제거하여 잡음이 제거된 개별 음성신호를 추출하는 적어도 하나 이상의 음성신호 검출부를 포함하는 것으로, 적어도 하나 이상의 음원이 동시에 입력되더라도 그 음원들을 각각 분리하여 별도로 저장 관리하거나 최초 음원을 저장 관리할 수 있는 이점이 있다.

대표도 - 도3



(72) 발명자

배정훈

서울특별시 중구 신당6동 58-44

이택진

서울특별시 관악구 신림4동 472-33

특허청구의 범위

청구항 1

서로 다른 하나이상의 음원을 분리하는 시스템에 있어서,

빔포밍된 적어도 하나 이상의 마이크 어레이를 통해 입력된 통합 음성신호에 윈도우를 씌우는 windowing 처리부;

상기 windowing 처리부를 통해 윈도우가 씌워진 신호를 주파수축으로 변환하는 DFT 변환부;

윈도우가 씌워진 신호로부터 상기 서로 다른 둘 이상의 개별 음성 신호의 특징 값을 갖는 트랜스퍼 평선을 추정하는 TF 추정부;

상기 TF 추정부를 통해 추정된 서로 다른 둘 이상의 개별 음성신호의 특징값을 갖는 트랜스퍼 평선으로부터 개별 음성 신호들의 잡음을 제거하는 잡음 추정부; 및

잡음 제거된 음성 신호로부터 상기 서로 다른 개별 음성신호를 추출하는 음성신호 검출부를 포함하는 빔포밍 기술을 이용한 음원 분리 시스템.

청구항 2

제 1항에 있어서,

상기 트랜스퍼 평선 추정부는,

DFT 변환부를 통해 변환된 값을 통해 얻어진 임펄스 응답을 이용해서 트랜스퍼 평선을 추정하는 것을 특징으로 하는 빔포밍 기술을 이용한 음원 분리 시스템.

청구항 3

제 1항에 있어서,

상기 DFT 변환부는,

Discrete 푸리에 변환(DFT)임을 특징으로 하는 빔포밍 기술을 이용한 음원 분리 시스템.

청구항 4

제 1항에 있어서,

상기 푸리에 변환부를 통해 변환된 값으로부터 음성신호의 트랜스퍼 평선을 추정하기 위한 TF 추정부는,

서로 다른 음원 수와 같은 것을 특징으로 하는 빔포밍 기술을 이용한 음원 분리 시스템.

청구항 5

제 1항에 있어서,

상기 시스템은,

상기 TF 추정부를 통해 제공되는 개별 음성신호들 중 추출하고자 하는 개별 음성신호를 제외한 나머지 개별 음성신호를 상기 DFT 변환부를 통해 제공되는 통합 음성신호에서 제거하는 적어도 하나 이상의 음성신호 추출부를 더 포함하는 빔포밍 기술을 이용한 음원 분리 시스템.

청구항 6

제 1항에 있어서,

상기 windowing 처리부는,

hanning 윈도우의 길이는 32ms이며, 이동구간은 16ms인 것을 특징으로 하는 빔포밍 기술을 이용한 음원 분리 시스템.

청구항 7

제 6항에 있어서,

상기 TF 추정부는,

기 설정된 방향의 음성신호에 대하여 임의의 시간동안 마이크 어레이 사이의 임펄스 응답(impulse response)을 구하여 트랜스퍼 기능(transfer function; TF)을 추정하는 것을 특징으로 하는 빔포밍 기술을 이용한 음원 분리 시스템.

청구항 8

제 7항에 있어서,

상기 음성신호 검출부는,

주파수 축의 개별 음성신호를 시간 축 개별 음성신호로 변환하는 IDFT 변환부를 더 포함하는 빔포밍 기술을 이용한 음원 분리 시스템.

청구항 9

제 8항에 있어서,

개별 음성신호의 트랜스퍼 기능(Transfer Function)을 검출한 후 해당 방향으로 신호를 보낸 후 신호의 값의 크기를 비교하여 신호이득이 1이 나오면 정확하게 원하는 방향의 신호를 받아들인 것으로 판단하고 1이 아닌 다른 값이 나오면 개별 음성신호의 트랜스퍼 기능(Transfer Function)이 부정확하다고 판단하는 TF 정확성 확인부를 더 포함하는 빔포밍 기술을 이용한 음원 분리 시스템.

청구항 10

빔포밍 기술을 이용하여 서로 다른 하나이상의 음원을 음원 분리 방법에 있어서,

빔포밍된 적어도 하나 이상의 마이크 어레이를 통해 입력된 통합 음성신호에 윈도우를 씌우는 windowing 처리 단계;

상기 windowing 처리 단계를 통해 윈도우가 씌워진 신호를 주파수축으로 변환하는 DFT 변환 단계;

윈도우가 씌워진 신호로부터 상기 서로 다른 둘 이상의 개별 음성 신호의 특징 값을 갖는 트랜스퍼 평선을 추정하는 TF 추정 단계;

상기 TF 추정 단계를 통해 추정된 서로 다른 둘 이상의 개별 음성신호의 특징값을 갖는 트랜스퍼 평선으로부터 개별 음성 신호들의 잡음을 제거하는 잡음 추정 단계; 및

잡음 제거된 음성 신호로부터 상기 서로 다른 개별 음성신호를 추출하는 음성신호 검출 단계를 포함하는 빔포밍 기술을 이용한 음원 분리 방법.

청구항 11

제 10항에 있어서,

상기 트랜스퍼 평선 추정 단계는,

DFT 변환부를 통해 변환된 값을 통해 얻어진 임펄스 응답을 이용해서 트랜스퍼 평선을 추정하는 것을 특징으로 하는 빔포밍 기술을 이용한 음원 분리 방법.

청구항 12

제 10항에 있어서,

상기 DFT 변환 단계는,

Discrete 푸리에 변환(DFT)임을 특징으로 하는 빔포밍 기술을 이용한 음원 분리 방법.

청구항 13

제 10항에 있어서,
상기 TF 추정 단계는,
서로 다른 음원 수 만큼 수행하는 것을 특징으로 하는 빔포밍 기술을 이용한 음원 분리 시스템.

청구항 14

제 10항에 있어서,
상기 방법은,
상기 TF 추정 단계를 통해 제공되는 개별 음성신호들 중 추출하고자 하는 개별 음성신호를 제외한 나머지 개별 음성신호를 상기 DFT 변환 단계를 통해 제공되는 통합 음성신호에서 제거하는 음성신호 추출 단계를 더 포함하는 빔포밍 기술을 이용한 음원 분리 방법.

청구항 15

제 10항에 있어서,
상기 windowing 처리 단계는,
hanning 윈도우의 길이는 32ms이며, 이동구간은 16ms인 것을 특징으로 하는 빔포밍 기술을 이용한 음원 분리 방법.

청구항 16

제 15항에 있어서,
상기 TF 추정 단계는,
기 설정된 방향의 음성신호에 대하여 임의의 시간동안 마이크 어레이 사이의 임펄스 응답(impulse response)을 구하여 트랜스퍼 기능(transfer function; TF)을 추정하는 것을 특징으로 하는 빔포밍 기술을 이용한 음원 분리 방법.

청구항 17

제 16항에 있어서,
상기 음성신호 검출 단계는,
주파수 축의 개별 음성신호를 시간 축 개별 음성신호로 변환하는 IDFT 처리 단계를 더 포함하는 빔포밍 기술을 이용한 음원 분리 방법.

청구항 18

제 17항에 있어서,
개별 음성신호의 트랜스퍼 기능(Transfer Function)을 검출한 후 해당 방향으로 신호를 보낸 후 신호의 값의 크기를 비교하여 신호이득이 1이 나오면 정확하게 원하는 방향의 신호를 받아들인 것으로 판단하고 1이 아닌 다른 값이 나오면 개별 음성신호의 트랜스퍼 기능(Transfer Function)이 부정확하다고 판단하는 TF 정확성 확인 단계를 더 포함하는 빔포밍 기술을 이용한 음원 분리 방법.

명세서

발명의 상세한 설명

기술 분야

본 발명은 음성통신/인식에 필요한 음원분리에 관한 것이다.

[0001]

[0002] 여기서 음원 분리란 입력장치(마이크 어레이)에 적어도 두 개 이상의 음원이 동시에 입력되는 경우 이를 각각 분리하는 것을 말한다.

배경 기술

[0003] 종래 마이크 어레이 어레이를 이용한 잡음 제거 시스템은 적어도 하나 이상의 마이크 어레이, 마이크 어레이에 각각 연결된 단 구간 분석부, 에코 제거부, 방향성 잡음 제거와, 정면소리 유/무를 기반으로 필터 weight update on/off하는 적응 빔포밍(adaptive beamforming) 처리부, 마이크간의 신호의 상관관계를 이용하여 정면소리를 정면소리 검출부, 잔여잡음 제거와 정면소리 유/무를 기반으로 잔여 잡음을 제어하는 포스트 필터링부(Post-filtering) 및 오버랩 및 추가(Overlap and add) 처리부를 포함한다.

[0004] 빔포밍(Beamforming)은 마이크 어레이를 사용하는 경우 각 마이크에 들어오는 신호의 차이 때문에 각도에 따른 입력신호의 이득이 달라진다. 이것은 각도에 따라서 방향성 패턴(directionality pattern)을 형성한다.

[0005] 도 1은 마이크 어레이 정면이 90도인 경우의 방향성 패턴(directionality pattern)이다.

[0006] 이는 하기 [수학식 1]과 같다.

[0007] [수학식 1]

$$D(f, \alpha_x) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e^{j2\pi \alpha_x n d},$$

[0008]

[0009] 여기서, f = 주파수, N = 마이크의 개수, d = 마이크 사이의 간격,

[0010] $w_n(f) = a_n(f) e^{j\varphi_n(f)}$, $a_n(f)$ 는 amplitude weight이고 $\varphi_n(f)$ 는 phase weight이다.

[0011] 따라서 빔포밍(Beamforming) 기술은 마이크 어레이를 사용함에 따라 생기는 방향성 패턴(directionality pattern)을 $a_n(f)$ 과 $\varphi_n(f)$ 을 조절함으로써 원하는 각도의 방향으로 바꿔주는 것이다.

[0012] 이러한 방법을 통해서 원하는 방향의 신호만을 받을 수 있다.

[0013] 이후, FDBSS(frequency domain blind source separation, 이하 "FDBSS"라 함) 기술을 수행한다.

[0014] FDBSS 기술은 두 개의 음원이 섞여서 들어오는 경우에 이를 분리하기 위한 기술로서 주파수축에서 수행된다. 주파수 축에서 하는 경우 알고리즘 구현이 보다 간단해지고 계산시간이 줄어드는 장점이 있다.

[0015] 입력된 두 개의 음원이 섞인 신호는 short-time Fourier transform(STFT)으로 주파수축으로 바뀐다. 그 후 independent component analysis(ICA)의 3개의 단계를 통해서 음원이 분리된 신호로 바뀐다.

[0016] 우선, 1단계는 linear transformation이다.

[0017] 마이크의 수가 음원의 수보다 많은 경우 transformation(V)을 통해서 입력신호의 차원을 음원의 차원으로 줄여주는 단계이다. 일반적으로 음원의 수보다 마이크의 수가 많기 때문에 범위 분류(dimension reduction) 부분이 ICA에 포함된다.

[0018] 2단계는 1단계에서 처리된 신호에 단일 행렬(unitary matrix)(B)를 곱해줌으로써 분리된 신호의 주파수 영역의 값을 구한다.

[0019] 3단계는 1단계와 2단계를 통하여 구한 분리 행렬(separation matrix)(V*B)를 연구를 통해 나타난 법칙(learning rule)을 이용하여 점진적으로 향상시킨다.

[0020] 위의 과정을 통해서 분리된 신호를 얻은 후 위치추정(localization)을 취해준다.

[0021] 위치추정은 ICA에 의해 분리된 음원이 어떤 방향에서 들어온 소리인지를 구분한다.

[0022] 다음 단계는 교환(permutation)이다.

[0023] 이 단계에서는 앞에서 분리한 음원의 방향이 바뀌지 않고 유지되도록 해주 는 역할을 한다.

- [0024] 그 다음은 마지막 단계로서 scaling&smoothing이다.
- [0025] 음원 분리된 신호의 크기가 왜곡되지 않도록 조정해주는 단계가 크기이다. 음원분리를 위해 사용한 분리 행렬(separation matrix)의 pseudo inverse를 계산함으로써 해결 할 수 있다.
- [0026] FDBSS에서 f_s/L (f_s : sampling frequency)의 간격을 갖는 L 포인트(point)들로 샘플링(sampling)된 주파수 응답(frequency response)들은 시간 축에서 주기 L/f_s 를 갖는 주기신호들로 표현된다.
- [0027] 이것은 periodic infinite-length filter로서 현실적이지 않다.
- [0028] 그래서 일반적으로 시간 축에서 신호가 하나의 주기를 갖는 필터를 사용한다.
- [0029] 이 필터를 사용하는 경우 신호(signal)에 손실이 생기고 분리(separation)의 성능이 낮아진다.
- [0030] 이것은 해결하기위해서 smoothing 단계가 필요하다.
- [0031] 이 단계에서 양 끝이 0으로 점점 부드럽게 가까워지는 Hanning window를 곱해줌으로써 주파수 응답이 부드럽게 하여 신호의 손실이 줄고 분리의 성능이 향상된다.
- [0032] 이런 방법을 통해서 음원 분리를 하는 것이 FDBSS이다.
- [0033] 그러나 종래 빔포밍(Beamforming) 기술은 마이크 어레이의 방향성 패턴(directionality pattern)을 조정함으로써 원하는 방향의 신호만을 받을 수 있는 방법이지만 그 방향 주변에 다른 음원이 있는 경우 성능이 하락되는 문제가 있었다. 종래 빔포밍(Beamforming) 기술은 어느 정도 원하는 방향으로의 방향성 패턴(directionality pattern) 조절은 가능하지만 정확히 우리가 원하는 방향을 향해 뽀족하게 나오도록 하는 것은 어렵기 때문이다.
- [0034] 그리고 FDBSS 기술은 음원의 개수, reverberation, 사용자 위치이동 같은 제약조건에 성능 차이가 나는 문제가 있다. 또한 음성인식 관점에서 missing feature compensation의 보상이 필요한 문제점이 있었다.
- [0035] 두 사람이 동시에 말하는 경우에 신호가 섞이게 되면서 음성인식의 성능이 현저하게 떨어지게 되는 문제점이 있었다.

발명의 내용

해결 하고자하는 과제

- [0036] 따라서 본 발명은 상기와 같은 종래의 문제점을 해결하기 위한 것으로서, 본 발명의 목적은 두 개의 음성이 동시에 들어오는 상황에서 음성통신/인식기의 성능 향상을 위해 각각의 음성을 분리하는 빔포밍 기술을 이용한 음원 분리 방법 및 시스템을 제공하는 데 있다.

과제 해결수단

- [0037] 상기한 목적을 달성하기 위한 본 발명에 따른 빔포밍 기술을 이용한 음원 분리 시스템의 일 측면에 따르면, 빔포밍된 적어도 하나 이상의 마이크 어레이를 통해 입력된 통합 음성신호에 윈도우를 씌우는 windowing 처리부; 상기 windowing 처리부를 통해 윈도우가 씌워진 신호를 주파수축으로 변환하는 DFT 변환부; 윈도우가 씌워진 신호로부터 상기 서로 다른 둘 이상의 개별 음성 신호의 특징 값을 갖는 트랜스퍼 평선을 추정하는 TF 추정부; 상기 TF 추정부를 통해 추정된 서로 다른 둘 이상의 개별 음성신호의 특징값을 갖는 트랜스퍼 평선으로부터 개별 음성 신호들의 잡음을 제거하는 잡음 추정부; 및 잡음 제거된 음성 신호로부터 상기 서로 다른 개별 음성신호를 추출하는 음성신호 검출부를 포함한다.
- [0038] 상기 트랜스퍼 평선 추정부는, DFT 변환부를 통해 변환된 값을 통해 얻어진 임펄스 응답을 이용해서 트랜스퍼 평선을 추정한다.
- [0039] 그리고 상기 DFT 변환부는, Discrete 푸리에 변환(DFT)이다.
- [0040] 상기 TF 추정부는, 서로 다른 음원 수와 같다.
- [0041] 상기 시스템은, 상기 TF 추정부를 통해 제공되는 개별 음성신호들 중 추출하고자 하는 개별 음성신호를 제외한 나머지 개별 음성신호를 상기 DFT 변환부를 통해 제공되는 통합 음성신호에서 제거하는 적어도 하나 이상의 음성신호 추출부를 더 포함한다.

- [0042] 상기 windowing 처리부는, hanning 윈도우의 길이는 32ms이며, 이동구간은 16ms이다.
- [0043] 그리고 상기 TF 추정부는, 기 설정된 방향의 음성신호에 대하여 임의의 시간동안 마이크 어레이 사이의 임펄스 응답(impulse response)을 구하여 트랜스퍼 기능(transfer function; TF)을 추정한다.
- [0044] 또한 상기 음성신호 검출부는, 주파수 축의 개별 음성신호를 시간 축 개별 음성신호로 변환하는 IDFT 변환부를 더 포함한다.
- [0045] 개별 음성신호의 트랜스퍼 평선(Transfer Function)을 검출한 후 해당 방향으로 신호를 보낸 후 신호의 값의 크기를 비교하여 신호이득이 1이 나오면 정확하게 원하는 방향의 신호를 받아들인 것으로 판단하고 1이 아닌 다른 값이 나오면 개별 음성신호의 트랜스퍼 평선(Transfer Function)이 부정확하다고 판단하는 TF 정확성 확인부를 더 포함할 수 있다.
- [0046] 본 발명에 따른 빔포밍 기술을 이용한 음원 분리 방법의 일 측면에 따르면, 빔포밍된 적어도 하나 이상의 마이크 어레이를 통해 입력된 통합 음성신호에 윈도우를 씌우는 windowing 처리 단계; 상기 windowing 처리 단계를 통해 윈도우가 씌워진 신호를 주파수축으로 변환하는 DFT 변환 단계; 윈도우가 씌워진 신호로부터 상기 서로 다른 둘 이상의 개별 음성 신호의 특징 값을 갖는 트랜스퍼 평선을 추정하는 TF 추정 단계; 상기 TF 추정 단계를 통해 추정된 서로 다른 둘 이상의 개별 음성신호의 특징값을 갖는 트랜스퍼 평선으로부터 개별 음성 신호들의 잡음을 제거하는 잡음 추정 단계; 및 잡음 제거된 음성 신호로부터 상기 서로 다른 개별 음성신호를 추출하는 음성신호 검출 단계를 포함한다.
- [0047] 상기 트랜스퍼 평선 추정 단계는, DFT 변환부를 통해 변환된 값을 통해 얻어진 임펄스 응답을 이용해서 트랜스퍼 평선을 추정한다.
- [0048] 여기서, 상기 DFT 변환 단계는, Discrete 푸리에 변환(DFT)이다.
- [0049] 한편, 상기 TF 추정 단계는, 서로 다른 음원 수 만큼 수행한다.
- [0050] 또한 상기 방법은, 상기 TF 추정 단계를 통해 제공되는 개별 음성신호들 중 추출하고자 하는 개별 음성신호를 제외한 나머지 개별 음성신호를 상기 DFT 변환 단계를 통해 제공되는 통합 음성신호에서 제거하는 음성신호 추출 단계를 더 포함한다.
- [0051] 한편, 상기 windowing 처리 단계는, hanning 윈도우의 길이는 32ms이며, 이동구간은 16ms이다.
- [0052] 그리고 상기 TF 추정 단계는, 기 설정된 방향의 음성신호에 대하여 임의의 시간동안 마이크 어레이 사이의 임펄스 응답(impulse response)을 구하여 트랜스퍼 기능(transfer function; TF)을 추정한다.
- [0053] 또한 상기 음성신호 검출 단계는, 주파수 축의 개별 음성신호를 시간 축 개별 음성신호로 변환하는 IDFT 처리 단계를 더 포함한다.
- [0054] 개별 음성신호의 트랜스퍼 평선(Transfer Function)을 검출한 후 해당 방향으로 신호를 보낸 후 신호의 값의 크기를 비교하여 신호이득이 1이 나오면 정확하게 원하는 방향의 신호를 받아들인 것으로 판단하고 1이 아닌 다른 값이 나오면 개별 음성신호의 트랜스퍼 평선(Transfer Function)이 부정확하다고 판단하는 TF 정확성 확인 단계를 더 포함할 수 있다.

효 과

- [0055] 상술한 바와 같이 본 발명에 의한 빔포밍 기술을 이용한 음원 분리 방법 및 시스템에 의하면, 적어도 하나 이상의 음원이 동시에 입력되더라도 그 음원들을 각각 분리하여 별도로 저장 관리하거나 최초 음원을 저장 관리할 수 있는 뛰어난 효과가 있다.

발명의 실시를 위한 구체적인 내용

- [0056] 이하, 본 발명에 따른 빔포밍 기술을 이용한 음원 분리 방법 및 시스템에 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 이해해야 할 것이다.
- [0057] 도 2는 종래 마이크 어레이 어레이를 이용한 방향성 잡음 제거 시스템을 나타낸 도면으로써, 적어도 하나 이상의 마이크 어레이(10), 마이크 어레이(10)에 각각 연결된 단구간 분석부(20), 에코 제거부(30), 방향성 잡음 제

거와, 정면소리 유/무를 기반으로 필터 weight update on/off하는 adaptive 빔포밍(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.

[0058] 각 마이크 어레이(10)을 통해 입력된 음성들은 각각의 단구간 분석부(20)를 통해 주파수 영역을 분석한다.

[0059] 예를 들어 설명하면, 하나의 프레임은 2.56ms에 해당하며 이동구간은 128ms이다. 따라서 16Khz 샘플링에서 256ms는 4,096개로 샘플링되며, Hanning window를 사용할 수 있다.

[0060] 이후, DFT를 하기 위해 real-FFT를 이용하며, 소스코드는 ETSI 표준 특징 추출 프로그램을 이용한다.

[0061] 적응 빔포밍(adaptive Beamforming) 처리부(40)를 통해 방향성 잡음이 제거한다.

[0062] 적응 빔포밍 처리부(40)는 generalized sidelobe canceller(이하 "GSC"라 함)를 이용한다.

[0063] 이후, Far-end 신호가 스피커에 어레이로 도착하는 경로를 추정하여 에코를 제거하는 것과 마찬가지로이다.

[0064] 도 3은 본 발명에 따른 빔포밍 기술을 이용한 음원 분리 시스템의 구성을 나타낸 도면으로서, 본 발명에 따른 빔포밍 기술을 이용한 음원 분리 시스템은 적어도 하나 이상의 마이크 어레이(10), 마이크 어레이(10)에 각각 연결된 단구간 분석부(20), 에코 제거부(30), 정면소리 검출부(50), Post-filtering부(60) 및 Overlap and add 처리부(70)가 포함되며, 범위설정(windowing)부(100), DFT 변환부(200), 적어도 하나 이상의 TF 추정부(300), 잡음 추정부(400), 적어도 하나 이상의 음성신호 추출부(500) 및 적어도 하나 이상의 음성신호 검출부(600)를 포함하며, 음성신호 검출부(600)는 IDFT 변환부(610)를 포함한다.

[0065] 범위설정(windowing)부(100)는 마이크 어레이(10)를 통해 적어도 하나 이상의 음성이 포함된 통합 음성신호가 입력되면 그 통합 음성신호에 대하여 기 설정된 방향으로 hanning 윈도우를 씌어 프레임으로 구분한다. 이때, 범위설정부(100)는 단구간 분석부(20)와 에코 제거부(30)를 통해 마이크 어레이(10)를 통해 통합 음성신호를 제공받을 수도 있다.

[0066] 이때, 범위설정(windowing)부(100)는 hanning 윈도우의 길이는 32ms이며, 이동구간은 16ms이다.

[0067] 그리고 DFT 변환부(200)는 범위설정(windowing)부(100)를 통해 프레임별로 구분된 각각의 개별 음성신호를 주파수축으로 변환한다.

[0068] 또한, 적어도 하나 이상의 TF 추정부(300)는 DFT 변환부(200)를 통해 주파수 축으로 변환된 프레임에 대한 임펄스 응답을 구하여 개별 음성신호의 트랜스퍼 평선(Transfer Function)을 추정한다. 이때, TF 추정부(300)는 기 설정된 방향의 음성신호에 대하여 임의의 시간동안 마이크 어레이 사이의 임펄스 응답(impulse response)을 구하여 트랜스퍼 기능(transfer function; TF)을 추정한다.

[0069] 그리고 잡음 추정부(400)는 DFT 변환부(200)를 통해 주파수 축으로 변환된 통합 음성신호에서 각 TF 추정부(300)를 통해 추정된 트랜스퍼 평선(Transfer Function)을 통해 검출한 개별 음성신호들을 제거하여 잡음 신호를 추정한다.

[0070] 또한 적어도 하나 이상의 음성신호 추출부(500)는 TF 추정부(300)를 통해 제공되는 개별 음성신호들 중 추출하고자 하는 개별 음성신호를 제외한 나머지 개별 음성신호를 상기 DFT 변환부(200)를 통해 제공되는 통합 음성신호에서 제거한다.

[0071] 그리고 적어도 하나 이상의 음성신호 검출부(600)는 트랜스퍼 평선(Transfer Function)을 통해 검출하고자 하는 개별 음성신호에서 잡음 추정부(400)를 통해 제공되는 잡음 부분을 제거하여 잡음이 제거된 개별 음성신호를 추출한다. 이때, 음성신호 검출부(600)는 주파수 축의 개별 음성신호를 시간 축 개별 음성신호로 변환하는 IDFT 변환부(610)를 더 포함한다.

[0072] 상기한 각 구성들에 대한 일반적인 기능 및 각각의 상세한 동작에 대하여는 그 설명을 생략하고, 본 발명에 상응하는 동작 위주로 그 동작들을 설명하기로 한다.

[0073] 먼저, 마이크 어레이(10)는 두개의 음성신호가 섞인 통합 음성신호를 입력받아 범위설정(windowing)부(100)로 제공한다. 이때, 마이크 어레이(10)를 통해 입력되는 신호는 마이크 어레이(10) 사이의 간격 때문에 약간씩 차이가 있는 음성신호이다.

[0074] 그러면, 범위설정(windowing)부(100)는 제공된 통합 음성신호에 대하여 기 설정된 방향으로 hanning 윈도우를 씌어 32ms 구간의 프레임으로 구분한다. 이 과정에서 구분된 프레임은 16ms 구간만큼 이동시키면서 차르게된

다.

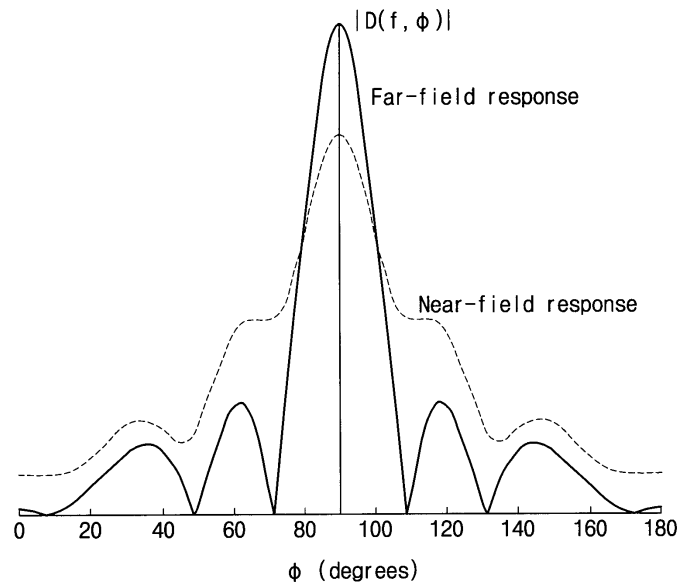
- [0075] 한편, 범위설정(windowing)부(100)는 hanning 윈도우를 씌우기 위한 방향은 기 설정되어 있으며, hanning 윈도우의 개수는 사람의 수에 따라 달라질 수 있는 것으로 이를 한정하지는 않는다.
- [0076] 이어서, DFT 변환부(200)는 범위설정(windowing)부(100)를 통해 프레임별로 구분된 각각의 개별 음성신호를 주파수축으로 변환한다.
- [0077] 이후, TF 추정부(300)는 DFT 변환부(200)를 통해 주파수 축으로 변환된 프레임에 대한 임펄스 응답을 구하여 개별 음성신호의 트랜스퍼 평선(Transfer Function)을 추정한다. 이때, TF 추정부(300)는 두 개의 개별 음성신호의 트랜스퍼 평선(Transfer Function)을 추정하거나, 두 개의 TF 추정부(300)를 통해 각각의 개별 음성신호의 트랜스퍼 평선(Transfer Function)을 추정할 수 있다. 이때, TF 추정부(300)는 기 설정된 방향의 음성신호에 대하여 임의의 시간동안 마이크 어레이 사이의 임펄스 응답(impulse response)을 구하여 트랜스퍼 기능(transfer function; TF)을 추정한다.
- [0078] 이렇게 TF 추정부(300) 또는 각각의 TF 추정부(300)를 통해 개별 음성신호의 트랜스퍼 평선(Transfer Function)이 추정되면, 잡음 추정부(400)는 DFT 변환부(200)를 통해 주파수 축으로 변환된 통합 음성신호에서 TF 추정부(300)를 통해 추정된 트랜스퍼 평선(Transfer Function)을 통해 검출한 개별 음성신호들을 제거하여 잡음 신호를 추정한다.
- [0079] 이후, 음성신호 추출부(500)는 TF 추정부(300)를 통해 제공되는 개별 음성신호의 트랜스퍼 평선(Transfer Function)들 중 추출하고자 하는 개별 음성신호의 트랜스퍼 평선(Transfer Function)을 제외한 나머지 개별 음성신호의 트랜스퍼 평선(Transfer Function)을 DFT 변환부(200)를 통해 제공되는 통합 음성신호에서 제거한다. 그러면, 추출하고자 하는 개별 음성신호를 추출할 수 있다.
- [0080] 이후, 적어도 하나 이상의 음성신호 검출부(600)는 트랜스퍼 평선(Transfer Function)을 통해 검출하고자 하는 개별 음성신호에서 상기 잡음 추정부(400)를 통해 제공되는 잡음 부분을 제거하여 잡음이 제거된 개별 음성신호를 추출한다. 이때, 음성신호 검출부(600)는 IDFT 변환부(610)를 통해 주파수 축의 개별 음성신호를 시간 축 개별 음성신호로 변환한다.
- [0081] 그러면, 상기와 같은 구성을 가지는 본 발명에 따른 빔포밍 기술을 이용한 음원 분리 방법에 대해서도 4를 참조하여 설명하기로 한다.
- [0082] 먼저, 마이크 어레이(10)를 통해 적어도 하나 이상의 음성이 포함된 통합 음성신호가 입력되면 기 설정된 방향으로 hanning 윈도우를 씌어 프레임으로 구분한다(S1). 여기서 windowing 처리 단계(S1)는 hanning 윈도우의 길이는 32ms이며, 이동구간은 16ms인 이다.
- [0083] 이어서, 프레임별로 구분된 각각의 개별 음성신호를 주파수축으로 변환한다(S2).
- [0084] 그런 후, 주파수 축으로 변환된 프레임에 대한 임펄스 응답을 구하여 개별 음성신호의 트랜스퍼 평선(Transfer Function)을 추정한다(S3). 한편, TF 추정 단계는 기 설정된 방향의 음성신호에 대하여 임의의 시간동안(5초) 마이크 어레이 사이의 임펄스 응답(impulse response)을 구하여 트랜스퍼 기능(transfer function; TF)을 추정한다.
- [0085] 이어서, 주파수 축으로 변환된 통합 음성신호에서 트랜스퍼 평선(Transfer Function)을 통해 검출한 개별 음성신호들을 제거하여 잡음 신호를 추정한다(S4).
- [0086] 이후, 개별 음성신호들 중 추출하고자 하는 개별 음성신호를 제외한 나머지 개별 음성신호를 통합 음성신호에서 제거한다(S5).
- [0087] 그러면, 트랜스퍼 평선(Transfer Function)을 통해 검출하고자 하는 개별 음성신호에서 잡음 부분을 제거하여 잡음이 제거된 개별 음성신호를 추출한다(S6). 한편 음성신호 검출 단계(S6)는 주파수 축의 개별 음성신호를 시간 축 개별 음성신호로 변환한다.
- [0088] 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 특허청구 범위에 속함은 당연한 것이다.

도면의 간단한 설명

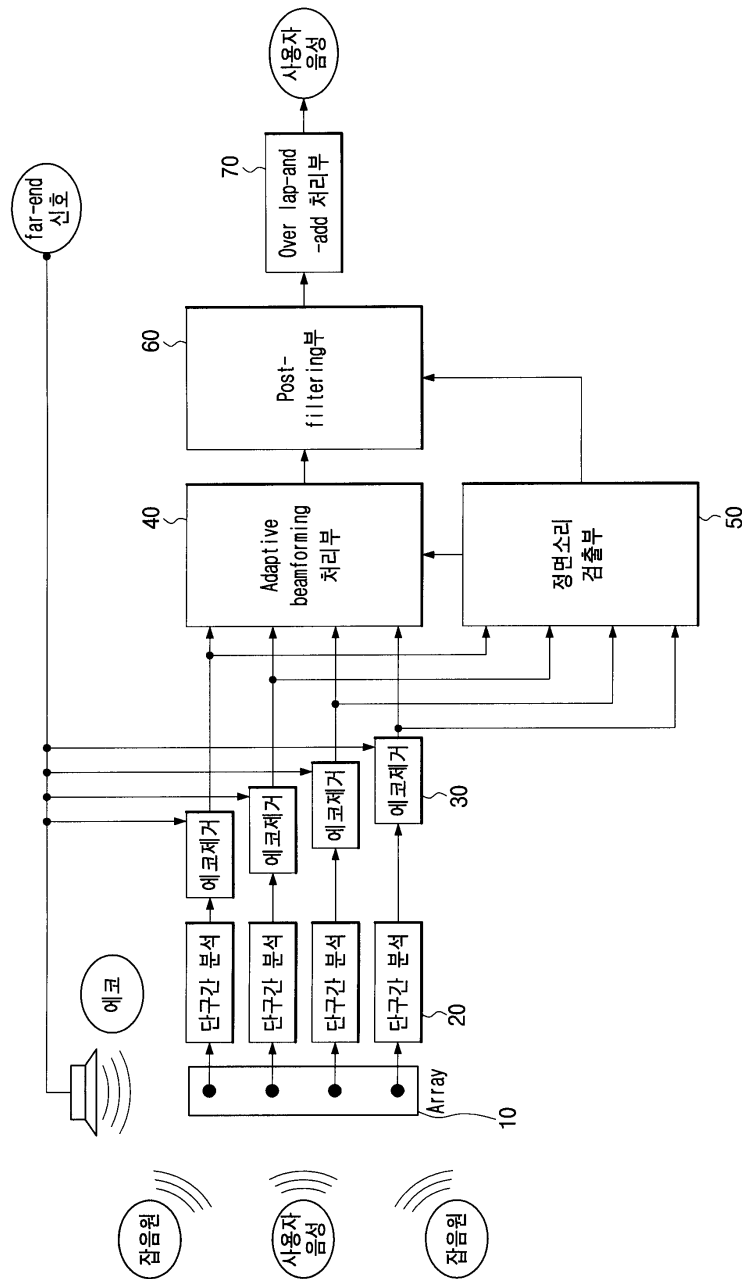
- [0089] 도 1은 종래 적응 빔포밍을 위한 사용자 방향의 소리 검출 시스템에서 마이크 어레이 정면이 90도인 경우의 방향성 패턴(directionality pattern)을 나타낸 도면.
- [0090] 도 2는 종래 적응 빔포밍을 위한 사용자 방향의 소리 검출 시스템을 나타낸 기능블록도.
- [0091] 도 3은 본 발명에 따른 빔포밍 기술을 이용한 음원 분리 시스템의 구성을 나타낸 기능블록도.
- [0092] 도 4는 본 발명에 따른 빔포밍 기술을 이용한 음원 분리 방법을 나타낸 순서도이다.
- [0093] <도면의 주요 부분에 대한 부호의 설명>
- [0094] 10 : 적어도 하나 이상의 마이크 어레이 20 : 단구간 분석부
- [0095] 30 : 에코 제거부 40 : adaptive beamforming 처리부
- [0096] 50 : 정면소리 검출부 60 : Post-filtering부
- [0097] 70 : Overlap and add 처리부
- [0098] 100 : 범위설정부 200 : DFT 변환부
- [0099] 300 : 적어도 하나 이상의 TF 추정부 400 : 잡음 추정부
- [0100] 500 : 적어도 하나 이상의 음성신호 추출부
- [0101] 600 : 적어도 하나 이상의 음성신호 검출부
- [0102] 610 : IDFT 변환부

도면

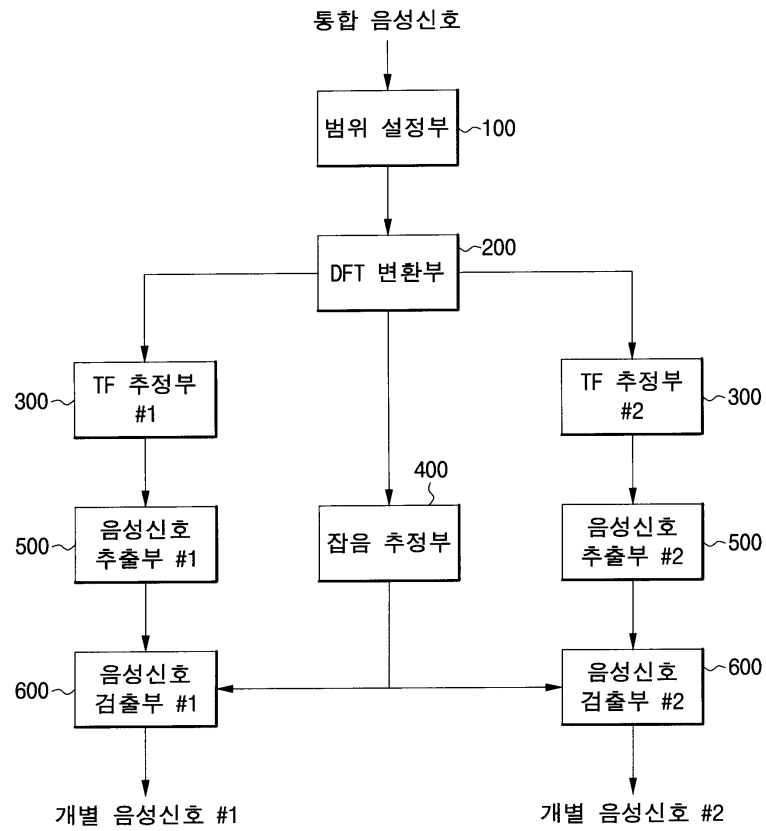
도면1



도면2



도면3



도면4

