(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2010년05월27일
(11) 등록번호 10-0959983
(24) 등록일자 2010년05월18일

(51) Int. Cl.
G10L 15/00 (2006.01) *G10L 15/20* (2006.01)
H04R 1/40 (2006.01) *H04R 3/00* (2006.01)
(21) 출원번호 10-2007-7026955
(22) 출원일자(국제출원일자) 2006년08월11일
심사청구일자 2007년11월20일
(85) 번역문제출일자 2007년11월20일
(65) 공개번호 10-2008-0009211
(43) 공개일자 2008년01월25일
(86) 국제출원번호 PCT/JP2006/315953
(87) 국제공개번호 WO 2007/018293
국제공개일자 2007년02월15일
(30) 우선권주장
JP-P-2005-00233195 2005년08월11일 일본(JP)
(56) 선행기술조사문헌
KR1020040035150 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
아사히 가세이 가부시키가이샤
일본 오사카후 오사카시 기타쿠 나카노시마 3쵸메 3방 23고
(72) 발명자
나가하마, 가쯔마사
일본 100-8440 도꾜도 지요다꾸 유라꾸쪼 1쪼메 1-2
마쯔이, 신야
일본 100-8440 도꾜도 지요다꾸 유라꾸쪼 1쪼메 1-2
(74) 대리인
양영준, 장수길

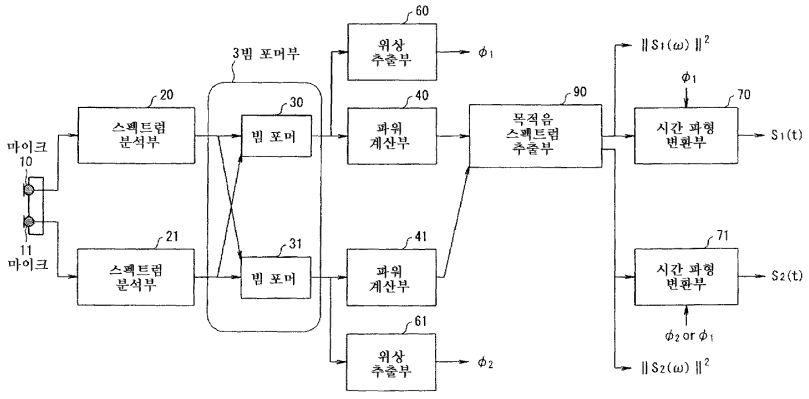
전체 청구항 수 : 총 13 항 심사관 : 신창우

(54) 음원 분리 장치, 음성 인식 장치, 휴대 전화기, 음원 분리방법, 및, 프로그램

(57) 요 약

마이크로폰 소자의 감도의 변동에 영향을 받지 않고, 복수의 음원으로부터 발하여진 음원 신호가 혼합된 혼합음 중에서, 목적 음원으로부터의 음원 신호를 분리하는 것을 가능하게 한다. 음원 분리 장치(1)의 빔 포머부(3)는, 스펙트럼 분석 후의 마이크로폰(10, 11)으로부터의 출력 신호에 대하여 복소공역의 관계에 있는 가중 계수를 승산함으로써, 2개의 마이크로폰(10, 11)을 연결하는 직선의 수선에 대칭인 방향으로부터 도래하는 음원 신호를 각각 감쇠시키기 위한 빔 포머 처리를 행한다. 파워 계산부(40, 41)는 파워 스펙트럼 정보를 계산하고, 목적음 스펙트럼 추출부(50, 51)는 파워 스펙트럼 정보끼리의 차분에 기초하여 목적 음원의 스펙트럼 정보를 추출한다.

대 표 도



특허청구의 범위

청구항 1

서로 이격하여 배치된 적어도 2개의 마이크로폰에 대하여 복수의 음원으로부터 발하여진 음원 신호가 혼합된 혼합음을 입력받고, 그 혼합음으로부터 목적 음원으로부터의 음원 신호를 분리하는 음원 분리 장치에 있어서,

상기 마이크로폰으로부터의 출력 신호에 대하여 제1 계수를 이용한 연산을 행함으로써, 소정의 방향으로부터 도래하는 음원 신호를 감쇠시키기 위한 제1 빔 포머 처리를 행하고,

또한,

상기 마이크로폰으로부터의 출력 신호에 대하여, 상기 제1 계수와 주파수 영역에서 복소공액의 관계에 있는 제2 계수를 이용한 연산을 행함으로써, 2개의 마이크로폰을 연결하는 직선의 수선에 대하여 상기 소정의 방향과 대칭인 방향으로부터 도래하는 음원 신호를 감쇠시키기 위한 제2 빔 포머 처리를 행하는 빔 포머 수단과,

상기 빔 포머 수단에 의해 얻어진 음원 신호 각각에 대한 파워 스펙트럼 정보를 계산하는 파워 계산 수단과,

상기 파워 계산 수단에 의해 계산된 파워 스펙트럼 정보끼리의 차분에 기초하여, 목적 음원의 스펙트럼 정보를 추출하는 목적음 스펙트럼 추출 수단

을 구비하는 것을 특징으로 하는 음원 분리 장치.

청구항 2

제1항에 있어서,

상기 빔 포머 수단은,

서로 이격하여 배치된 3개의 마이크로폰 중, 어느 2개의 마이크로폰의 조합 및 다른 2개의 마이크로폰의 조합 각각에 대해서, 상기 제1 빔 포머 처리 및 상기 제2 빔 포머 처리를 행하는 것을 특징으로 하는 음원 분리 장치.

청구항 3

제1항 또는 제2항에 있어서,

마이크로폰으로부터의 출력 신호에 대하여 지연을 부여하는 지향성 제어 수단을 더 구비하는 것을 특징으로 하는 음원 분리 장치.

청구항 4

제3항에 있어서,

상기 지향성 제어 수단은,

2개의 마이크로폰 중 적어도 한쪽의 마이크로폰으로부터의 출력 신호에 대하여 지연을 부여함으로써, 가상적으로 3개의 마이크로폰으로부터의 출력 신호를 생성하는 것을 특징으로 하는 음원 분리 장치.

청구항 5

제3항에 있어서,

상기 음원 신호가 도래하는 방향을 추정하는 도래 방향 추정 수단을 더 구비하고,

상기 지향성 제어 수단은,

상기 도래 방향 추정 수단에 의해 추정된 도래 방향에 기초하여, 2개의 음원의 위치가 가상적으로 2개의 마이크로폰을 연결하는 직선의 수선에 대하여 대칭으로 되도록, 그 마이크로폰으로부터의 출력 신호에 대하여 지연을 부여하는 것을 특징으로 하는 음원 분리 장치.

청구항 6

제1항 또는 제2항에 있어서,

상기 목적음 스펙트럼 추출 수단에 의해 추출된 파워 스펙트럼 정보에 대하여, 스펙트럼 서브트랙션 처리를 행하는 스펙트럼 서브트랙션 수단을 더 구비한 것을 특징으로 하는 음원 분리 장치.

청구항 7

제1항 또는 제2항에 있어서,

상기 빔 포머 수단에 의한 처리를 행하기 전에, 잡음을 저감하는 처리를 행하는 정상 잡음 저감 수단을 더 구비하는 것을 특징으로 하는 음원 분리 장치.

청구항 8

제1항 또는 제2항의 음원 분리 장치에 의해 분리된 음원 신호의 음성 인식을 행하는 음성 인식 수단을 구비한 음성 인식 장치.

청구항 9

제8항에 있어서,

차량의 운전석측으로부터 발하여지는 어휘의 후보의 일람인 운전석측 인식 어휘 리스트 및 조수석측으로부터 발하여지는 어휘의 후보의 일람인 조수석측 인식 어휘 리스트를 기억하는 인식 어휘 리스트 기억 수단을 더 구비하고,

상기 음성 인식 수단은,

상기 인식 어휘 리스트 기억 수단에 기억되어 있는 운전석측 인식 어휘 리스트 및 조수석측 인식 어휘 리스트에 기초하여, 상기 음원 분리 장치에 의해 분리된 음원 신호의 음성 인식 처리를 행하는 것을 특징으로 하는 음성 인식 장치.

청구항 10

제8항에 있어서,

현재의 차량의 상태를 관리하는 상태 천이 수단과,

차량의 상태에 따른 조수석측 및 운전석측에서의 유효한 어휘 리스트를 기억하는 유효 어휘 리스트 기억 수단과,

상기 상태 천이 수단에 의해 관리되고 있는 현재의 차량의 상태와, 상기 유효 어휘 리스트 기억 수단에 기억되어 있는 어휘 리스트에 기초하여, 상기 음성 인식 수단에 의해 인식된 어휘가 유효인지의 여부를 판단하고, 그 판단 결과에 따라서 제어를 행하는 제어 수단

을 더 구비하는 것을 특징으로 하는 음성 인식 장치.

청구항 11

제1항 또는 제2항의 음원 분리 장치를 구비한 휴대 전화기.

청구항 12

서로 이격하여 배치된 적어도 2개의 마이크로폰에 대하여, 복수의 음원으로부터 발하여진 음원 신호를 입력받는 음원 신호 수신 스텝과,

상기 마이크로폰으로부터의 출력 신호에 대하여, 주파수 영역에서 복소공액의 관계에 있는 2개의 가중 계수 각각을 이용하여 연산을 행함으로써, 2개의 마이크로폰을 연결하는 직선의 수선에 대하여 대칭인 소정의 방향으로부터 도래하는 음원 신호를 각각 감쇠시키기 위한 제1 빔 포머 처리 및 제2 빔 포머 처리를 행하는 빔 포머 처리 스텝과,

상기 빔 포머 처리 스텝에서 얻어진 음원 신호 각각에 대해서 파워 스펙트럼 정보를 산출하는 파워 계산 스텝과,

상기 파워 계산 스텝에서 계산된 파워 스펙트럼 정보끼리의 차분에 기초하여, 목적 음원의 스펙트럼 정보를 추출하는 목적음 스펙트럼 추출 스텝

을 구비하는 것을 특징으로 하는 음원 분리 방법.

청구항 13

컴퓨터에,

서로 이격하여 배치된 적어도 2개의 마이크로폰으로부터, 복수의 음원으로부터 발하여진 음원 신호가 혼합된 출력 신호를 취득하는 출력 신호 취득 스텝과,

상기 출력 신호 취득 스텝에서 취득된 출력 신호에 대하여, 주파수 영역에서 복소공액의 관계에 있는 2개의 가중 계수 각각을 이용하여 연산을 행함으로써, 2개의 마이크로폰을 연결하는 직선의 수선에 대하여 대칭인 소정의 방향으로부터 도래하는 음원 신호를 각각 감쇠시키기 위한 제1 빔 포머 처리 및 제2 빔 포머 처리를 행하는 빔 포머 처리 스텝과,

상기 빔 포머 처리 스텝에서 얻어진 음원 신호 각각에 대해서 파워 스펙트럼 정보를 산출하는 파워 계산 스텝과,

상기 파워 계산 스텝에서 계산된 파워 스펙트럼 정보끼리의 차분에 기초하여, 목적 음원의 스펙트럼 정보를 추출하는 목적음 스펙트럼 추출 스텝

를 실행시키기 위한 프로그램을 기록한 컴퓨터 판독 가능 기록 매체.

명 세 서

기 술 분 야

[0001] 본 발명은, 복수의 마이크로폰을 사용하고, 복수의 음원으로부터 발하여진 복수의 음성 신호나 각종 환경 잡음 등 복수의 음향 신호가 혼합된 신호로부터, 목적으로 하는 음원으로부터 도래하는 음원 신호를 분리하는 음원 분리 장치, 음성 인식 장치, 휴대 전화기, 음원 분리 방법, 및, 프로그램에 관한 것이다.

배 경 기 술

[0002] 다양한 환경 하에서 특정한 음성 신호 등을 수록하고자 하는 경우, 주위 환경에는 다양한 잡음원이 있기 때문에, 목적음으로 하는 신호만을 마이크로폰으로 수록하는 것은 곤란하여, 어떠한 잡음 저감 처리 혹은 음원 분리 처리가 필요로 된다.

[0003] 이들 처리가 특히 필요해지는 예로서, 예를 들면 자동차 환경 하를 들 수 있다. 자동차 환경 하에서, 휴대 전화의 보급에 의해 운전 중의 휴대 전화를 사용하는 통화는 차내에 이격하여 설치된 마이크를 사용하는 것이 일반적으로, 통화 품질을 현저하게 열화시키고 있다. 또한, 자동차 환경 하에서 운전 중에 음성 인식을 행하는 경우도 마찬가지의 상황에서 발화하기 때문에, 음성 인식 성능을 열화시키는 원인으로 되고 있다. 현재의 음성 인식 기술의 진보에 의해, 정상 잡음에 대한 음성 인식율의 열화의 문제에 대하여, 열화된 성능의 상당한 부분을 회복하는 것이 가능하게 되어 있다. 그러나, 현상의 음성 인식 기술로 대응이 어려운 것으로서, 복수발화자의 동시 발화 시의 인식 성능의 열화의 문제가 있다. 현재의 음성 인식의 기술로는 동시에 발화된 두 사람의 혼합 음성을 인식하는 기술이 낮기 때문에, 음성 인식 장치 사용시에는 발화자 이외의 동승자는 발화를 제한받아, 동승자의 행동을 제한하는 상황이 발생하고 있다. 이들 음원 분리의 방법으로서 독립 주성분 분석법 등을 사용하는 방법도 있지만, 연산량, 음원수의 변동 등의 문제가 있어 실용화되고 있지 않다.

[0004] 상기의 문제를 해결하기 위해서, 차내의 복수의 마이크를 부착하여 특정 방향으로부터의 음성만을 수록하는 방법이 각종 제안되어 있지만, 자동차 내에 다수의 마이크를 부착하기 위한 스페이스를 확보하는 것은 곤란하고, 또한 코스트적인 문제 때문에 특성을 갖춘 마이크를 사용하는 것도 어렵다. 이 때문에, 가능한 한 소수로 특성 변동이 있는 마이크를 사용하여 동작을 하는 방식이 요망되고 있다.

[0005] 일반적으로 복수의 마이크로폰을 사용하는 경우, 저코스트의 마이크로폰일수록 그 감도 특성의 변동이 커서 약 ±3dB 정도 각 주파수 특성에 변동이 있다고 말해지고 있다. 이들 특성의 변동이 마이크로폰 어레이 기술 중에서도 지연합 어레이 등의 가산형 어레이 처리에서는 마이크로폰 어레이 성능이 설계와 같은 특성을 실현할 수 없는 것뿐이지만, 적응 어레이 등의 소위 감산형 어레이에서는 하나의 마이크를 사용하는 경우보다 특히 1kHZ

이하 정도의 저역에서 성능을 악화시키는 경우가 있다.

[0006] 마이크로폰 어레이 기술에 있어서 센서로서의 마이크로폰의 특성의 변동은 큰 문제이며, 이들 대책으로서, 특허 문헌 1 내지 5 등에 복수의 마이크 소자의 감도를 맞추는 하는 방법이 제안되어 있다.

[0007] 종래, 소수의 마이크로폰수로 큰 효과가 얻어지는 적응 빔 포머 처리 기술을 이용한 마이크로폰 어레이에 관해서는, 예를 들면 비특허 문헌 1이나, 비특허 문헌 2에 기재되어 있는 바와 같이, 일반화 사이드로브 캔슬러(GSC), 프로스트형 빔 포머 및 참조 신호법 등, 다양한 방법이 알려져 있다.

[0008] 적응 빔 포머 처리는, 기본적으로는 잡음원의 방향으로 사각을 갖는 지향성 빔을 형성한 필터에 의해 잡음을 억압하는 처리이며, 그 중에서도 특히 일반화 사이드로브 캔슬러는, 비교적 성능이 좋은 것으로 알려져 있다. 그러나, GSC에서는, 설정된 목적 신호원의 방향과 어긋난 방향으로부터 목적 신호가 도래했을 때에, 목적 신호가 캔슬되어 열화된다고 하는 문제점이 있다. 이에 대하여, 특허 문헌 6, 7에서, 이것을 주파수 영역에서 동작시킴으로써 계산량을 삭감함과 함께, 주파수 영역의 필터 계수로부터 화자 방향과 특정 잡음 방향을 축차적으로 검지하고, 목적음과 목적음 이외의 잡음을 어느 정도 분리하고, 스펙트럼 서브트랙션과 병용함으로써 도래 방향 불분명한 잡음이나 확산성의 잡음을 저감시키는 방법이 개시되어 있다.

[0009] [특허 문헌 1] 일본 특개평5-131866호 공보

[0010] [특허 문헌 2] 일본 특개2002-99297호 공보

[0011] [특허 문헌 3] 일본 특개2003-153372호 공보

[0012] [특허 문헌 4] 일본 특개2004-343700호 공보

[0013] [특허 문헌 5] 일본 특개2004-289762호 공보

[0014] [특허 문헌 6] 일본 특개2001-100800호 공보

[0015] [특허 문헌 7] 일본 특개2000-47699호 공보

[0016] [비특허 문헌 1] 전자 정보 통신 학회편 「음향 시스템과 디지털 처리」

[0017] [비특허 문헌 2] HAykin저, "ADAptive Filter Theory(PrentiCe HAll)"

[0018] <발명의 개시>

[0019] [발명이 해결하고자 하는 과제]

[0020] 이 특허 문헌 6, 7에 기재되어 있는 바와 같은 적응 빔 포머와 스펙트럼 서브트랙션을 조합한 기술에서, 마이크로폰의 소자 감도의 변동이 존재하는 경우, 적응 필터부의 입력 신호로서 목적 신호를 억압한 참조 신호를 작성할 필요가 있지만, 목적 신호를 충분히 억압할 수 없어, 적응 필터부의 참조 신호와 목적 신호 쌍방에 목적 신호가 포함되기 때문에, 적응 필터 처리를 실시한 결과 목적 신호가 왜곡되어 음질이 열화된다. 마찬가지의 현상은, 초기 반사가 큰 폐공간에서도 참조 신호 내에 목적 신호가 누설되게 되어 발생한다.

[0021] 이것을 해결하기 위해서, 마이크로폰의 소자 감도 보정을 하기 위해서 제품 조립 시에 변동을 측정하고 보정 데이터를 작성하여 사용 시에 보정을 행하거나, 사용 시에 특정 방향으로부터의 참조 신호를 이용하여 마이크로폰의 소자 감도의 개체차, 주위 온도나 부품의 경년 변화에 의한 축차 변화를 측정하여 보정하는 방법이 있지만, 제조 코스트의 증가나, 참조 신호를 언제 재생시킬지, 혹은 수록한 신호가 정말로 참조 신호만인지의 여부를 어떻게 판단할지 등의 문제가 있다.

[0022] 또한 적응 필터의 동작에 주목하면, 일반적으로 적응 필터에서 참조 신호와 목표 신호가 높은 상관성을 갖는 경우, 적응 필터의 추정 동작이 잘 행해지지 않아 추정 정밀도를 현저하게 열화시키게 된다. 일례를 들면, 목적 신호와 잡음의 쌍방이 음성 신호인 경우 등이다.

[0023] 마찬가지의 문제는 에코 캔슬러에서도 발생하고, 원단 신호로부터 근단 신호에 혼입된 에코를 추정하는 적응 필터에서, 원단, 근단 모두 동시에 2명의 사람이 발화를 한 경우, 소위 더블 토크 상태가 발생하여 에코 캔슬러의 동작이 현저하게 열화된다. 이 때문에, 더블 토크 상태를 어떻게 정확하게 검출하고, 적응 필터의 파라미터 추정을 정지시키거나 추정 속도도 느리게 하거나 할지가 중요한 성능 유지의 조건으로 되어 있다.

[0024] 따라서, 본 적응 필터 동작에서도, 목적 신호와 특정 잡음이 쌍방 모두 음성인 경우, 쌍방 모두 발화하고 있는 상태를 검출할 필요가 있음과 동시에, 쌍방 모두 발화하고 있는 빈도가 높을수록 적응 필터 추정을 정지시키는

빈도가 높아져, 결과적으로 적응 필터의 잡음 제거 성능이 열화된다고 하는 문제가 있다.

[0025] 또한, 자동차의 주행 잡음 등 큰 확산성 잡음 등이 발생하는 상황에서는 확산성 잡음의 영향으로 적응 필터의 추정 정밀도가 열화되기 때문에, 결과적으로, 적응 필터가 추정한 필터 파라미터를 사용한 화자 방향, 특정 잡음 방향의 도래 방향 추정 정밀도가 열화되어, 시스템 전체의 성능이 열화된다고 하는 문제가 있다.

[0026] 또한, 상기의 특허 문헌 6, 7에 기재된 방식에서는, 자동차용 등에 저코스트의 마이크를 사용하여 잡음 성분 억압 장치를 구성하고자 한 경우, 차내 음원의 초기 반사가 크고, 확산성의 잡음 성분이 크며, 사용하는 마이크 소자의 변동이 크고(±3db 정도), 운전자와 동승자가 동시에 발화하여 상관성이 강한 목적음과 특정 잡음이 동시에 존재하는 등의 경우에, 적응 필터부가 원하는 동작을 하지 않아, 전체적으로 원하는 동작을 실현할 수 없다.

[0027] 본 발명은 상기 문제점을 감안하여 이루어진 것으로서, 마이크로폰 소자의 감도의 변동에 영향을 받지 않고, 복수의 음원으로부터 발하여진 음원 신호가 혼합된 혼합음 중에서, 목적 음원으로부터의 음원 신호를 분리하는 것을 가능하게 하는 음원 분리 장치, 음성 인식 장치, 휴대 전화기, 음원 분리 방법, 및, 프로그램을 제공하는 것을 목적으로 한다.

[0028] [과제를 해결하기 위한 수단]

[0029] 상기 과제를 해결하기 위해서, 청구항 1에 기재된 발명은, 서로 이격하여 배치된 적어도 2개의 마이크로폰에 대하여 복수의 음원으로부터 발하여진 음원 신호가 혼합된 혼합음을 입력받고, 그 혼합음으로부터 목적 음원으로부터의 음원 신호를 분리하는 음원 분리 장치로서, 상기 마이크로폰으로부터의 출력 신호에 대하여 제1 계수를 이용한 연산을 행함으로써, 소정의 방향으로부터 도래하는 음원 신호를 감쇠시키기 위한 제1 빔 포머 처리를 행하고, 또한, 상기 마이크로폰으로부터의 출력 신호에 대하여, 상기 제1 계수와 주파수 영역에서 복소공액의 관계에 있는 제2 계수를 이용한 연산을 행함으로써, 2개의 마이크로폰을 연결하는 직선의 수선에 대하여 상기 소정의 방향과 대칭인 방향으로부터 도래하는 음원 신호를 감쇠시키기 위한 제2 빔 포머 처리를 행하는 빔 포머 수단과, 상기 빔 포머 수단에 의해 얻어진 음원 신호 각각에 대한 파워 스펙트럼 정보를 계산하는 파워 계산 수단과, 상기 파워 계산 수단에 의해 계산된 파워 스펙트럼 정보끼리의 차분에 기초하여, 목적으로 하는 음원의 스펙트럼 정보를 추출하는 목적음 스펙트럼 추출 수단을 구비하는 것을 특징으로 하는 음원 분리 장치를 제공한다.

[0030] 본 발명에 따르면, 음원 분리 장치는, 지향 특성이 마이크로폰 소자의 감도에 영향을 받지 않는다고 하는 성질을 실현할 수 있고, 마이크로폰 소자의 감도의 변동에 영향을 받지 않아, 복수의 음원으로부터 발하여진 음원 신호가 혼합된 혼합음 중에서, 목적 음원으로부터의 음원 신호를 분리하는 것이 가능하게 된다.

[0031] 청구항 2에 기재된 발명은, 청구항 1에 기재된 음원 분리 장치에서, 상기 빔 포머 수단은, 서로 이격하여 배치된 3개의 마이크로폰 중, 어느 2개의 마이크로폰의 조합 및 다른 2개의 마이크로폰의 조합 각각에 대해서, 상기 제1 빔 포머 처리 및 상기 제2 빔 포머 처리를 행하는 것을 특징으로 한다.

[0032] 본 발명에 따르면, 3개의 마이크로폰 중, 2개의 마이크로폰을 연결하는 직선의 수선을 경계로 한 각 영역에 존재하는 목적 음원으로부터의 음원 신호를 추출할 수 있다.

[0033] 청구항 3에 기재된 발명은, 청구항 1 또는 청구항 2에 기재된 음원 분리 장치에서, 마이크로폰으로부터의 출력 신호에 대하여 지연을 부여하는 지향성 제어 수단을 더 구비하는 것을 특징으로 한다.

[0034] 본 발명에 따르면, 지연을 부여함으로써 지향 특성을 최적의 상태로 보정할 수 있어, 음원의 분리 성능을 향상시킬 수 있다.

[0035] 청구항 4에 기재된 발명은, 청구항 3에 기재된 음원 분리 장치에서, 상기 지향성 제어 수단은, 2개의 마이크로폰 중 적어도 한쪽의 마이크로폰으로부터의 출력 신호에 대하여 지연을 부여함으로써, 가상적으로 3개의 마이크로폰으로부터의 출력 신호를 생성하는 것을 특징으로 한다.

[0036] 본 발명에 따르면, 가상적으로 3개의 마이크로폰으로부터의 출력 신호를 생성할 수 있기 때문에, 2개의 마이크로폰만을 이용하여, 2개의 마이크로폰을 연결하는 직선의 수선 방향으로부터 도래하는 음원 신호를 분리 추출하는 것이 가능하게 된다.

[0037] 청구항 5에 기재된 발명은, 청구항 3 또는 청구항 4에 기재된 음원 분리 장치에서, 상기 음원 신호가 도래하는 방향을 추정하는 도래 방향 추정 수단을 더 구비하고, 상기 지향성 제어 수단은, 상기 도래 방향 추정 수단에

의해 추정된 도래 방향에 기초하여, 2개의 음원의 위치가 가상적으로 2개의 마이크로폰을 연결하는 직선의 수선에 대하여 대칭으로 되도록, 그 마이크로폰으로부터의 출력 신호에 대하여 지연을 부여하는 것을 특징으로 한다.

[0038] 본 발명에 따르면, 음원의 높은 분리 성능이 얻어지도록 지연 조작을 행하는 것이 가능하게 된다.

[0039] 청구항 6에 기재된 발명은, 청구항 1 내지 청구항 5 중 어느 한 항에 기재된 음원 분리 장치에서, 상기 목적음 추출 수단에 의해 추출된 파워 스펙트럼 정보에 대하여, 스펙트럼 서브트랙션 처리를 행하는 스펙트럼 서브트랙션 수단을 더 구비한 것을 특징으로 한다.

[0040] 본 발명에 따르면, 스펙트럼 서브트랙션 처리를 행함으로써 도래 방향 불명의 정상 잡음이나 확산성의 잡음 등을 제거하는 것이 가능하게 된다.

[0041] 청구항 7에 기재된 발명은, 청구항 1 내지 청구항 6 중 어느 한 항에 기재된 음원 분리 장치에서, 상기 빔 포머 수단에 의한 처리를 행하기 전에, 잡음을 저감하는 처리를 행하는 정상 잡음 저감 수단을 더 구비하는 것을 특징으로 한다.

[0042] 본 발명에 따르면, 뮤지컬 노이즈와 같은 청감상 불쾌한 잡음의 발생을 저감하는 것이 가능하게 된다.

[0043] 청구항 8에 기재된 발명은, 청구항 1 내지 청구항 7 중 어느 한 항에 기재된 음원 분리 장치에 의해 분리된 음원 신호의 음성 인식을 행하는 음성 인식 수단을 구비한 음성 인식 장치를 제공한다.

[0044] 본 발명에 따르면, 정밀도 좋게 분리된 음원 신호에 기초하여, 정밀도가 좋은 음성 인식을 행할 수 있다.

[0045] 청구항 9에 기재된 발명은, 청구항 8에 기재된 음성 인식 장치에서, 차량의 운전석측으로부터 발하여지는 어휘의 후보의 일람인 운전석측 인식 어휘 리스트 및 조수석측으로부터 발하여지는 어휘의 후보의 일람인 조수석측 인식 어휘 리스트를 기억하는 인식 어휘 리스트 기억 수단을 더 구비하고, 상기 음성 인식 수단은, 상기 인식 어휘 리스트 기억 수단에 기억되어 있는 운전석측 인식 어휘 리스트 및 조수석측 인식 어휘 리스트에 기초하여, 상기 음원 분리 장치에 의해 분리된 음원 신호의 음성 인식 처리를 행하는 것을 특징으로 한다.

[0046] 본 발명에 따르면, 음성 인식 장치는, 인식 어휘 리스트 기억 수단에 기억되어 있는 조수석측 인식 어휘 리스트 및 운전석측 인식 어휘 리스트에 기초하여 음성 인식 처리를 행하기 때문에, 운전석, 조수석별로 어휘 리스트 중에서 최적의 어휘를 선택할 수 있어, 정밀도 좋게 음성 인식을 행하는 것이 가능하게 된다.

[0047] 청구항 10에 기재된 발명은, 청구항 8 또는 청구항 9에 기재된 음성 인식 장치에서, 현재의 차량의 상태를 관리하는 상태 천이 수단과, 차량의 상태에 따른 조수석측 및 운전석측에서의 유효한 어휘 리스트를 기억하는 유효 어휘 리스트 기억 수단과, 상기 상태 천이 수단에 의해 관리되고 있는 현재의 차량의 상태와, 상기 유효 어휘 리스트 기억 수단에 기억되어 있는 어휘 리스트에 기초하여, 상기 음성 인식 수단에 의해 인식된 어휘가 유효한지의 여부를 판단하고, 그 판단 결과에 따라서 제어를 행하는 제어 수단을 더 구비하는 것을 특징으로 한다.

[0048] 본 발명에 따르면, 현재의 차량의 상태와 유효 어휘 리스트에 기초하여, 인식된 어휘가 유효한지의 여부를 판단하고, 판단 결과에 따라서 제어를 행할 수 있기 때문에, 차량 내의 탑승자에게 있어서 쾌적한 제어를 행할 수 있다. 또한, 유효 어휘 리스트나 제어 내용을 자유롭게 설계할 수 있기 때문에, 음성 인식을 사용한 어플리케이션 설계에 자유도를 부여할 수 있다.

[0049] 청구항 11에 기재된 발명은, 청구항 1 내지 청구항 7 중 어느 한 항에 기재된 음원 분리 장치를 구비한 휴대 전화기를 제공한다.

[0050] 본 발명에 따르면, 휴대 전화기를 중규모의 회의실 등에서 마이크로서 이용하는 것이 가능하다.

[0051] 청구항 12에 기재된 발명은, 서로 이격하여 배치된 적어도 2개의 마이크로폰에 대하여, 복수의 음원으로부터 발하여진 음원 신호를 입력받는 음원 신호 수신 스텝과, 상기 마이크로폰으로부터의 출력 신호에 대하여, 주파수 영역에서 복소공액의 관계에 있는 2개의 가중 계수 각각을 이용하여 연산을 행함으로써, 2개의 마이크로폰을 연결하는 직선의 수선에 대하여 대칭인 소정의 방향으로부터 도래하는 음원 신호를 각각 감쇠시키기 위한 제1 빔 포머 처리 및 제2 빔 포머 처리를 행하는 빔 포머 처리 스텝과, 상기 빔 포머 처리 스텝에서 얻어진 음원 신호 각각에 대해서 파워 스펙트럼 정보를 산출하는 파워 계산 스텝과, 상기 파워 계산 스텝에서 계산된 파워 스펙트럼 정보끼리의 차분에 기초하여, 목적 음원의 스펙트럼 정보를 추출하는 목적음 스펙트럼 추출 스텝을 구비하는 것을 특징으로 하는 음원 분리 방법을 제공한다.

[0052] 청구항 13에 기재된 발명은, 컴퓨터에, 서로 이격하여 배치된 적어도 2개의 마이크로폰으로부터, 복수의 음원으로부터 발하여진 음원 신호가 혼합된 출력 신호를 취득하는 출력 신호 취득 스텝과, 상기 출력 신호 취득 스텝에서 취득된 출력 신호에 대하여, 주파수 영역에서 복소공액의 관계에 있는 2개의 가중 계수 각각을 이용하여 연산을 행함으로써, 2개의 마이크로폰을 연결하는 직선의 수선에 대하여 대칭인 소정의 방향으로부터 도래하는 음원 신호를 각각 감쇠시키기 위한 제1 빔 포머 처리 및 제2 빔 포머 처리를 행하는 빔 포머 처리 스텝과, 상기 빔 포머 처리 스텝에서 얻어진 음원 신호 각각에 대하여 파워 스펙트럼 정보를 산출하는 파워 계산 스텝과, 상기 파워 계산 스텝에서 계산된 파워 스펙트럼 정보끼리의 차분에 기초하여, 목적 음원의 스펙트럼 정보를 추출하는 목적음 스펙트럼 추출 스텝을 실행시키기 위한 프로그램을 제공한다.

[0053] [발명의 효과]

[0054] 본 발명에 따르면, 주파수 영역에서 복소공액의 관계에 있는 2개의 가중 계수 각각을 이용하여 연산을 행함으로써, 2개의 마이크로폰을 연결하는 직선의 수선에 대하여 대칭인 소정의 방향으로부터 도래하는 음원 신호를 각각 감쇠시키기 위한 제1 빔 포머 처리 및 제2 빔 포머 처리를 행함으로써, 2개의 마이크로폰을 연결하는 직선의 수선에 대하여 대칭인 방향으로부터 도래하는 음원 신호를 각각 감쇠시키고, 상기 제1 빔 포머 처리 및 상기 제2 빔 포머 처리에 의해 얻어진 음원 신호 각각에 대한 파워 스펙트럼 정보끼리의 차분에 기초하여 목적 음원의 스펙트럼 정보를 추출함으로써, 지향 특성이 마이크로폰 소자의 감도에 영향을 받지 않는다고 하는 성질을 실현할 수 있고, 마이크로폰 소자의 감도의 변동에 영향을 받지 않아, 복수의 음원으로부터 발하여진 음원 신호가 혼합된 혼합음 중에서, 목적 음원으로부터의 음원 신호를 분리하는 것이 가능하게 된다.

산업상 이용 가능성

[0231] 음성 인식 장치, 카 네비게이션, 집음 장치, 녹음 장치, 음성 커맨드에 의한 기기의 제어 등, 음원을 정밀도 좋게 분리할 필요가 있는 모든 산업에 이용 가능하다.

도면의 간단한 설명

[0055] 도 1은 본 발명의 제1 실시 형태에 따른 음원 분리 시스템의 기본적 구성을 도시하는 도면.

[0056] 도 2는 동 실시 형태에 따른 마이크의 종류의 일례를 도시하는 도면.

[0057] 도 3은 동 실시 형태에 따른 빔 포머부의 구성을 도시하는 도면.

[0058] 도 4는 동 실시 형태에 따른 파워 계산부의 구성을 도시하는 도면.

[0059] 도 5는 동 실시 형태에 따른 목적음 스펙트럼 추출부의 구성을 도시하는 도면.

[0060] 도 6은 동 실시 형태에 따른 사각 제어형 빔 포머를 설명하기 위한 도면.

[0061] 도 7은 동 실시 형태에 따른 사각 제어형 빔 포머를 설명하기 위한 도면.

[0062] 도 8은 동 실시 형태에 따른 차분 결과의 지향 특성의 일례를 도시하는 도면.

[0063] 도 9는 동 실시 형태에 따른 마이크의 소자 감도차를 변동시킨 경우의 공액 빔 포머의 지향 특성을 도시하는 도면.

[0064] 도 10은 동 실시 형태에 따른 마이크의 소자 감도차를 변동시킨 경우의 빔 포머 단체의 지향 특성을 도시하는 도면.

[0065] 도 11은 동 실시 형태에 따른 음원 분리 장치의 360도의 지향 특성을 구한 그래프를 도시하는 도면.

[0066] 도 12는 동 실시 형태에 따른 음원 분리 장치의 지향 특성을 이용한 2화자 개별 녹음의 실시예를 도시하는 도면.

[0067] 도 13은 동 실시 형태에 따른 음원 분리 장치의 지향 특성을 이용한 간이 핸즈프리 장치 및 음성 인식 장치용 마이크의 실시예를 도시하는 도면.

[0068] 도 14는 동 실시 형태에 따른 평가 실험에서의 마이크 부착 개소를 도시하는 도면.

[0069] 도 15는 동 실시 형태에 따른 평가 실험의 조건을 도시하는 도면.

[0070] 도 16은 동 실시 형태에 따른 평가 실험 결과를 도시하는 도면.

[0071] 도 17은 제2 실시 형태에 따른 음원 분리 시스템의 구성을 도시하는 도면.

[0072] 도 18은 동 실시 형태에 따른 빔 포머의 구성을 도시하는 도면.

[0073] 도 19는 제3 실시 형태에 따른 음원 분리 시스템의 구성을 도시하는 도면.

[0074] 도 20은 동 실시 형태에 따른 다른 음원 분리 시스템의 구성을 도시하는 도면.

[0075] 도 21은 동 실시 형태에 따른 목적음 스펙트럼 추출부의 구성을 도시하는 도면.

[0076] 도 22는 제4 실시 형태에 따른 음원 분리 시스템의 구성을 도시하는 도면.

[0077] 도 23은 제5 실시 형태에 따른 음원 분리 시스템의 구성을 도시하는 도면.

[0078] 도 24는 동 실시 형태에 따른 지향성 제어의 예를 도시하는 도면.

[0079] 도 25는 동 실시 형태에 따른 지향성 제어부의 구성을 도시하는 도면.

[0080] 도 26은 제6 실시 형태에 따른 음원 분리 시스템의 구성을 도시하는 도면.

[0081] 도 27은 제7 실시 형태에 따른 음원 분리 시스템의 구성을 도시하는 도면.

[0082] 도 28은 동 실시 형태에 따른 분리되는 음원의 위치를 설명하기 위한 도면.

[0083] 도 29는 동 실시 형태에 따른 단일지향성 마이크의 설치예를 도시하는 도면.

[0084] 도 30은 동 실시 형태에 따른 목적음 추출부의 구성의 일례를 도시하는 도면.

[0085] 도 31은 동 실시 형태에 따른 목적음 추출부의 구성의 일례를 도시하는 도면.

[0086] 도 32는 동 실시 형태에 따른 목적음 추출부의 구성의 일례를 도시하는 도면.

[0087] 도 33은 동 실시 형태에 따른 음원 분리 장치를 이용한 퍼스널 컴퓨터에의 음성 입력의 예를 도시하는 도면.

[0088] 도 34는 제8 실시 형태에 따른 목적음 범위와 잡음 범위를 설명하기 위한 도면.

[0089] 도 35는 동 실시 형태에 따른 지연 조작을 설명하기 위한 도면.

[0090] 도 36은 동 실시 형태에 따른 지향성 제어 수단의 구성예를 도시하는 도면.

[0091] 도 37은 동 실시 형태에 따른 음원 분리 장치 시스템의 구성예를 도시하는 도면.

[0092] 도 38은 동 실시 형태에 따른 목적음 추출부에서의 처리 방식의 일례를 도시하는 도면.

[0093] 도 39는 동 실시 형태에 따른 목적음 추출부에서의 처리 방식의 일례를 도시하는 도면.

[0094] 도 40은 제9 실시 형태에 따른 차재 기기 제어용 음성 인식 시스템의 구성을 도시하는 도면.

[0095] 도 41은 제10 실시 형태에 따른 휴대 전화기를 도시하는 도면.

[0096] 도 42는 제11 실시 형태에 따른 음원 분리 시스템에서의 마이크 배열을 도시하는 도면.

[0097] 도 43은 제12 실시 형태에 따른 음원 분리 시스템이 적용되는 환경을 도시하는 도면.

[0098] 도 44는 동 실시 형태에 따른 음원 분리 시스템이 적용되는 발화음의 입력 상황을 도시하는 도면.

[0099] 도 45는 동 실시 형태에 따른 가이던스 음성 삭제부의 구성을 도시하는 도면.

[0100] 도 46은 동 실시 형태에 따른 목적 음성 추출부의 구성을 도시하는 도면.

[0101] 도 47은 동 실시 형태에 따른 가이던스 음성 삭제부의 다른 구성을 도시하는 도면.

[0102] 도 48은 동 실시 형태에 따른 가이던스 음성 삭제부의 다른 구성에서의 목적 음성 추출부의 구성을 도시하는 도면.

[0103] 도 49는 동 실시 형태에 따른 음원 분리 시스템이 적용되는 다른 환경을 도시하는 도면.

[0104] 도 50은 동 실시 형태에 따른 음원 분리 시스템이 적용되는 발화음의 다른 입력 상황을 도시하는 도면.

[0105] 도 51은 동 실시 형태에 따른 음원 분리 시스템의 다른 구성을 도시하는 도면.

[0106] <발명을 실시하기 위한 최량의 형태>

[0107] 이하, 도면을 참조하여, 본 발명에 따른 실시 형태에 대해서 설명한다.

[0108] [제1 실시 형태]

[0109] 도 1은, 본 발명의 제1 실시 형태에 따른 음원 분리 시스템의 기본적 구성을 도시하는 도면이다. 이 시스템은, 2개의 마이크로폰(이하 「마이크」라고 함)(10, 11)과, 음원 분리 장치(1)로 구성되어 있다. 이 음원 분리 장치(1)는, 도시하지 않는, 전체를 제어하고 연산 처리를 실행하는 CPU와, ROM, RAM, 하드디스크 장치 등의 기억 장치를 포함하는 하드웨어와, 기억 장치에 기억된 프로그램, 데이터 등을 포함하는 소프트웨어를 구비하고 있다. 이들 하드웨어 및 소프트웨어에 의해, 도 1에 도시하는 기능 블록이 실현된다.

[0110] 2개의 마이크(10, 11)는 무지향성의 마이크이며, 평면 상에 서로 수㎝ 정도 이격하여 설치되어 있다. 또한, 마이크(10, 11)는 무지향성인 것이 기본이지만, 도 2에 도시하는 바와 같은 단일 지향성 마이크도 사용 가능하다. 마이크(10, 11)는 2개의 음원 R1, R2로부터 발하여진 신호를 수신한다. 이 때, 이들 2개의 음원 R1, R2는, 2개의 마이크(10, 11)를 연결하는 직선에 대하여 그어진 수선을 경계로 하여 분할된 2개의 영역(이하 「수선의 좌우」라고 함)에 각각 위치하는 것으로 하지만, 반드시 수선에 대하여 좌우 대칭인 위치에 존재할 필요는 없다.

[0111] 이 마이크(10, 11)에서 얻은 2개의 음원 신호를, 스펙트럼 분석부(20, 21)에서 각각 마이크 출력마다 주파수 분석하고, 빔 포머부(3)에서 이들 주파수 분석된 신호를 2개의 마이크(10, 11)를 연결하는 직선에 대하여 그어진 수선의 좌우에 대칭으로 사각을 형성한 빔 포머(30, 31)에서 필터링을 행하고, 파워 계산부(40, 41)에서 그 필터 출력의 파워를 계산하고, 목적음 스펙트럼 추출부(50, 51)에서 각각의 차분을 계산한 결과에 대하여 임의의 일정값 이상의 값을 출력하고 그 이하를 제로로 하는 처리를 행한다. 이들 처리는, 통상의 특정 잡음에 관해서 사각을 형성하는 처리와는 달리, 빔 포머부(3)를 몇 개의 조건 하에서 형성하여 상기의 처리를 실시함으로써, 종래 문제이었던 마이크의 소자 감도의 변동에 기인하는 마이크로폰 어레이의 특성의 열화의 문제를 해결함과 동시에, 상기 수선을 중심으로 하여 좌우로부터의 음을 분리하는 지향 특성을 폭넓은 주파수 대역에 대하여 실현하는 것이 가능하게 된다. 이하, 각 기능 블록에 대해서 상세하게 설명한다.

[0112] [빔 포머부]

[0113] 우선, 도 3을 참조하여, 빔 포머부(3)의 구성을 설명한다. 도 3에서는, 스펙트럼 분석부(20), 스펙트럼 분석부(21)에서 주파수 성분마다 분해된 신호 $x_1(\omega)$, $x_2(\omega)$를 입력으로 하여, 승산기(100a, 100b, 100c, 100d)에서, 필터의 가중 계수$w_1(\omega)$, $w_2(\omega)$, $w_1^*(\omega)$, $w_2^*(\omega)$(*은 복소공액의 관계에 있는 것을 나타냄)와 승산을 각각 행하고, 가산기(100e, 100f)에서 2개의 승산 결과를 가산하고, 그 출력으로서 필터링 처리 결과 $ds_1(\omega)$, $ds_2(\omega)$를 출력한다. 이와 같이, 빔 포머부(3)는, 복소공액 필터 계수를 사용함으로써, 마이크(10, 11) 사이를 연결하는 직선의 수선을 중심으로 하여 대칭인 위치에 사각을 형성하고 있다.

[0114] [파워 계산부]

[0115] 다음으로, 도 4를 참조하여, 파워 계산부(40, 41)에 대해서 설명한다. 파워 계산부(40, 41)는, 빔 포머(30), 빔 포머(31)로부터의 출력 $ds_1(\omega)$, $ds_2(\omega)$를, 이하의 계산식에 의해, 파워 계산부(40), 파워 계산부(41)에서 파워 스펙트럼 정보$ps_1(\omega)$, $ps_2(\omega)$로 변환한다.

[0116] $$ps_1(\omega)=[Re(ds_1(\omega))]^2+[Im(ds_1(\omega))]^2$$

[0117] $$ps_2(\omega)=[Re(ds_2(\omega))]^2+[Im(ds_2(\omega))]^2$$

[0118] [목적음 스펙트럼 추출부]

[0119] 다음으로, 도 5를 참조하여, 목적음 스펙트럼 추출부(50, 51)에 관해서 설명하면, 파워 계산부(40, 41)의 출력 $ps_1(\omega)$, $ps_2(\omega)$는, 목적음 스펙트럼 추출부(50, 51)의 2개의 입력으로서 사용된다. 목적음 스펙트럼 추출부(50, 51)는, 2개의 빔 포머(30, 31) 출력의 파워 스펙트럼 정보를 입력으로 하여, 출력으로서 좌우의 목적음을 각각 출력하지만, 내부적으로는 차분 계산부(500, 510)와 계수 변환부(501, 511)로 구성된다.

[0120] 차분 계산부(500)는, 감산기(500a)에서 빔 포머(30)의 파워 스펙트럼 정보로부터 빔 포머(31)의 파워 스펙트럼

정보의 감산을 행하고, 마찬가지로, 차분 계산부(510)는, 감산기(510a)에서 빔 포머(31)의 파워 스펙트럼 정보로부터 빔 포머(30)의 파워 스펙트럼 정보의 감산을 행하고, 그 결과를 각각, 계수 변환부(501), 계수 변환부(511)에 입력한다. 계수 변환부(501), 계수 변환부(511)는, 각각 좌우의 음을 분리하기 위한 블록이며, 각각, 값이 특정한 임계값 이상의 값을, 목적으로 하는 방향으로부터의 신호로서 스펙트럼 정보를 출력한다. 여기서, 임계값의 값은 일반적으로는 "0"이지만, 사용 환경에 따라서 최적값을 실측으로부터 구하여 따로따로 설정하여도 된다.

[0121] [동작]

[0122] 다음으로, 도 1을 참조하여, 음원 분리 장치 시스템 전체의 동작을 설명한다.

[0123] 우선, 2개의 무지향성 혹은 지향성의 마이크(10, 11)를 서로 수㎝ 정도 이격하여 설치하고, 2개의 음원으로부터 발하여진 신호를 마이크(10), 마이크(11)에서 수신한다. 이 때, 각 마이크(10, 11)에서 수신된 2개의 음원 신호가 중첩된 신호는, 스펙트럼 분석부(20), 스펙트럼 분석부(21)에서 주파수 분석이 행해진다. 여기서는, 일반적으로는 고속 푸리에 변환 등의 방법이 이용되지만, 필터 뱅크 등의 주파수 분석 방법이어도 된다. 주파수 분석 처리는 10msec 정도의 고정 주기마다 행해진다.

[0124] 주파수 분석된 2개의 신호는, 마이크(10, 11)를 연결하는 직선에의 수선에 대하여 대칭으로 사각이 형성된 빔 포머(30), 빔 포머(31)에서 필터링되어, 특정 방향으로부터의 신호가 감쇠시켜진다. 그러나, 여기서는 특정한 음원 도래 방향을 정확하게 추측하고, 정확하게 추측된 음원 방향으로 사각을 향하게 하는 것을 목적으로는 하고 있지 않다. 이 2채널의 입력을 사용한 필터링은 각 주파수 성분마다 행해지고, 빔 포머(30), 빔 포머(31)의 출력은 파워 계산부(40), 파워 계산부(41)에서 스펙트럼 파워 정보로 변환됨과 동시에, 빔 포머(30), 빔 포머(31)의 출력으로부터 위상 추출부(60), 위상 추출부(61)에서 위상 정보 Φ_1, Φ_2가 추출된다. 다음으로, 파워 계산부(40), 파워 계산부(41)에서 스펙트럼 파워 정보로 변환된 빔 포머(30), 빔 포머(31)의 출력은, 목적음 스펙트럼 추출부(50), 목적음 스펙트럼 추출부(51)에 보내어져, 목적음 스펙트럼 추출부(50)에서 오른쪽 방향(0~90°)으로부터 온 음원 신호의 파워 스펙트럼 정보가 추출되고, 목적음 스펙트럼 추출부(51)에서 왼쪽 방향(-90° ~0)으로부터 온 음원 신호의 파워 스펙트럼 정보가 추출된다.

[0125] 목적음 스펙트럼 추출부(51)로부터 추출된 파워 스펙트럼 정보를 음성 인식 처리의 전처리로서 사용하는 경우에는, 그 파워 스펙트럼 정보는 도시하지 않은 음향 파라미터 해석부에 보내어져, 음향 분석 처리가 행해진다. 한편, 추출한 음원 신호의 파워 스펙트럼 정보를 시간 신호로 되돌릴 필요가 있는 경우에는, 위상 추출부(60), 위상 추출부(61)에서 추출한 위상 정보와 목적음 스펙트럼 추출부(50), 목적음 스펙트럼 추출부(51)에서 추출한 스펙트럼 정보를 시간 파형 변환부(70), 시간 파형 변환부(71)에 입력하여 시간 신호 정보로 되돌리는 처리를 행한다.

[0126] [사각 빔 포머의 설계예]

[0127] 다음으로, 빔 포머부(3) 내의 빔 포머(30), 빔 포머(31)에서, 2개의 마이크(10, 11)를 연결하는 직선의 수선에 대하여 대칭인 위치에 사각을 형성함으로써, 지향 특성(지향성)이 마이크 소자의 감도의 영향을 받지 않는 것에 대해서 증명한다.

[0128] 2개의 마이크 소자를 사용하는 경우에서, 도 6에 도시하는 바와 같이, 목적 방위 Θ_1에 대한 게인을 1로 하고, 타방향 Θ_2에 1개의 사각(게인 0)을 형성하는 사각 빔 포머의 설계예를 이하에 기재한다.

[0129] 사각 빔 포머의 출력 신호를 $S(f)=[s_1(f), s_2(f)]'$, 관측 신호를 $X(f)=[x_1(f), x_2(f)]'$로 했을 때, 임의의 주파수 f에서의 사각 빔 포머의 가중 계수 벡터 $W(f, \Theta_1, \Theta_2)=[w_1(f), w_2(f)]'$는 이하의 계산에 의해 구할 수 있다(단, '은 전치 조작을 나타냄).

수학식 1

[0130]
$$S(f,\theta_1,\theta_2)=W(f,\theta_1,\theta_2)X(f,\theta_1,\theta_2)$$

$$W(f,\theta_1,\theta_2)=\begin{bmatrix} w_1(f,\theta_1,\theta_2) \\ w_2(f,\theta_1,\theta_2) \end{bmatrix}$$

[0131] 이 때,

$$\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}\exp[j2\pi fd_1\sin\theta_1/c] & \exp[j2\pi fd_2\sin\theta_1/c]\\ \exp[j2\pi fd_1\sin\theta_2/c] & \exp[j2\pi fd_2\sin\theta_2/c]\end{bmatrix}\begin{bmatrix}w_1(f,\theta_1,\theta_2)\\ w_2(f,\theta_1,\theta_2)\end{bmatrix}$$

$$\begin{bmatrix}w_1(f,\theta_1,\theta_2)\\ w_2(f,\theta_1,\theta_2)\end{bmatrix}=\begin{bmatrix}\exp[j2\pi fd_1\sin\theta_1/c] & \exp[j2\pi fd_2\sin\theta_1/c]\\ \exp[j2\pi fd_1\sin\theta_2/c] & \exp[j2\pi fd_2\sin\theta_2/c]\end{bmatrix}^{-1}\begin{bmatrix}1\\0\end{bmatrix}$$

[0132]

[0133] 한편, 도 7에 도시하는 바와 같이, 목적 방향, 사각 방향을 2개의 마이크(10, 11)를 연결하는 직선의 수선을 중심으로 하여, 도 6에 도시하는 위치와는 선대칭인 위치로 설정할 때, 가중 계수 벡터 $W(f, -\Theta_1, -\Theta_2)=[w_1(f), w_2(f)]'$는 이하의 계산에 의해 구할 수 있다.

수학식 2

$$S(f,-\theta_1,-\theta_2)=W(f,-\theta_1,-\theta_2)X(f,-\theta_1,-\theta_2)$$

$$W(f,-\theta_1,-\theta_2)=\begin{bmatrix}w_1(f,-\theta_1,-\theta_2)\\ w_2(f,-\theta_1,-\theta_2)\end{bmatrix}$$

[0134]

[0135] 이 때,

수학식 3

$$\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}\exp[-j2\pi fd_1\sin\theta_1/c] & \exp[-j2\pi fd_2\sin\theta_1/c]\\ \exp[-j2\pi fd_1\sin\theta_2/c] & \exp[-j2\pi fd_2\sin\theta_2/c]\end{bmatrix}\begin{bmatrix}w_1(f,-\theta_1,-\theta_2)\\ w_2(f,-\theta_1,-\theta_2)\end{bmatrix}$$

$$\begin{bmatrix}w_1(f,-\theta_1,-\theta_2)\\ w_2(f,-\theta_1,-\theta_2)\end{bmatrix}=\begin{bmatrix}\exp[-j2\pi fd_1\sin\theta_1/c] & \exp[-j2\pi fd_2\sin\theta_1/c]\\ \exp[-j2\pi fd_1\sin\theta_2/c] & \exp[-j2\pi fd_2\sin\theta_2/c]\end{bmatrix}^{-1}\begin{bmatrix}1\\0\end{bmatrix}$$

[0136]

[0137] 으로 계산되고, 양자의 관계는,

수학식 4

$$\begin{bmatrix}w_1(f,-\theta_1,-\theta_2)\\ w_2(f,-\theta_1,-\theta_2)\end{bmatrix}=\begin{bmatrix}w_1(f,\theta_1,\theta_2)\\ w_2(f,\theta_1,\theta_2)\end{bmatrix}^*$$

[0138]

[0139] 로 되어, 각 가중 계수는 복소공액의 관계에 있다.

[0140] 다음으로, 파워 계산부(40, 41), 목적음 스펙트럼 추출부(50, 51)에서 지향 특성의 도출을 행한다. 지향 특성의 계산을 위해서, 가중 벡터 W, 및, 방위 벡터 V를 이하의 식으로 정의한다.

수학식 5

$$W = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \quad V = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

[0141]

[0142] 그러면, 파워 계산부(40)의 출력 $ps_1(\omega)$, 파워 계산부(41)의 출력 $ps_2(\omega)$의 지향 특성은 아래와 같이 표현할 수 있다.

수학식 6

$$ps_1(\omega) = \left[V^T W^*\right]\left[V^T W^*\right]^H = V^T W^* W^T V^*$$
$$ps_2(\omega) = \left[V^T (W^*)^*\right]\left[V^T (W^*)^*\right]^H = V^T W W^H V^*$$

[0143]

[0144] 단, *은 공액 조작, H는 공액 전치 조작을 나타낸다. 이것으로부터, 목적음 스펙트럼 추출부(50) 내의 차분 계산부(500)의 출력 $dr_1(\omega)$은 이하와 같이 구할 수 있다.

수학식 7

$$dr_1(\omega) = V^T W^* W^T V^* - V^T W W^H V^* = V^T \left(W^* W^T - W W^H\right) V^*$$
$$= 2 \times \mathrm{Re}\left[w_1^* w_2 v_1 v_2^*\right] - 2 \times \mathrm{Re}\left[w_1 w_2^* v_1 v_2^*\right]$$

[0145]

[0146] 여기서, 마이크의 소자 감도의 변동을 표현하는 파라미터로서 α를 도입하고, 한쪽의 마이크의 소자 감도가 한쪽의 마이크의 소자 감도의 α배이다라고 가정한다. 이 때에, 한쪽의 마이크 출력이 α배이다라고 하는 것과, 한쪽의 채널에 곱하는 가중치가 α배인 것과는 등가이기 때문에, 마이크 소자 감도의 변동을 고려하여 $W_2 = \alpha W_{org2}$로 하면,

수학식 8

$$dr_1(\omega) = \alpha\left\{2 \times \mathrm{Re}\left[w_1^* w_{org2} v_1 v_2^*\right] - 2 \times \mathrm{Re}\left[w_1 w_{org2}^* v_1 v_2^*\right]\right\}$$

[0147]

[0148] 로 되어, 마이크 소자 감도가 변화되어도 지향 특성은 변화되지 않는다.

[0149] 이 때 음원이 마이크에 대하여 충분히 이격되어 있는 경우, 즉 평면파의 경우에는, 방위 벡터 V는,

수학식 9

$$V = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \exp(-j \dfrac{1}{2\pi} d \sin\theta) \end{bmatrix}$$

[0150]

[0151] 로 표현되기 때문에, 결국,

수학식 10

$$dr(\omega)_1 = \alpha\left\{2\times \mathrm{Re}\left[w_1^* w_{org2}\exp\left(j\frac{2\pi}{\lambda}d\sin\theta\right)\right] - 2\times \mathrm{Re}\left[w_1 w_{org2}^*\exp\left(j\frac{2\pi}{\lambda}d\sin\theta\right)\right]\right\}$$

[0152]

[0153] 으로 된다. 단, 상기에 설명한 방법은 구면파인 경우도 마찬가지의 성질이 유지된다.

[0154] 도 8은, 사각을 상기 수선으로부터 ±45° 방향으로 형성할 때에, 빔 포머(30, 31)의 게인을 유지하는 구속 조건을 수선에 대칭인 위치(±45°)로 지정하여 설계한 경우의 차분 결과의 지향 특성예이다. 도 8로부터 알 수 있는 바와 같이, 0° 방향을 경계로 하여, 오른쪽 방향(마이크(10, 11)로부터 음원 R1, R2의 방향을 본 경우의 오른쪽 방향, 이하 마찬가지)(0~90°)으로부터 온 음원 신호는 플러스의 값을 취하고, 왼쪽 방향(-90° ~0)으로부터 온 음원 신호는 마이너스의 값을 취한다. 이에 의해, 각 주파수 성분에서 좌우 어느 쪽의 방향으로부터 도래했는지를 판단할 수 있다.

[0155] [지향 특성의 실험 결과]

[0156] 이상 설명해 온 바와 같이, 빔 포머(30), 빔 포머(31)의 승산기에서 사용하는 가중 계수를 각각 복소공액의 관계로 하여 상기의 처리를 행함으로써, 마이크의 소자 감도에 의해 어레이 마이크의 지향 특성이 영향을 받지 않는 것을 나타내기 위해서, 마이크의 소자 감도차 α를 0db, +6db, -6db로 변동시켜서 지향 특성의 계산을 행한 예를 도 9에 도시한다. 도 9에는 마이크(10, 11)를 연결하는 직선에의 수선에 대하여 대칭인 방향의 지향 특성이 도시되어 있지 않지만, 그 지향 특성은 도 9에 도시한 특성과 대칭인 특성을 갖는다. 도 9를 보고 알 수 있는 바와 같이, 게인의 변동이 있는 경우, 어레이 마이크의 출력 레벨의 변동은 발생하지만, 지향 특성은 변동을 받지 않는다. 이에 의해, 염가의 마이크를 사용하고 마이크의 소자 감도에 변동이 있는 경우라도 안정된 지향 특성을 실현하는 것이 가능하게 된다. 또한, 도 9에 도시한 지향 특성은 사각을 ±45° 방향으로 작성한 것이지만, 도 9로부터 알 수 있는 바와 같이 지향 특성에 폭이 있기 때문에, 사각을 실제의 목적 음원에 대하여 정확하게 작성할 필요는 없다. 또한, 마이크의 소자 감도차 α를 0db, +6db, -6db로 변동시킨 경우의 빔 포머 단체의 지향 특성을 도 10에 도시하지만, 도 10으로부터 마이크의 소자 감도가 6db 상이하면 특정한 방향으로 사각을 형성한다고 하는 빔 포머에 요망되고 있는 원하는 동작을 거의 행할 수 없는 것을 알 수 있다. 한편, 본 발명의 특징으로서 가장 주목할 점으로서, 도 10에 도시하는 바와 같은 지향 특성이 발생하도록 하는 지향 특성이 열화된 빔 포머를 사용하여도, 결과적으로 얻어지는 지향 특성은 마이크의 소자 감도가 균일한 경우와 동일한 것이다.

[0157] 특정 방향으로 빔 포머에 의해 예리한 사각을 형성하는 방법은 소수의 마이크로도 이론상 실현 가능하기 때문에, 측정한 신호는 SN비를 높이기 위해서 사용되거나, 적응 필터로서 빈번하게 사용되는 일반화 사이드로브 캔슬러의 참조 신호 생성부인 블로킹 매트릭스 부분에 사용되지만, 전술한 마이크의 소자 감도차에 의해 설계대로의 성능으로 사각을 생성할 수 없어, 양산을 목적으로 한 경우, 실제 환경에서 소정의 성능을 실현할 수 없는 큰 요인의 하나로 되고 있다.

[0158] 이와 같은 음원 분리 장치(1)를 이용하여, 360도의 지향 특성을 구한 예를 도 11에 도시한다. 도 11로부터 알 수 있는 바와 같이, 음원 분리 장치(1)는 180° 마다의 지향 특성을 가지며, 2개의 지향 특성이 그 경계에서 겹치지 않고 분리되어 있다. 또한, 그 지향 특성이 마이크의 소자 감도의 영향을 받지 않는 것을 부가적인 특징으로 하고 있다. 리니어 어레이의 경우, ±90° 이상에서는 0~±90° 까지의 특성과 대칭으로 되는 지향 특성으로 된다. 이와 같이, 지향 특성은, 2개의 마이크를 연결하는 선에 수직으로 그은 선을 경계로 하여 2개의 존으로 분할하는 것이 가능하게 된다.

[0159] 이 특성을 이용한 실시예를, 도 12 및 도 13에 도시한다. 도 12는 음성 메모 장치에의 응용이다. 종래, 회의나 협의 시에 회의의 메모의 목적으로 음성 메모 장치가 사용되어 왔다. 이와 같은 사용예의 경우, 주위의 소음이나 2명의 음성이 동시에 녹음되고 있기 때문에, 녹음된 내용을 알아듣기 어려운 경우가 있다. 이와 같은 경우, 2개의 마이크(10, 11)를 각각 2명의 발화자쪽으로 향하게 하여, 음원 분리 장치(1)에서 각각 한쪽의 회화의 내용을 강조하여 녹음함으로써 후일의 듣기를 쉽게 한다. 이와 같이 음원 분리 장치(1)를 이용함으로써, 대향하는 2명의 음성을 분리하여 수집하는 것이 가능하게 되어, 회의의 의사록용 음성 메모로서 이용하는 것이나,

회의 의사록에서 2명의 회화를 개별로 동시 음성 인식하는 것이 가능하게 된다.

[0160] 도 13은, 간이 핸즈프리 장치, 및, 음성 인식 장치용 마이크에의 응용예를 도시하는 도면이다. 최근, 개인용 컴퓨터(이하, 「퍼스널 컴퓨터」라고 함)를 사용하여 인터넷 회의 등이 행하여지도록 되어 오고 있지만, 자택 등에서 퍼스널 컴퓨터를 사용하여 인터넷 회의를 행하는 경우, 퍼스널 컴퓨터 내부에 에코 캔슬러 기능이 없기 때문에 헤드셋 마이크 등을 이용하여 스피커 음성으로부터 마이크에의 에코의 돌아 들어가는 양을 작게 하는 등의 대책이 행해지고 있지만, 헤드셋 마이크와 같이 신체의 일부에 기구를 밀착시키는 것이 좀처럼 받아들여지고 있지 않다. 이 대책으로서, 퍼스널 컴퓨터 내부에 에코 캔슬러 기능을 소프트웨어로서 실장하거나, 퍼스널 컴퓨터 외부에 에코 캔슬러 기능을 내장한 핸즈프리 장치를 접속하는 것도 가능하다. 퍼스널 컴퓨터 내부에 에코 캔슬러 기능을 실장하는 경우, 스피커에의 재생 음성과 마이크로부터의 입력 음성의 동기를 취할 필요가 있지만, 재생 음성과 입력 음성 사이의 지연이 크고, 또한, 퍼스널 컴퓨터 기종마다 지연의 크기가 서로 다른 등 실장상의 문제가 있다. 또한, 외부에 에코 캔슬러 기능을 내장한 핸즈프리 장치를 접속하는 경우, 코스트가 든다.

[0161] 한편, 본 실시예에서는, 마이크(10, 11)를 스피커와 발화자 사이에 둘 필요가 있지만, 스피커 재생 신호와의 동기를 취할 필요도 없다. 2개의 마이크(10, 11)를 준비하고, 마이크(10, 11)로부터의 신호를 퍼스널 컴퓨터에 취득하여, 퍼스널 컴퓨터에 기억되어 있는 소프트웨어에 기초하여 음원 분리를 실행할 수 있기 때문에, 실장이 용이하다.

[0162] 또한, 주택 내에서 음성 인식을 행하는 상황에서 가장 빈번하게 발생할 환경으로서, 텔레비전 그 밖의 피제어 기기를 향해서 음성 커맨드를 사용하는 것이 생각된다. 이 경우, 텔레비전 그 밖의 피제어 기기의 스피커는 피제어 기기측으로부터 발화자측을 향해서 설치하고 있고, 스피커로부터 음성이 흐르고 있거나 다양한 가이던스가 흐르고 있는 상황에서, 발화자는 음성으로 마이크 장치에 제어 커맨드를 입력하고자 하는 경우가 있다. 이 경우, 어떠한 수단으로 스피커의 음성이 마이크 장치에 돌아 들어가는 양을 감쇠시킬 필요가 있지만, 음원 분리 장치(1)를 사용함으로써 용이하게 피제어 기기로부터의 음성과 발화자로부터의 음성을 분리할 수 있어, 음성 인식 성능이 향상된다.

[0163] [평가 실험 결과]

[0164] 다음으로, 평가 실험 결과에 대해서 설명한다. 종래, 자동차 내에서의 기기 제어 그 밖의 목적으로 음성 인식이 사용되어 있다. 종래에는, 에어컨의 송풍 음, 로드 노이즈, 엔진음 등의 잡음을 저감하는 기술은 확립되어 있지만, 음성 등의 영향을 받지 않는 음성 인식 기술은 확립되어 있지 않다. 이들의 것을 근거로 하여, 이하의 점을 실현 가능하게 하는 음성 인식 기술을 제공하는 것이 중요하게 된다. (1) 운전석에 있는 운전자가 발한 음성, 및, 조수석에 있는 조수석자가 발한 음성(이하 「운전석과 조수석의 음성」이라고 함)의 분리, (2) 헤드 포지션의 이동을 허용 가능, (3) 넓은 빔 폭과 분리 성능의 양립, (4) 소수 마이크에서의 기능 제공, (5) 음성 인식 사용 시, 운전자 또는 조수석자에게 조용히 하고 있을 것을 부과하지 않음

[0165] 음원 분리 장치(1)의 유효성을 나타내기 위해서, 차내에서 2명의 발화자에게 동시 발화시켜, 2개의 마이크로 수록하고 음원 분리 장치에서 음원 분리를 행하여, 음성 인식 실험을 행하였다. 이 음원 분리 장치는, 2개의 음원의 분리를 목적으로 개발된 것이며, 자동차 내의 이용 방법으로서는, 예를 들면 운전석과 조수석의 음성을 분리하는 것이 가능하게 된다. 이를 위한 마이크의 설치 개소로서는, 도 14에 도시하는 바와 같이, 부착 개소 L1 또는 부착 개소 L2와 같이 자동차 내 중앙부가 알맞다. 또한, 부착 개소 L2에 설치한 경우에는, 룸 미러(400)가 운전자 쪽을 향하게 하기 때문에, 마이크 부착 시에, 마이크의 부착 방향이 대체로 정면으로 되도록 설치하여도 되고, 후술하는 지향성 제어 기능을 갖게 하여도 된다. 음원 분리 장치(1)에서는 본래 빔 폭을 넓게 취하고 있기 때문에 정확한 위치 정렬은 필요로 하지 않는다. 또한, 부착 개소 L2에 설치하는 경우에는, 배면으로부터의 반사를 억제하기 위해서 마이크 모듈을 궁리하거나 지향성 마이크를 이용하는 것이 유효한 경우가 있다.

[0166] 평가 실험의 조건을 도 15에, 음원 분리 후의 음성 인식 실험 결과를 도 16에 도시한다. 도 16의 (a)에 도시하는 바와 같이, 차내에서의 2화자 발화 시(차내 동시 발화)에서는, 1채널 마이크만을 사용한 종래법1에서의 미처리 시에는 29%(정차 시), 27%(60km/h 주행 시)이었지만, 본 음원 분리 방법을 적응함으로써, 78%(정차 시), 78%(60km/h 주행 시)로 개선되었다. 또한, 도 16의 (b)에 도시하는 바와 같이, 조수석측의 음성을 운전석측의 음성으로 잘못하여 음성 인식 결과를 나오게 했다거나, 혹은 운전석측의 음성을 조수석측의 음성으로 잘못하여 음성 인식 결과를 나오게 한 비율을 평가하면, 종래의 1채널 마이크만 사용 시(차내 1인 발화)에서는, 전체 발화의 93%(리젝션 성능 7%)에 대해서 어떠한 음성 인식 결과가 출력되게 되었지만, 본 방법을 적응한 경우에는,

어떠한 음성 인식 결과가 출력된 것은 0%(리젝션율 100%)이었다. 또한, 2마이크를 사용한 종래예로서 "잡음 성분 억압 처리 장치 및 잡음 성분 억압 처리 방법(특허 제3484112호)"을 종래법2로서 정지 상태에서의 성능 비교를 행하였다. 종래법2는 목적음, 잡음 도래 방향을 추정하면서 적응 빔 포밍 처리를 행하여 목적음 및 잡음을 강조하고, 목적음이 강조된 신호로부터 잡음이 강조된 신호를 주파수 영역에서 스펙트럼 서브트랙션하는 방식이지만, 도래 방향 추정 에러의 영향을 없애기 위해서 목적음, 잡음 모두 도래 방향을 기지(고정 방향으로부터 도래하는 것)로 하고, 적응 빔 포머부의 최적값을 구한 후에, 목적음(발화자1), 잡음(발화자2)을 동시에 재생하여 목적 음성을 추출하는 처리를 행하였다(도 16의 (a) 참조).

[0167] [제2 실시 형태]

[0168] 다음으로, 제2 실시 형태에 대해서 설명한다. 도 17에는, 제2 실시 형태에 따른 음원 분리 시스템의 구성을 도시한다. 전술한 제1 실시 형태에서는, 마이크(10, 11)로부터의 입력을, 우선 스펙트럼 분석부(20), 스펙트럼 분석부(21)에서 주파수 성분으로 변환했지만, 본 실시 형태에서는, 우선 시간 영역에서 빔 포머(80) 및 빔 포머(81)에서 사각을 생성하여 특정한 도래 방향으로부터의 신호를 감쇠시킨 신호를 생성하고, 그 후, 스펙트럼 분석부(20) 및 스펙트럼 분석부(21)에서 주파수 성분으로 변환하고 있다. 또한, 도 17에서는, 도 1과 동일한 기능을 갖는 것에는 동일한 번호가 붙여져 있다. 빔 포머(80), 빔 포머(81)의 구성은 도 18에 도시하는 바와 같은 FIR 필터 등의 형식으로 구성된 필터 처리를 실행함으로써 실현된다. 이 때, FIR 필터의 계수는, 도 3에서 도시한 주파수 영역에서의 복소공액의 관계에 있는 가중 계수를, 시간 영역의 필터 계수로 변환함으로써 구할 수 있다.

[0169] [제3 실시 형태]

[0170] 다음으로, 제3 실시 형태에 대해서 설명한다. 도 19 및 도 20은, 제3 실시 형태에 따른 음원 분리 시스템의 구성을 도시하는 도면이다. 전술한 바와 같이, 도 1 및 도 17에 도시하는 목적음 스펙트럼 추출부(50), 목적음 스펙트럼 추출부(51)는, 도 5에서 도시한 구성으로 실현되어 있고, 실험에 의해 구한 최적의 임계값을 사용하여 음원 분리 처리를 행하는 구성으로 되어 있다. 한편, 도 8에 도시한 바와 같이, 목적음 스펙트럼 추출부(50, 51) 내의 차분 계산부(500, 510)의 출력인 $dr_i(\omega)$(i=1, 2)는 정면 0°를 중심으로 하여 점대칭의 관계로 되어 있는 것을 알 수 있다. 이것으로부터, 목적음 스펙트럼 추출부(50, 51) 내의 계수 변환부(501, 511)에서 임계값을 "0"으로 하면 차분 계산부(500, 510)의 부호만을 보아 둠으로써, 플러스일 때에 오른쪽 방향(0~90°)으로부터 온 음원 신호의 파워 스펙트럼 정보로서 추출되고, 마이너스일 때에 왼쪽 방향(-90°~0)으로부터 온 음원 신호의 파워 스펙트럼 정보로서 추출된다. 이 때문에, 도 1 및 도 17에서 도시한 전체 구성은, 도 19 및 도 20에 도시하는 바와 같이 간략화하는 것이 가능하게 된다. 도 19 및 도 20에서의 목적음 스펙트럼 추출부(90)는, 도 21에 도시하는 구성으로 실현된다.

[0171] 도 21에서, 파워 계산부(40), 파워 계산부(41)에서 계산된 빔 포머(30), 빔 포머(31)의 파워 스펙트럼 정보는, 목적음 스펙트럼 추출부(90) 내부의 차분 계산부(900)에 입력된다. 그리고, 감산기(900a)에서 감산 처리가 행해지고, 계수 변환부(910), 계수 변환부(920)에서 각각 목적으로 하는 방향으로부터의 음원 신호만이 추출된다. 구체적으로는, 계수 변환부(910)는, 오른쪽 방향(0~90°)으로부터의 음원 추출을 행하기 위한 블록으로서, 입력이 플러스인 경우에는, 그 스펙트럼 정보는 오른쪽 방향(0~90°)으로부터 온 것으로서 출력하고, 마이너스인 경우에는 목적 방향 외로부터 도래한 음원의 스펙트럼 정보로서 출력하지 않는다. 한편, 계수 변환부(920)는, 왼쪽 방향(-90°~0)으로부터의 음원 추출을 행하기 위한 블록으로서, 입력이 마이너스인 경우에는, 그 스펙트럼 정보는 왼쪽 방향(-90°~0)으로부터 온 것으로서 출력하고, 플러스인 경우에는 목적 방향 외로부터 도래한 음원의 스펙트럼 정보로서 출력하지 않는다. 이상의 동작에 의해, 2개의 마이크(10, 11)를 연결하는 직선의 수선을 중심으로 하여 좌우의 방향으로부터 도래하는 음원 신호를 분리하는 것이 가능하게 된다.

[0172] 또한, 도 19에 도시하는 음원 분리 시스템과 도 20에 도시하는 음원 분리 시스템은, 빔 포머 처리를 주파수 영역에서 행할지 시간 영역에서 행할지의 차이가 구성으로서 서로 다르다. 도 19에서는 빔 포머 처리를 주파수 영역에서 행하고 있고, 도 20에서는 시간 영역에서 행하고 있다.

[0173] [제4 실시 형태]

[0174] 다음으로, 제4 실시 형태에 대해서 설명한다. 도 22는, 제4 실시 형태에 따른 음원 분리 시스템의 구성을 도시하는 도면이다. 도 22에 도시하는 스펙트럼 서브트랙션부(100, 101)가, 목적음 스펙트럼 추출부(50), 목적음 스펙트럼 추출부(51)에서 추출한 각 목적음에 중첩되어 있는 도래 방향 불명의 정상 잡음이나 확산성의 잡음 등을 제거하기 위한 스펙트럼 서브트랙션 처리를 행한다. 이와 같은 구성은, 회의실 내의 공조나 프로젝터 등의

팬의 음 등이 존재하는 환경에서의 사용에서도 유효하지만, 자동차 내에서의 사용 시에 특히 유효하다. 자동차 내에서 운전자와 조수석에 동승자가 탑승하고 있는 경우 등, 예를 들면 운전석과 조수석의 음성은, 전술한 방식을 사용함으로써 따로따로 취출하는 것이 가능하다. 그러나, 에어컨의 송풍음, 로드 노이즈, 바람을 가르는 소리 등, 도래 방향이 불분명한 것이나 확산성의 잡음은 전술한 방법으로는 제거할 수 없다. 이들 잡음의 영향은, 처리 후단에 스펙트럼 서브트랙션 처리를 넣음으로써 제거하는 것이 가능하다. 스펙트럼 서브트랙션 처리는, 1개의 마이크의 입력 신호를 사용하여 발화 구간 검출을 행하고, 무발화 구간에서 잡음 스펙트럼을 추정하고, 발화 구간에서 그 전에 추정한 잡음 성분을 스케일링하여 감산하는 타입과, 잡음이 우위에 수록된 신호와 음성이 우위에 수록된 신호를 이용하여, 음성이 우위에 수록된 신호로부터 잡음이 우위에 수록된 신호의 스펙트럼을 스케일링하여 차분하는 타입이 있다. 음성 인식의 경우, 1마이크 방식을 기본으로 한 처리로 충분한 경우가 많은데, 본 실시 형태에서도 이것을 채용하고 있으며, 본 실시 형태에 따른 음원 분리 시스템은, 제1 실시 형태에 따른 음원 분리 시스템에 대하여, 발화 구간 검출부(110)와 스펙트럼 서브트랙션부(100, 101)를 새롭게 추가함으로써 대응하고 있다.

[0175] 도 22에서, 음원 R1, 음원 R2는 목적음이며, 음원 R3이 도래 방향 불명의 정상 잡음이나 확산성의 잡음을 나타내고 있다. 이들 잡음의 대부분은 명확한 지향성을 갖지 않는 것이 많다. 이와 같은 잡음의 경우, 목적음 스펙트럼 추출부의 출력에서, 약한 지향성을 갖는 것은 그 방향의 음원을 추출하는 목적음 스펙트럼 추출부에 많이 나타나고, 전혀 지향성을 갖지 않는 것이나 도로의 이음매를 밟은 충격음 등은 좌우의 추출된 스펙트럼에 교대로 검출되는 경우가 많지만, 이들 잡음을 스펙트럼 서브트랙션부(100, 101)에서 제거 가능하다. 또한, 스펙트럼 서브트랙션으로서는, 발화 구간 검출이 불필요한 연속형 스펙트럼 서브트랙션을 사용하여도 된다.

[0176] [제5 실시 형태]

[0177] 다음으로, 제5 실시 형태에 대해서 설명한다. 도 23에는, 제5 실시 형태에 따른 음원 분리 시스템의 구성을 도시한다. 본 실시 형태에서는, 분리하고자 하는 2개의 목적 음원 R1, R2가 2개의 마이크(10, 11)를 연결하는 직선에의 수선에 대하여 대칭이라고 할 수 있는 위치로부터 크게 벗어나 있는 경우의 대책이 기재되어 있다. 본 실시 형태에서는, 이 대책을 위해서 2개의 목적 음원 R1, R2의 대략적인 위치를 검출하기 위한 도래 방향 추정부(120)와, 도래 방향 추정부(120)가 추정한 2개의 목적 음원 R1, R2의 대략의 음원 도래 방향 정보를 사용하여, 지향성 제어부(140)에서 분리하고자 하는 2개의 목적 음원 R1, R2가 가상적으로 가능한 한 2개의 마이크(10, 11)를 연결하는 직선에의 수선에 대하여 대칭으로 되도록, 한쪽의 마이크 입력에 지연 조작을 부여한다.

[0178] 도 24에는, 2개의 음원 R1, 음원 R2가 마이크(10), 마이크(11)를 연결하는 직선의 수선에 대하여 $\Theta\tau$만큼 회전한 직선에 대하여 좌우 대칭으로 되는 상황을 도시하고 있다. 이와 같은 경우, 한쪽의 마이크에서 취득한 신호에 일정 지연량 τd를 부여함으로써, $\Theta\tau$만큼 회전한 것과 등가의 상황을 실현 가능하다. 또한, 미리 몇갠가의 2개의 음원에 대해서, 2개의 마이크를 연결하는 직선의 수선에 대한 위치 관계로부터, 2음원을 분리하는 데에 최적의 빔 포머(30, 31)의 필터 파라미터를 준비해 두고, 도래 방향 추정부(120)로부터의 대략적인 2개의 음원의 도래 방향 정보에 기초하여 현재의 상황의 2개의 음원을 분리하는 데에 최적으로 생각되는 빔 포머의 필터 파라미터를 빔 포머 제어에서 선택하고, 선택한 빔 포머의 필터 파라미터를 빔 포머(30), 빔 포머(31)에 설정하도록 하여도 된다.

[0179] 도 25에 지향성 제어부(140)의 구성예를 도시한다. 도 25의 (a)에 도시하는 구성예는, 주파수 영역에서 각 주파수 성분마다 한쪽의 마이크에서 취득한 신호에 일정 지연량 τd를 부여할 때의 예를 나타내고 있다. 도 25의 (a)에 도시하는 구성예의 경우, $x_1(\omega)$에 대하여 승산기(100a)에서 $e^{-j\omega\tau}$와의 승산을 행함으로써 지연 조작을 실현하고 있다. 또한, 시간 영역에서 빔 포머를 행하고자 하는 경우에는, 도 25의 (b)에 도시하는 바와 같은 FIR 필터의 형식의 지연기(100b)에 의해 필터링 처리를 행하면 된다.

[0180] 또한, 한쪽의 마이크 입력에 지연을 부여하는 것이 아니라, 쌍방의 마이크 입력에 각각 절반씩의 지연을 부여하여 전체적으로 동량의 지연 조작을 실현하여도 된다. 즉, 한쪽의 마이크에서 취득한 신호에 지연량 지연 τd를 부여하는 것이 아니라, 한쪽의 마이크에서 취득한 신호에 지연량 $\tau d/2$, 다른 한쪽의 마이크에서 취득한 신호에 지연량 $-\tau d/2$을 부여함으로써, 전체의 지연차가 τd로 되도록 하여도 된다.

[0181] [제6 실시 형태]

[0182] 다음으로, 제6 실시 형태에 대해서 설명한다. 도 26은, 제6 실시 형태에 따른 음원 분리 시스템의 구성을 도시하는 도면이다. 본 실시 형태에 따른 음원 분리 시스템은 청감을 중시하기 위한 구성으로 되어 있고, 자동차

내에서의 핸즈프리 통화 등에의 응용을 고려하고 있다. 자동차 내에서 운전자와 조수석의 동승자가 탑승하고 있는 경우 등은, 예를 들면 운전석과 조수석의 음성은 전술한 방식을 사용함으로써 따로따로 취출하는 것이 가능하지만, 에어컨의 송풍음, 로드 노이즈, 바람을 가르는 소리 등 도래 방향이 불분명한 것이나 확산성의 잡음은 전술한 방법으로는 제거할 수 없다. 이들의 경우, 제4 실시 형태에서 설명한 바와 같이, 처리 후단에 스펙트럼 서브트랙션 처리를 넣음으로써 이들 잡음의 영향을 제거하는 것이 가능하여, 음성 인식 등의 청감상의 음질을 문제로 하지 않는 용도에는 최적이지만, 핸즈프리 통화 장치용의 마이크 등에 사용한 경우에, 청감상 뮤지컬 노이즈로 불리는 잡음의 소거 잔여가 문제로 되는 경우가 있다.

[0183] 본 발명에서는, 2개의 마이크(10, 11)를 연결하는 직선에의 수선을 중심으로 하여 좌우의 음을 분리하기 위해서, 확산성의 잡음 등 그 음의 지향성이 끊임없이 변동하도록 하는 잡음의 경우, 좌우의 분리 결과로 불규칙하게 분류되게 되어 음질을 열화시키는 경우가 있다.

[0184] 이 때문에, 본 실시 형태에서는, 본 발명에 따른 음원 분리 방식이 마이크 게인의 시간 변화에 영향을 받지 않는 것을 이용하여, 통상 어레이 마이크 처리의 후단에 사용하는 포스트 필터 처리를 빔 포머 처리의 전단에 넣어, 확산성 잡음, 정상 잡음 등을 저감하고, 음원 분리 후의 뮤지컬 노이즈의 발생을 방지하고 있다.

[0185] [제7 실시 형태]

[0186] 다음으로, 제7 실시 형태에 대해서 설명한다. 도 27은, 제7 실시 형태에 따른 음원 분리 시스템의 구성을 도시하는 도면이다. 이 음원 분리 시스템은, 3개의 마이크(10, 11, 12)를 사용하여 3음원을 분리하는 구성을 나타내고 있다. 도 27에서, 마이크(10)와 마이크(12), 마이크(11)와 마이크(12)를 사용하여, 각각 2개의 마이크를 연결하는 직선에의 수선을 중심으로 하여 좌우의 음원 신호를 분리하고, 2조의 마이크의 조를 사용하여 분리된 음원 신호 합계 4개를 사용하여, 목적음 추출부(160)에서 마이크(10), 마이크(11)의 정면 방향 근변으로부터 도래하는 음원 R3을 최종적으로 분리한다.

[0187] 도 28을 사용하여 본 구성에서 분리되는 음원의 설명을 행하면, 도 28에 도시하는 바와 같이 2개의 마이크를 연결하는 직선에의 수선 a와 수선 b를 중심으로 하여, 수선의 좌우로부터 도래하는 음원을 분리하는 것이 가능하게 된다. 도 28에서는, 존 A, B, C의 각 영역에 음원이 존재한다고 가정하면, 수선 a를 사용하여, 존 A와 존 B, C의 영역으로부터 도래하는 음원 신호를 분리할 수 있고, 수선 b를 사용하여, 존 A, B와 존 C로부터 도래하는 음원의 분리를 할 수 있다. 도 27에서, 이들 분리를 행하는 블록이 분리부 b0과 분리부 b1이다. 분리부 b0은, 3개의 음원이 중첩된 신호로부터, 존 A의 영역으로부터의 음원 신호 $S_A(\omega)$와, 존 B, C의 영역으로부터의 음원 신호 $S_{Bmix}(\omega)$, $S_{Cmix}(\omega)$가 혼합된 신호를 분리 가능하고, 분리부 b1은 마찬가지로, 존 A, B의 영역으로부터의 음원 신호 $S_{Amix}(\omega)$, $S_{Bmix}(\omega)$가 혼합된 신호와, 존 C의 영역으로부터의 음원 신호 $S_C(\omega)$를 분리 가능하고, 이 스텝에서, $S_A(\omega)$과 $S_C(\omega)$를 분리할 수 있다. 이와 같이 하여 얻어진 4개의 신호로부터 파워 스펙트럼 영역에서 목적음 추출부(160)에서 소정의 조작을 행함으로써, $S_B(\omega)$를 구할 수 있다. 단, 여기서, $S_{Amix}(\omega)$, $S_{Bmix}(\omega)$, $S_{Cmix}(\omega)$는 타신호와 혼합되었을 때의 각 신호를 나타낸다.

[0188] 상기에서는, 존 D에 음원 정보가 없거나 혹은 있어도 레벨적으로 작은 것이 전제이지만, 만약 존 D에 음원이 있는 경우에는, 3개의 마이크(10, 11, 12)에 지향성 마이크를 사용함으로써 존 D로부터의 음원 신호의 혼입을 대폭 경감하는 것이 가능하다.

[0189] 도 29에는 단일 지향성 마이크의 설치예를 도시하지만, 일반적으로 이와 같이 지향성 마이크를 사용하면 어레이 마이크의 빔 포머부에서 각 마이크 사이의 지향성의 변동 등에 의해 설계 시의 성능이 나오지 않는 경우가 일어나지만, 본 방식에서 원래 마이크의 소자 감도 변동에 영향을 받지 않는 것과 마찬가지로, 지향 특성의 변동에도 영향을 받지 않는 일정한 성능을 실현할 수 있다.

[0190] 도 30 내지 32를 참조하여, 목적음 추출부(160)에서의 처리 방식에 대해서 상세하게 설명한다. 도 30에 도시하는 처리 방식은, 2채널의 스펙트럼 서브트랙션 처리와 마찬가지의 처리 방식이다. 즉, 한쪽의 채널 신호로서도 목적음과 방해음의 파워 스펙트럼의 합이 구해져 있고, 다른 한쪽의 채널 신호로서도 방해음의 파워 스펙트럼이 구해져 있기 때문에, 이들을 감산기(100a)에서 감산함으로써 목적음 $S_B(\omega)$를 구할 수 있다.

[0191] 도 31에 도시하는 처리 방식은, 목적음에 방해음이 중첩된 것을 2개 구하는 것이 가능하기 때문에, 가산기(101b)에서 이들을 가산함으로써 목적음의 파워 스펙트럼 정보를 2배의 크기로 하고, 한편, 가산기(101a)에서 방해음끼리를 가산함으로써 방해음의 파워 스펙트럼을 구하고, 이것에 승산기(101c)에서 일정 팩터(O.E.F 1~

2)를 곱하고, 감산기(101d)에서 가산기(101b)의 출력으로부터 차분을 계산함으로써, 목적음을 추출하는 것이다. 또한, 감산기(101d)의 출력의 스텝에서 원음 신호보다 음량이 크기 때문에 레벨 조정부(101e)에서 레벨 조정을 행한다.

[0192] 도 32에 도시하는 처리 방식은, 도 31에 도시하는 가산기(101a, 101b) 대신에 최소값 계산(102a, 102b)을 사용하는 것이다. 또한, 도 31에서는, O.E.F는 1.0보다 큰 쪽이 음질이 좋은 경우가 많지만, 도 32에서는 O.E.F는 1.0 전후가 좋다. 또한, 도 32에서는 최소값 계산을 행하였지만, 최소값 계산 대신에 최대값 계산이어도 된다.

[0193] 또한, 분리하고자 하는 목적 음원의 위치가 본 방법에서 최적의 분리 성능이 얻어지는 위치로부터 크게 벗어나 있는 경우가 있지만, 제5 실시 형태에서 설명한 바와 같이, 마이크로부터 출력되는 입력 신호에 지연을 부여하여 음원의 도래 방향을 가상적으로 바꿈으로써, 가능한 한 최적의 분리 성능이 얻어지도록 조작하는 것이 가능하다.

[0194] 도 33에는, 본 실시 형태에 따른 음원 분리 시스템의 사용예를 도시한다. 퍼스널 컴퓨터에의 음성 입력의 용도에서, 3개의 마이크(10, 11, 12)를 사용하여, 작은 실장 면적에서, 퍼스널 컴퓨터 정면으로부터의 음성을, 지향성을 갖게 하여 취득하는 예를 나타낸다.

[0195] [제8 실시 형태]

[0196] 다음으로, 제8 실시 형태에 대해서 설명한다. 전술한 실시 형태에서는,

[0197] (1) 2개의 마이크를 사용하여 마이크 사이를 연결하는 직선을 중심으로 하여, 좌우로부터의 음을 분리하는 실시 형태 (2) 3개의 마이크를 사용하여 정면으로부터의 음과 그 좌우로부터의 음을 분리하는 실시 형태에 대해서 설명했지만, 도 34에 도시하는 바와 같이, 2개의 마이크(10, 11)를 사용하여 2개의 마이크(10, 11)를 연결하는 직선의 정면 방향으로부터의 음을 분리 추출하고자 하는 경우가 있다.

[0198] 이 경우, 지향성 제어 수단에 의해 2개의 마이크(10, 11) 중 한쪽 마이크로부터의 출력 신호에 도 35의 (b)에 도시하는 바와 같이 지연을 부여하여, 가상적으로 3채널째의 마이크 위치를 작성함으로써, 가상적으로 도 35의 (a)에 도시하는 3마이크 입력을 실현하는 것이 가능하다. 도 36에는, 도 35에서 도시한 지연 조작을 행하는 지향성 제어 수단의 구성예를 도시한다. 도면에서 Di(i=1, 2, 3, 4)는 지연 소자를 나타내지만, 실제의 지연 조작은 시간 영역에서 지연 조작을 행하여도 되고, 스펙트럼 분석 후에 주파수 영역에서 지연 조작을 행하여도 된다.

[0199] 도 37에는, 본 실시 형태에 따른 음원 분리 장치 시스템의 구성예를 도시한다. 지향성 제어 수단(141, 142)은, 스펙트럼 분석부(20, 21)와, 지연 처리를 행하는 지연 소자로 구성되어 있다. 처리순으로서는, 스펙트럼 분석 처리를 행하고 나서 지연 처리를 행하여도 되고(도면에서의 Type1), 지연 처리를 행하고 나서 스펙트럼 분석 처리를 행하여도 된다(도면에서의 Type2). 지향성 제어 수단(141, 142)의 출력 신호는, 본 방법 NBF 이후의 블록에서 빔 포머(30, 31), 파워 계산부(40, 41) 등에서의 처리가 행해지고, 처리 후의 신호가 목적음 추출부(52)에 입력된다.

[0200] 도 38 및 도 39에는, 목적음 추출부(52)에서의 처리 방식의 일례를 도시한다. 도 38은, Θ_1과 Θ_2가 마이크(11, 12)를 연결하는 직선의 수선에 대하여 대칭인 각도인 경우의 처리 방식의 일례를 도시하고, 도 39는, Θ_1과 Θ_2가 상기 수선에 대하여 대칭인 각도가 아닌 경우의 처리 방식의 일례를 도시한다.

[0201] [제9 실시 형태]

[0202] 다음으로, 제9 실시 형태에 대해서 설명한다. 도 40은, 제9 실시 형태에 따른 차재 기기 제어용 음성 인식 시스템의 구성을 도시하는 도면이다. 본 실시 형태에서는, 자동차 등의 차량에 설치된 차재 기기 제어용 음성 인식 시스템에 본 발명에 따른 음원 분리 장치(1)를 응용한 예를 나타내고 있다. 본 응용예에서는, 운전석과 조수석의 음성을 2개의 마이크(10, 11)에서 취득하고, 음원 분리 장치(1)에서 운전석과 조수석의 음성을 분리하고, 분리한 운전석과 조수석의 음성을 각각 발화 구간 검출, 음성 인식 처리, 음성 인식 결과와 자동차 주행 상태나 그 밖의 운전 상태에 따라서 유효한 인식 결과만을 사용하여, 기기의 제어, 시스템의 응답 등을 행함으로써, 차재 기기 제어용 음성 인식 시스템의 신뢰성의 향상과 차재 기기 제어용 음성 인식 시스템의 응답의 자유성의 확장 등을 제공하는 것이다.

[0203] 차재 기기 제어용 음성 인식 시스템은, 본 시스템에 특징적인 데이터로서, 기억 장치에 조수석측 인식 어휘 리스트(190), 운전석측 인식 어휘 리스트(191), 조수석측 유효 어휘 리스느(210), 및, 운전석측 유효 어휘 리스트

(211)를 기억하고 있다. 운전석측 인식 어휘 리스트(191)는 운전석측으로부터 발하여지는 어휘의 후보의 일람이며, 조수석측 인식 어휘 리스트(190)는 조수석측으로부터 발하여지는 어휘의 후보의 일람이다. 운전석측 유효 어휘 리스트(211)는, 차량의 상태(자동차 주행 상태나 그 밖의 운전 상태)에 따른 운전석측에서의 유효한 어휘 리스트이다. 조수석측 유효 어휘 리스트(210)는, 차량의 상태에 따른 조수석측에서의 유효한 어휘 리스트이다. 여기서, 「유효」란, 어휘(음성 커맨드)에 따른 제어 명령을 출력하는 것이 허용되어 있는 상태를 말한다.

[0204] 본 시스템의 동작을 도 40을 이용하여 설명하면, 운전자 및 조수석의 동승자가 발화한 음성은 2개의 마이크(10, 11)에서 수음되고, 음원 분리 장치(1)에서 운전석과 조수석의 음성으로 분리된 후, 각각 운전자용, 조수석의 동승자용으로 준비된 발화 구간 검출부(170, 171), 음성 인식부(180, 181)에 입력된다. 이 때, 본 발명에 따른 음원 분리 장치(1)의 출력에서 2명의 음성은 정밀도 좋게 분리되어 있기 때문에, 조수석측 발화 구간 검출부(170), 운전석측 발화 구간 검출부(171)에서 양자의 발화 구간을 정밀도 좋게 분리할 수 있음과 동시에, 조수석측 음성 인식부(180), 운전석측 음성 인식부(181)에 상대의 음성을 억압한 정보를 부여하는 것이 가능하여, 상대의 발화에 영향을 받지 않는 음성 인식 처리를 정밀도 좋게 행할 수 있다.

[0205] 본 응용예에서는, 각 음성 인식부(180, 181)에 각각 전용으로 시스템의 상태에 관계없이 어떤 어휘를 인식할지를 지시하기 위한 조수석측 인식 어휘 리스트(190), 운전석측 인식 어휘 리스트(191)가 제공되어 있고, 각 음성 인식부(180, 181)는 이 어휘 리스트에 따라서 음성 인식 처리를 행하여, 음성 인식 결과를 제어부·상태 천이부(200)에 출력한다.

[0206] 제어부·상태 천이부(200)가 구비하는 상태 천이부(201)는, 음성 인식 결과 및 현재의 상태에 기초하여 다음의 상태로 천이할 수 있도록 되어 있다. 제어부·상태 천이부(200)가 구비하는 제어부(202)에서는, 상태 천이부(201)로부터 얻어지는 현재의 상태와, 음성 인식부(180, 181)로부터의 음성 인식 결과에 기초하여, 조수석측, 운전석측 각각에 준비된 조수석측 유효 어휘 리스트(210), 운전석측 유효 어휘 리스트(211)에 기초하여, 어느 음성 커맨드에 응답(제어 명령을 출력)하면 좋을지를 조사한다. 도 40에 도시하는 유효 어휘 리스트(210, 211)에서는, 상태와 음성 커맨드의 조합에 대하여 "○"가 대응지어져 있는 경우에, 음성 커맨드에 응답가능한 것을 나타내고 있다. 예를 들면, 운전석측에서는, 상태가 "운전 중"인 경우에 응답이 허용되어 있는 음성 커맨드는 "좀더", "라이트 켜고", "시원하게"이며, 또한, "에어컨 조작", "CD", "MD"는 금지되어 있다.

[0207] 그리고, 허용되어 있는 음성 커맨드에만 응답하여, 에어컨을 조작하거나 라이트를 켜거나 하는 제어를 행함으로써, 차량의 탑승자는 차내에서 쾌적하게 지낼 수 있다. 또한, 차재 기기 제어용 음성 인식 시스템의 신뢰성을 높일 수 있음과 동시에, 음성 인식을 사용한 어플리케이션 작성에 사양 설계의 자유도를 보다 많이 제공하는 것이 가능하게 된다.

[0208] 이상의 사용예에 따르면, 운전석과 조수석의 동승자로부터 동시에 발화된 음성을 동시에 인식하는 것이 가능하게 되거나, 어느 1명이 발화한 경우라도 그것이 운전석측으로부터의 발화인지 조수석측으로부터의 발화인지를 확실하게 검출하여 인식하는 것이 가능하게 되기 때문에, 동승자의 행동을 제한하지 않고, 발화자와 그 발화자의 음성 커맨드에 대한 응답을 개별로 설계할 수 있게 된다.

[0209] [제10 실시 형태]

[0210] 다음으로, 제10 실시 형태에 대해서 설명한다. 도 41에는, 본 실시 형태에 따른 휴대 전화기(300)를 도시하고 있다. 휴대 전화기(300)에는, 마이크(10, 11) 및 도시하지 않은 음원 분리 장치가 탑재되어 있다. 이 휴대 전화기(300)는, 통상은 화상 전화용이지만, 모드를 절환하여 마이크로서도 사용 가능하다. 도 41의 (a)는 마이크(10, 11)가 화상 전화용 마이크로서 동작하고 있을 때의 모습을 도시하는 도면이며, 도 41의 (b)는 마이크(10, 11)가 마이크로서 동작하고 있을 때의 모습을 도시하는 도면이다. 중규모의 회의실 등에서, 발표자가 마이크를 사용할 만큼 대규모의 회의실도 아닌데, 그러나, 방이 넓어 작은 소리이면 발표자의 소리가 알아듣기 어려운 장면에서 사용할 수 있다.

[0211] 이상 설명한 바와 같이, 최소 2개의 마이크를 서로 이격하여 설치하고, 2개의 마이크를 연결하는 직선의 수선을 중심으로 하여 좌우 대칭인 각도에 시간 영역 혹은 주파수 영역에서 빔 포머에 의해 사각을 형성하고, 시간 영역에서 사각을 작성한 경우에는 주파수 영역으로 변환하고, 쌍방의 빔 포머의 파워 스펙트럼의 차분을 계산하고, 얻어진 결과의 계수 변환을 행함으로써, 좌우의 사각을 중심으로 하여 폭을 갖게 한 지향 특성을 형성하고, 음원 분리를 행하는 것이 가능하게 된다. 이와 같이, 마이크의 소자 감도에 지향 특성이 영향을 받지 않는다고 하는 성질을 실현할 수 있어, 마이크의 소자 감도의 변동에 영향을 받지 않고, 음원 도래 방향의 상정

방향과의 어긋남이나 큰 초기 반사에 관해서도 적절히 넓은 지향 특성에 의해 커버되어, 안정된 2개의 음원의 분리 특성을 실현하는 것이 가능하게 된다.

[0212] [제11 실시 형태]

[0213] 다음으로, 제11 실시 형태에 대해서 설명한다. 도 42에는 추출하고자 하는 목적음이 존 A(예를 들면 운전석의 존)에 존재하고, 그 이외의 장소(존 B, 존 C, 존 D)에 방해음이 존재하는 상황 하에서, 존 A의 목적음을 추출하는 예가 도시되어 있다. 본 방법을 사용한 어레이 마이크(2001)(예를 들면 차실 내에서 전방으로 배치(예를 들면 룸 미러에 설정))를 사용한 경우, 존 A/C(예를 들면 운전석 및 그 후부 좌석의 존)와 존 B/D(예를 들면 조수석 및 그 후부 좌석의 존)에 존재하는 음을 분리하는 것은 가능하지만 존 A(예를 들면 운전석의 존)과 존 C(예를 들면 그 후부 좌석의 존)에 존재하는 음을 분리할 수는 없다. 그러나, 본 방법을 사용한 어레이 마이크(2002)를 도 42에 기재되는 바와 같이 존 A/B와 존 C/D의 경계의 위치에 배치함으로써 존 A/B와 존 C/D에 존재하는 음을 분리하는 것이 가능하게 되기 때문에 존 A의 음만을 추출하는 것이 가능하게 된다.

[0214] 구체적으로는, 존 A, B, C, D에 있는 발화자 A, B, C, D가 동시에 발화한 경우, 우선 존 A/B와 존 C/D의 경계에 배치한 어레이 마이크(2002)를 사용함으로써 존 A/B로부터의 음과 존 C/D로부터의 음을 분리하는 것이 가능하게 된다. 다음으로 어레이 마이크(2001)에 의해 존 A/C로부터의 음과 존 B/D로부터의 음을 분리할 수 있다. 마지막으로 어레이 마이크(2001)를 사용하여 구한 존 A/C로부터의 음과 어레이 마이크(2002)를 사용하여 구한 존 A/B로부터의 음을 각 주파수 영역에서 비교하여 쌍방에 공통으로 있는 주파수 성분을 존 A로부터의 음으로서 분리하는 것이 가능하게 된다. 마찬가지의 처리에 의해, 각 존 B, C, D로부터의 음도 개별로 구하는 것이 가능하다.

[0215] [제12 실시 형태]

[0216] 다음으로, 제12 실시 형태에 대해서 설명한다. 도 43은, 자동차 등의 환경하에서 음성 인식에 의해 기기 조작을 행하는 상황을 상정하고 있다. 도 44는, 그때의 기기 조작의 가이던스 음성과 발화자의 음성의 관계를 도시한다.

[0217] 이와 같은 경우, 스피커(15)로부터 발화자 A의 발화의 타이밍을 재촉할 목적으로 "목적지를 말해 주세요" 등의 가이던스 음이 흐른 후, 기계음 예를 들면 "삐이" 등과 같은 음이 흐르고, 그 후에 발화자 A는 음성 커맨드를 발성한다. 그러나, 유저가 음성 커맨드에 의한 기기 조작에 익숙해짐에 따라, 도 44에 도시하는 바와 같이, 가이던스 음성 중에 발화자 A가 발화를 시작하게 되는 상황이 발생하게 되며, 이것이 음성 인식 성능을 떨어트리는 요인으로 되고 있다.

[0218] 이와 같은 상황에의 대책으로서, 일반적으로는 에코 캔슬러를 사용하여 마이크(10)로부터의 수록음에 혼재되는 가이던스 음성을 적응적으로 추정하여 제거함으로써 대응하고 있다. 또한 다른 대응책으로서, 도 45 내지 도 48에 도시하는 바와 같이, 주파수 분석 후의 마이크(10)에의 1입력 신호에 관해서, 스펙트럼 서브트랙션을 행하거나(도 45, 도 46), 각 주파수 성분마다 가이던스 음성, 발화자 A의 음성 중 어느 쪽이 많이 포함되는지를 추정하여 발화자 A의 음성만이 많이 포함되는 주파수 성분만을 발화자 A의 음성으로서 취출하는 것이 행해지고 있다(도 47, 도 48).

[0219] 여기서, 도 45, 도 47에서, 필터부(1001)는, 스피커(15)로부터 마이크(10)까지의 음향적인 반사 패스를 모의한 필터로서, 미리 구해 둔 스피커(15)로부터 마이크(10)까지의 임펄스 응답을 사용하여도 되고, 적응 필터 처리에 의해 동적으로 구해 두어도 된다.

[0220] 도 45에서의, 게인 조작부(1002)는, 스펙트럼 서브트랙션을 행할 때에 사용하는 오버 서브트랙션 팩터를 결정하는 부분으로서, 스피커(15)의 음량에 따라서 1 내지 10 정도의 게인 중에서 선택하여 사용한다.

[0221] 또한, 도 45에서의, 목적 음성 추출부(1003)는, 게인 조작부(1002) 및 스펙트럼 분석부(21)의 출력에 기초하여, 도 46에 도시하는 바와 같은 처리를 행하고, 처리 결과의 신호를, 시간 파형 변환부(1004)에 출력한다.

[0222] 도 47에서의, 임계값 계산부(1011)는, 가이던스 음성의 평균 에너지에 기초하여 임계값 th를 결정한다.

[0223] 또한, 도 47에서의, 목적 음성 추출부(1012)는, 임계값 계산부(1011) 및 스펙트럼 분석부(21)의 출력에 기초하여, 도 48에 도시하는 바와 같은 처리를 행하고, 처리 결과의 신호를, 시간 파형 변환부(1004)에 출력한다. 또한, 도 48에 도시하는 th_{min}은, 도 48에 도시하는 $X_{Far}(\omega)$이 유효한 입력이다라고 판단하기 위한 임계값으로 된다.

[0224] 또한, 시간 파형 변환부(1004)는, 제1 실시 형태에서의 시간 파형 변환부(70, 71)와 마찬가지의 처리를 행한다.

[0225] 그러나, 종래 방법에서는 도 43과 같이, 발화자 A만이 발화하는 상황에서는 전술한 바와 같은 구성에 의해 대응 가능하지만, 도 49에 도시하는 바와 같이 발화자 A(예를 들면 운전석의 승원)뿐만 아니라 발화자 B(예를 들면 조수석의 승원)가 있는 경우에, 도 50에 도시하는 바와 같이, 발화자 A뿐만 아니라 발화자 B도 어떠한 발화를 하는 경우가 일어날 수 있지만, 이와 같은 상황에는 대응할 수 없다.

[0226] 이와 같은 상황에 대하여, 도 51에 도시하는 바와 같이 본 방법과 도 45 또는 도 47에서 도시한 가이던스 음성 삭제부(1021, 1022)를 조합함으로써 대응이 가능하게 된다.

[0227] 구체적으로는, 도 51에서 가이던스 음성은 차내의 스피커(15)로부터 재생되고, 동시에 발화자 A, 발화자 B 모두 발화하고 있는 상황 하에서, 마이크(10) 및 마이크(11)에는, 가이던스 음성, 발화자 A의 발화 음성, 발화자 B의 발화 음성이 중첩되어 입력되어 있다. 이 때, 가이던스 음성 삭제부(1021) 및 가이던스 음성 삭제부(1022)에서는 도 45 또는 도 47에 도시한 방법에 의해, 가이던스 음성을 제거하고, 결과로서 발화자 A의 음성과 발화자 B의 음성의 양자의 음성이 중첩된 신호를 결과로서 출력한다. 또한, 여기서는 후처리로서 사용하는 본 방법에의 입력시의 계산의 낭비를 없애기 위해서, 시간 파형으로는 되돌리지 않고, 주파수 성분 그대로 본 방법(도 1)에 입력한다. 후단의 본 방법에서도 입력으로서 주파수 성분 정보가 입력되기 때문에 스펙트럼 분석부의 처리를 생략하고, 빔 포머부에 직접 입력을 행하고, 본 방법을 적용하여 처리를 행함으로써 출력 결과로서, 발화자 A의 음성과 발화자 B의 음성을 개별로 얻는 것이 가능하게 되어, 음성 인식 장치의 신뢰성, 성능의 향상과 어플리케이션의 자유도를 대폭 확대할 수 있다.

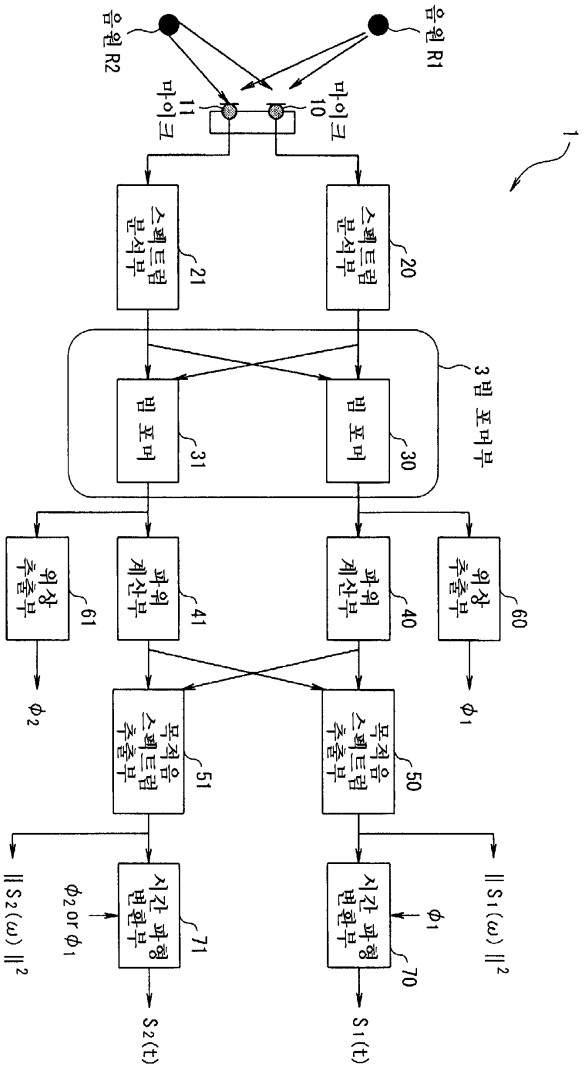
[0228] 또한, 전술한 각종 기능의 조합이나 마이크로부터의 신호의 지연 조작에 의해, 정면 방향으로 좁은 지향 특성을 실현하거나, 특정 방향으로부터의 음원 신호만을 검출하는 것이 가능하게 된다.

[0229] 덧붙여, 마이크 간격을 넓히지 않더라도 저역까지 높은 분리 성능을 확보할 수 있기 때문에, 실장 스페이스를 삭감하는 것이 가능하여, 휴대 기기 등에의 사용도 가능하게 된다.

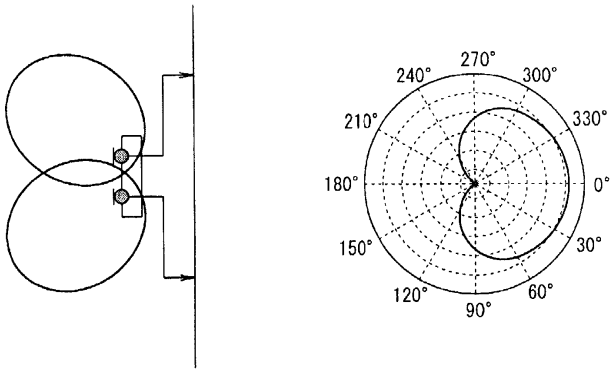
[0230] 또한, 전술한 실시 형태에서는, 음원 분리 시스템의 각 기능 블록을 프로그램으로 실현하는 것으로서 설명했지만, 회로 등을 이용함으로써 하드웨어로 실현하는 것도 가능하다.

도면

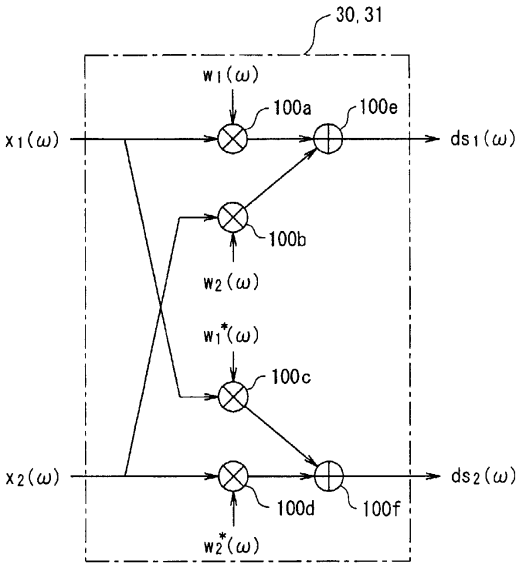
도면1


음원 R2
음원 R1
마이크 11
마이크 10
1
스펙트럼 분석부 21
스펙트럼 분석부 20
3 빔 포머부
빔 포머 31
빔 포머 30
위상 추출부 61
파워 계산부 41
파워 계산부 40
위상 추출부 60
ϕ_2
ϕ_1
목적음 스펙트럼 추출부 51
목적음 스펙트럼 추출부 50
$\|S_2(\omega)\|^2$
ϕ_2 or ϕ_1
시간 파형 변환부 71
시간 파형 변환부 70
ϕ_1
$\|S_1(\omega)\|^2$
$S_2(t)$
$S_1(t)$

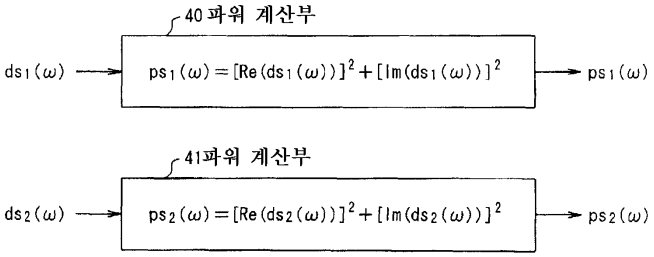
도면2


270°
240°
300°
210°
330°
180°
0°
150°
30°
120°
60°
90°

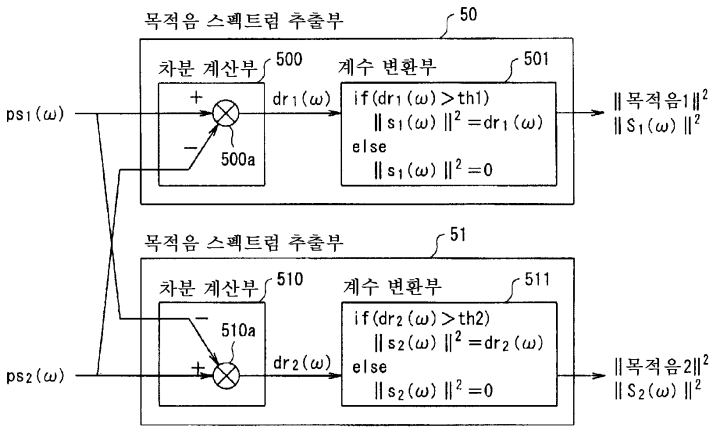
도면3


30, 31
w1(ω)
100a
100e
x1(ω)
ds1(ω)
100b
w2(ω)
w1*(ω)
100c
x2(ω)
ds2(ω)
100d
100f
w2*(ω)

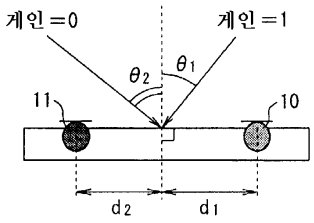
도면4


40 파워 계산부
ds1(ω)
ps1(ω)=[Re(ds1(ω))]2+[Im(ds1(ω))]2
ps1(ω)
41파워 계산부
ds2(ω)
ps2(ω)=[Re(ds2(ω))]2+[Im(ds2(ω))]2
ps2(ω)

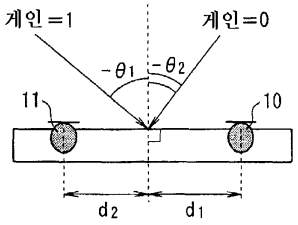
도면5


목적음 스펙트럼 추출부
50
차분 계산부
500
계수 변환부
501
ps1(ω)
dr1(ω)
500a
if(dr1(ω)>th1)
‖s1(ω)‖2=dr1(ω)
else
‖s1(ω)‖2=0
‖목적음1‖2
‖S1(ω)‖2
목적음 스펙트럼 추출부
51
차분 계산부
510
계수 변환부
511
510a
ps2(ω)
dr2(ω)
if(dr2(ω)>th2)
‖s2(ω)‖2=dr2(ω)
else
‖s2(ω)‖2=0
‖목적음2‖2
‖S2(ω)‖2

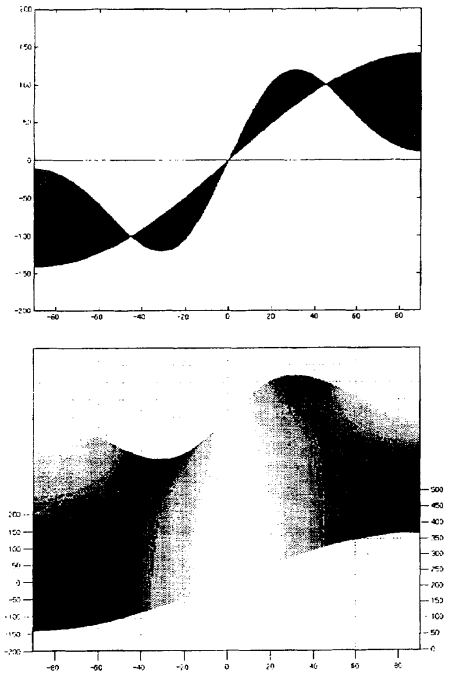
도면6



도면7



도면8



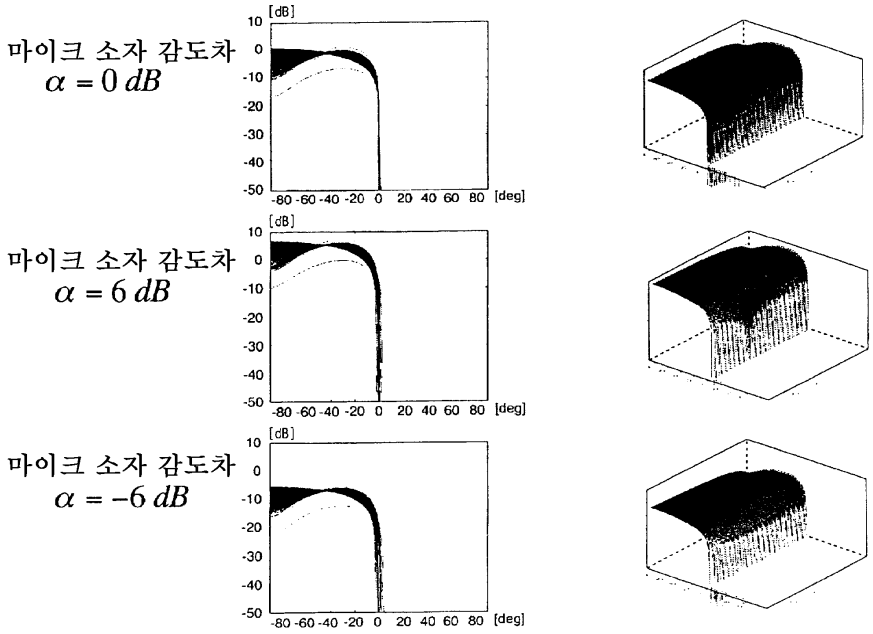
MicGain차 0.0dB

구속 조건

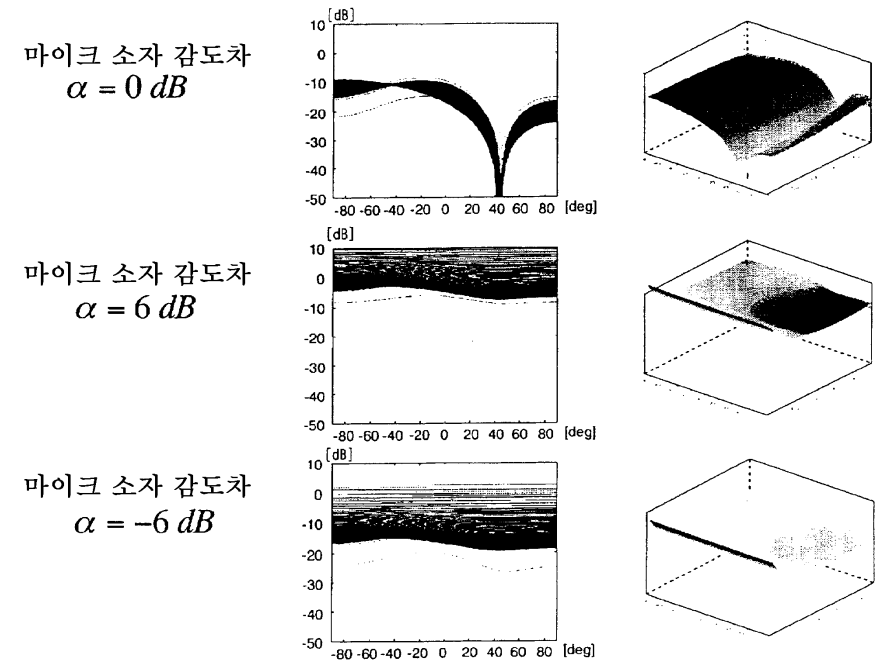
-45도 방향 1.0

45도 방향 0.0

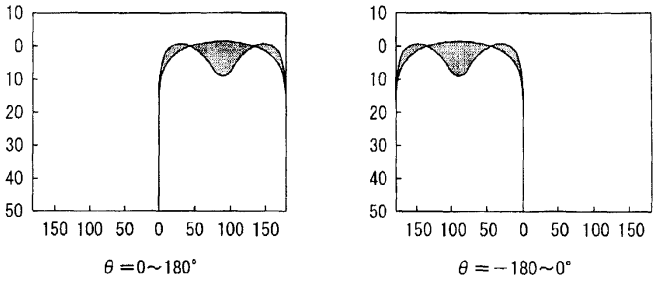
도면9


마이크 소자 감도차
α = 0 dB
마이크 소자 감도차
α = 6 dB
마이크 소자 감도차
α = −6 dB
[dB]
10
0
-10
-20
-30
-40
-50
-80 -60 -40 -20 0 20 40 60 80
[deg]

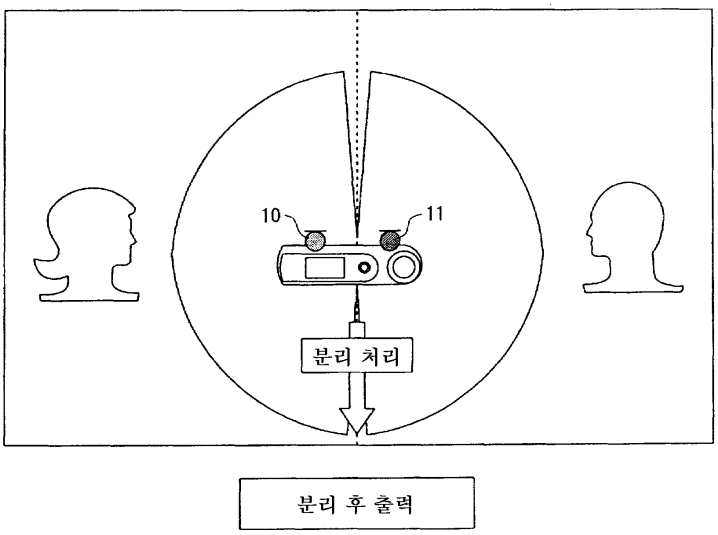
도면10


마이크 소자 감도차
α = 0 dB
마이크 소자 감도차
α = 6 dB
마이크 소자 감도차
α = −6 dB
[dB]
10
0
-10
-20
-30
-40
-50
-80 -60 -40 -20 0 20 40 60 80
[deg]

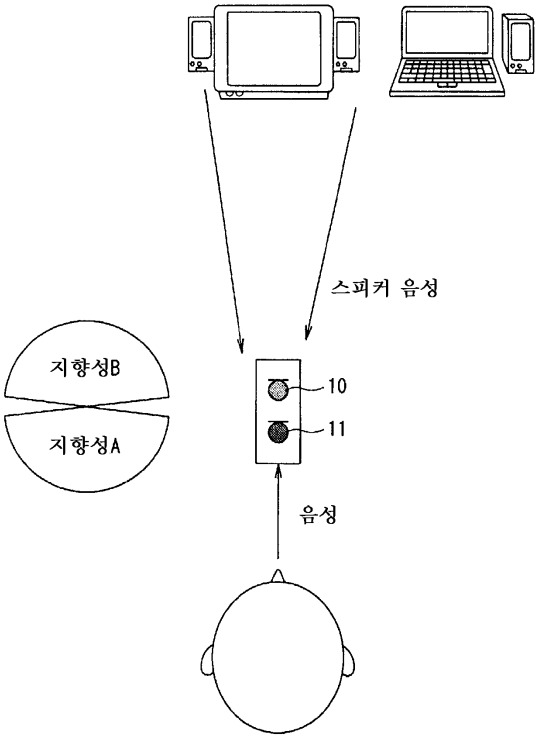
도면11


10
0
10
20
30
40
50
150 100 50 0 50 100 150
θ=0~180°
10
0
10
20
30
40
50
150 100 50 0 50 100 150
θ=-180~0°

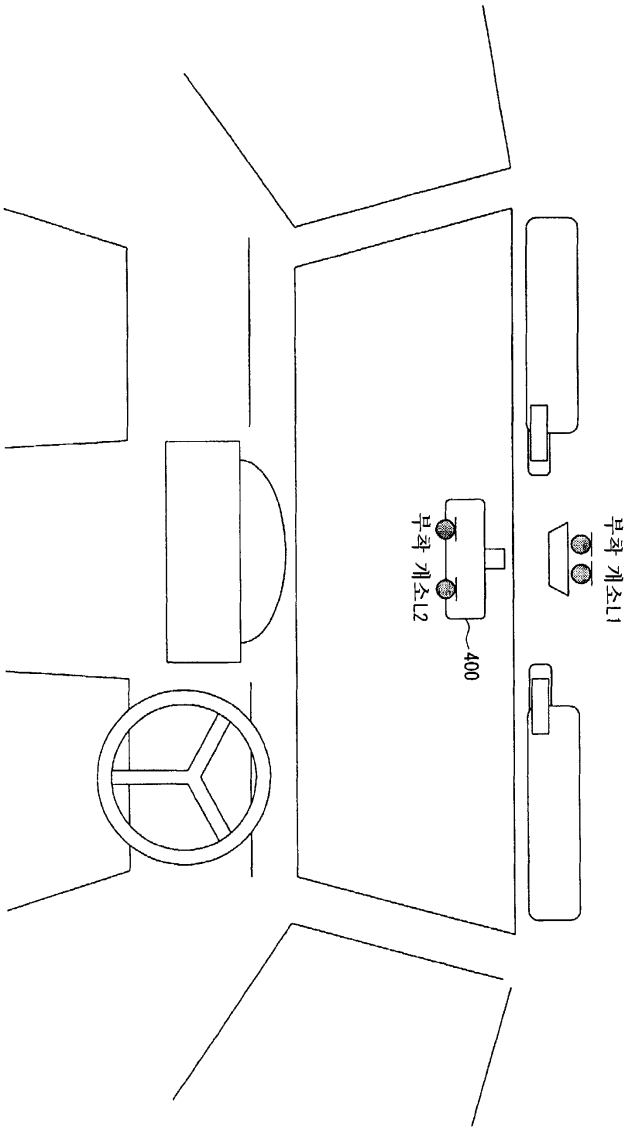
도면12


10
11
분리 처리
분리 후 출력

도면13



도면14



도면15

• 발화자	더미 헤드
• 발화자용	
— 운적석	정면 방향
— 조수석	정면 방향
• 인식 음성	제어어 128단어 운전석, 조수석 동시 발화(각 남녀 2조) 계 12패턴의 데이터
• 어휘 네트워크	128단어 병렬 네트워크(Garbage 모델 첨부)
• 음원 분리법	본 발명 방법
• 샘플링 주파수	11.025kHz

• 분리된 운전석 발화 음성, 및 조수석 발화 음성에 대하여 음성 인식률을 평가

도면16

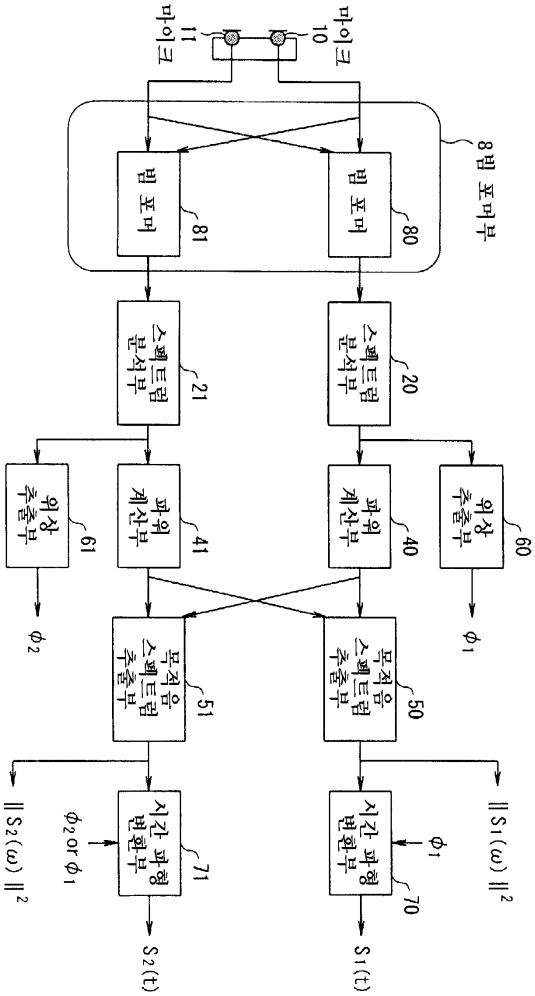
(a)

		정지 상태	주행 상태	비고
종래법1	차내 동시 발화 음성	29%	27%	1화자 대상
종래법2	최적값으로부터 갱신·목적음 추출	36%	–	이동 음원 대응 방법
본 방법	제안법 분리 처리후 음성	78%	78%	

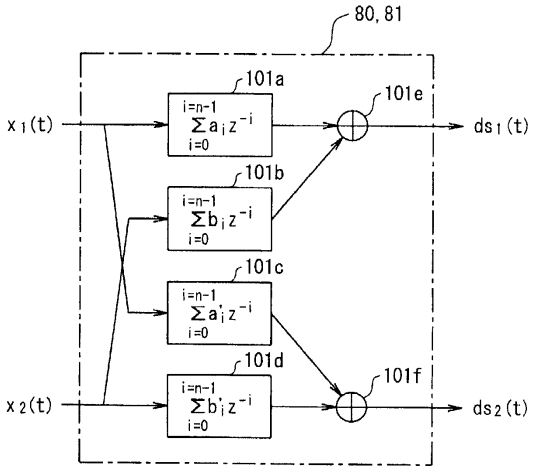
(b)

		정지 상태	비고
종래법1	차내 동시 발화 음성	93%	1화자 대상
본 방법	제안법 분리 처리후 음성	0%	

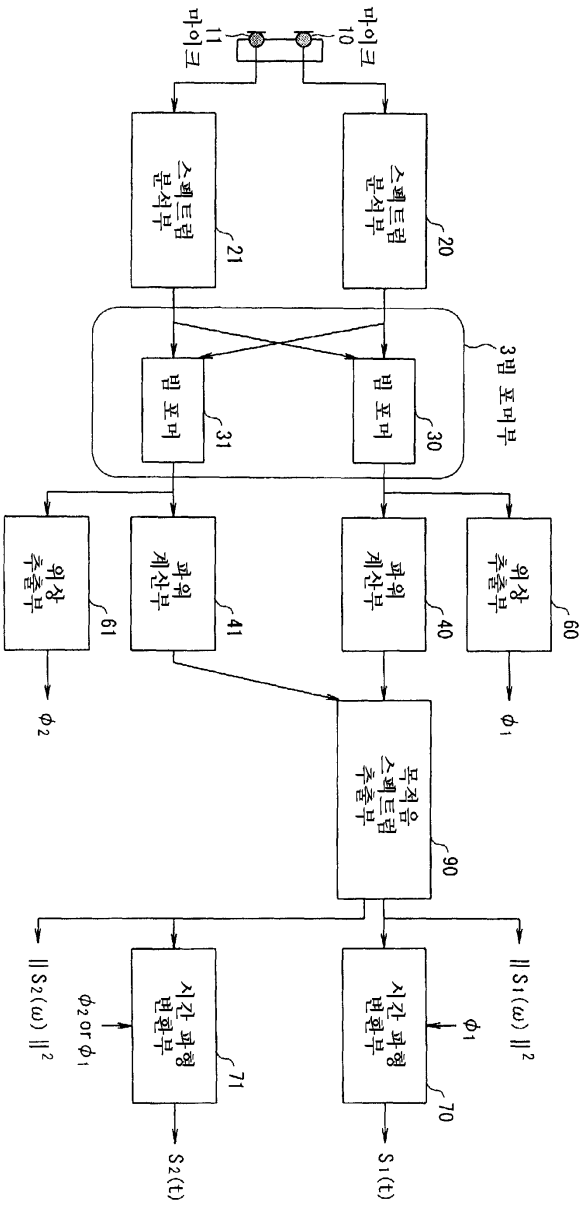
도면17



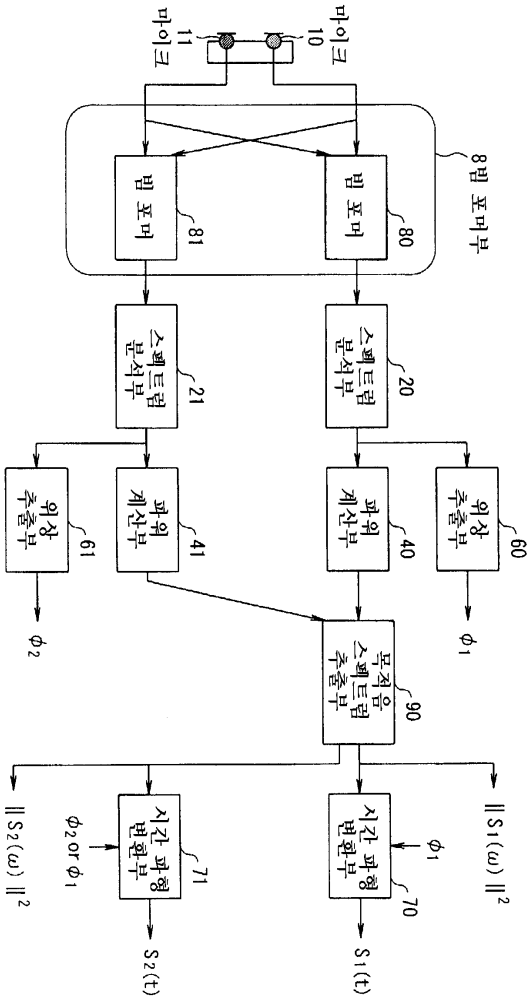
도면18



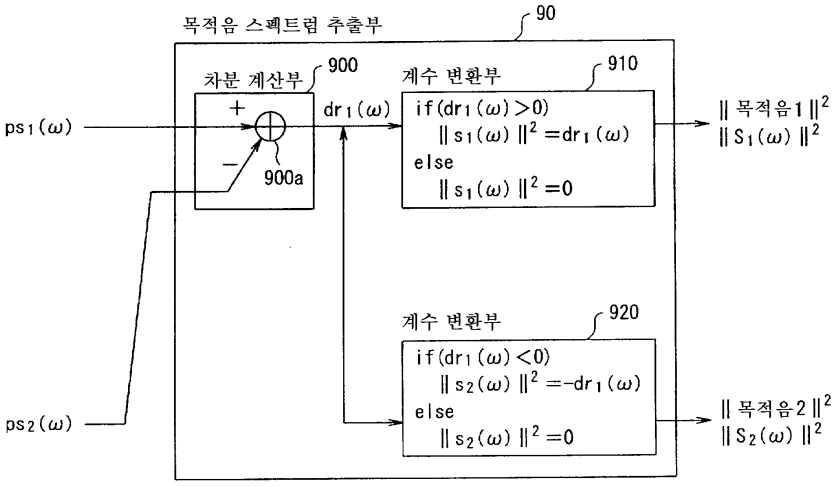
도면19


마이크 11
마이크 10
스펙트럼 분석부 21
스펙트럼 분석부 20
3빔 포머부
빔 포머 31
빔 포머 30
위상 추출부 61
파워 계산부 41
파워 계산부 40
위상 추출부 60
ϕ_2
ϕ_1
무잡음 스펙트럼 추출부 90
$\| S_2(\omega) \|^2$
ϕ_2 or ϕ_1
시간 파형 변환부 71
시간 파형 변환부 70
$\| S_1(\omega) \|^2$
ϕ_1
$S_2(t)$
$S_1(t)$

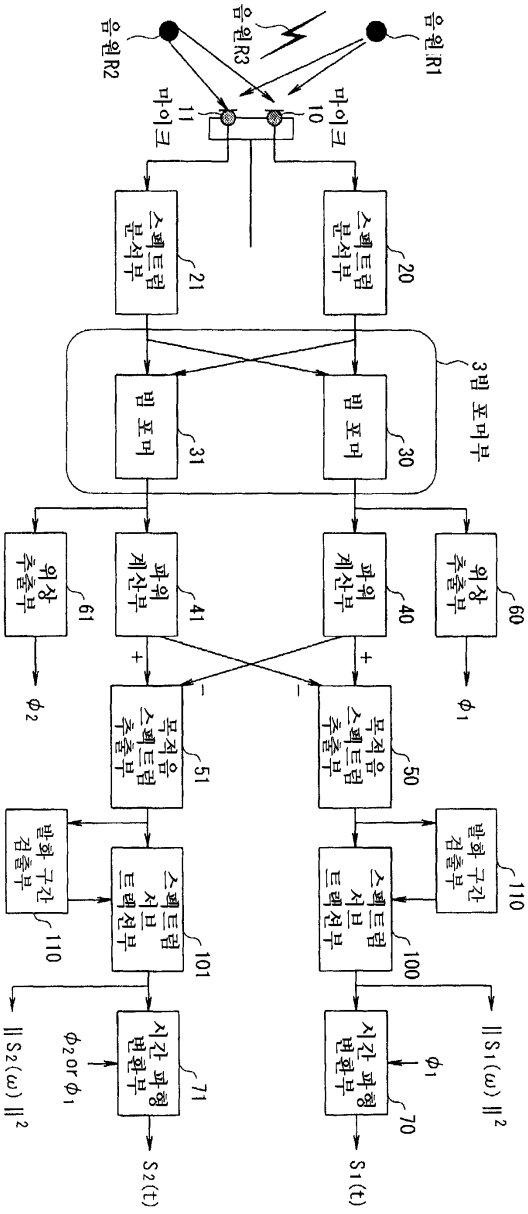
도면20


마이크 11
마이크 10
8빔 포머부
빔 포머 81
빔 포머 80
스펙트럼 분석부 21
스펙트럼 분석부 20
위상 추출부 61
파워 계산부 41
파워 계산부 40
위상 추출부 60
ϕ_2
ϕ_1
목적음 스펙트럼 추출부 90
$\|S_2(\omega)\|^2$
ϕ_2 or ϕ_1
시간 파형 변환부 71
시간 파형 변환부 70
ϕ_1
$\|S_1(\omega)\|^2$
$S_2(t)$
$S_1(t)$

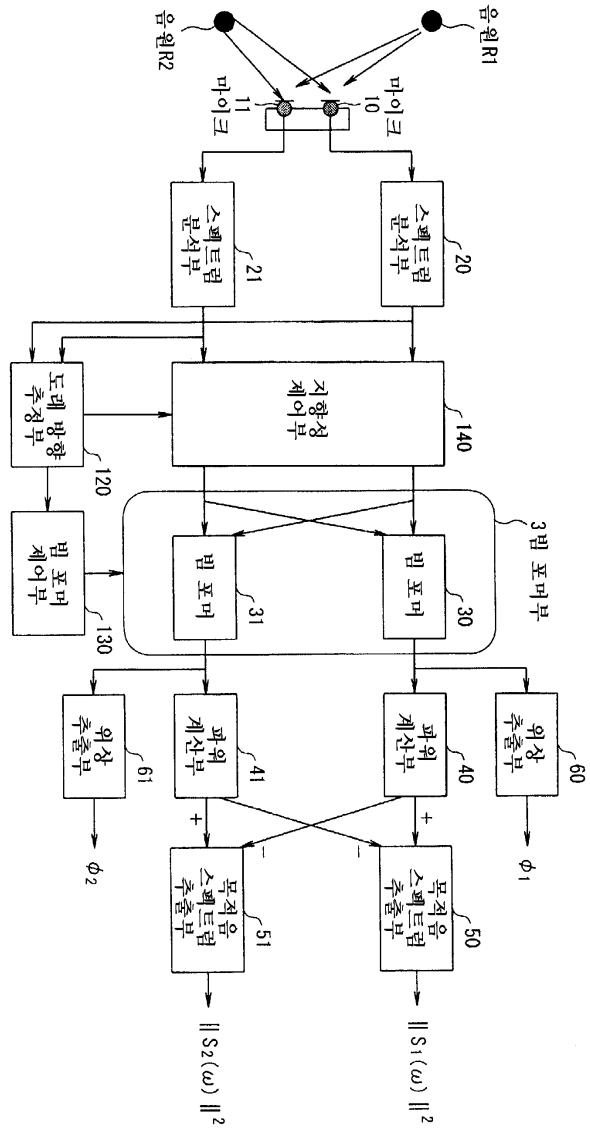
도면21


목적음 스펙트럼 추출부 90
차분 계산부 900
900a
$dr_1(\omega)$
계수 변환부 910
if($dr_1(\omega)>0$)
$\|s_1(\omega)\|^2 = dr_1(\omega)$
else
$\|s_1(\omega)\|^2 = 0$
$ps_1(\omega)$
$\|$목적음1$\|^2$
$\|S_1(\omega)\|^2$
계수 변환부 920
if($dr_1(\omega)<0$)
$\|s_2(\omega)\|^2 = -dr_1(\omega)$
else
$\|s_2(\omega)\|^2 = 0$
$ps_2(\omega)$
$\|$목적음2$\|^2$
$\|S_2(\omega)\|^2$

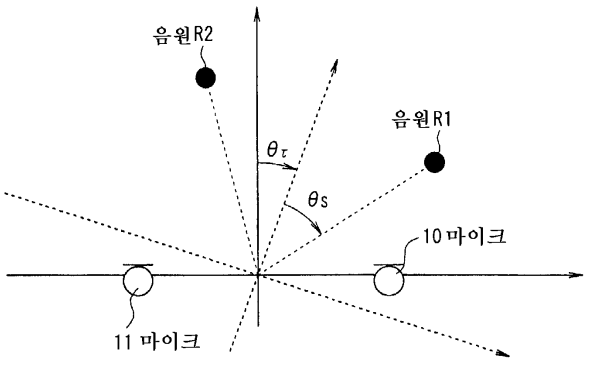
도면22


음원R1
음원R3
음원R2
마이크 10
마이크 11
스펙트럼 분석부 20
스펙트럼 분석부 21
3입 포머부
빔 포머 30
빔 포머 31
위상 추출부 60
파워 계산부 40
파워 계산부 41
위상 추출부 61
목적음 스펙트럼 추출부 50
목적음 스펙트럼 추출부 51
발화 구간 검출부 110
발화 구간 검출부 110
스펙트럼 서브 트랙션부 100
스펙트럼 서브 트랙션부 101
시간 파형 변환부 70
시간 파형 변환부 71
ϕ_1
ϕ_2
ϕ_2 or ϕ_1
$\|S_1(\omega)\|^2$
$\|S_2(\omega)\|^2$
$S_1(t)$
$S_2(t)$

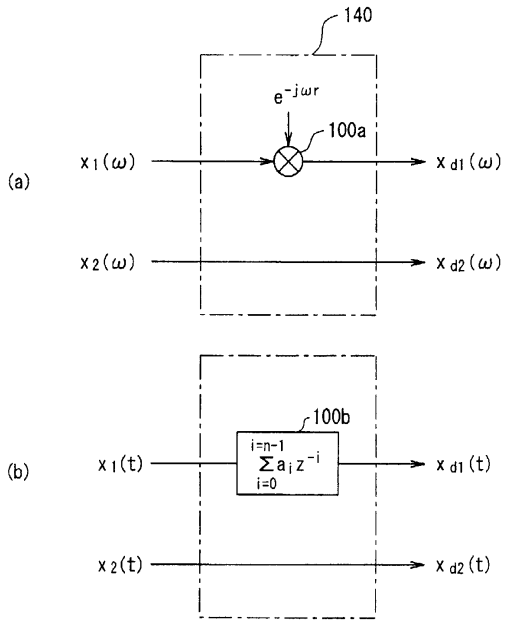
도면23


음원R2
음원R1
11 마이크
10 마이크
스펙트럼 분석부 21
스펙트럼 분석부 20
도래 방향 추정부 120
지향성 제어부 140
빔 포머 제어부 130
빔 포머 31
빔 포머 30
3 빔 포머부
위상 추출부 61
파워 계산부 41
파워 계산부 40
위상 추출부 60
목적음 스펙트럼 추출부 51
목적음 스펙트럼 추출부 50
ϕ_2
ϕ_1
$\|S_2(\omega)\|^2$
$\|S_1(\omega)\|^2$

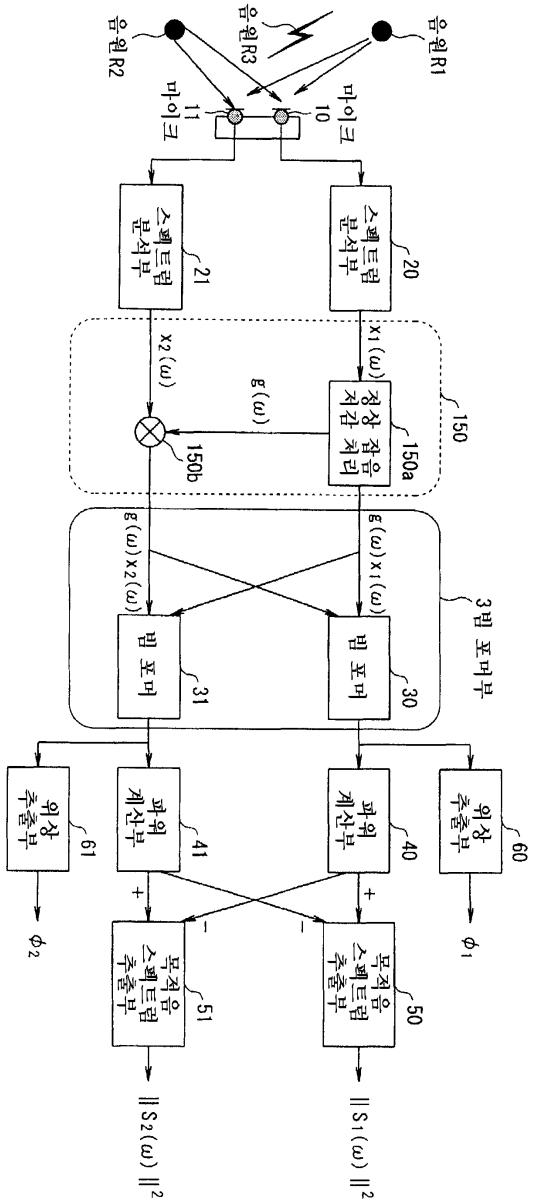
도면24


음원R2
음원R1
θ_τ
θ_S
10 마이크
11 마이크

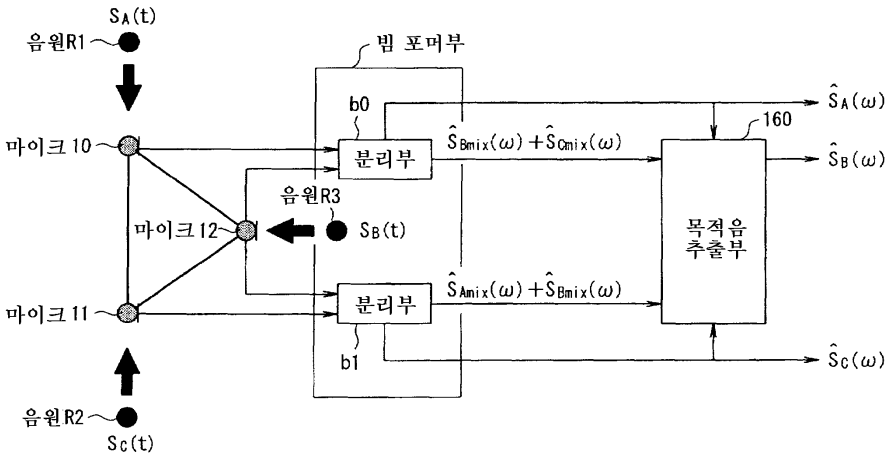
도면25


140
$e^{-j\omega r}$
100a
(a)
$x_1(\omega)$
$x_{d1}(\omega)$
$x_2(\omega)$
$x_{d2}(\omega)$
100b
(b)
$x_1(t)$
$\sum_{i=0}^{i=n-1} a_i z^{-i}$
$x_{d1}(t)$
$x_2(t)$
$x_{d2}(t)$

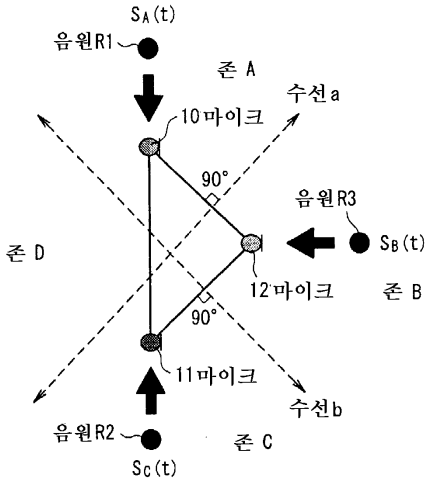
도면26


음원R1
음원R3
음원R2
마이크 10
마이크 11
스펙트럼 분석부 20
스펙트럼 분석부 21
$x_1(\omega)$
$x_2(\omega)$
$g(\omega)$
정상 잡음 처리 150a
150
150b
$g(\omega)x_1(\omega)$
$g(\omega)x_2(\omega)$
빔 포머 30
빔 포머 31
3빔 포머부
위상 추출부 60
파워 계산부 40
파워 계산부 41
위상 추출부 61
목적음 스펙트럼 추출부 50
목적음 스펙트럼 추출부 51
ϕ_1
ϕ_2
$\|S_1(\omega)\|^2$
$\|S_2(\omega)\|^2$

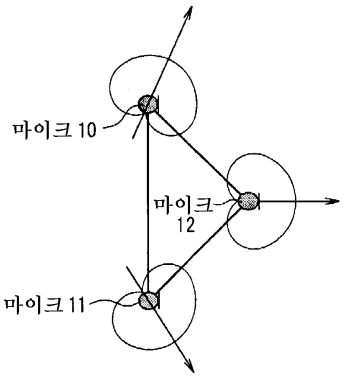
도면27



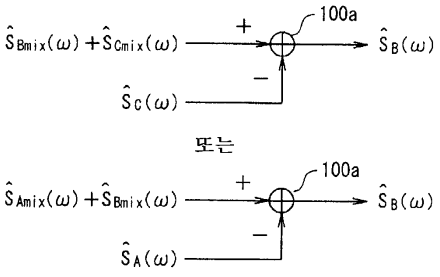
도면28



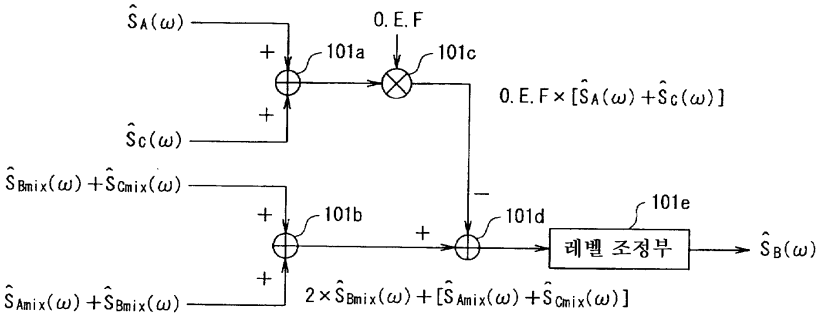
도면29



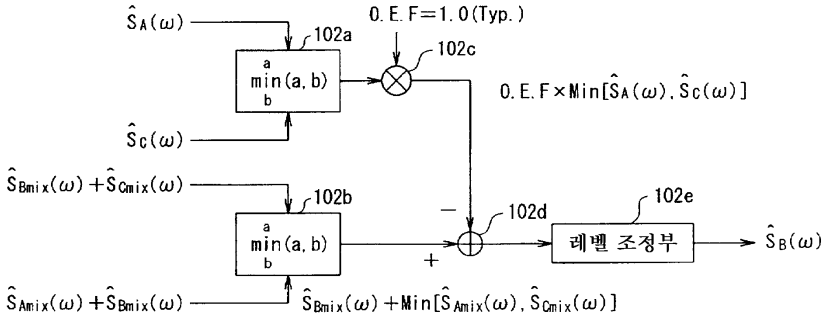
도면30



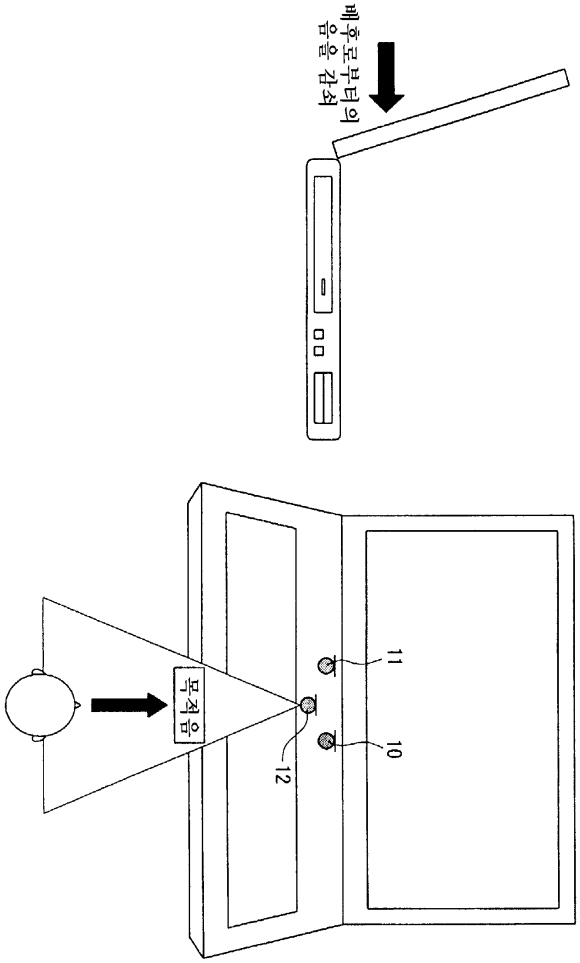
도면31



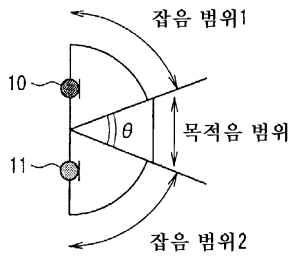
도면32



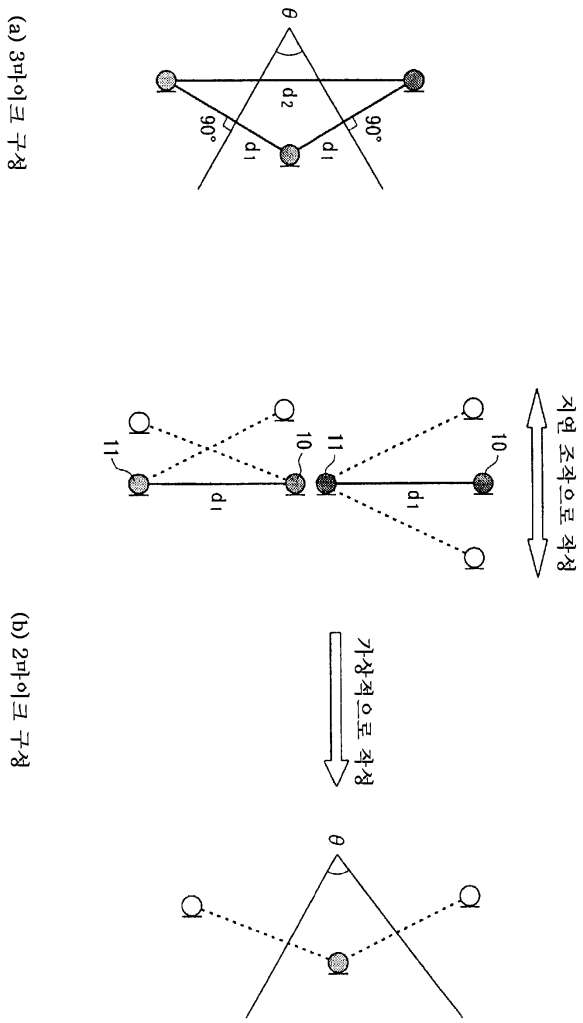
도면33



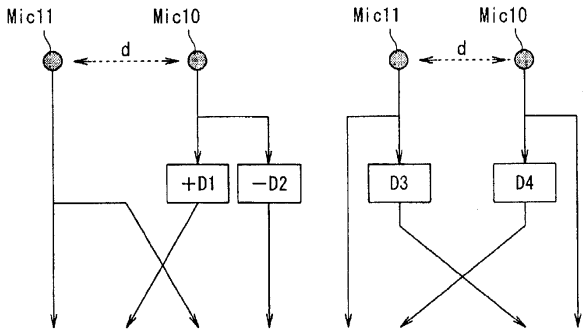
도면34



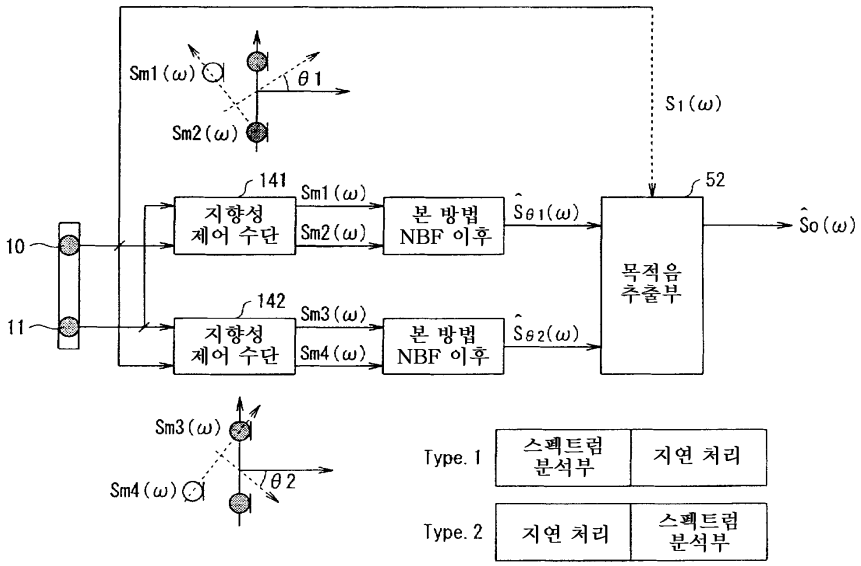
도면35


(a) 3마이크 구성
θ
d2
90°
d1
90°
d1
(b) 2마이크 구성
지연 조작으로 작성
11
10
11
10
d1
d1
가상적으로 작성
θ

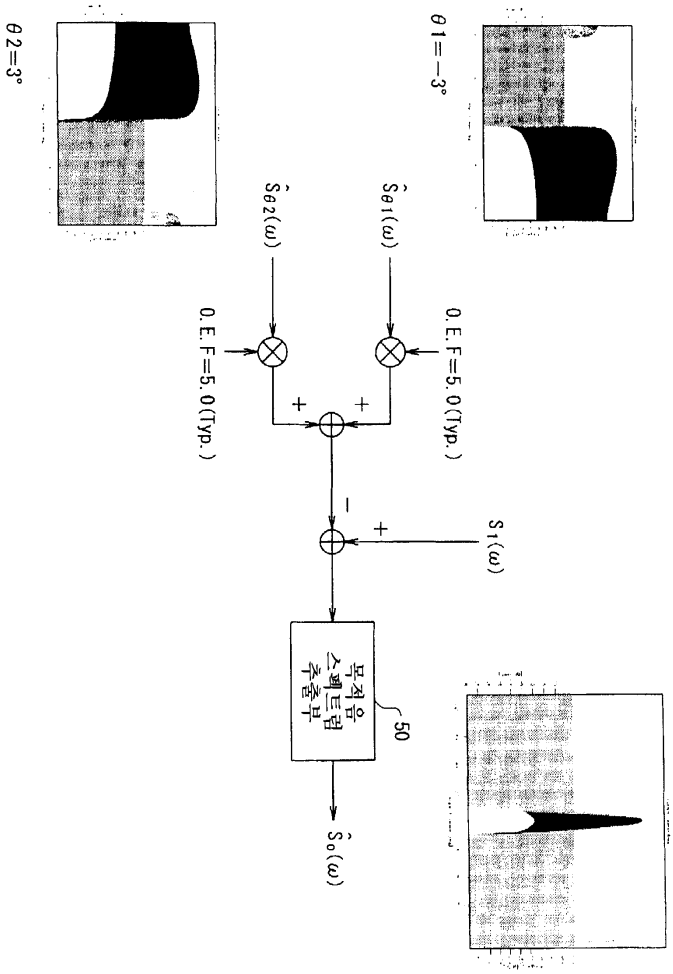
도면36


Mic11
Mic10
d
+D1
−D2
Mic11
Mic10
d
D3
D4

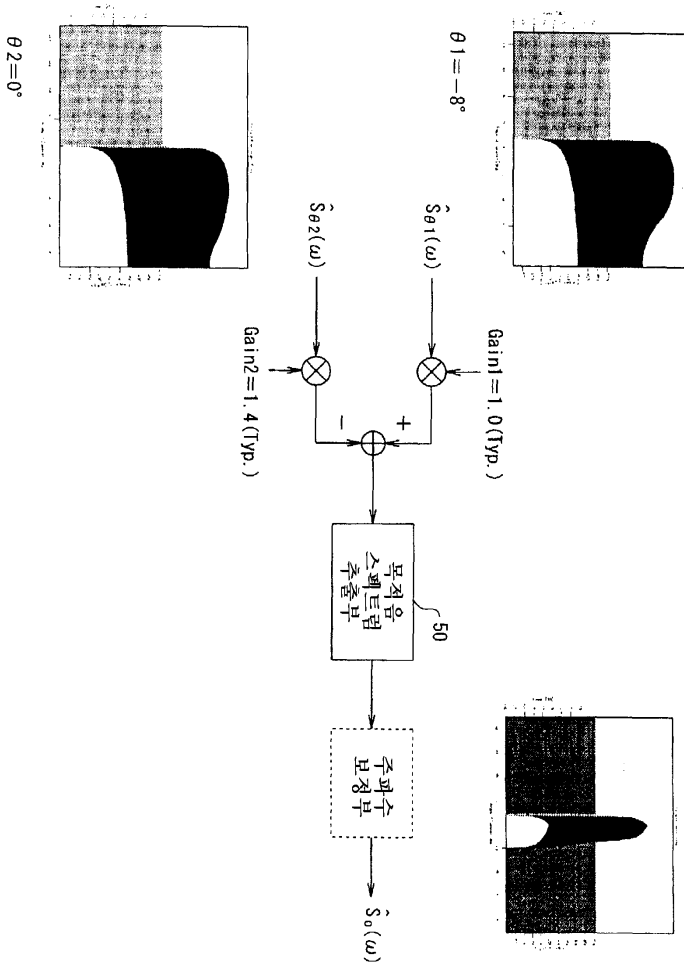
도면37


Sm1(ω)
θ1
Sm2(ω)
S1(ω)
141
Sm1(ω)
지향성 제어 수단
Sm2(ω)
본 방법 NBF 이후
Sθ1(ω)
52
So(ω)
목적음 추출부
10
11
142
Sm3(ω)
지향성 제어 수단
Sm4(ω)
본 방법 NBF 이후
Sθ2(ω)
Sm3(ω)
Sm4(ω)
θ2
Type. 1
스펙트럼 분석부
지연 처리
Type. 2
지연 처리
스펙트럼 분석부

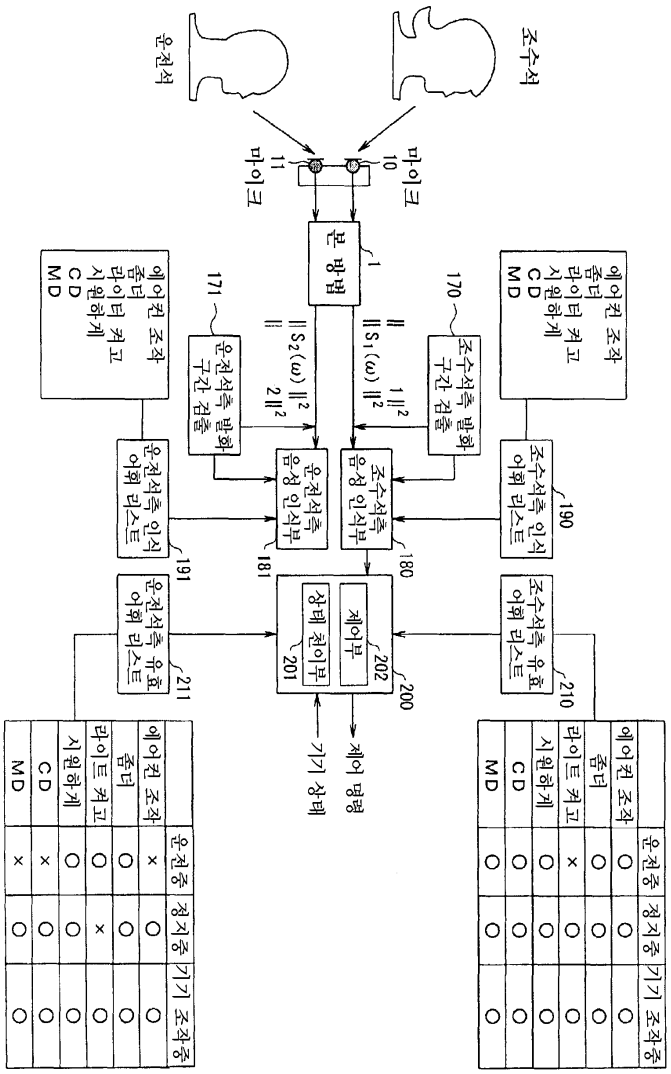
도면38


θ2=3°
θ1=−3°
Sθ2(ω)
Sθ1(ω)
0.E.F=5.0(Typ.)
0.E.F=5.0(Typ.)
+
+
−
+
S1(ω)
목적음 스펙트럼 추출부
50
So(ω)

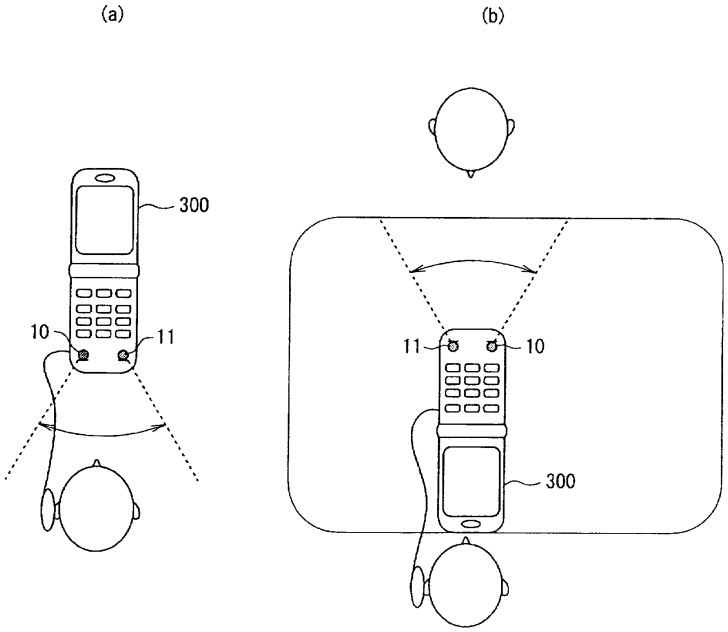
도면39


θ1=-8°
θ2=0°
$\hat{S}_{\theta 1}(\omega)$
$\hat{S}_{\theta 2}(\omega)$
Gain1=1.0(Typ.)
Gain2=1.4(Typ.)
+
-
목적음 스펙트럼 추출부
50
주파수 보정부
$\hat{S}_o(\omega)$

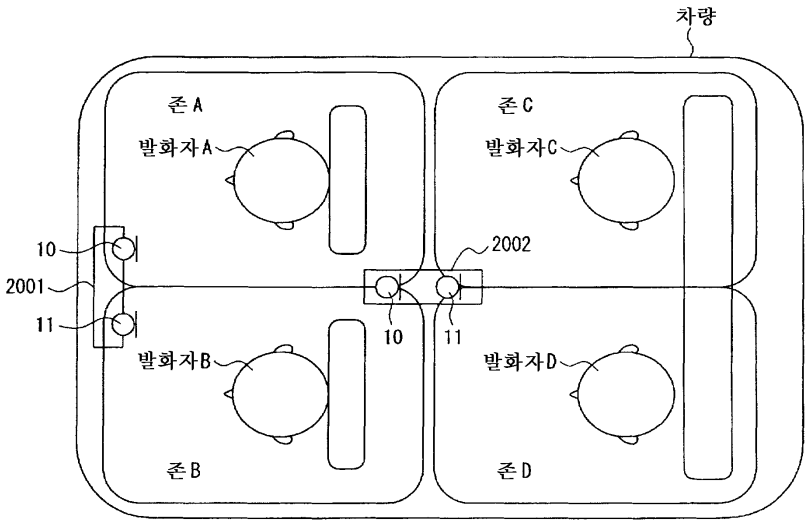
도면40


조수석
운전석
마이크 10
마이크 11
본 방법 1
조수석측 발화 구간 검출 170
운전석측 발화 구간 검출 171
조수석측 음성 인식부 180
운전석측 음성 인식부 181
상태 천이부 201
제어부 202
제어부 200
조수석측 유효 어휘 리스트 210
운전석측 유효 어휘 리스트 211
조수석측 인식 어휘 리스트 190
운전석측 인식 어휘 리스트 191
기기 상태
제어 명령
에어컨 조작
줌터
라이트 켜고
시원하게
CD
MD
운전중
정지중
기기 조작중

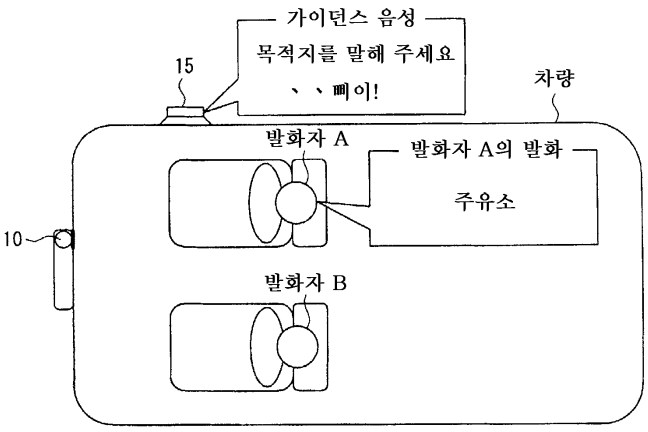
도면41


(a)
(b)
300
10
11
11
10
300

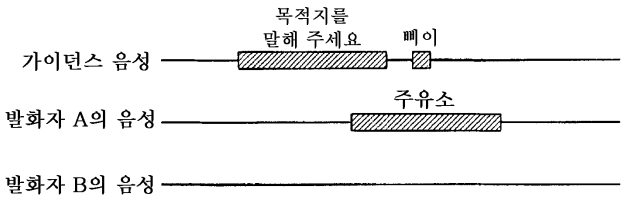
도면42


차량
존 A
발화자A
존 C
발화자C
10
2001
11
2002
10
11
발화자B
발화자D
존 B
존 D

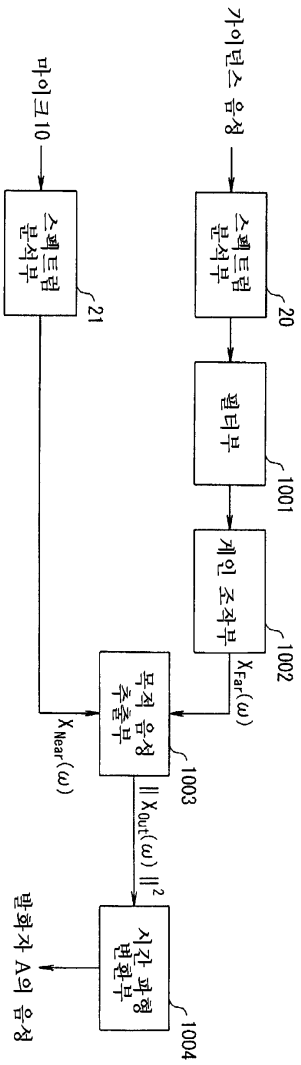
도면43


가이던스 음성
목적지를 말해 주세요
、、삐이!
15
차량
발화자 A
발화자 A의 발화
주유소
10
발화자 B

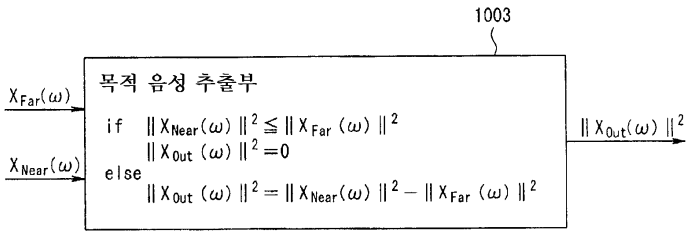
도면44


목적지를
말해 주세요
삐이
가이던스 음성
주유소
발화자 A의 음성
발화자 B의 음성

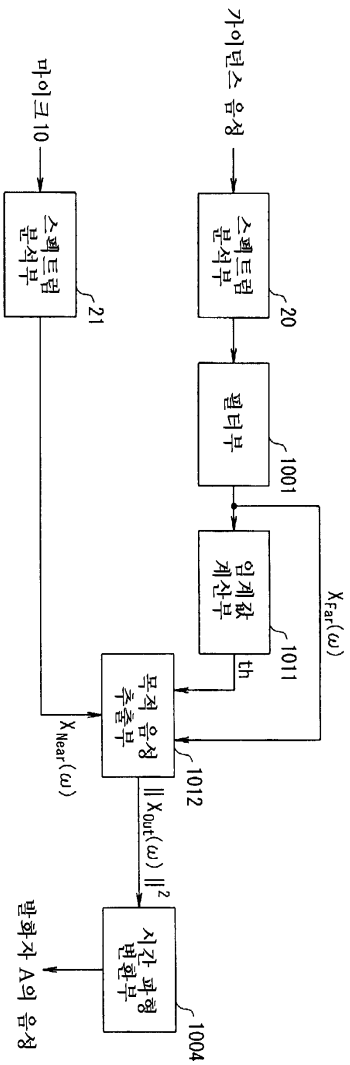
도면45


가이던스 음성
마이크10
스펙트럼 분석부
20
스펙트럼 분석부
21
필터부
1001
게인 조작부
1002
$X_{Far}(\omega)$
목적 음성 추출부
1003
$X_{Near}(\omega)$
$\| X_{Out}(\omega) \|^2$
시간 파형 변환부
1004
발화자 A의 음성

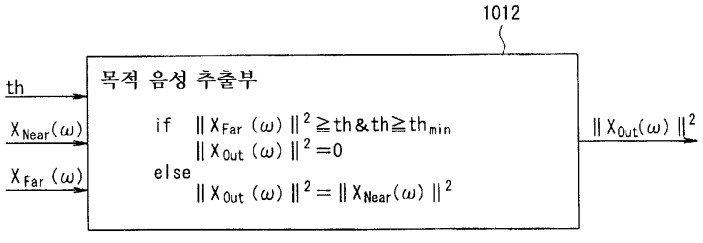
도면46


1003
목적 음성 추출부
$X_{Far}(\omega)$
$X_{Near}(\omega)$
if $\| X_{Near}(\omega) \|^2 \leq \| X_{Far}(\omega) \|^2$
$\| X_{Out}(\omega) \|^2 = 0$
else
$\| X_{Out}(\omega) \|^2 = \| X_{Near}(\omega) \|^2 - \| X_{Far}(\omega) \|^2$
$\| X_{Out}(\omega) \|^2$

도면47

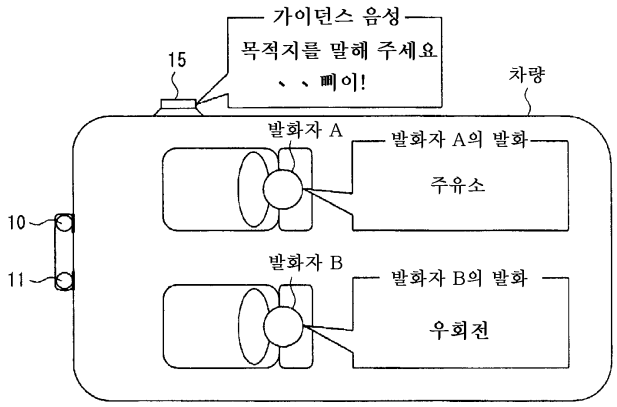


도면48

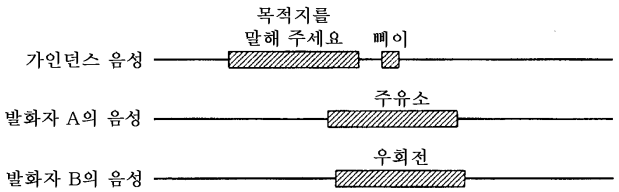
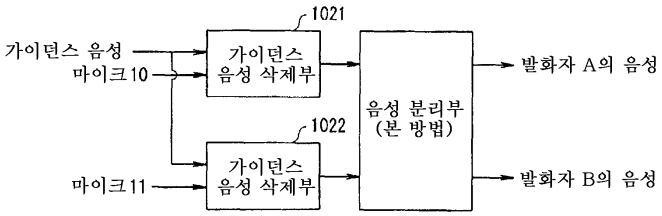


th_{min} : $X_{Far}(\omega)$이 유효한 입력이라고 판단하기 위한 임계값

도면49



도면50



도면51