

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일

2022년 5월 12일 (12.05.2022)



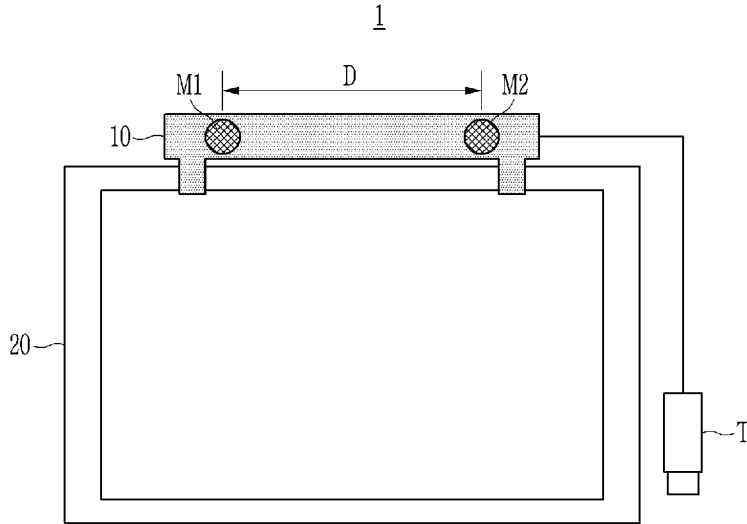
(10) 국제공개번호

WO 2022/097919 A1

- (51) 국제특허분류: *H04B 7/06* (2006.01) *G06N 3/04* (2006.01) *G06N 3/08* (2006.01) (PARK, Sang-Hyun); 41190 대구시 동구 신암북로7길 59, 303호, Daegu (KR).
- (21) 국제출원번호: PCT/KR2021/013328 (74) 대리인: 유미특허법인 (YOU ME PATENT AND LAW FIRM); 06134 서울시 강남구 테헤란로 115, Seoul (KR).
- (22) 국제출원일: 2021년 9월 29일 (29.09.2021) (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (25) 출원언어: 한국어 (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM,
- (26) 공개언어: 한국어
- (30) 우선권정보: 10-2020-0146191 2020년 11월 4일 (04.11.2020) KR
- (71) 출원인: 주식회사 딥히어링 (DEEPHEARING INC.) [KR/KR]; 34134 대전시 유성구 대학로 99, 2521호, Daejeon (KR).
- (72) 발명자: 안강현 (AHN, Kang Hun); 34140 대전시 유성구 어은로 57, 116동 205호, Daejeon (KR). 박상현

(54) Title: BEAMFORMING METHOD AND BEAMFORMING SYSTEM USING NEURAL NETWORK

(54) 발명의 명칭: 뉴럴 네트워크를 이용한 빔포밍 방법 및 빔포밍 시스템



(57) Abstract: Provided are a beamforming method and a beamforming system. The beamforming method may comprise the steps of: receiving a first sound signal and a second sound signal by respectively using a first microphone and a second microphone spaced apart from the first microphone by a pre-determined distance; obtaining a Fourier transform result for each of the first sound signal and the second sound signal; obtaining a phase difference between the first sound signal and the second sound signal from the Fourier transform result; performing an arithmetic operation by inputting the phase difference to a beamforming model by using a neural network; performing element multiplication on an operation result of the neural processor and the Fourier transform result for the first sound signal; and outputting a result of the element multiplication.

(57) 요약서: 빔포밍 방법 및 빔포밍 시스템이 제공된다. 빔포밍 방법은, 제1 마이크와, 상기 제1 마이크로부터 미리 정해진 거리만큼 이격되어 배치된 제2 마이크를 이용하여 제1 소리 신호 및 제2 소리 신호를 각각 수신하는 단계; 상기 제1 소리 신호 및 상기 제2 소리 신호 각각에 대한 푸리에 변환 결과를 획득하는 단계; 상기 푸리에 변환 결과로부터 상기 제1 소리 신호와 상기 제2 소리 신호의 위상차를 획득하는 단계; 뉴럴 프로세서를 이용하여 상기 위상차를 빔포밍 모델에 입력하여 연산하는 단계; 상기 뉴럴 프로세서의 연산 결과와 상기 제1 소리 신호에 대한 푸리에 변환 결과에 대해 원소곱을 수행하는 단계; 및 상기 원소곱 결과를 출력하는 단계를 포함할 수 있다.

[다음 쪽 계속]

WO 2022/097919 A1

ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 뉴럴 네트워크를 이용한 빔포밍 방법 및 빔포밍 시스템

기술분야

- [1] 본 발명은 뉴럴 네트워크를 이용한 빔포밍 방법 및 빔포밍 시스템에 관한 것이다.

배경기술

- [2] 칵테일 파티 효과(cocktail party effect)는 파티의 참석자들이 시끄러운 주변 소음이 있는 방에 있음에도 불구하고 대화자와의 이야기를 선택적으로 집중하여 잘 받아들이는 현상을 말한다. 기계에서 이러한 능력, 즉 빔포밍(beamforming)을 구현하는 것은 칵테일 파티 문제로 알려져 왔으며 최근에는 칵테일 파티 문제 해결을 위해 뉴럴 네트워크를 이용하려는 시도가 늘어나고 있다. 빔포밍 기법의 성능을 높이는 것은 오디오 관련 전자 제품의 성능과 밀접한 관계가 있으며, 특히 보청기와도 관련이 있기 때문에 매우 중요한 사회 문제라는 의미도 가질 수 있다.
- [3] 빔포밍은 2 개 이상의 마이크를 이용하여 획득한 소리를 분석해 대상 신호를 재구성하는 과정을 의미할 수 있다. 이를 위해, 각 마이크에서 오는 신호에 대해 인위적인 지연 시간을 만들어 합치는 것과 같이, 주어진 제한조건을 만족시키면서 들어오는 소리의 크기를 최소화하는 기법이 오래전부터 사용되어 왔다. 최근에는 뉴럴 네트워크를 이용한 MVDR(minimum variance distortionless response) 빔포머의 성능 향상에 관한 연구 또는 빔포머를 구현하는 뉴럴 네트워크 모델의 훈련방법 등이 활발히 연구되고 있다.

발명의 상세한 설명

기술적 과제

- [4] 본 발명이 해결하고자 하는 과제는, 규칙기반 빔포밍에서 공간정보를 얻기위해 많은 양의 연산이 요구되는 한계를 극복하기 위해 뉴럴 네트워크를 사용하되, 빔포밍에 최적화되도록 뉴럴 네트워크 구조를 설계함으로써 파라미터량 및 연산량을 최소화할 수 있는, 뉴럴 네트워크를 이용한 빔포밍 방법 및 빔포밍 시스템을 제공하기 위한 것이다.

과제 해결 수단

- [5] 본 발명의 일 실시 예에 따른 빔포밍 방법은, 제1 마이크와, 상기 제1 마이크로부터 미리 정해진 거리만큼 이격되어 배치된 제2 마이크를 이용하여 제1 소리 신호 및 제2 소리 신호를 각각 수신하는 단계; 상기 제1 소리 신호 및 상기 제2 소리 신호 각각에 대한 푸리에 변환 결과를 획득하는 단계; 상기 푸리에 변환 결과로부터 상기 제1 소리 신호와 상기 제2 소리 신호의 위상차를 획득하는 단계; 뉴럴 프로세서를 이용하여 상기 위상차를 빔포밍 모델에 입력하여

연산하는 단계; 상기 뉴럴 프로세서의 연산 결과와 상기 제1 소리 신호에 대한 푸리에 변환 결과에 대해 원소곱을 수행하는 단계; 및 상기 원소곱 결과를 출력하는 단계를 포함할 수 있다.

- [6] 상기 원소곱을 수행하는 단계는, 상기 원소곱을 수행하기 전에 상기 연산 결과에 대해 마스크(mask)를 적용하는 단계를 더 포함할 수 있다.
- [7] 상기 원소곱을 수행하는 단계는, 상기 원소곱을 수행한 후에 이득 제어(gain control)를 수행하는 단계를 더 포함할 수 있다.
- [8] 상기 미리 정해진 거리는 10 cm 내지 14 cm일 수 있다.
- [9] 상기 빔포밍 방법은, 상기 위상차를 이용하여 상기 빔포밍 모델을 학습시키는 단계를 더 포함할 수 있다.
- [10] 본 발명의 일 실시 예에 따른 빔포밍 시스템은, 제1 소리 신호를 수신하는 제1 마이크; 상기 제1 마이크로부터 미리 정해진 거리만큼 이격되어 배치되어 제2 소리 신호를 수신하는 제2 마이크; 상기 제1 소리 신호에 대한 푸리에 변환 결과를 획득하는 제1 STFT 모듈; 상기 제2 소리 신호에 대한 푸리에 변환 결과를 획득하는 제2 STFT 모듈; 상기 푸리에 변환 결과로부터 상기 제1 소리 신호와 상기 제2 소리 신호의 위상차를 획득하는 위상차 획득 모듈; 상기 위상차를 입력받아 빔포밍 모델을 이용하여 뉴럴 네트워크 연산을 수행하는 뉴럴 프로세서; 상기 뉴럴 프로세서의 연산 결과와 상기 제1 소리 신호에 대한 푸리에 변환 결과에 대해 원소곱을 수행하는 원소곱 모듈; 및 상기 원소곱 결과를 출력하는 출력 모듈을 포함할 수 있다.
- [11] 상기 빔포밍 시스템은, 상기 원소곱을 수행하기 전에 상기 연산 결과에 대해 마스크를 적용하는 마스크 모듈을 더 포함할 수 있다.
- [12] 상기 빔포밍 시스템은, 상기 원소곱을 수행한 후에 이득 제어를 수행하는 이득 제어 모듈을 더 포함할 수 있다.
- [13] 상기 미리 정해진 거리는 10 cm 내지 14 cm일 수 있다.
- [14] 상기 빔포밍 시스템은, 상기 위상차를 이용하여 상기 빔포밍 모델을 학습시키는 학습 모듈을 더 포함할 수 있다.

발명의 효과

- [15] 본 발명의 실시 예들에 따르면, 다양한 소음 환경에 대한 조향 벡터와 공간상관행렬을 연산할 필요없이 위상차만을 이용하여 마이크에서 수신되는 음성을 복원할 수 있어, 빔포밍을 효율적으로 구현할 수 있다. 또한, 경량화된 오토인코더의 적용이 가능하여 경량화 및 소형화를 가능하게 할 수 있다.

도면의 간단한 설명

- [16] 도 1은 본 발명의 일 실시 예에 따른 빔포밍 시스템을 설명하기 위한 도면이다.
- [17] 도 2는 본 발명의 실시 예들에 따른 빔포밍 장치를 구현하기 위한 컴퓨터 장치를 설명하기 위한 도면이다.
- [18] 도 3은 본 발명의 일 실시 예에 따른 빔포밍 방법을 설명하기 위한 도면이다.

- [19] 도 4는 본 발명의 일 실시 예에 따른 빔포밍 방법을 설명하기 위한 도면이다.
- [20] 도 5는 본 발명의 일 실시 예에 따른 빔포밍 방법에 대한 일 구현 예를 설명하기 위한 도면이다.
- [21] 도 6은 본 발명의 일 실시 예에 따른 빔포밍 시스템을 설명하기 위한 도면이다.
- [22] 도 7은 본 발명의 일 실시 예에 따른 빔포밍 시스템을 설명하기 위한 도면이다.
- [23] 도 8은 본 발명의 일 실시 예에 따른 빔포밍 시스템을 설명하기 위한 도면이다.
- [24] 도 9 및 도 10은 본 발명의 실시 예들에 따른 빔포밍 방법 및 빔포밍 시스템의 유리한 효과를 설명하기 위한 도면이다.

발명의 실시를 위한 형태

- [25] 아래에서는 첨부한 도면을 참조하여 본 발명의 실시 예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [26] 명세서 및 청구범위 전체에서, 어떤 부분이 어떤 구성 요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성 요소를 제외하는 것이 아니라 다른 구성 요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다. 또한, 아래에서 설명되는 실시 예들에 따른 빔포밍 방법, 빔포밍 장치 및 빔포밍 시스템 중 적어도 일부는 프로그램 또는 소프트웨어로 구현될 수 있고, 프로그램 또는 소프트웨어는 컴퓨터로 판독 가능한 매체에 저장될 수 있다.
- [27] 도 1은 본 발명의 일 실시 예에 따른 빔포밍 시스템을 설명하기 위한 도면이다.
- [28] 도 1을 참조하면, 본 발명의 일 실시 예에 따른 빔포밍 시스템(1)은 제1 마이크(M1), 제2 마이크(M2) 및 연결 단자(T)를 포함하는 빔포밍 장치(10)와 모니터(20)를 포함할 수 있다.
- [29] 빔포밍 장치(10)는 모니터(20)에 부착되어 마이크(M1, M2)를 이용하여 소리를 수신할 수 있다. 예를 들어, 빔포밍 장치(10)는 마이크(M1, M2)를 이용하여 모니터(20) 앞에서 화상 회의에 참여하는 사람의 목소리를 수신할 수 있다. 특히, 빔포밍 장치(10)는 주변 소음이 많은 환경에서 화상 회의를 참여하는 사람의 목소리를 수신할 수 있다.
- [30] 빔포밍 장치(10)는, 마이크(M1, M2)를 이용하여 수신한 소리 신호에 대해 빔포밍을 수행한 후, 그 결과로서 획득한 빔포밍된 소리 신호를 출력할 수 있다. 예를 들어, 빔포밍 장치(10)는 주변 소음이 많은 환경에서 화상 회의에 참여하는 사람의 목소리를 분별하여 다른 컴퓨팅 장치(예컨대 모니터(20))가 연결된

개인용 컴퓨터)에 제공할 수 있다. 그러면 해당 컴퓨터 장치는 분별된 사람의 목소리를, 예컨대 다른 화상 회의 참여자에게 제공할 수 있다. 빔포밍 장치(10)가 사람의 목소리를 분별한 출력 신호를 다른 컴퓨팅 장치에 제공하기 위해 연결 단자(T)가 사용될 수 있으며, 본 실시 예에서 연결 단자(T)는 USB(Universal Serial Bus) 단자일 수 있으나, 본 발명의 범위가 이에 제한되는 것은 아니다.

- [31] 본 실시 예에서, 제1 마이크(M1)와 제2 마이크(M2)는 미리 정해진 거리(D)만큼 이격되어 배치될 수 있다. 예를 들어, 제1 마이크(M1)는 제1 측(예를 들어 좌측)에서 화상 회의에 참여하는 사람의 목소리와 주변 소음(즉, 제1 소리 신호)을 수신할 수 있고, 제2 마이크(M2)는, 제1 마이크(M1)로부터 미리 정해진 거리(D)만큼 떨어진 제2 측(예를 들어 우측)에서 화상 회의에 참여하는 사람의 목소리와 주변 소음(즉, 제2 소리 신호)을 수신할 수 있다.
- [32] 본 실시 예에서, 제1 마이크(M1)와 제2 마이크(M2) 사이의 미리 정해진 거리(D)는 10 cm 내지 14 cm, 바람직하게는 12 cm일 수 있으나, 본 발명의 범위가 이에 제한되는 것은 아니다.
- [33] 도 2는 본 발명의 실시 예들에 따른 빔포밍 장치를 구현하기 위한 컴퓨팅 장치를 설명하기 위한 도면이다.
- [34] 도 2를 참조하면, 본 발명의 실시 예들에 따른 빔포밍 장치(10)를 구현하기 위한 컴퓨팅 장치는 프로세서(100), 뉴럴 프로세서(110), 메모리(120), 출력 모듈(130), 제1 마이크(M1) 및 제2 마이크(M2)를 포함할 수 있으며, 해당 컴퓨팅 장치는 본 발명의 실시 예들에 따른 빔포밍 방법을 수행하기 위해 동작할 수도 있다. 프로세서(100), 뉴럴 프로세서(110), 메모리(120), 출력 모듈(130), 제1 마이크(M1) 및 제2 마이크(M2)는 버스(190)를 통해 서로 데이터를 주고 받을 수 있다.
- [35] 프로세서(100)는 빔포밍 장치(10)에 대한 전반적인 제어를 수행하며, 뉴럴 프로세서(110)와 함께 또는 뉴럴 프로세서(110)와 독립적으로, 본 명세서에서 설명하는 기능들 및 방법들을 수행할 수 있다. 프로세서(100)는, AP(Application Processor), CPU(Central Processing Unit), GPU(Graphic Processing Unit) 등과 같은 다양한 종류의 프로세서로 구현될 수 있으며, 본 발명의 범위는 특정 프로세서로 제한되지 않는다.
- [36] 뉴럴 프로세서(110)는 본 명세서에서 설명하는 기능들 및 방법들 중 특히 뉴럴 네트워크 연산을 수행할 수 있다. 예를 들어, 뉴럴 프로세서(110)는 본 명세서에서 설명하는 빔포밍 모델을 이용한 연산을 수행할 수 있다. 여기서 뉴럴 네트워크는 CNN(Convolutional Neural Network)를 포함할 수 있으나, 본 발명의 범위가 이에 제한되는 것은 아니다.
- [37] 메모리(120)는 프로세서(100) 또는 뉴럴 프로세서(110)가 연산을 위해 필요로 하는 데이터 또는 구동되는 애플리케이션을 적재할 수 있다. DRAM(Dynamic Random Access Memory), SRAM(Static Random Access Memory)를 비롯한 다양한 종류의 휘발성 메모리를 포함할 수 있거나, 플래시 메모리와 같은 비휘발성 메모리를 포함할 수 있거나, 휘발성 메모리와 비휘발성 메모리의 조합을 포함할

수도 있다.

- [38] 출력 모듈(130)은 빔포밍 장치(10)가 마이크(M1, M2)를 이용하여 수신한 소리 신호에 대해 빔포밍을 수행한 후, 그 결과로서 획득한 빔포밍된 소리 신호를 출력하기 위한 임의의 입출력 인터페이스 장치를 포함할 수 있다.
- [39] 또한, 본 발명의 실시 예들에 따른 빔포밍 방법, 빔포밍 장치 및 빔포밍 시스템 중 적어도 일부는 키 제어 방법, 키 제어 장치 및 사용자 기기 중 적어도 일부는 컴퓨팅 장치에서 실행되는 프로그램 또는 소프트웨어로 구현될 수 있고, 프로그램 또는 소프트웨어는 컴퓨터로 판독 가능한 매체에 저장될 수 있다. 또한, 본 발명의 실시 예들에 따른 빔포밍 방법, 빔포밍 장치 및 빔포밍 시스템 중 적어도 일부는 컴퓨팅 장치와 전기적으로 접속될 수 있는 하드웨어로 구현될 수도 있다.
- [40] 도 1에서 설명한 바와 같은 빔포밍 장치(10)는 모니터(20)에 부착되도록 구현되며, 사람의 목소리를 분별한 출력 신호를 제공하기 위해 다른 컴퓨팅 장치에 연결될 수 있다. 여기서 특히 주목할 점은, 빔포밍 장치(10)는 뉴럴 프로세서(110)를 자체적으로 보유하고 있어, 다른 컴퓨팅 장치의 연산 자원을 사용하지 않고도, 자신의 뉴럴 프로세서(110)를 이용하여 주변 소음이 많은 환경에서 화상 회의에 참여하는 사람의 목소리를 분별하기 위해 뉴럴 네트워크 연산을 수행할 수 있다는 점이다.
- [41] 이제 도 3 내지 도 5를 참조하여, 도 3은 본 발명의 실시 예들에 따른 빔포밍 방법에 대해 설명하도록 한다.
- [42] 도 3은 본 발명의 일 실시 예에 따른 빔포밍 방법을 설명하기 위한 도면이다.
- [43] 도 3을 참조하면, 본 발명의 일 실시 예에 따른 빔포밍 방법(30)에서, 제1 마이크(M1)는 제1 측으로부터 제1 소리 신호(S1)를 수신할 수 있다. 예를 들어, 제1 마이크(M1)는 제1 측으로부터 화상 회의에 참여하는 사람의 목소리와 주변 소음을 포함하는 제1 소리 신호(S1)를 수신하여, 이를 제1 STFT 모듈(300)(도면 상에서는 L_STFT 모듈로 표시함)에 전달할 수 있다.
- [44] 제1 STFT 모듈(300)은, 제1 마이크(M1)로부터 수신한 제1 소리 신호(S1)에 대해 푸리에 변환 연산을 수행하고, 제1 소리 신호(S1)에 대한 푸리에 변환 결과(P1)를 획득할 수 있다.
- [45] 한편, 제1 마이크(M1)로부터 미리 정해진 거리(D)만큼 이격되어 배치되는 제2 마이크(M2)는 제2 측으로부터 제2 소리 신호(S2)를 수신할 수 있다. 예를 들어, 제2 마이크(M2)는 제2 측으로부터 화상 회의에 참여하는 사람의 목소리와 주변 소음을 포함하는 제2 소리 신호(S2)를 수신하여, 이를 제2 STFT 모듈(301)(도면 상에서는 R_STFT 모듈로 표시함)에 전달할 수 있다.
- [46] 제2 STFT 모듈(301)은, 제2 마이크(M2)로부터 수신한 제2 소리 신호(S2)에 대해 푸리에 변환 연산을 수행하고, 제2 소리 신호(S2)에 대한 푸리에 변환 결과(P2)를 획득할 수 있다.
- [47] 위상차 획득 모듈(302)은, 제1 STFT 모듈(300)로부터 제공받은 푸리에 변환

- 결과(P1) 및 제2 STFT 모듈(301)로부터 제공받은 푸리에 변환 결과(P2)로부터, 제1 소리 신호(S1)와 제2 소리 신호(S2)의 위상차(dP)를 획득할 수 있다.
- [48] 학습 모듈(303)은, 제1 소리 신호(S1)와 제2 소리 신호(S2)의 위상차(dP)를 이용하여 빔포밍 모델(304)을 학습시킬 수 있다. 이에 따라, 빔포밍 모델(304)은, 제1 마이크(M1)와 제2 마이크(M2)를 통해 수신되는 2 가지 소리 신호의 위상차만으로 빔포밍을 수행할 수 있도록 학습될 수 있다.
- [49] 학습 시, 제1 마이크(M1)와 제2 마이크(M2) 사이의 미리 정해진 거리(D)는 10 cm 내지 14 cm, 바람직하게는 12 cm일 수 있으나, 본 발명의 범위가 이에 제한되는 것은 아니다. 미리 정해진 거리(D)가 12 cm일 때 학습된 빔포밍 모델(304)의 성능은, 추론 시 제1 마이크(M1)와 제2 마이크(M2) 사이의 거리가 10 cm 내지 14 cm인 경우에까지 성능을 발휘할 수 있다.
- [50] 도 4는 본 발명의 일 실시 예에 따른 빔포밍 방법을 설명하기 위한 도면이다.
- [51] 도 4를 참조하면, 본 발명의 일 실시 예에 따른 빔포밍 방법(31)에서, 제1 마이크(M1)는 제1 측으로부터 제1 소리 신호(S1)를 수신할 수 있다. 예를 들어, 제1 마이크(M1)는 제1 측으로부터 화상 회의에 참여하는 사람의 목소리와 주변 소음을 포함하는 제1 소리 신호(S1)를 수신하여, 이를 제1 STFT 모듈(310)(도면 상에서는 L_STFT 모듈로 표시함)에 전달할 수 있다.
- [52] 제1 STFT 모듈(310)은, 제1 마이크(M1)로부터 수신한 제1 소리 신호(S1)에 대해 푸리에 변환 연산을 수행하고, 제1 소리 신호(S1)에 대한 푸리에 변환 결과(P1)를 획득할 수 있다.
- [53] 한편, 제1 마이크(M1)로부터 미리 정해진 거리(D)만큼 이격되어 배치되는 제2 마이크(M2)는 제2 측으로부터 제2 소리 신호(S2)를 수신할 수 있다. 예를 들어, 제2 마이크(M2)는 제2 측으로부터 화상 회의에 참여하는 사람의 목소리와 주변 소음을 포함하는 제2 소리 신호(S2)를 수신하여, 이를 제2 STFT 모듈(311)(도면 상에서는 R_STFT 모듈로 표시함)에 전달할 수 있다.
- [54] 제2 STFT 모듈(311)은, 제2 마이크(M2)로부터 수신한 제2 소리 신호(S2)에 대해 푸리에 변환 연산을 수행하고, 제2 소리 신호(S2)에 대한 푸리에 변환 결과(P2)를 획득할 수 있다.
- [55] 위상차 획득 모듈(312)은, 제1 STFT 모듈(310)로부터 제공받은 푸리에 변환 결과(P1) 및 제2 STFT 모듈(311)로부터 제공받은 푸리에 변환 결과(P2)로부터, 제1 소리 신호(S1)와 제2 소리 신호(S2)의 위상차(dP)를 획득할 수 있다.
- [56] 도 3에서 설명한 바와 같이 학습이 완료된 빔포밍 모델(314)은 제1 소리 신호(S1)와 제2 소리 신호(S2)의 위상차(dP)를 입력으로 받아 뉴럴 네트워크 연산(즉, 추론 연산)을 수행할 수 있다.
- [57] 마스킹 모듈(315)은 추론 연산 결과에 대해 마스크를 적용할 수 있으며, 이어서 원소곱 모듈(316)은, 추론 연산 결과(또는 마스크가 적용된 결과)와 제2 STFT 모듈(311)로부터 제공받은 제2 소리 신호(S2)에 대한 푸리에 변환 결과(P2)에 대해 원소곱을 수행할 수 있다. 여기서 원소곱은 동일한 크기의 2 개 행렬의 각

성분을 곱하는 연산일 수 있다.

[58] 출력 모듈(317)은 원소곱 모듈(316)로부터 제공받은 원소곱 결과(S3)를 출력할 수 있다. 예를 들어, 출력 모듈(317)은, 빔포밍 모델(314)를 이용하여 마이크(M1, M2)를 이용하여 수신한 소리 신호에 대해 빔포밍을 수행한 결과로서 획득한 빔포밍된 소리 신호(S3)를 출력할 수 있다. 예를 들어, 빔포밍된 소리 신호(S3)는 주변 소음이 많은 환경에서 화상 회의에 참여하는 사람의 목소리를 분별한 것일 수 있고, 다른 컴퓨팅 장치(예컨대 모니터(20)가 연결된 개인용 컴퓨터)에 제공되고, 이어서 다른 화상 회의 참여자에게 제공될 수 있다.

[59] 도 5는 본 발명의 일 실시 예에 따른 빔포밍 방법에 대한 일 구현 예를 설명하기 위한 도면이다.

[60] 방향성 청력을 위해서는 기본적으로 2 개 이상의 마이크가 필요하다. 마이크 배열의 기하학적 형태는 각 마이크에 의해 수신되는 신호가 정렬되는 공간적 특성을 가지고 있다. 빔포밍을 위한 마스크를 획득하는 과정은 다음과 같이 공식화할 수 있다. 우선, 다수의 마이크로부터 수신된 각각의 신호를 STFT(short time Fourier transform)하여 스펙트로그램을 얻을 수 있다고 하자. $S_t^{(l)}(\omega)$ 는 얻고자 하는 음성이고 아래와 같이 표현된다.

$$[61] \quad S_t^{(l)}(\omega) = a^{(l)}(\omega)A_l(\omega)$$

$$[62] \quad a^{(l)}(\omega) = \exp((-1)^l i \frac{\omega d}{2c} \cos\theta)$$

[63] 또 $N_t^{(l)}(\omega)$ 는 잡음이라고 할 때, 음성과 잡음이 동시에 들어오는 마이크 배열의 입력은 $X_t^{(l)}(\omega) = S_t^{(l)}(\omega) + N_t^{(l)}(\omega)$ 로 나타낼 수 있다.

[64] $a^{(l)}(\omega)$ 는 각도에 따른 마이크 배열의 특성을 나타낸다. 이 때 입력된 스펙트로그램에 필터 $W_l(\omega)$ 를 취해준다고 하면, 그 결과는 $W_l(\omega)X_t^{(l)}(\omega) = W_l(\omega)a^{(l)}(\omega)A_l(\omega) + W_l(\omega)N_t^{(l)}(\omega)$ 가 된다. 따라서 입력 신호의 크기를 최소화하되 얻고자 하는 신호는 남기는 방향으로 $W_l(\omega)$ 를 디자인할 수 있다. 굵은 문자로 표시한 것이 벡터라 하면 $Y_l(\omega) = W^H(\omega) \cdot X_t(\omega)$ 일 때

$$[65] \quad \text{minimizing } E[|Y_l(\omega)|^2] \text{ constraint of } W^H(\omega) \cdot a(\omega) = 1$$

[66] 가 되고 이때 해는

$$[67] \quad W(\omega) = \frac{R_{XX}^{-1}a(\omega)}{a^H(\omega)R_{XX}^{-1}a(\omega)}$$

[68] 로 구해질 수 있다. 여기서, 윗첨자 H 는 켈레 복소수를 취해준 뒤 전치한 행렬을 말한다. 따라서 구해진 필터를 이용해서 얻고자 하는 음성의

스펙트로그램은 $S_r(\omega) = W^H(\omega) X_r(\omega)$ 와 같이 얻을 수 있다. 이러한 방식을 사용하여 빔포밍을 구현하는 경우에 가장 중요한 부분은 바로 조향벡터 $a(\omega)$ 와 공간상관행렬 R_{xx} 를 정확하게 구하는 것이다.

- [69] 이와 같이, 뉴럴 네트워크를 사용하지 않는 MVDR 방식에서, MVDR을 정확하게 만족하기 위해서는 음원에서 발생한 소리가 각각의 마이크에 도달하기까지의 경로를 수학적으로 모델링하는 조향벡터가 필요한데, 화상대화와 같이 마이크와 사용자의 입이 가까운 거리, 예를 들어 1m 내외에 위치하고, 마이크의 간격 또한 근접한 거리, 예를 들어 수cm 또는 십 수cm 정도인 경우에는 수학적 모델링이 매우 어렵다. 또한, 실제 화상대화 환경에서와 같이 사용자가 불규칙하게 움직이는 경우, 조향벡터를 고정 값을 둘 경우 사용자의 불편도는 더욱 올라가게 되는 한계점들이 있다. 이러한 한계점들을 극복하기 위해, 조향벡터가 아닌 뉴럴 네트워크를 이용할 수 있다.
- [70] 또 본 방식에서 뉴럴 네트워크의 입력으로 사용하는 위상차 행렬은 기존의 방식에서 음원의 위치 정보를 얻기 위해 사용하는 공간상관행렬보다 구하는 과정이 간단하여 더욱 쉽게 음원의 위치정보를 얻을 수 있다. 본 발명의 일 실시 예에 따른 빔포밍 방법에서는, 단지 2 개의 마이크에서 공간 정보를 위한 뉴럴 네트워크를 훈련시키고 이를 통해 미리 정해진 방향으로 소리를 획득할 수 있다. 따라서 아래와 같은 이점을 가진다.
- [71] 1. 실제 상황에서 고정된 마이크로 측정할 때 움직이는 대상에 대한 조향벡터를 정확하게 정의할 필요가 없다.
- [72] 2. 선형 필터를 구하는 과정에서 발생하는 특이점(Singular point)의 발산이 사라진다. 이는 정면 이외에서 $Y_r(\omega)$ 의 파워인 사이드 로브(side lobe)를 줄여주어 자동으로 사이드 로브 제거(side lobe cancelation) 효과를 가져온다.
- [73] 3. 정면 이외의 각도에서 복원되는 음성의 STOI값으로 볼 때 여러 방향에서 음성 신호가 들어올 경우 MVDR의 측면 STOI가 뉴럴 네트워크의 STOI보다 더 크게 나타나며, 즉, 뉴럴 네트워크를 이용할 때 정면의 음성신호를 더 정확히 구분해 낸다.
- [74] 음원과 마이크 사이에 장애물이 없을 때는 음원의 각도에 의해서만 음의 이동 경로가 영향을 받는다. 양쪽 마이크와 음원의 거리가 같기 때문에 앞에서 들려오는 소리는 서로 같다. 즉, 전방 소스의 도착 시간 차이(TDOA)는 0에 근접한다. 이 점을 이용하면 앞에서 들려오는 소리를 그대로 남길 수 있다.
- [75] 임의의 위치에 있는 음원에서 발생하여 두 마이크에 도달한 소리를 $X_r^{(1)}(\omega)$, $X_r^{(2)}(\omega)$ 라 하면 각각의 신호로부터 2 개의 위상행렬 $\Phi_r^{(1)}(\omega)$, $\Phi_r^{(2)}(\omega)$ 를 구할 수 있다. 이 위상행렬간의 차이가 0에 가까울수록 정면에서 가까운 음원이 만들어낸 소리를 의미한다. 이것을 구분하기 위하여

입력이 0에서 멀어질 수록 0에 가까운 출력을 내고 입력이 0에 가까울 수록 1을 출력으로 내는 함수를 생각해 본다면 정면에서 오는 신호만을 남기는 마스크를 만드는 것이 가능하다. 따라서 아래의 오차 L 을 최소화하는 마스크 $M_A(\omega)$ 을 찾으면 정면에서 오는 신호를 복원할 수 있다.

$$[76] \quad L = \frac{1}{T} \sum_{t=0}^{T-1} (s(t) - \text{ISTFT}(M_A(\omega) \odot X_t^{(1)}(\omega)))^2$$

[77] 푸리에 변환을 사용할 때 신호는 각 주파수의 단순한 퓨어 톤(pure tone)의 합이라고 생각할 수 있지만 동일한 주파수의 두 퓨어 톤은 위상 지연을 통해 합쳐질 경우 전혀 새로운 퓨어 톤 신호를 만든다. 이 문제는 양쪽 마이크의 위상 차이가 음원의 각도에 정확히 비례하지 않는 상황을 만들 수 있으므로 단순히 위상차에만 의존하는 규칙에 기반하여 마스크를 만드는 것은 부정확하다. 음성을 얻는 일은 광범위한 주파수를 처리하는 필터가 필요하므로 이 문제를 해결해야 한다. 이를 극복하기 위한 방안으로 본 발명의 실시 예들은 뉴럴 네트워크와 기준 마이크의 스펙트로그램 패턴을 통해 위상 차이를 인식하는 방법을 제공한다.

[78] 도 6을 참조하면, 설계하고자 하는 마스크는 실수로 정의된 0과 1 사이의 마스크이기 때문에 IBM(Ideal Binary Mask)와 유사한 기능을 수행할 수 있다. 다만 그 사이값이 있기 때문에 SBM(Soft Binary Mask)가 될 수 있다. 기존 IBM를 이용한 뉴럴 네트워크에 대한 연구를 보면, 시끄러운 위상을 그대로 사용했다. 이로부터, 노이즈가 있는 스펙트로그램으로 스피치를 재구성할 때, 스피치의 구성과 관련된 주파수 영역을 남겨두는 것만으로 신호의 스피치 재구성이 가능하다. 그러므로 SBM 방식의 뉴럴 네트워크로 구한 마스크는 위상차로부터 매그니튜드에 적용되는 방향성 청각 마스크를 만들어 내는데, 이는 위상차가 나는 원소더라도 신호를 남기는 기능을 수행한다. 한편, TDOA로부터 방향 정보를 보다 명확하게 얻기 위해 가정된 시간 지연을 최적화하는 것을 제시하는 방법이 있으나, 본 발명의 다양한 실시 예에서는 CNN을 사용한다. CNN은 이미지와 같은 2D 행렬에 더 효율적이고, 계산 능력이 덜 필요하기 때문이다. 또한, 콘볼루션 필터는 이상적인 마스크의 위상차에 따른 감소폭에 최적화되어 있다. 한편, 뉴럴 네트워크에 위상차만이 입력으로 들어가는 경우 목표하는 음성의 패턴을 학습할 수 없다는 문제점을 방지하기 위해 역전파를 이용할 수 있다. 시간 영역에서 구해지는 MSE(mean squared error)는 위상 정보가 있는 마스크를 훈련시키는 손실 함수(loss function)로 사용되는데 앞서 말한 방식은 이 방법과 비슷한 방식이다. 역 STFT(ISTFT)로 마스크된 스펙트로그램을 복원함으로써, 손실 함수를 이용하여 시간 영역의 깨끗한 표적 소리와 비교할 수 있다. 이 경우 업데이트할 그레디언트값은 기준 마이크에 대한 음성 패턴 정보를 포함할 수 있다.

[79] 다양한 상황의 데이터로 모델을 훈련시키기 위해, 시뮬레이션 방법을 사용하여

소정의 데이터 세트를 생성하고, 뉴럴 네트워크를 이용하여 대상의 음성을 재구성하였으며, 스테레오 채널 음원은 10x10x10m 공간 시뮬레이션을 통해 생성되었다. 마이크의 높이는 2m로 좌우 6cm 간격으로 2 개 위치(9, 5.06, 2)와 (9, 4.94, 2)로 지정된다. 음원은 지름 1미터의 반원 상에 위치하며 이때 반원의 중심은 마이크의 중심과 동일하다. 얻고자 하는 음원의 위치는 정면인 90° , (7, 5, 2)에 위치한다. 그리고 반원 상의 4개의 구간

$[0^\circ, 40^\circ)$, $[40^\circ, 80^\circ)$, $[100^\circ, 140^\circ)$, $[140^\circ, 180^\circ)$ 을 나누어 각 구간별 무작위 위치에 잡음의 음원을 배치한다. 음성 데이터 세트 중 일부는 훈련 데이터에 사용되었으며, 다른 일부는 시험 데이터로 사용되었다. 소음 데이터 세트는 두 개의 인공 소음(speech shape noise 및 babble noise)과 DEMAND database를 사용했으며, 이는 13개의 녹음된 소음이다. 훈련 데이터는 2개의 인공 소음과 8개의 녹음된 소음(카페, 자동차, 키친, 미팅, 메트로, 레스토랑, 역, 교통)으로 구성된다. 시험 데이터는 5개의 녹음된 소음(버스, 카페테리아, 거실, 사무실 및 공용 광장)으로 구성된다. 소음을 내기 위해 동일한 소음 신호의 무작위 지점 4개에서 음성의 길이만큼 소음이 추출되었다. 그 후 네 군데의 음원으로 부터 소리가 나오는 것으로 시뮬레이션 하여, 마이크 2 개로 수용하는 방식을 채택했다.

- [80] 훈련 데이터의 경우 40가지 조건 즉 소음 상황 10개와 SNR 4개(0dB, 5dB, 10dB, 15dB)의 조합을 사용했다. 시험 데이터는 20개의 다른 조건 즉 5개의 소음 상황과 4개의 SNR(2.5dB, 7.5dB, 12.5dB, 17.5dB)의 조합을 사용했다. 이때는 image-source method의 Room impulse response를 이용하여 거리에 따른 에너지 감소만을 시뮬레이션 하였다. 잔향은 고려하지 않기 위하여 잔향시간(T_{60})은 0으로 설정했다.
- [81] 2 개의 마이크가 수신한 각 신호는 STFT를 통해 푸리에 영역으로 변환되었다. STFT는 16kHz 샘플 레이트(sample rate)의 신호에 256 포인트 해밍 윈도우(hamming window)를 사용한다. 윈도우 시프트(window shift)는 128 포인트(128 포인트의 겹치짐)을 사용한다. 뉴럴 네트워크 조작이 완료된 후 ISTFT 작업을 수행할 때도 동일한 조건이 사용된다. 모델 구조는 표 1의 구조를 따른다.
- [82] 입력 값은 STFT 결과 중 낮은 주파수 128 개만 포함한 것이다. 뉴럴 네트워크의 결과로 얻어진 마스크와 입력된 128 개 주파수의 스펙트로그램을 곱하고, 129번째 주파수를 0으로 채워서 ISTFT하여 복원된 신호를 얻는다.

[83] [표1]

Name	Input	Layer	Output
Conv1	B,128,T,1	(5,3),(1,1),(2,0)	B,128,T,5
Conv2	B,128,T,5	(4,1),(2,1),(1,0)	B,64,T,5
Conv3	B,64,T,5	(5,3),(1,1),(2,0)	B,64,T,10
Conv4	B,64,T,10	(4,1),(2,1),(1,0)	B,32,T,10
Conv5	B,32,T,10	(5,3),(1,1),(2,0)	B,32,T,18
Conv6	B,32,T,18	(4,1),(2,1),(1,0)	B,16,T,18
Conv7	B,16,T,18	(5,3),(1,1),(2,0)	B,16,T,32
Conv8	B,16,T,32	(4,1),(2,1),(1,0)	B,8,T,32
Conv9	B,8,T,32	(5,3),(1,1),(2,0)	B,8,T,32
deConv1	B,8,T,32	(4,1),(2,1),(1,0)	B,16,T,64
Conv10	B,16,T,64	(5,3),(1,1),(2,0)	B,16,T,18
deConv2	B,16,T,18	(4,1),(2,1),(1,0)	B,32,T,36
Conv11	B,32,T,36	(5,3),(1,1),(2,0)	B,32,T,10
deConv3	B,32,T,10	(4,1),(2,1),(1,0)	B,64,T,20
Conv12	B,64,T,20	(5,3),(1,1),(2,0)	B,64,T,5
deConv4	B,64,T,5	(4,1),(2,1),(1,0)	B,128,T,10
Conv13	B,128,T,10	(5,3),(1,1),(2,0)	B,128,T,5
Conv14	B,128,T,5	(1,1),(1,1),(0,0)	B,128,T,2

- [84] 여기서, 입력은 (batch, frequency, time step, channel)로 이루어진다. 그리고 컨벌루션 레이어는 (filter height, filter width),(stride height, stride width),(padding height, padding width) 로 구성된다. 출력은 (batch, frequency, time step, channel)로 구성된다. 활성화함수는 모두 PReLU가 사용되었다. 가장 마지막 레이어의 활성화함수는 시그모이드 함수를 사용하여 1채널과 2채널을 각각 마스크의 실수부, 허수부로 사용했다.
- [85] 본 발명의 실시 예들에 따른 유리한 효과는 도 9 및 도 10과 관련하여 후술하도록 한다.
- [86] 도 6은 본 발명의 일 실시 예에 따른 빔포밍 시스템을 설명하기 위한 도면이고, 도 7은 본 발명의 일 실시 예에 따른 빔포밍 시스템을 설명하기 위한 도면이고, 도 8은 본 발명의 일 실시 예에 따른 빔포밍 시스템을 설명하기 위한 도면이다.
- [87] 도 6을 참조하면, 본 발명의 일 실시 예에 따른 빔포밍 시스템(2)은, 제1 마이크(M1) 및 제2 마이크(M2)를 구비하는 모니터로 구현될 수 있다.

- [88] 도 7을 참조하면, 본 발명의 일 실시 예에 따른 빔포밍 시스템(2)은, 스마트폰을 비롯한 포터블 컴퓨팅 장치(22)를 거치할 수 있는 거치대형 장치(14)를 포함하도록 구현될 수 있다. 그리고 마이크(M1, M2)를 이용하여 수신한 소리 신호에 대해 빔포밍을 수행한 후, 그 결과로서 획득한 빔포밍된 소리 신호는 임의의 접속 수단을 통해 포터블 컴퓨팅 장치(22)에 제공될 수 있다. 그러면 컴퓨터 장치(22)는 분별된 사람의 목소리를, 예컨대 다른 화상 회의 참여자에게 제공할 수 있다.
- [89] 도 8을 참조하면, 본 발명의 일 실시 예에 따른 빔포밍 시스템(4)은, 스마트폰을 비롯한 포터블 컴퓨팅 장치(22)에 부착할 수 있는 부착형 장치(16)로 구현될 수 있다. 그리고 마이크(M1, M2)를 이용하여 수신한 소리 신호에 대해 빔포밍을 수행한 후, 그 결과로서 획득한 빔포밍된 소리 신호는 임의의 접속 수단을 통해 포터블 컴퓨팅 장치(22)에 제공될 수 있다. 그러면 컴퓨터 장치(22)는 분별된 사람의 목소리를, 예컨대 다른 화상 회의 참여자에게 제공할 수 있다.
- [90] 도 9 및 도 10은 본 발명의 실시 예들에 따른 빔포밍 방법 및 빔포밍 시스템의 유리한 효과를 설명하기 위한 도면이다.
- [91] 제1 마이크(M1)와 제2 마이크(M2) 사이의 미리 정해진 거리(D)가 12 cm인 경우의 빔포밍 알고리즘에 8, 10, 12, 14, 16 cm의 마이크 거리로 시뮬레이션한 데이터를 넣어서 측정한 결과로, 도 9는 고전적인 방식(MVDR)을 이용한 빔포밍의 각도별 소리 크기(윗줄)와 STOI(short-time objective intelligibility) 점수(아래줄)를 나타낸 것이고, 도 10은 본 발명의 다양한 실시 예들에 따른 뉴럴 네트워크를 이용한 빔포밍의 각도별 소리 크기(윗줄)와 STOI점수(아래줄)를 나타낸 것이다.
- [92] STOI는 소리의 크기와 무관하게 소리의 복원된 정도와 관련 있는 지표로써 고전적인 방식의 경우 듣고자 하는 방향이 아닌 방향에서 STOI가 높게 나타나는 반면, 본 발명의 다양한 실시 예들에 따른 뉴럴 네트워크 방식의 측면에서는 낮게 나타나며, 이를 통해, 본 발명의 다양한 실시 예들에 따른 뉴럴 네트워크 방식이, 원하는 방향이 아닌 곳의 발화자가 말한 음성을 더 확실하게 분리하며 더 적게 방해받을 수 있다는 것을 알 수 있다.
- [93] 이제까지 설명한 본 발명의 실시 예들에 따르면, 다양한 소음 환경에 대한 조향 벡터를 연산할 필요없이 위상차만을 이용하여 마이크에서 수신되는 음성을 복원할 수 있어, 빔포밍을 효율적으로 구현할 수 있다.
- [94] 기존 MVDR의 경우 잔향이 심한 환경에서는 조향벡터를 얻기 힘들고 잔향이 없는 환경이라 하더라도 화상대화의 경우 원거리의 음원을 가정한 꼴을 구할 수 없어 간단한 꼴의 조향벡터를 구하는 것이 힘들어진다. 또 실제 상황에서는 화자가 움직이는 경우가 많은데 조향벡터를 정확하게 예측할수록 이러한 환경에서 사용하는 것이 어려워진다. 또한 위에서 R_{xx} 로 표현한 공간상관행렬을 구할 때 잡음이 섞인 소리를 이용하므로 정확한 예측이

불가능하다. 하지만, 본 방식은 뉴럴 네트워크를 사용하여 많은 양의 데이터로부터 공간정보를 학습하므로 조향벡터를 계산하지 않고 깨끗한 음성신호를 표적음성으로 하여 학습하는 방법을 사용하기 때문에, 정확한 학습이 가능하다. 본 방법을 통해서 훈련한 인공지능 모델과 기존의 방식인 MVDR방식으로 824개의 시험데이터를 계산하고 구해진 정면 소리의 평균 SSNR(Segmented Signal to Noise Ratio)지표를 보면 잡음이 포함된 소리는 -1.557, MVDR을 이용한 소리는 -0.120, 뉴럴 네트워크를 통과한 모델은 2.165으로 뉴럴 네트워크 모델이 가장 높은 것을 알 수 있다.

- [95] 이상에서 본 발명의 실시 예에 대하여 상세하게 설명하였지만 본 발명의 권리 범위는 이에 한정되는 것은 아니고, 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자의 여러 변형 및 개량 형태 또한 본 발명의 권리 범위에 속한다.

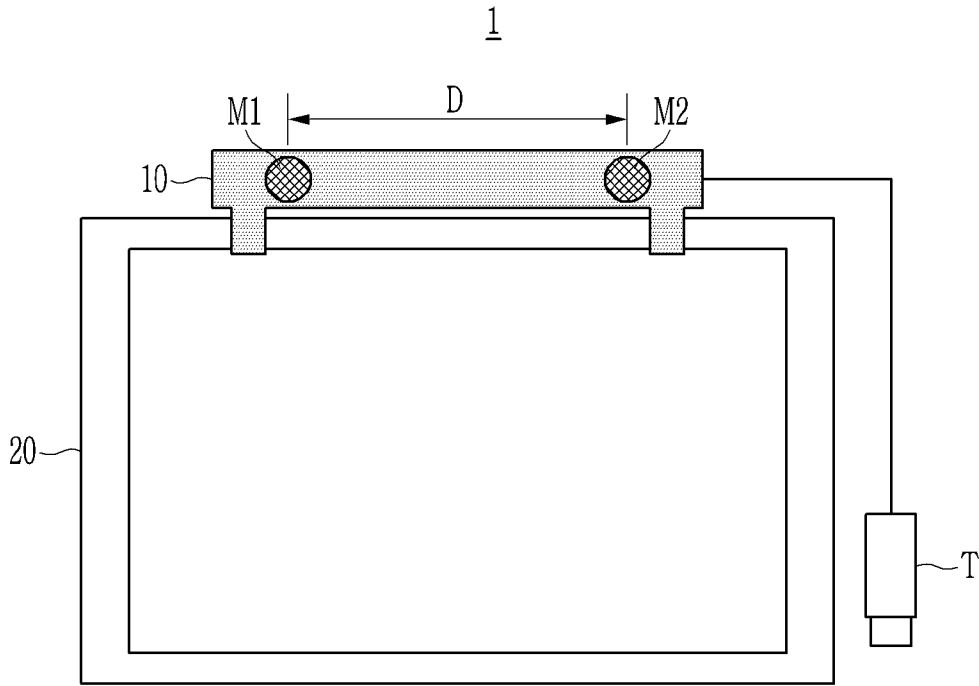
청구범위

- [청구항 1] 제1 마이크와, 상기 제1 마이크로부터 미리 정해진 거리만큼 이격되어 배치된 제2 마이크를 이용하여 제1 소리 신호 및 제2 소리 신호를 각각 수신하는 단계;
상기 제1 소리 신호 및 상기 제2 소리 신호 각각에 대한 푸리에 변환 결과를 획득하는 단계;
상기 푸리에 변환 결과로부터 상기 제1 소리 신호와 상기 제2 소리 신호의 위상차를 획득하는 단계;
뉴럴 프로세서를 이용하여 상기 위상차를 빔포밍 모델에 입력하여 연산하는 단계;
상기 뉴럴 프로세서의 연산 결과와 상기 제1 소리 신호에 대한 푸리에 변환 결과에 대해 원소곱을 수행하는 단계; 및
상기 원소곱 결과를 출력하는 단계를 포함하는 빔포밍 방법.
- [청구항 2] 제1항에 있어서,
상기 원소곱을 수행하는 단계는,
상기 원소곱을 수행하기 전에 상기 연산 결과에 대해 마스크(mask)를 적용하는 단계를 더 포함하는 빔포밍 방법.
- [청구항 3] 제1항에 있어서,
상기 원소곱을 수행하는 단계는,
상기 원소곱을 수행한 후에 이득 제어(gain control)를 수행하는 단계를 더 포함하는 빔포밍 방법.
- [청구항 4] 제1항에 있어서,
상기 미리 정해진 거리는 10 cm 내지 14 cm인, 빔포밍 방법.
- [청구항 5] 제1항에 있어서,
상기 위상차를 이용하여 상기 빔포밍 모델을 학습시키는 단계를 더 포함하는 빔포밍 방법.
- [청구항 6] 제1 소리 신호를 수신하는 제1 마이크;
상기 제1 마이크로부터 미리 정해진 거리만큼 이격되어 배치되어 제2 소리 신호를 수신하는 제2 마이크;
상기 제1 소리 신호에 대한 푸리에 변환 결과를 획득하는 제1 STFT 모듈;
상기 제2 소리 신호에 대한 푸리에 변환 결과를 획득하는 제2 STFT 모듈;
상기 푸리에 변환 결과로부터 상기 제1 소리 신호와 상기 제2 소리 신호의 위상차를 획득하는 위상차 획득 모듈;
상기 위상차를 입력받아 빔포밍 모델을 이용하여 뉴럴 네트워크 연산을 수행하는 뉴럴 프로세서;
상기 뉴럴 프로세서의 연산 결과와 상기 제1 소리 신호에 대한 푸리에

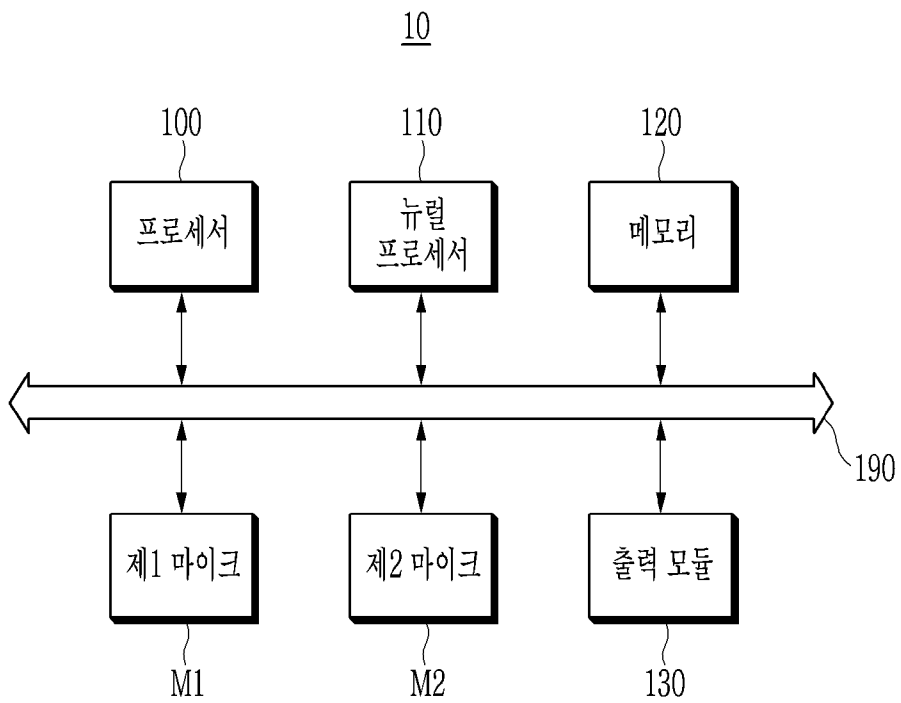
변환 결과에 대해 원소곱을 수행하는 원소곱 모듈; 및
상기 원소곱 결과를 출력하는 출력 모듈을 포함하는
빔포밍 시스템.

- [청구항 7] 제6항에 있어서,
상기 원소곱을 수행하기 전에 상기 연산 결과에 대해 마스크를 적용하는
마스킹 모듈을 더 포함하는 빔포밍 시스템.
- [청구항 8] 제6항에 있어서,
상기 원소곱을 수행한 후에 이득 제어를 수행하는 이득 제어 모듈을 더
포함하는 빔포밍 시스템.
- [청구항 9] 제6항에 있어서,
상기 미리 정해진 거리는 10 cm 내지 14 cm인, 빔포밍 시스템.
- [청구항 10] 제6항에 있어서,
상기 위상차를 이용하여 상기 빔포밍 모델을 학습시키는 학습 모듈을 더
포함하는 빔포밍 시스템.

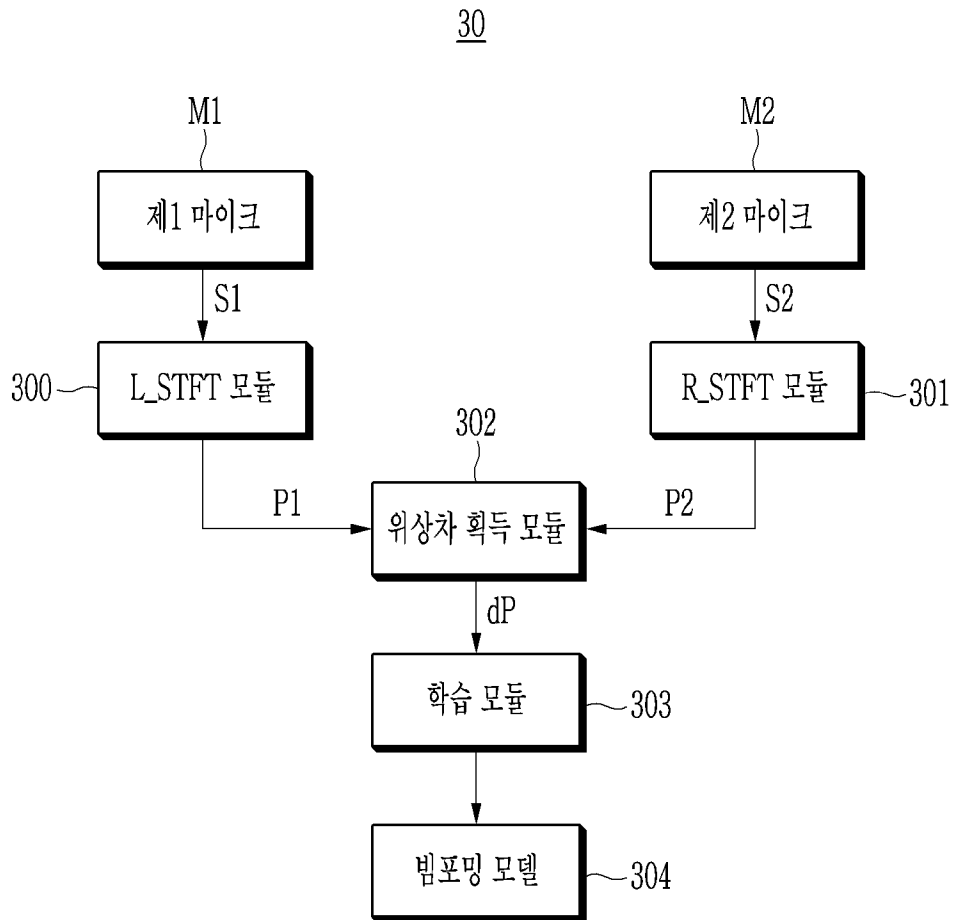
[도1]



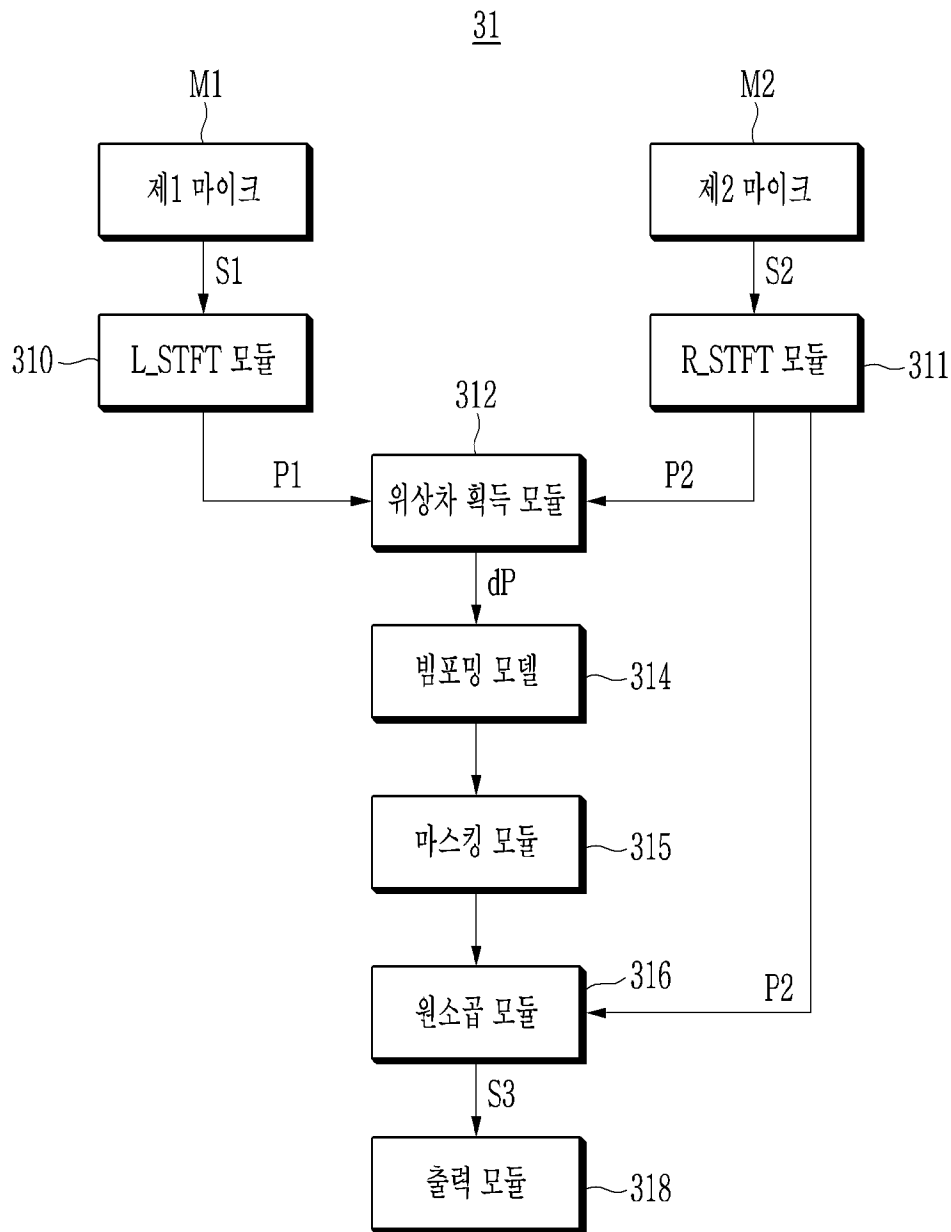
[도2]



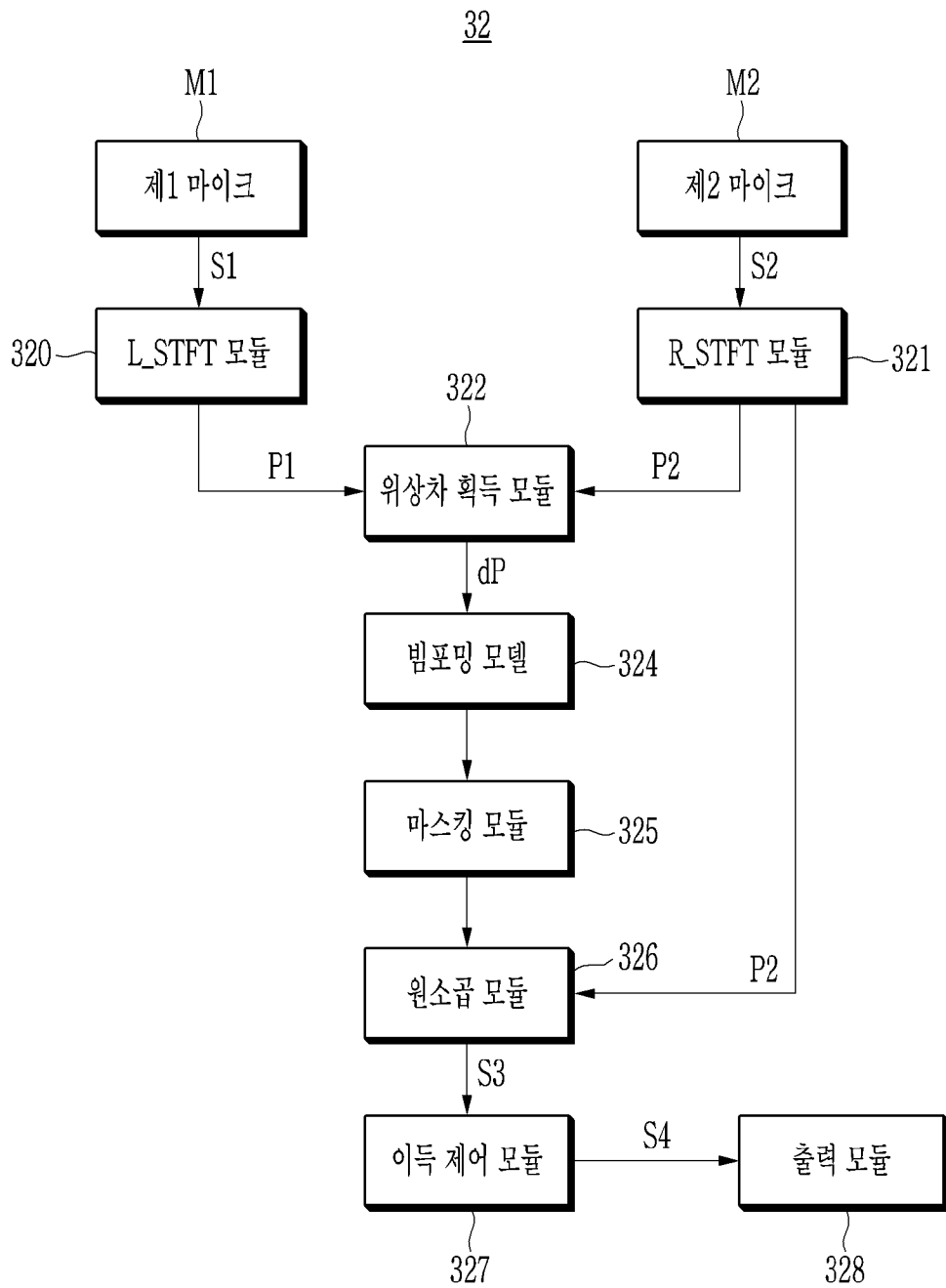
[도3]



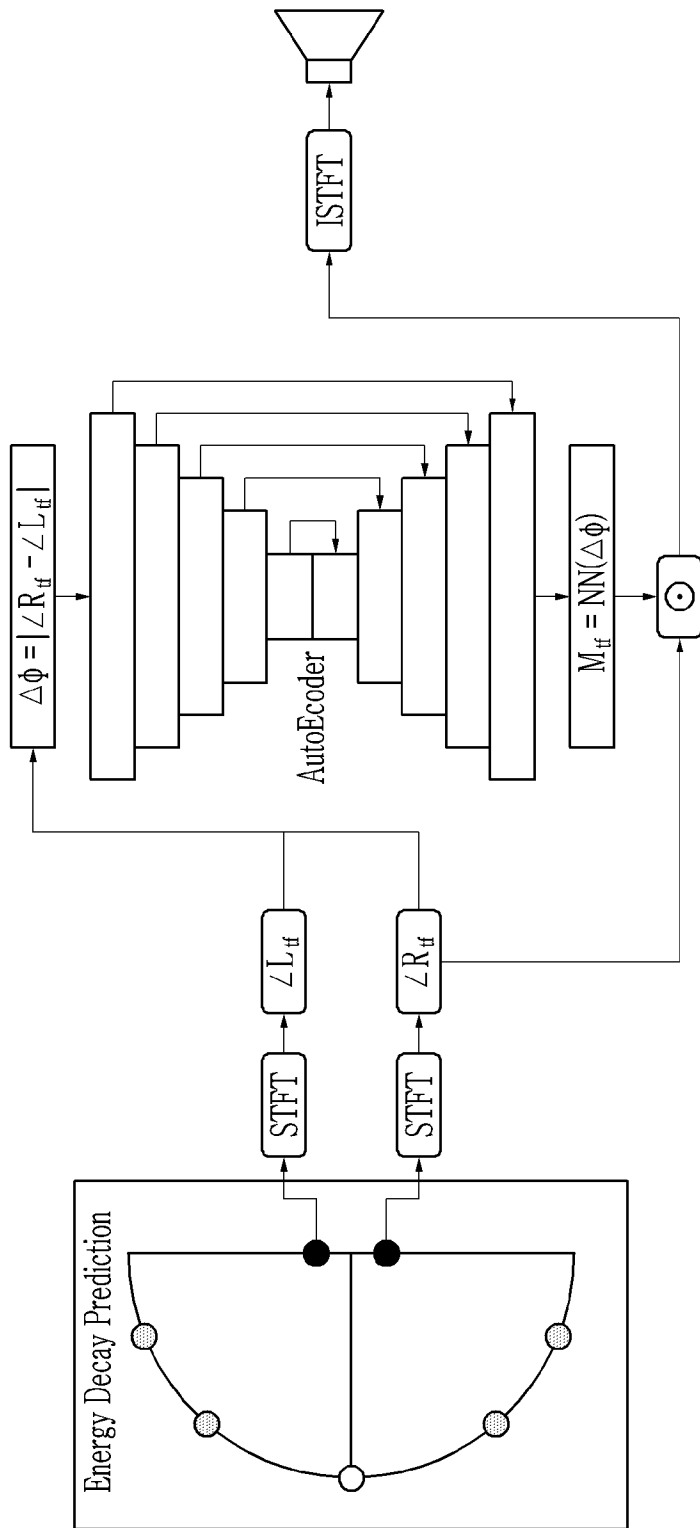
[도4]



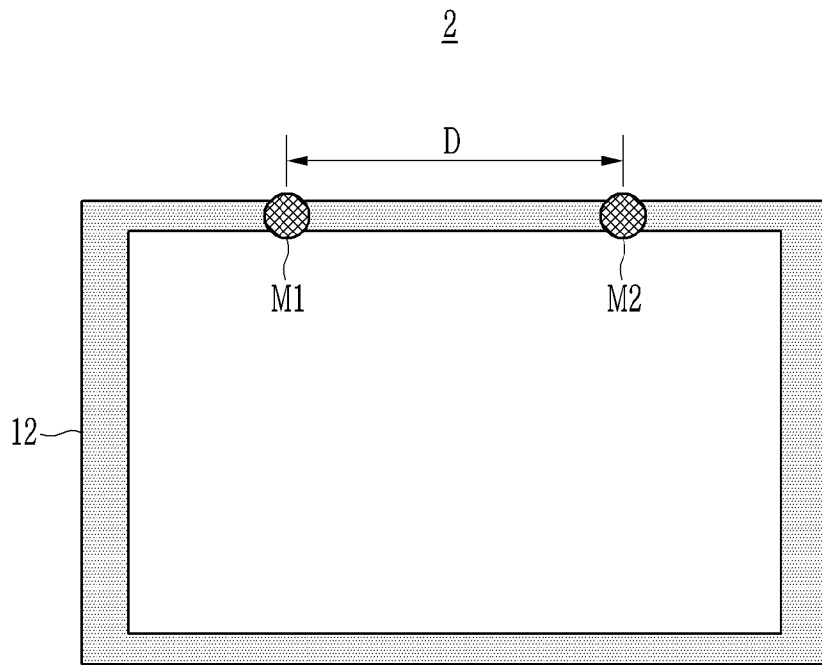
[도5]



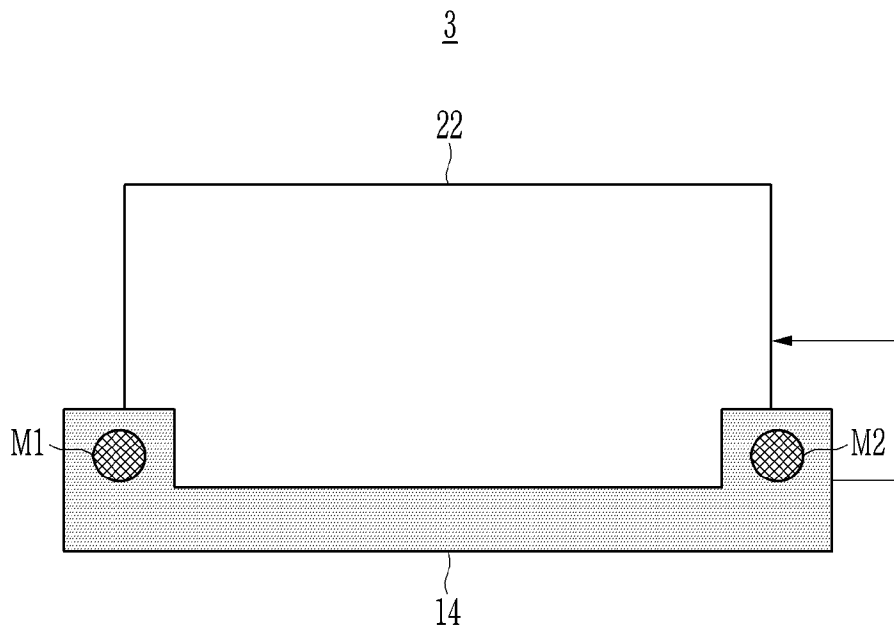
[도6]



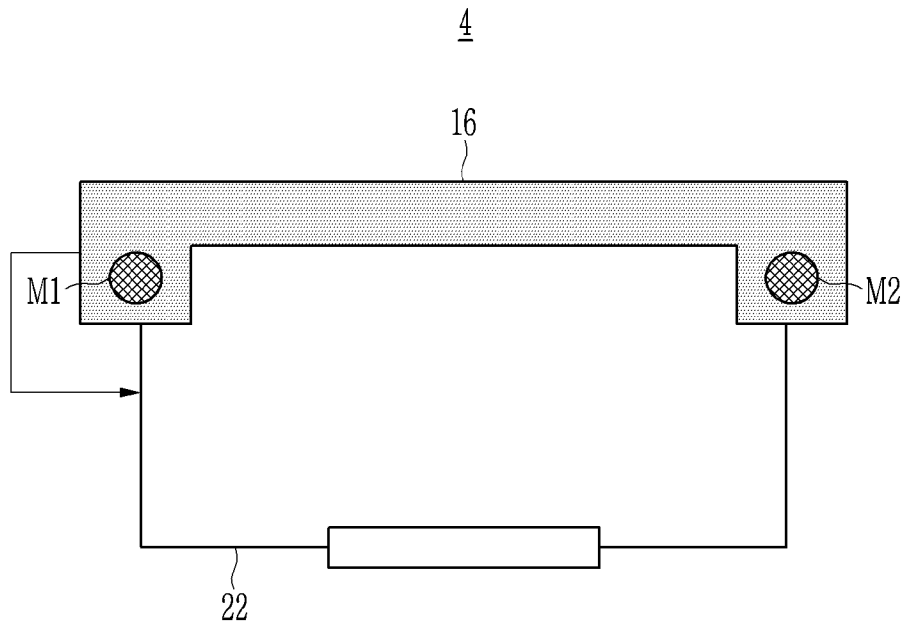
[도7]



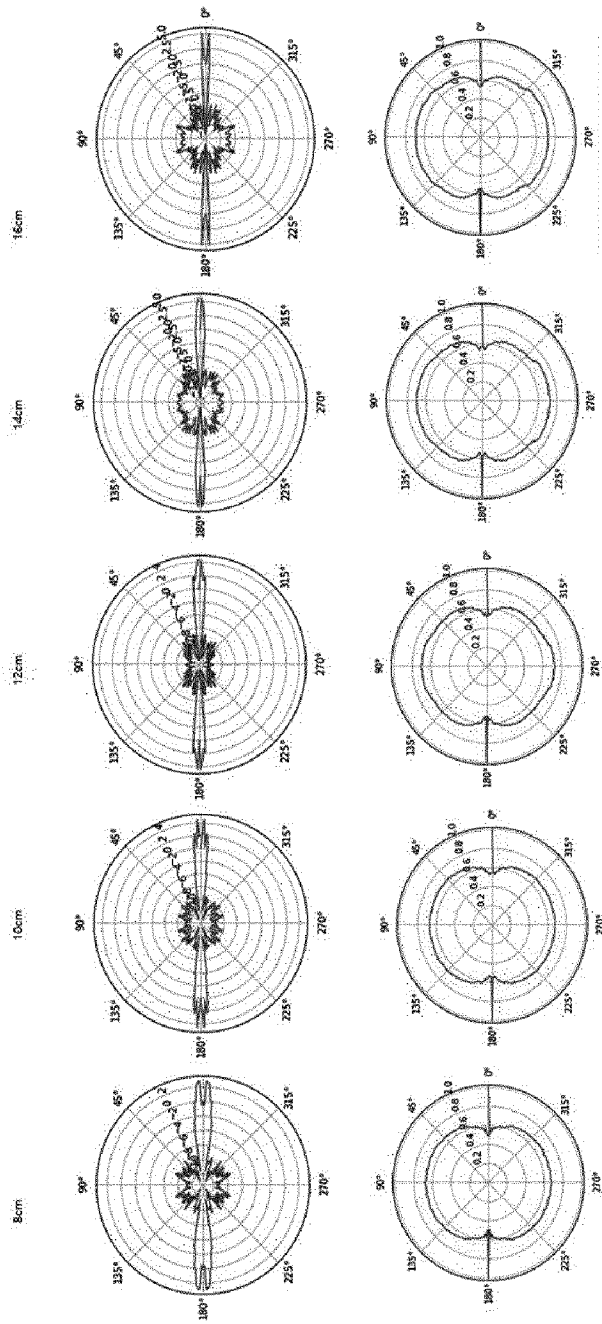
[도8]



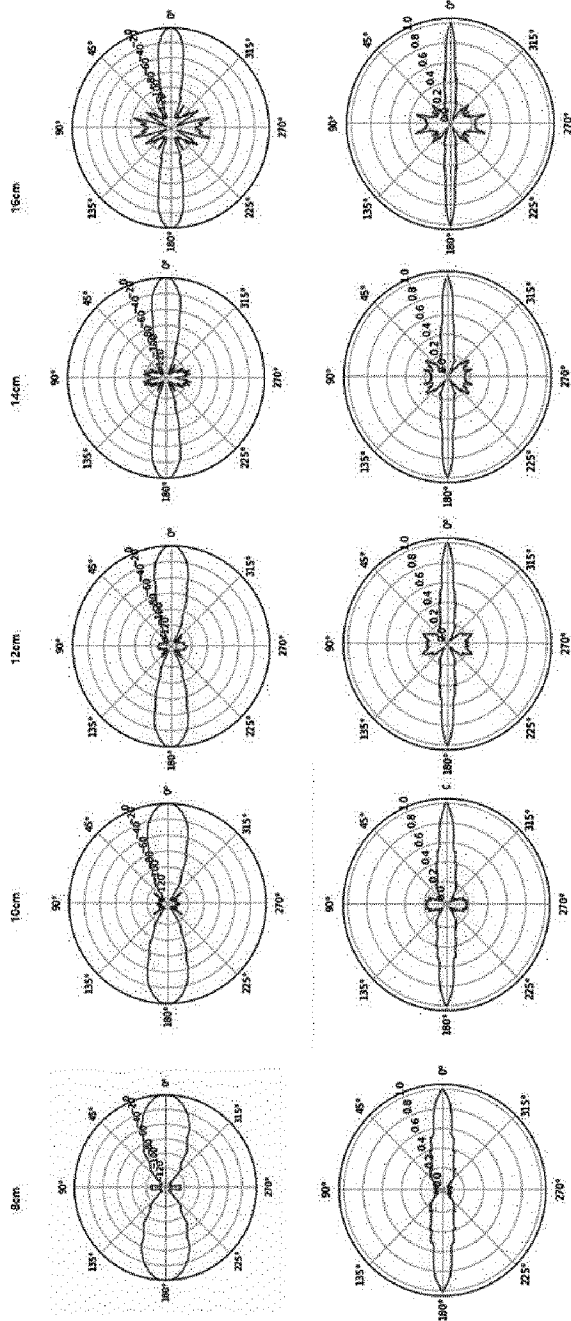
[도9]



[도 10]



[도 11]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2021/013328

A. CLASSIFICATION OF SUBJECT MATTER		
H04B 7/06(2006.01)i; G06N 3/08(2006.01)i; G06N 3/04(2006.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) H04B 7/06(2006.01); G10H 1/00(2006.01); G10L 15/16(2006.01); G10L 21/0208(2013.01); G10L 21/0216(2013.01); G10L 21/0264(2013.01); G10L 25/18(2013.01); G10L 25/30(2013.01)		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Korean utility models and applications for utility models: IPC as above Japanese utility models and applications for utility models: IPC as above		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) eKOMPASS (KIPO internal) & keywords: 마이크(mike), 소리 신호(sound signal), 푸리에 변환(fourier transform), 위상 (phase), 뉴럴 네트워크(neural network), 빔포밍(beamforming)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2020-0342891 A1 (BEIJING DIDI INFINITY TECHNOLOGY AND DEVELOPMENT CO., LTD.) 29 October 2020 (2020-10-29) See paragraphs [0017], [0020] and [0036]-[0038]; and claim 18.	1-10
Y	US 2020-0111483 A1 (GOOGLE LLC) 09 April 2020 (2020-04-09) See paragraphs [0097], [0113] and [0138].	1-10
A	WO 2019-199554 A1 (MICROSOFT TECHNOLOGY LICENSING, LLC) 17 October 2019 (2019-10-17) See paragraphs [0023]-[0030]; and figure 3.	1-10
A	KR 10-2018-0111271 A (SAMSUNG ELECTRONICS CO., LTD.) 11 October 2018 (2018-10-11) See paragraphs [0049]-[0054]; and figure 3.	1-10
A	KR 10-2018-0115984 A (IUCF-HYU (INDUSTRY-UNIVERSITY COOPERATION FOUNDATION HANYANG UNIVERSITY)) 24 October 2018 (2018-10-24) See paragraphs [0096]-[0098]; and figures 6-7.	1-10
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 29 December 2021		Date of mailing of the international search report 03 January 2022
Name and mailing address of the ISA/KR Korean Intellectual Property Office Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208 Facsimile No. +82-42-481-8578		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2021/013328

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2020-0342891	A1	29 October 2020	CN	111863015	A	30 October 2020
US	2020-0111483	A1	09 April 2020	US	10140980	B2	27 November 2018
				US	10529320	B2	07 January 2020
				US	10714078	B2	14 July 2020
				US	11069344	B2	20 July 2021
				US	2018-0174575	A1	21 June 2018
				US	2019-0115013	A1	18 April 2019
				US	2019-0156819	A1	23 May 2019
				US	2020-0286468	A1	10 September 2020
WO	2019-199554	A1	17 October 2019	EP	3776535	A1	17 February 2021
				US	10957337	B2	23 March 2021
				US	2019-0318757	A1	17 October 2019
KR	10-2018-0111271	A	11 October 2018	US	10593347	B2	17 March 2020
				US	2018-0286425	A1	04 October 2018
KR	10-2018-0115984	A	24 October 2018	KR	10-1934636	B1	02 January 2019
				US	11017791	B2	25 May 2021
				US	2020-0105287	A1	02 April 2020
				WO	2018-190547	A1	18 October 2018

A. 발명이 속하는 기술분류(국제특허분류(IPC)) H04B 7/06(2006.01)i; G06N 3/08(2006.01)i; G06N 3/04(2006.01)i		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) H04B 7/06(2006.01); G10H 1/00(2006.01); G10L 15/16(2006.01); G10L 21/0208(2013.01); G10L 21/0216(2013.01); G10L 21/0264(2013.01); G10L 25/18(2013.01); G10L 25/30(2013.01) 조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 마이크(mike), 소리 신호(sound signal), 푸리에 변환(fourier transform), 위상(phase), 뉴럴 네트워크(neural network), 빔포밍(beamforming)		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	US 2020-0342891 A1 (BEIJING DIDI INFINITY TECHNOLOGY AND DEVELOPMENT CO., LTD.) 2020.10.29 단락 [0017], [0020], [0036]-[0038]; 및 청구항 18	1-10
Y	US 2020-0111483 A1 (GOOGLE LLC) 2020.04.09 단락 [0097], [0113], [0138]	1-10
A	WO 2019-199554 A1 (MICROSOFT TECHNOLOGY LICENSING, LLC) 2019.10.17 단락 [0023]-[0030]; 및 도면 3	1-10
A	KR 10-2018-0111271 A (삼성전자주식회사) 2018.10.11 단락 [0049]-[0054]; 및 도면 3	1-10
A	KR 10-2018-0115984 A (한양대학교 산학협력단) 2018.10.24 단락 [0096]-[0098]; 및 도면 6-7	1-10
<input type="checkbox"/> 추가 문헌이 C(계속)에 기재되어 있습니다. <input checked="" type="checkbox"/> 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: "A" 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 "D" 본 국제출원에서 출원인이 인용한 문헌 "E" 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌 "L" 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 "O" 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 "P" 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌		
국제조사의 실제 완료일 2021년12월29일(29.12.2021)		국제조사보고서 발송일 2022년01월03일(03.01.2022)
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578		심사관 양정록 전화번호 +82-42-481-5709

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
US 2020-0342891 A1	2020/10/29	CN 111863015 A	2020/10/30
US 2020-0111483 A1	2020/04/09	US 10140980 B2	2018/11/27
		US 10529320 B2	2020/01/07
		US 10714078 B2	2020/07/14
		US 11069344 B2	2021/07/20
		US 2018-0174575 A1	2018/06/21
		US 2019-0115013 A1	2019/04/18
		US 2019-0156819 A1	2019/05/23
		US 2020-0286468 A1	2020/09/10
WO 2019-199554 A1	2019/10/17	EP 3776535 A1	2021/02/17
		US 10957337 B2	2021/03/23
		US 2019-0318757 A1	2019/10/17
KR 10-2018-0111271 A	2018/10/11	US 10593347 B2	2020/03/17
		US 2018-0286425 A1	2018/10/04
KR 10-2018-0115984 A	2018/10/24	KR 10-1934636 B1	2019/01/02
		US 11017791 B2	2021/05/25
		US 2020-0105287 A1	2020/04/02
		WO 2018-190547 A1	2018/10/18