

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일

2024년 1월 4일 (04.01.2024)



(10) 국제공개번호

WO 2024/005403 A1

(51) 국제특허분류:

G10L 21/02 (2006.01)

G10L 25/30 (2013.01)

G10L 15/02 (2006.01)

G06N 3/08 (2006.01)

(21) 국제출원번호:

PCT/KR2023/008049

(22) 국제출원일:

2023년 6월 12일 (12.06.2023)

(25) 출원언어:

한국어

(26) 공개언어:

한국어

(30) 우선권정보:

10-2022-0078040 2022년 6월 27일 (27.06.2022) KR

(71) 출원인: 한양대학교 산학협력단 (IUCF-HYU (INDUSTRY-UNIVERSITY COOPERATION FOUNDATION HANYANG UNIVERSITY)) [KR/KR]; 04763 서울특별시 성동구 왕십리로 222, Seoul (KR).

(72) 발명자: 장준혁 (CHANG, Joon-Hyuk); 04763 서울특별시 성동구 왕십리로 222, Seoul (KR). 최정환 (CHOI, Jeonghwan); 04763 서울특별시 성동구 왕십리로 222, Seoul (KR).

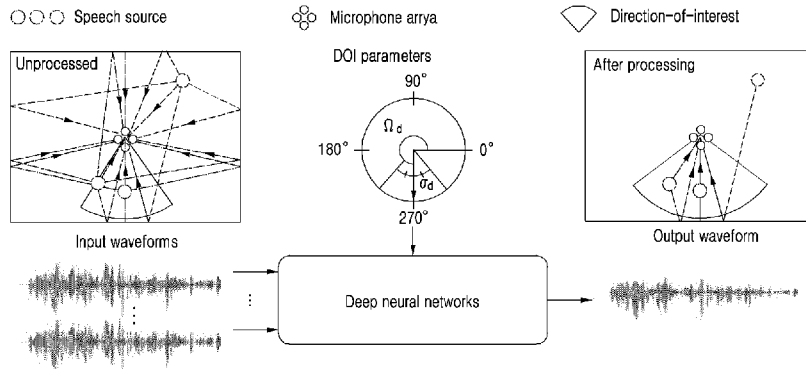
(74) 대리인: 양성보 (YANG, Sungbo); 06099 서울특별시 강남구 선릉로125길 14 삼성빌딩 2층 피앤티특허법률사무소, Seoul (KR).

(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, 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, ME, 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).

(54) Title: SUPERVISED LEARNING METHOD AND SYSTEM FOR EXPLICIT SPATIAL FILTERING OF SPEECH

(54) 발명의 명칭: 음성의 명시적 공간 필터링을 위한 지도 학습 방법 및 시스템



(57) Abstract: A supervised learning method and system for explicit spatial filtering of speech are disclosed. A supervised learning method for spatial filtering of speech performed by a beamformer learning system, according to one embodiment, comprises the steps of: receiving a multi-channel speech signal incident on a microphone array in an environment where reverberation exists in a neural network-based beamformer model, and a beam condition expressing a direction of interest; and outputting a request signal corresponding to the beam condition from the multi-channel speech signal by using the neural network-based beamformer model, wherein the neural network-based beamformer model may be trained so as to extract speech signals of azimuth and elevation angles set in the beam condition by using training data.

(57) 요약서: 음성의 명시적 공간 필터링을 위한 지도 학습 방법 및 시스템이 개시된다. 일 실시예에 따른 빔포머 학습 시스템에 의해 수행되는 음성의 공간 필터링을 위한 지도 학습 방법은, 뉴럴 네트워크 기반의 빔포머 모델에 잔향이 존재하는 환경에서 마이크어레이로 입사되는 다채널의 음성 신호와 관심 방향을 표현하는 빔 컨디션(Beam condition)을 입력받는 단계; 및 상기 뉴럴 네트워크 기반의 빔포머 모델을 이용하여 상기 다채널의 음성 신호로부터 상기 빔 컨디션에 대응되는 요구 신호를 출력하는 단계를 포함하고, 상기 뉴럴 네트워크 기반의 빔포머 모델은, 학습 데이터를 이용하여 상기 빔 컨디션에 설정된 방위각 및 고도각의 음성 신호를 추출하도록 학습된 것일 수 있다.

[다음 쪽 계속]

규칙 4.17에 의한 선언서:

- 신규성을 해치지 아니하는 개시 또는 신규성 상실의
예외에 관한 선언 (규칙 4.17(v))

공개:

- 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 음성의 명시적 공간 필터링을 위한 지도 학습 방법 및 시스템

기술분야

- [1] 아래의 설명은 음성의 공간 필터링을 위한 학습 기술에 관한 것이다.

배경기술

- [2] 뉴럴 빔포머(Neural Beamformers)는 음성 신호 처리에서 널리 연구되고 있다. 뉴럴 빔포머는 음성 신호의 품질 대신 인식 성능을 향상시키기 위해 최적화되었다. 뉴럴 빔포머는 자동 음성 인식 시스템의 전처리기로 제안되었으며 신경망 기반 음향 모델과 공동으로 학습되어 음성 신호의 품질 대신 인식 성능을 향상시켰다. 또한, 음성 분리 또는 향상을 위한 뉴럴 빔포머 기술이 제시된 바 있다. 대부분이 평가 메트릭의 성능을 향상시키기 위한 네트워크 아키텍처 설계에 중점을 두었으며 공간 필터링에 대한 영향은 자세히 논의되지 않았다. 또한, 특정 방향에서 음성 신호 사건을 추출하기 위한 뉴럴 빔포머에 대한 여러 연구가 제시된 바 있다. 이러한 뉴럴 빔포머는 도착 방향(Direction-Of-Arrival; DOA) 정보가 필요하며, 목표 신호를 지정하고 시간-주파수 마스크 추정을 위해 도착 방향(DOA)을 기반으로 하는 방향성 특징을 활용한다. 그러나, 정확한 도착 방향 정보가 필요하며 잘못 추정된 도착 방향에서 추론된 출력 신호의 열화를 예측하기 어려울 수 있다. 이에, 뉴럴 빔포머는 도착 방향을 정확하게 추정하는 대신 미리 정의된 보기 방향을 사용하여 목표 도착 방향에 가장 가까운 음성 신호를 추출하도록 학습되는 기술이 제안된 바 있다. 또한, COSNet은 모든 방향으로 조향하고 빔 폭을 조정하는 데 도움이 될 수 있다. 앞서 언급한 방법들과 달리 원하는 방향에 대한 시간 샘플을 정렬하여 빔폭과 스테리링을 조절하여 분리를 위한 공간 범위를 지정할 수 있다. 그러나 고도에 따라 방위각 조향을 위한 시간 지연이 고려된다. 더욱이, 시간 영역에서 샘플을 정확하게 정렬하려면 높은 샘플링 속도가 필요하며 이는 인접한 마이크 사이의 간격에 따라 달라진다.

- [3] 이전 연구들에서는 목표 신호를 잔향 신호로 설정했으나, 이는 초기 반사가 잔향 환경에서 직접 경로만큼 방향성이 있기 때문에 공간 필터링 문제를 복잡하게 만든다. 이와 같이, 기존의 뉴럴 네트워크 기반의 다중 채널 발화 강화(multi-channel speech enhancement) 연구들의 경우, 공간적 필터링을 위한 명시적인 학습 방법에 대한 논의가 부족했다.

발명의 상세한 설명

기술적 과제

- [4] 방위각 및 고도각으로 특정되는 임의의 방향으로부터 입사되는 음성 신호를 추출하는 뉴럴 네트워크 기반의 빔포머 모델을 학습시키는 방법 및 시스템을 제공할 수 있다.

- [5] 직접 경로뿐만 아니라 초기 반사음의 방향성 또한 고려하여 잔향 환경에서의 공간 필터링(spatial filtering)을 위한 요구 신호(desired signal)를 정의하는 방법 및 시스템을 제공할 수 있다.

과제 해결 수단

- [6] 빔포머 학습 시스템에 의해 수행되는 음성의 공간 필터링을 위한 지도 학습 방법은, 뉴럴 네트워크 기반의 빔포머 모델에 잔향이 존재하는 환경에서 마이크어레이로 입사되는 다채널의 음성 신호와 관심 방향을 표현하는 빔 컨디션(Beam condition)을 입력받는 단계; 및 상기 뉴럴 네트워크 기반의 빔포머 모델을 이용하여 상기 다채널의 음성 신호로부터 상기 빔 컨디션에 대응되는 요구 신호를 출력하는 단계를 포함하고, 상기 뉴럴 네트워크 기반의 빔포머 모델은 학습 데이터를 이용하여 상기 빔 컨디션에 설정된 방위각 및 고도각의 음성 신호를 추출하도록 학습된 것일 수 있다.
- [7] 음성의 공간 필터링을 위한 지도 학습 방법은, 상기 빔 컨디션에 따라 결정되는 요구 신호(desired signal)를 정의하기 위해 공간 이득 함수(spatial gain function)가 구성되고, 상기 공간 이득 함수는, 하드 이득 함수(hard gain function)와 소프트 이득 함수(soft gain function)를 포함할 수 있다.
- [8] 상기 입력받는 단계는, 지도 학습 방법을 사용하여 공간 필터를 뉴럴 네트워크 기반의 빔포머 모델에 학습시키기 위해 학습 데이터를 생성하는 단계를 포함할 수 있다.
- [9] 상기 입력받는 단계는, 초기 반사에 공간 이득을 곱한 소스 위치 및 관심 방향의 파라미터를 통해 보기 방향과 빔폭을 위한 빔 컨디션을 결정하는 단계를 포함할 수 있다.
- [10] 상기 입력받는 단계는, 다중 경로에서 직접 경로의 도착 방향과, 이미지 방법(image method)을 이용하여 초기 반사의 단일 경로 전파를 획득하는 단계를 포함할 수 있다.
- [11] 상기 입력받는 단계는, 3차원 공간에서의 방향 정보와 관심 범위를 설정하기 위한 관심 방향(direction-of-interest) 정보를 정의하고, 상기 정의된 관심 방향 정보를 빔 컨디션 벡터로 변환하는 단계를 포함할 수 있다.
- [12] 빔포머 학습 시스템은, 뉴럴 네트워크 기반의 빔포머 모델에 잔향이 존재하는 환경에서 마이크어레이로 입사되는 다채널의 음성 신호와 관심 방향을 표현하는 빔 컨디션(Beam condition)을 입력받는 빔 컨디션 입력부; 및 상기 뉴럴 네트워크 기반의 빔포머 모델을 이용하여 상기 다채널의 음성 신호로부터 상기 빔 컨디션에 대응되는 요구 신호를 출력하는 모델 학습부를 포함하고, 상기 뉴럴 네트워크 기반의 빔포머 모델은 학습 데이터를 이용하여 상기 빔 컨디션에 설정된 방위각 및 고도각의 음성 신호를 추출하도록 학습된 것일 수 있다.

발명의 효과

- [13] 종래 기술의 경우 잔향 환경에서의 방향성을 직접 경로(direct-path)만을 고려하여 결정함에 따라 임의의 방향으로부터 입사되는 소리를 추출하는 것이 아니라 해당 방향과 공간적으로 가장 가까운 소리를 추출하도록 뉴럴 네트워크 기반의 빔포머 모델이 학습되었다. 실시예에 따르면, 초기 반사의 방향성을 고려하여 공간적으로 명시적인 학습 방법을 제안함에 따라 특정 방향으로부터 입사되는 소리들을 사용자가 조정하면서 청취할 수 있다는 장점이 있다.

도면의 간단한 설명

- [14] 도 1은 일 실시예에 있어서, 뉴럴 네트워크 기반의 빔포머 모델을 이용하여 빔 컨디션에 대한 웨이브 폼을 출력하는 개괄적인 동작을 설명하기 위한 도면이다.
- [15] 도 2는 일 실시예에 있어서, 빔포머 학습 시스템의 구성을 설명하기 위한 블록도이다.
- [16] 도 3은 일 실시예에 있어서, 빔포머 학습 시스템에서 음성의 공간 필터링을 위한 지도 학습 방법을 설명하기 위한 흐름도이다.
- [17] 도 4는 일 실시예에 있어서, 뉴럴 네트워크 기반의 빔포머 모델의 구조를 설명하기 위한 예이다.
- [18] 도 5는 일 실시예에 있어서, 공간 이득 함수를 설명하기 위한 도면이다.
- [19] 도 6은 일 실시예에 있어서, 실시예에서 제안된 방법으로 학습된 뉴럴 네트워크 기반의 빔포머 모델의 실환경 데이터 셋에 대한 출력을 나타낸 도면이다.

발명의 실시를 위한 최선의 형태

- [20] 이하, 실시예를 첨부한 도면을 참조하여 상세히 설명한다.
- [21]
- [22] 실시예에서는 방위각 및 고도각으로 특정되는 임의의 방향으로부터 입사되는 음성 신호를 추출하는 뉴럴 네트워크 기반의 빔포머 모델을 명시적으로 학습시키는 동작에 대하여 설명하기로 한다. 마이크로폰 어레이(microphone array)가 장착된 전자 기기에서의 활용이 가능하다. 이를 위해, 3차원 공간에서의 특정 방향 및 관심 범위를 특정짓기 위한 관심 방향(Direction-Of-Interest; DOI) 정보를 정의하고, 정의된 관심 방향 정보를 빔 컨디션 벡터 형태로 모델에 컨디셔닝할 수 있다. 또한, 공간적으로 다양한 데이터를 생성하기 위한 학습 데이터 생성 동작에 대하여 설명하기로 한다.
- [23] 도 1은 일 실시예에 있어서, 뉴럴 네트워크 기반의 빔포머 모델을 이용하여 빔 컨디션에 대한 웨이브 폼을 출력하는 개괄적인 동작을 설명하기 위한 도면이다.
- [24] 잔향이 존재하는 음성 신호가 마이크로폰 어레이로 입사될 때, 빔 컨디션이라는 조건이 뉴럴 네트워크 기반의 빔포머 모델에 입력될 수 있다. 이때, 빔 컨디션은 방위각과 고도각을 갖는 관심 방향의 파라미터들이 조절될 수 있다. 빔포머 모델은 빔 컨디션에 대응되는 웨이브폼(waveform)을 결과로서 출력할 수 있다.

[25] 도 1에서는 잔향이 존재하는 방환경에서 N명의 화자가 M개의 마이크로 구성된 평면 어레이에 대해 발화하는 상황을 고려하기로 한다. 시간 t에서 m번째 마이크로에서 캡처된 신호는 다음과 같이 표현될 수 있다.

$$[26] \quad y_m(t) = \sum_{n=1}^N h_m(r_n, t) * s_n(t) + v_m(t) \quad (1)$$

[27] 여기서, $s_n(t)$ 는 n번째 화자로부터 발화되는 음성 소스, $h_m(r_n, t)$ 는 레퍼런스 마이크로 $s_n(t)$ 의 상대적인 위치를 나타내는 r_n 로부터 m번째 마이크로 of 다중 경로 음향 전파(multipath acoustic propagation), $v_m(t)$ 는 공간적으로 상관되지 않은 마이크로 자체 노이즈를 의미한다.

[28] 실내 음향에서 $h_m(r_n, t)$ 는 다음과 같이 직접, 초기 반사 및 후기 잔향 성분으로 분해될 수 있다.

$$[29] \quad h_m(r_n, t) = h_m^{\text{direct}}(r_n, t) + h_m^{\text{early}}(r_n, t) + h_m^{\text{late}}(r_n, t) \quad (2)$$

[30] 여기서, 앞의 두 텀은 입의 도착 방향(direction-of-arrival; DOA) $\Omega \triangleq (\theta, \varnothing)$ (여기서, $\theta \in [-180^\circ, 180^\circ]$ 과 $\varnothing \in [0^\circ, 90^\circ]$ 는 각각 방위각 및 고도각을 의미)를 갖는 뚜렷한 방향성이 있는 성분으로 구성되어 있다. 이러한 관점에서 직접 반사 경로(direct reflection paths)와 초기 반사 경로(early reflection paths)는 다음과 같이 표현될 수 있다.

$$[31] \quad h_m^{\text{direct}}(r_n, t) + h_m^{\text{early}}(r_n, t) = \sum_{i=1}^I h_m^{\Omega_i}(r_n, t) \quad (3)$$

[32] 여기서, $h_m^{\Omega_i}(r_n, t)$ 는 입사각 Ω_i 을 갖는 단일 경로 전파를 의미하며, I는 실시예에서 고려되는 총 경로의 수를 의미한다. 다음으로 요구 관심 방향(desired direction-of-interest; DOI) Ω_d 는 요구 도착 방향(desired DOA) Ω_d 및 빔-앵글(beam-angle) σ_d 을 기반으로 하는 셋으로 정의할 수 있다.

$$[33] \quad \Omega_d = \{ \Omega_i | \Delta \sigma_{i,d} < \sigma_d \} \quad (4)$$

[34] where

$$[35] \quad \Delta \sigma_{i,d} = \arccos(u_{\Omega_i} \cdot u_{\Omega_d}) \quad (5)$$

[36] 수학식 5는 Ω_i 와 Ω_d 의 각도 차이를 의미하고,

$u_{\Omega} = [\cos\theta\cos\varnothing, \sin\theta\cos\varnothing, \sin\varnothing]^T$ 는 각도 Ω 에 해당하는 단위 벡터를 의미한다. 입의 첫 번째 마이크로를 레퍼런스로 결정하면, 요구 신호(desired signal)를 다음과 같이 정의할 수 있다.

[37]
$$z_d(t) = \sum_{n=1}^N h_1^{\Omega_d}(r_n, t) * s_1(t) \quad (6)$$

[38] where

[39]
$$h_1^{\Omega_d}(r_n, t) = \sum_{i=1}^I g(\Delta\sigma_{i,d}) \cdot h_1^{\Omega_d}(r_n, t) \quad (7)$$

[40] 수학식 7은 임의의 각도 차이에 기반한 공간 이득(spatial gain) $g(\cdot)$ 이 곱해진 관심 방향(DOI) 내의 단일 경로(single path)들의 합을 의미한다. 본 발명에서의 목표는 $y_1(t), \dots, y_m(t)$ 에서 Ω_d 및 σ_d 로 특정되는 요구 빔(desired beam)에 해당하는 $z_d(t)$ 를 추출하는 것이다.

[41] 도 2는 일 실시예에 있어서, 빔포머 학습 시스템의 구성을 설명하기 위한 블록도이고, 도 3은 일 실시예에 있어서, 빔포머 학습 시스템에서 음성의 공간 필터링을 위한 지도 학습 방법을 설명하기 위한 흐름도이다.

[42] 빔포머 학습 시스템(100)의 프로세서는 빔 컨디션 입력부(210) 및 신호 출력부(220)를 포함할 수 있다. 이러한 프로세서의 구성요소들은 빔포머 학습 시스템에 저장된 프로그램 코드가 제공하는 제어 명령에 따라 프로세서에 의해 수행되는 서로 다른 기능들(different functions)의 표현들일 수 있다. 프로세서 및 프로세서의 구성요소들은 도 3의 음성의 공간 필터링을 위한 지도 학습 방법이 포함하는 단계들(310 내지 320)을 수행하도록 빔포머 학습 시스템을 제어할 수 있다. 이때, 프로세서 및 프로세서의 구성요소들은 메모리가 포함하는 운영체제의 코드와 적어도 하나의 프로그램의 코드에 따른 명령(instruction)을 실행하도록 구현될 수 있다.

[43] 프로세서는 음성의 공간 필터링을 위한 지도 학습 방법을 위한 프로그램의 파일에 저장된 프로그램 코드를 메모리에 로딩할 수 있다. 예를 들면, 빔포머 학습 시스템에서 프로그램이 실행되면, 프로세서는 운영체제의 제어에 따라 프로그램의 파일로부터 프로그램 코드를 메모리에 로딩하도록 빔포머 학습 시스템을 제어할 수 있다. 이때, 프로세서의 빔 컨디션 입력부(210) 및 신호 출력부(220) 각각은 메모리에 로딩된 프로그램 코드 중 대응하는 부분의 명령을 실행하여 이후 단계들(310 내지 320)을 실행하기 위한 프로세서의 서로 다른 기능적 표현들일 수 있다.

[44] 단계(310)에서 빔 컨디션 입력부(210)는 뉴럴 네트워크 기반의 빔포머 모델에 잔향이 존재하는 환경에서 마이크어레이로 입사되는 다채널의 음성 신호와 관심 방향을 표현하는 빔 컨디션(Beam condition)을 입력받을 수 있다. 빔 컨디션 입력부(210)는 3차원 공간에서의 방향 정보와 관심 범위를 특정짓기 위한 관심 방향(direction-of-interest) 정보를 정의하고, 정의된 관심 방향 정보를 빔 컨디션 벡터로 변환할 수 있다. 빔 컨디션 입력부(210)는 지도 학습 방법을 사용하여 공간 필터를 뉴럴 네트워크 기반의 빔포머 모델에 학습시키기 위해 학습 데이터를 생

성할 수 있다. 빔 컨디션 입력부(210)는 초기 반사에 공간 이득을 곱한 소스 위치 및 관심 방향의 파라미터의 조합을 통해 보기 방향과 빔폭을 위한 빔 컨디션을 결정할 수 있다. 빔 컨디션 입력부(210)는 다중 경로에서 직접 경로의 도착 방향과, 이미지 방법(image method)을 이용하여 초기 반사의 단일 경로 전파를 획득할 수 있다.

- [45] 단계(320)에서 신호 출력부(220)는 뉴럴 네트워크 기반의 빔포머 모델을 이용하여 다채널의 음성 신호로부터 빔 컨디션에 대응되는 요구 신호를 출력할 수 있다. 신호 출력부(220)는 입력받은 빔 컨디션에 기초하여 직접 경로와 초기 반사의 방향성을 고려한 잔향 환경에서의 공간적 필터링을 위한 요구 신호를 추출할 수 있다.
- [46] 도 4는 일 실시예에 있어서, 뉴럴 네트워크 기반의 빔포머 모델의 구조를 설명하기로 한다. 이때, 뉴럴 네트워크 기반의 빔포머 모델은 FiLM 을 포함한 구조로 구성된 것일 수 있다. 뉴럴 네트워크 기반의 빔포머 모델은 빔 컨디션을 입력받을 수 있는 임의의 발화 분리(speech separation) 모듈에 적용 가능하다. 설명의 이해를 위하여, 도 4와 같이 인코더, 디코더, 추정기를 포함하는 뉴럴 네트워크 기반의 빔포머 모델의 구조를 예를 들어 설명하기로 한다.
- [47] 뉴럴 네트워크 기반의 빔포머 모델은 Conv-TasNet 아키텍처가 채택되고 관심 방향(DOI) 정보를 사용하도록 수정된 것일 수 있다. 네트워크 $F(\cdot)$ 는 1차원 컨볼루션 인코더(Conv1D) $\varepsilon(\cdot)$, 1차원 전치된 컨볼루션 디코더(1D transposed convolutional decoder; TConv1D) $D(\cdot)$ 및 조건부 마스크 추정기(conditional mask estimator) $M(\cdot)$ 로 구성될 수 있다.
- [48] 조건부 마스크 추정기 $M(\cdot)$ 의 주요 부분은 R번 반복되는 다른 팽창 계수를 가진 S개의 연속 1차원 컨볼루션 블록이 있는 시간 컨볼루션 네트워크(Temporal convolutional network; TCN)이다. 조건부 마스크 추정기 $M(\cdot)$ 에 요구되는 관심 방향 (DOI) 정보를 부과하기 위해 모든 1차원 컨볼루션 블록 다음에 특징별 선형 변조(feature-wise linear modulation; FiLM)를 추가하여 시간 컨볼루션이 수정될 수 있다. $y_m \in \mathbb{R}^{1 \times T}$ 를 $y_m(t)$ 에서 길이 T의 청크라고 하자. y_m 의 잠재 표현은 $Y_m = \varepsilon(y_m) \in \mathbb{R}^{K \times L}$ 로 1차원 컨볼루션 인코더(Conv1D) $\varepsilon(\cdot)$ 에 입력됨에 따라 획득될 수 있다. 여기서, K와 L은 각각 컨볼루션 커널과 프레임의 수를 나타낸다. $Y_1, \dots, Y_M$ 는 커널 차원 $Y \in \mathbb{R}^{M \times K \times L}$ 을 따라 연결되고, 계층 정규화에 이어 지점별 컨볼루션(pointwise convolutional; PointConv) 계층에 제공되어 커널 차원 MK를 B로 변환한다. $F_{s,r} \in \mathbb{R}^{B \times L}$ 를 s번째 스택 및 r번째 반복된 1차원 컨볼루션 블록의 출력이라고 하자.

- [49] Ω_d 와 σ_d 는 DOI 벡터 $b_d = [u_{\Omega_d}^T, \tilde{\sigma}_d]^T \in \mathbb{R}^{4 \times 1}$ 로 변환되며, 여기서,
 $\tilde{\sigma}_d = (\sigma_d - \sigma_{\min}) / (\sigma_{\max} - \sigma_{\min})$ 는 0과 1을 갖는 정규화된 빔폭(normalized beamwidth)이고, $\sigma_{\min}$ 와 $\sigma_{\max}$ 는 σ_d 의 최소값과 최대값인 빔폭 파라미터를 나타낸다. 네트워크는 $\sigma_{\min}$ 및 $\sigma_{\max}$ 에 의해 결정되는 광범위한 빔폭이 처리되도록 학습될 수 있다. b_d 는 시간 컨볼루션 네트워크(TCN)의 각 FiLM 계층에 공급되고, 관심 방향(DOI) 정보를 부과하기 위해 $F_{s,r}$ 을 변조하는 $\beta_{s,r}$ 와 $\gamma_{s,r}$ 로 변환된다.
- [50] 구체적으로, $\beta_{s,r}$ 와 $\gamma_{s,r}$ 는 컨볼루션 커널 B 가 있는 지점별 컨볼루션 계층에 b_d 를 전달하여 획득되고, $f_{s,r,l} = F_{s,r} e_l \in \mathbb{R}^{B \times 1}$ 에 적용되며, 이는 $F_{s,r}$ 의 l 번째 프레임 벡터이다. 여기서, $e_l = [0, \dots, \underbrace{1}_{l\text{th}}, \dots, 0]^T \in \mathbb{R}^{L \times 1}$ 이다.
- FiLM $(f_{s,r,l} | \gamma_{s,r}, \beta_{s,r}) = \gamma_{s,r} \odot f_{s,r,l} + \beta_{s,r}$ 이다. 여기서, $\odot$ 는 요소별 곱셈을 나타낸다. 요구 마스크는 $M_d = M(Y | b_d) \in \mathbb{R}^{K \times L}$ 으로부터 획득하고, 요구되는 신호의 잠재 표현은 $Z_d = Y_1 \odot M_d$ 으로 계산될 수 있다. 마침내, 요구되는 신호의 청크는 다음과 같이 Z_d 를 $D(\cdot)$ 에 전달하여 다음과 같이 재구성될 수 있다($Z_d = D(Z_d) \in \mathbb{R}^{1 \times T}$).
- [51] 도 5는 일 실시예에 있어서, 공간 이득 함수를 설명하기 위한 도면이다.
- [52] 빔 조건에 따라 결정되는 요구 신호(desired signal)를 정의하기 위해 공간 이득 함수(spatial gain function)가 구성될 수 있다. 도 5(a)는 공간 이득 함수, 도 5(b)는 Ω_d 값이 다른 반구를 시각화한 것이다.
- [53] 두 가지 타입의 공간 이득 함수가 고려될 수 있다. 하드 이득 함수는 이상적인 공간 필터에 대응되며, 다음과 같이 표현될 수 있다.
- [54]
$$g_{\text{hard}}(\sigma) = \begin{cases} 1 & \sigma \leq \sigma_d \\ 0 & \text{otherwise} \end{cases} \quad (8)$$
- [55] g_{hard} 의 사용은 직관적으로 이상적이지만, 성능이 마이크의 개수에 따라 한정적이므로 다음과 같이 소프트 이득 함수가 정의될 수 있다. 소프트 이득 함수를 사용하여 다음과 같이 σ_d 에서 급격한 변화를 완화할 수 있다.
- [56]
$$g_{\text{soft}}(\sigma) = \exp(k_d (\cos \sigma - 1)) \quad (9)$$
- [57] 여기서, $k_d = \ln(0.7071) / (\cos \sigma_d - 1)$ 는 하드 이득 함수의 3dB 감쇄에 해당하는 경계로 설정된 파라미터이다.

- [58] 지도 학습 방법을 사용하여 명시적인 공간 필터를 뉴럴 네트워크 기반의 빔포머 모델에 학습시키기 위해 초기 반사에 공간 이득을 곱한 소스 위치 및 관심 방향(DOI) 파라미터에 대한 다양한 조합의 빔 컨디션 및 빔 컨디션에 대한 다채널 데이터가 필요하다.
- [59] 이에, 학습 데이터 생성 동작에 대하여 설명하기로 한다. 우선적으로, 초기 반사 및 도착 방향(DOA)에 대하여 설명하기로 한다. $h_m(r_n, t)$ 의 직접 경로는 $h_m^{\text{direct}}(r_n, t) = \delta(t - \tau_{m,n}) / 4\pi d_{m,n}$ 로 표현될 수 있다. 여기서, $\delta(\cdot)$ 는 크로네커 델타 함수(Kronecker delta function)를 나타내고, $d_{m,n}$ 및 $\tau_{m,n}$ 은 각각 m 번째 마이크로폰과 r_n 사이의 거리와 시간 지연이다. 직접 경로의 도착 방향(DOA)은 다음과 같이 계산될 수 있다.
- [60] $\Omega = (\text{atan2}(r_{y_n}, r_{x_n}), \arcsin(r_{z_n} / \|r_n\|))$ (10)
- [61] 여기서, atan2 는 2-argument arctangent를 의미하고, $\|r_n\|$ 는 r_n 의 길이를 나타낸다.
- [62] 또한, 이미지 방법에서는 공간(방)이 완벽하게 반사되는 단단한 벽으로 둘러싸여 있다고 가정한다. 이는 이미지 소스의 위치가 방에 대한 소스 위치의 대칭 전치에 의해 계산될 수 있음을 의미한다. 단순화를 위해, 모든 방이 동일한 반사 계수 ρ 를 나타낸다고 가정하면, 초기 반사의 단일 경로 전파는 다음과 같이 획득될 수 있다.
- [63] $h_m^{\Omega}(r_n, t) \equiv \rho^{N(i)} \cdot h_m^{\text{direct}}(r_n^{(i)}, t)$ (11)
- [64] 여기서, $N(i)$ 는 i 번째 경로의 반사 횟수이고, $r_n^{(i)}$ 는 i 번째 이미지 소스의 위치를 나타낸다. 따라서, 수학적 식 10과 11을 사용하여 I 개의 이미지 소스 및 해당하는 도착 방향(DOA) 즉, $\{h_m^{\Omega_i}(r, t) \Omega_i\}_{i=1}^I$ 의 단일 경로 전파를 계산할 수 있다.
- [65] 다음으로 요구 신호 생성에 대하여 설명하기로 한다. 빔 컨디션의 경우 확률 분포에서 관심 방향(DOI) 파라미터를 이용하여 보기 방향과 빔폭을 위한 다양한 조합을 생성할 수 있다. 먼저, σ_d 는 아래와 같이 $\sigma_{\min}$ 에서 $\sigma_{\max}$ 까지 N_{σ} 개의 등간격의 원소로 구성된 다음의 셋 $\{\sigma_n\}_{n=1}^{N_{\sigma}}$ 으로부터 샘플링될 수 있다.
- [66] $\sigma_d \sim u(\{\sigma_n\}_{n=1}^{N_{\sigma}})$ (12)
- [67] 여기서, $u(\{\cdot\})$ 는 셋 $\{\cdot\}$ 에 대한 균일 분포를 나타낸다. u_{Ω_d} 는 다양한 보기 방향에 대한 학습 샘플을 생성하기 위해 von-Mises Fisher(vMF) 분포로부터 샘플링할 수 있다. 여기서, 평균 방향은 소스 위치 중에서 무작위로 선택된 정규화된 벡터이고, k_d 는 컨센트레이션(concentration)이다.

[68] $u\Omega_d \sim \text{VMF}(r_k/\|r_k\|, k_d) \quad (13)$

[69] 여기서, $k \sim u\{1, \dots, N\}$ 이다. 각 보기 방향(시선 방향)에 해당하는 다양한 요구 신호로 네트워크를 학습시킬 수 있다. 관심 방향(DOI)의 파라미터가 결정된 후, 요구 신호(목표 신호)는 알고리즘 1에 따라 계산될 수 있다.

[70] 알고리즘 1:

[71]

Algorithm 1 On-the fly generation of training data.

```
// Generate input signal and DOI parameters
Sample N source position from the set  $\{r_n\}_{n=1}^{N_0}$ 
Load  $\{h_m(r_1, t)\}_{m=1}^M, \dots, \{h_m(r_N, t)\}_{m=1}^M$ 
Compute mixture signals  $y_1(t), \dots, y_M(t)$  using Eq.(1)
Sample DOI parameters  $\sigma_d$  and  $\Omega_d$  using Eqs. (12) and (13)
// Generate target signal
Load  $\{h_1^{\Omega_d}(r_1, t)\}_{i=1}^I, \dots, \{h_1^{\Omega_d}(r_N, t)\}_{i=1}^I$ 
Compute the difference in angles  $(\Delta\sigma_{i,d})$  using Eq. (5)
Compute the spatial gains  $g(\Delta\sigma_{i,d})$  using Eq. (8) or (9)
Compute  $h_1^{\Omega_d}(r_n, t)$  using Eq. (7)
Compute target signal  $z_d(t)$  using Eq. (6)
```

[72] 도 6은 일 실시예에 있어서, 실시예에서 제안된 방법으로 학습된 뉴럴 네트워크 기반의 빔포머 모델의 실환경 데이터 셋에 대한 출력을 나타낸 도면이다.

[73] 실시예에 따르면, python 프로그래밍 언어로 구현될 수 있고, pytorch 모듈이 활용될 수 있다. 학습을 위해 음성 데이터 셋 및 room impulse response(RIR) 생성기를 사용하여 데이터 셋이 생성될 수 있다. 마이크 잡음의 경우 white Gaussian noise를 음성 신호에 25~30 dB signal-to-noise ratio(SNR)로 섞어 준다. 학습에 사용된 RIR의 경우 잔향을 나타내는 reverberation time(RT60)가 0.29 ~ 0.7 초 범위로, 가로, 세로, 높이는 [3.5, 3.5, 2.5] ~ [10, 8, 4] m의 방 환경이 시뮬레이션될 수 있다. 반지름이 3.25cm인 4개의 원형 마이크 어레이를 x-y평면에 배치하였으며 각 방마다 음원의 위치는 약 600개 정도의 서로 다른 도착 방향(DOA)를 고려하여 설정할 수 있다.

[74] 반사 경로(reflected path) I는 총 337개로 설정될 수 있다. 빔 조건 파라미터는 각각 5도, 40도, 20으로 설정될 수 있다. 하이퍼 파라미터로는 Adam optimizer를 learning rate 0.001, batch size 24로 200 에포크까지 설정하여 뉴럴 네트워크 기반의 빔포머 모델이 학습될 수 있다.

[75] 두 개의 실제 녹음(왼쪽 상단)과 각 녹음(왼쪽 중간 및 왼쪽 하단)의 혼합에 대한 스펙트로그램이다. 고정 $\tilde{\sigma}_d = 0.5$ (상단-중간), $\tilde{\sigma}_d = 0.1$ (상단-우측)로 $u\Omega_d$ 변경하여 얻은 혼합물의 SR맵이다. x는 각 신호에 대해 추정된 도착 방향을 의미

하는 로컬 최대값을 나타낸다. 각 추정 도착 방향을 네트워크에 입력하여 획득된 각 출력에 대한 스펙트로그램은 도 6과 같다. 실시예에서 제안된 방법으로 학습된 뉴럴 네트워크 기반의 빔포머 모델이 실환경 데이터 셋에서도 기대대로 동작될 수 있다.

발명의 실시를 위한 형태

- [76] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.
- [77] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.
- [78] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크,

플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

- [79] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.
- [80] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

청구범위

- [청구항 1] 빔포머 학습 시스템에 의해 수행되는 음성의 공간 필터링을 위한 지도 학습 방법에 있어서,
 뉴럴 네트워크 기반의 빔포머 모델에 잔향이 존재하는 환경에서 마이크 어레이로 입사되는 다채널의 음성 신호와 관심 방향을 표현하는 빔 컨디션(Beam condition)을 입력받는 단계; 및
 상기 뉴럴 네트워크 기반의 빔포머 모델을 이용하여 상기 다채널의 음성 신호로부터 상기 빔 컨디션에 대응되는 요구 신호를 출력하는 단계를 포함하고,
 상기 뉴럴 네트워크 기반의 빔포머 모델은, 학습 데이터를 이용하여 상기 빔 컨디션에 설정된 방위각 및 고도각의 음성 신호를 추출하도록 학습된 것을 특징으로 하는 음성의 공간 필터링을 위한 지도 학습 방법.
- [청구항 2] 제1항에 있어서,
 상기 빔 컨디션에 따라 결정되는 요구 신호(desired signal)를 정의하기 위해 공간 이득 함수(spatial gain function)가 구성되고,
 상기 공간 이득 함수는, 하드 이득 함수(hard gain function)와 소프트 이득 함수(soft gain function)를 포함하는, 것을 특징으로 하는 음성의 공간 필터링을 위한 지도 학습 방법.
- [청구항 3] 제1항에 있어서,
 상기 입력받는 단계는,
 지도 학습 방법을 사용하여 공간 필터를 뉴럴 네트워크 기반의 빔포머 모델에 학습시키기 위해 학습 데이터를 생성하는 단계를 포함하는 음성의 공간 필터링을 위한 지도 학습 방법.
- [청구항 4] 제3항에 있어서,
 상기 입력받는 단계는,
 초기 반사에 공간 이득을 곱한 소스 위치 및 관심 방향의 파라미터의 조합을 통해 보기 방향과 빔폭을 위한 빔 컨디션을 결정하는 단계를 포함하는 음성의 공간 필터링을 위한 지도 학습 방법.
- [청구항 5] 제1항에 있어서,
 상기 입력받는 단계는,
 다중 경로에서 직접 경로의 도착 방향과, 이미지 방법(image method)을 이용하여 초기 반사의 단일 경로 전파를 획득하는 단계를 포함하는 음성의 공간 필터링을 위한 지도 학습 방법.
- [청구항 6] 제1항에 있어서,
 상기 입력받는 단계는,

3차원 공간에서의 방향 정보와 관심 범위를 설정하기 위한 관심 방향 (direction-of-interest) 정보를 정의하고, 상기 정의된 관심 방향 정보를 빔 컨디션 벡터로 변환하는 단계

를 포함하는 음성의 공간 필터링을 위한 지도 학습 방법.

[청구항 7]

빔포머 학습 시스템에 있어서,

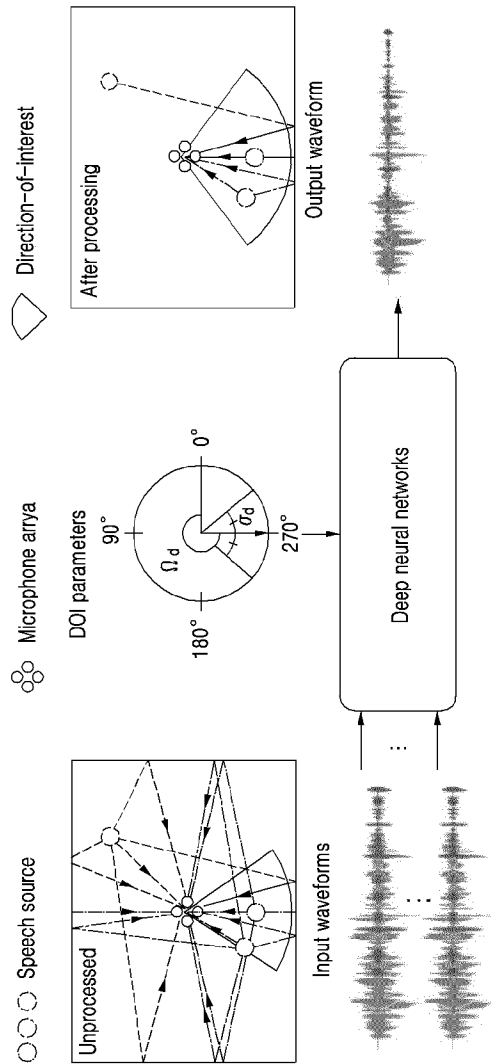
뉴럴 네트워크 기반의 빔포머 모델에 잔향이 존재하는 환경에서 마이크 어레이로 입사되는 다채널의 음성 신호와 관심 방향을 표현하는 빔 컨디션(Beam condition)을 입력받는 빔 컨디션 입력부; 및

상기 뉴럴 네트워크 기반의 빔포머 모델을 이용하여 상기 다채널의 음성 신호로부터 상기 빔 컨디션에 대응되는 요구 신호를 출력하는 신호 출력부

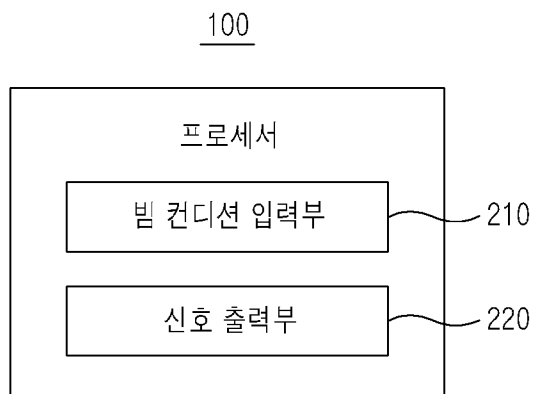
를 포함하고,

뉴럴 네트워크 기반의 빔포머 모델은, 학습 데이터를 이용하여 상기 빔 컨디션에 설정된 방위각 및 고도각의 음성 신호를 추출하도록 학습된 것을 특징으로 하는 빔포머 학습 시스템.

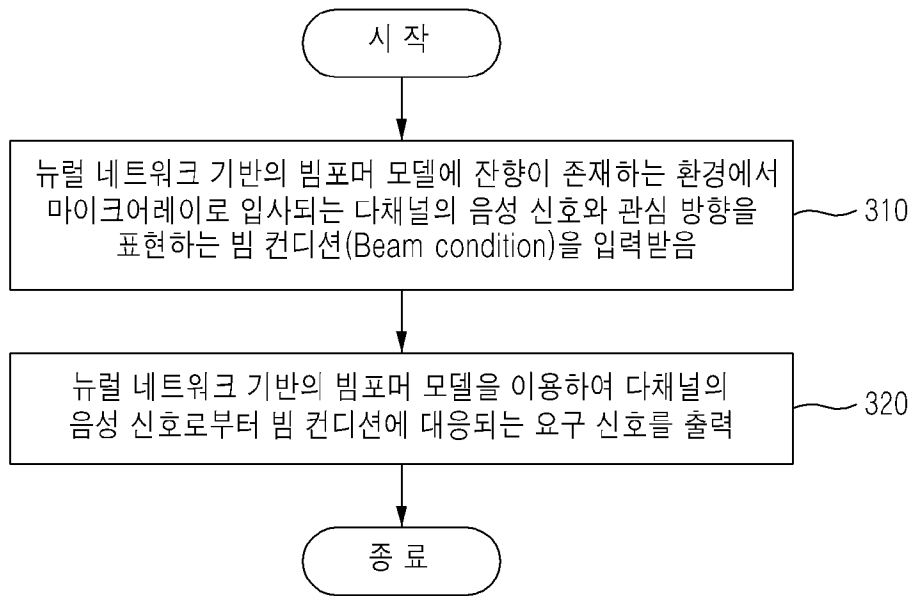
[도1]



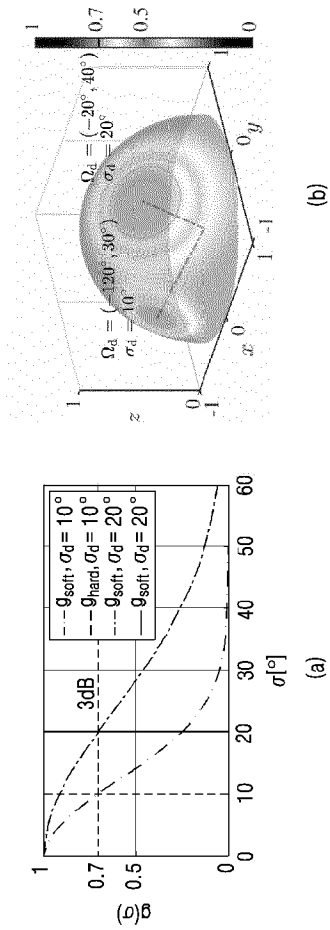
[도2]



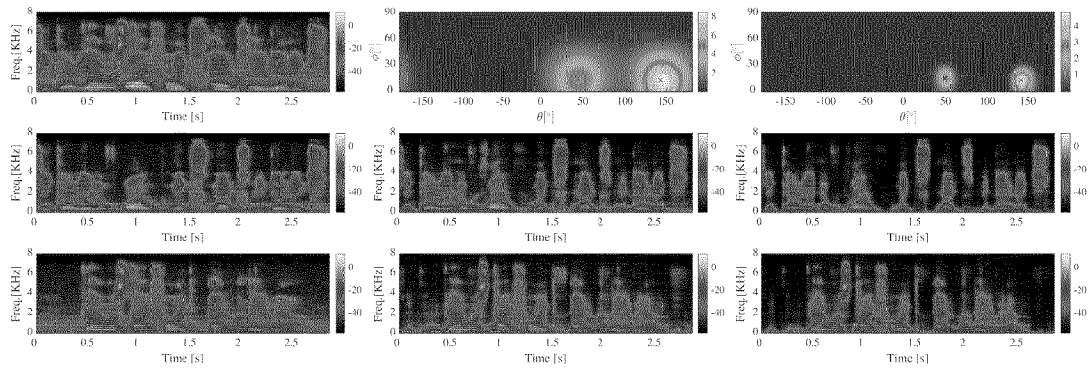
[도3]



[도5]



[도6]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2023/008049

A. CLASSIFICATION OF SUBJECT MATTER		
G10L 21/02(2006.01)i; G10L 25/30(2013.01)i; G10L 15/02(2006.01)i; G06N 3/08(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) G10L 21/02(2006.01); G06N 3/02(2006.01); G10L 21/0208(2013.01); G10L 21/0216(2013.01); G10L 25/12(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: 빔포머 모델(beamformer model), 음성(voice), 공간 필터링(space filtering), 관심 방향(direction of interest), 빔 컨디션(beam condition), 방위각(direction angle), 고도각(altitude angle)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YU, Meng et al. End-to-End Multi-Look Keyword Spotting. arXiv:2005.10386v1 [eess.AS]. 20 May 2020. [Retrieved on 04 September 2023]. Retrieved from < https://www.semanticscholar.org/reader/0b3b732864e65c8882d4d9b014596de127c616a7 >.	1,7
A	See pages 1-2; and figure 1.	2-6
A	KR 10-2020-0145219 A (IUCF-HYU (INDUSTRY-UNIVERSITY COOPERATION FOUNDATION HANYANG UNIVERSITY)) 30 December 2020 (2020-12-30) See paragraphs [0033]-[0036]; and figure 1.	1-7
A	KR 10-2019-0108711 A (IUCF-HYU (INDUSTRY-UNIVERSITY COOPERATION FOUNDATION HANYANG UNIVERSITY)) 25 September 2019 (2019-09-25) See paragraphs [0047]-[0063]; and figure 2.	1-7
A	KR 10-2022-0022286 A (PUSAN NATIONAL UNIVERSITY INDUSTRY-UNIVERSITY COOPERATION FOUNDATION) 25 February 2022 (2022-02-25) See paragraphs [0035]-[0042]; and figure 2.	1-7
☒ Further documents are listed in the continuation of Box C. ☒ 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 14 September 2023		Date of mailing of the international search report 14 September 2023
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

International application No.

PCT/KR2023/008049

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-2021-0017252 A (SAMSUNG ELECTRONICS CO., LTD.) 17 February 2021 (2021-02-17) See paragraphs [0060]-[0076]; and figure 4.	1-7
X	CHOI, Jeonghwan et al. Supervised Learning Approach for Explicit Spatial Filtering of Speech. IEEE Signal Processing Letters. pp. 1-5, 13 June 2022. [Retrieved on 04 September 2023]. Retrieved from < https://ieeexplore.ieee.org/document/9794625 >, doi: <10.1109/LSP.2022.3181971>. See pages 1-5. ※ This document is a document declaring 'exceptions to lack of novelty' by the applicant.	1-7

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2023/008049

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
KR	10-2020-0145219	A	30 December 2020	KR	10-2316537	B1	22 October 2021
				US	2022-0199095	A1	23 June 2022
				WO	2020-256257	A2	24 December 2020
				WO	2020-256257	A3	11 March 2021
KR	10-2019-0108711	A	25 September 2019	KR	10-2087307	B1	10 March 2020
KR	10-2022-0022286	A	25 February 2022	KR	10-2410850	B1	20 June 2022
KR	10-2021-0017252	A	17 February 2021	US	11308973	B2	19 April 2022
				US	2021-0043220	A1	11 February 2021
				WO	2021-025515	A1	11 February 2021

A. 발명이 속하는 기술분류(국제특허분류(IPC)) G10L 21/02(2006.01)i; G10L 25/30(2013.01)i; G10L 15/02(2006.01)i; G06N 3/08(2006.01)i		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) G10L 21/02(2006.01); G06N 3/02(2006.01); G10L 21/0208(2013.01); G10L 21/0216(2013.01); G10L 25/12(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: 빔포머 모델(beamformer model), 음성(voice), 공간 필터링(space filtering), 관심 방향(direction of interest), 빔 컨디션(beam condition), 방위각(direction angle), 고도각(altitude angle)		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	MENG YU 등, End-to-End Multi-Look Keyword Spotting, arXiv:2005.10386v1 [eess.AS], 2020.05.20 [검색일: 2023-09-04], 출처: < https://www.semanticscholar.org/reader/0b3b732864e65c8882d4d9b014596de127c616a7 > 페이지 1-2; 및 도면 1	1,7
A		2-6
A	KR 10-2020-0145219 A (한양대학교 산학협력단) 2020.12.30 단락 [0033]-[0036]; 및 도면 1	1-7
A	KR 10-2019-0108711 A (한양대학교 산학협력단) 2019.09.25 단락 [0047]-[0063]; 및 도면 2	1-7
A	KR 10-2022-0022286 A (부산대학교 산학협력단) 2022.02.25 단락 [0035]-[0042]; 및 도면 2	1-7
A	KR 10-2021-0017252 A (삼성전자주식회사) 2021.02.17 단락 [0060]-[0076]; 및 도면 4	1-7
☒ 추가 문헌이 C(계속)에 기재되어 있습니다. ☒ 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 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. “&” 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2023년09월14일(14.09.2023)		국제조사보고서 발송일 2023년09월14일(14.09.2023)
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578		심사관 변성철 전화번호 +82-42-481-8262

C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	JEONGHWAN CHOI 등, Supervised Learning Approach for Explicit Spatial Filtering of Speech, IEEE Signal Processing Letters, pp. 1-5, 2022.06.13 [검색일: 2023-09-04], 출처: <https://ieeexplore.ieee.org/document/9794625>, doi: <10.1109/LSP.2022.3181971> 페이지 1-5 ※ 본 문헌은 출원인에 의해 ‘신규성 상실의 예외’가 선언된 문헌임.	1-7

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2020-0145219 A	2020/12/30	KR 10-2316537 B1	2021/10/22
		US 2022-0199095 A1	2022/06/23
		WO 2020-256257 A2	2020/12/24
		WO 2020-256257 A3	2021/03/11
KR 10-2019-0108711 A	2019/09/25	KR 10-2087307 B1	2020/03/10
KR 10-2022-0022286 A	2022/02/25	KR 10-2410850 B1	2022/06/20
KR 10-2021-0017252 A	2021/02/17	US 11308973 B2	2022/04/19
		US 2021-0043220 A1	2021/02/11
		WO 2021-025515 A1	2021/02/11