

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2015년 7월 9일 (09.07.2015)



(10) 국제공개번호
WO 2015/102127 A1

- (51) 국제특허분류:
G10L 15/06 (2013.01) *G10L 15/28* (2013.01)
- (21) 국제출원번호: PCT/KR2013/012415
- (22) 국제출원일: 2013년 12월 31일 (31.12.2013)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (71) 출원인: 엘지전자 주식회사 (LG ELECTRONICS INC.) [KR/KR]; 150-721 서울시 영등포구 여의도동 20, Seoul (KR). 한국과학기술원 (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY) [KR/KR]; 305-701 대전시 유성구 대학로 291, Daejeon (KR).
- (72) 발명자: 윤환식 (YUN, Hwansik); 137-724 서울시 서초구 우면동 16번지 엘지전자 특허센터, Seoul (KR). 정규준 (JUNG, Guejun); 137-724 서울시 서초구 우면동 16번지 엘지전자 특허센터, Seoul (KR). 이청안 (LEE, Cheongan); 156-708 서울시 동작구 서달로 1길 17 106동 1104호, Seoul (KR). 이수영 (LEE, Sooyoung); 135-871 서울시 강남구 삼성동 78-4 청구아파트 102-1101, Seoul (KR). 김호경 (KIM, Hogyong); 704-140 대구시 달서구 이곡동 성서우방타운 103-703, Daegu (KR). 김

유진 (KIM, Eugene); 137-724 서울시 서초구 우면동 16번지 엘지전자 특허센터, Seoul (KR).

(74) 대리인: 김용인 (KIM, Yong In) 등; 138-861 서울시 송파구 잠실동 175-9 현대빌딩 7층 KBK 특허법률사무소, Seoul (KR).

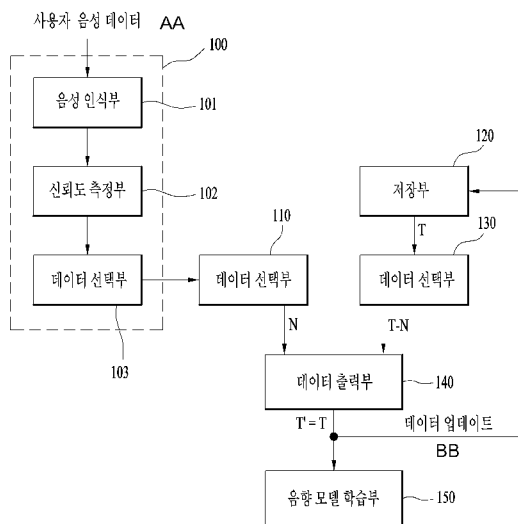
(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, 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, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR),

[다음 쪽 계속]

(54) Title: SYSTEM AND METHOD FOR VOICE RECOGNITION

(54) 발명의 명칭 : 음성 인식 시스템 및 방법



101 ... Voice recognition unit
102 ... Reliability measurement unit
103, 110, 130 ... Data selection unit
120 ... Storage unit
140 ... Data output unit
150 ... Acoustic model learning unit
AA ... User voice data
BB ... Update data

(57) Abstract: Disclosed are a system and a method for voice recognition. The system for voice recognition may comprise: a labeling unit for recognizing inputted user voice data, performing automatic labeling for the recognized user voice data and then primarily selecting and outputting a part of the user voice data containing label information; a first data selection unit for secondarily selecting a part of the user voice data that is primarily selected and outputted by the labeling unit; a second data selection unit for selecting and outputting a part of seed data that was previously used for acoustic model learning; and an acoustic model learning unit for learning an acoustic model using the user voice data that is secondarily selected and outputted by the first data selection unit and the seed data that is selected and outputted by the second data selection unit.

(57) 요약서: 음성 인식 시스템 및 방법이 개시된다. 음성 인식 시스템은 입력되는 사용자 음성 데이터를 인식하고, 인식된 사용자 음성 데이터에 대해 자동 레이블링을 수행한 후 라벨 정보가 있는 사용자 음성 데이터 중 일부를 1차로 선택하여 출력하는 레이블링부, 상기 레이블링부에서 1차로 선택되어 출력되는 사용자 음성 데이터 중 일부를 2차로 선택하여 출력하는 제 1 데이터 선택부, 이전에 음향 모델 학습에 이용되었던 시드 데이터의 일부를 선택하여 출력하는 제 2 데이터 선택부, 및 상기 제 1 데이터 선택부에서 2차로 선택되어 출력되는 사용자 음성 데이터와 상기 제 2 데이터 선택부에서 선택되어 출력되는 시드 데이터를 이용하여 음향 모델을 학습하는 음향 모델 학습부를 포함할 수 있다.



OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, **공개:**
ML, MR, NE, SN, TD, TG).

— 국제조사보고서와 함께 (조약 제 21 조(3))

명세서

발명의 명칭: 음성 인식 시스템 및 방법

기술분야

- [1] 본 발명은 음성 인식에 관한 것으로서, 더욱 상세하게는 음성 인식 성능을 점진적으로 향상시키는 음성 인식 시스템 및 방법에 관한 것이다.

배경기술

- [2] 음성 인식 기술은 프로세서의 고속화, 메모리 양의 증가, 병렬처리 기법, 음성언어 자원의 증가 등으로 인해 인식 성능이 지속적으로 향상되고 있다. 또, 음성 인식 시스템은 서버급 컴퓨터로부터 소형 휴대 단말기 또는 가전기기 등과 같이 다양한 하드웨어 플랫폼 상에 탑재가 되고 있다.
- [3] 음성 인식이란 컴퓨터가 키보드나 마우스 등의 기계적인 조작이 아닌 인간의 음성을 듣고 실행을 하는 기술이다.
- [4] 음성 인식은 흔히 가장 직관적인 사용자 인터페이스로 불린다. 생각을 표현함에 있어서 손이나 몸을 사용하는 것보다 목소리를 사용하는 것이 가장 자연스럽도록 진화된 인간의 신체적 특성에 기인함이다. 키보드나 마우스 등 기존의 많은 인터페이스들이 생각을 직접적으로 표현하기 이전에 문서용 언어로 혹은 미리 정의된 컴퓨터가 알 수 있는 명령으로 변환하는 과정을 거쳐야 하고 그것은 추가적인 사고과정과 더불어 기계에 익숙해지기 위한 추가적인 노력 또한 필요로 한다.
- [5] 음성 인식 기술은 이러한 추가적인 노력을 최소로 줄이고 인간에게 보다 본질적이고 창의적인 사고에 집중할 수 있는 길을 열어 준다.

발명의 상세한 설명

기술적 과제

- [6] 본 발명의 목적은 사용자의 음성 입력을 사용하여 음향 모델을 다시 학습함으로써, 음성 인식 성능을 점진적으로 향상시키는 음성 인식 시스템 및 방법을 제공함에 있다.

과제 해결 수단

- [7] 상기와 같은 목적을 달성하기 위한 본 발명에 따른 음성 인식 시스템은 입력되는 사용자 음성 데이터를 인식하고, 인식된 사용자 음성 데이터에 대해 자동 레이블링을 수행한 후 라벨 정보가 있는 사용자 음성 데이터 중 일부를 1차로 선택하여 출력하는 레이블링부, 상기 레이블링부에서 1차로 선택되어 출력되는 사용자 음성 데이터 중 일부를 2차로 선택하여 출력하는 제1 데이터 선택부, 이전에 음향 모델 학습에 이용되었던 시드 데이터의 일부를 선택하여 출력하는 제2 데이터 선택부, 및 상기 제1 데이터 선택부에서 2차로 선택되어 출력되는 사용자 음성 데이터와 상기 제2 데이터 선택부에서 선택되어 출력되는 시드 데이터를 이용하여 음향 모델을 학습하는 음향 모델 학습부를 포함하는

- 것을 일 실시예로 한다.
- [8] 상기 레이블링부는 상기 라벨 정보가 있는 사용자 음성 데이터의 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 상기 라벨 정보가 있는 사용자 음성 데이터의 일부를 1차로 선택하는 것을 일 실시예로 한다.
- [9] 상기 제1 데이터 선택부는 상기 레이블링부에서 선택되어 출력되는 사용자 음성 데이터의 라벨 정보를 이용하여 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 N (여기서 N 은 1보다 큼)개의 사용자 음성 데이터를 2차로 선택하는 것을 일 실시예로 한다.
- [10] 상기 시드 데이터를 저장하는 저장부를 더 구비하며, 상기 저장부에 저장된 시드 데이터는 상기 음향 모델 학습부에서 음향 모델 학습에 사용된 데이터로 업데이트되는 것을 일 실시예로 한다.
- [11] 상기 제2 데이터 선택부는 상기 저장부에 저장된 시드 데이터의 라벨 정보를 이용하여 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 M (여기서 M 은 1보다 큼)개의 시드 데이터를 선택하는 것을 일 실시예로 한다.
- [12] 상기 제1 데이터 선택부에서 선택되는 사용자 음성 데이터의 개수와 상기 제2 데이터 선택부에서 선택되는 시드 데이터의 개수의 합은 상기 저장부에 저장된 시드 데이터의 개수와 같은 것을 일 실시예로 한다.
- [13] 본 발명에 따른 음성 인식 방법은 입력되는 사용자 음성 데이터를 인식하고, 인식된 사용자 음성 데이터에 대해 자동 레이블링을 수행하여 라벨 정보가 있는 사용자 음성 데이터를 출력하는 단계, 상기 라벨 정보가 있는 사용자 음성 데이터 중 일부를 1차로 선택하는 단계, 상기 1차로 선택된 사용자 음성 데이터 중 일부를 2차로 선택하는 단계, 이전에 음향 모델 학습에 이용되었던 시드 데이터의 일부를 선택하는 단계, 및 상기 2차로 선택된 사용자 음성 데이터와 상기 선택된 시드 데이터를 이용하여 음향 모델을 학습하는 단계를 포함하는 것을 일 실시예로 한다.
- [14] 상기 1차 선택 단계는 상기 라벨 정보가 있는 사용자 음성 데이터의 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 상기 라벨 정보가 있는 사용자 음성 데이터의 일부를 1차로 선택하는 것을 일 실시예로 한다.
- [15] 상기 2차 선택 단계는 상기 1차로 선택된 사용자 음성 데이터의 라벨 정보를 이용하여 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 N (여기서 N 은 1보다 큼)개의 사용자 음성 데이터를 2차로 선택하는 것을 일 실시예로 한다.
- [16] 상기 2차 선택 단계는 상기 2차로 선택되는 사용자 음성 데이터에 대해 가중치 팩터를 적용하여 출력하는 단계를 더 포함하는 것을 일 실시예로 한다.
- [17] 상기 시드 데이터는 저장부에 저장되며, 상기 저장부에 저장된 시드 데이터는 상기 음향 모델 학습에 사용된 데이터로 업데이트되는 것을 일 실시예로 한다.
- [18] 상기 시드 데이터 선택 단계는 상기 저장부에 저장된 시드 데이터의 라벨 정보를 이용하여 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 M (여기서 M 은 1보다 큼)개의 시드 데이터를 선택하는 것을 일 실시예로 한다.

- [19] 상기 2차로 선택되는 사용자 음성 데이터의 개수와 상기 선택되는 시드 데이터의 개수의 합은 상기 저장부에 저장된 시드 데이터의 개수와 같은 것을 일 실시예로 한다.

발명의 효과

- [20] 본 발명은 음향 모델 학습에 이용하였던 기존 데이터(Seed data)와 사용자의 음성 데이터를 모두 이용하여 음향 모델을 다시 학습함으로써, 음성 인식 성능을 점진적으로 향상시키는 효과가 있다.
- [21] 특히 본 발명은 일정량의 사용자 음성 데이터가 모이면 자동적으로 음향 모델을 다시 학습하도록 함으로써, 음성 인식 성능을 점진적으로 향상시키는 효과와 함께 음향 모델을 학습하는 시간을 단축시킬 수 있다. 또한 사용자 음성 데이터 중 일부와 음향 모델 학습에 이용하였던 기존 데이터 중 일부를 선택하여 음향 모델 학습에 이용함으로써, 총 데이터 양을 유지하여 매번 학습 시 필요한 시간을 일정하게 유지할 수 있다.

도면의 간단한 설명

- [22] 도 1은 본 발명에 따른 음성 인식 장치의 일 실시예를 보인 구성 블록도
- [23] 도 2는 본 발명에 따른 음성 인식 방법의 일 실시예를 보인 흐름도

발명의 실시를 위한 최선의 형태

- [24] 이하 상기의 목적을 구체적으로 실현할 수 있는 본 발명의 바람직한 실시예를 첨부한 도면을 참조하여 상세히 설명한다. 이때 도면에 도시되고 또 이것에 의해서 설명되는 본 발명의 구성과 작용은 적어도 하나의 실시예로서 설명되는 것이며, 이것에 의해서 본 발명의 기술적 사상과 그 핵심 구성 및 작용이 제한되지는 않는다.
- [25] 본 발명에서 사용되는 용어는 본 발명에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어를 선택하였으나, 이는 당분야에 종사하는 기술자의 의도 또는 관례 또는 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 발명에서 사용되는 용어는 단순한 용어의 명칭이 아닌 그 용어가 가지는 의미와 본 발명의 전반에 걸친 내용을 토대로 정의되어야 함을 밝혀두고자 한다.
- [26] 또한 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시 예들에 대해서 특정한 구조적 내지 기능적 설명들은 단지 본 발명의 개념에 따른 실시 예를 설명하기 위한 목적으로 예시된 것으로, 본 발명의 개념에 따른 실시 예들은 다양한 형태로 실시될 수 있으며 본 명세서에 설명된 실시 예들에 한정되는 것으로 해석되어서는 아니된다.
- [27] 본 발명의 개념에 따른 실시 예는 다양한 변경을 가할 수 있고 여러 가지 형태를 가질 수 있으므로 특정 실시예들을 도면에 예시하고 본 명세서에 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시 예를 특정한 개시

형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

- [28] 그리고 본 발명에서 제1 및/또는 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로만, 예컨대 본 발명의 개념에 따른 권리 범위로부터 이탈되지 않은 채, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.
- [29] 또한 명세서 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. 그리고 명세서에 기재된 “...부” 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [30] 일반적인 음성 인식 시스템에서는 음향 모델 학습에 사용된 음성 데이터의 녹음 환경과 실제 사용 환경이 달라 인식 성능 저하가 발생한다. 본 발명은 이를 개선하기 위해 사용자의 음성 데이터를 음향 모델 학습에 사용하는 것을 일 실시예로 한다. 하지만 사용자 음성 데이터는 시간이 지남에 따라 계속 추가되어 그 양이 방대해지기 때문에 본 발명은 일정량의 사용자 음성 데이터가 모이면 자동적으로 음향 모델을 다시 학습하는 것을 일 실시예로 한다. 이때 데이터를 많이 모을수록 성능은 좋아지지만 음향 모델을 학습하는데 시간이 많이 걸린다. 본 발명은 또한 이를 개선하기 위해 사용자의 음성 데이터에서 일부를 선정하고 기존 음향 모델 학습에 사용했던 시드 데이터(Seed data)에서 일부를 선정하여 음향 모델 학습에 사용하는 것을 일 실시예로 한다. 따라서 본 발명은 데이터 선별을 통해 총 데이터 양을 유지함으로써, 매번 음향 모델 학습 시 필요한 시간을 일정하게 유지할 수 있게 된다.
- [31] 도 1은 본 발명에 따른 음성 인식 시스템의 일 실시예를 보인 구성 블록도로서, 자동 레이블링부(Labeling unit)(100), 제1 데이터 선택부(110), 저장부(120), 제2 데이터 선택부(130), 데이터 출력부(140), 및 음향 모델 학습부(150)를 포함하는 것을 일 실시예로 한다.
- [32] 상기 자동 레이블링부(100)는 사용자 음성 데이터를 인식하고 인식된 사용자 음성 데이터에 레이블링 작업을 수행하는 음성 인식부(101), 상기 음성 인식부(101)에서 인식된 사용자 음성 데이터의 신뢰도(confidence measure)를 계산하는 신뢰도 측정부(102), 및 상기 신뢰도 측정부(102)의 계산 결과를 기반으로 신뢰도가 높은 사용자 음성 데이터를 선택하여 출력하는 선택부(103)를 포함하는 것을 일 실시예로 한다.
- [33] 이와 같이 구성된 본 발명에서 상기 자동 레이블링부(100)는 사용자 음성 데이터에 대해 자동적으로 레이블링을 수행한다. 즉, 본 발명은 사람에 의한

레이블링이 아닌 자동 레이블링부(100)를 이용하여 기계적으로 레이블링을 수행한다. 다시 말해, 음향 모델 학습을 위해 사용자가 발성한 음성 데이터에 대한 텍스트 데이터가 필요하기 때문에 레이블링 작업이 필요하며, 본 발명은 점진 학습을 자동적으로 수행하기 위해서 음성 인식 결과를 사용하여 레이블링을 수행한다.

- [34] 이를 위해 상기 자동 레이블링부(100)의 음성 인식부(101)는 사용자가 발성하는 사용자 음성 데이터를 인식하고, 인식된 사용자 음성 데이터에 대한 레이블링 작업을 수행하여 라벨 정보가 있는 사용자 음성 데이터를 신뢰도 측정부(102)로 출력한다.
- [35] 상기 신뢰도 측정부(102)는 라벨 정보가 있는 사용자 음성 데이터의 신뢰도를 계산하여 데이터 선택부(103)로 출력하고, 상기 데이터 선택부(103)는 신뢰도가 높은 사용자 음성 데이터만을 선택한 후 음향 모델 학습을 위해 제1 데이터 선택부(110)로 출력한다.
- [36] 예를 들어, 사용자가 “학교에 간다”라고 발성했을 때 음성 인식부(101)의 후보에 “학교에 간다”가 있으면 사용자가 발성한 음성 데이터에 “학교에 간다”라는 라벨이 붙는다. 그런데 상기 음성 인식부(101)의 후보에 “학교에 간다”가 없고 “전화해”가 있으면, 사용자는 “학교에 간다”라고 발성했어도 라벨은 “전화해”가 붙을 수 있다. 이것은 사용자 음성 데이터에 대한 레이블링 작업이 잘못 수행된 예이며, 이러한 사용자 음성 데이터가 음향 모델 학습에 이용되면 결국 인식 성능 저하가 발생하게 된다. 따라서 본 발명에서는 이를 해결하기 위해 신뢰도 측정부(102)와 선택부(103)가 추가되어 신뢰도가 높은 데이터를 선택하여 출력하는 과정이 추가로 수행된다. 즉, 상기 신뢰도 측정부(102)와 선택부(103)는 음성 인식부(101)에서 출력되는 라벨이 있는 사용자 음성 데이터 중 실제 사용자가 발성한 사용자 음성 데이터와 일치할 확률이 높은 사용자 음성 데이터만을 선택하여 출력하게 된다. 다시 말해, 상기 신뢰도 측정부(102)와 선택부(103)는 레이블링 작업이 제대로 수행된 사용자 음성 데이터만을 선별하여 출력하는 과정을 수행하게 된다.
- [37] 상기 제1 데이터 선택부(110)는 자동 레이블링부(100)에서 출력되는 사용자 음성 데이터 중 일부를 음향 모델 학습에 이용하기 위해 선택하여 데이터 출력부(140)로 출력한다. 또한 상기 제2 데이터 선택부(130)는 상기 저장부(120)에 저장된 시드 데이터(seed data) 중 일부를 음향 모델 학습에 이용하기 위해 선택하여 상기 데이터 출력부(140)로 출력한다. 이때 상기 제1 데이터 선택부(110)와 제2 데이터 선택부(130)에서 선택하는 데이터 양을 일정하게 유지시킴으로써, 매번 음향 모델 학습시 필요한 시간을 일정하게 유지할 수 있게 된다.
- [38] 일 예로, 상기 제1 데이터 선택부(110)는 음향 모델 성능 개선에 도움이 될거라고 예측되는 사용자 음성 데이터를 선택하여 음향 모델 학습에 반영한다.
- [39] 이때, 제1 데이터 선택부(110)로 입력되는 사용자 음성 데이터는 발성(utterance)

단위로 되어있다고 가정한다. 그리고 발성 단위로 사용자 음성 데이터를 선택한다고 가정할 때 상기 제1 데이터 선택부(110)는 각 발성에 대해 다음의 수학적 식 1과 같이 신뢰도(confidence measure) 계산을 수행하여 음향 모델 학습부(150)로 출력할 데이터를 결정하게 된다.

[40] 수학적 식 1

$$J_{cs}^u = \sum_{k=1}^N |J_k^u|, \quad J_k^u = g(h_1^k - h_2^k),$$

where

$$h_1^k = -\max_l \log p(f_k^u | \theta_l), \quad h_2^k = -\max_{l, l \neq l^*} \log p(f_k^u | \theta_l),$$

$$l^* = \arg \max_l \log p(f_k^u | \theta_l).$$

[41] 상기 수학적 식 1에서 각 발성은 음소 단위로 세그먼트되어 있다고 가정할 때, f_k^u 는 u번째 발성의 k번째 음소에 대한 특징 벡터(feature vector, 여기서 특징 벡터는 음성 데이터를 인식 및 학습에 적합하도록 변환한 값), θ_l 은 l번째 클래스에 대한 음향 모델 파라미터를 의미한다. 따라서 h_1^k 는 k번째 음소에 가장 적합한 클래스에 대한 확률 값이고, h_2^k 는 k번째 음소에 두 번째로 적합한 클래스에 대한 확률 값이다. 결국 두 값의 차이인 J_k^u 가 작을수록 k번째 음소가 어느 클래스에 속하는지 구분하기 어려우며 인식 오류가 발생할 확률이 높다는 의미가 된다.

[42] 상기 수학적 식 1을 적용하여 한 발성에 속한 각 음소의 J_{cs}^u 값에 대한 절대값의 합계를 구해 그 발성을 선택할지 결정하게 된다.

[43] 즉, J_{cs}^u 값이 작은 데이터가 인식 오류가 발생할 확률이 높은 데이터이고, 이 데이터를 음향 모델 학습에 반영할 경우 기존 음향 모델의 취약한 부분을 개선시킬 수 있게 된다.

[44] 상기 수학적 식 1은 상기 제1 데이터 선택부(110)로 입력되는 사용자 음성 데이터에 라벨이 없다고 가정하고, 라벨이 있는 데이터와 라벨이 없는 데이터로 나누어 신뢰도 계산을 하고 있다.

[45] 이때 본 발명은 자동 레이블링부(100)에서 사용자 음성 데이터에 대해 레이블링 작업을 수행하였기 때문에 상기 제1 데이터 선택부(110)로 출력되는 데이터는 라벨 정보가 있는 데이터이다. 그러므로, 상기 제1 데이터 선택부(110)는 데이터를 선택할 때 하기 수학적 식 2와 같이 라벨 정보를 이용할 수 있다.

[46] 수학적 식 2

$$I_{cs}^u = \sum_{k=1}^N |I_k^u|, \quad I_k^u = g(e_1^k - e_2^k),$$

$$e_1^k = -\log p(\mathbf{f}_k^u | \theta_{i^u}), \quad e_2^k = -\max_{l, l \neq i^u} \log p(\mathbf{f}_k^u | \theta_l).$$

- [47] 상기 수학식 2에서 e_1^k 는 k번째 음소에 가장 적합한 클래스에 대한 확률 값이고, e_2^k 는 k번째 음소에 두 번째로 적합한 클래스에 대한 확률 값이다. 그리고 라벨(Label) 정보에 의해 k번째 음소가 어떤 클래스에 속하는지 이미 알고 있으므로 e_1^k 을 구하는 방식이 상기 수학식 1과 다르며, 나머지 수식은 수학식 1과 동일하다.
- [48] 상기 제2 데이터 선택부(130)는 상기 저장부(120)에 저장된 시드 데이터 중 일부를 선택하여 음향 모델 학습부(150)로 출력한다. 즉, 기존에 음향 모델의 학습에 사용되었던 시드 데이터를 모두 사용할 경우 데이터가 점점 늘어나기 때문에 본 발명은 일정한 수의 데이터를 선택하여 다음 음향 모델 학습에 사용한다. 다시 말해, 기존의 음향 모델 학습에 사용되었던 데이터는 레이블링 및 신뢰도 계산이 수행된 데이터이므로 많이 모을수록 음향 모델 성능을 더욱 높일 수 있지만 데이터가 많아질수록 음향 모델을 학습하는데 시간이 많이 걸린다. 그러므로, 제2 데이터 선택부(130)에서도 상기 저장부(120)에 저장된 데이터 중 일부를 선택하여 출력하는 과정을 수행한다.
- [49] 상기 제2 데이터 선택부(130)에서도 수학식 1을 적용하여 데이터를 선택할 수 있다. 또한 상기 저장부(120)에 저장된 데이터는 이미 음향 모델 학습에 사용했던 데이터이므로, 각 데이터에 라벨 정보가 있다.
- [50] 따라서 상기 제2 데이터 선택부(130)에서도 상기 수학식 2를 적용하여 데이터를 선택할 수 있다.
- [51] 본 발명은 음향 모델 학습에 이용될 총 데이터량을 일정하게 유지하기 위하여 상기 제1 데이터 선택부(110)에서 N개의 사용자 음성 데이터를 선택하였다면, 상기 제2 데이터 선택부(130)에서는 T-N개의 시드 데이터를 선택하는 것을 일 실시예로 한다. 여기서 T는 상기 저장부(120)에 저장된 시드 데이터의 개수이다.
- [52] 이때 상기 제1 데이터 선택부(110)에서 선택되는 데이터는 기존 데이터의 양에 비해 상대적으로 작아 음향 모델 학습에 영향을 주기 어려울 수 있다. 이를 해결하기 위해 상기 제1 데이터 선택부(110)는 가중치 팩터(weighting factor)를 적용하여 제1 데이터 선택부(110)에서 출력되는 사용자 음성 데이터의 특성이 음향 모델에 잘 반영될 수 있도록 하는 것을 일 실시예로 한다. 이때 처음에는

- 가중치 팩터(weighting factor) 값을 크게 하고 반복(iteration)이 증가할수록 기존 데이터와 같은 영향력을 갖도록 가중치 팩터를 줄이는 것을 일 실시예로 한다.
- [53] 상기 제1 데이터 선택부(110)에서 선택된 사용자 음성 데이터와 상기 제2 데이터 선택부(130)에서 선택된 시드 데이터는 데이터 출력부(140)에서 합성되어 음향 모델 학습부(150)와 저장부(120)로 출력된다. 상기 저장부(120)는 데이터 출력부(140)에서 출력되는 데이터로 기 저장된 시드 데이터를 업데이트한다.
- [54] 상기 음향 모델 학습부(150)는 상기 데이터 출력부(140)에서 출력되는 데이터를 이용하여 음향 모델 학습을 수행한다. 이때 상기 음향 모델 학습부(150)는 상기 저장부(120)에 저장된 시드 데이터를 이용하여 초기 음향 모델을 학습한다. 초기 음향 모델을 학습한 이후에는 음향 모델이 수렴할 때까지 사용자 음성 데이터의 일부와 시드 데이터의 일부를 선택하여 음향 모델을 학습하고, 음향 모델 학습에 이용된 데이터로 시드 데이터를 업데이트하는 과정을 반복한다.
- [55] 상기 음향 모델 학습부(150)는 dHMM(Discrete Hidden Markov Model) 알고리즘을 적용하여 음향 모델 학습을 수행하는 것을 일 실시예로 한다. 여기서 dHMM는 각 음소의 확률 분포를 이산적인 값으로 모델링한다. 상기 dHMM 알고리즘 대신 HMM 알고리즘이 사용될 수도 있다. 상기 HMM은 통계적 패턴 인식을 기반으로 하며 단어 단위 검색과 문장 단위 검색 과정이 하나의 최적화 과정으로 통합된 방식이다. 상기 dHMM 알고리즘과 HMM 알고리즘은 본 발명의 이해를 돕기 위한 실시예들이며, 음향 모델을 학습하기 위한 알고리즘은 어느 것이나 사용될 수 있다. 그리고 상기 음향 모델 학습부(150)에서 사용하는 알고리즘에 따라 자동 레이블링부(100), 제1 데이터 선택부(110), 및 제2 데이터 선택부(130)에서 신뢰도를 계산하는 방법이 달라질 수 있다.
- [56] 도 2는 본 발명에 따른 음성 인식 방법의 일 실시예를 보인 흐름도이다. 즉, 사용자가 발성하는 사용자 음성 데이터를 인식하고, 인식된 사용자 음성 데이터에 대해 레이블링 작업을 수행한다(S201). 그리고 상기 레이블링 작업에 의해 라벨이 있는 사용자 음성 데이터의 신뢰도를 계산하고, 이 중 신뢰도가 높은 사용자 음성 데이터를 1차로 선택한다(S202). 상기 신뢰도는 수학적 1 및/또는 수학적 2를 적용하여 계산할 수 있다.
- [57] 상기 단계 S202에서 1차로 선택된 사용자 음성 데이터에 대해 다시 신뢰도를 계산하고, 계산된 신뢰도를 이용하여 1차로 선택된 사용자 음성 데이터 중 일부를 2차로 선택한다(S203).
- [58] 또한 저장부(120)에 저장된 시드 데이터에 대해 신뢰도를 계산하고, 계산된 신뢰도를 이용하여 일부 시드 데이터를 선택한다(S205).
- [59] 상기 단계 S203에서 선택된 사용자 음성 데이터와 상기 단계 S205에서 선택된 시드 데이터를 이용하여 음향 모델 학습을 수행한다(S206). 이때 상기 음향 모델 학습에 이용된 데이터는 단계 S204로 진행하여 상기 저장부(120)에 저장된 시드

데이터를 업데이트한다.

- [60] 상기 단계들은 음향 모델이 수렴할 때까지 반복되며, 음향 모델이 수렴되면 음향 모델 학습 과정이 종료되는 것을 일 실시예로 한다. 이때 상기 음향 모델 학습 과정에서 음향 모델이 수렴하지 않는 경우는 학습 과정에서의 예외 사항이므로, 이 과정은 생략될 수도 있다.

- [61] 지금까지 설명한 본 발명은 상술한 실시예에 한정되지 않으며, 첨부된 청구범위에서 알 수 있는 바와 같이 본 발명이 속한 분야의 통상의 지식을 가진 자에 의해 변형이 가능하고 이러한 변형은 본 발명의 범위에 속한다.

발명의 실시를 위한 형태

- [62] 전술한 바와 같이, 상기 발명의 실시를 위한 최선의 형태에서, 관련된 사항을 기술하였다.

산업상 이용가능성

- [63] 전술한 바와 같이, 본 발명은 음성 인식 기술을 적용하는 장치 및 단말기에 모두 적용될 수 있다.

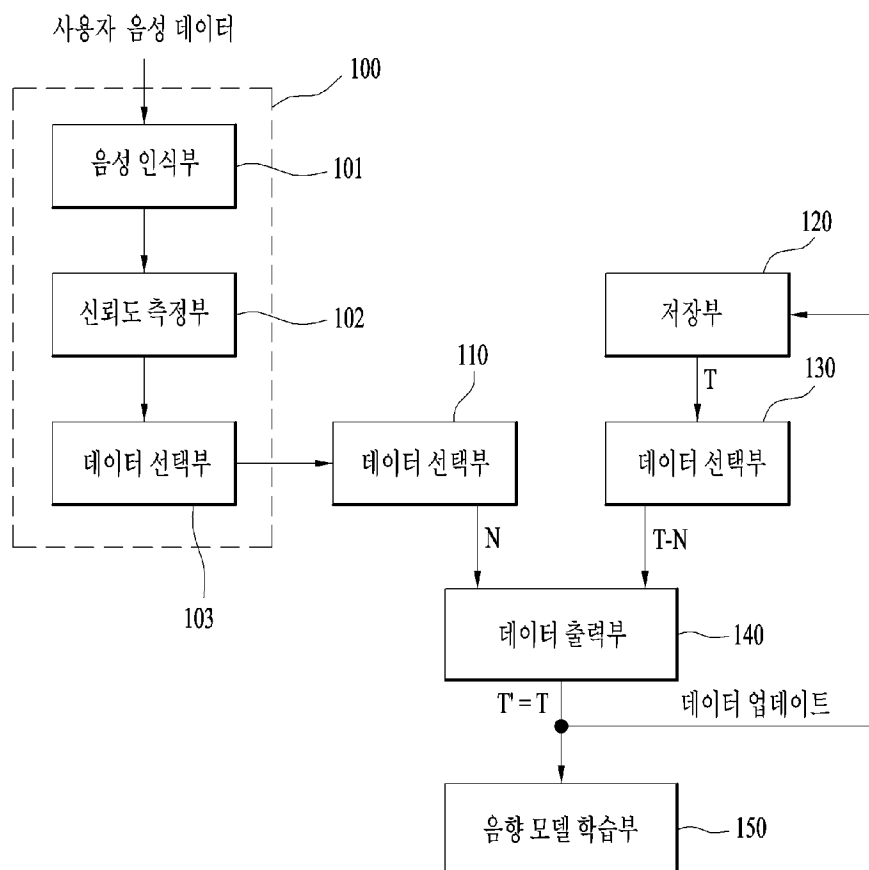
청구범위

- [청구항 1] 입력되는 사용자 음성 데이터를 인식하고, 인식된 사용자 음성 데이터에 대해 자동 레이블링을 수행한 후 라벨 정보가 있는 사용자 음성 데이터 중 일부를 1차로 선택하여 출력하는 레이블링부;
상기 레이블링부에서 1차로 선택되어 출력되는 사용자 음성 데이터 중 일부를 2차로 선택하여 출력하는 제1 데이터 선택부;
이전에 음향 모델 학습에 이용되었던 시드 데이터의 일부를 선택하여 출력하는 제2 데이터 선택부; 및
상기 제1 데이터 선택부에서 2차로 선택되어 출력되는 사용자 음성 데이터와 상기 제2 데이터 선택부에서 선택되어 출력되는 시드 데이터를 이용하여 음향 모델을 학습하는 음향 모델 학습부를 포함하는 것을 특징으로 하는 음성 인식 시스템.
- [청구항 2] 제 1 항에 있어서, 상기 레이블링부는
상기 라벨 정보가 있는 사용자 음성 데이터의 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 상기 라벨 정보가 있는 사용자 음성 데이터의 일부를 1차로 선택하는 것을 특징으로 하는 음성 인식 시스템.
- [청구항 3] 제 2 항에 있어서, 상기 제1 데이터 선택부는
상기 레이블링부에서 선택되어 출력되는 사용자 음성 데이터의 라벨 정보를 이용하여 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 N (여기서 N 은 1보다 큼)개의 사용자 음성 데이터를 2차로 선택하는 것을 특징으로 하는 음성 인식 시스템.
- [청구항 4] 제 3 항에 있어서,
상기 시드 데이터를 저장하는 저장부를 더 구비하며,
상기 저장부에 저장된 시드 데이터는 상기 음향 모델 학습부에서 음향 모델 학습에 사용된 데이터로 업데이트되는 것을 특징으로 하는 음성 인식 시스템.
- [청구항 5] 제 4 항에 있어서, 상기 제2 데이터 선택부는
상기 저장부에 저장된 시드 데이터의 라벨 정보를 이용하여 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 M (여기서 M 은 1보다 큼)개의 시드 데이터를 선택하는 것을 특징으로 하는 음성 인식 시스템.
- [청구항 6] 제 5 항에 있어서,
상기 제1 데이터 선택부에서 선택되는 사용자 음성 데이터의 개수와 상기 제2 데이터 선택부에서 선택되는 시드 데이터의 개수의 합은 상기 저장부에 저장된 시드 데이터의 개수와 같은

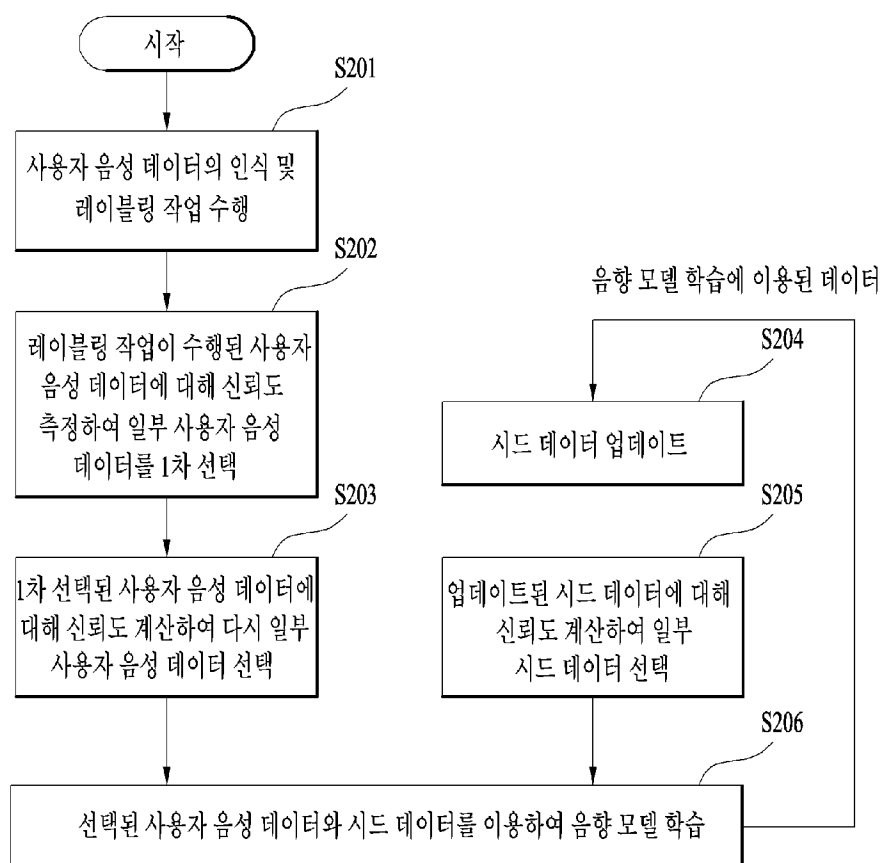
- 것을 특징으로 하는 음성 인식 시스템.
- [청구항 7] 입력되는 사용자 음성 데이터를 인식하고, 인식된 사용자 음성 데이터에 대해 자동 레이블링을 수행하여 라벨 정보가 있는 사용자 음성 데이터를 출력하는 단계;
상기 라벨 정보가 있는 사용자 음성 데이터 중 일부를 1차로 선택하는 단계;
상기 1차로 선택된 사용자 음성 데이터 중 일부를 2차로 선택하는 단계;
이전에 음향 모델 학습에 이용되었던 시드 데이터의 일부를 선택하는 단계; 및
상기 2차로 선택된 사용자 음성 데이터와 상기 선택된 시드 데이터를 이용하여 음향 모델을 학습하는 단계를 포함하는 것을 특징으로 하는 음성 인식 방법.
- [청구항 8] 제 7 항에 있어서, 상기 1차 선택 단계는
상기 라벨 정보가 있는 사용자 음성 데이터의 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 상기 라벨 정보가 있는 사용자 음성 데이터의 일부를 1차로 선택하는 것을 특징으로 하는 음성 인식 방법.
- [청구항 9] 제 8 항에 있어서, 상기 2차 선택 단계는
상기 1차로 선택된 사용자 음성 데이터의 라벨 정보를 이용하여 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 N (여기서 N 은 1보다 큼)개의 사용자 음성 데이터를 2차로 선택하는 것을 특징으로 하는 음성 인식 방법.
- [청구항 10] 제 9 항에 있어서, 상기 2차 선택 단계는
상기 2차로 선택되는 사용자 음성 데이터에 대해 가중치 팩터를 적용하여 출력하는 단계를 더 포함하는 것을 특징으로 하는 음성 인식 방법.
- [청구항 11] 제 9 항에 있어서,
상기 시드 데이터는 저장부에 저장되며, 상기 저장부에 저장된 시드 데이터는 상기 음향 모델 학습에 사용된 데이터로 업데이트되는 것을 특징으로 하는 음성 인식 방법.
- [청구항 12] 제 11 항에 있어서, 상기 시드 데이터 선택 단계는
상기 저장부에 저장된 시드 데이터의 라벨 정보를 이용하여 신뢰도를 계산하고, 계산된 신뢰도를 기반으로 M (여기서 M 은 1보다 큼)개의 시드 데이터를 선택하는 것을 특징으로 하는 음성 인식 방법.
- [청구항 13] 제 12 항에 있어서,
상기 2차로 선택되는 사용자 음성 데이터의 개수와 상기 선택되는

시드 데이터의 개수의 합은 상기 저장부에 저장된 시드 데이터의 개수와 같은 것을 특징으로 하는 음성 인식 방법.

[Fig. 1]



[Fig. 2]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2013/012415**A. CLASSIFICATION OF SUBJECT MATTER*****G10L 15/06(2006.01)i, G10L 15/28(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 15/06; G10L 15/00; G10L 15/18; G10L 15/28

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: voice recognition, labeling unit, sound model learning unit, seed data, data selection unit

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2009-0171662 A1 (HUANG, Jun et al.) 02 July 2009 See paragraphs [0021]-[0068]; and figures 1-7.	1-13
A	US 7149687 B1 (GORIN, Allen Louis et al.) 12 December 2006 See column 2, line 42 - column 6, line 40; and figures 1-5B.	1-13
A	JP 2010-243914 A (KYOTO UNIV.) 28 October 2010 See paragraphs [0035]-[0127]; and figures 1-10.	1-13
A	JP 2012-177815 A (NATIONAL INSTITUTE OF INFORMATION & COMMUNICATION TECHNOLOGY) 13 September 2012 See paragraphs [0012]-[0094]; and figures 1-8.	1-13
A	KR 10-1183344 B1 (MICROSOFT CORPORATION) 14 September 2012 See paragraphs [0027]-[0058]; and figures 1-4.	1-13



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

"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

"I" 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

24 SEPTEMBER 2014 (24.09.2014)

Date of mailing of the international search report

24 SEPTEMBER 2014 (24.09.2014)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 189 Seonsa-ro, Daejeon 302-701,
Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2013/012415

Patent document cited in search report	Publication date	Patent family member	Publication date
US 2009-0171662 A1	02/07/2009	US 8583416 B2	12/11/2013
US 7149687 B1	12/12/2006	US 2014-156275 A1 US 8650033 B1	05/06/2014 11/02/2014
JP 2010-243914 A	28/10/2010	JP 05366050 B2	11/12/2013
JP 2012-177815 A	13/09/2012	NONE	
KR 10-1183344 B1	14/09/2012	CN 1645477 A CN 1645477 B EP 1557822 A1 EP 1557822 B1 JP 04657736 B2 JP 2005-208643 A KR 10-2005-0076697 A US 2005-0159949 A1 US 2011-0015927 A1 US 8019602 B2 US 8280733 B2	27/07/2005 11/01/2012 27/07/2005 25/05/2011 23/03/2011 04/08/2005 26/07/2005 21/07/2005 20/01/2011 13/09/2011 02/10/2012

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G10L 15/06(2006.01)i, G10L 15/28(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G10L 15/06; G10L 15/00; G10L 15/18; G10L 15/28

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: 음성 인식, 레이블링부, 음향 모델 학습부, 시드 데이터, 데이터 선택부

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	US 2009-0171662 A1 (JUN HUANG et al.) 2009.07.02 단락 [0021]-[0068]; 및 도면 1-7 참조.	1-13
A	US 7149687 B1 (ALLEN LOUIS GORIN et al.) 2006.12.12 컬럼 2, 라인 42 - 컬럼 6, 라인 40 ; 및 도면 1-5B 참조.	1-13
A	JP 2010-243914 A (KYOTO UNIV.) 2010.10.28 단락 [0035]-[0127]; 및 도면 1-10 참조.	1-13
A	JP 2012-177815 A (NATIONAL INSTITUTE OF INFORMATION & COMMUNICATION TECHNOLOGY) 2012.09.13 단락 [0012]-[0094]; 및 도면 1-8 참조.	1-13
A	KR 10-1183344 B1 (마이크로소프트 코퍼레이션) 2012.09.14 단락 [0027]-[0058]; 및 도면 1-4 참조.	1-13

☐ 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

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌

“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“&” 동일한 대응특허문헌에 속하는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

국제조사의 실제 완료일

2014년 09월 24일 (24.09.2014)

국제조사보고서 발송일

2014년 09월 24일 (24.09.2014)

ISA/KR의 명칭 및 우편주소

 대한민국 특허청
(302-701) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 +82-42-472-7140

심사관

김도원

전화번호 +82-42-481-5560



국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
US 2009-0171662 A1	2009/07/02	US 8583416 B2	2013/11/12
US 7149687 B1	2006/12/12	US 2014-156275 A1 US 8650033 B1	2014/06/05 2014/02/11
JP 2010-243914 A	2010/10/28	JP 05366050 B2	2013/12/11
JP 2012-177815 A	2012/09/13	없음	
KR 10-1183344 B1	2012/09/14	CN 1645477 A CN 1645477 B EP 1557822 A1 EP 1557822 B1 JP 04657736 B2 JP 2005-208643 A KR 10-2005-0076697 A US 2005-0159949 A1 US 2011-0015927 A1 US 8019602 B2 US 8280733 B2	2005/07/27 2012/01/11 2005/07/27 2011/05/25 2011/03/23 2005/08/04 2005/07/26 2005/07/21 2011/01/20 2011/09/13 2012/10/02