



**(19) 대한민국특허청(KR)**  
**(12) 등록특허공보(B1)**

(45) 공고일자 2014년08월01일  
 (11) 등록번호 10-1424496  
 (24) 등록일자 2014년07월23일

(51) 국제특허분류(Int. Cl.)  
 G10L 15/14 (2006.01)  
 (21) 출원번호 10-2013-0077676  
 (22) 출원일자 2013년07월03일  
 심사청구일자 2013년07월03일  
 (56) 선행기술조사문헌  
 KR101229108 B1\*  
 \*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
 에스케이텔레콤 주식회사  
 서울특별시 중구 을지로 65 (을지로2가)  
 (72) 발명자  
 광철  
 경상북도 문경시 문경읍 교촌담골길 9 (교촌리)  
 (74) 대리인  
 박종한

전체 청구항 수 : 총 6 항

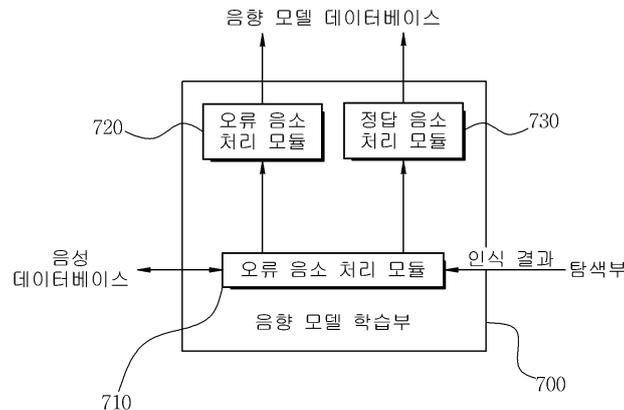
심사관 : 이수철

(54) 발명의 명칭 **음향 모델 학습을 위한 장치 및 이를 위한 방법이 기록된 컴퓨터 판독 가능한 기록매체**

**(57) 요약**

본 발명은 음향 모델 학습을 위한 장치 및 이를 위한 방법이 기록된 컴퓨터 판독 가능한 기록매체에 관한 것으로, 이러한 본 발명은 복수의 음소에 대한 음성 인식 결과로부터 상기 복수의 음소를 정답 인식 결과인 정답 음소와 및 오류 인식 결과인 오류 음소로 구분하는 단계와, 상기 오류 음소의 로그우도를 가중치로 적용하여, 상호 정보량이 최소화되도록 오류 음소의 음소 모델 파라미터 분포를 업데이트하는 단계를 포함하는 음향 모델 학습을 위한 방법을 수행하는 장치와 이 방법이 기록된 컴퓨터 판독 가능한 기록매체를 제공한다.

**대표도** - 도3



**특허청구의 범위**

**청구항 1**

복수의 음소에 대한 음성 인식 결과로부터 상기 복수의 음소를 정답 인식 결과인 정답 음소와 및 오류 인식 결과인 오류 음소로 구분하는 인식 결과 분류 모듈; 및

상기 정답 음소의 음소 모델 파라미터 분포와 상기 오류 음소의 음소 모델 파라미터 분포 간의 중첩되는 확률 분포가 최소화되도록 상기 오류 음소의 로그우도를 가중치로 적용하여 상기 오류 음소의 음소 모델 파라미터 분포를 업데이트하는 오류 음소 처리 모듈;을 포함하는 것을 특징으로 하는 음향 모델 학습을 위한 장치.

**청구항 2**

제1항에 있어서,

상기 오류 음소 처리 모듈은

최대 로그우도(ML, maximum likelihood) 추정 방법을 이용하여 상기 로그우도를 산출하는 것을 특징으로 하는 음향 모델 학습을 위한 장치.

**청구항 3**

제1항에 있어서,

상호 정보량이 최소화되도록 정답 음소의 음소 모델 파라미터 분포를 업데이트하는 정답 음소 처리 모듈;을 더 포함하는 것을 특징으로 하는 음향 모델 학습을 위한 장치.

**청구항 4**

제3항에 있어서,

상기 정답 음소 처리 모듈은

최대 상호정보량(MMI, maximum mutual information) 추정 방법을 이용하여 상기 정답 음소의 음소 모델 파라미터 분포를 업데이트하는 것을 특징으로 하는 음향 모델 학습을 위한 장치.

**청구항 5**

복수의 음소에 대한 음성 인식 결과로부터 상기 복수의 음소를 정답 인식 결과인 정답 음소와 및 오류 인식 결과인 오류 음소로 구분하는 단계; 및

상기 정답 음소의 음소 모델 파라미터 분포와 상기 오류 음소의 음소 모델 파라미터 분포 간의 중첩되는 확률 분포가 최소화되도록 상기 오류 음소의 로그우도를 가중치로 적용하여, 상기 오류 음소의 음소 모델 파라미터 분포를 업데이트하는 단계;를 포함하는 음향 모델 학습을 위한 방법이 기록된 컴퓨터 판독 가능한 기록매체.

**청구항 6**

제5항에 있어서,

상호 정보량이 최소화되도록 정답 음소의 음소 모델 파라미터 분포를 업데이트하는 단계;를 더 포함하는 음향 모델 학습을 위한 방법이 기록된 컴퓨터 판독 가능한 기록매체.

**명세서**

**기술분야**

본 발명은 음성 인식 기술에 관한 것으로, 더욱 상세하게는, 음소 단위 모델 파라미터 분포의 상호 정보량을 최소화할 수 있는 음향 모델 학습을 위한 장치 및 이러한 학습 방법이 기록된 컴퓨터 판독 가능한 기록매체에 관한 것이다.

[0001]

**배경 기술**

- [0002] 음성 인식은 자동적 수단에 의하여 음성으로부터 언어적 의미 내용을 식별하는 것. 구체적으로 음성과형을 입력하여 단어나 단어열을 식별하고 의미를 추출하는 처리 과정이다. 이러한 음성 인식은 크게 음성 분석, 음소 인식, 단어 인식, 문장 해석, 의미 추출의 5가지로 분류될 수 있다. 음성 인식은 좁은 의미로 음성 분석에서 단어 인식까지를 말하는 경우가 있다.
- [0003] 인간-기계 인터페이스 개선의 하나로 음성으로 정보를 입력하는 음성 인식과 음성으로 정보를 출력하는 음성 합성 기술의 연구 개발이 오랫동안 진행되어 왔다. 대형 장치를 필요로 하는 음성 인식 장치와 음성 합성 장치를 대규모 집적 회로(LSI, large scale integrated circuit)의 발달에 따라 가로세로 수 mm 크기의 집적 회로 위에 실현할 수 있게 됨으로써 음성 입출력 장치가 실용화되었다.
- [0004] 현재 전화에 의한 은행 잔액 조회, 증권 시세 조회, 통신 판매의 신청, 신용 카드 조회, 호텔이나 항공기 좌석 예약 등에 사용된다. 그러나 이들 서비스는 제한된 수의 단어를 하나하나 떼어서 발음하는 음성을 인식하는 단어 음성 인식 장치를 사용한다. 음성 인식의 궁극적인 목표는 자연스러운 발성에 의한 음성을 인식하여 실행 명령어로서 받아들이거나 자료로서 문서에 입력하는 완전한 음성 혹은 텍스트 변환의 실현이다. 이는 단지 단어를 인식할 뿐 아니라 구문 정보, 의미 정보, 작업에 관련된 정보와 지식 등을 이용하여 연속 음성 또는 문장의 의미 내용을 정확하게 추출하는 음성 이해 시스템을 개발하는 것이다. 이러한 시스템의 연구 개발이 활발하게 진행되고 있다.

**선행기술문헌**

**특허문헌**

- [0005] (특허문헌 0001) 한국공개특허 제2012-0045582호, 2012.05.09 공개 (명칭: 음향 모델 생성 장치 및 방법)

**발명의 내용**

**해결하려는 과제**

- [0006] 본 발명의 목적은 음성 인식 시스템에서 오류 인식 결과를 가지는 음소에 대한 음소 모델 파라미터 분포의 상호 정보량을 최소화할 수 있는 방법 및 장치를 제공함에 있다.

**과제의 해결 수단**

- [0007] 상술한 바와 같은 목적을 달성하기 위한 본 발명의 바람직한 실시예에 따른 음향 모델 학습을 위한 장치는 복수의 음소에 대한 음성 인식 결과로부터 상기 복수의 음소를 정답 인식 결과인 정답 음소와 및 오류 인식 결과인 오류 음소로 구분하는 인식 결과 분류 모듈과, 상기 오류 음소의 로그우도를 가중치로 적용하여, 상호 정보량이 최소화되도록 오류 음소의 음소 모델 파라미터 분포를 업데이트하는 오류 음소 처리 모듈을 포함한다.
- [0008] 상술한 바와 같은 목적을 달성하기 위한 본 발명의 바람직한 실시예에 따른 컴퓨터 판독 가능한 기록매체는, 복수의 음소에 대한 음성 인식 결과로부터 상기 복수의 음소를 정답 인식 결과인 정답 음소와 및 오류 인식 결과인 오류 음소로 구분하는 단계와, 상기 오류 음소의 로그우도를 가중치로 적용하여, 상호 정보량이 최소화되도록 오류 음소의 음소 모델 파라미터 분포를 업데이트하는 단계를 포함하는 음향 모델 학습을 위한 방법이 기록된다.

**발명의 효과**

- [0009] 상술한 바와 같이 본 발명은 음향 모델의 음소 모델 파라미터 분포를 업데이트할 때, 오류 인식 결과로 구분된 음소의 로그우도를 가중치로 적용하여, 오류 인식 결과도 상호 정보량을 최소화하도록 반영함으로써, 변별 학습(discriminative training)에서 효율적으로 상호 정보량을 최소화시킬 수 있다.

**도면의 간단한 설명**

- [0010] 본 발명에 관한 이해를 돕기 위해 상세한 설명의 일부로 포함되는, 첨부 도면은 본 발명에 대한 실시예를 제공

하고, 상세한 설명과 함께 본 발명의 기술적 특징을 설명한다.

도 1은 본 발명의 실시예에 따른 학습 방법의 개념을 설명하기 위한 개념도이다.

도 2는 본 발명의 실시예에 따른 음성 인식 시스템을 설명하기 위한 도면이다.

도 3은 본 발명의 실시예에 따른 음향 모델 학습부의 내부 구성을 설명하기 위한 도면이다.

도 4는 본 발명의 실시예에 따른 음향 모델 학습 방법을 설명하기 위한 흐름도이다.

도 5는 본 발명의 실시예에 따른 음향 모델 학습부의 음향 모델 학습 방법을 설명하기 위한 흐름도이다.

**발명을 실시하기 위한 구체적인 내용**

[0011] 이하 첨부된 도면을 참조하여 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있는 바람직한 실시예를 상세히 설명한다. 다만, 본 발명의 바람직한 실시예에 대한 동작 원리를 상세하게 설명함에 있어 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 이는 불필요한 설명을 생략함으로써 본 발명의 핵심을 흐리지 않고 더욱 명확히 전달하기 위한 것이다.

[0012] 또한, 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 또한, 본 명세서에서 사용한 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 또한, 본 명세서에서 기술되는 "포함 한다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0013] 그리고, 제1, 제2 등의 용어는 다양한 구성요소들을 설명하기 위해 사용하는 것으로, 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용될 뿐, 상기 구성요소들을 한정하기 위해 사용되지 않는다. 이때, 도면 전체에 걸쳐 유사한 기능 및 작용을 하는 부분에 대해서는 동일한 도면 부호를 사용하며, 이에 대한 중복되는 설명은 생략하기로 한다.

[0014] 도 1a 내지 도 1d는 본 발명의 실시예에 따른 학습 방법의 개념을 설명하기 위한 개념도이다.

[0015] 도 1a 내지 도 1d의 설명에 앞서 본 발명의 실시예에 따른 음성 인식 방법을 개략적으로 살펴보면, 다음과 같다. 음성 인식은 입력된 음성에 대한 음향이 분석되고, 그 음성의 특징량을 나타내는 소정 차원의 특징 벡터의 추출이 이루어진다. 그 후, 특징 벡터와 음향 모델과의 매칭 처리가 이루어진다. 본 발명의 실시예에 따르면, 그 매칭 처리는 음소 단위로 이루어진다. 그 매칭 처리 결과, 특징 벡터에 매칭되는 음향 모델의 음소가 음성 인식 결과(인식 결과)가 된다. 매칭 처리에서는 음향 모델을 구성하는 확률 분포(음소 모델 파라미터 분포)를 이용하여, 음성 인식 결과의 복수의 후보로, 해당 음향 모델에 대한 특징 벡터가 관측되는 로그우도가 계산된다. 예컨대, 입력된 단어(열)이 3개의 음소 A, B 및 C로 이루어진 경우, 다음의 <표 1>과 같은 결과가 출력될 수 있다.

**표 1**

[0016]

	음소 A 인식 결과 (로그우도)	음소 B 인식 결과 (로그우도)	음소 C 인식 결과 (로그우도)
후보 1	E(0.1)	V(0.2)	E(0.1)
후보 2	A(1.2)	V(0.2)	C(1.4)
후보 3	A(1.2)	B(2.2)	C(1.4)

[0017] <표 1>에서, 로그우도는 인식된 음소가 음향 모델에서 해당 음소일 확률이다. 다른 말로, 로그우도는 음향 모델의 음소 모델 파라미터 분포와 입력된 음성의 음소의 유사도(확률)이다. 각 후보의 인식 결과가 산출되면, 그 로그우도에 기초하여 복수의 후보 중에서, 최종적인 음성 인식 결과가 결정된다. <표 1>에서는 후보 3이 선택될 것이다. 즉, 복수의 후보 중 로그우도가 가장 높은 후보가 입력된 음성에 가장 가까운 것으로 선택되고, 그 후

보를 구성하는 음향 모델에 대응하는 단어열이 음성 인식 결과로서 출력될 것이다.

- [0018] 도 1a를 참조하면, <표 1>과 같이 음향 모델에 음소 A, B, C가 존재한다고 가정한다. 도면 부호 10, 20 및 30은 음소 A, B 및 C의 음소 모델 파라미터 분포를 도식화한 것이다. 음소 모델 파라미터 분포는 해당 음소의 확률 분포가 될 수 있다. 여기서, 각 음소의 음소 모델 파라미터 분포가 중첩된 부분인 도면 부호 40은 상호 정보량을 나타낸다.
- [0019] 앞서 설명된 음성 인식 방법을 기반으로 하는 음성 인식 방법에서, 상호 정보량이 많은 경우, 입력된 음성의 음소들을 명확히 구분하는 인식정확도를 저하시키는 문제가 발생한다. 따라서 본 발명의 실시예는 이러한 "상호 정보량을 최소화"하도록 음소 모델 파라미터 분포를 업데이트한다.
- [0020] 본 발명의 실시예에 따른 "상호 정보량 최소화"의 의미에 대해 설명하면 다음과 같다. 도 1b를 참조하면, 2 개의 음소, 즉, 음소 M 및 N에 대한 음소 모델 파라미터 분포가 도시되었다.
- [0021] 도면 부호 50 및 60은 각각 현재 음향 모델의 음소 M 및 N에 대한 음소 모델 파라미터 분포이며, 도면 부호 70 및 80은 각각 음소 M 및 N에 대한 상호 정보량이 최소화된 이상적인 음소 모델 파라미터 분포라고 가정한다.
- [0022] 도면 부호 90은 현재 음향 모델의 음소 M 및 N에 대한 음소 모델 파라미터 분포(50, 60) 간의 상호 정보량이다. 이러한 상호 정보량(90)으로 인하여 음소 M이 입력되었음에도 불구하고, 음소 N으로 인식될 수 있다. 따라서 이러한 상호 정보량을 최소화시켜야 한다. 상호 정보량(90)을 최소화시키기 위해서는 현재 음향 모델의 음소 M의 음소 모델 파라미터 분포(50)가 음소 M의 이상적인 음소 모델 파라미터 분포(70)로 이동되어야 한다. 혹은, 현재 음향 모델의 음소 N의 음소 모델 파라미터 분포(60)가 음소 N의 이상적인 음소 모델 파라미터 분포(80)로 이동되어야 한다.
- [0023] 앞서 설명된 바와 같이, 음성 인식 결과는 입력된 음성 신호의 음소가 음향 모델에 저장된 음소와 동일한 음소 일 확률(유사도)로 출력된다. 훈련 혹은 학습을 위한 음성은 알려져 있는 음성을 이용한다. 따라서 각 음소에 대한 인식 결과가 정답인지 혹은 오류인지 여부를 알 수 있다. 즉, 입력된 음성 신호의 음성 인식 결과 출력된 음소가 정답인 경우 정답 인식 결과라고 하고, 음성 인식 결과 출력된 음소가 오류인 경우 오류 인식 결과라고 한다.
- [0024] 예컨대, 음소 M이 입력되었을 때, 음소 M으로 인식한 경우, 정답 인식 결과라고 하며, 음소 M을 정답 음소라고 칭한다. 또한, 음소 M이 입력되었을 때, 음소 N으로 인식한 경우, 오류 인식 결과라고 하며, 음소 N을 오류 음소라고 칭한다.
- [0025] 도 1c를 참조하면, 음소 M에 대한 음성 인식 결과는 정답 인식 결과가 출력되었다고 가정한다. 따라서 음성 인식 결과가 정답인 음소 M의 경우, 음성 인식 결과를 음향 모델의 음소 모델 파라미터 분포에 반영시켜 업데이트 하면, 현재 음향 모델의 음소 M에 대한 음소 모델 파라미터 분포(50)는 음소 M에 대한 이상적인 음소 모델 파라미터 분포(70)로 이동할 것이다. 이에 따라, 음소 M의 음소 N에 대한 상호 정보량(90)이 줄어든다. 즉, 상호 정보량(90)이 최소화된다. 도 1c와 같은 경우에는 상호 정보량이 전혀 없는 이상적인 형태로 최소화되었다.
- [0026] 한편, 도 1d를 참조하면, 음소 M에 대한 음성 인식 결과가 오류 인식 결과가 출력되었다고 가정한다. 예컨대, 음소 M에 대한 음성 인식 결과가 N으로 출력되었다고 가정한다. 이는 도면 부호 90이 나타내는 상호 정보량에 의한 오류 인식 결과이다. 이러한 경우, 종래에는 오류 인식 결과는 무시되었기 때문에 음향 모델의 업데이트는 없었다. 하지만, 본 발명의 실시예에 따르면, 음성 인식 결과가 오류인 음소 N의 로그우도를 산출하고, 음성 인식 결과를 음향 모델의 음소 모델 파라미터 분포에 반영시켜 업데이트한다. 이에 따라, 현재 음향 모델의 음소 N에 대한 음소 모델 파라미터 분포(60)는 음소 N에 대한 이상적인 음소 모델 파라미터 분포(80)로 이동할 것이다. 이는 상호 정보량에 의해 오류가 나타나지 않도록 오류 인식 결과로 나타난 음소 N의 음소 모델 파라미터 분포(60)를 음소 M과의 상호 정보량이 줄어들도록 이동시키는 것이다. 이에 따라, 음소 N의 음소 M에 대한 상호 정보량(90)이 도면 부호 93에 의해 지시되는 바와 같이 줄어든다. 즉, 상호 정보량이 최소화된다.
- [0027] 이때, 음소 N의 음소 모델 파라미터 분포(60)는 로그우도의 크기에 따라 그 이동의 정도를 달리한다. 다른 말로, 음소 N의 음소 모델 파라미터 분포(60)가 이동하는 정도는 로그우도의 값에 따른 가중치에 따른다. 상술한 바와 같이, 본 발명의 실시예에 따른 상호 정보량 최소화는 서로 다른 음소의 음소 모델 파라미터 분포에서 중첩되는 확률 분포 부분(90)을 줄이는 것을 의미한다.
- [0028] 상술한 바와 같은 방법에 따라, 음향 모델에서 음소 M과 음소 N의 상호 정보량(90)을 최소화시킬 수 있다. 특히, 본 발명에 따르면, 정답(correct) 인식 결과의 음소들에 대한 모델 파라미터 분포를 업데이트하고, 추가

로, 오류(incorrect) 인식 결과의 음소들에 대한 모델 파라미터 분포를 업데이트함으로써, 변별 학습의 성능을 향상시킬 수 있다.

- [0029] 도 2는 본 발명의 실시예에 따른 음향 모델 학습을 위한 장치를 포함하는 음성 인식 시스템을 설명하기 위한 도면이다.
- [0030] 도 2를 참조하면, 음성 인식 시스템은 특징 추출부(100), 탐색부(200), 음향 모델 데이터베이스(300), 발음 사전 데이터베이스(400), 언어 모델 데이터베이스(500), 음성 데이터베이스(600), 음향 모델 학습부(700), 텍스트 데이터베이스(800) 및 언어 모델 학습부(900)를 포함한다.
- [0031] 특징 추출부(100)는 입력된 음성 신호로부터 음성 신호의 특징을 추출하기 위한 것이다. 여기서, 음성 신호는 음성 입력 장치나 파일을 통해 입력될 수 있다. 특징 추출부(100)는 입력된 음성 신호에서 잡음을 제거하거나 음성 인식 성능을 높이기 위한 신호 처리를 수행 한다. 그런 다음, 특징 추출부(100)는 신호 처리된 음성 구간의 음성 신호에서 특징 벡터를 추출하여, 탐색부(200)에 제공한다.
- [0032] 탐색부(200)는 음향 모델, 언어 모델과 발음 사전을 통해 탐색 공간을 형성하고, 형성된 탐색공간과 입력된 음성으로부터 특징 추출부(100)가 구한 특징 벡터를 사용하여 음성 인식을 수행한다.
- [0033] 본 발명의 실시예에서 탐색부(200)는 미리 학습된 모델에 대한 유사도 값을 인식 결과로 출력할 수 있다. 탐색부(200)는 음성 인식을 통해 1-best 인식 결과와 격자(lattice) 형태의 인식결과를 얻을 수 있으며, 격자(lattice) 형태의 인식결과로부터 N-best의 인식 결과를 얻을 수 있다. 이를 위하여 탐색부(200)는 비터비(Viterbi) 알고리즘 또는 DTW(Dynamic Time Warping)와 같은 패턴 정합 알고리즘을 이용할 수 있다. 예컨대, 탐색 공간은 명령어 인식 및 숫자음 인식과 같은 적은 어휘의 인식을 위한 FSN(Finite state network) 형태의 탐색 공간과 대어휘 인식과 빠른 인식을 위한 트리(tree) 형태의 탐색 공간을 포함할 수 있다.
- [0034] 음향 모델 데이터베이스(300)는 음향 모델을 저장한다. 여기서, 음향 모델은 음향 모델은 시간적으로 변화하는 음성신호의 특징을 모델링한다. 음향 모델링 방법은 HMM, Continuous HMM, 신경회로망(NN) 등을 예시할 수 있다. 본 발명의 실시예에 따른 음향 모델 데이터베이스는 각 음소 별로 음소 모델 파라미터 분포를 저장할 수 있다.
- [0035] 발음 사전 데이터베이스(400)는 발음 사전을 저장한다. 발음 사전은 음성에 대한 발음을 저장한다. 발음 사전은 음향 모델과 연결하여 특정 음성에 대한 다중의 발음들을 저장한다.
- [0036] 언어 모델 데이터베이스(500) 언어 모델은 단어간의 문법을 고려하여 인식 후보에 가중치를 줌으로써 문법에 맞는 문장이 더 높은 점수를 얻도록 함으로써 인식률을 향상시킨다. 최적의 인식 단어열을 찾기 위한 탐색에서는 비교하여야 할 후보의 개수를 줄이는 역할도 하게 된다. 인식되는 대상 어휘의 수와 인식 속도, 인식 성능을 고려하여 언어 모델을 선택할 수 있다.
- [0037] 상술한 음향 모델 데이터베이스(300)의 음향 모델, 발음 사전 데이터베이스(400)의 발음 사전, 및 언어 모델 데이터베이스(500)의 언어 모델을 이용하여 음성 인식에 필요한 탐색 공간을 형성한다.
- [0038] 음성 데이터베이스(600)는 학습을 위한 음성 및 그 음성에 대한 텍스트(전사 데이터)를 저장할 수 있다. 이때, 그 음성에 대한 텍스트는 생략될 수 있다. 음향 모델 학습부(700)는 음성 데이터베이스(600)에 저장된 음성을 통해 음성 인식 결과의 음소들이 정답인지 혹은 오류인지 여부를 확인할 수 있다.
- [0039] 음향 모델 학습부(700)는 탐색부(200)로부터 음성 인식 결과를 수신하여, 음성 데이터베이스(600)에 저장된 음성과 비교하여 음성 인식 결과의 음소들이 정답인 인식 결과를 가지는지 혹은 오류인 인식 결과를 가지는지 여부를 판단할 수 있다. 예컨대, 음소 M이 입력되었을 때, M으로 인식한 경우, 그 음소 M을 정답 인식 결과로 판단하며, 정답 음소라고 한다. 그리고 음소 M이 입력되었을 때, N으로 인식한 경우, 그 음소 N을 오류 인식 결과로 판단하며, 오류 음소라고 한다. 오류 인식 결과인 경우, 도 1d에 도시된 바와 같이, 음향 모델 학습부(700)는 오류 음소의 로그우도를 가중치로 적용하여, 음향 모델 데이터베이스(300)의 오류 음소에 대한 음소 모델 파라미터 분포를 업데이트하여 상호 정보량을 최소화한다. 이때, 로그우도는 최대 로그우도(ML, maximum likelihood) 추정 방법을 이용하여 산출된다. 반면, 정답 인식 결과의 음소의 경우, 도 1c에 도시된 바와 같이, 음향 모델 학습부(700)는 음향 모델 데이터베이스(300)의 정답 음소에 대한 음소 모델 파라미터 분포를 업데이트하여, 정답 음소의 음소 모델 파라미터 분포의 상호 정보량을 최소화한다. 이때, 상기 정답 음소에 대한 음향 모델의 음소 모델 파라미터 분포를 업데이트하는 것은 최대 상호정보량(MMI, maximum mutual information) 추정 방법을 통해 수행할 수 있다.

- [0040] 텍스트 데이터베이스(800)는 언어 모델을 생성하기 위한 텍스트들을 저장한다.
- [0041] 언어 모델 학습부(900)는 텍스트 데이터베이스(800)에 저장된 텍스트들을 통해 언어 모델을 생성 혹은 업데이트한다.
- [0042] 도 3은 본 발명의 실시예에 따른 음향 모델 학습부의 내부 구성을 설명하기 위한 도면이다.
- [0043] 도 3을 참조하면, 음향 모델 학습부(700)는 인식 결과 분류 모듈(710), 오류 음소 처리 모듈(720) 및 정답 음소 처리 모듈(730)을 포함한다.
- [0044] 인식 결과 분류 모듈(710)은 탐색부(200)로부터 음성 인식 결과를 수신하면, 음성 데이터베이스(600)에 저장된 음성과 비교하여, 음성 인식 결과의 음소들이 정답 인식 결과를 가지는지 혹은 오류 인식 결과를 가지는지 여부를 구분한다. 그리고 인식 결과 분류 모듈(710)은 정답 인식 결과를 정답 음소 처리 모듈(730)에 제공하고, 오류 인식 결과를 오류 음소 처리 모듈(720)에 제공한다.
- [0045] 오류 음소 처리 모듈(720)은 인식 결과 분류 모듈(710)로부터 오류 인식 결과를 입력 받는다. 그러면, 오류 음소 처리 모듈(720)은 음향 모델 데이터베이스(300)에 저장된 음향 모델에서, 오류 음소의 음소 모델 파라미터 분포를 업데이트한다. 이때, 오류 음소 처리 모듈(720)은 상기 오류 음소의 로그우도를 가중치로 적용하여, 음향 모델의 음소 모델 파라미터 분포를 업데이트한다. 이때, 로그우도는 최대 로그우도(ML) 추정 방법을 이용하여 산출된다. 이에 따라, 도 1d에서 설명된 바와 같이, 입력된 음소 M이 아니라, 오류 음소인 음소 N의 음소 모델 파라미터 분포(60)가 음소 M의 음소 모델 파라미터 분포(50)와의 상호 정보량이 줄어드는 방향으로 이동할 것이다. 이때, 음소 N에 대한 음소 모델 파라미터 분포(60)는 음소 N에 대한 이상적인 음소 모델 파라미터 분포(80)로 이동할 것이다. 또한, 음소 N에 대한 음소 모델 파라미터 분포(60)는 로그우도의 크기에 비례하여 이동되는 정도가 결정된다.
- [0046] 정답 음소 처리 모듈(730)은 인식 결과 분류 모듈(710)로부터 정답 인식 결과를 입력 받는다. 그러면, 정답 음소 처리 모듈(730)은 정답 인식 결과를 반영하여 음향 모델 데이터베이스(300)에 저장된 음향 모델에서, 정답 음소의 음소 모델 파라미터 분포를 업데이트한다. 이때, 상기 정답 음소에 대한 음향 모델의 음소 모델 파라미터 분포를 업데이트하는 것은 최대 상호정보량(MMI) 추정 방법을 통해 수행할 수 있다. 예컨대, 도 1c에서 설명된 바와 같이, 정답 인식 결과를 음향 모델의 음소 모델 파라미터 분포에 반영시켜 업데이트하면, 현재 음향 모델의 음소 M에 대한 음소 모델 파라미터 분포(50)는 음소 M에 대한 이상적인 음소 모델 파라미터 분포(70)로 이동할 것이다.
- [0047] 도 4는 본 발명의 실시예에 따른 음향 모델 학습을 위한 방법을 설명하기 위한 흐름도이다.
- [0048] 도 4를 참조하면, 특징 추출부(100)는 S410 단계에서 음성 신호가 입력되면 음성 신호의 특징 벡터를 추출하고, S420 단계에서 추출된 음성의 특징 벡터를 탐색부(200)에 제공한다.
- [0049] 탐색부(200)는 S430 단계에서 입력된 음성 벡터에 대해 음향 모델, 발음 사전 및 언어 모델을 기초로 형성된 탐색 공간에서 음성 인식을 수행한다. 이러한 음성 인식은 음소 단위로 이루어진다. 또한, 음성 인식의 결과는 1-best 또는 N-best의 인식 결과가 될 수 있으나, N-best의 인식 결과가 바람직하다. 음소 단위의 음성 인식을 수행한 후, 탐색부(200)는 S440 단계에서 음소 단위의 음성 인식 결과를 음향 모델 학습부(700)에 제공한다.
- [0050] 음향 모델 학습부(700)는 음소 단위의 음성 인식 결과를 수신하여, S450 단계에서 음성 인식 결과의 음소들이 정답 인식 결과를 가지는지 혹은 오류 인식 결과를 가지는지 여부를 구분한다. 이때, 음향 모델 학습부(700)는 음성 데이터베이스(600)에 미리 저장된 음성을 이용한다.
- [0051] 음향 모델 학습부(700)는 S460 단계에서 오류 인식 결과와 정답 인식 결과에 따라 음향 모델을 업데이트한다. 도 1d를 참조하면, 이러한 S460 단계에서 음향 모델 학습부(700)는 오류 인식 결과로 구분된 오류 음소의 음향 모델의 음소 모델 파라미터 분포를 업데이트한다. 이때, 음향 모델 학습부(700)는 오류 음소의 로그우도(likelihood)를 가중치로 적용하여, 음향 모델의 음소 모델 파라미터 분포를 업데이트한다. 이때, 로그우도는 최대 로그우도(ML) 추정 방법을 이용하여 산출될 수 있다. 이와 동시에, 도 1c를 참조하면, S460 단계에서 음향 모델 학습부(700)는 정답 인식 결과를 반영하여 정답 음소에 대한 음향 모델의 음소 모델 파라미터 분포를 업데이트한다. 이때, 정답 음소에 대한 음향 모델의 음소 모델 파라미터 분포를 업데이트하는 것은 최대 상호정보량(MMI) 추정 방법을 통해 수행할 수 있다.
- [0052] 상술한 바와 같이 본 발명의 실시예에 따르면, 변별학습을 이용하여, 정답 인식 결과의 음소들에 대한 모델 파라미터를 업데이트하고, 추가로, 오류 인식 결과의 음소들에 대한 모델 파라미터 분포를 반영함으로써, 음소 단

위의 모델 파라미터 분포의 상호 정보량을 최소화할 수 있다.

- [0053] 도 5는 본 발명의 실시예에 따른 음향 모델 학습부의 음향 모델 학습 방법을 설명하기 위한 흐름도이다.
- [0054] 인식 결과 분류 모듈(710)은 S510 단계에서 탐색부(200)로부터 음성 인식 결과를 수신한다. 탐색부(200)는 음소 단위의 음성 인식을 수행하며, 인식 결과 또한 음소 단위로 제공될 수 있다. 따라서 인식 결과 분류 모듈(710)은 음소 단위의 음성 인식 결과를 수신한다. 이러한 음성 인식 결과는 N-best 인식 결과로 제공되는 것이 바람직하다. 예컨대, N-best 인식 결과는 앞서 설명된 <표 1>과 같이 출력될 수 있다.
- [0055] 음성 인식 결과가 수신되면, 인식 결과 분류 모듈(710)은 S520 단계에서 음성 인식 결과의 음소들이 정답 인식 결과를 가지는지 혹은 오류 인식 결과를 가지는지 여부를 구분한다. 이를 위하여, 인식 결과 분류 모듈(710)은 음성 데이터베이스(600)에 미리 저장된 음성과 비교하여, 각 음소가 정답 혹은 오류 인식 결과를 가지는지 여부를 판별할 수 있다.
- [0056] 오류 음소 처리 모듈(720)은 S530 단계에서 오류 인식 결과로 구분된 오류 음소에 대한 로그우도를 도출한다. 이때, 로그우도는 최대 로그우도(ML, maximum likelihood) 추정 방법을 이용하여 산출된다.
- [0057] 그런 다음, 오류 음소 처리 모듈(720)은 S540 단계에서 음향 모델 데이터베이스(300)의 오류 음소에 대한 음소 모델 파라미터 분포를 업데이트한다. 하여, 상호 정보량을 최소화한다. 이때, 오류 음소 처리 모듈(720)은 상기 오류 음소의 로그우도를 가중치로 적용하여, 오류 음소에 대한 음소 모델 파라미터 분포를 업데이트한다. 이에 따라, 오류 음소에 대한 음소 모델 파라미터 분포의 상호 정보량이 최소화된다.
- [0058] 예컨대, 도 1d에서 설명된 바와 같이, 음소 M이 입력되었지만 음소 N으로 인식된 경우, 음소 M과 음소 N의 상호 정보량에 기인한 오류일 수 있다. 따라서 본 발명은 오류 음소인 음소 N의 음소 모델 파라미터 분포(60)를 이동시켜, 음소 M과의 상호 정보량을 최소화한다. 이때, 로그우도의 크기에 따라 음소 N의 음소 모델 파라미터 분포(60)는 음소 N의 이상적인 음소 모델 파라미터 분포(80)로 이동할 것이다. 이에 따라, 음소 M의 음소 N에 대한 상호 정보량이 줄어든다. 즉, 상호 정보량이 최소화된다.
- [0059] 정답 음소 처리 모듈(730)은 S550 단계에서 인식 결과 분류 모듈(710)로부터 정답 인식 결과를 입력 받아, 음향 모델 데이터베이스(300)의 정답 음소에 대한 음소 모델 파라미터 분포를 업데이트한다. 이때, 정답 음소 처리 모듈(730)은 해당 정답 음소의 음소 모델 파라미터 분포의 상호 정보량을 최소화한다.
- [0060] 이때, 상기 정답 음소에 대한 음향 모델의 음소 모델 파라미터 분포를 업데이트하는 것은 최대 상호 정보량(MMI) 추정 방법을 통해 수행할 수 있다.
- [0061] 예컨대, 도 1c에서 설명된 바와 같이, 정답 인식 결과를 음향 모델의 음소 모델 파라미터 분포에 반영시켜 업데이트하면, 현재 음향 모델의 음소 M에 대한 음소 모델 파라미터 분포(50)는 음소 M에 대한 이상적인 음소 모델 파라미터 분포(70)로 이동할 것이다. 이에 따라, 음소 M의 음소 N에 대한 상호 정보량이 줄어든다. 즉, 상호 정보량이 최소화된다.
- [0062] 상술한 오류 인식 결과 및 정답 인식 결과를 적용하여 음향 모델 데이터베이스(300)의 음소 모델 파라미터 분포를 업데이트하는 것이 순차로 수행되는 것과 같이 기술되었지만, 병렬적으로 수행되는 것이 바람직하다.
- [0063] 상술한 바와 같은, 본 발명의 실시 예에 따른 음향 모델 학습 방법은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있으며, 컴퓨터 시스템에 의해 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 기록매체의 예로는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM(Compact Disk Read Only Memory), DVD(Digital Video Disk)와 같은 광기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-Optical Media) 및 롬(ROM, Read Only Memory), 램(RAM, Random Access Memory), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함한다.
- [0064] 또한, 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다. 그리고, 본 발명을 구현하기 위한 기능적인(functional) 프로그램, 코드 및 코드 세그먼트들은 본 발명이 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있다.
- [0065] 이상과 같이, 본 명세서와 도면에는 본 발명의 바람직한 실시예에 대하여 개시하였으나, 여기에 개시된 실시예 외에도 본 발명의 기술적 사상에 바탕을 둔 다른 변형 예들이 실시 가능하다는 것은 본 발명이 속하는 기술 분

야에서 통상의 지식을 가진 자에게 자명한 것이다. 또한, 본 명세서와 도면에서 특정 용어들이 사용되었으나, 이는 단지 본 발명의 기술 내용을 쉽게 설명하고 발명의 이해를 돕기 위한 일반적인 의미에서 사용된 것이지, 본 발명의 범위를 한정하고자 하는 것은 아니다. 따라서, 상술한 상세한 설명은 모든 면에서 제한적으로 해석되어서는 아니 되고 예시적인 것으로 고려되어야 한다. 본 발명의 범위는 첨부된 청구항의 합리적 해석에 의해 선정되어야 하고, 본 발명의 등가적 범위 내에서의 모든 변경은 본 발명의 범위에 포함된다.

**산업상 이용가능성**

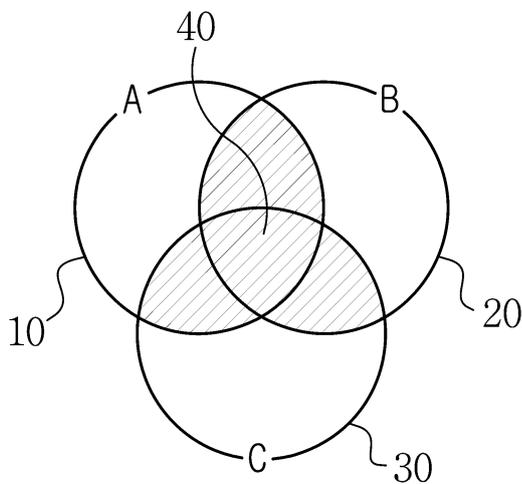
[0066] 본 발명은 음향 모델 학습을 위한 장치 및 이를 위한 방법이 기록된 컴퓨터 판독 가능한 기록매체에 관한 것이다. 이러한 본 발명은 단어열을 음소 단위로 음성 인식한 결과로부터 상기 단어열을 구성하는 음소들을 정답 인식 결과 및 오류 인식 결과로 구분하고, 상기 오류 인식 결과로 구분된 오류 음소에 대한 음향 모델의 음소 모델 파라미터 분포에 상기 오류 인식 결과를 반영하여 업데이트하되, 상기 오류 음소의 로그우도를 가중치로 적용하며, 상기 정답 인식 결과로 구분된 정답 음소에 대한 음향 모델의 음소 모델 파라미터 분포에 상기 정답 인식 결과를 반영하여 업데이트한다. 상술한 바와 같이 본 발명은 음향 모델의 음소 모델 파라미터 분포를 업데이트할 때, 오류 인식 결과로 구분된 음소의 로그우도를 가중치로 적용하여, 오류 인식 결과도 상호 정보량을 최소화하도록 반영함으로써, 변별 학습에서 효율적으로 상호 정보량을 최소화시킬 수 있다. 이러한 본 발명은 시판 또는 영업의 가능성이 충분할 뿐만 아니라 현실적으로 명백하게 반복 실시할 수 있는 정도이므로 산업상 이용가능성이 있다.

**부호의 설명**

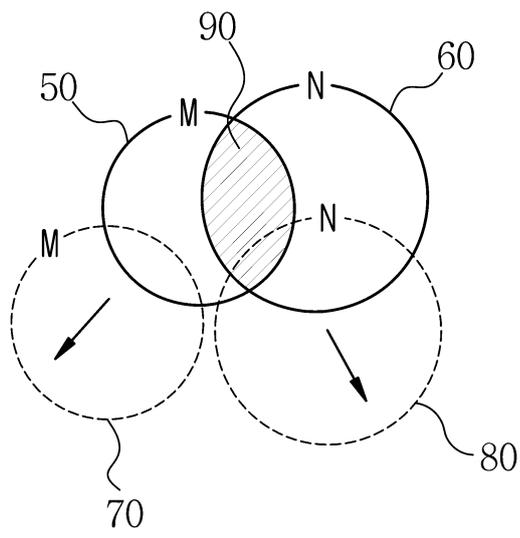
- |        |                   |                   |
|--------|-------------------|-------------------|
| [0067] | 100: 특징 추출부       | 200: 탐색부          |
|        | 300: 음향 모델 데이터베이스 | 400: 발음 사전 데이터베이스 |
|        | 500: 언어 모델 데이터베이스 | 600: 음성 데이터베이스    |
|        | 700: 음향 모델 학습부    | 710: 인식 결과 분류 모듈  |
|        | 720: 오류 음소 처리 모듈  | 730: 정답 음소 처리 모듈  |
|        | 800: 텍스트 데이터베이스   | 900: 언어 모델 학습부    |

**도면**

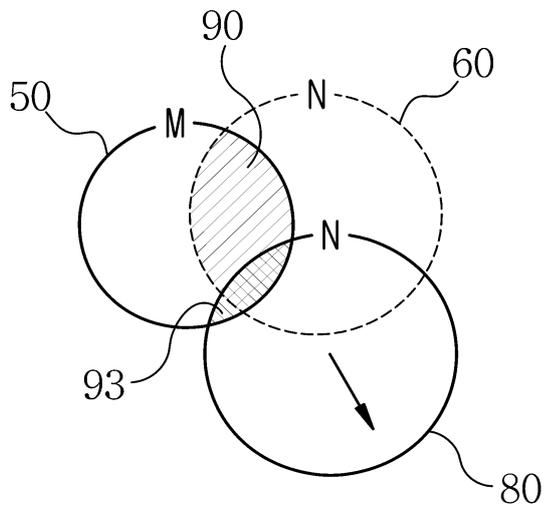
**도면1a**



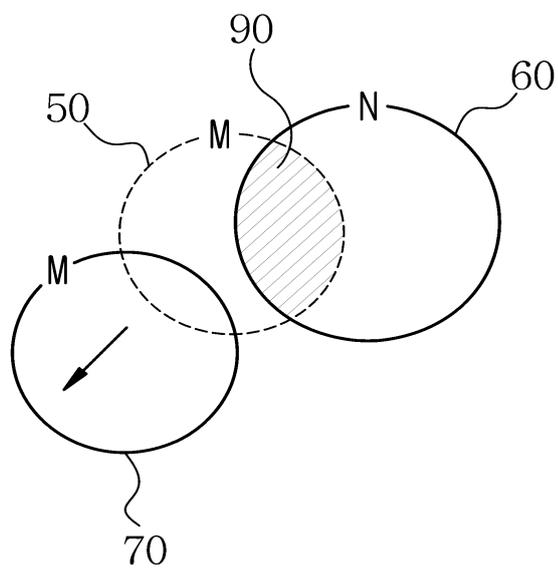
도면1b



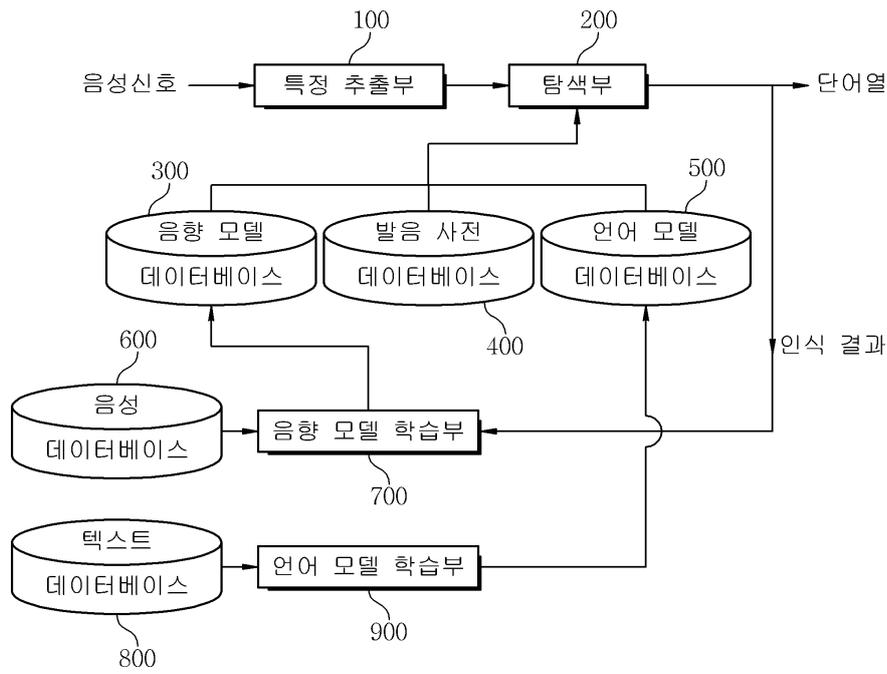
도면1c



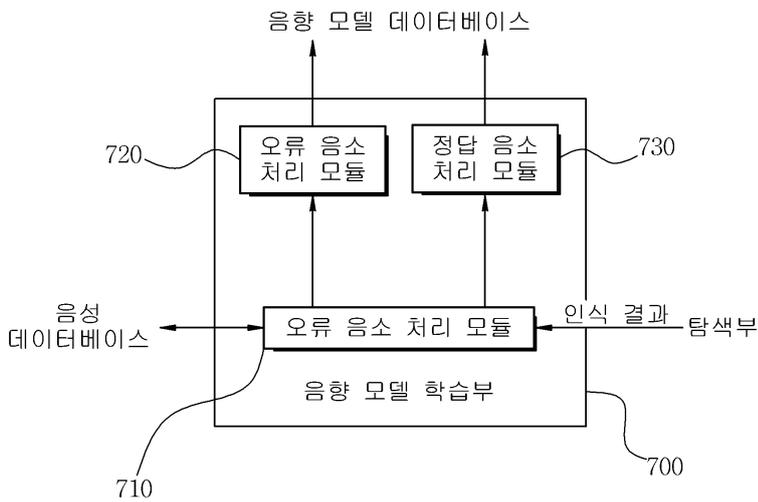
도면1d



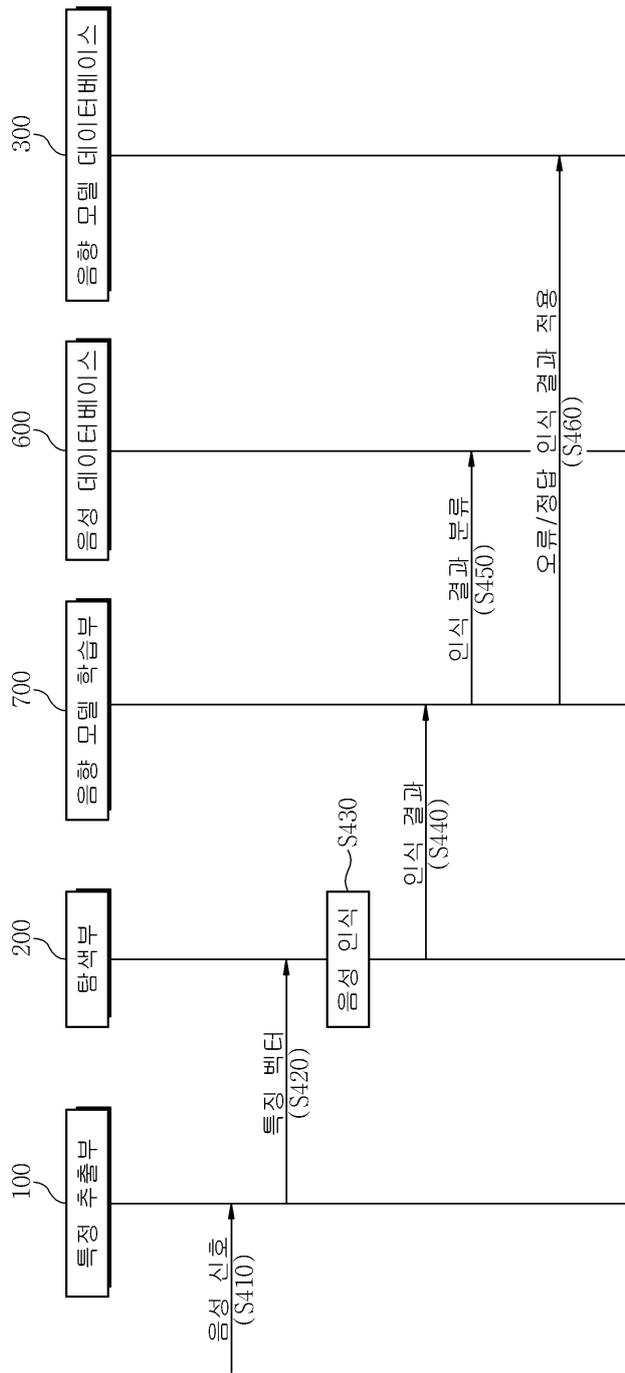
도면2



도면3



도면4



도면5

