



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0077780
(43) 공개일자 2014년06월24일

(51) 국제특허분류(Int. Cl.)
G10L 15/14 (2006.01)

(21) 출원번호 10-2012-0146911

(22) 출원일자 2012년12월14일

심사청구일자 없음

(71) 출원인

한국전자통신연구원

대전광역시 유성구 가정로 218 (가정동)

(72) 발명자

정훈

강원 홍천군 홍천읍 연봉중앙로 18, 103동 1101호
(현대아파트)

전형배

대전 서구 청사로 65, 109동 605호 (월평동, 황실
타운)

(뒷면에 계속)

(74) 대리인

한양특허법인

전체 청구항 수 : 총 1 항

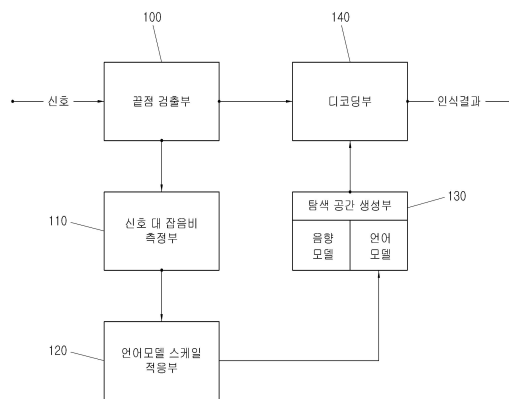
(54) 발명의 명칭 **신호 대 잡음비를 이용한 언어 모델 스케일 적응 장치**

(57) 요약

본 발명은 음성 인식 시스템에 관한 것으로서, 더 상세하게는 음성 인식 시스템에서 음성 인식 성능을 높이기 위한 언어 모델 스케일 적응 장치에 관한 것이다.

본 발명에 따르면, 신호 대 잡음비가 낮은 음성 신호에 대해서는 언어 모델의 변별력에 가중치를 뒀으로써 잡음 환경에 대한 인식 성능을 개선하게 된다.

대표도 - 도1



(72) 발명자

박진규

대전 유성구 대덕대로541번길 68, 103동 205호 (도룡동, 현대아파트)

오유리

대전 유성구 가정로 270, 한국전자통신연구원기숙사 신관 339 (가정동)

강점자

대전 서구 둔산로 155, 119동 407호 (둔산동, 크로바아파트)

이윤근

대전 서구 청사서로 11, 103동 1406호 (월평동, 무지개아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호 10035252

부처명 지식경제부

연구사업명 산업원천기술개발사업(정보통신)

연구과제명 모바일 플랫폼 기반 대화모델 적용 자연어 음성 인터페이스 기술 개발

기여율 1/1

주관기관 한국전자통신연구원

연구기간 2010.03.01 ~ 2015.02.28

특허청구의 범위

청구항 1

음성 인식 방식의 신호 대 잡음비를 이용한 언어 모델 스케일 적응 장치에 있어서,

상기 신호대 잡음비에 기반하여 음향 모델의 확률값에 다른 가중치를 주어 언어 모델 스케일을 조정하는 것을 특징으로 신호 대 잡음비를 이용한 언어 모델 스케일 적응 장치.

명세서

기술분야

[0001] 본 발명은 음성 인식 시스템에 관한 것으로서, 더 상세하게는 음성 인식 시스템에서 음성 인식 성능을 높이기 위한 언어 모델 스케일 적응 장치에 관한 것이다.

배경기술

[0002] 음성인식 기술이 비교적 보편화되어 여러 응용분야에서 활용되고 있다. 그런데, 고립단어 수준의 음성인식 기술이 상용화되다 보니, 사용자 입장에서도 보다 높은 기능을 갖는 음성인식 제품에 대한 요구가 늘고 있다.

[0003] 즉, 인식대상 단어 전후에 다른 말을 포함시켜도 인식이 가능한 키워드 스포팅(key word spotting) 기술이나 자연스러운 문장 형태의 인식이 가능한 연속 음성인식 기술이 요구되고 있다.

[0004] 그러나, 연속 음성인식의 경우 아직 사용자의 기대수준에 도달하지 못하고 있으며, 이를 잘 구현하기 위해서는 사람이 이용하는 언어정보를 잘 활용하여야 한다.

[0005] 즉, 음향모델의 성능 이외에 얼마나 좋은 언어모델을 적용할 수 있는가 하는 문제가 남게 된다.

[0006] 대부분의 경우 언어모델은 텍스트 데이터를 활용하여 구축하게 되는데, 여러 가지 텍스트 데이터를 구할 수 있는 텍스트 코퍼스(corpus)를 활용하여 구축하게 된다.

[0007] 예를 들어, 구술(Dictation)과 같이 범용성을 띠는 경우, 신문 기사, 소설, 기타 인터넷 상에서 얻을 수 있는 자료들을 활용하게 된다. 하지만, 이 경우 자료를 이용하여 만든 언어 모델의 성능은 그 한계가 있다.

[0008] 특히, 어떤 특정 응용분야의 경우 충분한 언어모델이 확보되지 않을 경우, 사용자가 기대하는 성능을 얻기 어렵다.

[0009] 가장 이상적인 방법은 그 응용 분야에 적합한 발성 내용의 텍스트 자료를 구하는 것이나, 이는 현실적으로 어려움이 많다.

[0010] 이러한 문제점들을 극복하기 위한 노력이 여러 방식으로 진행되어왔다. 이중 언어모델 적응도 이러한 노력의 하나로 볼 수 있다.

[0011] 그러나, 음향 모델과 언어 모델은 모델링 방법의 차이로 인해 서로 다른 범위의 확률값을 가지게 되고 이 차이를 보정하는 역할을 언어 모델 스케일이 하게 된다.

[0012] 일반적으로 최적의 언어 모델 스케일은 실험을 통해 구하게 되는데 주어진 평가 코퍼스와 시스템에 대해 속도 대 성능이 최적인 값을 사용하게 된다.

[0013] 일반적으로 신호 대 잡음 비가 좋은 경우에는 음향 모델간의 변별력이 좋아지나 신호대 잡음비가 나쁜 경우에는 음향 모델간의 변별력이 나빠지게 된다.

[0014] 그러나, 언어 모델의 확률값 혹은 변별력은 입력 신호의 품질에 무관하게 유지된다는 문제점이 있다.

선행기술문헌

특허문헌

[0015] (특허문헌 0001) 1. 한국공개특허 제10-2012-0066530호

발명의 내용

해결하려는 과제

[0016] 본 발명은 배경기술에서 기술된 문제점을 해소하기 위해 제안된 것으로서, 잡음 환경에서도 안정적인 인식 성능을 유지하기 위해 입력되는 신호의 잡음 정도에 따라 언어 모델 스케일을 조정한다.

[0017] 일반적으로 신호 대 잡음 비가 좋은 경우에는 음향 모델간의 변별력이 좋아지나 신호대 잡음비가 나쁜 경우에는 음향 모델간의 변별력이 나빠지게 된다.

[0018] 그러나, 언어 모델의 확률값 혹은 변별력은 입력 신호의 품질에 무관하게 유지된다.

[0019] 따라서, 신호 대 잡음비가 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 두고 그렇지 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 뒀으로써 잡음 환경에서는 언어 모델의 변별력을 좀더 사용하도록 언어 모델 스케일을 조정함으로써 잡음 환경에서 인식 성능을 개선하는 신호 대 잡음비를 이용한 언어 모델 스케일 적응 장치를 제공하는 데 그 목적이 있다.

과제의 해결 수단

[0020] 본 발명은 배경기술에서 제기된 문제점을 해소하기 위해, 신호 대 잡음비가 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 두고 그렇지 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 뒀으로써 잡음 환경에서는 언어 모델의 변별력을 좀더 사용하도록 언어 모델 스케일을 조정함으로써 잡음 환경에서 인식 성능을 개선하는 신호 대 잡음비를 이용한 언어 모델 스케일 적응 장치를 제공한다.

[0021] 상기 언어 모델 스케일 적응 장치는, 음성 인식 방식의 신호 대 잡음비를 이용한 언어 모델 스케일 적응 장치에 있어서, 상기 신호대 잡음비에 기반하여 음향 모델의 확률값에 다른 가중치를 주어 언어 모델 스케일을 조정하는 것을 특징으로 할 수 있다.

[0022] 다른 한편으로, 본 발명의 다른 일실시예는, 음성 신호가 입력되는 음성 신호 입력 단계; 입력된 음성 신호의 끝점을 검출하는 끝점 검출 단계; 끝점이 검출됨에 따라 음성 신호에 대한 신호 대 잡음비(SNR: Signal to Noise Ratio)를 측정하는 신호 대 잡음비 측정 단계; 측정된 신호 대 잡음비에 따라 신호 대 잡음비가 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 두고 그렇지 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 주어 언어 모델 스케일을 적응하는 언어 모델 스케일 적응 단계; 언어 모델 스케일이 적용됨에 따라 상기 음성 신호에 대한 탐색 공간을 생성하는 탐색 공간 생성 단계; 및 탐색 공간된 신호를 디코딩하여 최종 음성 인식 결과를 생성하는 디코딩 단계를 포함하는 신호 대 잡음비를 이용한 언어 모델 스케일 적응 방법을 제공한다.

발명의 효과

[0023] 본 발명에 따르면, 신호 대 잡음비가 낮은 음성 신호에 대해서는 언어 모델의 변별력에 가중치를 뒀으로써 잡음 환경에 대한 인식 성능을 개선하게 된다.

[0024] 즉, 신호 대 잡음비가 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 두고 그렇지 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 뒀으로써 잡음 환경에서는 언어 모델의 변별력을 좀더 사용하도록 언어 모델 스케일을 조정함으로써 잡음 환경에서 인식 성능을 개선할 수 있다.

도면의 간단한 설명

[0025] 도 1은 본 발명의 일실시예에 따른 신호 대 잡음비를 이용한 언어 모델 스케일 적응 장치의 구성도를 보여주는 도면이다.

도 2는 본 발명의 일실시예에 따른 신호 대 잡음비를 이용한 언어 모델 스케일 적응 처리 과정을 보여주는 순서도이다.

발명을 실시하기 위한 구체적인 내용

- [0026] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는바, 특정 실시예들을 도면에 예시하고 상세한 설명에 구체적으로 설명하고자 한다. 그러나 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0027] 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용한다.
- [0028] 제 1, 제 2등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.
- [0029] 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제 1 구성요소는 제 2 구성요소로 명명될 수 있고, 유사하게 제 2 구성요소도 제 1 구성요소로 명명될 수 있다. "및/또는" 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [0030] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미가 있다.
- [0031] 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않아야 한다.
- [0032] 이하 첨부된 도면을 참조하여 본 발명의 일실시예에 따른 신호 대 잡음비를 이용한 언어 모델 스케일 적응 장치에 대하여 상세하게 설명하기로 한다.

[0033] 도 1은 본 발명의 일실시예에 따른 신호 대 잡음비를 이용한 언어 모델 스케일 적응 장치의 구성도를 보여주는 도면이다. 도 1을 참조하면, 상기 언어 모델 스케일 적응 장치는, 입력된 음성 신호의 끝점을 검출하는 끝점 검출부(100)와, 끝점이 검출됨에 따라 음성 신호에 대한 신호 대 잡음비(SNR: Signal to Noise Ratio)를 측정하는 신호 대 잡음비 측정부(110)와, 측정된 신호 대 잡음비에 따라 신호 대 잡음비가 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 두고 그렇지 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 주어 언어 모델 스케일을 적응하는 언어 모델 스케일 적응부(120)와, 언어 모델 스케일 적응부(120)의 언어 모델 스케일 적응에 따라 탐색 공간을 생성하는 탐색 공간 생성부(130)와, 탐색 공간된 신호를 디코딩하여 최종 음성 인식 결과를 생성하는 디코딩부(140) 등으로 구성된다.

[0034] 일반적으로, 확률 기반의 음성 인식 시스템은 수학적 식 1과 같이 입력된 음성 신호 X 에 대해 최대 우도 확률값(maximum likelihood a posteriori probability; ML-APP)을 가지는 단어열 W 를 다음식을 이용하여 구하게 된다.

수학적 식 1

$$\bar{W} = \operatorname{argmax} P(X|W)P(W)^\alpha$$

- [0035]
- [0036] 이때, $P(X|W)$ 를 음향 모델, $P(W)$ 언어 모델, α 를 언어 모델 스케일이라 한다.
- [0037] 음향 모델은 각 단어나 음소가 특정 음성 신호를 발생할 확률이고 언어 모델은 연속된 단어에 대한 발생 확률이 된다.
- [0038] 음향 모델과 언어 모델은 모델링 방법의 차이로 인해 서로 다른 범위의 확률값을 가지게 되고 이 차이를 보정하는 역할을 언어 모델 스케일이 하게 된다.
- [0039] 본 발명의 일실시예에서는 신호대 잡음비에 기반한 언어 모델 스케일 적응 방식이 사용되며, 이를 수학적식으로 표현하면 다음식과 같다. 수학적 식 2와 같이 언어 모델 스케일을 시간 t 와 신호 대 잡음 비의 함수가 되도록 한다.

수학식 2

$$\overline{W} = \operatorname{argmax} P(X|W)P(W)^{\alpha(t)}$$

[0040]

수학식 3

$$\alpha(t) = \alpha - \beta \cdot \operatorname{sigmoid}(SNR(t))$$

[0041]

[0042] 여기서, $SNR(t)$ 는 시간 프레임 t 에서의 신호 대 잡음비이고 α 는 실험을 통해 구한 최적 언어 모델 스케일이며 β 는 가중치 인자(weighting factor)로 실험을 통해 구하게 된다. 이때, 시그모이드(sigmoid) 함수는 다음식과 같이 구한다.

수학식 4

$$\operatorname{sigmoid}(x) = \frac{1}{1 + e^{-(x-b)}}$$

[0043]

[0044] 또한, 도 2는 본 발명의 일실시예에 따른 신호 대 잡음비를 이용한 언어 모델 스케일 적용 처리 과정을 보여주는 순서도이다.

[0045] 도 2를 참조하면, 상기 언어 모델 스케일 적용 과정은, 음성 신호가 입력되는 음성 신호 입력 단계(S200)와, 입력된 음성 신호의 끝점을 검출하는 끝점 검출 단계(S210)와, 끝점이 검출됨에 따라 음성 신호에 대한 신호 대 잡음비(SNR: Signal to Noise Ratio)를 측정하는 신호 대 잡음비 측정 단계(S220)와, 측정된 신호 대 잡음비에 따라 신호 대 잡음비가 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 두고 그렇지 좋은 경우에는 음향 모델의 확률값에 좀더 가중치를 주어 언어 모델 스케일을 적용하는 언어 모델 스케일 적용 단계(S230)와, 언어 모델 스케일이 적용됨에 따라 상기 음성 신호에 대한 탐색 공간을 생성하는 탐색 공간 생성 단계(S240)와, 탐색 공간된 신호를 디코딩하여 최종 음성 인식 결과를 생성하는 디코딩 단계(S250) 등으로 구성된다.

[0046] 특히, 특히, 본 발명의 일실시예에 따른 신호 대 잡음비를 이용한 언어 모델 스케일 적용 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 코드 형태로 구현되어 컴퓨터 판독 가능 저장 매체에 기록될 수 있다.

[0047] 상기 컴퓨터 판독 가능 저장 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다.

[0048] 상기 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.

[0049] 컴퓨터 판독 가능 저장 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-광매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다.

[0050] 상기 매체는 프로그램 명령, 데이터 구조 등을 지정하는 신호를 전송하는 반송파를 포함하는 광 또는 금속선, 도파관등의 전송 매체일 수도 있다.

[0051] 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 본 발명의 동작을

수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0052] 또한, 본 발명의 일실시예는 하드웨어, 소프트웨어 또는 이들의 조합으로 구현될 수 있다. 하드웨어 구현에 있어, 상술한 기능을 수행하기 위해 디자인된 ASIC(application specific integrated circuit), DSP(digital signal processing), PLD(programmable logic device), FPGA(field programmable gate array), 프로세서, 제어기, 마이크로프로세서, 다른 전자 유닛 또는 이들의 조합으로 구현될 수 있다.

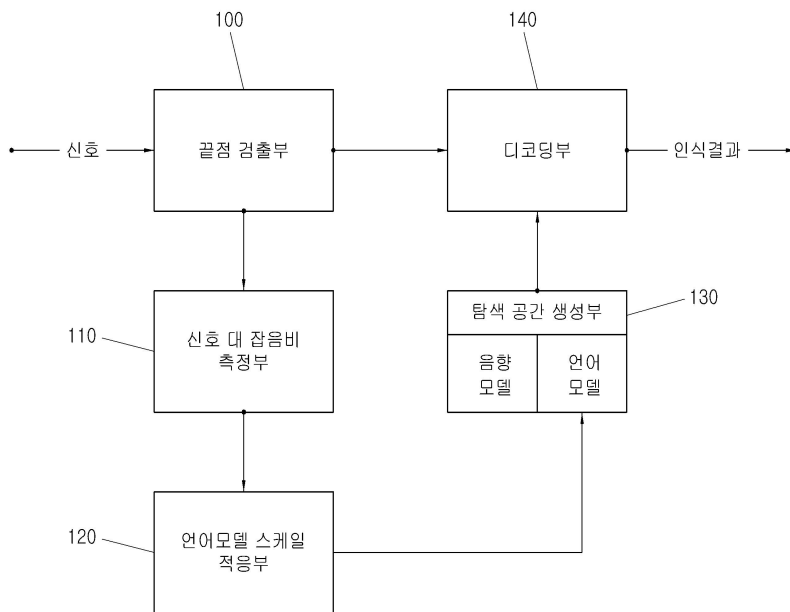
[0053] 소프트웨어적인 구현에 있어, 상술한 기능을 수행하는 모듈로 구현될 수 있다. 소프트웨어는 메모리 유닛에 저장될 수 있고, 프로세서에 의해 실행된다. 메모리 유닛이나 프로세서는 당업자에게 잘 알려진 다양한 수단을 채용할 수 있다.

부호의 설명

- [0054] 100: 끝점 검출기
- 110: 신호대 잡음비 측정부
- 120: 언어모델 스케일 적응부
- 130: 탐색 공간 생성부
- 140: 디코딩부

도면

도면1



도면2

