



(19) 대한민국특허청(KR)

(12) 공개특허공보(A)

(11) 공개번호

10-2014-0028174

(43) 공개일자

2014년03월10일

(51) 국제특허분류(Int. Cl.)

G10L 15/06 (2006.01)

G10L 25/00 (2014.01)

(21) 출원번호

10-2012-0076809

(22) 출원일자

2012년07월13일

심사청구일자

없음

(71) 출원인

삼성전자주식회사

경기도 수원시 영통구 삼성로 129 (매탄동)

고려대학교 산학협력단

서울시 성북구 안암로 145 (안암동)

(72) 발명자

이재원

서울특별시 강남구 선릉로126길 22 롯데캐슬프레미어아파트 102동 1001호

육동석

서울특별시 성북구 안암로 145 고려대학교

(뒷면에 계속)

(74) 대리인

정홍식, 김태현, 이현수

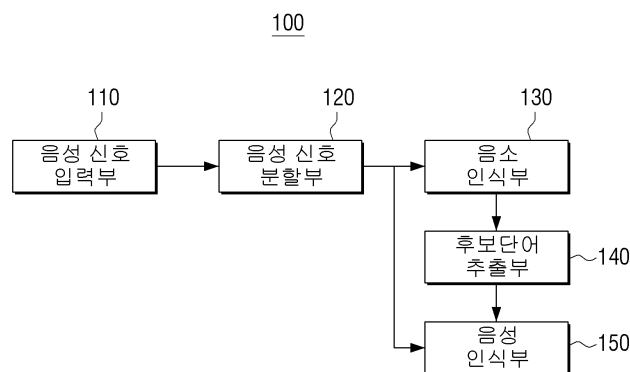
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 음성 인식 방법 및 이를 적용한 전자 장치

(57) 요약

음성 인식 방법 및 이를 적용한 전자 장치가 제공된다. 본 전자 장치의 음성 인식 방법은, 음성 신호가 입력되면, 음성 신호를 기 설정된 시간 간격의 복수의 구간으로 분할하고, 제1 음향 모델을 이용하여 복수의 구간 중 일 구간의 음성 신호에 대한 음소 인식을 수행하며, 음소 인식 결과를 이용하여 일 구간의 음성 신호에 대한 후보 단어를 추출하고, 후보 단어를 이용하여 일 구간의 음성 신호에 대한 음성 인식을 수행한다. 이에 의해, 전자 장치는 사용자가 발화한 음성을 더욱 신속하게 인식할 수 있게 된다.

대표도 - 도1



(72) 발명자

임현택

서울특별시 성북구 안암로 145 고려대학교

김태윤

서울특별시 서초구 잠원로 150 한신아파트 6동
1405호

특허청구의 범위

청구항 1

전자 장치의 음성 인식 방법에 있어서,

음성 신호가 입력되면, 상기 음성 신호를 기 설정된 시간 간격의 복수의 구간으로 분할하는 단계;

제1 음향 모델을 이용하여 상기 복수의 구간 중 일 구간의 음성 신호에 대한 음소 인식을 수행하는 단계;

상기 음소 인식 결과를 이용하여 상기 일 구간의 음성 신호에 대한 후보 단어를 추출하는 단계; 및

상기 후보 단어를 이용하여 상기 일 구간의 음성 신호에 대한 음성 인식을 수행하는 단계;를 포함하는 음성 인식 방법.

청구항 2

제1항에 있어서,

상기 음소 인식을 수행하는 단계는,

분할 비터비(segmented viterbi) 알고리즘을 이용하여 상기 일 구간의 음성 신호에 포함된 복수의 음소 중 끝 부분의 적어도 하나의 음소를 삭제하는 단계;를 더 포함하고,

상기 삭제된 적어도 하나의 음소는,

다음 구간의 음성 신호에 대한 음소 인식을 수행할 때 이용되는 것을 특징으로 하는 음성 인식 방법.

청구항 3

제1항에 있어서,

상기 추출하는 단계는,

상기 인식된 음소와 발음이 유사한 유사 음소를 추출하는 단계; 및

상기 인식된 음소 및 상기 유사 음소를 이용하여 상기 일 구간의 후보 단어를 추출하기 위한 단어 그래프를 생성하는 단계;를 포함하는 것을 특징으로 하는 음성 인식 방법.

청구항 4

제3항에 있어서,

상기 음성 인식을 수행하는 단계는,

제2 음향 모델을 이용하여 상기 일 구간의 음성 신호에 대한 가우시안 확률을 계산하는 단계;

상기 제2 음향 모델 및 언어 모델을 이용하여 상기 단어 그래프 내에 확률이 가장 높은 단어열을 출력하는 단계;를 포함하는 음성 인식 방법.

청구항 5

제4항에 있어서,

상기 제1 음향 모델 및 상기 제2 음향 모델은 서로 상이한 음향 모델인 것을 특징으로 하는 음성 인식 방법.

청구항 6

제1항에 있어서,

상기 음소 인식을 수행하는 단계, 상기 추출하는 단계 및 상기 음성 인식을 수행하는 단계는 각기 상이한 코어에 의해 병렬적으로 수행되는 것을 특징으로 하는 음성 인식 방법.

청구항 7

전자 장치에 있어서,

음성 신호를 입력받는 음성 신호 입력부;

상기 입력부를 통해 입력된 음성 신호를 기 설정된 시간 간격의 복수의 구간으로 분할하는 음성 신호 분할부;

제1 음향 모델을 이용하여 상기 복수의 구간 중 일 구간의 음성 신호에 대한 음소 인식을 수행하는 음소 인식부;

상기 음소 인식 결과를 이용하여 상기 일 구간의 음성 신호에 대한 후보 단어를 추출하는 후보 단어 추출부; 및
상기 후보 단어를 이용하여 상기 일 구간의 음성 신호에 대한 음성 인식을 수행하는 음성 인식부;를 포함하는 전자 장치.

청구항 8

제7항에 있어서,

상기 음소 인식부는,

분할 비터비 알고리즘을 이용하여 상기 일 구간의 음성 신호에 포함된 복수의 음소 중 끝 부분의 적어도 하나의 음소를 삭제하여 음소 인식을 수행하며,

상기 삭제된 적어도 하나의 음소는,

다음 구간의 음성 신호에 대한 음소 인식을 수행할 때 이용되는 것을 특징으로 하는 전자 장치.

청구항 9

제7항에 있어서,

상기 후보 단어 추출부는,

상기 인식된 음소와 발음이 유사한 유사 음소를 추출하고, 상기 인식된 음소 및 상기 유사 음소를 이용하여 상기 일 구간의 후보 단어를 추출하기 위한 단어 그래프를 생성하는 것을 특징으로 하는 전자 장치.

청구항 10

제9항에 있어서,

상기 음성 인식부는,

제2 음향 모델을 이용하여 상기 일 구간의 음성 신호에 대한 가우시안 확률을 계산하고, 상기 제2 음향 모델 및 언어 모델을 이용하여 상기 단어 그래프 내에 확률이 가장 높은 단어열을 출력하여 음성 인식을 수행하는 것을 특징으로 하는 전자 장치.

청구항 11

제10항에 있어서,

상기 음소 인식부의 제1 음향 모델 및 상기 음성 인식부의 제2 음향 모델은 서로 상이한 음향 모델인 것을 특징으로 하는 전자 장치.

청구항 12

제7항에 있어서,

상기 음소 인식부, 상기 후보 단어 추출부 및 상기 음성 인식부는 상이한 코어로 구현되는 것을 특징으로 하는 전자 장치.

명세서

기술분야

[0001] 본 발명은 음성 인식 방법 및 이를 적용한 전자 장치에 관한 것으로, 더욱 상세하게는 음향 모델, 언어 모델 및 발음 사전을 이용하여 사용자가 발화한 음성을 인식하는 음성 인식 방법 및 이를 적용한 전자 장치에 관한 것이다.

배경기술

[0002] 근래에는 스마트 폰, 네비게이션과 같은 전자 장치를 음성 인식을 이용하여 제어하고 있다. 특히, 최근 전자 장치의 하드웨어의 성능향상 및 사용자의 음성 인식 요구 수준이 높아짐에 따라, 기존의 수십 단어 이내의 명령어 위주로 사용자 음성을 인식하는 고립 단어 인식에서 대용량의 자연어 인식하는 연속어 음성 인식 방법으로 사용자 환경이 변화되고 있다.

[0003] 기존의 연속어 음성 인식은 수십, 수백만 개의 단어 중 적어도 하나의 단어를 포함하는 단어열을 인식하는 것으로, 가능한 모든 어휘에 대해서 탐색 공간을 구성하고, 해당 발화가 어떠한 문장인지를 판별하기 위해 음향 모델, 언어 모델 및 발음 사전으로 구성된 정보를 이용하여 확률을 계산하며, 계산 결과에 따라 인식된 문장을 획득하게 된다.

[0004] 그러나, 기존의 연속어 음성 인식은 수 많은 단어를 인식하기 위하여, 탐색 공간이 거대화됨에 따라 메모리 요구량이 증가되며, 계산량 증가로 인해 인식이 불가능하거나 인식 속도가 현저히 떨어지는 문제점이 존재한다.

[0005] 따라서, 대용량의 자연어를 더욱 신속하게 인식하기 위한 음성 인식 방안의 모색이 요청된다.

발명의 내용

[0006] 본 발명은 상술한 문제점을 해결하기 위한 것으로, 본 발명의 목적은 대용량의 자연어를 더욱 신속하게 인식하기 위한 음성 인식 방법 및 이를 적용한 전자 장치를 제공함에 있다.

[0007] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른, 전자 장치의 음성 인식 방법은, 음성 신호가 입력되면, 상기 음성 신호를 기 설정된 시간 간격의 복수의 구간으로 분할하는 단계; 제1 음향 모델을 이용하여 상기 복수의 구간 중 일 구간의 음성 신호에 대한 음소 인식을 수행하는 단계; 상기 음소 인식 결과를 이용하여 상기 일 구간의 음성 신호에 대한 후보 단어를 추출하는 단계; 및 상기 후보 단어를 이용하여 상기 일 구간의 음성 신호에 대한 음성 인식을 수행하는 단계;를 포함한다.

[0008] 그리고, 상기 음소 인식을 수행하는 단계는, 분할 비터비 알고리즘을 이용하여 상기 일 구간의 음성 신호에 포함된 복수의 음소 중 끝 부분의 적어도 하나의 음소를 삭제하는 단계;를 더 포함하고, 상기 삭제된 적어도 하나의 음소는, 다음 구간의 음성 신호에 대한 음소 인식을 수행할 때 이용될 수 있다.

[0009] 또한, 상기 추출하는 단계는, 상기 인식된 음소와 발음이 유사한 유사 음소를 추출하는 단계; 및 상기 인식된 음소 및 상기 유사 음소를 이용하여 상기 일 구간의 후보 단어를 추출하기 위한 단어 그래프를 생성하는 단계;를 포함할 수 있다.

[0010] 그리고, 상기 음성 인식을 수행하는 단계는, 제2 음향 모델을 이용하여 상기 일 구간의 음성 신호에 대한 가우시안 확률을 계산하는 단계; 상기 제2 음향 모델 및 언어 모델을 이용하여 상기 단어 그래프 내에 확률이 가장 높은 단어열을 출력하는 단계;를 포함할 수 있다.

[0011] 또한, 상기 제1 음향 모델 및 상기 제2 음향 모델은 서로 상이한 음향 모델일 수 있다.

[0012] 그리고, 상기 음소 인식을 수행하는 단계, 상기 추출하는 단계 및 상기 음성 인식을 수행하는 단계는 각기 상이한 코어에 의해 병렬적으로 수행될 수 있다.

[0013] 한편, 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른, 전자 장치는, 음성 신호를 입력받는 음성 신호 입력부; 상기 입력부를 통해 입력된 음성 신호를 기 설정된 시간 간격의 복수의 구간으로 분할하는 음성 신호 분할부; 제1 음향 모델을 이용하여 상기 복수의 구간 중 일 구간의 음성 신호에 대한 음소 인식을 수행하는 음소 인식부; 상기 음소 인식 결과를 이용하여 상기 일 구간의 음성 신호에 대한 후보 단어를 추출하는 후보 단어 추출부; 및 상기 후보 단어를 이용하여 상기 일 구간의 음성 신호에 대한 음성 인식을 수행하는 음성 인식부;를 포함한다.

[0014] 그리고, 상기 음소 인식부는, 분할 비터비 알고리즘을 이용하여 상기 일 구간의 음성 신호에 포함된 복수의 음

소 중 끝 부분의 적어도 하나의 음소를 삭제하여 음소 인식을 수행하며, 상기 삭제된 적어도 하나의 음소는, 다음 구간의 음성 신호에 대한 음소 인식을 수행할 때 이용될 수 있다.

[0015] 또한, 상기 후보 단어 추출부는, 상기 인식된 음소와 발음이 유사한 유사 음소를 추출하고, 상기 인식된 음소 및 상기 유사 음소를 이용하여 상기 일 구간의 후보 단어를 추출하기 위한 단어 그래프를 생성할 수 있다.

[0016] 그리고, 상기 음성 인식부는, 제2 음향 모델을 이용하여 상기 일 구간의 음성 신호에 대한 가우시안 확률을 계산하고, 상기 제2 음향 모델 및 언어 모델을 이용하여 상기 단어 그래프 내에 확률이 가장 높은 단어열을 출력하여 음성 인식을 수행할 수 있다.

[0017] 또한, 상기 음소 인식부의 제1 음향 모델 및 상기 음성 인식부의 제2 음향 모델은 서로 상이한 음향 모델일 수 있다.

[0018] 그리고, 상기 음소 인식부, 상기 후보 단어 추출부 및 상기 음성 인식부는 상이한 코어로 구현될 수 있다.

[0019] 상술한 바와 같은 본 발명의 다양한 실시예에 의해, 전자 장치는 사용자가 발화한 음성을 더욱 신속하게 인식할 수 있게 된다.

도면의 간단한 설명

[0020] 도 1은 본 발명의 일 실시예에 따른, 음성 인식을 수행하기 위한 전자 장치의 구성을 간략히 도시한 도면,
도 2는 본 발명의 일 실시예에 따른, 음성 인식을 수행하기 위한 전자 장치의 구성을 자세히 도시한 도면,
도 3은 본 발명의 일 실시예에 따른, 병렬적인 음성 인식 처리 방법을 설명하기 위한 도면, 그리고
도 4는 본 발명의 일 실시예에 따른, 음성 인식 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0021] 이하에서는 도면을 참조하여 본 발명에 대해 더욱 상세히 설명하도록 한다.

[0022] 도 1은 본 발명의 일 실시예에 따른, 전자 장치(100)의 음성 인식을 수행하기 위한 구성을 나타내는 블록도이다. 도 1에 도시된 바와 같이, 전자 장치(100)는 음성 신호 입력부(110), 음성 신호 분할부(120), 음소 인식부(130), 후보 단어 추출부(140) 및 음성 인식부(150)를 포함한다. 본 발명의 일 실시예에 따른, 전자 장치(100)는 스마트 폰, 스마트 TV, 데스크 탑 PC, 태블릿 PC 등과 같은 다양한 전자 장치로 구현될 수 있다.

[0023] 음성 신호 입력부(110)는 사용자가 발화한 음성에 대응되는 음성 신호를 입력받는다. 이때, 음성 신호 입력부(110)는 마이크와 같은 음성 입력 수단 및 입력된 음성을 증폭하기 위한 앰프를 포함할 수 있다. 그러나, 마이크를 이용하여 실시간으로 음성 신호를 입력받는 것은 일 실시예에 불과할 뿐, 음성 신호 입력부(110)는 기 저장된 파일을 통해 음성 신호를 입력받을 수 있다.

[0024] 음성 신호 분할부(120)는 입력된 음성 신호를 복수의 구간으로 분할한다. 구체적으로, 음성 신호 분할부(120)는 입력된 음성 신호를 기설정된 시간 간격(예를 들어, 0.1초)의 복수의 구간으로 분할할 수 있다.

[0025] 음소 인식부(130)는 음성 신호 분할부(120)에서 분할된 복수의 구간 중 일 구간의 음성 신호에 대한 음소 인식을 수행한다. 이때, 구체적으로, 음소 인식부(130)는 음소 인식을 위한 제1 음향 모델을 이용하여 입력된 일 구간의 음성 신호에 대응되는 특징 벡터에 대한 가우시안 확률 분포를 계산하고, 최적의 음소를 선택한다.

[0026] 이때, 음소 인식부(130)는 분할 비터비(segmented viterbi) 알고리즘을 이용하여 일 구간의 음성 신호에 포함된 복수의 음소 중 끝 부분의 적어도 하나의 음소를 삭제할 수 있다. 구체적으로, 음소 단위로 음성 신호를 분할하는 것이 아닌 시간 단위로 음성신호를 분할하므로, 음소 인식부(130)는 일 구간의 끝 부분에 위치하는 음소를 제대로 인식할 수 없게 된다. 따라서, 음소 인식부(130)는 일 구간의 끝 부분에 위치하는 적어도 하나의 음소를 삭제하고, 삭제된 적어도 하나의 음소가 다음 구간의 음소 인식시 이용될 수 있도록 음성 신호 분할부(120)로 출력한다.

[0027] 후보 단어 추출부(140)는 음소 인식부(130)에서 출력된 음소 인식 결과를 이용하여 인식된 음소에 대한 후보 단어를 추출한다. 구체적으로, 후보 단어 추출부(140)는 음소 인식부(130)에서 출력된 음소와 발음이 유사한 유사 음소를 추출하고, 유사 음소를 이용하여 후보 단어를 추출하기 위한 일 구간의 음성 신호에 대한 단어 그래프를 생성한다. 그러나, 후보 단어를 추출하기 위해, 일 구간의 음성 신호에 대한 단어 그래프를 생성하는 것은 일

실시예에 불과할 뿐, 일 구간의 음성 신호에 대한 후보 단어 리스트를 생성할 수 있다. 그리고, 후보 단어 추출부(130)는 생성된 일 구간의 음성 신호에 대한 단어 그래프를 음성 인식부(150)로 출력한다.

- [0028] 음성 인식부(150)는 후보 단어 추출부(140)에서 추출된 후보 단어를 이용하여 일 구간의 음성 신호에 대한 음성 인식을 수행한다. 구체적으로, 음성 인식부(150)는 음성 신호 분할부(120)에서 출력된 일 구간의 음성 신호에 대하여 후보 단어 추출부(140)에서 출력된 단어 그래프 중에서 최적 경로를 탐색하여 음성 인식을 수행할 수 있다.
- [0029] 이때, 음소 인식부(130), 후보 단어 추출부(140) 및 음성 인식부(150)는 서로 다른 코어에서 병렬적으로 동작할 수 있다. 즉, 음소 인식부(130)에서 제1 구간의 음성 신호에 대한 음소 인식이 수행되면, 제1 구간의 음소 인식 결과를 후보 단어 추출부(140)로 전송하고, 바로 제2 구간의 음성 신호에 대한 음소 인식을 수행한다. 그리고, 후보 단어 추출부(140)는 제1 구간의 음소 인식 결과를 바탕으로 후보 단어를 추출하고, 추출된 후보 단어를 음성 인식부(150)로 출력하며, 바로 음소 인식부(130)에서 출력된 제2 구간의 음성 인식에 대한 음소 인식 결과를 이용하여 후보 단어를 추출한다. 그리고, 음성 인식부(150)는 후보 단어 추출부(140)에서 추출된 제1 구간의 음성 신호에 대한 후보 단어를 이용하여 제1 구간의 음성 신호에 대한 음성 인식을 수행하고, 바로 후보 단어 추출부(140)에서 추출된 제2 구간의 음성 신호에 대한 후보 단어를 이용하여 제2 구간의 음성 신호에 대한 음성 인식을 수행한다.
- [0030] 상술한 바와 같은 전자 장치에 의해, 전자 장치(100)는 상대적으로 계산량이 적은 음소인식을 빠르게 수행하고, 인식된 음소 인식 결과를 바탕으로 적은 양의 후보 단어를 추출한 후, 기존의 대용량 연속어 인식 방법에 비해 현저히 적은 숫자의 후보 단어를 이용하여 음성 인식을 수행할 수 있게 된다. 또한, 음소 인식, 후보 단어 추출, 음성 인식을 병렬적으로 수행함으로써, 사용자는 더욱 신속하게 음성 인식을 수행할 수 있게 된다.
- [0031] 도 2는 본 발명의 일 실시예에 따른, 전자 장치(100)의 음성 인식을 위한 구성을 자세히 도시한 블록도이다. 도 2에 도시된 바와 같이, 전자 장치(100)는 음성 신호 입력부(110), 음성 신호 분할부(120), 음소 인식부(130), 후보 단어 추출부(140) 및 음성 인식부(150)를 포함한다.
- [0032] 음성 입력부(110)는 사용자 음성에 대응되는 음성 신호를 입력받는다. 이때, 음성 입력부(110)는 마이크와 같은 음성 입력 장치로부터 실시간으로 음성 신호를 입력받을 수 있으나, 이는 일 실시예에 불과할 뿐, 전자 장치(100)의 저장부(미도시)에 저장된 파일로부터 음성 신호를 입력받을 수 있다.
- [0033] 음성 신호 분할부(120)는 음성 신호를 기 설정된 시간의 복수의 구간으로 분할한다. 이때, 음성 신호 분할부(120)는 구간 분할부(121), 전처리부(122), 특징 벡터 추출부(123)를 포함한다.
- [0034] 구간 분할부(121)는 기 설정된 시간 간격(예를 들어, 0.1초)으로 음성 신호 입력부(110)로부터 출력된 음성 신호를 분할한다.
- [0035] 전처리부(122)는 분할된 일 구간의 음성 신호에 대한 노이즈 제거와 같은 신호 처리를 수행한다.
- [0036] 특징 벡터 추출부(123)는 전처리가 수행된 일 구간의 음성 신호로부터 특징 벡터를 추출한다. 그리고, 특징 벡터 추출부(123)는 추출된 일 구간의 음성 신호에 대한 특징 벡터를 음소 인식부(130) 및 음성 인식부(150)에 각각 출력한다.
- [0037] 음소 인식부(130)는 특징 벡터 추출부(123)에서 추출된 특징 벡터를 이용하여 음소 인식을 수행한다. 이때, 음소 인식부(130)는 제1 가우시안 확률 계산부(131), 제1 음향 모델(132), 최적 후보 탐색부(133) 및 구간 분할 오류 보정부(134)를 포함한다.
- [0038] 제1 가우시안 확률 계산부(131)는 제1 음향 모델(132)을 이용하여 입력된 일 구간의 음성 신호에 대한 특징 벡터의 가우시안 확률 계산한다.
- [0039] 이때, 제1 음향 모델(132)은 음소 인식을 위한 음향 모델로서, 한국어의 경우, 40~50개 정도의 음소에 대한 정보가 저장된다. 그리고, 제1 음향 모델은 HMM(hidden Markov model) 음향 모델을 이용할 수 있다. 특히, 제1 음향 모델(132)은 기존의 대용량 연속어 인식 방법에 적용되는 음향 모델보다 간단하게 구현될 수 있어, 빠른 음소 인식이 가능하다.
- [0040] 최적 후보 탐색부(133)는 제1 음향 모델(132) 및 제1 가우시안 확률 계산부(131)의 계산 결과를 바탕으로 일 구간의 음성 신호에 포함된 최적의 음소들을 선택한다.

- [0041] 구간 분할 오류 수정부(134)는 최적 후보 탐색부(133)에서 선택된 복수의 음소 중 끝 부분의 적어도 하나의 음소를 삭제한다. 구체적으로, 본 발명의 일 실시예에 따른, 음성 신호 분할부(120)는 음소 기준으로 음성 신호를 분할하는 것이 아닌 시간 기준으로 음성 신호를 분할한다. 이에 의해, 음소 인식부(130)에 입력되는 일 구간의 음성 신호 중 끝 부분의 음소들에 대해서는 모든 데이터가 입력되지 못하는 경우가 발생하므로, 최적 후보 탐색부(133)에서 선택된 복수의 음소 중 끝부분의 적어도 하나의 음소는 잘못 선택된 음소일 수 있다. 따라서, 구간 분할 오류 수정부(134)는 최적 후보 탐색부(133)에서 선택된 복수의 음소 중 끝 부분의 적어도 하나의 음소를 삭제하고, 삭제되지 않은 음소들을 후보 단어 추출부(140)로 출력한다. 그리고, 구간 분할 오류 수정부(134)는 삭제된 적어도 하나의 음소를 다음 구간에서 인식할 수 있도록 구간 분할부(121)로 출력한다.
- [0042] 한편, 본 발명의 일 실시예에 따른 음소 인식부(130)는 구간 분할 오류 수정부(134)를 통해 최적 후보 탐색부(133)에서 선택된 복수의 음소 중 끝 부분의 적어도 하나의 음소를 삭제하여 구간 분할 오류를 수정할 수 있으나, 이는 일 실시예에 불과할 뿐, HMM 스테이트 위치 파악 또는 신호처리적인 기법을 이용하여 음소의 끝부분을 찾아 구간 분할 오류를 최소화할 수 있다.
- [0043] 후보 단어 추출부(140)는 음소 인식부(130)에서 인식된 일 구간의 음성 신호에 대한 음소를 바탕으로 후보 단어를 추출한다. 이때, 후보 단어 추출부(140)는 유사도 계산부(141) 및 구간 단어 그래프 생성부(142)를 포함한다.
- [0044] 유사도 계산부(141)는 발음 사전을 이용하여 입력된 일 구간의 음성 신호에 대한 음소와 다른 음소들간의 발음 유사도를 계산하여, 입력된 일 구간의 음성 신호에 대한 음소와 발음이 유사한 유사 음소를 추출한다.
- [0045] 구간 단어 그래프 생성부(142)는 추출된 유사 음소들을 바탕으로 후보 단어를 생성할 수 있는 구간 단어 그래프를 생성한다. 이때, 구간 단어 그래프는 음소 인식부(130)에서 인식된 음소들과 유사 음소들이 연결된 네트워크 형태의 그래프일 수 있다. 그리고, 구간 단어 그래프 생성부(142)는 생성된 일 구간의 음성 신호에 대하여 후보 단어를 추출하기 위한 단어 그래프를 최적 단어 그래프 경로 탐색부(153)로 출력한다.
- [0046] 한편, 상술한 실시예에서는 후보 단어 추출부(140)가 구간 단어 그래프를 생성하는 것으로 설명하였으나, 이는 일 실시예에 불과할 뿐, 후보단어들을 추출하여 후보단어 리스트를 생성하는 실시예 역시 본 발명의 기술적 사항이 적용될 수 있다.
- [0047] 음성 인식부(150)는 후보 단어 추출부(140)에서 출력한 후보 단어들을 이용하여 일 구간에 대한 음성 인식을 수행한다. 이때, 음성 인식부(150)는 제2 가우시안 확률 계산부(151), 제2 음향 모델(152), 최적 단어 그래프 경로 탐색부(153), 언어 모델(154) 및 음성 인식 출력부(155)를 포함한다.
- [0048] 제2 가우시안 확률 계산부(151)는 제2 음향 모델(152)을 이용하여 일 구간의 음성 신호에 대한 가우시안 확률 분포를 계산한다.
- [0049] 이때, 제2 음향 모델(152)은 일반적인 대용량 연속어 인식 방법에 사용되는 음향 모델로서, 트라이폰(triphone)을 이용한 음향 모델일 수 있다. 특히, 복잡한 음성 인식을 수행하기 위해, 제1 음향 모델(152)은 제1 음향 모델(132)에 비하여 많은 양의 데이터를 저장한다.
- [0050] 최적 단어 그래프 경로 탐색부(153)는 제2 가우시안 확률 계산부(151)의 계산 결과를 바탕으로 구간 단어 그래프 생성부(142)에서 출력된 단어 그래프 중 음성 신호에 대응되는 최적의 경로를 탐색한다. 이때, 최적 단어 그래프 경로 탐색부(153)는 문장 인식을 더욱 정확히 하기 위하여, 문법 및 문장 구조가 저장된 언어 모델(154)을 이용하여 음성 인식을 수행할 수 있다. 즉, 제1 음향모델(132)은 고속의 음소인식에 특화된 음향 모델을 사용하고, 제2 음향모델(152)은 연속어 음성인식의 성능을 높이기 위한 정교한 모델을 사용할 수 있다.
- [0051] 음성 인식 출력부(155)는 최적 단어 그래프 경로 탐색부(153)에서 탐색된 최적 경로에 의해 생성된 단어열(문장)을 출력한다.
- [0052] 한편, 음소 인식부(120), 후보 단어 추출부(130) 및 음성 인식부(140)는 각각 상이한 코어에 의해 병렬적으로 동작하는 파이프라인 형태로 구성될 수 있다. 구체적으로, 도 3에 도시된 바와 같이, 음성 신호 분할부(120)는 음성 신호를 N개의 구간으로 나누고, 각각의 구간에 대한 음성 신호를 음소 인식부(130)에 출력한다. 그리고, t_1 시간에는 음소 인식부(130)가 제1 구간에 대한 음성 인식을 수행한다. 그리고, t_2 시간에는 음소 인식부(130)가 제2 구간에 대한 음성 인식을 수행하고, 후보 단어 추출부(140)가 제1 구간에 대한 후보 단어를 추출한다. 그리고, t_3 시간에는 음소 인식부(130)가 제3 구간에 대한 음성 인식을 수행하고, 후보 단어 추출부(140)가 제2

구간에 대한 후보 단어를 추출하며, 음성 인식부(150)가 제1 구간에 대한 음성 인식을 수행한다. 이와 같은 방식으로, 각각의 시간에 대해 음소 인식부(120), 후보 단어 추출부(130) 및 음성 인식부(140)는 병렬적으로 동작으로 수행하게 된다. 그리고, 음성 인식부(150)는 사용자가 발화를 마친 시점(t_n)으로부터 짧은 시간($t_{n+2} - t_n$) 후 모든 구간의 음성 신호에 대한 음성 인식을 수행하여 출력한다.

[0053] 상술한 바와 같이 전자 장치에 의해, 음소 인식 동작, 음소 인식을 이용한 후보 단어 추출 동작, 및 후보 단어를 이용한 음성 인식 동작이 병렬적으로 수행됨으로써, 전자 장치(100)는 기존의 대용량 연속어 인식 방법에 비해 더욱 신속하게 음성 인식을 수행할 수 있다.

[0054] 이하에서는 도 4를 참조하여, 본 발명의 일 실시예에 따른, 전자 장치(100)의 음성 인식 방법을 설명하도록 한다.

[0055] 전자 장치(100)는 음성 신호를 입력받는지 여부를 판단한다(S410). 이때, 음성 신호는 마이크와 같은 음성 입력 장치를 통해 실시간으로 입력될 수 있으며, 기 저장된 파일을 통해 입력될 수 있다.

[0056] 음성 신호가 입력되면(S410-Y), 전자 장치(100)는 음성 신호를 기 설정된 시간 간격으로 복수의 구간으로 분할한다(S420). 구체적으로, 전자 장치(100)는 입력된 음성 신호를 기 설정된 시간 간격(예를 들어, 0.1초)의 복수의 구간으로 분할하고, 분할된 일 구간의 음성 신호에 대해 신호 처리를 수행하여 특징 벡터를 추출한다.

[0057] 그리고, 전자 장치(100)는 일 구간의 음성 신호에 대한 음소를 인식한다(S430). 구체적으로, 전자 장치(100)는 제1 음향 모델을 이용하여 분할된 일 구간의 음성 신호에 대한 음소를 인식한다. 이때, 더욱 정확한 음소 인식을 위하여, 전자 장치(100)는 인식된 복수의 음소 중 끝 부분의 적어도 하나의 음소를 삭제하고, 삭제된 적어도 하나의 음소를 다음 구간의 음성 신호에 대한 음소 인식시 이용한다.

[0058] 그리고, 전자 장치(100)는 음소 인식 결과를 이용하여 일 구간의 음성 신호에 대한 후보 단어를 추출한다(S440). 구체적으로, 전자 장치(100)는 인식된 복수의 음소 각각에 대한 유사 음소를 추출한 후, 후보 단어를 추출하기 위한 단어 그래프를 생성한다. 이때, 단어 그래프는 인식된 음소들과 유사 음소들이 각각 연결된 네트워크 형태의 그래프이다.

[0059] 그리고, 전자 장치(100)는 후보 단어를 이용하여 일 구간의 음성 신호에 대한 음성 인식을 수행한다(S450). 구체적으로, 전자 장치(100)는 S440 단계에서 추출된 후보 단어(단어 그래프) 내에서 제2 음형 모델 및 언어 모델을 이용하여, 일 구간의 음성 신호에 대한 음성 인식을 수행한다.

[0060] 그리고, 전자 장치(100)는 다음 구간들의 음성 신호에 대해 S430 내지 S450 동작을 반복적으로 수행할 수 있다. 그리고, 전자 장치(100)는 S430 내지 S450 동작을 반복적으로 수행할 때, S430 내지 S450 동작을 각각 상이한 코어를 이용하여 병렬적으로 수행할 수 있다.

[0061] 상술한 바와 같은 음성 인식 방법에 의해, 전자 장치는 기존의 대용량 연속어 인식 방법에 비하여 더욱 신속하고 정확한 음성 인식을 수행할 수 있다.

[0062] 한편, 이상과 같은 다양한 실시 예에 따른 음성 인식 방법을 수행하기 위한 프로그램 코드는 컴퓨터에 의해 관
독가능한 비일시적인(non-transitory) 기록 매체에 저장될 수 있다. 예를 들어, 하드디스크, 리무버블 디스크,
메모리 카드, USB 메모리, CD-ROM 등과 같이, 단말기에서 관독 가능한 다양한 유형의 기록 매체에 저장되어 있
을 수 있다.

[0063] 이상에서는 본 발명의 바람직한 실시 예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특정의 실시 예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어져서는 안될 것이다.

부호의 설명

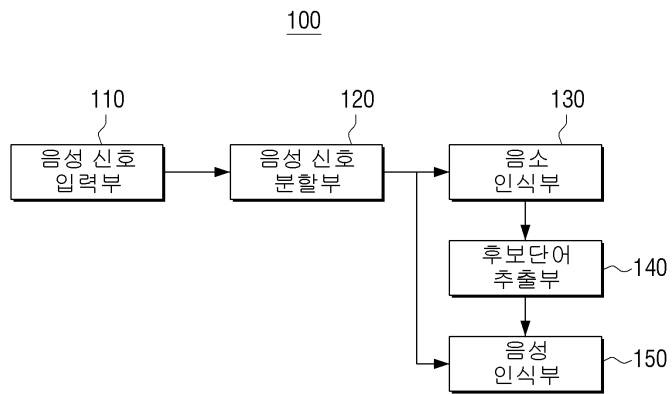
[0064] 110: 음성 신호 입력부 120: 음성 신호 분할부

 130: 음소 인식부 140: 후보 단어 추출부

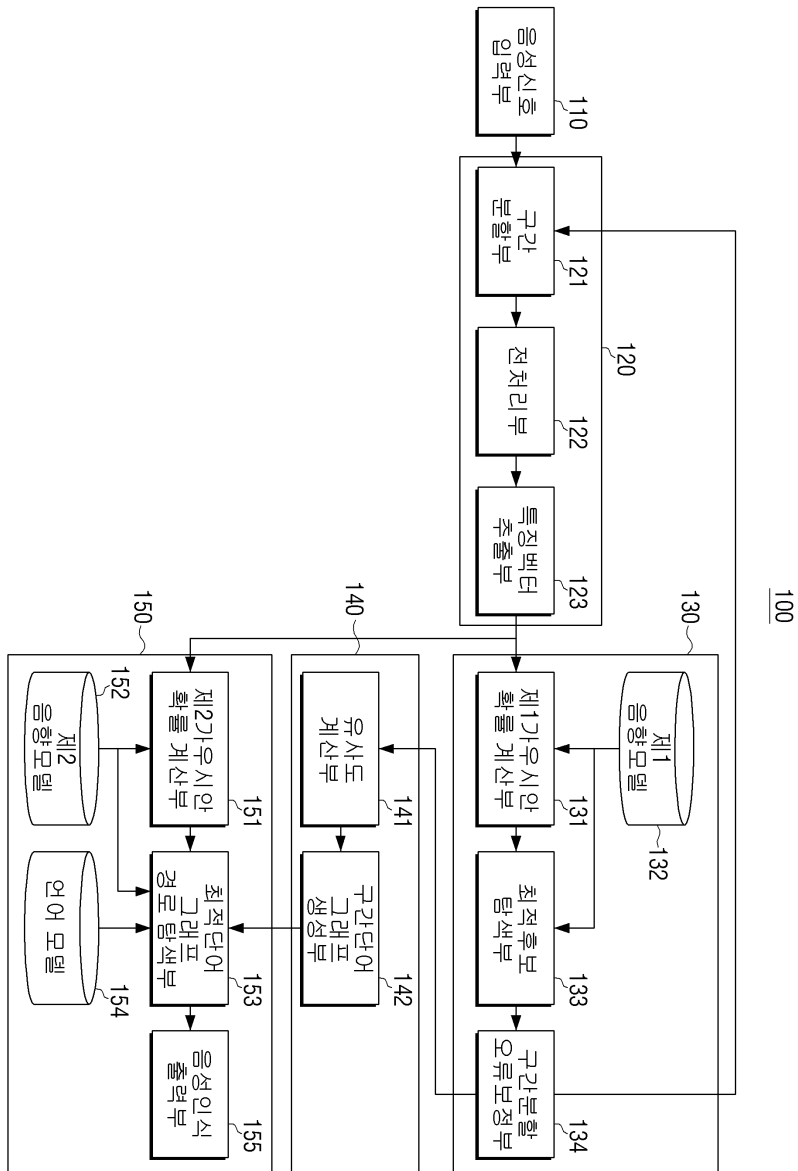
150: 음성 인식부

도면

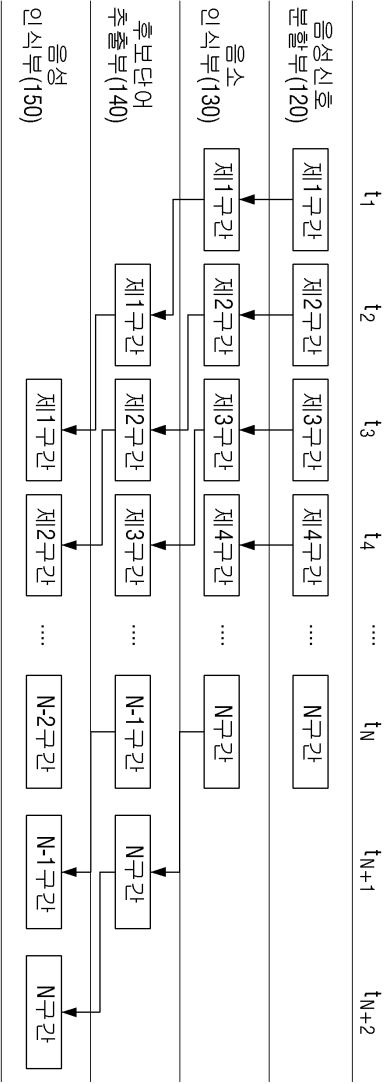
도면1



도면2



도면3



도면4

