



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0119152
(43) 공개일자 2017년10월26일

- | | |
|--|---|
| <p>(51) 국제특허분류(Int. Cl.)
 <i>G10L 15/16</i> (2006.01) <i>G06N 3/02</i> (2006.01)
 <i>G10L 15/02</i> (2006.01) <i>G10L 17/04</i> (2013.01)
 <i>G10L 19/035</i> (2013.01)</p> <p>(52) CPC특허분류
 <i>G10L 15/16</i> (2013.01)
 <i>G06N 3/02</i> (2013.01)</p> <p>(21) 출원번호 10-2016-0046952
 (22) 출원일자 2016년04월18일
 심사청구일자 2016년04월18일</p> | <p>(71) 출원인
 한양대학교 산학협력단
 서울특별시 성동구 왕십리로 222(행당동, 한양대학교내)</p> <p>(72) 발명자
 이지혜
 서울특별시 서초구 방배로 21, 5동 803호(방배동, 강남아파트)</p> <p>장준혁
 서울특별시 강남구 언주로30길 26, G동 5007호 (도곡동, 타워팰리스)</p> <p>(74) 대리인
 양성보</p> |
|--|---|

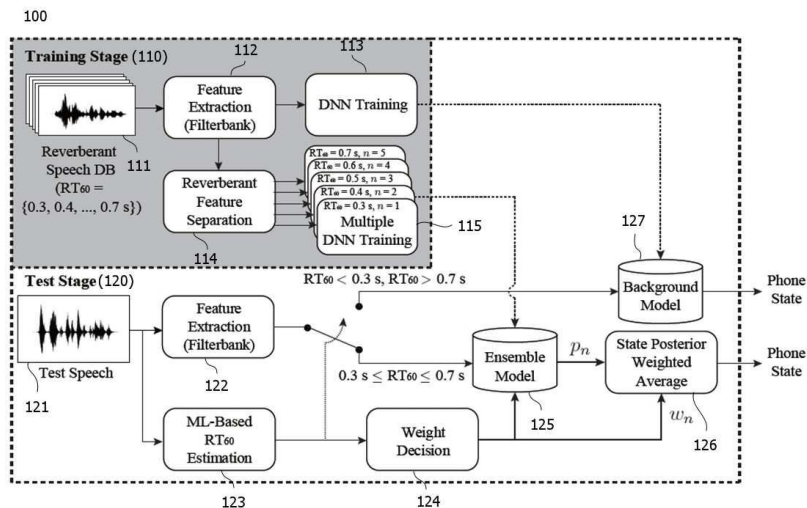
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델 및 이를 이용한 음성인식 방법**

(57) 요약

잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델 및 이를 이용한 음성인식 방법이 제시된다. 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법에 있어서, 입력되는 음성 신호로부터 특징 벡터를 추출하는 단계; 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 심화신경망 앙상블 기반의 음향 모델의 앙상블을 이용하여 결합하는 단계; 및 음소를 분류하여 음성을 인식하는 단계를 포함하고, 상기 미리 학습된 심화신경망 앙상블 기반의 음향 모델은, 다수의 잔향 환경에서 각각의 잔향 환경에 대해 음소 확률을 추정할 수 있다.

대표도 - 도1



(52) CPC특허분류

G10L 15/02 (2013.01)

G10L 17/04 (2013.01)

G10L 19/035 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711030639

부처명 미래창조과학부

연구관리전문기관 (재)한국연구재단

연구사업명 이공분야 기초연구사업 / 중견연구자지원사업 / 도약연구(전략-연구분야지정)

연구과제명 미래 웨어러블기기용 고성능 음성신호처리 핵심기술개발

기 여 율 1/1

주관기관 한양대학교 산학협력단

연구기간 2015.11.01 ~ 2016.10.31

명세서

청구범위

청구항 1

잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법에 있어서,

입력되는 음성 신호로부터 특징 벡터를 추출하는 단계;

상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 심화신경망 앙상블 기반의 음향 모델의 앙상블을 이용하여 결합하는 단계; 및

음소를 분류하여 음성을 인식하는 단계

를 포함하고,

상기 미리 학습된 심화신경망 앙상블 기반의 음향 모델은,

다수의 잔향 환경에서 각각의 잔향 환경에 대해 음소 확률을 추정하는 것

을 특징으로 하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법.

청구항 2

제1항에 있어서,

상기 음성 신호로부터 잔향 시간을 추정하는 단계; 및

상기 잔향 시간을 기반으로 가중치를 산출하는 단계

를 더 포함하고,

상기 음소를 분류하여 음성을 인식하는 단계는,

산출된 상기 가중치를 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델의 앙상블 결합에 적용하는 단계

를 포함하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법.

청구항 3

제2항에 있어서,

상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 심화신경망 앙상블 기반의 음향 모델의 앙상블을 이용하여 결합하는 단계는,

상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델에 기반한 앙상블 모델을 통과시켜 음소 확률을 추정하는 것

을 특징으로 하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법.

청구항 4

제3항에 있어서,

상기 음소를 분류하여 음성을 인식하는 단계는,

상기 음소 확률과 상기 가중치를 이용하여 상기 심화신경망 앙상블 기반의 음향 모델의 사후 확률(posterior probability)을 산출하여 상기 음소를 분류하는 것

을 특징으로 하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법.

청구항 5

제2항에 있어서,

상기 음성 신호로부터 잔향 시간을 추정하는 단계는,

상기 심화신경망 앙상블 기반의 음향 모델 중 우도비가 가장 큰 두 개의 음향 모델을 선택하여 최대 우도법(maximum likelihood)을 통한 상기 잔향 시간을 추정하는 것

을 특징으로 하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법.

청구항 6

제1항에 있어서,

상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 단계

를 더 포함하고,

상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 단계는,

학습 단계에서, 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출하는 단계;

추출된 상기 특징 벡터의 잔향 환경의 특징을 분리하는 단계; 및

추출된 상기 특징 벡터를 심화신경망을 통하여 각각의 상기 잔향 환경에 대해 학습시키는 단계

를 포함하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법.

청구항 7

제1항에 있어서,

상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 단계

를 더 포함하고,

상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 단계는,

학습 단계에서, 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출하는 단계;

추출된 상기 특징 벡터의 잔향 환경의 특징을 분리하는 단계;

추출된 상기 특징 벡터를 잔향이 없는 음성 특징으로 맵핑(mapping)시키는 상기 특징 맵핑 심화신경망을 통과시키는 단계;

상기 특징 맵핑 심화신경망의 출력을 이용하여 상기 음향 모델링 심화신경망을 학습시키는 단계; 및

상기 음향 모델링 심화신경망은 상기 특징 맵핑 심화신경망 위에 바로 쌓이고, 결합된 상기 심화신경망 앙상블 기반의 음향 모델은 연결되어 재학습되는 단계

를 포함하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법.

청구항 8

잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템에 있어서,

입력되는 음성 신호로부터 특징 벡터를 추출하는 특징 벡터 추출부;

상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 심화신경망 앙상블 기반의 음향 모델의 앙상블을 이용하여 결합하는 앙상블 모델; 및

음소를 분류하여 음성을 인식하는 음소 분류부

를 포함하고,

상기 미리 학습된 심화신경망 앙상블 기반의 음향 모델은,

다수의 잔향 환경에서 각각의 잔향 환경에 대해 음소 확률을 추정하는 것

을 특징으로 하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템.

청구항 9

제8항에 있어서,

상기 음성 신호로부터 잔향 시간을 추정하는 잔향 시간 예측부; 및

상기 잔향 시간을 기반으로 가중치를 산출하는 가중치 결정부

를 더 포함하고,

상기 음소 분류부는,

산출된 상기 가중치를 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델의 앙상블 결합에 적용하는 것

을 특징으로 하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템.

청구항 10

제9항에 있어서,

상기 앙상블 모델은,

상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델에 기반한 앙상블 모델을 통과시켜 음소 확률을 추정하는 것

을 특징으로 하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템.

청구항 11

제10항에 있어서,

상기 음소 분류부는,

상기 음소 확률과 상기 가중치를 이용하여 상기 심화신경망 앙상블 기반의 음향 모델의 사후 확률(posterior probability)을 산출하여 상기 음소를 분류하는 것

을 특징으로 하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템.

청구항 12

제9항에 있어서,

상기 잔향 시간 예측부는,

상기 심화신경망 앙상블 기반의 음향 모델 중 우도비가 가장 큰 두 개의 음향 모델을 선택하여 최대 우도법(maximum likelihood)을 통한 상기 잔향 시간을 추정하는 것

을 특징으로 하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한

음성인식 시스템.

청구항 13

제8항에 있어서,

상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 학습부

를 더 포함하고,

상기 학습부는,

학습 단계에서, 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출하는 특징 벡터 추출부;

추출된 상기 특징 벡터의 잔향 환경의 특징을 분리하는 잔향 특징 분류부; 및

추출된 상기 특징 벡터를 심화신경망을 통하여 각각의 상기 잔향 환경에 대해 학습시키는 다중 심화신경망 학습부

를 포함하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템.

청구항 14

제8항에 있어서,

상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 학습부

를 더 포함하고,

상기 학습부는,

잔향이 없는 음성 특징으로 맵핑(mapping)시키는 특징 맵핑 심화신경망과 음향 모델링 심화신경망을 결합하는 구조를 이용하여, 서로 다른 잔향 환경에 대해 학습된 음향 모델을 구성하는 것

을 특징으로 하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템.

청구항 15

제14항에 있어서,

상기 학습부는,

학습 단계에서, 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출하는 특징 벡터 추출부;

추출된 상기 특징 벡터의 잔향 환경의 특징을 분리하는 잔향 특징 분류부;

추출된 상기 특징 벡터를 잔향이 없는 음성 특징으로 맵핑(mapping)시키는 상기 특징 맵핑 심화신경망; 및

상기 특징 맵핑 심화신경망의 출력을 이용하여 학습시키는 상기 음향 모델링 심화신경망

을 포함하고,

상기 음향 모델링 심화신경망은 상기 특징 맵핑 심화신경망 위에 바로 쌓이고, 결합된 상기 심화신경망 앙상블 기반의 음향 모델은 연결되어 재학습되는 것

을 특징으로 하는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템.

발명의 설명

기술 분야

[0001]

아래의 실시예들은 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델 및 이를 이용한 음성인식 방법에 관한 것이다.

배경 기술

- [0002] 음성 인식(Speech recognition) 기술은 자동적 수단에 의하여 음성으로부터 언어적 의미 내용을 식별하는 것으로, 음성파형을 입력하여 단어나 단어열을 식별하고 의미를 추출하는 처리 과정이며, 크게 음성 분석, 음소 인식, 단어 인식, 문장 해석, 의미 추출의 5 가지로 분류된다.
- [0003] 최근의 음성 인식 기술은 HMM(Hidden Markov Model)을 기반으로 구현되고 있다. HMM 기반의 음성 인식 기술은 각 HMM 상태의 가우시안 성분(Gaussian mixture component) 중 확률적 분포가 가장 유사한 두 개의 가우시안 성분들을 점진적으로 통합하여 이진 트리를 구성한 뒤, 구성된 이진 트리를 적정 수준에서 가지치기하여 최적의 음향 모델을 생성하고, 생성된 음향 모델을 기반으로 음성 인식을 수행한다.
- [0004] 그런데 최근 음성 인식 기술의 응용 분야가 점차 확대되면서 음성 인식시스템이 적용되는 환경이 다양해지고 있다. 이렇게 주변 환경이 변경될 경우, 주변 잡음의 크기 및 종류, 그리고 음성 파형이 달라질 수 있기 때문에 적용되는 환경에 맞추어 음향 모델을 재 학습하는 과정이 필요하다.
- [0005] 종래의 심화신경망 기법을 이용한 음향 모델은 단순한 구조를 적용하였기 때문에 다양한 실제 환경에 대해 효과적으로 모델링되지 않는다. 또한 다양한 환경의 음성 신호 데이터를 사용하더라도 한계가 존재하고, 학습 데이터와 테스트 환경이 다를 경우 인식성능 저하를 초래할 수 있다.
- [0006] 한국공개특허 10-2016-0015005호는 이러한 클래스 기반 음향 모델의 변별 학습 방법 및 장치, 그리고 이를 이용한 음성 인식 장치에 관한 기술을 기재하고 있다.

발명의 내용

해결하려는 과제

- [0007] 실시예들은 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델 및 이를 이용한 음성인식 방법에 관하여 기술하며, 보다 구체적으로 다양한 잔향 환경의 음성 신호를 효과적으로 모델링하기 위해서 잔향 시간 추정을 이용한 음향 모델 앙상블 구조와 결합 학습된 심화신경망 모델의 앙상블 구조를 이용하는 기술을 제공한다.
- [0008] 실시예들은 심화신경망 앙상블 구조를 기반으로 테스트 환경에 최적화된 음향 모델의 음소 확률을 추정함으로써, 다양한 잔향 환경에서 우수한 인식성능을 가지는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델 및 이를 이용한 음성인식 방법을 제공하는데 있다.
- [0009] 또한, 실시예들은 결합 학습된 심화신경망 모델을 앙상블에 이용하여 잔향 제거와 음향 모델이 통합된 하나의 심화신경망을 생성함으로써, 잔향 환경에 더욱 효과적인 음향 모델 및 이를 이용한 음성인식 방법을 제공하는데 있다.

과제의 해결 수단

- [0010] 일 실시예에 따른 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법에 있어서, 입력되는 음성 신호로부터 특징 벡터를 추출하는 단계; 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 심화신경망 앙상블 기반의 음향 모델의 앙상블을 이용하여 결합하는 단계; 및 음소를 분류하여 음성을 인식하는 단계를 포함하고, 상기 미리 학습된 심화신경망 앙상블 기반의 음향 모델은, 다수의 잔향 환경에서 각각의 잔향 환경에 대해 음소 확률을 추정할 수 있다.
- [0011] 상기 음성 신호로부터 잔향 시간을 추정하는 단계; 및 상기 잔향 시간을 기반으로 가중치를 산출하는 단계를 더 포함하고, 상기 음소를 분류하여 음성을 인식하는 단계는, 산출된 상기 가중치를 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델의 앙상블 결합에 적용하는 단계를 포함할 수 있다.
- [0012] 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 심화신경망 앙상블 기반의 음향 모델의 앙상블을 이용하여 결합하는 단계는, 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델에 기반한 앙상블 모델을 통과시켜 음소 확률을 추정할 수 있다.
- [0013] 상기 음소를 분류하여 음성을 인식하는 단계는, 상기 음소 확률과 상기 가중치를 이용하여 상기 심화신경망 앙상블 기반의 음향 모델의 사후 확률(posterior probability)을 산출하여 상기 음소를 분류할 수 있다.

- [0014] 상기 음성 신호로부터 잔향 시간을 추정하는 단계는, 상기 심화신경망 앙상블 기반의 음향 모델 중 우도비가 가장 큰 두 개의 음향 모델을 선택하여 최대 우도법(maximum likelihood)을 통한 상기 잔향 시간을 추정할 수 있다.
- [0015] 상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 단계를 더 포함하고, 상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 단계는, 학습 단계에서, 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출하는 단계; 추출된 상기 특징 벡터의 잔향 환경의 특징을 분리하는 단계; 및 추출된 상기 특징 벡터를 심화신경망을 통하여 각각의 상기 잔향 환경에 대해 학습시키는 단계를 포함할 수 있다.
- [0016] 상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 단계를 더 포함하고, 상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 단계는, 학습 단계에서, 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출하는 단계; 추출된 상기 특징 벡터의 잔향 환경의 특징을 분리하는 단계; 추출된 상기 특징 벡터를 잔향이 없는 음성 특징으로 맵핑(mapping)시키는 상기 특징 맵핑 심화신경망을 통과시키는 단계; 상기 특징 맵핑 심화신경망의 출력을 이용하여 상기 음향 모델링 심화신경망을 학습시키는 단계; 및 상기 음향 모델링 심화신경망은 상기 특징 맵핑 심화신경망 위에 바로 쌓이고, 결합된 상기 심화신경망 앙상블 기반의 음향 모델은 연결되어 재학습되는 단계를 포함할 수 있다.
- [0017] 다른 실시예에 따른 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델에 있어서, 입력되는 음성 신호로부터 특징 벡터를 추출하는 특징 벡터 추출부; 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 심화신경망 앙상블 기반의 음향 모델의 앙상블을 이용하여 결합하는 앙상블 모델; 및 음소를 분류하여 음성을 인식하는 음소 분류부를 포함하고, 상기 미리 학습된 심화신경망 앙상블 기반의 음향 모델은, 다수의 잔향 환경에서 각각의 잔향 환경에 대해 음소 확률을 추정할 수 있다.
- [0018] 상기 음성 신호로부터 잔향 시간을 추정하는 잔향 시간 예측부; 및 상기 잔향 시간을 기반으로 가중치를 산출하는 가중치 결정부를 더 포함하고, 상기 음소 분류부는, 산출된 상기 가중치를 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델의 앙상블 결합에 적용할 수 있다.
- [0019] 상기 앙상블 모델은, 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델에 기반한 앙상블 모델을 통과시켜 음소 확률을 추정할 수 있다.
- [0020] 상기 음소 분류부는, 상기 음소 확률과 상기 가중치를 이용하여 상기 심화신경망 앙상블 기반의 음향 모델의 사후 확률(posterior probability)을 산출하여 상기 음소를 분류할 수 있다.
- [0021] 상기 잔향 시간 예측부는, 상기 심화신경망 앙상블 기반의 음향 모델 중 우도비가 가장 큰 두 개의 음향 모델을 선택하여 최대 우도법(maximum likelihood)을 통한 상기 잔향 시간을 추정할 수 있다.
- [0022] 상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 학습부를 더 포함하고, 상기 학습부는, 학습 단계에서, 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출하는 특징 벡터 추출부; 추출된 상기 특징 벡터의 잔향 환경의 특징을 분리하는 잔향 특징 분류부; 및 추출된 상기 특징 벡터를 심화신경망을 통하여 각각의 상기 잔향 환경에 대해 학습시키는 다중 심화신경망 학습부를 포함할 수 있다.
- [0023] 상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 학습부를 더 포함하고, 상기 학습부는, 잔향이 없는 음성 특징으로 맵핑(mapping)시키는 특징 맵핑 심화신경망과 음향 모델링 심화신경망을 결합하는 구조를 이용하여, 서로 다른 잔향 환경에 대해 학습된 음향 모델을 구성할 수 있다.
- [0024] 상기 학습부는, 학습 단계에서, 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출하는 특징 벡터 추출부; 추출된 상기 특징 벡터의 잔향 환경의 특징을 분리하는 잔향 특징 분류부; 추출된 상기 특징 벡터를 잔향이 없는 음성 특징으로 맵핑(mapping)시키는 상기 특징 맵핑 심화신경망; 및 상기 특징 맵핑 심화신경망의 출력을 이용하여 학습시키는 상기 음향 모델링 심화신경망을 포함하고, 상기 음향 모델링 심화신경망은 상기 특징 맵핑 심화신경망 위에 바로 쌓이고, 결합된 상기 심화신경망 앙상블 기반의 음향 모델은 연결되어 재학습될 수 있다.

발명의 효과

- [0025] 실시예들에 따르면 심화신경망 앙상블 구조를 기반으로 테스트 환경에 최적화된 음향 모델의 음소 확률을 추정함으로써, 다양한 잔향 환경에서 우수한 인식성능을 가지는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델 및 이를 이용한 음성인식 방법을 제공할 수 있다.

[0026] 또한, 실시예들에 따르면 결합 학습된 심화신경망 모델을 앙상블에 이용하여 잔향 제거와 음향 모델이 통합된 하나의 심화신경망을 생성함으로써, 잔향 환경에 더욱 효과적인 음향 모델 및 이를 이용한 음성인식 방법을 제공할 수 있다.

도면의 간단한 설명

[0027] 도 1은 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템을 개략적으로 나타내는 블록도이다.

도 2는 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법을 나타내는 흐름도이다.

도 3은 다른 실시예에 따른 결합 학습 구조의 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템을 개략적으로 나타내는 블록도이다.

도 4는 다른 실시예에 따른 특징 맵핑 심화신경망의 입력과 출력을 설명하기 위한 도면이다.

도 5는 일 실시예에 따른 잔향 시뮬레이션 환경을 나타내는 도면이다.

도 6은 일 실시예에 따른 다양한 잔향 환경에서 종래의 배경 모델과 제안된 앙상블 모델의 단어 오인지율의 비교를 나타내는 그래프이다.

도 7은 일 실시예에 따른 잔향 환경에서의 음성에 대해 추정된 잔향 시간에 대한 최대 우도비 그래프를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0028] 이하, 첨부된 도면을 참조하여 실시예들을 설명한다. 그러나, 기술되는 실시예들은 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 이하 설명되는 실시예들에 의하여 한정되는 것은 아니다. 또한, 여러 실시예들은 당해 기술분야에서 평균적인 지식을 가진 자에게 본 발명을 더욱 완전하게 설명하기 위해서 제공되는 것이다. 도면에서 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다.

[0029] 아래의 실시예들은 결합 학습된 심화신경망 앙상블 음향 모델을 이용한 잔향 환경에서의 음성인식 방법에 관한 것으로서, 다양한 잔향 환경의 음성 신호를 효과적으로 모델링하기 위해서 잔향 시간 추정을 이용한 음향 모델 앙상블 구조와 결합 학습된 심화신경망 모델의 앙상블 구조를 이용한다.

[0030] 여기에서, 음향 모델(Acoustic Model)이란 음성 신호를 음소 데이터로 분류하는 기술로 음성인식의 필수적인 요소이다. 예를 들어, 음향 모델은 음성인식을 이용하여 편리한 기기 조작, 화자의 고유 정보 전달을 가능하게 하여 편의성을 증대시키며 보안, 금융, 의료 등 개인별 서비스 제공을 가능하게 하는 역할을 한다. 또한 음향 모델은 번역기 및 검색 서비스의 경우 음성인식 시스템을 이용하여 효율적인 실시간 정보 처리를 할 수 있다. 잔향 환경에 적합한 음향 모델은 실제 환경에서의 음성인식 성능을 향상시킬 수 있기 때문에 널리 연구되고 있다. 잔향에 의해 왜곡된 입력신호들은 단순한 심화신경망 구조 기반의 음향 모델로 충분히 모델링하는데 한계가 존재한다.

[0031] 이에 따라 실시예들은 입력신호들로부터 추출한 음성 특징들의 왜곡을 제거하는 특징 매핑 심화신경망과 음향 모델링 심화신경망을 통합하는 결합 학습 구조를 이용하여 각각의 잔향 환경에 대한 모델들의 음소 확률을 추정하고, 계산된 각 모델의 음소 확률의 앙상블을 이용하여 결합함으로써 음소를 분류할 수 있다.

[0032] 실시예들에 따르면 음성 신호로부터 추출한 특징 벡터들을 결합 구조 심화신경망을 통하여 각각의 잔향 환경에 대해 모델링하고, 이를 기반으로 테스트 환경에서의 최대 우도법(maximum likelihood)을 통한 잔향 시간 추정을 이용하여 음소 확률을 앙상블함으로써, 다양한 잔향 환경에서도 우수한 성능의 음성인식이 가능하다.

[0033] 아래에서 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델의 앙상블에 대해 설명한다.

[0034] 도 1은 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템을 개략적으로 나타내는 블록도이다.

[0035] 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템은 다양한 잔향 환경의 음성

신호를 효과적으로 모델링하기 위해서 잔향 시간 추정을 이용한 음향 모델 앙상블 구조를 이용할 수 있다. 이러한 심화신경망 앙상블 기반의 음향 모델을 기반으로 테스트 환경에 최적화된 음향 모델의 음소 확률을 추정함으로써, 다양한 잔향 환경에서 우수한 인식성능을 가지는 음성인식 방법을 제공할 수 있다.

- [0036] 특히, 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템은 최대 우도법을 이용한 실제 잔향 환경 추정을 기반으로 상이한 가중치를 산출하고, 이를 각각의 잔향 환경에 대해 학습된 음향 모델의 결합에 적용하는 앙상블 모델을 이용할 수 있다.
- [0037] 도 1을 참조하면, 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템(100)은 학습부(110) 및 테스트부(120)를 포함하여 이루어질 수 있다.
- [0038] 학습부(110)는 상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 것으로, 특징 벡터 추출부(112), 잔향 특징 분류부(114), 및 다중 심화신경망 학습부(115)를 포함하여 이루어질 수 있다. 또한 실시예에 따라 학습부(110)는 음성 입력부(111) 및 심화신경망 학습부(113)를 더 포함하여 이루어질 수 있다.
- [0039] 특징 벡터 추출부(112)는 학습 단계에서 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출할 수 있다.
- [0040] 잔향 특징 분류부(114)는 학습 단계에서 추출된 상기 특징 벡터의 잔향 환경의 특징을 분리시킬 수 있다.
- [0041] 다중 심화신경망 학습부(115)는 학습 단계에서 추출된 상기 특징 벡터를 심화신경망을 통하여 각각의 상기 잔향 환경에 대해 학습시킬 수 있다.
- [0042] 테스트부(120)는 학습부(110)에서 미리 학습된 심화신경망 앙상블 기반의 음향 모델을 기반으로 최적화된 음향 모델의 음소 확률을 추정함으로써, 다양한 잔향 환경에서 우수한 인식성능을 가지는 음성인식 방법을 제공할 수 있다. 이러한 테스트부(120)는 특징 벡터 추출부(122), 잔향 시간 예측부(123), 가중치 결정부(124), 앙상블 모델(125), 및 음소 분류부(126)를 포함하여 이루어질 수 있다. 또한 실시예에 따라 테스트부(120)는 음성 입력부(121) 및 배경 모델(127)을 더 포함하여 이루어질 수 있다.
- [0043] 특징 벡터 추출부(122)는 입력되는 음성 신호로부터 특징 벡터를 추출할 수 있다.
- [0044] 잔향 시간 예측부(123)는 음성 신호로부터 잔향 시간을 추정할 수 있으며, 상기 심화신경망 앙상블 기반의 음향 모델 중 우도비가 가장 큰 두 개의 음향 모델을 선택하여 최대 우도법(maximum likelihood)을 통한 상기 잔향 시간을 추정할 수 있다.
- [0045] 가중치 결정부(124)는 잔향 시간을 기반으로 가중치를 산출할 수 있다.
- [0046] 앙상블 모델(125)은 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 심화신경망 앙상블 기반의 음향 모델의 앙상블을 이용하여 결합할 수 있다.
- [0047] 여기에서 미리 학습된 심화신경망 앙상블 기반의 음향 모델은 다수의 잔향 환경에서 각각의 잔향 환경에 대해 음소 확률을 추정할 수 있다.
- [0048] 앙상블 모델(125)은 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델에 기반한 앙상블 모델을 통과시켜 음소 확률을 추정할 수 있다.
- [0049] 음소 분류부(126)는 음소를 분류하여 음성을 인식할 수 있으며, 음소 분류부(126)는 가중치 결정부(124)에서 산출된 상기 가중치를 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델의 앙상블 결합에 적용할 수 있다.
- [0050] 이러한 음소 분류부(126)는 상기 음소 확률과 상기 가중치를 이용하여 상기 심화신경망 앙상블 기반의 음향 모델의 사후 확률(posterior probability)을 산출하여 상기 음소를 분류할 수 있다.
- [0051] 아래에서 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델에 대해 하나의 예를 들어 더 구체적으로 설명한다.
- [0052] 도 2는 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법을 나타내는 흐름도이다.
- [0053] 일 실시예에 따른 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법은 입력되는 음성 신호로부터 특징 벡터를 추출하는 단계(210), 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 심화신경망 앙상블 기반의 음향 모델의 앙상블을 이용하여 결합하는 단계(240), 및 음소

를 분류하여 음성을 인식하는 단계(250)를 포함하여 이루어질 수 있다. 여기에서 미리 학습된 심화신경망 앙상블 기반의 음향 모델은 다수의 잔향 환경에서 각각의 잔향 환경에 대해 음소 확률을 추정할 수 있다.

- [0054] 그리고, 일 실시예에 따른 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법은 상기 음성 신호로부터 잔향 시간을 추정하는 단계(220) 및 상기 잔향 시간을 기반으로 가중치를 산출하는 단계(230)를 더 포함하여 이루어질 수 있다.
- [0055] 이때, 상기 음소를 분류하여 음성을 인식하기 위해서 산출된 상기 가중치를 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델의 앙상블 결합에 적용할 수 있다.
- [0056] 이와 같이 실시예들은 심화신경망 앙상블 구조를 기반으로 테스트 환경에 최적화된 음향 모델의 음소 확률을 추정함으로써, 다양한 잔향 환경에서 우수한 인식성능을 가지는 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델 및 이를 이용한 음성인식 방법을 제공하는데 있다.
- [0057] 아래에서는 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법의 각 단계에 대해 상세히 설명하기로 한다. 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법은 도 1에서 설명한 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템을 이용하여 더 구체적으로 설명할 수 있다.
- [0058] 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템(100)은 학습부(110) 및 테스트부(120)를 포함하여 이루어질 수 있다.
- [0059] 여기에서 학습부(110)는 상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 것으로, 특징 벡터 추출부(112), 잔향 특징 분류부(114), 및 다중 심화신경망 학습부(115)를 포함하여 이루어질 수 있다. 또한 실시예에 따라 학습부(110)는 음성 입력부(111) 및 심화신경망 학습부(113)를 더 포함하여 이루어질 수 있다.
- [0060] 테스트부(120)는 학습부(110)에서 미리 학습된 심화신경망 앙상블 기반의 음향 모델을 기반으로 최적화된 음향 모델의 음소 확률을 추정하는 것으로, 특징 벡터 추출부(122), 잔향 시간 예측부(123), 가중치 결정부(124), 앙상블 모델(125), 및 음소 분류부(126)를 포함하여 이루어질 수 있다. 또한 실시예에 따라 테스트부(120)는 음성 입력부(121) 및 배경 모델(127)을 더 포함하여 이루어질 수 있다.
- [0061] 먼저, 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법은 상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 단계를 포함하여 이루어질 수 있다.
- [0062] 학습 단계에서, 특징 벡터 추출부(112)는 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출할 수 있으며, 잔향 특징 분류부(114)는 추출된 상기 특징 벡터의 잔향 환경의 특징을 분리시킬 수 있다. 이후, 다중 심화신경망 학습부(115)는 추출된 상기 특징 벡터를 심화신경망을 통하여 각각의 상기 잔향 환경에 대해 학습시킬 수 있다.
- [0063] 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법은 다양한 잔향 환경의 음성 신호를 효과적으로 모델링하기 위해서 잔향 시간 추정을 이용한 음향 모델 앙상블 구조를 이용할 수 있다. 이후, 이러한 심화신경망 앙상블 기반의 음향 모델을 기반으로 테스트 환경에 최적화된 음향 모델의 음소 확률을 추정함으로써, 다양한 잔향 환경에서 우수한 인식성능을 가지는 음성인식 방법을 제공할 수 있다.
- [0064] 단계(210)에서, 테스트부(120)의 특징 벡터 추출부(122)는 입력되는 음성 신호로부터 특징 벡터를 추출할 수 있다.
- [0065] 한편, 단계(220)에서 잔향 시간 예측부(123)는 음성 신호로부터 잔향 시간을 추정할 수 있다. 더 구체적으로 잔향 시간 예측부(123)는 심화신경망 앙상블 기반의 음향 모델 중 우도비가 가장 큰 두 개의 음향 모델을 선택하여 최대 우도법(maximum likelihood)을 통한 상기 잔향 시간을 추정할 수 있다.
- [0066] 또한 단계(230)에서 가중치 결정부(124)는 잔향 시간을 기반으로 가중치를 산출할 수 있다.
- [0067] 단계(240)에서 앙상블 모델(125)은 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 심화신경망 앙상블 기반의 음향 모델의 앙상블을 이용하여 결합할 수 있다.
- [0068] 여기에서 미리 학습된 심화신경망 앙상블 기반의 음향 모델은 다수의 잔향 환경에서 각각의 잔향 환경에 대해 음소 확률을 추정할 수 있다.

- [0069] 양상블 모델(125)은 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 상기 심화신경망 양상블 기반의 음향 모델에 기반한 양상블 모델을 통과시켜 음소 확률을 추정할 수 있다.
- [0070] 단계(250)에서 음소 분류부(126)는 음소를 분류하여 음성을 인식할 수 있으며, 더 구체적으로 음소 분류부(126)는 가중치 결정부(124)에서 산출된 상기 가중치를 미리 학습된 상기 심화신경망 양상블 기반의 음향 모델의 양상블 결합에 적용할 수 있다.
- [0071] 이러한 음소 분류부(126)는 상기 음소 확률과 상기 가중치를 이용하여 상기 심화신경망 양상블 기반의 음향 모델의 사후 확률(posterior probability)을 산출하여 상기 음소를 분류할 수 있다.
- [0072] 서로 다른 잔향 환경에 대해 각각 학습된 N 개의 음향 모델의 양상블을 이용한 양상블 음향 모델의 최종 사후 확률(posterior probability)은 다음 수학적 식 1과 같이 표현될 수 있다.

수학적 식 1

$$P(k_{\text{EAM}} | x) = \sum_{n=1}^N w_n p_n$$

- [0073]
- [0074] 여기에서 n 은 음향 모델 인덱스, k 는 히든 마르코브 모델(Hidden Markov Model, HMM) 상태(state), x 는 음성 특징 벡터 인덱스를 각각 나타낼 수 있다. $p_n = p(k_n | x)$ 와 w_n 은 n 번째 음향 모델의 출력 확률과 가중치(weight)를 나타내며, 가중치 w_n 은 잔향 시간 추정에 의해서 결정될 수 있다.
- [0075] 잔향은 수학적 식 2와 같이 간단한 잡음 감소 곡선(noise decay curve)으로 수학적으로 모델링 가능하다.

수학적 식 2

$$d(k) = A_r v(k) e^{-\rho k T_s} \epsilon(k)$$

- [0076]
- [0077] 여기에서 A_r , ρ , $\epsilon(k)$ 는 각각 실제 진폭, 감소율, 단위 계단 시퀀스를 의미할 수 있다. 또한 T_s 는 샘플링 단위, $v(k)$ 는 확률 변수 시퀀스를 나타낼 수 있다.
- [0078] 그리고 $d(k)$ 는 $k \in \{0, 1, \dots, N-1\}$ 에 대해 N 개의 독립적 확률 변수로 모델링 되므로, 최대 우도법을 잔향 시간 추정에 사용 가능하다. 따라서 로그 우도비 함수는 다음 수학적 식 3과 같이 나타낼 수 있다.

수학적 식 3

$$\mathcal{L}(\rho) = -\frac{N}{2} \left((N-1) \ln(a) + \ln \left(\frac{2\pi}{N} \sum_{i=1}^{N-1} a^{-2i} d^2(i) \right) + 1 \right)$$

- [0079]
- [0080] 이 때, $a = e^{-T_s \rho}$ 이며, 감소율 ρ 는 로그 우도비의 최대값을 찾음으로써 추정 가능하고 다음 수학적 식 4와 같이 구할 수 있다.

수학식 4

$$\hat{\rho}^{(\text{ML})} = \max_{\rho} \{ \mathcal{L}(\rho) \}$$

여기에서 $\hat{\rho}^{(\text{ML})}$ 은 수학식 5와 같이 잔향 시간으로 변환할 수 있다.

수학식 5

$$\text{RT}_{60} = \frac{3}{\rho \log_{10} e} \approx \frac{6.908}{\rho}$$

이에 따라 서로 다른 잔향 환경에 대해 학습된 N 개의 음향 모델 중 우도비가 가장 큰 두 개의 음향 모델 m_1 과 m_2 이 실제 잔향 시간에 가장 가깝다. 두 개의 음향 모델 각각에 대한 가중치는 수학식 3을 이용하여 구한 우도비의 비율로 결정되며 다음 수학식 6과 같이 표현될 수 있다.

수학식 6

$$w_{m_1} = \frac{\mathcal{L}(\rho_{m_1})}{\mathcal{L}(\rho_{m_1}) + \mathcal{L}(\rho_{m_2})}$$

$$w_{m_2} = \frac{\mathcal{L}(\rho_{m_2})}{\mathcal{L}(\rho_{m_1}) + \mathcal{L}(\rho_{m_2})}$$

여기에서 w_{m_1} 는 우도비가 가장 큰 음향 모델 m_1 의 가중치이고, w_{m_2} 는 우도비가 두 번째로 큰 음향 모델 m_2 의 가중치가 될 수 있다. 이 때, $m_1, m_2 \in \{1, 2, \dots, N\}$ 이며, N 은 5가 될 수 있다.

아래에서 다른 실시예에 따른 결합 학습 구조의 심화신경망 앙상블 기반의 음향 모델의 앙상블에 대해 설명한다.

도 3은 다른 실시예에 따른 결합 학습 구조의 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템을 개략적으로 나타내는 블록도이다.

다른 실시예에 따른 결합 학습 구조의 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템은 다양한 잔향 환경의 음성 신호를 효과적으로 모델링하기 위해서 잔향 시간 추정을 이용한 결합 학습된 심화신경망 모델의 앙상블 구조를 이용할 수 있다. 이러한 결합 학습된 심화신경망 모델을 앙상블에 이용하여 잔향 제거와 음향 모델이 통합된 하나의 심화신경망을 생성함으로써, 잔향 환경에 더욱 효과적인 모델을 구성할 수 있다.

특히, 다른 실시예에 따른 결합 학습 구조의 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템은 최대 우도법을 이용한 실제 잔향 환경 추정을 기반으로 상이한 가중치를 산출하고 이를 각각의 잔향 환경에 대해 결합 학습된 모델의 결합에 적용하는 앙상블 결합 모델을 이용할 수 있다.

도 3을 참조하면, 다른 실시예에 따른 결합 학습 구조의 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인

식 시스템은 서로 다른 잔향 환경에 대해 학습된 N 개의 음향 모델을 구성할 때, 기존의 음향 모델 학습법과 달리 결합 학습법을 이용할 수 있다. 결합 학습법은 잔향 음성 특징을 잔향이 없는 음성 특징으로 맵핑(mapping)시키는 특징 맵핑 심화신경망과 음향 모델링 심화신경망을 결합하는 구조이다.

- [0092] 다른 실시예에 따른 결합 학습 구조의 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템(300)은 학습부(310) 및 테스트부(320)를 포함하여 이루어질 수 있다.
- [0093] 학습부(310)는 상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 것으로, 특징 벡터 추출부(312), 잔향 특징 분류부(314), 및 다중 결합 학습부(315)를 포함하여 이루어질 수 있다. 또한 실시예에 따라 학습부(310)는 음성 입력부(311) 및 결합 학습부(313)를 더 포함하여 이루어질 수 있다. 결합 학습부(313)는 하나의 특징 맵핑 심화신경망과 음향 모델링 심화신경망을 포함하여 이루어질 수 있다.
- [0094] 이러한 학습부(310)는 잔향이 없는 음성 특징으로 맵핑(mapping)시키는 특징 맵핑 심화신경망과 음향 모델링 심화신경망을 결합하는 구조를 이용하여, 서로 다른 잔향 환경에 대해 학습된 음향 모델을 구성할 수 있다.
- [0095] 특징 벡터 추출부(312)는 학습 단계에서 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출할 수 있다.
- [0096] 잔향 특징 분류부(314)는 학습 단계에서 추출된 상기 특징 벡터의 잔향 환경의 특징을 분리시킬 수 있다.
- [0097] 다중 결합 학습부(315)는 특징 맵핑 심화신경망(315a)과 음향 모델링 심화신경망(315b)을 포함하여 이루어질 수 있다.
- [0098] 특징 맵핑 심화신경망(315a)은 추출된 상기 특징 벡터를 잔향이 없는 음성 특징으로 맵핑(mapping)시킬 수 있다. 그리고 음향 모델링 심화신경망(315b)은 특징 맵핑 심화신경망(315a)의 출력을 이용하여 학습시킬 수 있다.
- [0099] 음향 모델링 심화신경망(315b)은 특징 맵핑 심화신경망(315a) 위에 바로 쌓이고, 결합된 상기 심화신경망 앙상블 기반의 음향 모델은 연결되어 재학습될 수 있다.
- [0100] 테스트부(320)는 학습부(310)에서 미리 학습된 심화신경망 앙상블 기반의 음향 모델을 기반으로 최적화된 음향 모델의 음소 확률을 추정함으로써, 다양한 잔향 환경에서 우수한 인식성능을 가지는 음성인식 방법을 제공할 수 있다. 이러한 테스트부(320)는 특징 벡터 추출부(322), 잔향 시간 예측부(323), 가중치 결정부(324), 앙상블 모델(325), 및 음소 분류부(326)를 포함하여 이루어질 수 있다. 또한 실시예에 따라 테스트부(320)는 음성 입력부(321) 및 배경 모델(327)을 더 포함하여 이루어질 수 있다.
- [0101] 특징 벡터 추출부(322)는 입력되는 음성 신호로부터 특징 벡터를 추출할 수 있다.
- [0102] 잔향 시간 예측부(323)는 음성 신호로부터 잔향 시간을 추정할 수 있으며, 상기 심화신경망 앙상블 기반의 음향 모델 중 우도비가 가장 큰 두 개의 음향 모델을 선택하여 최대 우도법(maximum likelihood)을 통한 상기 잔향 시간을 추정할 수 있다.
- [0103] 가중치 결정부(324)는 잔향 시간을 기반으로 가중치를 산출할 수 있다.
- [0104] 앙상블 모델(325)은 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 심화신경망 앙상블 기반의 음향 모델의 앙상블을 이용하여 결합할 수 있다.
- [0105] 여기에서 미리 학습된 심화신경망 앙상블 기반의 음향 모델은 다수의 잔향 환경에서 각각의 잔향 환경에 대해 음소 확률을 추정할 수 있다.
- [0106] 앙상블 모델(325)은 상기 특징 벡터를 각각의 잔향 환경에 대해 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델에 기반한 앙상블 모델을 통과시켜 음소 확률을 추정할 수 있다.
- [0107] 음소 분류부(326)는 음소를 분류하여 음성을 인식할 수 있으며, 음소 분류부(326)는 가중치 결정부(324)에서 산출된 상기 가중치를 미리 학습된 상기 심화신경망 앙상블 기반의 음향 모델의 앙상블 결합에 적용할 수 있다.
- [0108] 이러한 음소 분류부(326)는 상기 음소 확률과 상기 가중치를 이용하여 상기 심화신경망 앙상블 기반의 음향 모델의 사후 확률(posterior probability)을 산출하여 상기 음소를 분류할 수 있다.
- [0109] 이와 같이 최대 우도법을 이용한 잔향 시간 추정을 통해 다양한 잔향 환경에 대해 학습된 모델들을 가중치를 두어 결합하는 앙상블 구조와 이를 확장하여 종래의 음향 모델 학습법 대신 결합 학습법을 적용하여 학습시킨 모

델들을 결합하는 결합 앙상블 구조를 이용하여 음성인식 성능을 확인할 수 있다.

- [0110] 한편, 다른 실시예에 따른 결합 학습 구조의 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법은 다른 실시예에 따른 결합 학습 구조의 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 시스템을 이용하여 수행될 수 있다. 다른 실시예에 따른 결합 학습 구조의 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법은 도 2에서 설명한 일 실시예에 따른 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법과 유사하므로, 차이점을 중심으로 설명하기로 한다.
- [0111] 다른 실시예에 따른 결합 학습 구조의 심화신경망 앙상블 기반의 음향 모델을 이용한 음성인식 방법은 상기 심화신경망 앙상블 기반의 음향 모델을 학습시키는 단계를 포함하여 이루어질 수 있다.
- [0112] 학습 단계에서, 특징 벡터 추출부(312)는 다수의 잔향 환경에서의 음성 신호를 입력 받아 특징 벡터를 추출할 수 있으며, 잔향 특징 분류부(314)는 추출된 상기 특징 벡터의 잔향 환경의 특징을 분리시킬 수 있다.
- [0113] 이후, 특징 맵핑 심화신경망(315a)은 추출된 상기 특징 벡터를 잔향이 없는 음성 특징으로 맵핑(mapping)시킬 수 있다. 그리고 음향 모델링 심화신경망(315b)은 특징 맵핑 심화신경망(315a)의 출력을 이용하여 학습시킬 수 있다.
- [0114] 이러한 음향 모델링 심화신경망(315b)은 특징 맵핑 심화신경망(315a) 위에 바로 쌓이고, 결합된 상기 심화신경망 앙상블 기반의 음향 모델은 연결되어 재학습될 수 있다. 이때 다중 결합 학습부(315)는 특징 맵핑 심화신경망(315a)과 음향 모델링 심화신경망(315b)을 포함하여 이루어질 수 있다.
- [0115] 도 4는 다른 실시예에 따른 특징 맵핑 심화신경망의 입력과 출력을 설명하기 위한 도면이다.
- [0116] 도 4를 참조하면, (a) 잔향이 없는 음성의 특징, (b) 잔향 시간이 0.3 s인 음성의 특징, (c) 잔향 시간이 0.5 s인 음성의 특징, (d) 잔향 시간이 0.7 s인 음성의 특징, (e) 잔향 시간이 0.3 s인 음성의 특징 맵핑 심화신경망 출력, (f) 잔향 시간이 0.5 s인 음성의 특징 맵핑 심화신경망 출력, 및 (g) 잔향 시간이 0.7 s인 음성의 특징 맵핑 심화신경망 출력을 나타내는 것을 확인할 수 있다.
- [0117] 모든 잔향 환경에서 특징 맵핑 층을 통과하면서 잔향에 의해 발생한 번짐이 감소하고, 프레임간의 경계가 회복될 수 있다.
- [0118] 다음으로, 특징 맵핑 층의 출력을 이용하여 음향 모델링 층을 학습시킬 수 있다. 음향 모델링 심화신경망은 특징 맵핑 심화신경망 위에 바로 쌓이고, 결합된 심화신경망은 연결되어 재학습될 수 있다. 이렇게 결합 학습법을 이용하여 학습된 N 개의 심화신경망 모델은 앙상블 구조에 적용되고, 앙상블 가중치는 수학적식 7과 같이 나타낼 수 있다.

수학적식 7

$$w_{m_1} = \frac{\mathcal{L}(\rho_{m_1})}{\mathcal{L}(\rho_{m_1}) + \mathcal{L}(\rho_{m_2})}$$

$$w_{m_2} = \frac{\mathcal{L}(\rho_{m_2})}{\mathcal{L}(\rho_{m_1}) + \mathcal{L}(\rho_{m_2})}$$

- [0119]
- [0120] 여기에서 w_{m_1} 는 우도비가 가장 큰 음향 모델 m_1 의 가중치이고, w_{m_2} 는 우도비가 두 번째로 큰 음향 모델 m_2 의 가중치가 될 수 있다. 이 때, $m_1, m_2 \in \{1, 2, \dots, N\}$ 이며, N 은 5가 될 수 있다.
- [0121] 그리고 결합 학습된 앙상블 모델의 최종 사후 확률은 수학적식 8과 같이 표현될 수 있으며, 이러한 최종 사후 확률은 음소 구분에 사용할 수 있다.

수학식 8

$$P(k_{\text{EAM}} | x) = \sum_{n=1}^N w_n p_n$$

여기에서 n 은 음향 모델 인덱스, k 는 히든 마르코브 모델(Hidden Markov Model, HMM) 상태(state), x 는 음성 특징 벡터 인덱스를 각각 나타낼 수 있다. $p_n = p(k_n | x)$ 와 w_n 은 n 번째 음향 모델의 출력 확률과 가중치(weight)를 나타내며, 가중치 w_n 은 잔향 시간 추정에 의해서 결정될 수 있다.

본 실시예의 구조의 성능을 검증하기 위해 다양한 잔향 환경에서 실험을 진행할 수 있다.

도 5는 일 실시예에 따른 잔향 시뮬레이션 환경을 나타내는 도면이다.

도 5를 참조하면, 훈련을 위해 5 가지의 잔향 시간 0.3 s, 0.4 s, 0.5 s, 0.6 s, 0.7 s에서 각각의 환경마다 3,320개의 발화를 이용하고, 타당성 확인을 위해 동일한 5 가지 잔향 시간에 대해 각각 376개의 발화를 이용할 수 있다. 실험을 위해서는 잔향 시간 0.3 s부터 0.7 s까지 0.01 s 간격으로 50 가지의 잔향 환경에 대해 각각 192개의 발화를 이용할 수 있다. 특징 벡터로 72 차원(dimension) 필터뱅크(filterbank) 특징을 사용하고, 프레임 길이는 25 ms, 프레임 이동은 10 ms를 사용할 수 있다. 또한 2,021개의 tied-state triphone HMM을 이용할 수 있다. 음성 인식 성능 실험은 칼디(Kaldi)를 사용할 수 있다.

표 1은 다양한 잔향 환경에서 단일 모델과 기본 모델의 단어 오인율 비교를 나타내는 것이다.

표 1

Model	Background Model	Base Acoustic Model					Base Joint Model				
RT ₆₀ Condition	0.3 s - 0.7 s	0.3 s	0.4 s	0.5 s	0.6 s	0.7 s	0.3 s	0.4 s	0.5 s	0.6 s	0.7 s
0.3 s	29.5	27.0	32.9	38.6	42.1	44.9	26.5	32.7	39.0	43.5	46.1
0.4 s	31.2	34.2	28.9	32.8	36.6	39.7	34.0	28.5	32.9	37.8	40.2
0.5 s	33.5	42.7	33.7	31.1	33.5	36.5	43.0	33.9	30.9	34.6	37.4
0.6 s	36.2	48.3	40.6	34.3	33.3	34.4	48.4	41.5	34.3	33.7	35.4
0.7 s	38.9	52.0	45.4	39.5	35.3	35.2	52.4	46.2	39.8	36.3	35.4

표 1을 참조하면, 다양한 잔향 환경에서 기존 단일 음향 모델과 각 잔향 환경에 대한 기본 모델들의 단어 오인율(word error rate)(%) 성능을 비교할 수 있다.

배경 모델(background model)은 기존의 단일 음향 모델이고, 기본 음향 모델(base acoustic model)은 각각의 잔향 환경에 대한 음향 모델이며, 기본 결합 모델(base joint model)은 결합 학습법을 이용하여 각각의 잔향 환경에 대해 학습시킨 모델을 의미한다. 표 1에서 가장 뛰어난 성능을 보이는 경우는 진하게 표시되었다.

배경 모델과 각각의 기본 음향 모델은 7개의 은닉층과 1개의 소프트맥스(softmax) 층으로 구성되어 있으며, 각 은닉층은 2,048개의 노드로 이루어질 수 있다.

기본 결합 모델은 2,048개의 노드로 이루어진 3개의 은닉층으로 구성된 특징 맵핑 층, 2,048개의 노드로 이루어진 7개의 은닉층, 1개의 소프트맥스(softmax) 층으로 구성될 수 있다.

모든 잔향 환경에서 종래의 단일 음향 모델에 비해 각각의 잔향 환경에 대해 구성한 기본 모델들의 성능이 뛰어나다는 것을 알 수 있으며, 그 중에서도 특히 일반적인 잔향 환경에서 결합 학습된 모델의 성능이 뛰어남을 확인할 수 있다.

- [0134] 도 6은 일 실시예에 따른 다양한 잔향 환경에서 종래의 배경 모델과 제안된 앙상블 모델의 단어 오인지율의 비교를 나타내는 그래프이다.
- [0135] 도 6을 참조하면, 기존의 단일 음향 모델과 제안하는 앙상블 음향 모델, 결합 학습된 앙상블 모델에 대한 단어 오인지율 비교할 수 있다. 이 그래프는 단어 오인지율과 잔향 시간을 각각 y축과 x축으로 하여, 보다 그래프가 낮을수록 높은 인식 정확도를 나타낼 수 있다.
- [0136] 본래의 배경 모델(original background model)은 기존의 단일 음향 모델을 나타내며, 깊은 배경 모델(deep background model)은 앙상블 모델과 연산량을 맞추기 위해 3,390개의 노드와 10개의 은닉층으로 증가시킨 단일 음향 모델을 나타내고, 앙상블 모델(ensemble model)은 기본 음향 모델들의 앙상블 구조를 나타내며, 앙상블 결합 모델(ensemble joint model)은 결합 학습된 모델들의 앙상블 구조를 의미할 수 있다.
- [0137] 모든 잡음 상황에서 제안한 앙상블 모델과 결합 학습된 앙상블 모델이 뛰어난 성능을 나타내는 것을 확인할 수 있다. 같은 연산량에도 불구하고 깊은 배경 모델에 비해 앙상블 모델의 인식 성능이 더 뛰어났고, 앙상블 구조에 결합 학습법을 적용함으로써 일반적인 잔향 환경에서의 성능을 더욱 향상시킬 수 있다.
- [0138] 도 7은 일 실시예에 따른 잔향 환경에서의 음성에 대해 추정된 잔향 시간에 대한 최대 우도비 그래프를 나타낸다.
- [0139] 도 7을 참조하면, 잔향 환경에서의 음성에 대해 추정된 잔향 시간에 대한 최대 우도비 그래프는 최대 우도비와 잔향 시간을 각각 y축과 x축으로 하여, 보다 그래프가 위에 위치할수록 높은 확률을 의미할 수 있다.
- [0140] 그래프는 5 가지의 잔향 시간에 대한 각각의 최대 우도비를 나타내며, (a) 잔향 시간이 0.61 s인 음성에 대한 최대 우도비, (b) 잔향 시간이 0.63 s인 음성에 대한 최대 우도비, (c) 잔향 시간이 0.67 s인 음성에 대한 최대 우도비를 보인다.
- [0141] 모든 그래프에서 5가지 잔향 시간 중에서 잔향 시간이 0.6 s일 확률과 0.7 s일 확률이 가장 높다. 잔향 시간이 0.61 s인 음성과 0.63 s인 음성에서는 0.6 s일 확률이 더 높고, 0.67 s인 음성에서는 0.7 s일 확률이 더 높다. 잔향 환경 추정은 일반적으로 정확하므로 이를 앙상블 가중치로 사용할 수 있다.
- [0142] 기존 단일 음향 모델 구조는 학습에 사용하지 않은 잔향 환경에서 음성인식성능을 저하시킨다. 실시예들은 앙상블 구조와 결합 학습법을 적용함으로써 보다 다양한 잔향 환경에 강인하게 함으로써 다양한 기기 및 시스템에 적용하여 음성인식의 정확도를 높일 수 있을 것이다.
- [0143] 음향 모델은 음성인식에 적용되어 실제 잔향 및 잡음이 존재하는 환경에서의 인식성능을 향상시키는 역할을 할 수 있다. 또한 여러 기기들과 시스템에 적용되어 다양한 환경에서 음성인식을 이용한 편리한 조작과 서비스 제공이 가능하다.
- [0144] 따라서, 실시예들에 따른 잔향 환경에서의 음성인식을 위한 결합 학습된 심화신경망 앙상블 기반의 음향 모델 및 이를 이용한 음성인식 방법은 휴대폰 단말기, 무선통신사업자, 카카오톡 등의 음성통화 서비스, 구글 보이스, 시리 등의 음성인식 서비스뿐만 아니라 다양한 실제 환경에서의 음성 신호처리 분야에 적용되어 보다 우수한 성능을 도출할 수 있다.
- [0145] 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, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable 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 장치

는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0146] 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 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 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.

[0147] 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. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0148] 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.

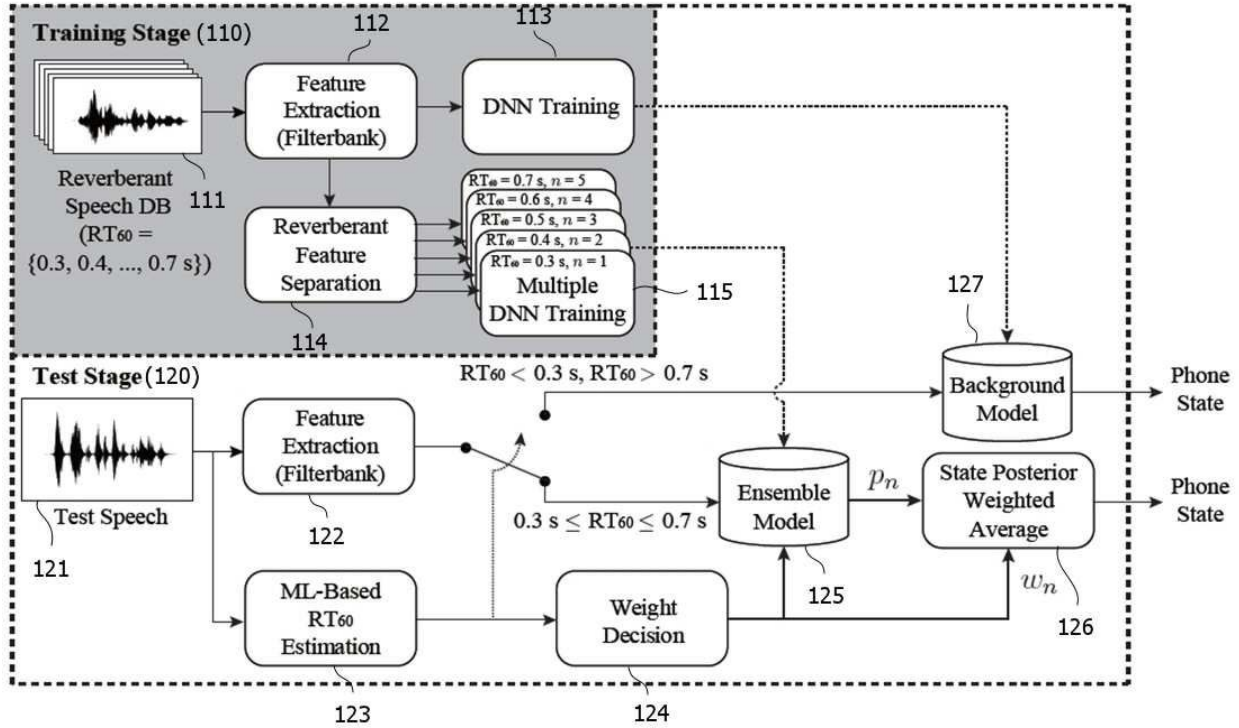
[0149] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

[0150]

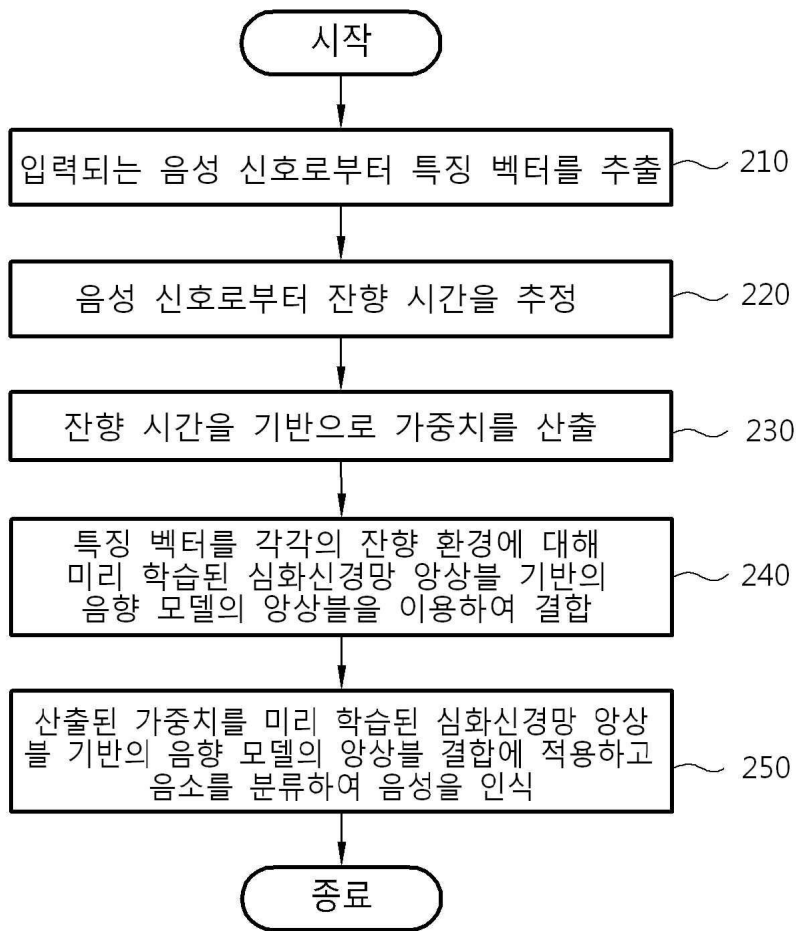
도면

도면1

100

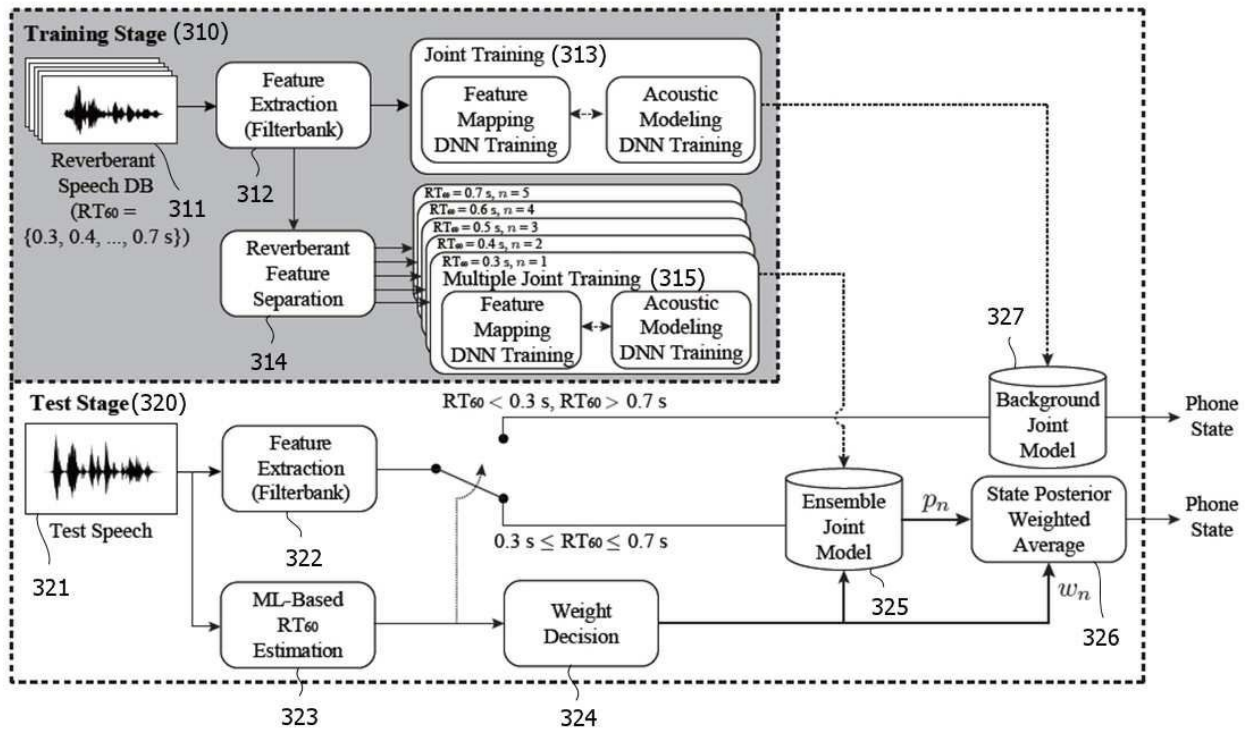


도면2

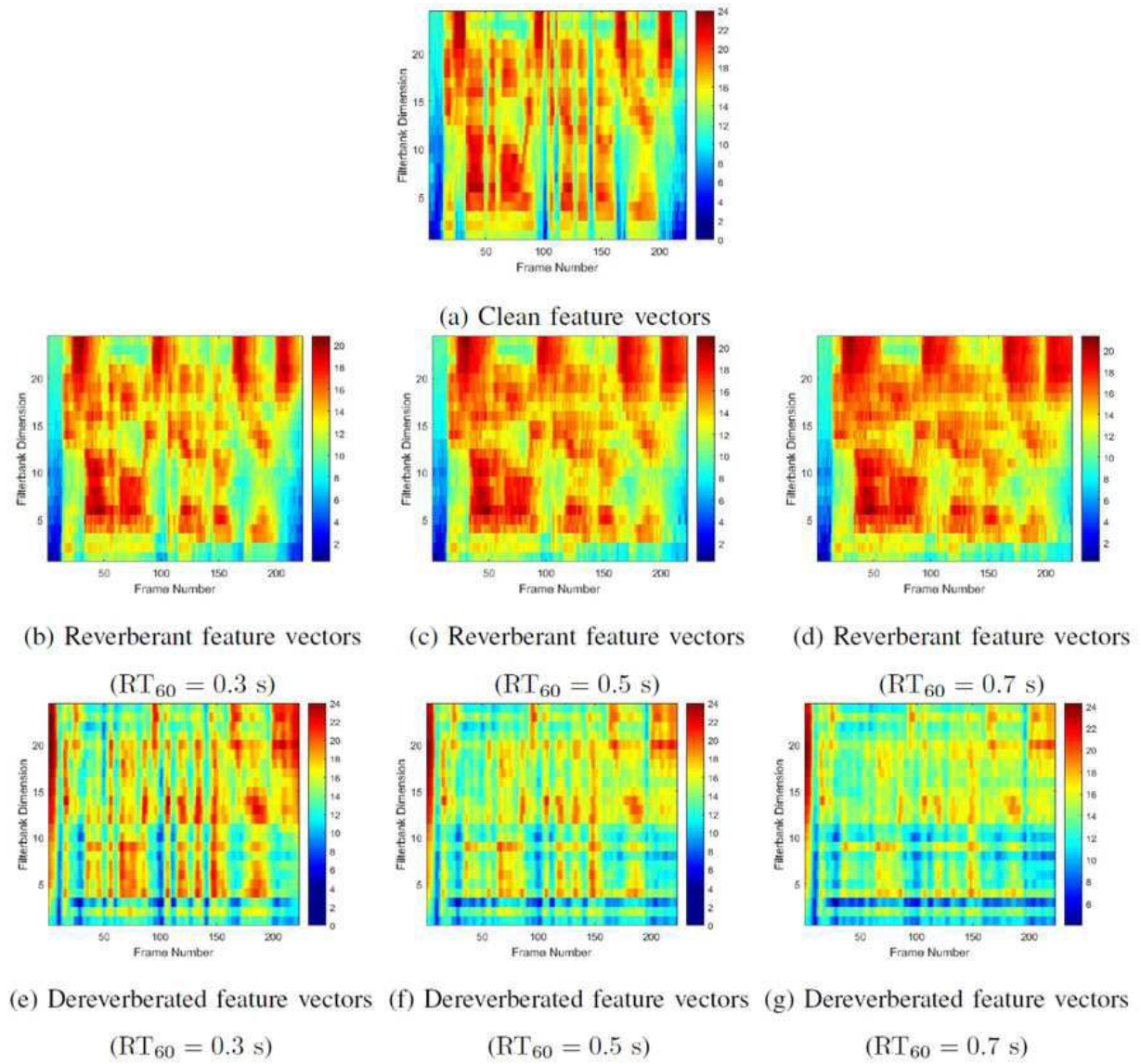


도면3

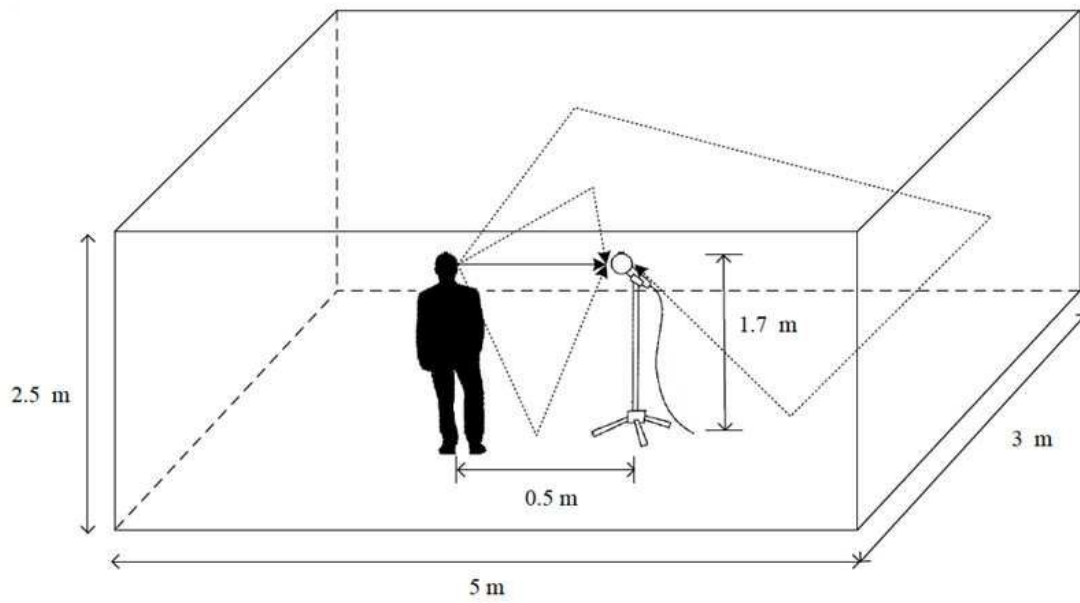
300



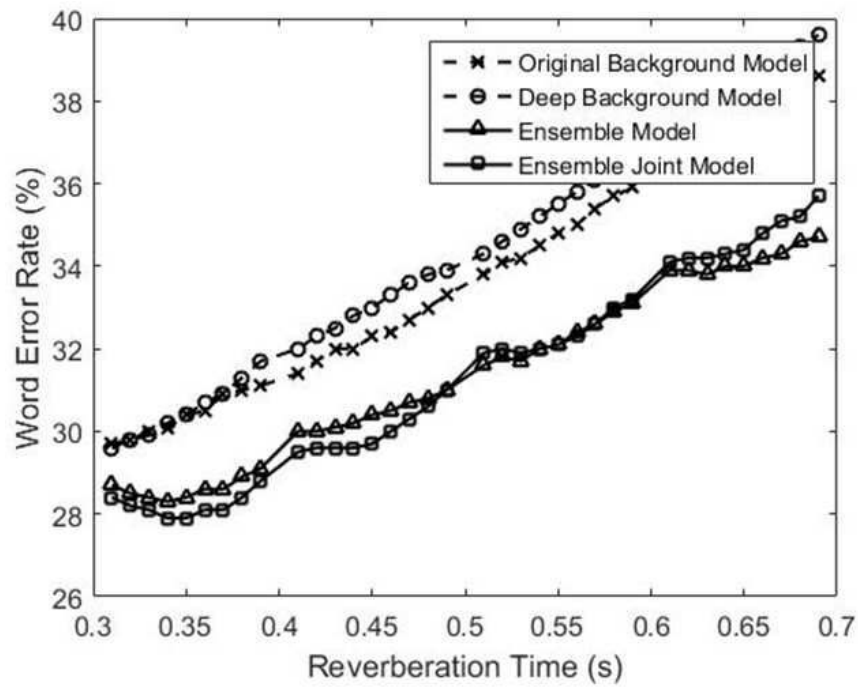
도면4



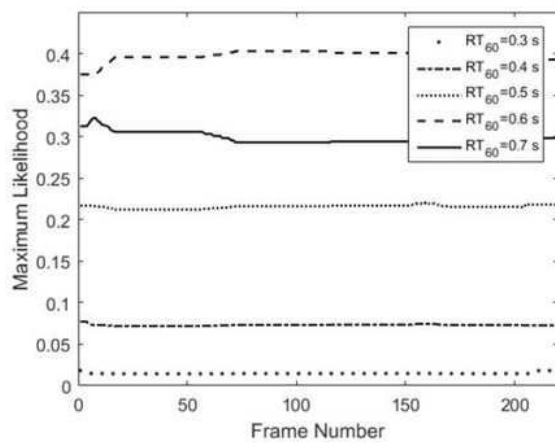
도면5



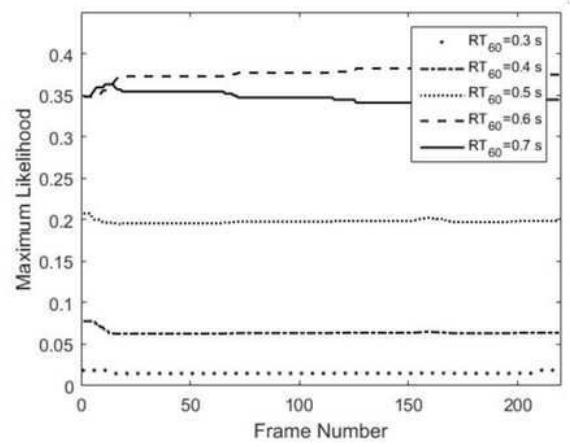
도면6



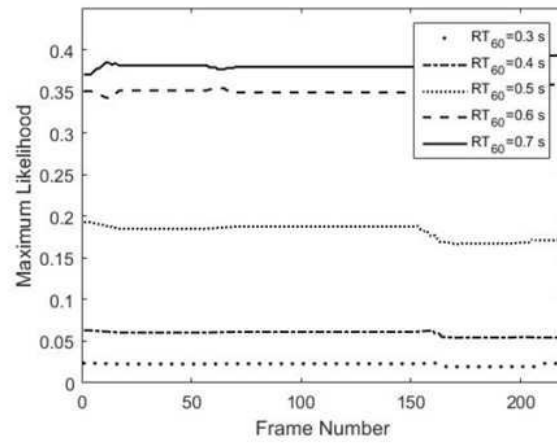
도면7



(a) ML ($RT_{60} = 0.61$ s)



(b) ML ($RT_{60} = 0.63$ s)



(c) ML ($RT_{60} = 0.67$ s)