



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년07월21일
(11) 등록번호 10-2136464
(24) 등록일자 2020년07월15일

(51) 국제특허분류(Int. Cl.)
G10L 25/18 (2013.01) G10L 19/038 (2013.01)
(52) CPC특허분류
G10L 25/18 (2013.01)
G10L 19/038 (2013.01)
(21) 출원번호 10-2018-0089545
(22) 출원일자 2018년07월31일
심사청구일자 2018년09월12일
(65) 공개번호 10-2020-0014101
(43) 공개일자 2020년02월10일
(56) 선행기술조사문헌
KR101674851 B1*
KR1020170107683 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
전자부품연구원
경기도 성남시 분당구 새나리로 25 (야탑동)
(72) 발명자
이영한
경기도 용인시 기흥구 예현로35번길 21, 105동 105호
양중열
경기도 용인시 수지구 신봉1로 27, 502동 704호
(뒷면에 계속)
(74) 대리인
남충우

전체 청구항 수 : 총 10 항

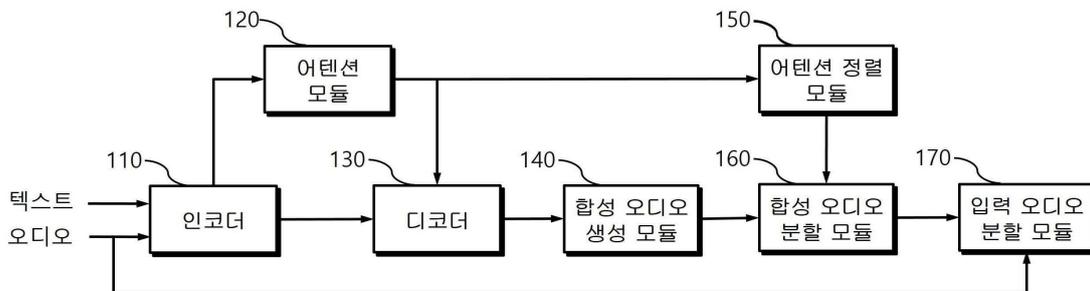
심사관 : 노지명

(54) 발명의 명칭 어텐션 메커니즘 기반의 오디오 분할 방법

(57) 요약

어텐션 메커니즘 기반의 오디오 분할 방법이 제공된다. 본 발명의 실시예에 따른 오디오 분할 방법은, 입력된 텍스트와 입력된 텍스트를 이용하여 자동으로 합성하는 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터 간의 매핑 관계를 획득하고, 매핑 관계를 이용하여 입력된 오디오 신호를 분할한다. 이에 따라, 어텐션 메커니즘을 활용한 오디오 분할을 통해, 고품질이면서도 노력, 시간 및 비용은 획기적으로 절감시킬 수 있게 된다.

대표도



(72) 발명자

조충상

경기도 성남시 수정구 위례동로 61, 5613동 303호

정혜동

서울특별시 송파구 위례광장로 185, 104동 702호

이 발명을 지원한 국가연구개발사업

과제고유번호 401B8036

부처명 자체

연구관리전문기관 전자부품연구원

연구사업명 기본연구사업

연구과제명 대화 인터페이스를 위한 음성 합성 기술 개발

기 여 율 1/1

주관기관 전자부품연구원

연구기간 2017.09.01 ~ 2020.08.31

명세서

청구범위

청구항 1

오디오 신호를 입력받는 단계;

오디오 신호에 대한 텍스트를 입력받는 단계;

'입력된 텍스트'와 입력된 텍스트를 이용하여 자동으로 합성하는 '텍스트에 대한 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터' 간의 매핑 관계를 획득하는 단계; 및

매핑 관계를 이용하여, 입력된 오디오 신호를 분할하는 단계;를 포함하고,

분할 단계는,

오디오 스펙트럼 특징벡터를 이용하여 합성한 오디오 신호를 매핑 관계를 참고하여 분할하는 제1 분할 단계;

제1 분할 단계에서의 분할 결과를 참조하여, 입력된 오디오 신호를 분할하는 제2 분할 단계;를 포함하며,

제2 분할 단계는,

합성한 오디오 신호와 입력된 오디오 신호를 컨볼루션 연산하여, 합성한 오디오 신호의 분할 구간들과 동등한 입력된 오디오 신호의 구간들을 검색하여, 입력된 오디오 신호를 다수의 구간들로 분할하는 것을 특징으로 하는 오디오 분할 방법.

청구항 2

청구항 1에 있어서,

획득 단계는,

입력된 텍스트를 이용하여 '텍스트에 대한 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터' 자동으로 합성하는 인공지능 메커니즘에서, '입력된 텍스트'와 '오디오 스펙트럼 특징벡터' 간의 매핑 관계를 학습하는 인공지능 모듈로부터, 매핑 관계를 획득하는 것을 특징으로 하는 오디오 분할 방법.

청구항 3

청구항 2에 있어서,

매핑 관계는,

'입력된 텍스트를 구성하는 라벨들' 각각과 '오디오 스펙트럼 특징벡터를 구성하는 오디오 스펙트럼 특징들' 각각에 대한 매핑도들을 나타낸 맵인 것을 특징으로 하는 오디오 분할 방법.

청구항 4

청구항 1에 있어서,

획득된 매핑 관계를 후처리 하는 단계;를 더 포함하고,

분할 단계는,

후처리된 매핑 관계를 이용하여, 입력된 오디오 신호를 분할하는 것을 특징으로 하는 오디오 분할 방법.

청구항 5

청구항 4에 있어서,

후처리 단계는,

매핑도를 기초로, '오디오 스펙트럼 특징들' 각각을 '입력된 텍스트를 구성하는 라벨들' 중 하나에 매핑시키는 것을 특징으로 하는 오디오 분할 방법.

청구항 6

청구항 5에 있어서,

후처리 단계는,

최대 매핑도, 오디오 스펙트럼 특징들 각각을 입력된 텍스트를 구성하는 라벨들 중 하나에 매핑시키기 위한 weight matrix 및 함수 중 어느 하나를 이용하는 것을 특징으로 하는 오디오 분할 방법.

청구항 7

청구항 5에 있어서,

후처리 단계는,

연음에 해당하는 라벨들의 오디오 스펙트럼 특징들을 통합하는 것을 특징으로 하는 오디오 분할 방법.

청구항 8

삭제

청구항 9

삭제

청구항 10

오디오 신호와 오디오 신호에 대한 텍스트를 입력받는 입력부;

'입력된 텍스트'와 입력된 텍스트를 이용하여 자동으로 합성하는 '텍스트에 대한 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터' 간의 매핑 관계를 획득하고, 매핑 관계를 이용하여 입력된 오디오 신호를 분할하는 프로세서;를 포함하고,

프로세서는,

오디오 스펙트럼 특징벡터를 이용하여 합성한 오디오 신호를 매핑 관계를 참고하여 분할하며,

분할 결과를 참조하여 입력된 오디오 신호를 분할하되, 합성한 오디오 신호와 입력된 오디오 신호를 컨볼루션 연산하여, 합성한 오디오 신호의 분할 구간들과 동등한 입력된 오디오 신호의 구간들을 검색하여, 입력된 오디오 신호를 다수의 구간들로 분할하는 것을 특징으로 하는 오디오 분할 시스템.

청구항 11

'입력된 텍스트'와 입력된 텍스트를 이용하여 자동으로 합성하는 '텍스트에 대한 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터' 간의 매핑 관계를 획득하는 단계;

매핑 관계를 이용하여, 입력된 오디오 신호를 분할하는 단계; 및

입력된 오디오 신호의 분할 결과를 출력하는 단계;를 포함하고,

분할 단계는,

오디오 스펙트럼 특징벡터를 이용하여 합성한 오디오 신호를 매핑 관계를 참고하여 분할하는 제1 분할 단계;

제1 분할 단계에서의 분할 결과를 참조하여, 입력된 오디오 신호를 분할하는 제2 분할 단계;를 포함하며,

제2 분할 단계는,

합성한 오디오 신호와 입력된 오디오 신호를 컨볼루션 연산하여, 합성한 오디오 신호의 분할 구간들과 동등한 입력된 오디오 신호의 구간들을 검색하여, 입력된 오디오 신호를 다수의 구간들로 분할하는 것을 특징으로 하는 오디오 분할 방법.

청구항 12

'입력된 텍스트'와 입력된 텍스트를 이용하여 자동으로 합성하는 '텍스트에 대한 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터' 간의 매핑 관계를 획득하고, 매핑 관계를 이용하여 입력된 오디오 신호를 분할하는 프로세서; 및

입력된 오디오 신호의 분할 결과를 출력하는 출력부;를 포함하고,

프로세서는,

오디오 스펙트럼 특징벡터를 이용하여 합성한 오디오 신호를 매핑 관계를 참고하여 분할하며,

분할 결과를 참조하여 입력된 오디오 신호를 분할하되, 합성한 오디오 신호와 입력된 오디오 신호를 컨볼루션 연산하여, 합성한 오디오 신호의 분할 구간들과 동등한 입력된 오디오 신호의 구간들을 검색하여, 입력된 오디오 신호를 다수의 구간들로 분할하는 것을 특징으로 하는 오디오 분할 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 오디오 분할 방법에 관한 것으로, 더욱 상세하게는 인공지능 기술을 이용하여 오디오를 자동으로 분할하는 방법에 관한 것이다.

배경 기술

[0003] 오디오 분할을 위해서는 오디오 데이터를 구분하는 작업이 필요한데, 수작업으로 수행하거나, 비교사 학습 기법(Unsupervised learning)을 활용하는 기법이 활용되고 있다.

[0004] 하지만, 수작업에 의한 오디오 분할은 전문가가 직접 오디오 데이터를 청취하면서 구분할 시점을 표기해야 하기 때문에, 작업에 소요되는 노력, 시간 및 비용이 매우 높다는 문제가 있다.

[0005] 그리고, 비교사 학습 기법에 의한 오디오 분할은, 분할 오차가 빈번하게 발생하여 최종 결과물의 품질을 떨어뜨리는 문제가 있다.

[0006] 이에 따라, 품질을 보장할 수 있으며, 노력과 시간 및 비용을 절감할 수 있는 오디오 분할 방법의 모색이 요청된다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 상기와 같은 문제점을 해결하기 위하여 안출된 것으로서, 본 발명의 목적은, 품질을 보장할 수 있으며, 노력과 시간 및 비용을 절감하기 위한 방안으로, 어텐션 메커니즘을 활용한 오디오 분할 방법을 제공함에 있다.

과제의 해결 수단

- [0010] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른, 오디오 분할 방법은, 오디오 신호를 입력받는 단계; 오디오 신호에 대한 텍스트를 입력받는 단계; '입력된 텍스트'와 입력된 텍스트를 이용하여 자동으로 합성하는 '텍스트에 대한 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터' 간의 매핑 관계를 획득하는 단계; 및 매핑 관계를 이용하여, 입력된 오디오 신호를 분할하는 단계;를 포함한다.
- [0011] 그리고, 획득 단계는, 입력된 텍스트를 이용하여 '텍스트에 대한 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터' 자동으로 합성하는 인공지능 메커니즘에서, '입력된 텍스트'와 '오디오 스펙트럼 특징벡터' 간의 매핑 관계를 학습하는 인공지능 모듈로부터, 매핑 관계를 획득할 수 있다.
- [0012] 또한, 매핑 관계는, '입력된 텍스트를 구성하는 라벨들' 각각과 '오디오 스펙트럼 특징벡터를 구성하는 오디오 스펙트럼 특징들' 각각에 대한 매핑도들을 나타낸 맵일 수 있다.
- [0013] 그리고, 본 발명의 실시예에 따른 오디오 분할 방법은, 획득된 매핑 관계를 후처리 하는 단계;를 더 포함하고, 분할 단계는, 처리된 매핑 관계를 이용하여, 입력된 오디오 신호를 분할할 수 있다.
- [0014] 또한, 후처리 단계는, 매핑도를 기초로, '오디오 스펙트럼 특징들' 각각을 '입력된 텍스트를 구성하는 라벨들' 중 하나에 매핑시킬 수 있다.
- [0015] 그리고, 후처리 단계는, 최대 매핑도, 오디오 스펙트럼 특징들 각각을 입력된 텍스트를 구성하는 라벨들 중 하나에 매핑시키기 위한 weight matrix 및 함수 중 어느 하나를 이용할 수 있다.
- [0016] 또한, 후처리 단계는, 연음에 해당하는 라벨들의 오디오 스펙트럼 특징들을 통합할 수 있다.
- [0017] 그리고, 분할 단계는, 오디오 스펙트럼 특징벡터를 이용하여 합성한 오디오 신호를 매핑 관계를 참고하여 분할하는 제1 분할 단계; 제1 분할 단계에서의 분할 결과를 참조하여, 입력된 오디오 신호를 분할하는 제2 분할 단계;를 포함할 수 있다.
- [0018] 또한, 제2 분할 단계는, 합성한 오디오 신호와 입력된 오디오 신호를 컨볼루션 연산하여, 합성한 오디오 신호의 분할 구간들과 동등한 입력된 오디오 신호의 구간들을 검색하여, 입력된 오디오 신호를 다수의 구간들로 분할할 수 있다.
- [0019] 한편, 본 발명의 다른 실시예에 따른, 오디오 분할 시스템은, 오디오 신호와 오디오 신호에 대한 텍스트를 입력받는 입력부; '입력된 텍스트'와 입력된 텍스트를 이용하여 자동으로 합성하는 '텍스트에 대한 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터' 간의 매핑 관계를 획득하고, 매핑 관계를 이용하여 입력된 오디오 신호를 분할하는 프로세서;를 포함한다.
- [0020] 한편, 본 발명의 다른 실시예에 따른, 오디오 분할 방법은, '입력된 텍스트'와 입력된 텍스트를 이용하여 자동으로 합성하는 '텍스트에 대한 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터' 간의 매핑 관계를 획득하는 단계; 매핑 관계를 이용하여, 입력된 오디오 신호를 분할하는 단계; 및 입력된 오디오 신호의 분할 결과를 출력하는 단계;를 포함한다.
- [0021] 한편, 본 발명의 다른 실시예에 따른, 오디오 분할 시스템은, '입력된 텍스트'와 입력된 텍스트를 이용하여 자동으로 합성하는 '텍스트에 대한 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터' 간의 매핑 관계를 획득하고, 매핑 관계를 이용하여 입력된 오디오 신호를 분할하는 프로세서; 및 입력된 오디오 신호의 분할 결과를 출력하는 출력부;를 포함한다.

발명의 효과

- [0023] 이상 설명한 바와 같이, 본 발명의 실시예들에 따르면, 어텐션 메커니즘을 활용한 오디오 분할을 통해, 고품질 이면서도 노력, 시간 및 비용은 획기적으로 절감시킬 수 있게 된다.

도면의 간단한 설명

- [0025] 도 1은 수작업에 의한 오디오 분할에 이용되는 작업 화면이 나타난 도면,
- 도 2는 GMM-HMM 기반의 학습기를 이용한 오디오 분할 기법을 나타낸 도면,
- 도 3은 Encoder-Decoder 형태로 구성된 seq-to-seq model의 구조도,

도 4는 본 발명의 실시예에 따른 어텐션 메커니즘 기반의 오디오 분할 방법의 설명에 제공되는 도면,
 도 5는 본 발명의 다른 실시예에 따른 오디오 분할 방법의 설명에 제공되는 흐름도,
 도 6은 본 발명의 또 다른 실시예에 따른 오디오 분할 시스템의 하드웨어 구성을 나타낸 블록도이다.

발명을 실시하기 위한 구체적인 내용

- [0026] 이하에서는 도면을 참조하여 본 발명을 보다 상세하게 설명한다.
- [0027] 도 1은 수작업에 의한 오디오 분할에 이용되는 작업 화면이 나타난 도면이다. 수작업에 의한 오디오 분할은, 전문가가 오디오를 직접 듣고, 도 1에 나타난 바와 같이, 오디오를 텍스트 라벨들에 맞추어 직접 구분하여 오디오를 분할하는 방식이다.
- [0028] 전문가의 능력 및 정책 등에 따라 분할 성능/품질의 편차가 크고, 데이터베이스의 크기에 비례하여 시간과 비용이 증가하기 때문에, 고용량 오디오 데이터에 대한 분할에 특히 부담이 될 수 있다.
- [0029] 도 2는 GMM-HMM 기반의 학습기를 이용한 오디오 분할 기법을 나타낸 도면이다. GMM-HMM 기반의 학습에 의한 오디오 분할할 경우, 데이터베이스의 크기에 따라 기계 학습에 소요되는 시간이 증가할 수 있지만, 수작업을 하지 않아도 되기 때문에 시간과 비용 측면에서 유리한 장점이 있다. 하지만, 분할 오차가 크기 때문에, 이에 따른 분할 오디오 데이터를 사용하기에 부적절한 부분이 존재한다.
- [0030] 도 3은 Encoder-Decoder 형태로 구성된 seq(quence)-to-seq(quence) model의 구조도이다. 본 발명의 실시예에 따른 오디오 분할 방법에서 활용할 어텐션 메커니즘은 딥러닝 기술 중 하나인 seq-to-seq model을 기반으로 한다.
- [0031] 도 3에 도시된 바와 같이, seq-to-seq model은 시계열 데이터를 입력으로 하여 시계열 데이터를 출력하는 응용 분야에 적용하기 위한 딥러닝 기술이다.
- [0032] 본 발명의 실시예에 따른 오디오 분할 방법은, 시계열 데이터 중의 하나인 오디오 데이터를 입력 받아 그에 해당하는 연속된 분할 데이터들을 생성하기 위해, 어텐션 메커니즘을 활용한다.
- [0033] 이하에서, 본 발명의 실시예에 따른 오디오 분할 방법에 대해, 도 4를 참조하여 상세히 설명한다. 도 4는 본 발명의 실시예에 따른 어텐션 메커니즘 기반의 오디오 분할 방법의 설명에 제공되는 도면이다.
- [0034] 본 발명의 실시예에 따른 오디오 분할 방법은, End-to-End 음성 합성 기법에 이용되는 어텐션 메커니즘을 이용하여 분할 오차를 줄이면서 높은 작업 효율로 오디오 분할을 수행한다.
- [0035] 이를 위해, 본 발명의 실시예에 따른 오디오 분할 방법에서는, 어텐션 정보를 활용하여 텍스트로부터 합성한 오디오를 텍스트 라벨들에 맞춰 분할하고, 분할한 합성 오디오를 참조하여 입력 오디오를 분할한다. 입력 오디오에 대한 구체적인 분할 과정에 대해 이하에서 설명한다.
- [0036] 도 4에 도시된 바와 같이, 먼저, 오디오 신호와 오디오 신호에 대한 텍스트를 입력받는다. 그러면, 인코더(110)는 입력된 텍스트에 대한 특징벡터를 추출한다.
- [0037] 디코더(130)는 인코더(110)가 추출한 특징벡터를 이용하여, 입력 텍스트에 대한 오디오 스펙트럼 특징벡터를 자동으로 생성한다. 오디오 스펙트럼 특징벡터로부터 텍스트에 대한 오디오 신호가 합성될 수 있다.
- [0038] 한편, 어텐션 모듈(120)은 인코더(110)에 입력된 텍스트와 디코더(130)에서 생성하는 오디오 스펙트럼 특징벡터 간의 매핑 관계를 생성한다. 어텐션 모듈(120)에 의해 생성되는 매핑 관계는, '입력 텍스트를 구성하는 라벨들' 각각과 '오디오 스펙트럼 특징벡터를 구성하는 오디오 스펙트럼 특징들' 각각에 대한 매핑도를 나타낸 맵이다.
- [0039] 이와 같이, 어텐션 모듈(120)은 입력 텍스트를 이용하여 텍스트에 대한 오디오 신호를 생성하기 위한 오디오 스펙트럼 특징벡터를 자동으로 합성하는 인공지능 메커니즘에서, '입력 텍스트'와 '오디오 스펙트럼 특징벡터' 간의 상관 관계를 학습하는 인공지능 모듈에 해당한다.
- [0040] 입력 텍스트에 대한 오디오 스펙트럼 특징벡터를 생성함에 있어, 디코더(130)는 어텐션 모듈(120)에 의해 생성된 매핑 관계를 참조한다. 텍스트 라벨들이 오디오 신호에 적절하게 분배되도록 하기 위함이다.
- [0041] 합성 오디오 생성 모듈(140)은 디코더(130)에 의해 생성된 오디오 스펙트럼 특징벡터로부터 입력 텍스트에 대한 오디오 신호를 합성한다.
- [0042] 어텐션 정렬 모듈(150)은 어텐션 모듈(120)에 의해 생성된 매핑 관계를 후처리 하기 위한 모듈로, 매핑 관계에

나타난 매핑도를 기초로, '오디오 스펙트럼 특징들' 각각을 '입력된 텍스트를 구성하는 라벨들' 중 최적의 단 하나에만 매핑시킨다.

- [0043] 구체적으로, 어텐션 정렬 모듈(150)은 오디오 스펙트럼 특징을 매핑도가 가장 큰 텍스트 라벨에 매핑시킬 수 있다. 다른 방법으로, 어텐션 정렬 모듈(150)은 양자를 매핑시키기 위한 weight matrix, 함수 등을 이용할 수 있으며, 이들은 학습을 통해 훈련될 수 있다.
- [0044] 또한, 어텐션 정렬 모듈(150)은 연음에 해당하는 텍스트 라벨들의 오디오 스펙트럼 특징들을 통합할 수 있다.
- [0045] 합성 오디오 분할 모듈(160)은 어텐션 정렬 모듈(150)에 의해 후처리된 매핑 관계를 이용하여, 합성 오디오 생성 모듈(140)에서 생성된 합성 오디오 신호를 분할한다.
- [0046] 입력 오디오 분할 모듈(170)은 합성 오디오 분할 모듈(160)에 의해 분할된 합성 오디오 신호를 참조하여, 입력 오디오 신호를 분할한다.
- [0047] 구체적으로, 입력 오디오 분할 모듈(170)은 합성 오디오 신호와 입력 오디오 신호를 컨볼루션 연산하여, 합성 오디오 신호의 분할 구간들과 동등한 입력 오디오 신호의 구간들을 검색하여, 입력 오디오 신호를 다수의 구간들로 구분한다.
- [0048] 도 5는 본 발명의 다른 실시예에 따른 오디오 분할 방법의 설명에 제공되는 흐름도이다. 입력 오디오를 분할하기 위해서는, 입력 오디오 외에 입력 오디오에 대한 텍스트가 필요하다.
- [0049] 먼저, 도 5에 도시된 바와 같이, 인코더(110)가 입력 텍스트에 대한 특징벡터를 추출한다(S210). 그리고, 어텐션 모듈(120)은 입력 텍스트와 합성 오디오를 생성하기 위한 오디오 스펙트럼 특징벡터 간의 매핑 관계를 생성한다(S220).
- [0050] 디코더(130)는 S210단계에서 추출된 특징벡터를 이용하여, S220단계에서 생성된 매핑 관계를 참고로, 입력 텍스트에 대한 오디오 스펙트럼 특징벡터를 생성한다(S230).
- [0051] 그러면, 합성 오디오 생성 모듈(140)이 S230단계에서 생성된 오디오 스펙트럼 특징벡터로부터 입력 텍스트에 대한 오디오 신호를 합성한다(S240).
- [0052] 한편, 어텐션 정렬 모듈(150)은 S230단계에서 생성된 매핑 관계를 후처리하여, '오디오 스펙트럼 특징들' 각각을 '입력된 텍스트를 구성하는 라벨들' 중 최적의 단 하나에만 매핑시킨다(S250). 나아가, 어텐션 정렬 모듈(150)은 연음에 해당하는 라벨들의 오디오 스펙트럼 특징들을 통합할 수 있다.
- [0053] 그러면, 합성 오디오 분할 모듈(160)은 S250단계에서 후처리된 매핑 관계를 이용하여, S240단계에서 생성된 합성 오디오 신호를 분할한다(S260).
- [0054] 그리고, 입력 오디오 분할 모듈(170)은 S260단계에서 분할된 합성 오디오 신호를 참조하여, 입력 오디오 신호를 분할한다(S270).
- [0055] 도 6은 본 발명의 또 다른 실시예에 따른 오디오 분할 시스템의 하드웨어 구성을 나타낸 블럭도이다. 본 발명의 다른 실시예에 따른 오디오 분할 시스템은, 도 6에 도시된 바와 같이, 통신부(310), 출력부(320), 프로세서(330), 입력부(340) 및 저장부(350)를 포함하는 컴퓨팅 시스템으로 구현할 수 있다.
- [0056] 통신부(310)는 외부 기기와 외부 네트워크로부터 오디오와 오디오에 대한 텍스트를 입력받기 위한 통신 수단이다.
- [0057] 입력부(340)는 사용자 설정 명령을 입력받기 위한 입력 수단이고, 출력부(320)는 오디오 분할 과정 및 오디오 분할 결과를 출력하기 위한 디스플레이이다.
- [0058] 프로세서(330)는 도 3 및 도 4에 도시된 방법을 실행하여 오디오 분할을 수행한다.
- [0059] 저장부(350)는 프로세서(330)가 동작함에 있어 필요한 저장 공간을 제공한다. 또한, 저장부(350)는 텍스트와 오디오를 저장하고 있을 수도 있다.
- [0060] 지금까지, 어텐션 메커니즘 기반의 오디오 분할 방법 및 시스템에 대해 바람직한 실시예들을 들어 상세히 설명하였다.
- [0061] 본 발명의 실시예에서는, 어텐션 메커니즘을 활용한 오디오 분할을 통해, 고품질이면서도 노력, 시간 및 비용은 획기적으로 절감시킬 수 있는 기법을 제시하였다.

[0062] 위 실시예에서 언급한 오디오 분할은 설명의 편의를 위한 일 예에 불과하다. 음성 등의 오디오 이외의 다른 시계열 데이터를 분할하는 경우에 대해서도 본 발명의 기술적 사상이 적용될 수 있음은 물론이다.

[0063] 한편, 본 실시예에 따른 장치와 방법의 기능을 수행하게 하는 컴퓨터 프로그램을 수록한 컴퓨터로 읽을 수 있는 기록매체에도 본 발명의 기술적 사상이 적용될 수 있음은 물론이다. 또한, 본 발명의 다양한 실시예에 따른 기술적 사상은 컴퓨터로 읽을 수 있는 기록매체에 기록된 컴퓨터로 읽을 수 있는 코드 형태로 구현될 수도 있다. 컴퓨터로 읽을 수 있는 기록매체는 컴퓨터에 의해 읽을 수 있고 데이터를 저장할 수 있는 어떤 데이터 저장 장치이더라도 가능하다. 예를 들어, 컴퓨터로 읽을 수 있는 기록매체는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광디스크, 하드 디스크 드라이브, 등이 될 수 있음은 물론이다. 또한, 컴퓨터로 읽을 수 있는 기록매체에 저장된 컴퓨터로 읽을 수 있는 코드 또는 프로그램은 컴퓨터간에 연결된 네트워크를 통해 전송될 수도 있다.

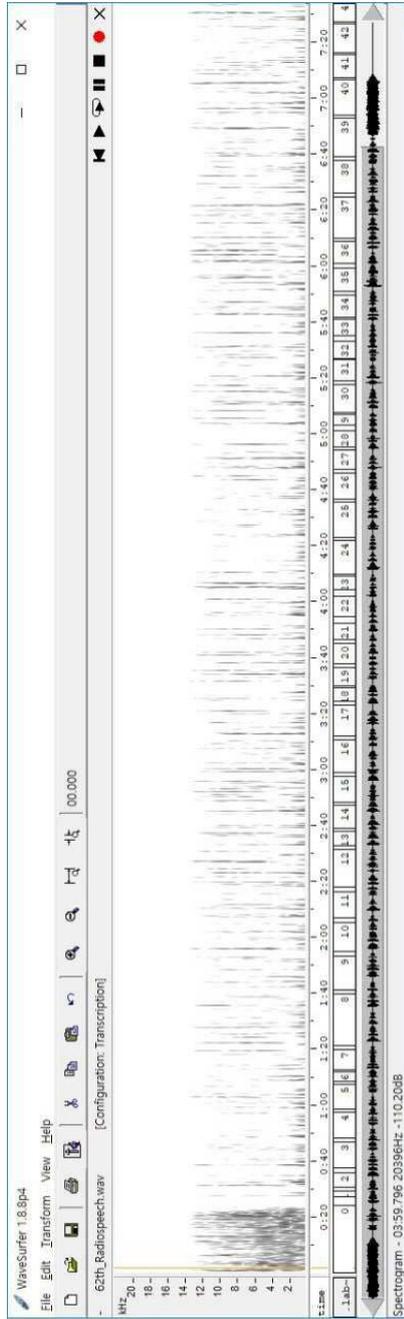
[0064] 또한, 이상에서는 본 발명의 바람직한 실시예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특정의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어져서는 안될 것이다.

부호의 설명

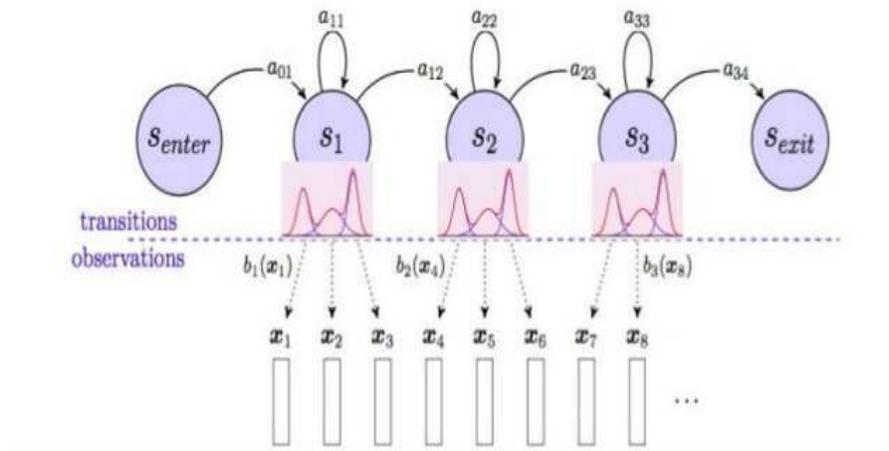
- [0066] 110 : 인코더
- 120 : 어텐션 모듈
- 130 : 디코더
- 140 : 합성 오디오 생성 모듈
- 150 : 어텐션 정렬 모듈
- 160 : 합성 오디오 분할 모듈
- 170 : 입력 오디오 분할 모듈

도면

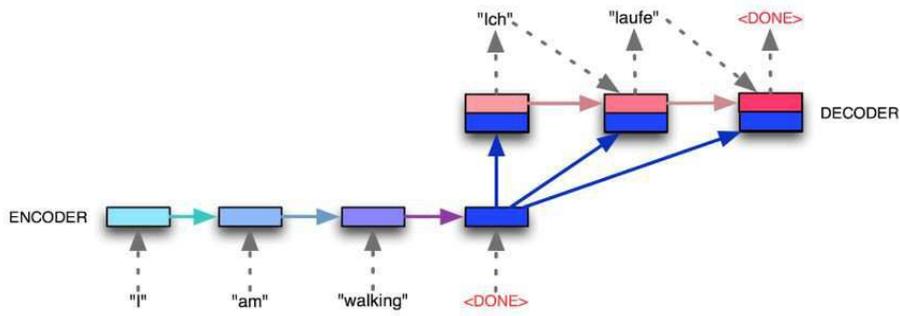
도면1



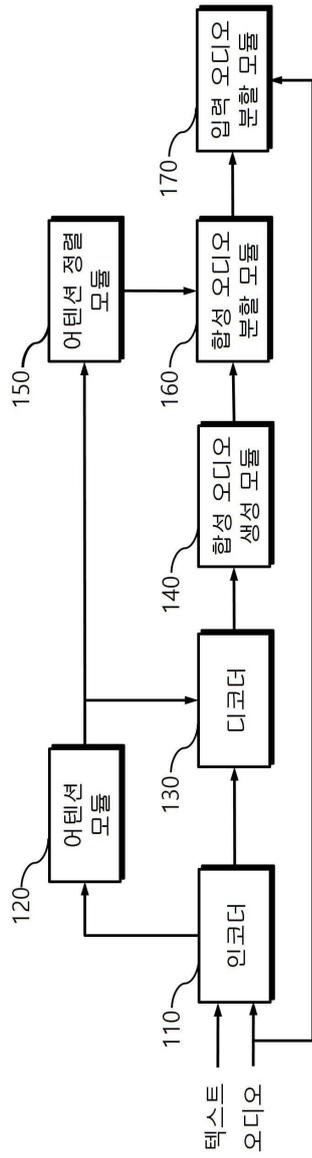
도면2



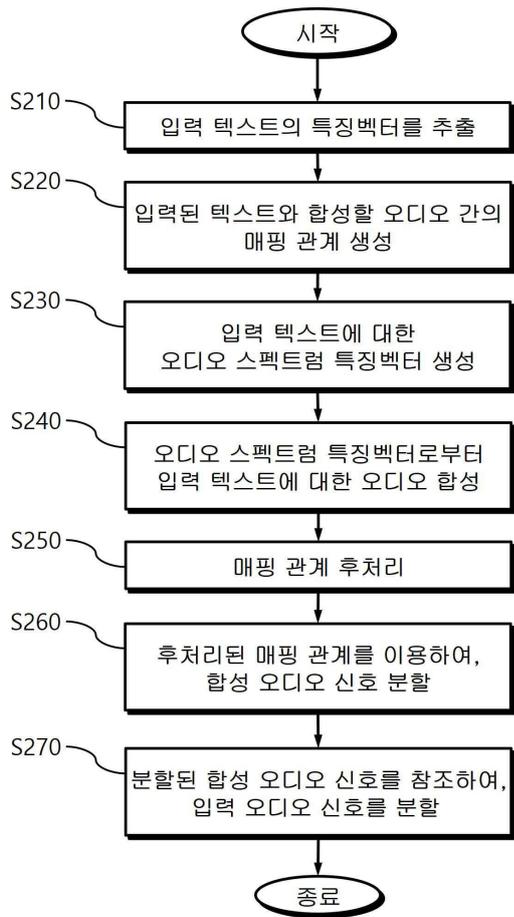
도면3



도면4



도면5



도면6

