



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0088133
(43) 공개일자 2024년06월20일

(51) 국제특허분류(Int. Cl.)

G10L 13/02 (2006.01) A61B 5/377 (2021.01)
G06F 3/01 (2006.01) G10L 15/04 (2006.01)
G10L 15/06 (2006.01) G10L 15/26 (2006.01)
G10L 25/18 (2013.01)

(52) CPC특허분류

G10L 13/02 (2013.01)
A61B 5/377 (2022.01)

(21) 출원번호 10-2022-0173551

(22) 출원일자 2022년12월13일

심사청구일자 2022년12월13일

(71) 출원인

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

이성환

서울특별시 강남구 언주로30길 13 (대림아크로빌, 대림아크로텔) B-2902

이영은

경기도 성남시 분당구 중앙공원로 17 시범단지한양아파트 311동 701호

(뒷면에 계속)

(74) 대리인

특허법인위솔

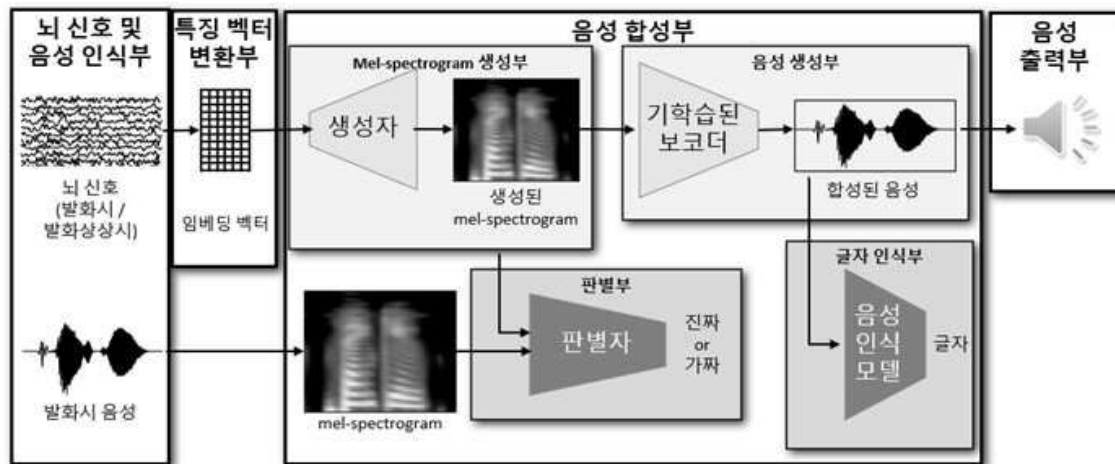
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 발화 상상 시 뇌파 기반 음성 합성 방법 및 장치

(57) 요약

본 발명은 발화 상상 시 뇌파 기반 음성 합성 방법 및 장치에 관한 것이다. 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 방법은, 사용자의 발화 상상 뇌파를 획득하는 단계, 획득된 상기 발화 상상 뇌파를 임베딩 벡터로 변환하는 단계, 상기 임베딩 벡터를 기반으로 mel-spectrogram을 생성하는 단계, 상기 mel-spectrogram을 이용하여 음성을 생성하는 단계 및 생성된 상기 음성을 출력하는 단계를 포함할 수 있다.

대표도 - 도5



(52) CPC특허분류

G06F 3/015 (2013.01)

G10L 15/04 (2013.01)

G10L 15/063 (2013.01)

G10L 15/26 (2013.01)

G10L 25/18 (2013.01)

(72) 발명자

이서현

서울특별시 강남구 압구정로 321 한양아파트 9동 403호

김수원

대구광역시 수성구 신천동로 320 신세계타운 11동 1005호

김상호

서울특별시 서초구 효령로33길 50 (방배동 방배서리풀이편한세상) 104동 201호

고병관

강원도 동해시 평릉길 60-1 사랑드림아파트 102동 504호

이지원

서울특별시 도봉구 도봉로180길 6-23 동아에코빌아파트 106동 2203호

이정선

경상남도 창원시 진해구 조천북로 131 경화베스티움 102동 803호

이 발명을 지원한 국가연구개발사업

과제고유번호	1711152863
과제번호	2021-0-02068-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	인공지능 혁신 허브 연구 개발
기 여 율	1/2
과제수행기관명	고려대학교산학협력단
연구기간	2022.01.01 ~ 2022.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711093609
과제번호	2019-0-00079-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	인공지능핵심고급인재양성(R&D)
연구과제명	인공지능대학원지원(고려대학교)
기 여 율	1/2
과제수행기관명	고려대학교산학협력단
연구기간	2022.01.01 ~ 2022.12.31

명세서

청구범위

청구항 1

발화 상상 시 뇌파 기반 음성 합성 방법에 있어서,
 사용자의 발화 상상 뇌파를 획득하는 단계;
 획득된 상기 발화 상상 뇌파를 임베딩 벡터로 변환하는 단계;
 상기 임베딩 벡터를 기반으로 mel-spectrogram을 생성하는 단계;
 상기 mel-spectrogram을 이용하여 음성을 생성하는 단계; 및
 생성된 상기 음성을 출력하는 단계를 포함하는 발화 상상 시 뇌파 기반 음성 합성 방법.

청구항 2

제1항에 있어서,
 생성자 및 판별자를 학습하는 단계를 더 포함하고,
 여기서 상기 생성자는 임베딩 벡터를 이용하여 mel-spectrogram을 생성하고, 상기 판별자는 뇌파로부터 생성된 mel-spectrogram과 음성으로부터 생성된 mel-spectrogram을 구분하는 것인 발화 상상 시 뇌파 기반 음성 합성 방법.

청구항 3

제2항에 있어서,
 상기 생성자 및 판별자를 학습하는 단계는,
 발화 시 발생하는 뇌파 및 음성 신호를 기반으로 상기 생성자 및 판별자를 학습하는 단계;
 상기 발화 시 발생하는 뇌파 및 음성 신호를 기반으로 학습된 상기 생성자 및 판별자에 대해 전이 학습을 수행하는 단계; 및
 발화 상상 시 발생하는 뇌파를 기반으로 상기 생성자 및 판별자를 재학습하는 단계를 포함하는 발화 상상 시 뇌파 기반 음성 합성 방법.

청구항 4

제3항에 있어서,
 상기 발화 시 발생하는 뇌파 및 음성 신호를 기반으로 상기 생성자 및 판별자를 학습하는 단계는,
 상기 발화 시 발생하는 뇌파 및 음성 신호를 획득하는 단계;
 상기 발화 시 발생하는 뇌파를 임베딩 벡터로 변환하는 단계;
 상기 발화 시 발생하는 뇌파로부터 변환된 임베딩 벡터를 기반으로 mel-spectrogram을 생성하는 단계; 및
 상기 발화 시 발생하는 뇌파로부터 변환된 임베딩 벡터를 기반으로 mel-spectrogram과 상기 음성 신호의 mel-spectrogram을 기반으로 상기 생성자 및 판별자를 학습하는 단계를 포함하는 발화 상상 시 뇌파 기반 음성 합성 방법.

청구항 5

제1항에 있어서,
 상기 사용자의 발화 상상 뇌파를 임베딩 벡터 변환 단계는, 발화 상상 뇌파의 음소 별 또는 단어 별 차이를 극

대화하는 임베딩 공간으로 상기 사용자의 발화 상상 뇌파를 변환하는 것인 발화 상상 시 뇌파 기반 음성 합성 방법.

청구항 6

제1항에 있어서,

음성 인식 모델을 기반으로 상기 음성을 글자, 음소, 또는 발음으로 변환하는 단계를 더 포함하는 발화 상상 시 뇌파 기반 음성 합성 방법.

청구항 7

제6항에 있어서,

변환된 상기 글자, 음소, 또는 발음은 생성자의 학습에 이용되는 발화 상상 시 뇌파 기반 음성 합성 방법.

청구항 8

발화 상상 시 뇌파 기반 음성 합성 장치에 있어서,

사용자의 발화 상상 뇌파를 획득하는 뇌 신호 및 음성 인식부;

획득된 상기 발화 상상 뇌파를 임베딩 벡터로 변환하는 특징 벡터 변환부;

상기 임베딩 벡터를 기반으로 mel-spectrogram을 생성하고, 상기 mel-spectrogram을 이용하여 음성을 생성하는 음성 합성부; 및

생성된 상기 음성을 출력하는 음성 출력부를 포함하는 발화 상상 시 뇌파 기반 음성 합성 장치.

청구항 9

제8항에 있어서,

상기 음성 합성부는,

임베딩 벡터를 이용하여 mel-spectrogram을 생성하는 생성자를 포함하는 mel-spectrogram 생성부; 및

뇌파로부터 생성된 mel-spectrogram과 음성으로부터 생성된 mel-spectrogram을 구분하는 판별자를 포함하는 판별부를 포함하는 발화 상상 시 뇌파 기반 음성 합성 장치.

청구항 10

제9항에 있어서,

발화 시 발생하는 뇌파 및 음성 신호를 기반으로 상기 생성자 및 판별자가 학습되고,

상기 발화 시 발생하는 뇌파 및 음성 신호를 기반으로 학습된 상기 생성자 및 판별자에 대해 전이 학습을 수행되고,

발화 상상 시 발생하는 뇌파를 기반으로 상기 생성자 및 판별자가 재학습되는 발화 상상 시 뇌파 기반 음성 합성 장치.

청구항 11

제10항에 있어서,

발화 시 발생하는 뇌파 및 음성 신호를 기반으로 상기 생성자 및 판별자가 학습되는 것은,

상기 발화 시 발생하는 뇌파 및 음성 신호가 획득되고,

상기 발화 시 발생하는 뇌파를 임베딩 벡터로 변환되고,

상기 발화 시 발생하는 뇌파로부터 변환된 임베딩 벡터를 기반으로 mel-spectrogram을 생성되고,

상기 발화 시 발생하는 뇌파로부터 변환된 임베딩 벡터를 기반으로 mel-spectrogram과 상기 음성 신호의 mel-spectrogram을 기반으로 상기 생성자 및 판별자가 학습되는 것인 발화 상상 시 뇌파 기반 음성 합성 장치.

청구항 12

제8항에 있어서,

특정 벡터 변환부는, 발화 상상 뇌파의 음소 별 또는 단어 별 차이를 극대화하는 임베딩 공간으로 상기 사용자의 발화 상상 뇌파를 변환하는 발화 상상 시 뇌파 기반 음성 합성 장치.

청구항 13

제8항에 있어서,

상기 음성 합성부는, 음성 인식 모델을 기반으로 상기 음성을 글자, 음소, 또는 발음으로 변환하는 글자 인식 부를 포함하는 발화 상상 시 뇌파 기반 음성 합성 장치.

청구항 14

제13항에 있어서,

변환된 상기 글자, 음소, 또는 발음은 생성자의 학습에 이용되는 발화 상상 시 뇌파 기반 음성 합성 장치.

발명의 설명

기술 분야

[0001] 본 발명은 발화 상상 시 뇌파 기반 음성 합성 방법 및 장치에 관한 것이다. 보다 상세하게는, 단어 또는 문장 등에 대한 발화 상상을 수행할 때의 뇌 신호를 측정하여 사용자의 뇌파로부터 사용자의 목소리를 합성하는 것을 가능케 하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 뇌-기계 인터페이스(Brain-machine interface) 기술은 사람의 뇌에서 일어나는 신경 세포들의 활동에 의해 발생하는 신호들을 측정하여 특정한 입력장치 없이 뇌 신호만으로 컴퓨터 및 기계와 같은 출력 장치의 제어가 가능하게 하는 기술이다. 기존에는 척추 부상이나 퇴행성 신경질환으로 인해 신체가 자유롭지 못하거나 마비된 환자들을 위한 보철기구 제어, 의사소통을 위한 스펠러 등과 같은 의료 분야에서 주로 사용되어 왔다. 그러나 최근 뇌 신호 분석 기술의 발달로 인하여 환자가 아닌 일반인들의 일상생활 보조 서비스를 비롯한 다양한 분야에서 비침습적 뇌-컴퓨터 인터페이스가 적용되고 있는 추세이다. 뇌-컴퓨터 인터페이스의 작동 방식으로는 외재적인 방식과 내재적인 방식이 있다. 외재적인 방식은 외부의 자극을 통해 유발된 뇌파를 기반으로 기기를 제어하는 방식이고, 내재적인 방식은 사용자가 외부의 자극 없이 상상을 통해서 유발되는 뇌파를 분석해 기기를 제어하는 방식이다. 내재적인 방식은 별도의 외부의 자극 장치가 필요 없고, 오히려 사용자의 상상 의도가 반영된 직관적 방식이기 때문에, 최근 뇌-컴퓨터 인터페이스 기술에서 더 많이 사용되고 있다.

[0003] 뇌-기계 인터페이스는 주로 신체부위의 동작 상상(Motor imagery, MI), 정상 상태 유발 전위(Steady state evoked potential, SSEP), 사건 유발 전위(Event related potential, ERP), 발화 상상(Imagined speech) 등의 특정한 뇌 신호 패턴을 유도하여 사용자의 의도를 인식한다. 현재 광범위하게 많이 쓰이는 동작 상상 또는 사건 유발전위 기반의 의사소통 시스템은 입력하고자 하는 철자가 제한적이고 직관성이 떨어진다는 단점이 있다.

[0004] 발화 상상은 실제로 소리를 내거나 조음기관을 움직이지 않고 말하는 상상을 통해 뇌 신호를 유도해 사용자의 의도를 인식하는 최신 패러다임이다. 발화 상상은 사용자가 말하는 것을 상상할 때 말하는 기능을 담당하는 브로카 영역(Broca's area)과 언어의 해석을 담당하는 베르니케 영역(Wernicke's area)에서 특정 패턴의 뇌 신호가 발현되는 것이다. 발화 상상은 말하는 상상으로서 언어적 다양성으로 인해 더욱 높은 자유도의 클래스 구분이 가능하며, 사용자의 대화 의도를 전달함에 있어서 직접적으로 말하는 상상을 하게 됨으로써 직관적인 의사소통이 가능하다. 이러한 발화 상상 기반의 뇌-기계 인터페이스는 말하는 능력에 장애가 있는 환자들이나 일반인들에게 의사소통 시스템으로 사용될 수 있다.

[0005] 뇌파를 통해 사용자 의도를 검출하는 데에 있어서 현재는 클래스 단위의 분류 위주로 진행되고 있다. 하지만, 사용자의 의사소통 시스템으로 사용되기 위해서는 자유도가 높은 기술을 활용할 필요가 있다. 따라서, 단순 클래스를 분류하는 시스템이 아닌, 뇌파로부터 자유도 높게 음성을 합성하는 기술을 통해 자유도 높은 의사소통이 가능하다. 본 발명은 이에 대한 것이다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 대한민국 공개특허공보 제10-2020-0052807호(2020.05.15)

발명의 내용

해결하려는 과제

[0007] 뇌 신호로부터 사용자의 의도를 인식하는 기술은 최근 많이 개발되고 있고, 분류 인식 성능을 높이기 위한 딥러닝 기반의 방법론들이 많이 개발되고 있다. 그 중에 발화 상상 기반의 사용자 의도 인식 기술은 클래스의 수를 늘릴 수 있다는 점에서 장점이 있다. 하지만, 분류 기반의 사용자 의도 인식은 분류 클래스 수로 자유도가 한정된다는 점에서 한계가 있다. 따라서, 본 발명에서는 뇌 신호로부터 음성을 합성하는 음성 합성 방법론을 제안한다.

[0008] 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0009] 전술한 기술적 과제를 달성하기 위한 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 방법은, 사용자의 발화 상상 뇌파를 획득하는 단계, 획득된 상기 발화 상상 뇌파를 임베딩 벡터로 변환하는 단계, 상기 임베딩 벡터를 기반으로 mel-spectrogram을 생성하는 단계, 상기 mel-spectrogram을 이용하여 음성을 생성하는 단계 및 생성된 상기 음성을 출력하는 단계를 포함할 수 있다.

[0010] 일 실시 예에 따르면, 생성자 및 판별자를 학습하는 단계를 더 포함하고, 여기서 상기 생성자는 임베딩 벡터를 이용하여 mel-spectrogram을 생성하고, 상기 판별자는 뇌파로부터 생성된 mel-spectrogram과 음성으로부터 생성된 mel-spectrogram을 구분하는 것일 수 있다.

[0011] 일 실시 예에 따르면, 상기 생성자 및 판별자를 학습하는 단계는, 발화 시 발생하는 뇌파 및 음성 신호를 기반으로 상기 생성자 및 판별자를 학습하는 단계, 상기 발화 시 발생하는 뇌파 및 음성 신호를 기반으로 학습된 상기 생성자 및 판별자에 대해 전이 학습을 수행하는 단계 및 발화 상상 시 발생하는 뇌파를 기반으로 상기 생성자 및 판별자를 재학습하는 단계를 포함할 수 있다.

[0012] 일 실시 예에 따르면, 상기 발화 시 발생하는 뇌파 및 음성 신호를 기반으로 상기 생성자 및 판별자를 학습하는 단계는, 상기 발화 시 발생하는 뇌파 및 음성 신호를 획득하는 단계, 상기 발화 시 발생하는 뇌파를 임베딩 벡터로 변환하는 단계, 상기 발화 시 발생하는 뇌파로부터 변환된 임베딩 벡터를 기반으로 mel-spectrogram을 생성하는 단계 및 상기 발화 시 발생하는 뇌파로부터 변환된 임베딩 벡터를 기반으로 mel-spectrogram과 상기 음성 신호의 mel-spectrogram을 기반으로 상기 생성자 및 판별자를 학습하는 단계를 포함할 수 있다.

[0013] 일 실시 예에 따르면, 상기 사용자의 발화 상상 뇌파를 임베딩 벡터 변환 단계는, 발화 상상 뇌파의 음소 별 또는 단어 별 차이를 극대화하는 임베딩 공간으로 상기 사용자의 발화 상상 뇌파를 변환하는 것일 수 있다.

[0014] 일 실시 예에 따르면, 음성 인식 모델을 기반으로 상기 음성을 글자, 음소, 또는 발음으로 변환하는 단계를 더 포함할 수 있다.

[0015] 일 실시 예에 따르면, 변환된 상기 글자, 음소, 또는 발음은 생성자의 학습에 이용될 수 있다.

[0016] 한편 전술한 기술적 과제를 달성하기 위한 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 장치는, 사용자의 발화 상상 뇌파를 획득하는 뇌 신호 및 음성 인식부, 획득된 상기 발화 상상 뇌파를 임베딩 벡터로 변환하는 특징 벡터 변환부, 상기 임베딩 벡터를 기반으로 mel-spectrogram을 생성하고, 상기 mel-spectrogram을 이용하여 음성을 생성하는 음성 합성부 및 생성된 상기 음성을 출력하는 음성 출력부를 포함할 수 있다.

[0017] 일 실시 예에 따르면, 상기 음성 합성부는, 임베딩 벡터를 이용하여 mel-spectrogram을 생성하는 생성자를 포함하는 mel-spectrogram 생성부 및 뇌파로부터 생성된 mel-spectrogram과 음성으로부터 생성된 mel-

spectrogram을 구분하는 판별자를 포함하는 판별부를 포함하는할 수 있다.

[0018] 일 실시 예에 따르면, 발화 시 발생하는 뇌파 및 음성 신호를 기반으로 상기 생성자 및 판별자가 학습되고, 상기 발화 시 발생하는 뇌파 및 음성 신호를 기반으로 학습된 상기 생성자 및 판별자에 대해 전이 학습을 수행되고, 발화 상상 시 발생하는 뇌파를 기반으로 상기 생성자 및 판별자가 재학습될 수 있다.

[0019] 일 실시 예에 따르면, 발화 시 발생하는 뇌파 및 음성 신호를 기반으로 상기 생성자 및 판별자가 학습되는 것은, 상기 발화 시 발생하는 뇌파 및 음성 신호가 획득되고, 상기 발화 시 발생하는 뇌파를 임베딩 벡터로 변환되고, 상기 발화 시 발생하는 뇌파로부터 변환된 임베딩 벡터를 기반으로 mel-spectrogram을 생성되고, 상기 발화 시 발생하는 뇌파로부터 변환된 임베딩 벡터를 기반으로 mel-spectrogram과 상기 음성 신호의 mel-spectrogram을 기반으로 상기 생성자 및 판별자가 학습되는 것일 수 있다.

[0020] 일 실시 예에 따르면, 특징 벡터 변환부는, 발화 상상 뇌파의 음소 별 또는 단어 별 차이를 극대화하는 임베딩 공간으로 상기 사용자의 발화 상상 뇌파를 변환할 수 있다.

[0021] 일 실시 예에 따르면, 상기 음성 합성부는, 음성 인식 모델을 기반으로 상기 음성을 글자, 음소, 또는 발음으로 변환하는 글자 인식부를 포함할 수 있다.

[0022] 일 실시 예에 따르면, 변환된 상기 글자, 음소, 또는 발음은 생성자의 학습에 이용될 수 있다.

발명의 효과

[0023] 본 발명의 다양한 실시 예들에 따르면, 생성 모델과 음성 인식의 방법론을 사용함으로써 뇌파로부터 자유도 높은 사용자의 음성을 합성하여 자유도 높은 의사소통이 가능한 효과가 있다.

[0024] 본 발명의 효과들은 이상에서 언급한 효과들로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해 될 수 있을 것이다.

도면의 간단한 설명

[0025] 도 1은 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 방법의 학습 과정을 나타낸 도면이다.

도 2는 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 방법의 사용 과정을 나타낸 도면이다.

도 3은 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 장치의 구성을 나타낸 도면이다.

도 4는 본 발명의 일 실시 예에 따른 음성 합성부의 구체적인 구조를 나타낸 도면이다.

도 5는 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 장치의 구체적인 동작을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0026] 본 발명의 목적과 기술적 구성 및 그에 따른 작용 효과에 관한 자세한 사항은 본 발명의 명세서에 첨부된 도면에 의거한 이하의 상세한 설명에 의해 보다 명확하게 이해될 것이다. 첨부된 도면을 참조하여 본 발명에 따른 실시 예를 상세하게 설명한다.

[0027] 본 명세서에서 개시되는 실시 예들은 본 발명의 범위를 한정하는 것으로 해석되거나 이용되지 않아야 할 것이다. 이 분야의 통상의 기술자에게 본 명세서의 실시 예를 포함한 설명은 다양한 응용을 갖는다는 것이 당연하다. 따라서, 본 발명의 상세한 설명에 기재된 임의의 실시 예들은 본 발명을 보다 잘 설명하기 위한 예시적인 것이며 본 발명의 범위가 실시 예들로 한정되는 것을 의도하지 않는다.

[0028] 도면에 표시되고 아래에 설명되는 기능 블록들은 가능한 구현의 예들일 뿐이다. 다른 구현들에서는 상세한 설명의 사상 및 범위를 벗어나지 않는 범위에서 다른 기능 블록들이 사용될 수 있다. 또한, 본 발명의 하나 이상의 기능 블록이 개별 블록들로 표시되지만, 본 발명의 기능 블록들 중 하나 이상은 동일 기능을 실행하는 다양한 하드웨어 및 소프트웨어 구성들의 조합일 수 있다.

[0029] 또한, 어떤 구성요소들을 포함한다는 표현은 "개방형"의 표현으로서 해당 구성요소들이 존재하는 것을 단순히 지칭할 뿐이며, 추가적인 구성요소들을 배제하는 것으로 이해되어서는 안 된다.

[0030] 나아가 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급될 때에는, 그 다른 구성

요소에 직접적으로 연결 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 한다.

- [0031] 이하 도면을 참조하여 본 발명의 다양한 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 방법 및 장치에 대하여 상세히 설명한다. 이하 도면을 참조하여 설명할 본 발명의 다양한 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 방법의 각 단계는 발화 상상 시 뇌파 기반 음성 합성 장치에 의해 수행되는 것으로 이해될 수 있다. 또한 이하 본 발명을 설명함에 있어서 별도로 정의하지 않는 한 "장치"는 본 발명의 다양한 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 장치를 의미하는 것으로 이해될 수 있다.
- [0032] 뇌 신호로부터 사용자의 의도를 인식하는 기술은 최근 많이 개발되고 있고, 분류 인식 성능을 높이기 위한 딥러닝 기반의 방법론들이 많이 개발되고 있다. 그 중에 발화 상상 기반의 사용자 의도 인식 기술은 클래스의 수를 늘릴 수 있다는 점에서 장점이 있다. 하지만, 분류 기반의 사용자 의도 인식은 분류 클래스 수로 자유도가 한정된다는 점에서 한계가 있다. 따라서, 본 발명에서는 뇌 신호로부터 음성을 합성하는 음성 합성 방법론을 제안한다.
- [0033] 실제 발화 시 음성은 발화 상상 시의 뇌 신호와 차이가 있을 수 있다. 한편 실제 발화 시의 뇌 신호는 발화 시 음성과 시간적으로 동일하게 발생하면서, 발화 상상 시 발생하는 뇌 신호와 뇌파 활동 특징이 유사하다. 따라서, 실제 발화 시 발생하는 뇌 신호는 발화 상상 시 발생하는 뇌 신호와 발화 시 음성 사이를 연결하는 매개체 역할로 기능할 수 있다.
- [0034] 따라서, 본 발명은 실제 발화 시 음성은 발화 상상 시의 뇌 신호 간 차이를 보완하기 위한 방법으로서, 실제 발화 시 발생하는 뇌 신호와 음성 신호를 학습에 이용하는 방법을 제안한다. 실제 발화 시의 뇌 신호와 발화 시의 음성 신호를 획득한 후 발화 시 음성 신호를 ground truth(정답지)로 이용하여 발화 시 뇌 신호를 기반으로 생성자와 판별자를 학습시킨 후, 학습 결과를 바탕으로 발화 상상 시 뇌 신호를 이용하여 낮은 learning rate로 생성자와 분류자를 fine-tuning 기법을 통해 재학습 시킬 수 있다. 이하 도 1을 참조하여 상세히 설명한다.
- [0035] 도 1은 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 방법의 학습 과정을 나타낸 도면이다.
- [0036] 도 1을 참조하면, 장치는 사용자가 발화를 수행할 때 발생하는 사용자의 뇌파를 획득할 수 있다(S110). 여기서 발화 뇌파는 휴지 상태, 청각 또는 시각적 큐 제공 및 발화 상상의 과정을 반복하면서 획득될 수 있다. 일 예로, 장치는 2 초 간의 휴지 상태 사용자의 뇌파를 측정된 뒤, 청각 또는 시각적 큐로 발화를 수행할 클래스에 대한 정보를 2초간 사용자에게 제공할 수 있다. 이어지는 2초의 시간 동안 사용자는 주어진 클래스에 대해 실제 발화를 수행하며, 장치는 실제 발화를 수행하는 사용자의 뇌파와 함께 발화되는 음성 신호를 획득할 수 있다. 장치는 이러한 휴지 상태, 청각 또는 시각적 큐 제공 및 발화 상상의 과정을 n 회 반복하면서 사용자의 발화 뇌파 및 음성 신호를 획득할 수 있다. 발화 뇌파는, 뇌파가 특징적으로 발현되는 브로카-베르니케 영역을 포함하도록 채널을 배치하여 측정될 수 있으며, 샘플링 주파수는 1000Hz의 주파수로 뇌 신호를 수집할 수 있다. 다만, 이는 바람직한 예시를 설명하기 위한 것일 뿐 본 발명의 범위를 한정하지 않으며 채널 위치 또는 샘플링 주파수는 상황에 따라 변동될 수 있음에 유의한다. 뇌파에 포함되는 잡음은 전처리 과정을 통해 제거할 수 있다. 사용자가 발화를 수행할 때 발생하는 뇌파의 특징은 시간-주파수 특징을 활용하여 추출할 수 있다.
- [0037] 장치는 획득된 발화 시 뇌파를 기반으로 음소, 단어 별 구분을 극대화 하기 위해서 공통 공간 패턴(Common spatial pattern, CSP) 또는 log variance 등의 임베딩 변환 과정을 거쳐 임베딩 벡터를 생성할 수 있다(S120). 이러한 임베딩 벡터 변환 과정은 발화 상상의 음소 별 또는 단어 별 차이를 극대화할 수 있다. 또한 발화 시 뇌 신호와 발화 상상 시의 뇌 신호의 임베딩 변환 행렬을 공유함으로써, 이후 설명할 전이 학습을 용이하게 할 수 있다.
- [0038] 장치는 mel-spectrogram을 생성하기 위한 생성자(generator) 및 판별자를 학습 시킬 수 있다(S130). 여기서 생성자는 임베딩 벡터를 기반으로 mel-spectrogram을 생성한다. 생성자는 여러 컨볼루션 레이어(convolution layer), 레지듀얼 블록(residual block), 또는 어텐션 모듈(attention module) 등으로 이루어질 수 있다. 이러한 mel-spectrogram을 생성하는 생성자의 학습을 위해, 장치는 사용자 또는 다른 사용자의 음성 신호의 mel-spectrogram을 ground truth (정답지)로 사용하여 생성자를 학습시킬 수 있다. 즉, 생성자는 ground-truth로 주어지는 음성의 mel-spectrogram과 유사한 뇌 신호 기반의 mel-spectrogram을 생성할 수 있도록 학습될 수 있다. 또한, 장치는 뇌 신호로부터 생성된 mel-spectrogram과 실제 음성 신호를 기반으로 생성된 mel-spectrogram의 분류를 하는 판별자(discriminator)를 학습 시킬 수 있다. 판별자는 Adversarial learning 방식으로 뇌 신호로부터 생성된 mel-spectrogram과 음성 신호를 기반으로 생성된 mel-spectrogram을 구분하도록 학습되며, 생

성자는 이러한 판별자가 뇌 신호로부터 생성된 mel-spectrogram과 음성 신호를 기반으로 생성된 mel-spectrogram을 구분하지 못하도록 뇌 신호로부터 생성된 mel-spectrogram과 음성 신호를 기반으로 생성된 mel-spectrogram과 충분히 유사하도록 학습될 수 있다.

[0039] 생성자를 통해 생성된 mel-spectrogram은 보코더를 통해 음성을 생성할 수 있다(S140). 추가적으로, 생성된 음성 신호로부터 음성인식 모델을 통해 글자, 음소, 또는 발음으로 변환하고 실제 글자와의 차이를 통해 손실 함수를 구하고 이를 앞서 언급한 생성자의 학습에 반영할 수 있다. 여기서 보코더는 특정 mel-spectrogram으로부터 음성을 생성하도록 미리 학습된 보코더이며, 음성인식 모델은 음성으로부터 글자, 음소 등을 인식하도록 미리 학습된 음성인식 모델일 수 있으며, 본 발명을 실시 함에 있어서 보코더 및 음성 인식 모델에 대해 별도 학습이 필요치는 않을 수 있다.

[0040] 이상 실제 발화 시 뇌 신호와 발화 음성 신호를 기반으로 장치의 학습을 수행하는 설명하였으며, 아래에서는 발화 상상 시 뇌 신호를 기반의 학습 결과를 기반으로 발화 상상 뇌파를 이용하여 장치의 학습을 수행하는 방법을 설명한다.

[0041] 장치는 사용자가 발화 상상을 수행할 때 발생하는 사용자의 뇌파를 획득할 수 있다(S150). 여기서 발화 상상 뇌파는 휴지 상태, 청각 또는 시각적 큐 제공 및 발화 상상의 과정을 반복하면서 획득될 수 있다. 일 예로, 장치는 2 초 간의 휴지 상태 사용자의 뇌파를 측정된 뒤, 청각 또는 시각적 큐로 발화 상상을 수행할 클래스에 대한 정보를 2초간 사용자에게 제공할 수 있다. 이어지는 2초의 시간 동안 사용자는 주어진 클래스에 대한 발화 상상을 수행할 수 있다. 장치는 발화 상상을 수행하는 사용자의 뇌파를 수집할 수 있다. 장치는 이러한 휴지 상태, 청각 또는 시각적 큐 제공 및 발화 상상의 과정을 n 회 반복하면서 사용자의 발화 상상 뇌파를 획득할 수 있다. 발화 상상 뇌파는, 발화 상상 시 뇌파가 특징적으로 발현되는 브로카-베르니케 영역을 포함하도록 채널을 배치하여 측정될 수 있으며, 샘플링 주파수는 1000Hz의 주파수로 뇌 신호를 수집할 수 있다. 다만, 이는 바람직한 예시를 설명하기 위한 것일 뿐 본 발명의 범위를 한정하지 않으며 채널 위치 또는 샘플링 주파수는 상황에 따라 변동될 수 있음에 유의한다. 뇌파에 포함되는 잡음은 전처리 과정을 통해 제거할 수 있다. 사용자가 발화 상상 시 발생하는 뇌파의 특징은 시간-주파수 특징을 활용하여 추출할 수 있다.

[0042] 장치는 획득된 뇌파를 기반으로 음소, 단어 별 구분을 극대화 하기 위해서 공통 공간 패턴(Common spatial pattern, CSP) 또는 log variance 등의 임베딩 변환 과정을 거쳐 임베딩 벡터를 생성할 수 있다(S160). 이러한 임베딩 벡터 변환 과정은 발화 상상의 음소 별 또는 단어 별 차이를 극대화할 수 있다. 또한 발화 시 뇌 신호와 발화 상상 시의 뇌 신호의 임베딩 변환 행렬을 공유함으로써, 이후 설명할 전이 학습을 용이하게 할 수 있다.

[0043] 장치는 mel-spectrogram을 생성하기 위한 생성자(generator) 및 판별자를 학습 시킬 수 있다(S170). 여기서 생성자는 임베딩 벡터를 기반으로 mel-spectrogram을 생성한다. 생성자는 여러 컨볼루션 레이어(convolution layer), 레지듀얼 블록(residual block), 또는 어텐션 모듈(attention module) 등으로 이루어질 수 있다. 일 실시 예에 따르면, 발화 상상 뇌파를 기반으로 한 학습 과정에서는 앞서 설명한 실제 발화 시 뇌 신호 및 실제 발화 시 음성 신호를 기반으로 학습한 생성자 및 판별자를 전이 학습을 통해 재학습시키며, 전이 학습을 통해 재학습된 생성자를 통해 장치는 발화 상상 뇌파를 이용해 생성된 임베딩 벡터를 기반으로 mel-spectrogram을 생성할 수 있다. 이처럼 먼저 실제 발화 시의 뇌 신호와 실제 발화 시 음성 신호를 기반으로 학습된 생성자 및 판별자를 이용해 전이 학습을 수행함으로써 발화 상상 시 뇌 신호를 이용하여 낮은 learning rate로 생성자와 분류자를 fine-tuning 기법을 통해 생성자를 재학습 시킬 수 있다

[0044] 생성자를 통해 생성된 mel-spectrogram은 보코더를 통해 음성을 생성할 수 있다(S180). 추가적으로, 생성된 음성 신호로부터 음성인식 모델을 통해 글자, 음소, 또는 발음으로 변환하고 실제 글자와의 차이를 통해 손실 함수를 구하고 이를 이용해 생성자를 재학습할 수 있다. 여기서 보코더는 특정 mel-spectrogram으로부터 음성을 생성하도록 미리 학습된 보코더이며, 음성인식 모델은 음성으로부터 글자, 음소 등을 인식하도록 미리 학습된 음성인식 모델일 수 있으며, 본 발명을 실시 함에 있어서 보코더 및 음성 인식 모델에 대해 별도 학습이 필요치는 않을 수 있다.

[0045] 도 2는 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 방법의 사용 과정을 나타낸 도면이다.

[0046] 장치는 사용자가 발화 상상을 수행할 때 발생하는 사용자의 뇌파를 획득할 수 있다(S210). 여기서 발화 상상 뇌파는 사용자의 의도가 포함된 발화 상상을 수행할 경우 발생하는 뇌파일 수 있다. 발화 상상 뇌파의 획득은 학습 과정에서와 마찬가지로 발화 상상 시 뇌파가 특징적으로 발현되는 브로카-베르니케 영역을 포함하도록 채널을 배치하여 측정될 수 있으며, 샘플링 주파수는 1000Hz의 주파수로 뇌 신호를 수집할 수 있다. 다만, 이는 바

람직한 예시를 설명하기 위한 것일 뿐 본 발명의 범위를 한정하지 않으며 채널 위치 또는 샘플링 주파수는 상황에 따라 변동될 수 있음에 유의한다. 뇌파에 포함되는 잡음은 전처리 과정을 통해 제거할 수 있다. 사용자가 발화 상상 시 발생하는 뇌파의 특징은 시간-주파수 특징을 활용하여 추출할 수 있다.

- [0047] 장치는 획득된 뇌파를 기반으로 학습 과정과 동일하게 임베딩 변환 행렬을 이용하여 임베딩 벡터를 생성할 수 있다(S220).
- [0048] 장치는 학습 과정에서 학습시킨 생성자를 기반으로, 생성된 임베딩 벡터를 이용하여 mel-spectrogram을 생성할 수 있다(S230).
- [0049] 생성자를 통해 생성된 mel-spectrogram은 보코더를 통해 음성을 생성하여 출력할 수 있다(S240). 여기서 보코더는 특정 mel-spectrogram으로부터 음성을 생성하도록 미리 학습된 보코더이며 바람직하게는 학습 과정에서 사용된 것과 동일하게 학습된 보코더일 수 있다.
- [0050] 이상 도 1 및 도 2를 참조하여 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 방법을 실시하기 위한 학습 과정 및 사용 과정을 설명하였다. 아래에서는 전술한 방법을 실시하기 위한 뇌파 기반 음성 합성 장치의 구성에 대하여 설명한다.
- [0051] 도 3은 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 장치의 구성을 나타낸 도면이다.
- [0052] 도 3을 참조하면, 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 장치(300)는 뇌 신호 및 음성 인식부(310), 특징 벡터 변환부(320), 음성 합성부(330) 및 음성 출력부(340)를 포함할 수 있다.
- [0053] 뇌 신호 및 음성 인식부(310)는 실제 발화 시의 뇌 신호 및 음성, 그리고 발화 상상 시의 뇌 신호를 획득할 수 있다. 예를 들어, 학습 과정에서 장치(300)는 학습을 위해 사용자의 발화 상상 뇌 신호 패턴을 분석하기 위한 다양한 발화 및 발화 상상을 유도하며, 뇌 신호 및 음성 인식부(310)는 실제 발화 시 발생하는 뇌 신호, 실제 발화 음성 신호, 그리고 발화 상상 시 발생하는 뇌 신호를 사용자로부터 획득할 수 있다. 또한, 사용 과정에서 뇌 신호 및 음성 인식부(310)는 사용자의 의도가 포함된 발화 상상 뇌파를 획득할 수 있다.
- [0054] 특징 벡터 변환부(320)는 획득된 뇌 신호를 임베딩 벡터로 변환하며, 획득된 뇌 신호로부터 음소 별 또는 단어 별 차이를 극대화하는 임베딩 공간으로 뇌 신호를 변환할 수 있다. 일 실시 예에서, 특징 벡터 변환부(320)는 공통 공간 패턴(common spatial pattern, CSP) 또는 log variance 등을 기반으로 뇌 신호를 임베딩 벡터로 변환할 수 있다.
- [0055] 음성 합성부(330)는 특징 벡터 변환부(320)를 통해 변환된 임베딩 벡터를 기반으로 mel-spectrogram을 생성하고, 생성된 mel-spectrogram을 기반으로 음성을 생성할 수 있다. 음성 합성부(330)의 구체적인 구성은 이후 도 4를 참조하여 후술한다.
- [0056] 음성 출력부(340)는 음성 합성부(330)를 통해 합성된 음성을 출력할 수 있다.
- [0057] 도 4는 본 발명의 일 실시 예에 따른 음성 합성부의 세부 구성을 나타낸 도면이다.
- [0058] 도 4를 참조하면, 본 발명의 일 실시 예에 따른 장치(300)의 음성 합성부(330)는 mel-spectrogram 생성부(410), 판별부(420), 음성 생성부(430), 글자 인식부(440)를 포함할 수 있다.
- [0059] Mel-spectrogram 생성부(410)는 임베딩 벡터를 이용하여 mel-spectrogram을 생성하는 생성자(generator)를 포함하며, 특징 벡터 변환부(320)를 통해 생성된 임베딩 벡터를 이용하여 mel-spectrogram을 생성할 수 있다. Mel-spectrogram 생성부(410)의 생성자는 학습 과정에서 실제 발화 시 발생하는 뇌 신호 및 실제 발화 음성 신호를 기반으로 학습되고, 이후 전이 학습을 통해 발화 상상 시 발생하는 뇌 신호를 기반으로 재학습이 수행될 수 있다.
- [0060] 판별부(420)는 생성된 mel-spectrogram이 뇌파를 기반으로 생성된 것인지, 또는 음성 신호를 기반으로 생성된 것인지 여부를 판별하며 이를 위해 판별자(discriminator)를 포함할 수 있다. 판별부(420)에 포함된 판별자는 판별자는 Adversarial learning 방식으로 뇌 신호로부터 생성된 mel-spectrogram과 음성 신호를 기반으로 생성된 mel-spectrogram을 구분하도록 학습될 수 있다.
- [0061] 음성 생성부(430)는 mel-spectrogram 생성부(410)를 통해 생성된 mel-spectrogram을 이용하여 기학습된 보코더를 통해 음성을 생성하며, 생성된 음성을 음성 출력부(340)로 전달하여 음성으로 출력되도록 할 수 있다.
- [0062] 글자 인식부(440)는 음성 생성부(430)를 통해 생성된 음성을 기학습된 음성 인식 모델을 통하여 글자, 음소, 또

는 발음으로 변환하며, 이를 mel-spectrogram 생성부(410)의 생성자의 재학습에 이용할 수 있다.

[0063] 한편, 이상과 같은 음성 합성부(330)에서 이루어지는 학습은 글자 또는 음소 단위로 진행될 수 있으며, 그에 따라 학습된 글자 또는 음소의 조합을 통해 학습되지 않은 단어 및 문장들도 합성할 수 있다 할 것이다.

[0064] 도 5는 본 발명의 일 실시 예에 따른 발화 상상 시 뇌파 기반 음성 합성 장치의 구체적인 동작을 나타낸 도면이다.

[0065] 학습 과정 또는 사용 과정에서 장치는 뇌 신호 및 음성 인식부를 통해 사용자의 실제 발화 시의 뇌 신호와 음성 신호, 그리고 발화 상상 시의 뇌 신호를 획득할 수 있다. 이 중 실제 발화 시의 뇌 신호와 발화 상상 시의 뇌 신호는 특징 벡터 변환부를 통해 임베딩 벡터로 변환되며, 변환된 임베딩 벡터는 음성 합성부에 포함된 mel-spectrogram 생성부의 생성자를 통해 mel-spectrogram으로 변환될 수 있다. 임베딩 벡터를 mel-spectrogram으로 변환하는 생성자의 학습 방법은 앞서 설명하였으므로 여기에서는 상세한 설명은 설명하도록 한다. 생성된 mel-spectrogram은 기 학습된 보코더를 이용해 음성 생성부에서 음성으로 생성되며, 생성된 음성은 음성 출력부를 통해 출력된다.

[0066] 한편, 실제 발화 시의 음성 신호는 직접 mel-spectrogram으로 변환되며, 판별부의 판별자는 생성자를 통해 뇌파를 기반으로 생성된 mel-spectrogram과 실제 음성 신호를 기반으로 생성된 mel-spectrogram을 구분하도록 학습되며, 따라서 판별자가 뇌 신호로부터 생성된 mel-spectrogram과 음성 신호를 기반으로 생성된 mel-spectrogram을 구분하지 못하게 mel-spectrogram을 생성하도록 생성자를 학습 시킬 수 있다.

[0067] 음성 생성부에서 생성된 음성은 글자 인식부에서 기학습된 음성 인식 모델을 통하여 글자, 음소, 또는 발음 등으로 변환되며, 이를 생성자의 학습에 이용할 수 있다.

[0068] 한편 도 3 내지 도 5에 도시되지 않았더라도 장치는 장치에 포함된 일련의 구성을 제어하거나 동작시키기 위해 필요한 적어도 하나의 프로세서, 상기 프로세서에 의해 수행되는 컴퓨터 프로그램을 로드(load)하기 위한 메모리, 그리고 상기 컴퓨터 프로그램을 저장하기 위한 스토리지(storage) 등을 더 포함할 수 있음에 유의한다. 일 예시로, 프로세서는 CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit) 또는 본 발명이 속하는 기술 분야에서 널리 알려져 있는 형태의 프로세서 중 어느 하나일 수 있다. 메모리는 각종 정보, 명령 및/또는 정보를 저장하며, 본 발명의 일 실시 예에 따른 뇌파 기반의 사용자 보조 방법을 수행하기 위해 스토리지로부터 하나 이상의 컴퓨터 프로그램을 로드할 수 있다. 스토리지는 하나 이상의 컴퓨터 프로그램 및 대용량 네트워크 정보를 비임시적으로 저장할 수 있다. 이러한 스토리지는 ROM(Read Only Memory), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM), 플래시 메모리 등과 같은 비휘발성 메모리, 하드 디스크(HDD), 보조 저장 매체(SSD), 착탈형 디스크, 또는 본 발명이 속하는 기술 분야에서 널리 알려져 있는 임의의 형태의 컴퓨터로 읽을 수 있는 기록 매체 중 어느 하나일 수 있다. 컴퓨터 프로그램은 메모리에 로드되어 적어도 하나의 프로세서에 의해 실행되며, 본 발명의 일 실시 예에 따른 뇌파 기반의 사용자 보조 방법을 수행할 수 있다.

[0069] 이상 첨부된 도면을 참조하여 본 발명의 실시 예들을 설명하였지만, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시 예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

부호의 설명

[0070] 300: 발화 상상 시 뇌파 기반 음성 합성 장치

310: 뇌 신호 및 음성 인식부

320: 특징 벡터 변환부

330: 음성 합성부

340: 음성 출력부

410: mel-spectrogram 생성부

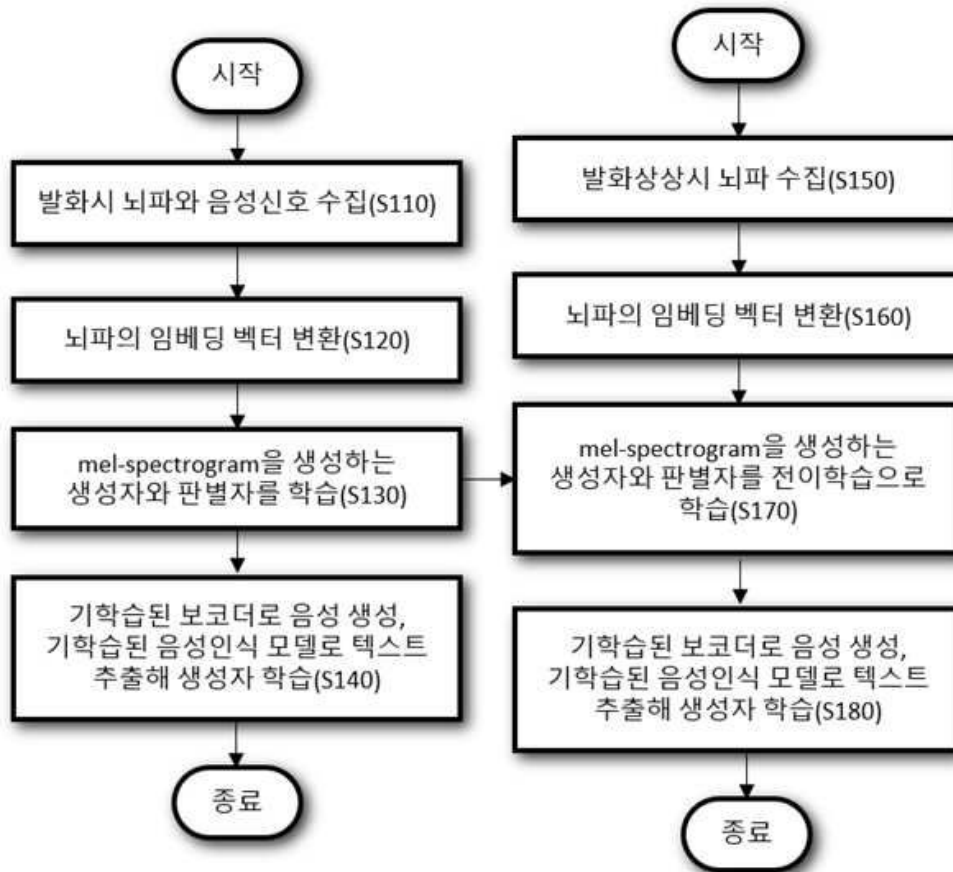
420: 판별부

430: 음성 생성부

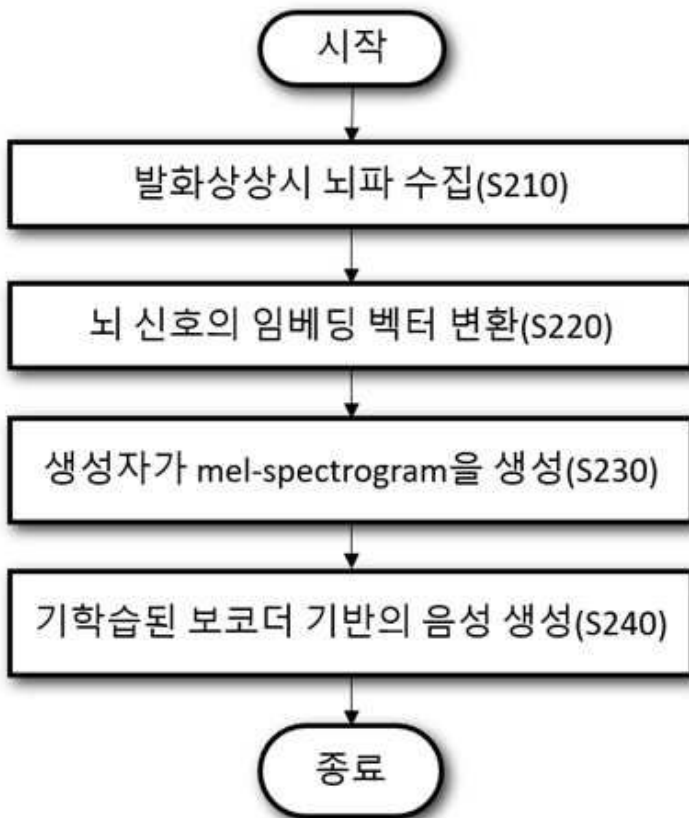
440: 글자 인식부

도면

도면1



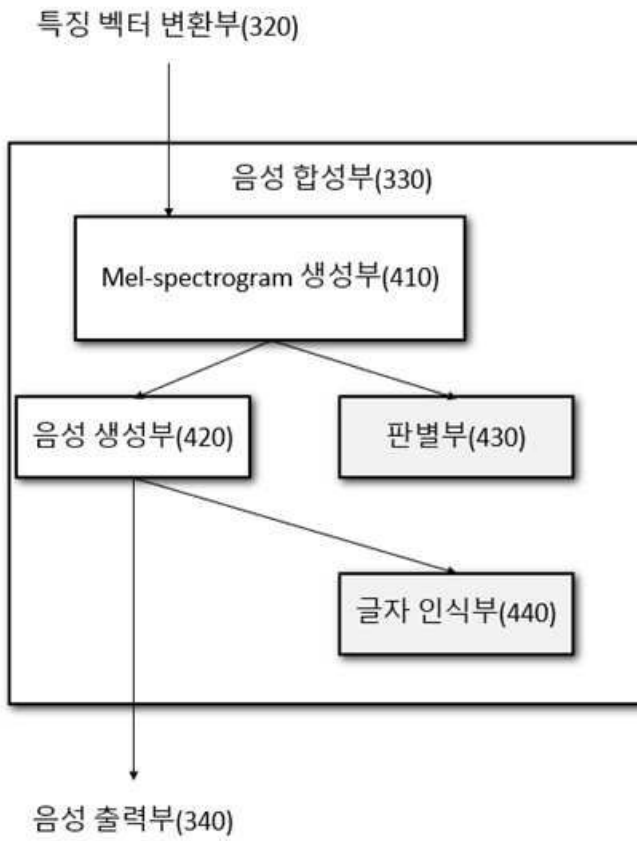
도면2



도면3



도면4



도면5

