



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0107676
(43) 공개일자 2010년10월06일

(51) Int. Cl.

G10L 21/06 (2006.01) G10L 11/00 (2006.01)

(21) 출원번호 10-2009-0025889

(22) 출원일자 2009년03월26일

심사청구일자 2009년03월26일

(71) 출원인

고려대학교 산학협력단

서울 성북구 안암동5가 1

(72) 발명자

육동석

서울 용산구 서빙고동 신동아아파트 1동 102호

임현택

경기도 안산시 상록구 성포동 선경아파트 17동 801호

(74) 대리인

특허법인충현

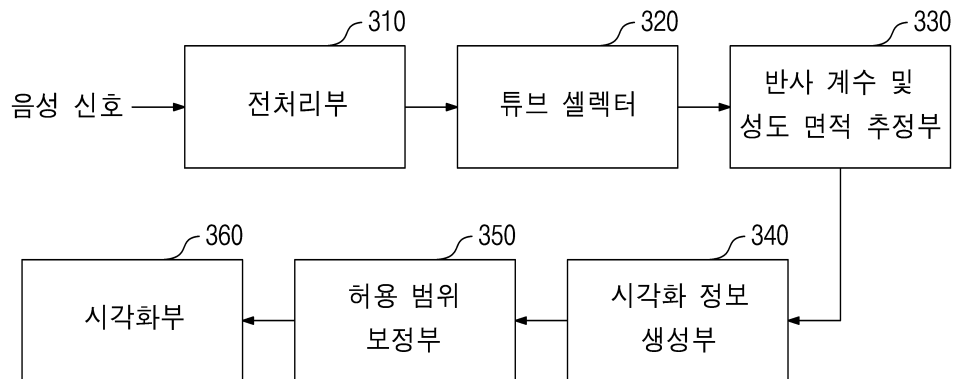
전체 청구항 수 : 총 16 항

(54) 음성 인식을 이용한 음성 시각화 방법 및 장치

(57) 요약

본 발명은 음성 인식을 이용한 음성 시각화 방법에 관한 것으로서 마이크를 통해 음성 신호를 입력받는 단계; 상기 입력된 음성 신호로부터 인식된 음운에 대응하는 하나 이상의 튜브들을 병합하여 멀티 튜브를 산출하는 단계; 상기 멀티 튜브를 이용하여 성도 면적을 추정하는 단계; 및 음운마다 해당하는 조음 구조에 대한 시각 정보의 데이터베이스를 이용하여 상기 추정된 성도 면적을 보정하는 단계를 포함하는 것을 특징으로 하며, 멀티 튜브 모델 및 보간 처리를 이용하여 음성을 시각화함으로써, 많은 비용이 드는 MRI, 초음파와 같은 측정 수단을 통하지 않고 일반적인 컴퓨터를 이용하여 화자 자신의 발화를 손쉽게 정확하게 시각화하여 표시할 수 있으므로, 청각 장애를 가진 사람이나 외국어 언어 학습자 등이 언어 학습 능력을 향상시킬 수 있다.

대표도 - 도3



특허청구의 범위

청구항 1

실시간으로 발화되는 음성에 대응하여 조음 기관의 이미지를 생성하는 방법에 있어서,

마이크를 통해 음성 신호를 입력받는 단계;

상기 입력된 음성 신호로부터 인식된 음운에 대응하는 하나 이상의 튜브들을 병합하여 멀티 튜브를 산출하는 단계;

상기 멀티 튜브를 이용하여 성도 면적을 추정하는 단계; 및

음운마다 해당하는 조음 구조에 대한 시각 정보의 데이터베이스를 이용하여 상기 추정된 성도 면적을 보정하는 단계를 포함하는, 음성 인식을 이용한 음성 시각화 방법.

청구항 2

제 1 항에 있어서,

상기 멀티 튜브를 산출하는 단계는,

상기 음성 신호에서 음성 구간을 검출하는 단계를 포함하는 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 방법.

청구항 3

제 1 항에 있어서,

상기 시각 정보의 데이터베이스는,

성도에 대한 초음파 이미지, MRI 사진 또는 X레이 사진 중 적어도 한 형태의 이미지를 각 음운에 매칭시킨 데이터베이스인 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 방법.

청구항 4

제 1 항에 있어서,

상기 추정된 성도 면적을 보정하는 단계는,

상기 음운에서 모음의 포먼트를 측정하는 단계;

상기 측정된 포먼트에 해당하는 혀의 모양을 산출하는 단계; 및

상기 산출된 혀의 모양을 이용하여 상기 추정된 성도 면적을 보정하는 단계를 포함하는 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 방법.

청구항 5

제 4 항에 있어서,

상기 혀의 모양을 산출하는 단계는,

각 모음마다 포먼트 값을 정의한 제1음운론 데이터베이스에서 상기 측정된 포먼트에 해당하는 포먼트 값의 모음 정보를 검색하는 단계; 및

모음 정보마다 혀의 모양을 시각화한 제2음운론 데이터베이스에서 상기 검색된 모음 정보에 대응하는 혀의 모양을 검색하는 단계를 포함하는 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 방법.

청구항 6

제 1 항에 있어서,

상기 보정된 성도 면적을 이용하여 상기 음운에 대해 추정된 발음 구조 이미지를 생성하는 단계를 더 포함하는 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 방법.

청구항 7

제 6 항에 있어서,

상기 추정된 발음 구조 이미지는,

상기 보정된 성도 면적에 따른 조음 기관의 이미지에 혀의 모양 및 혀의 위치를 표시한 이미지인 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 방법.

청구항 8

제1항 내지 제7항 중 어느 한 항의 방법을 컴퓨터 시스템에서 실행하기 위한 프로그램이 기록된, 컴퓨터 시스템이 판독할 수 있는 기록매체.

청구항 9

실시간으로 발화되는 음성에 대응하여 조음 기관의 이미지를 생성하는 장치에 있어서,

음성 신호를 입력받는 마이크;

상기 입력된 음성 신호로부터 인식된 음운에 대응하는 하나 이상의 튜브들을 병합하여 멀티 튜브를 산출하는 튜브 선택터;

상기 멀티 튜브를 이용하여 반사 계수와 상기 반사 계수에 따른 성도 면적을 추정하는 반사 계수 및 성도 면적 추정부;

음운마다 해당하는 조음 구조에 대한 시각 정보를 저장하는 시각 정보 데이터베이스; 및

상기 시각 정보 데이터베이스를 이용하여 상기 추정된 성도 면적을 보정하는 허용 범위 보정부를 포함하는, 음성 인식을 이용한 음성 시각화 장치.

청구항 10

제 9 항에 있어서,

상기 음성 신호에서 음성 구간을 검출하고 검출된 음성 구간 내의 음성 신호를 상기 튜브 선택터에 전달하는 전처리부를 더 포함하는 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 장치.

청구항 11

제 9 항에 있어서,

상기 시각 정보 데이터베이스는,

성도에 대한 초음파 이미지, MRI 사진 또는 X레이 사진 중 적어도 한 형태의 이미지를 각 음운에 매칭시킨 데이터베이스인 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 장치.

청구항 12

제 9 항에 있어서,

상기 허용 범위 보정부는,

상기 음운에서 모음의 포먼트를 측정하고, 상기 측정된 포먼트에 해당하는 혀의 모양을 산출하며, 상기 산출된 혀의 모양을 이용하여 상기 추정된 성도 면적을 보정하는 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 장치.

청구항 13

제 12 항에 있어서,

상기 허용 범위 보정부는,

각 모음마다 포먼트 값을 정의한 제1음운론 데이터베이스에서 상기 측정된 포먼트에 해당하는 포먼트 값의 모음

정보를 검색하고, 모음 정보마다 혀의 모양을 시각화한 제2음운론 데이터베이스에서 상기 검색된 모음 정보에 대응하는 혀의 모양을 검색하는 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 장치.

청구항 14

제 9 항에 있어서,

상기 보정된 성도 면적을 이용하여 상기 음운에 대해 추정된 발음 구조 이미지를 생성하는 시각화부를 더 포함하는 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 장치.

청구항 15

제 14 항에 있어서,

상기 추정된 발음 구조 이미지는,

상기 보정된 성도 면적에 따른 조음 기관의 이미지에 혀의 모양 및 혀의 위치를 표시한 이미지인 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 장치.

청구항 16

제 9 항에 있어서,

상기 추정된 성도 면적을 이용하여 예비적인 조음 구조 이미지를 생성하고, 생성된 조음 구조 이미지를 상기 허용 범위 보정부의 성도 면적 보정에 적용하는 시각화 정보 생성부를 더 포함하는 것을 특징으로 하는, 음성 인식을 이용한 음성 시각화 장치.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 음성 인식을 이용한 음성 시각화 방법에 관한 것으로, 더욱 상세하게는, 화자 자신의 발화를 손쉽게 정확하게 시각화하여 표시할 수 있는 음성 인식을 이용한 음성 시각화 방법, 그 기록 매체, 및 그 장치에 관한 것이다.

배경 기술

[0002] 음성 시각화란 음성이 발생될 때의 입술의 모양, 혀의 위치, 혀의 모양, 성도의 움직임 등을 시각적으로 보여주는 기술이다. 음성 시각화에서는 다양한 방법으로 측정된 성도의 면적을 통해 사용자가 이해하기 쉬운 그림이나 동영상으로 성도의 모습을 표현하는 것이 중요하다.

[0003] 성도의 면적을 측정하기 위한 방법에는 음성신호를 기반으로 하여 LPC (Linear Prediction Coefficient) 모델을 적용한 기법과 MRI (Magnetic Resonance Imaging), 초음파 장치 또는 X-레이와 같은 고가의 영상장치를 이용한 측정법이 있다. 영상 장치를 이용한 측정 방법은 많은 시간과 비용이 소모가 된다. 또한 영상을 분석해야 하는 시간이 필요 하므로 실시간으로 본인의 발화시 성도 모습을 확인하여야 하는 언어 학습자가 사용하기에는 부적합 하다. 음성신호를 기반으로 성도의 면적을 추정하는 방법은 실시간으로 수치화된 성도의 면적 값을 얻을 수 있지만 언어학습자가 이해하기 쉬운 형태의 시각화 모델로 표현되지 않는 단점을 지니고 있다.

[0004] 영상장치 또는 음성신호를 기반으로 하여 측정된 성도의 면적을 통해 시각화하는 기법으로 위드와이즈 매핑 (Widthwise Mapping) 기법이 있다. 하지만 이 방법은 추정된 면적 값을 단순히 히스토그램화 하여 표현한다. 이 방법으로는 혀의 위치, 구강의 모습들을 포함하여 정확한 성도의 모습을 나타내지 못한다. 또한, 비음을 발화할 때는 구강뿐만 아니라 비강을 함께 이용하여 발화를 하게 된다. LPC 기법으로 추정된 성도 면적에는 비강의 넓이와 구강의 넓이가 함께 더 해져서 나타나게 된다.

[0005] 한편, 실제 음성신호에서 비음 구간을 검출 하기 위하여서 HMM (Hidden Markov Models)을 이용하여 음성 인식을 한다. HMM은 음의 상태가 한 상태에서 다음 상태로 바뀌는 것을 천이 확률로 표현한다. HMM은 음성 신호의 시간적인 통계적 특성을 이용하여 훈련 데이터로부터 이들을 대표하는 모델을 구성한 후 실제 음성 신호와 유사도가

높은 확률 모델을 인식 결과로 채택하는 방법이다. 이 방법은 단독음이나 연결음, 연속음 인식에까지 구현이 용이하며 좋은 인식 성능을 나타내어 여러 가지 응용 분야에 많이 이용되고 있다. 실제로 음성 인식 기술이 대중화된 계기는 HMM의 등장이라고 할 수 있다. HMM은 수학적인 배경에서 개발된 알고리즘으로 전통적인 확률분포를 이용하며, 시간 정보와 잘 연동되기 때문에 화자 독립, 대화체 음성 인식 등 많은 장점을 갖고 있다. 또한 대어휘에서 DTW(Dynamic Time Warping)보다는 계산량이 적은 장점을 갖고 있다. 그러나 학습 데이터가 부족할 경우, 모델간의 변별력이 부족하고 음성 신호간의 연관성을 무시하는 경향이 있다.

- [0006] 이와 같이, 음성의 시각화는 1970년대 이후 관련 연구분야에 대해 많은 연구가 있었다. 하지만 가장 큰 문제점은 소리 정보를 이용하여 인간의 성문에서부터 입에 이르는 부분을 정확히 묘사할 수 없는 문제점이 있었다.
- [0007] 최근에는 소리를 통한 언어 학습이 아닌 MRI, 초음파를 이용하여 조음 구조에 대한 시각 정보를 얻고 있지만, 이는 많은 비용이 소요되고 장비 특성상 장시간 이용에 어려움이 있으므로, 청각 장애를 가진 일반적인 언어 학습자에게는 용이한 방법이 아니다.

발명의 내용

해결 하고자 하는 과제

- [0008] 소리 정보를 이용하여 인간의 성문에서 입술에 이르는 구간의 모양을 특수한 장비 없이 손쉽게 정확하게 표현할 수 있는 방법이 필요하다. 따라서, 본 발명이 해결하고자 하는 첫 번째 기술적 과제는 일반적인 컴퓨터를 이용하여 화자 자신의 발화를 손쉽게 정확하게 시각화할 수 있는 음성 인식을 이용한 음성 시각화 방법을 제공하는 데 있다.
- [0009] 본 발명이 해결하고자 하는 두 번째 기술적 과제는 많은 비용이 드는 MRI, 초음파와 같은 측정 수단을 통하지 않고 일반적인 컴퓨터를 이용하여 화자 자신의 발화를 손쉽게 정확하게 시각화할 수 있는 음성 인식을 이용한 음성 시각화 장치를 제공하는 데 있다.

과제 해결수단

- [0010] 본 발명은 상기 첫 번째 과제를 달성하기 위하여, 마이크를 통해 음성 신호를 입력받는 단계; 상기 입력된 음성 신호로부터 인식된 음운에 대응하는 하나 이상의 튜브들을 병합하여 멀티 튜브를 산출하는 단계; 상기 멀티 튜브를 이용하여 성도 면적을 추정하는 단계; 및 음운마다 해당하는 조음 구조에 대한 시각 정보의 데이터베이스를 이용하여 상기 추정된 성도 면적을 보정하는 단계를 포함하는 음성 인식을 이용한 음성 시각화 방법을 제공한다.
- [0011] 본 발명의 일 실시 예에 의하면, 상기 멀티 튜브를 산출하는 단계는 상기 음성 신호에서 음성 구간을 검출하는 단계를 포함할 수 있다.
- [0012] 본 발명의 다른 실시 예에 의하면, 상기 시각 정보의 데이터베이스는 성도에 대한 초음파 이미지, MRI 사진 또는 X레이 사진 중 적어도 한 형태의 이미지를 각 음운에 매칭시킨 데이터베이스일 수 있다.
- [0013] 본 발명의 또 다른 실시 예에 의하면, 상기 추정된 성도 면적을 보정하는 단계는 상기 음운에서 모음의 포먼트를 측정하는 단계, 상기 측정된 포먼트에 해당하는 혀의 모양을 산출하는 단계, 및 상기 산출된 혀의 모양을 이용하여 상기 추정된 성도 면적을 보정하는 단계를 포함할 수 있다. 보다 구체적으로, 상기 혀의 모양을 산출하는 단계는 각 모음마다 포먼트 값을 정의한 제1음운론 데이터베이스에서 상기 측정된 포먼트에 해당하는 포먼트 값의 모음 정보를 검색하는 단계, 및 모음 정보마다 혀의 모양을 시각화한 제2음운론 데이터베이스에서 상기 검색된 모음 정보에 대응하는 혀의 모양을 검색하는 단계를 포함할 수 있다.
- [0014] 본 발명의 또 다른 실시 예에 의하면, 상기 보정된 성도 면적을 이용하여 상기 음운에 대해 추정된 발음 구조 이미지를 생성하는 단계를 더 포함할 수 있다. 이 경우, 상기 추정된 발음 구조 이미지는 상기 보정된 성도 면적에 따른 조음 기관의 이미지에 혀의 모양 및 혀의 위치를 표시한 이미지일 수 있다.
- [0015] 본 발명은 상기 첫 번째 과제를 달성하기 위하여, 음성 신호를 입력받는 마이크; 상기 입력된 음성 신호로부터 인식된 음운에 대응하는 하나 이상의 튜브들을 병합하여 멀티 튜브를 산출하는 튜브 셀렉터; 상기 멀티 튜브를 이용하여 반사 계수와 상기 반사 계수에 따른 성도 면적을 추정하는 반사 계수 및 성도 면적 추정부; 음운마다 해당하는 조음 구조에 대한 시각 정보를 저장하는 시각 정보 데이터베이스; 및 상기 시각 정보 데이터베이스를 이용하여 상기 추정된 성도 면적을 보정하는 허용 범위 보정부를 포함하는 음성 인식을 이용한 음성 시각화 장

치를 제공한다.

- [0016] 본 발명의 일 실시 예에 의하면, 상기 음성 신호에서 음성 구간을 검출하고 검출된 음성 구간 내의 음성 신호를 상기 튜브 셀렉터에 전달하는 전처리부를 더 포함한다.
- [0017] 본 발명의 다른 실시 예에 의하면, 상기 시각 정보 데이터베이스는 성도에 대한 초음파 이미지, MRI 사진 또는 X레이 사진 중 적어도 한 형태의 이미지를 각 음운에 매칭시킨 데이터베이스일 수 있다.
- [0018] 본 발명의 또 다른 실시 예에 의하면, 상기 허용 범위 보정부는 상기 음운에서 모음의 포먼트를 측정하고, 상기 측정된 포먼트에 해당하는 혀의 모양을 산출하며, 상기 산출된 혀의 모양을 이용하여 상기 추정된 성도 면적을 보정할 수 있다. 이 경우, 보다 구체적으로, 상기 허용 범위 보정부는 각 모음마다 포먼트 값을 정의한 제1음운론 데이터베이스에서 상기 측정된 포먼트에 해당하는 포먼트 값의 모음 정보를 검색하고, 모음 정보마다 혀의 모양을 시각화한 제2음운론 데이터베이스에서 상기 검색된 모음 정보에 대응하는 혀의 모양을 검색할 수 있다.
- [0019] 본 발명의 또 다른 실시 예에 의하면, 상기 보정된 성도 면적을 이용하여 상기 음운에 대해 추정된 발음 구조 이미지를 생성하는 시각화부를 더 포함할 수 있다. 여기서, 상기 추정된 발음 구조 이미지는 상기 보정된 성도 면적에 따른 조음 기관의 이미지에 혀의 모양 및 혀의 위치를 표시한 이미지일 수 있다.
- [0020] 본 발명의 또 다른 실시 예에 의하면, 상기 추정된 성도 면적을 이용하여 예비적인 조음 구조 이미지를 생성하고, 생성된 조음 구조 이미지를 상기 허용 범위 보정부의 성도 면적 보정에 적용하는 시각화 정보 생성부를 더 포함할 수 있다.

효 과

- [0021] 본 발명에 의하면, 멀티 튜브 모델 및 보간 처리를 이용하여 음성을 시각화함으로써, 많은 비용이 드는 MRI, 초음파와 같은 측정 수단을 통하지 않고 일반적인 컴퓨터를 이용하여 화자 자신의 발화를 손쉽게 정확하게 시각화하여 표시할 수 있으므로, 청각 장애를 가진 사람이나 외국어 언어 학습자 등이 언어 학습 능력을 향상시킬 수 있다.

발명의 실시를 위한 구체적인 내용

- [0022] 이하에서는 도면을 참조하여 본 발명의 바람직한 실시 예를 설명하기로 한다. 그러나, 다음에 예시하는 본 발명의 실시 예는 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 다음에 상술하는 실시 예에 한정되는 것은 아니다.
- [0023] 음성 인식은 발화한 음성을 텍스트로 맵핑하는 일종의 패턴 인식의 한 분야이다. 본 발명에서는 기본적인 인식 이외에도 각 음소별 최적 튜브 모델 탐색 및 인식 결과를 음운론에 기반 이미지 정보를 보정 처리 하는 과정에 음성 인식을 이용한다.
- [0024] 본 발명에서는 음성 인식을 위해 HMM을 이용하는데, 다양한 음성정보 및 최적 튜브 모델 정보 그리고 보정을 위한 이미지 정보를 이용하여 음성 모델을 훈련할 필요가 있다.
- [0025] 이러한 과정을 거친 후, 테스트 데이터를 이용하여 성능을 측정한다. 이때 특정 음성 인식 과정에서 어떠한 튜브 모델이 적용되는지, 음운론을 적용시킬 음소 정보는 무엇인지를 확인한다. 다음으로 음성 모델에 대해 훈련 단계 및 특징 추출과정을 거친다.
- [0026] 도 1은 음성 인식을 이용한 음성 시각화의 개략적인 방법을 도시한 것이다.
- [0027] 먼저, 음성 구간 검출(VAD)을 통하여 음성 구간을 검출하고 여기 검출된 음성구간에 대해서 음성을 인식한다. 음성 인식 과정에서 인식된 결과 이외에 최적 튜브 파라미터가 나오게 된다. 여기서 얻어진 비균일 튜브 모델을 이용하여 성도면적 추정 알고리즘에 적용하게 되면, 특정 음성에 최적화된 추정 성도의 면적이 나오게 된다.
- [0028] 추정 면적의 정확성은 음성 정보만을 기준으로 판단되기 때문에 많이 떨어지게 된다. 따라서 여기에 음운 정보를 적용하게 된다. 이때 음성 시각화 장치는 음운론에 입각한 다양한 발성구조에 대한 기본 정보를 가지고 있다. 음성 시각화 장치는 음성 인식의 결과로부터 해당 정보가 어떠한 음운론의 특성에 맞는지를 찾고 여기서 얻어진 정보를 토대로 추정된 면적의 이미지를 보정 처리한다.
- [0029] 본 발명의 일 실시 예에 따른 음성 시각화 방법에서는 입력된 소리 중에서 발화 음성에 대해서만 시각화를 하기 위해 음성구간 검출을 할 수 있다. 검출된 음성은 도 2와 같이 폐에서 발생하여 입을 통해 발성될 때까지 어떤

형태의 터널을 통해 지나왔는지 성도 모델링을 통해 예측된다. 상술한 바와 같이, 예측된 성도 모델은 단순히 소리가 어떤 형태의 터널을 통과했는지를 예측한 것이므로, 좀 더 자세한 조음기관 모델링을 위해 발성된 음성이 코를 통해 발성된 비음인지 혀를 경구개에 붙이고 발성된 파찰음인지 등을 예측하여 이미지를 보정 처리할 필요가 있다.

- [0030] 본 발명의 일 실시 예에 따른 음성의 시각화를 위해 훈련 절차를 미리 거칠 수 있다.
- [0031] 훈련 절차에서는 먼저, 미리 입력된 음성데이터 및 MRI영상을 이용하여 특징벡터 및 다중 튜브 모델을 생성한다. 음성 데이터를 이용하여 특징벡터를 생성하는 과정은 입력된 음성 신호로부터 인식에 유효한 특징 파라미터를 뽑아내는 과정이다. 동일한 단어를 여러 사람이 발음하였을 경우 단어의 의미가 동일하더라도 음성 파형은 동일하지 않으며, 동일한 사람이 동일한 단어를 동일한 시간에 연속으로 발음하였다고 하여도 음성 파형은 동일하지 않다. 음성파형에서는 음성의 의미 정보 이외에도 화자의 음색, 감정 상태 등과 같은 정보도 포함하고 있기 때문이다. 그러므로 음성의 특징 추출이란 음성으로부터 의미 정보를 나타내어주는 특징을 추출하는 것으로 일종의 음성 압축 부분이며 한편으로 인간의 발성기관을 모델링하는 부분이라고 생각할 수 있다.
- [0032] 음성의 시각화를 위해서는 튜브 모델이 사용된다. 음성 신호로부터 성도의 단면적을 추정하기 위하여 성도를 구간마다 면적이 다른 음향 튜브 모델을 연속적으로 나열한 비균일 모델로 가정하여 각 튜브들의 단면적을 추정하여 성도의 모양을 표현할 수 있다. 본 발명에서는 복수의 튜브에 각각의 가중치를 부여하여 병합한 멀티 튜브를 이용한다.
- [0033] 인간의 성도 모양은 복잡한 음향학적 구조와 많은 굴곡으로 이루어져 음성 신호만으로 정확하게 추정하기가 어렵다. 따라서 적용된 면적의 보정처리를 위한 음운보정 처리가 있는데 이는 음운론에 입각한 혀의 위치 등에 대한 정보에 대해서 학습을 시키게 된다. 음운론은 하나의 언어 안에서 의미가 다른 두 소리를 구별해주는 음성적 요소에 대한 것으로, 여러 가지의 서로 다른 개인차 (음색의 차이, 미세한 발음의 차이)에도 불구하고 동일한 의미전달을 가능하게 하는 음성적 요소를 주제로 한다.
- [0034] 훈련 절차에서는 음성과 발화시의 시각 정보를 매칭시킨 데이터 베이스를 구축할 수 있다. 이렇게 훈련 절차를 통해 생성된 각종 데이터 베이스는 실제 발화에 대해 적용된다.
- [0035] 도 3은 본 발명의 일 실시 예에 따른 음성 인식을 이용한 음성 시각화 장치의 블록도이다.
- [0036] 본 발명의 일 실시 예에 따른 음성 인식을 이용한 음성 시각화 장치는 예를 들어, 컴퓨터, 마이크, 발음 구조를 화면으로 표시하기 위한 모니터 등으로 구성될 수 있다. 상기 컴퓨터에는 HMM이 적용될 수 있다. 음성정보에 대해 훈련 과정을 거치면 HMM에 적용할 특징 벡터를 얻게 된다. 이때 부가적으로 음소마다 표준 발음 정보의 이미지를 얻어 이를 튜브화 하여 보관하고, 많은 훈련과정을 거쳐 보다 정확한 모델로 상기 컴퓨터에 저장할 수 있다.
- [0037] 청각장애인 또는 일반인이 마이크를 통하여 발화를 하게 되면 음성을 통하여 정보를 시각화하여 모니터에 표시하면, 화자가 자신의 발음 구조에 존재하는 문제점을 파악하고 많은 연습을 통해 정확한 발음을 하게 유도하게 할 수 있다.
- [0038] 이하에서는 본 발명의 일 실시 예에 따른 음성 인식을 이용한 음성 시각화 장치의 구체적인 동작을 설명한다.
- [0039] 음성 신호가 사람의 발화를 통해 전처리부(310)에 입력되면, 전처리 과정을 통해 음성을 이후의 계산에 적용할 수 있는 형태로 변환시킨다. 전처리부(310)는 음성 신호에서 음성 구간을 검출하고 검출된 음성 구간 내의 음성 신호를 튜브 셀렉터(320)에 전달한다.
- [0040] 튜브 셀렉터(320)는 음성 인식 과정을 통하여 유사 튜브 모델을 선택하는데, 다수의 튜브 모델 후보군이 발생하게 되면 병합 과정을 거치게 되며 새로운 엔트리로 추가 하게 된다. 여러 튜브의 후보군을 얻는 과정과 이를 병합하여 비균일 튜브 모델을 생성하는 과정은 실시간으로 이루어진다. 도 4는 튜브들의 병합 과정을 도시한 것이다. 이와 같이 병합 과정을 거친 새로운 엔트리를 멀티 튜브로 정의한다.
- [0041] 반사 계수 및 성도 면적 추정부(330)는 튜브 셀렉터(320)에서 선택된 튜브 모델을 기준으로 반사계수를 구하기 위한 알고리즘을 적용하고, 이를 통하여 성도 면적을 추정한다. 도 5는 성도 면적을 추정하기 위해 구간마다 면적이 다른 음향 튜브 모델을 연속적으로 나열한 비균일 모델을 성도로 가정하고, 각 튜브들의 단면적을 추정하는 예를 도시한 것이다. 도 6은 성도 면적 추정을 통해 얻어진 성도 그림을 도시한 것이다.
- [0042] 시각화 정보 생성부(340)는 음성 시각화를 구현하는 방법으로 성도의 한쪽 끝을 고정하고 음성신호를 통하여 얻

어진 성도면적을 채워나가는 방법을 적용할 수 있다. 이 경우, 사람이 실제 발화할 때 턱 또는 혀 보다 상대적으로 움직임이 적은 입천장 부위의 이미지에 고정시켜서 시각화를 한다. 즉, 시각화 정보 생성부(340)는 음성 시각화를 위해 고정된 입천장의 이미지를 기준으로 하여 수직 방향으로 추정된 성도의 면적을 채워 나가는 방법을 이용할 수 있다. 인간의 성도는 매우 복잡한 모습을 나타낸다. 따라서 성도의 곡선을 함수로 추정하기는 매우 어렵다. 이에 성도의 모습을 실제로 모델링 하기 위하여서 초음파 이미지, MRI 혹은 X레이 등의 영상 장비 등을 통해 얻어진 이미지 모델을 사용할 수 있다. 도 7은 MRI를 이용한 성도 이미지의 예를 도시한 것이다.

[0043] 허용 범위 보정부(350)는 시각 정보 데이터베이스를 이용하여 반사 계수 및 성도 면적 추정부(330)에서 추정된 성도 면적을 보정한다. 허용 범위 보정부(350)는 음운론 적용 유무를 판별한다. 음운론을 이용하면, 보다 정확한 성도의 면적을 추정할 수 있고 이미지 보간 처리가 가능하다. 보정의 일 예로, 시각화하려는 음운이 미리 학습된 정보에 해당하는 음운에 해당할 경우 전설모음, 후설모음 등에 따른 혀의 위치를 보정하거나 고모음, 중모음, 저모음 등에 따른 높이를 보정할 수 있다. 또한, 허용 범위 보정부(350)는 비음에 따른 정보를 통해 성도 면적을 보정할 수 있다.

[0044] 보다 정확도를 높이기 위해 혀의 유동 범위를 계산하여 좀 더 정확한 혀의 모양을 표시할 수 있다. 허용 범위 보정부(350)는 반사 계수 및 성도 면적 추정부(330)에서 추정된 성도 면적이나 기타 계산된 정보의 값이 일반적인 혀의 움직임 범위 밖에 있을 경우 해당부분에서 사용되는 가중치들을 재분배한 후 재계산을 수행할 수도 있다.

[0045] 시각화부(360)는 허용 범위 보정부(350)에서 보정된 성도 면적을 최종적인 조음 구조의 이미지로 변환하거나, 예비적인 조음 구조의 이미지를 허용 범위 보정부(350)에서 보정한 정보를 반영한 최종 이미지를 생성할 수 있다. 시각화부(360)는 이와 같은 최종 이미지를 모니터 등의 화면에 표시할 수 있다.

[0046] 본 발명의 일 실시 예에 따라 추정된 성도 면적을 보정하는 구체적인 방법으로 음운론에 기초한 데이터베이스를 이용할 수 있다.

[0047] 도 8은 음운론에 따라 표시된 각 모음에 대한 혀의 위치를 도시한 것이다.

[0048] 도 9는 이러한 음운론에 따른 모음 삼각도를 도시한 것이다.

[0049] 각각의 모음은 특유의 포먼트 진동수를 갖는데, 현재 발생된 모음의 포먼트를 측정하여 각 모음들의 포먼트 값에 따라 어떤 모음이 발음되었는지 모음 삼각도를 통해 판단할 수 있다. 본 발명에 이용되는 제1음운론 데이터베이스는 도 9의 내용을 반영한 것으로, 각 모음마다 포먼트 값을 정의한다.

[0050] 도 10은 각 모음에 대한 혀의 위치 및 모양을 도시한 것이다. 본 발명에 이용되는 제2음운론 데이터베이스는 도 9의 내용을 반영한 것으로, 모음 정보마다 혀의 모양을 시각화한 정보를 포함한다.

[0051] 보다 구체적으로, 허용 범위 보정부(350)는 화자로부터 발생된 모음의 포먼트를 측정한 후, 각 모음들의 포먼트 값에 따라 어떤 모음이 발음되었는지를 제1음운론 데이터베이스를 통해 판단하고, 해당하는 혀 위치를 제2음운론 데이터베이스에서 검색할 수 있다. 이렇게 검색된 혀의 위치는 본 발명의 일 실시 예에 따른 음성 인식을 이용한 음성 시각화 장치의 성도 추정 알고리즘을 보정하기 위해 사용될 수도 있다.

[0052] 예를 들어, 한글 '이' 모음의 경우 포먼트 F1에서 약 300Hz, 포먼트 F2에서 약 2,300Hz의 값을 갖는다. 이러한 포먼트 값에 대응하는 모음을 도 9의 모음 삼각도를 이용하여 판단하고, 도 10과 같은 시각 정보에서 해당 모음의 혀모양을 찾아 앞서 추정된 성도 면적을 보정한다.

[0053] 도 11은 실제 음성 시각화의 결과 이미지를 도시한 것이다.

[0054] 도 11은 특히, 마찰음을 발생했을 그 음성을 인식하여 성도 모델링에 반영하여 혀의 위치와 혀의 모양 등을 정확하게 표시한 것이다. 음성인식을 이용하여 발생된 음성이 파열음인지, 마찰음인지, 비음인지 등을 인식하여 혀가 윗니에 붙어서 발생된 것인지 코를 통해 발생된 것인지 두 입술이 붙었다가 떨어지는 소리인지를 파악할 수 있다. 이를 시각화에 반영하면 정확한 성도 모델링을 할 수 있는 것이다.

[0055] 본 발명은 소프트웨어를 통해 실행될 수 있다. 바람직하게는, 본 발명의 일 실시 예에 따른 음성 인식을 이용한 음성 시각화 방법을 컴퓨터에서 실행시키기 위한 프로그램을 컴퓨터로 읽을 수 있는 기록매체에 기록하여 제공할 수 있다. 소프트웨어로 실행될 때, 본 발명의 구성 수단들은 필요한 작업을 실행하는 코드 세그먼트들이다. 프로그램 또는 코드 세그먼트들은 프로세서 판독 가능 매체에 저장되거나 전송 매체 또는 통신망에서 반송파와 결합된 컴퓨터 데이터 신호에 의하여 전송될 수 있다.

[0056] 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 장치의 예로는 ROM, RAM, CD-ROM, DVD±ROM, DVD-RAM, 자기 테이프, 플로피 디스크, 하드 디스크(hard disk), 광데이터 저장장치 등이 있다. 또한, 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 장치에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.

[0057] 본 발명은 도면에 도시된 일 실시 예를 참고로 하여 설명하였으나 이는 예시적인 것에 불과하며 당해 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 실시 예의 변형이 가능하다는 점을 이해할 것이다. 그리고, 이와 같은 변형은 본 발명의 기술적 보호범위 내에 있다고 보아야 한다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위의 기술적 사상에 의해서 정해져야 할 것이다.

산업이용 가능성

[0058] 본 발명은 성도의 면적을 추정을 통해 성문에서부터 입술에 이르는 면적을 시각화하는 기술에 관한 것으로, 시각화 정보를 실제 모델에 적용하고 음운론 보정처리 및 멀티 튜브 모델의 이용 등 조음 구조를 정확하게 묘사하기 위한 장치를 구비한 것이다. 본 발명은 언어장애인들을 위한 조음훈련, 외국어 학습자를 위한 발음 교정, 언어학, 음운학 등의 연구 보조 수단, 애니메이션의 자동 립싱크 기술 등 교육, 학문 연구, 콘텐츠 제작 분야에 적용될 수 있다.

도면의 간단한 설명

[0059] 도 1은 음성 인식을 이용한 음성 시각화의 개략적인 방법을 도시한 것이다.

[0060] 도 2는 음성이 폐에서 발생하여 입을 통해 발성될 때까지 어떤 형태의 터널을 통해 지나왔는지를 나타내는 성도 모델링을 나타낸 것이다.

[0061] 도 3은 본 발명의 일 실시 예에 따른 음성 인식을 이용한 음성 시각화 장치의 블록도이다.

[0062] 도 4는 멀티 튜브 생성을 위한 튜브들의 병합 과정을 도시한 것이다.

[0063] 도 5는 각 튜브들의 단면적을 추정하는 예를 도시한 것이다.

[0064] 도 6은 성도 면적 추정을 통해 얻어진 성도 그림을 도시한 것이다.

[0065] 도 7은 MRI를 이용한 성도 이미지의 예를 도시한 것이다.

[0066] 도 8은 음운론에 따라 표시된 각 모음에 대한 혀의 위치를 도시한 것이다.

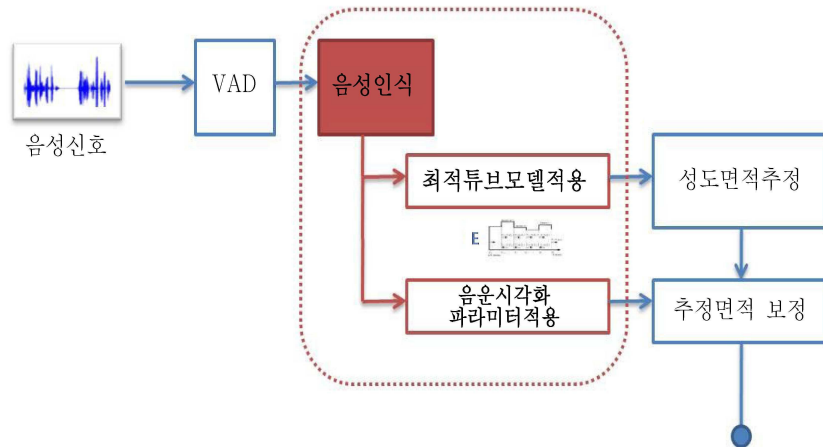
[0067] 도 9는 도 8의 음운론에 따른 모음 삼각도를 도시한 것이다.

[0068] 도 10은 각 모음에 대한 혀의 위치 및 모양을 도시한 것이다.

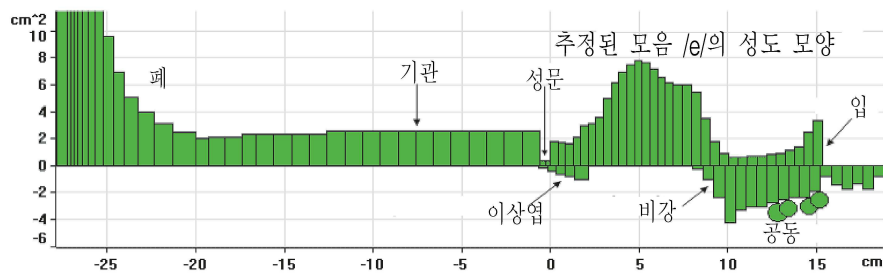
[0069] 도 11은 실제 음성 시각화의 결과 이미지를 도시한 것이다.

도면

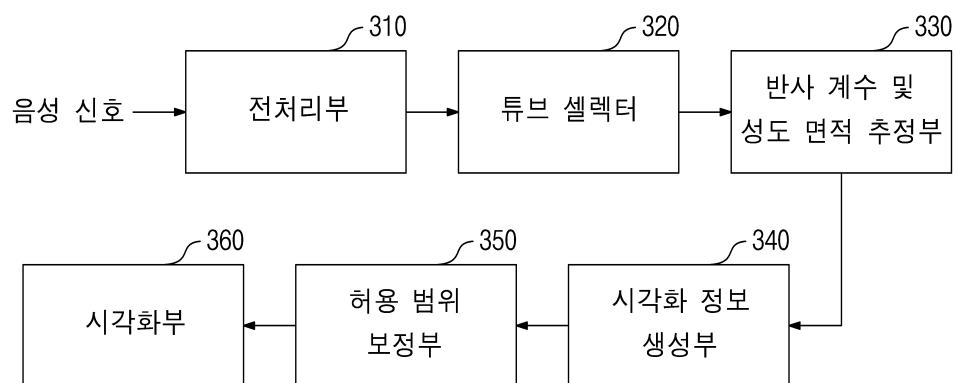
도면1



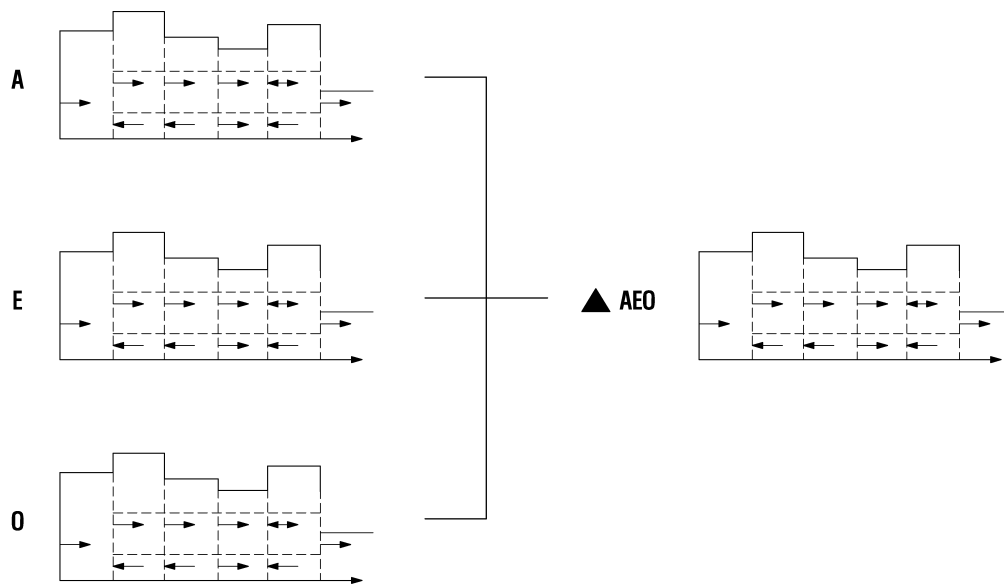
도면2



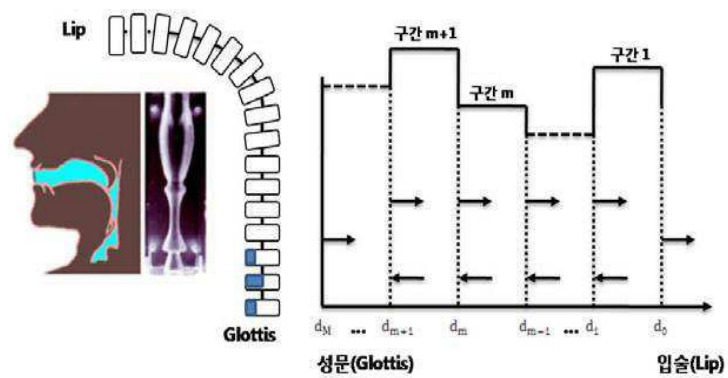
도면3



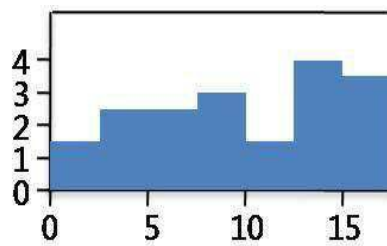
도면4



도면5



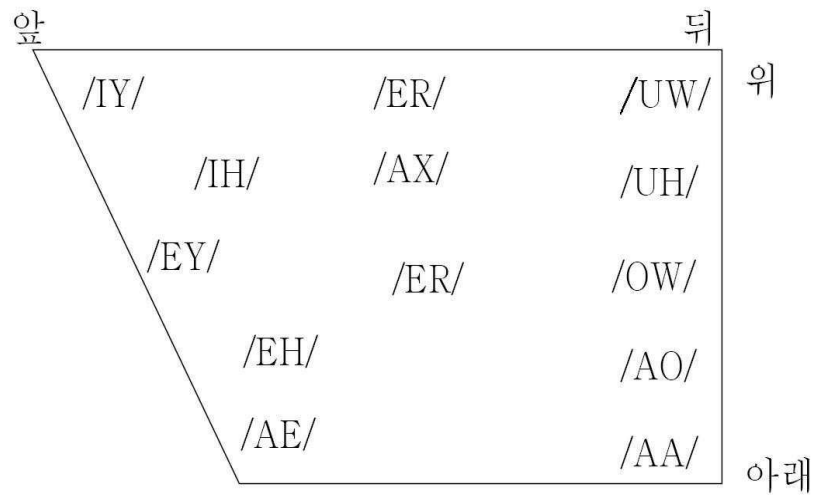
도면6



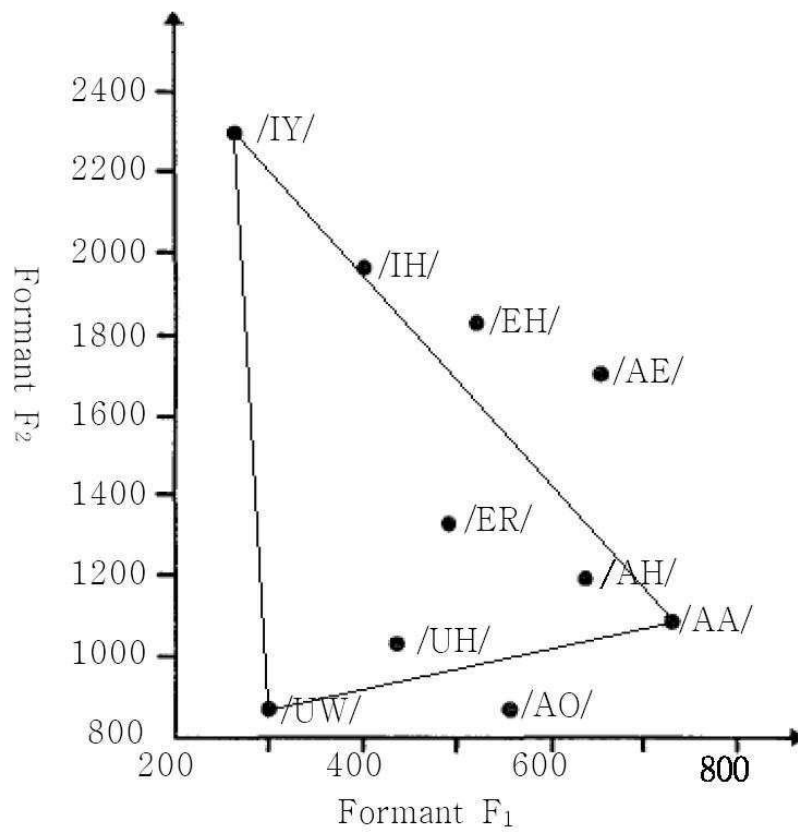
도면7



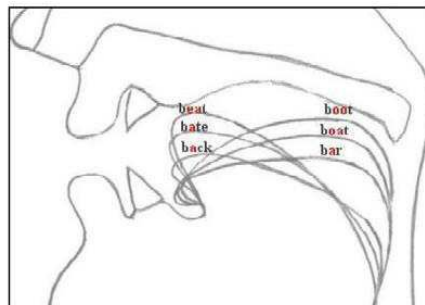
도면8



도면9



도면10



도면11

