

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0031066

(43) 공개일자 2025년03월06일

(51) 국제특허분류(Int. Cl.)

G10L 15/06 (2006.01) G01H 15/00 (2006.01)

G01P 15/18 (2006.01) G06N 3/0464 (2023.01)

G10L 15/16 (2006.01) G10L 21/10 (2013.01)

(52) CPC특허분류

G10L 15/063 (2013.01)

G01H 15/00 (2013.01)

(21) 출원번호 10-2023-0111769

(22) 출원일자 2023년08월25일

심사청구일자 2023년08월25일

(71) 출원인

인하대학교 산학협력단

인천광역시 미추홀구 인하로 100(용현동, 인하대학교)

(72) 발명자

김기우

인천광역시 서구 봉오대로283번길 6, 1동 1413호
(가정동, 가정뉴서울아파트)

(74) 대리인

특허법인아이더스

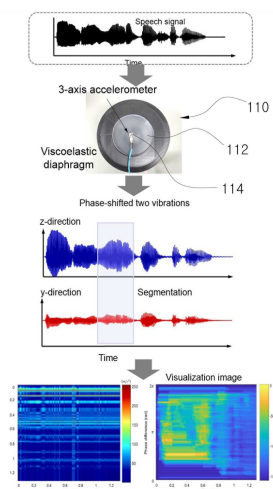
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 음성 인식을 위한 학습 데이터 생성 시스템 및 방법

(57) 요약

본 발명은 음성 인식을 위한 학습 데이터 생성 시스템 및 방법에 관한 것으로, 보다 상세하게는 사람의 고막과 유사한 점탄성 특성을 갖는 음향 센서를 이용하여 수신한 음향 신호를 이용하여 음성 인식을 위한 CNN 학습에 입력 데이터로 사용될 수 있는 시각화된 이미지 형태의 학습 데이터를 생성할 수 있도록 하는 음성 인식을 위한 학습 데이터 생성 시스템 및 방법에 관한 것이다.

대표도 - 도2



(52) CPC특허분류

G01P 15/18 (2013.01)

G06N 3/0464 (2023.01)

G10L 15/16 (2013.01)

G10L 21/10 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711191054
과제번호	2021R1A2C1003411
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)
연구과제명	압전 포토닉 복합재 기반 음향 가시화를 이용한 음성 인식
기 여 율	1/1
과제수행기관명	인하대학교
연구기간	2023.03.01 ~ 2024.02.29

명세서

청구범위

청구항 1

음향신호 발생장치의 전방에 일정 거리 이격되도록 설치되는 점탄성 재질의 음향 센서와,
상기 음향 센서에 의해 측정된 음향 신호를 수신하여 위상 변이된 2축 진동 신호로 변환시키는 신호변환장치와,
상기 신호변환장치에 의해 변환된 음향 신호를 이미지로 시각화하는 시각화장치를 포함하는 음성 인식을 위한 학습 데이터 생성 시스템.

청구항 2

제 1항에 있어서,
상기 음향 센서는 얇은 점탄성 재질로 이루어져 음파에 의해 진동하는 점탄성 다이어프램과,
상기 점탄성 다이어프램에 구비되어 음파에 의해 진동하는 점탄성 다이어프램의 각 방향 가속도를 측정하는 3축 가속도 센서를 포함하는 음성 인식을 위한 학습 데이터 생성 시스템.

청구항 3

제 2항에 있어서,
상기 신호변환장치는 3축 가속도 센서에 의해 측정된 점탄성 다이어프램의 가속도 신호를 수신하는 신호수신부와, 상기 신호수신부를 통해 수신된 가속도 신호를 이진화하여 위상 변이된 횡방향 진동과 종방향 진동으로 변환시키는 신호변환부를 포함하는 음성 인식을 위한 학습 데이터 생성 시스템.

청구항 4

제 1항에 있어서,
상기 시각화장치는,
신호변환장치에 의해 변환된 신호를 재구성하여 2차원 공간에서의 위상 궤적 포트레이트(phase trajectory portrait)를 생성하는 포트레이트 생성부와,
상기 위상 궤적 포트레이트에 포함된 위상 궤적을 CRP(Cross Recurrence Plot)를 이용하여 시각화하는 이미지 생성부를 포함하는 음성 인식을 위한 학습 데이터 생성 시스템.

청구항 5

음파에 의해 진동하는 점탄성 다이어프램의 동적 거동 특성을 CRP(Cross Recurrence Plot)를 통해 시각화하여 CNN 학습에 입력 데이터로 사용될 수 있는 이미지 형태의 학습 데이터를 생성하는 것을 특징으로 하는 음성 인식을 위한 학습 데이터 생성 방법.

청구항 6

제 5항에 있어서,

상기 점탄성 다이어프램의 동적 거동 특성은,

점탄성 다이어프램을 포함하는 음향 센서에 의해 음향신호 발생장치로부터 발생하는 음향 신호를 수신하는 음향 신호 수신단계와,

3축 가속도 센서를 이용하여 수신된 음향 신호의 음파에 의해 진동하는 점탄성 다이어프램의 가속도를 측정하는 가속도 측정단계 및

상기 가속도 측정단계에서 측정된 가속도를 이진화하여 위상 변이된 횡방향 진동과 종방향 진동으로 변환시키는 신호변환단계를 통해 획득되는 것을 특징으로 하는 음성 인식을 위한 학습 데이터 생성 방법.

청구항 7

제 6항에 있어서,

상기 CRP(Cross Recurrence Plot)를 통한 시각화 방법은,

상기 신호변환단계에서 변환된 위상 변이된 횡방향 진동과 종방향 진동을 재구성하여 2차원 공간에서의 위상 궤적 포트레이트(phase trajectory portrait)를 생성하는 포트레이트 생성단계와,

상기 위상 궤적 포트레이트에 포함된 위상 궤적을 거리 행렬로 변환시키는 행렬 변환단계 및

상기 거리 행렬의 강도에 따라 픽셀 별로 색상을 부여하여 최종적인 CRP 이미지를 생성하는 색상 표현단계를 포함하는 음성 인식을 위한 학습 데이터 생성 방법.

발명의 설명

기술 분야

- [0001] 본 발명은 음성 인식을 위한 학습 데이터 생성 시스템 및 방법에 관한 것으로, 보다 상세하게는 사람의 고막과 유사한 점탄성 특성을 갖는 음향 센서를 이용하여 수신한 음향 신호를 이용하여 음성 인식을 위한 CNN 학습에 입력 데이터로 사용될 수 있는 시각화된 이미지 형태의 학습 데이터를 생성할 수 있도록 하는 음성 인식을 위한 학습 데이터 생성 시스템 및 방법에 관한 것이다.

배경 기술

- [0003] 음성 인식(STT; Speech To Text)은 프로그램을 이용하여 사람의 음성을 인식하고 문자 형식으로 변환하는 프로세스를 의미한다.
- [0004] 이러한 음성 인식 기술을 사용할 경우, 기존의 터치 기반 휴먼 머신 인터페이스(HMI; Human Machine Interface) 장치를 대체하거나 보완할 수 있을 뿐만 아니라, 양방향 통신이 가능하다는 점에서 스마트 홈 오토메이션과 관련된 인공지능 서비스 및 사물인터넷의 핵심 기술로 주목받고 있다.
- [0005] 일반적으로 음성 인식 시스템에는 크게 음향 센서와 음성 인식 소프트웨어라는 두 가지 구성요소가 포함될 수 있는데, 이 중 음향 센서는 사람이 말하는 음압에 반응하여 음향 에너지를 아날로그 전기 신호로 변환시키는 역할을 하는 센서를 의미하고, 음성 인식 소프트웨어는 변환된 전기 신호를 이용하여 음성을 인식함으로써 마이크로 컴퓨터가 구두 명령, 즉 음성에 의한 명령을 실행할 수 있도록 짜여진 프로그램을 의미한다.
- [0006] 상기 음향 센서로는 두 개의 전도성 다이어프램 사이의 커패시턴스 차이를 이용하는 콘덴서 타입의 마이크로폰(microphone)이 주로 사용되는데, 그 응답(response)은 주로 시계열 데이터로 구성되고, 고속푸리에변환(FFT; Fast Fourier Transform)을 이용하여 주파수 도메인으로 변환된다.
- [0007] 최근에는 음향 센서에 의해 수집된 음향 신호를 전기신호로 바꾼 후 STFT(Short Time Fourier Transform, 이하 'STFT'라 한다)를 이용하여 주파수 도메인으로 변환시키는 방식이 주로 사용되는데, 상기 STFT는 시간에 대한 데이터를 스펙트로그램(spectrogram)이라고 하는 더 짧은 구간으로 나누어 푸리에 변환을 하는 방법으로 시간에 따라 주파수 성분이 변화하는 신호의 정보를 보다 효율적으로 분석할 수 있다.

- [0008] 이때, 상기 스펙트로그램은 음향 신호에 STFT를 적용한 결과에서의 시간과 주파수를 각각 가로축과 세로축으로 설정하고, 각 주파수의 세기를 dB 단위의 색변화로 표현한 그래프를 의미하는 것으로, 이러한 STFT 및 스펙트로그램을 이용할 경우, 음향 센서에 의해 수집된 음향 신호를 가로축에는 시간정보가 표현되고, 세로축에는 주파수의 크기가 dB 단위로 표현되는 2차원 이미지로 표현할 수 있다.
- [0009] 또한, 상기와 같이 생성된 2차원 이미지는 이미지 처리에 사용되는 CNN(Convolution Neural Network)을 기반으로 하는 딥러닝 프로그램, 즉 이미지를 이용하여 음성 특징을 추출하고 이를 바탕으로 단어를 분류할 수 있도록 하는 딥러닝 프로그램의 학습 데이터(input data)로 사용될 수 있다.
- [0010] 하지만, 상기 STFT의 경우 설정된 보다 짧은 시간 간격(window)에 대하여 각각 푸리에 변환을 하여야만 스펙트로그램을 얻을 수 있으므로, 2차원 이미지를 생성하기 위한 연산량이 상당하므로 학습 데이터의 생성에 소요되는 시간이 증가할 뿐만 아니라, 주파수 분해능(resolution)의 악화로 인해 정확도가 떨어질 수 있는 단점이 있다.
- [0011] 따라서, CNN 기반의 음성 인식을 위한 학습 데이터 생성에 사용될 수 있는 보다 간단하고 효율적인 시스템 및 방법의 개발이 요구되고 있는 실정이다.

선행기술문헌

특허문헌

- [0013] (특허문헌 0001) 1. 대한민국 공개특허공보 제10-2020-0033058호(2020. 03. 27. 공개)
- (특허문헌 0002) 2. 대한민국 등록특허공보 제10-2151682호(2020. 08. 28. 등록)

발명의 내용

해결하려는 과제

- [0014] 본 발명은 상기와 같은 종래기술의 문제점들을 해결하기 위하여 안출된 것으로, 본 발명의 목적은 사람의 고막과 유사한 점탄성 특성을 갖는 음향센서를 이용하여 수신한 음향 신호를 CRP를 통해 시각화함으로써 음성 인식을 위한 CNN 학습에 입력 데이터로 사용될 수 있는 학습 데이터를 생성할 수 있도록 하는 음성 인식을 위한 학습 데이터 생성 시스템 및 방법을 제공함에 있다.

과제의 해결 수단

- [0016] 상기와 같은 목적들을 달성하기 위한 본 발명은,
- [0017] 음향신호 발생장치의 전방에 일정 거리 이격되도록 설치되는 점탄성 재질의 음향 센서와, 상기 음향 센서에 의해 측정된 음향 신호를 수신하여 위상 변이된 2축 진동 신호로 변환시키는 신호변환장치와, 상기 신호변환장치에 의해 변환된 음향 신호를 이미지로 시각화하는 시각화장치를 포함할 수 있다.
- [0018] 이때, 상기 음향 센서는 얇은 점탄성 재질로 이루어져 음파에 의해 진동하는 점탄성 다이어프램과, 상기 점탄성 다이어프램에 구비되어 음파에 의해 진동하는 점탄성 다이어프램의 각 방향 가속도를 측정하는 3축 가속도 센서를 포함할 수 있다.
- [0019] 또한, 상기 신호변환장치는 3축 가속도 센서에 의해 측정된 점탄성 다이어프램의 가속도 신호를 수신하는 신호수신부와, 상기 신호수신부를 통해 수신된 가속도 신호를 이진화하여 위상 변이된 횡방향 진동과 종방향 진동으로 변환시키는 신호변환부를 포함할 수 있다.
- [0020] 그리고, 상기 시각화장치는, 신호변환장치에 의해 변환된 신호를 재구성하여 2차원 공간에서의 위상 궤적 포트레이트(phase trajectory portrait)를 생성하는 포트레이트 생성부와, 상기 위상 궤적 포트레이트에 포함된 위상 궤적을 CRP(Cross Recurrence Plot)를 이용하여 시각화하는 이미지 생성부를 포함할 수 있다.

- [0021] 한편, 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 방법은,
- [0022] 음파에 의해 진동하는 점탄성 다이어프램의 동적 거동 특성을 CRP(Cross Recurrence Plot)를 통해 시각화하여 CNN 학습에 입력 데이터로 사용될 수 있는 이미지 형태의 학습 데이터를 생성하는 것을 특징으로 한다.
- [0023] 이때, 상기 점탄성 다이어프램의 동적 거동 특성은, 점탄성 다이어프램을 포함하는 음향 센서에 의해 음향신호 발생장치로부터 발생하는 음향 신호를 수신하는 음향신호 수신단계와, 3축 가속도 센서를 이용하여 수신된 음향 신호의 음파에 의해 진동하는 점탄성 다이어프램의 가속도를 측정하는 가속도 측정단계 및 상기 가속도 측정단계에서 측정된 가속도를 이진화하여 위상 변이된 횡방향 진동과 종방향 진동으로 변환시키는 신호변환단계를 통해 획득되는 것을 특징으로 한다.
- [0024] 또한, 상기 CRP(Cross Recurrence Plot)를 통한 시각화 방법은, 상기 신호변환단계에서 변환된 위상 변이된 횡방향 진동과 종방향 진동을 재구성하여 2차원 공간에서의 위상 궤적 포트레이트(phase trajectory portrait)를 생성하는 포트레이트 생성단계와, 상기 위상 궤적 포트레이트에 포함된 위상 궤적을 거리 행렬로 변환시키는 행렬 변환단계 및 상기 거리 행렬의 강도에 따라 픽셀 별로 색상을 부여하여 최종적인 CRP 이미지를 생성하는 색상 표현단계를 포함할 수 있다.

발명의 효과

- [0026] 본 발명에 따르면, 사람의 고막과 유사한 점탄성 특성을 갖는 다이어프램의 음파에 의한 동적 거동 특성 및 이를 이용한 CRP 이미지 생성을 통해 STFT(Short Time Fourier Transform)를 사용하지 않고서도 음성 인식을 위한 CNN 학습을 위한 입력 데이터를 생성할 수 있는 효과를 갖는다.
- [0027] 또한, 본 발명에 따르면 종래의 STFT 방식에 비해 CNN 학습 데이터 생성을 위해 필요한 연산에 대한 부담을 현저하게 줄일 수 있고, 그에 따라 학습 데이터의 생성에 소요되는 시간을 대폭 단축시킬 수 있는 효과를 추가적으로 갖는다.

도면의 간단한 설명

- [0029] 도 1은 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 시스템을 개념적으로 나타낸 도면.
- 도 2는 도 1에 나타난 본 발명을 이용하여 음성 인식을 위한 학습 데이터를 생성하는 과정을 개략적으로 나타낸 도면.
- 도 3의 (a),(b)는 종래의 STFT를 이용하여 생성된 스펙트로그램 이미지와, 본 발명에 의해 생성된 CRP 이미지를 비교하여 나타낸 도면.
- 도 4는 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 방법을 순차적으로 나타낸 흐름도.
- 도 5의 (a)~(e)는 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 방법을 개념적으로 나타낸 도면.
- 도 6은 도 4에 나타난 본 발명 중 신호변환단계에서 이진화되는 두 개의 진동신호를 예시적으로 나타낸 도면.

발명을 실시하기 위한 구체적인 내용

- [0030] 이하, 첨부된 도면들을 참고로 하여 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 시스템 및 방법의 바람직한 실시예들을 상세히 설명하기로 한다.
- [0032] 이하의 내용은 단지 본 발명의 원리를 예시하는 것이므로 본 명세서에 열거된 모든 조건부 용어 및 실시예들은 원칙적으로, 본 발명의 개념이 이해되도록 하기 위한 목적으로만 명백히 의도되고, 이와 같이 특별히 열거된 실시예들 및 상태들에 제한적이지 않는 것으로 이해되어야 한다.
- [0033] 또한, 본 명세서의 블록도는 본 발명의 원리를 구체화하는 예시적인 회로의 개념적인 관점을 나타내는 것으로 이해되어야 하며, 이와 유사하게, 모든 흐름도 등은 컴퓨터가 판독 가능한 매체에 실질적으로 나타낼 수 있고 컴퓨터 또는 프로세서가 명백히 도시되었는지 여부를 불문하고 컴퓨터 또는 프로세서에 의해 수행되는 다양한

프로세스를 나타내는 것으로 이해되어야 한다.

- [0034] 프로세서 또는 이와 유사한 개념으로 표시된 기능 블록을 포함하는 도면에 도시된 다양한 구성요소의 기능은 전용 하드웨어뿐만 아니라 적절한 소프트웨어와 관련하여 소프트웨어를 실행할 능력을 가진 하드웨어의 사용으로 제공될 수 있으며, 상기 기능은 단일 전용 프로세서, 단일 공유 프로세서 또는 복수의 개별적 프로세서에 의해 제공될 수 있고, 이들 중 일부는 공유될 수 있다.
- [0035] 또한 프로세서, 제어 또는 이와 유사한 개념으로 제시되는 용어의 명확한 사용은 소프트웨어를 실행할 능력을 가진 하드웨어를 배타적으로 인용하여 해석되어서는 아니되고, 제한 없이 디지털 신호 프로세서(DSP) 하드웨어, 소프트웨어를 저장하기 위한 롬(ROM), 램(RAM) 및 비 휘발성 메모리를 암시적으로 포함하는 것으로 이해되어야 한다. 주지관용의 다른 하드웨어도 포함될 수 있다.
- [0036] 본 명세서의 청구범위에서, 상세한 설명에 기재된 기능을 수행하기 위한 수단으로 표현된 구성요소는 예를 들어 상기 기능을 수행하는 모든 형식의 소프트웨어 또는 상기 기능을 수행하는 모든 방법을 포함하는 것으로 의도되었으며, 상기 기능을 수행하도록 상기 소프트웨어를 실행하기 위한 적절한 회로와 결합될 수 있다. 이러한 청구범위에 의해 정의되는 본 발명은 다양하게 열거된 수단에 의해 제공되는 기능들이 결합되고 청구항이 요구하는 방식과 결합되기 때문에 상기 기능을 제공할 수 있는 어떠한 수단도 본 명세서로부터 파악되는 것과 균등한 것으로 이해되어야 한다.
- [0038] 도 1은 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 시스템을 개념적으로 나타낸 도면이고, 도 2는 도 1에 나타낸 본 발명을 이용하여 음성 인식을 위한 학습 데이터를 생성하는 과정을 개략적으로 나타낸 도면이며, 도 3의 (a),(b)는 종래의 STFT를 이용하여 생성된 스펙트로그램 이미지와, 본 발명에 의해 생성된 CRP 이미지를 비교하여 나타낸 도면이고, 도 4는 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 방법을 순차적으로 나타낸 흐름도이며, 도 5의 (a)~(e)는 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 방법을 개념적으로 나타낸 도면이고, 도 6은 도 4에 나타낸 본 발명 중 신호변환단계에서 이진화되는 두 개의 진동신호를 예시적으로 나타낸 도면이다.
- [0040] 본 발명은 사람의 고막과 유사한 점탄성 특성을 갖는 음향 센서를 이용하여 수신한 음향 신호를 이용하여 음성 인식을 위한 CNN 학습에 입력 데이터로 사용될 수 있는 시각화된 이미지 형태의 학습 데이터를 생성할 수 있도록 하는 음성 인식을 위한 학습 데이터 생성 시스템 및 방법에 관한 것으로, 먼저 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 시스템(100)은 도 1에 나타낸 바와 같이, 크게 음향 센서(110), 신호변환장치(120) 및 시각화장치(130)를 포함할 수 있다.
- [0041] 즉, 본 발명은 종래의 음성 인식을 위한 CNN 학습에 입력 데이터, 즉 학습 데이터로 사용되는 STFT를 이용하여 생성된 스펙트로그램 이미지를 대체할 수 있는 새로운 방식의 학습 데이터를 생성하기 위한 시스템(100)에 관한 것으로, 사람의 고막과 유사한 점탄성 재질의 음향 센서(110)를 통해 수집된 음향 신호의 가속도 성분을 이진화하고, 이를 후술할 CRP(Cross Recurrence Plot, 이하 'CRP'라 한다)를 이용하여 시각화함으로써 음성 인식을 위한 CNN 학습의 입력 데이터로 사용될 수 있는 이미지 형태의 학습 데이터 생성에 소요되는 시간을 종래의 STFT에 비해 대폭 단축시킬 수 있도록 한 것에 기술적 특징이 있는 것이다.
- [0042] 보다 상세히 설명하면, 상기 음향 센서(110)는 스피커 등의 음향신호 발생장치(미도시)의 전방으로 일정 거리 이격되도록 설치되어 음향신호 발생장치로부터 출력되는 음향 신호를 수신하는 역할을 하는 것으로, 그 구성은 사람의 고막과 유사한 점탄성 재질로 이루어지는 점탄성 다이어프램(Viscoelastic diaphragm)(112)과, 상기 점탄성 다이어프램(112)의 중심부에 설치되는 3축 가속도 센서(114)를 포함할 수 있다.
- [0043] 먼저, 상기 점탄성 다이어프램(112)은 음향신호 발생장치로부터 출력되는 음향 신호의 음파에 의해 구동되어 진동하는 역할을 하는 것으로, 약 0.1mm의 얇은 두께를 갖는 실리콘 고무 필름 등의 점탄성 재질로 이루어질 수 있으며, 상기 점탄성 다이어프램(112)의 가장자리 부분은 음파에 의한 진동이 원활하게 이루어질 수 있도록 고정될 수 있다.
- [0044] 즉, 상기 음향신호 발생장치로부터 발생된 음향신호가 가장자리 부분이 고정된 팽팽한 상태의 점탄성 다이어프램(112)에 도달하면, 도 6에 나타낸 바와 같이, 음파의 압력에 의해 점탄성 다이어프램(112)이 수직 방향, 즉 z 방향으로 횡방향 진동을 하게 된다.

- [0045] 상기 점탄성 다이어프램(112)이 횡방향 진동을 하면, 평형 위치, 즉 중심에서 벗어난 특정 지점이 불균형 상태가 되면서 횡방향 진동과 음파의 압력이 결합된 힘(combined force)으로 인해 평형 위치로부터 반경 방향, 즉 가장자리 방향으로의 위치 이동이 발생된다.
- [0046] 그에 따라, 점탄성 다이어프램(112)의 장력이 복원력으로 작용하게 되어 점탄성 다이어프램(112)에 평행한 반경 방향의 진동, 즉 종방향 진동이 발생되고, 결과적으로 음향신호 발생장치로부터 발생된 음향신호는 점탄성 다이어프램(112)에 z축 방향으로의 횡방향 진동(이하, '횡방향 진동'이라 한다)과, y축 방향으로의 종방향 진동(이하, '종방향 진동'이라 한다)을 발생시킬 수 있다.
- [0047] 또한, 점탄성 재료의 동적 응답은 일반적으로 저장 및 손실 탄성률(storage and loss modulus)을 기준으로 특성화되므로, 레오미터(rheometer)를 이용하여 주파수 진폭 및 진동 스위프 테스트(sweep test)를 통해 상기 저장 및 손실 탄성률을 측정할 수 있고, 이러한 점탄성 재료의 저장 및 손실 탄성률로부터 상기 횡방향 진동과 종방향 진동 사이의 예상되는 위상변이(phase shift), 즉 위상각 차이가 도출될 수 있다.
- [0048] 다음, 상기 3축 가속도 센서(114)는 음향신호의 음파에 의해 진동하는 점탄성 다이어프램(112)의 각 방향 가속도를 측정하는 역할을 하는 것으로, 도 2에 나타난 바와 같이, 점탄성 다이어프램(112)의 중심부에 설치될 수 있다.
- [0049] 즉, 상기 점탄성 다이어프램(112)의 중심부에 질량이 약 1g인 소형 3축 가속도 센서(114)를 설치함으로써 음파에 의해 진동하는 점탄성 다이어프램(112)의 횡방향 진동 및 종방향 진동 특성, 즉 각 방향으로의 가속도를 측정할 수 있다.
- [0050] 다음, 상기 신호변환장치(120)는 음향 센서(110)에 의해 측정된 음향 신호, 즉 3축 가속도 센서(114)에 의해 측정된 각 방향으로의 가속도 신호를 수신하여 위상변이된 2축 진동 신호, 즉 횡방향 진동 신호와 종방향 진동 신호로 변환시키는 역할을 하는 것으로, 그 구성은 신호수신부(122)와 신호변환부(124)를 포함할 수 있다.
- [0051] 상기 신호수신부(122)는 3축 가속도 센서(114)에 의해 측정된 점탄성 다이어프램(112)의 각 방향으로의 가속도 신호를 수신하는 역할을 할 수 있고, 상기 신호변환부(124)는 신호수신부(122)를 통해 수신된 각 방향으로의 가속도 신호를 이진화(binimize)하여 위상 변이된 횡방향 진동과 종방향 진동으로 변환시키는 역할을 할 수 있다.
- [0052] 즉, 전술한 바와 같이 음파에 의한 점탄성 다이어프램(112)의 진동은 횡방향 진동과 종방향 진동으로 나타나게 되고, 점탄성 특성으로 인해 두 진동 사이에 위상 변이가 발생하게 되는데, 상기 신호변환부(124)에서는 3축 가속도 센서(114)로부터 수신한 가속도 신호, 즉 주파수 응답 함수(frequency response function)의 이중 적분을 통해 두 진동 사이의 위상변이를 연산할 수 있고, 이를 통해 음향 신호를 두 개의 위상변이된 시계열 형태의 진동신호, 즉 횡방향 진동 신호와 종방향 진동 신호로 변환할 수 있다.
- [0053] 다음, 상기 시각화장치(130)는 신호변환부(124)에 의해 변환된 음향 신호를 CRP를 이용하여 CNN 학습에서의 입력 데이터로 사용될 수 있는 이미지 형태의 학습 데이터로 시각화하는 역할을 하는 것으로, 그 구성은 포트레이트 생성부(132)와 이미지 생성부(134)를 포함할 수 있다.
- [0054] 보다 상세히 설명하면, 상기 포트레이트 생성부(132)는 신호변환장치(120)에 의해 변환된 신호, 즉 횡방향 진동 신호와 종방향 진동 신호를 짧은 구간(window)으로 분할한 후 2차원 위상 공간에서의 위상 궤적 포트레이트(phase trajectory portrait)로 재구성하는 역할을 하는 것으로, 상기 포트레이트 생성부(132)에 의해 생성된 위상 궤적 포트레이트는 후술할 CRP 생성을 위한 거리 벡터(distance matrix)로의 변환을 위한 연산에 사용될 수 있다.
- [0055] 다음, 상기 이미지 생성부(134)는 위상 궤적 포트레이트에 포함된 횡방향 진동 신호와 종방향 진동 신호의 위상 궤적을 CRP를 이용하여 시각화함으로써 이미지 형태의 학습 데이터를 생성하는 역할을 하는 것으로, 그 구성은 행렬변환부(134a)와 색상표현부(134b)를 포함할 수 있다.
- [0056] 이때, 상기 CRP(Cross Recurrent Plot)는, 시계열 데이터, 즉 궤적 데이터의 행동을 시각화하기 위해 사용되는 공지의 RP(Recurrent Plot)를 이변량(bivariate)으로 확장할 수 있도록 진화시킨 시각화 방법으로, CRP를 이용하면 두 개의 다른 동적 시스템 사이의 관계를 시각화할 수 있다.
- [0057] 먼저, 상기 행렬변환부(134a)는 위상 궤적 포트레이트에 포함된 위상 궤적을 연산에 의해 거리 행렬(distance matrix)로 변환하는 역할을 하는 것이고, 상기 색상표현부(134b)는 연산된 거리 행렬의 강도(intensity), 즉 궤적 사이의 거리에 따라 픽셀들에 기설정된 색상을 부여하여 최종적인 CRP 이미지를 생성하는 역할을 하는 것이다.

다.

[0058] 상기 시각화장치(130)를 이용한 최종적인 CRP 이미지 생성과 관련된 보다 구체적인 내용들은 후술할 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 방법에서 보다 상세히 설명하기로 한다.

[0059] 한편, 도 3의 (a), (b)는 본 발명에 따른 학습 데이터 생성 시스템(100)에 의해 생성된 학습 데이터인 CRP 이미지(도 3의 (b)에 해당)를 종래의 STFT를 이용하여 생성된 스펙트로그램 이미지(도 3의 (a)에 해당)와 비교하여 나타낸 것으로, 본 발명에 의하면 시계열 데이터를 단순한 연산에 의해 직접 이미지화할 수 있으므로 종래의 STFT를 이용한 스펙트로그램 이미지 생성에 비해 연산에 소요되는 시간을 대폭 단축시킬 수 있다.

[0060] 즉, 아래의 (표 1)은 1.2초 동안 출력되는 'Inha University'라는 음성에 대한 컬러 이미지 생성에 소요되는 CPU(11th Gen Intel Core i5-11400H) 실행시간을 픽셀 크기, 즉 해상도(resolution)에 따라 비교하여 나타낸 것으로, 본 발명에 의해 생성된 CRP 이미지의 경우 해상도 134 × 134에서 CPU 실행시간이 0.005초 임에 비해, 종래의 STFT의 경우 이미지 생성에 소요되는 CPU 실행시간이 0.150초로 본 발명에 비해 약 30배 정도의 시간이 소요되는 것을 확인할 수 있다.

표 1

Pixel size(해상도)	STFT(sec)	CRP(sec)
134 × 134	0.150	0.005
333 × 333	0.179	0.025
666 × 666	0.230	0.102
1138 × 1138	0.279	0.307

[0061]

[0062] 단, 본 발명에 따른 CRP 이미지 생성에 소요되는 시간은 해상도가 높아질 수록 크게 증가하여 해상도가 1138 × 1138인 경우 오히려 종래의 STFT에 비해 이미지 생성에 더 오랜 시간이 소요되는 것으로 확인되었는데, STFT의 경우 각 시점에서의 윈도우(window) 크기에 따라 연산시간이 달라지는 점을 고려하면 그 차이가 미비하다고 할 수 있다.

[0063] 이미지의 해상도가 크지 않은 경우에는 CRP 이미지 생성에 소요되는 시간이 훨씬 짧으므로, 본 발명에 의하면 종래의 STFT에 비해 학습 데이터 생성을 위해 필요한 연산에 대한 부담을 현저하게 줄일 수 있는 효과를 보일 수 있다.

[0065] 다음, 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 방법은 전술한 학습 데이터 생성 시스템(100)을 이용하여 음성 인식을 위한 CNN 학습에 사용될 수 있는 학습 데이터, 즉 CRP 이미지를 생성하는 방법에 관한 것으로, 전술한 바와 같이 음파에 의해 진동하는 점탄성 다이어프램(112)의 동적 거동 특성을 CRP(Cross Recurrence Plot)를 통해 시각화하여 CNN 학습에 입력 데이터로 사용될 수 있는 이미지 형태의 학습 데이터, 즉 CRP 이미지를 형성시킬 수 있도록 한 것에 그 특징이 있는 것이다.

[0066] 즉, 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 방법은 도 4에 나타난 바와 같이, 크게 음향신호 수신 단계(S10), 가속도 측정단계(S20), 신호변환단계(S30) 및 이미지 생성단계(S40)를 포함할 수 있는데, 상기 음향신호 수신단계(S10), 가속도 측정단계(S20) 및 신호변환단계(S30)를 통해 음파에 의해 진동하는 점탄성 다이어프램(112)의 동적 거동 특성을 획득하고, 이미지 생성단계(S40)에서는 획득된 점탄성 다이어프램(112)의 동적 거동 특성을 이용하여 CRP 이미지를 형성시킬 수 있다.

[0067] 먼저, 상기 음향신호 수신단계(S10)는 스피커 등의 음향신호 발생장치의 전방에 점탄성 다이어프램(112)과 3축 가속도 센서(114)를 포함하는 음향 센서(110)를 설치하여 음향신호 발생장치로부터 발생하는 음향 신호를 수신하는 단계이고, 상기 가속도 측정단계(S20)는 점탄성 다이어프램(112)의 중심부에 설치되는 3축 가속도 센서(114)를 이용하여 음향 신호의 음파에 의해 진동되는 점탄성 다이어프램(112)의 가속도를 측정하는 단계이다.

(도 5의 (a) 참조)

[0068] 또한, 상기 신호변환단계(S30)는 도 5의 (b)에 나타난 바와 같이, 가속도 측정단계(S20)에서 측정된 가속도를 이진화하여 위상변이된 2축 진동 신호, 즉 횡방향 진동 신호와 종방향 진동 신호로 변환시키는 과정으로, 그 구체적인 내용은 전술한 학습 데이터 생성 시스템(100)의 신호변환장치(120)에서 설명한 바와 같으므로 이에 대한 상세한 설명은 생략하기로 한다.

[0069] 다음, 상기 이미지 생성단계(S40)는 신호변환단계(S30)에서 변환된 위상변이된 두 개의 시계열 데이터, 즉 횡방향 진동 신호와 종방향 진동 신호를 CRP(Cross Recurrence Plot)를 통해 시각화하여 CNN 학습에 입력 데이터로 사용될 수 있는 이미지 형태의 학습 데이터, 즉 CRP 이미지를 생성하는 과정으로, 포트레이트 생성단계(S42), 행렬변환단계(S44) 및 색상표현단계(S46)를 포함할 수 있다.

[0070] 먼저, 상기 포트레이트 생성단계(S42)는 신호변환단계(S30)에서 변환된 위상 변이된 횡방향 진동과 종방향 진동을 재구성하여 2차원 공간에서의 위상 궤적 포트레이트(phase trajectory portrait)를 생성하기 위한 과정으로, 전술한 포트레이트 생성부(132)에서 설명한 바와 같이 위상 변이된 횡방향 진동 신호와 종방향 진동 신호를 짧은 윈도우(window)로 분할한 다음 도 5의 (c)에 나타난 바와 같이, 2차원 상평면 상에 위상 궤적 포트레이트로 재구성할 수 있는데, 이와 같은 위상 궤적 포트레이트로의 재구성 방법은 이미 공지된 것이므로 이에 대한 상세한 설명은 생략하기로 한다.

[0071] 다음, 상기 행렬변환단계(S44)는 도 5의 (d)에 나타난 바와 같이, 위상 궤적 포트레이트에 포함된 횡방향 진동 신호와 종방향 진동 신호의 위상 궤적을 거리 행렬로 변환시키는 과정이고, 상기 색상표현단계(S46)는 도 5의 (e)에 나타난 바와 같이, 거리 행렬의 강도에 따라 픽셀 별로 색상을 부여하여 최종적인 CRP 이미지를 생성하는 과정으로 각각 시각화장치(130)의 행렬변환부(134a) 및 색상표현부(134b)에 의해 수행될 수 있다.

[0072] 보다 상세히 설명하면, 위상 궤적 포트레이트에 포함된 두 진동의 시계열 데이터, 즉 횡방향 진동 신호와 종방향 진동 신호의 위상 궤적은 각각 아래의 (1) 및 (2)식과 같이 표현될 수 있다.

$$x(t) = \vec{x}_i = \{u_1, u_2, u_3, \dots, u_n\} \quad \dots (1)$$

$$y(t) = \vec{y}_j = \{v_1, v_2, v_3, \dots, v_n\} \quad \dots (2)$$

[0075] 여기서, $x(t)$ 와 $y(t)$ 는 각각 시계열을 의미하고, i, j 는 각각 i 번째, j 번째 카운트를 의미한다.

[0076] 상기 행렬변환단계(S44)에서는 (1) 및 (2)식에 해당하는 두 시계열 데이터의 거리행렬을 연산할 수 있는데, 공학용 소프트웨어인 MATLAB의 툴박스 중 "Cross recurrence plot and quantification analysis"를 이용하면 두 시계열 데이터의 거리 행렬인 CR_{ij} 를 다음의 (3)식에 의해 연산할 수 있다.

$$CR_{ij} = \Theta(\varepsilon - \|\vec{x}_i - \vec{y}_j\|) \quad \dots (3)$$

[0078] 여기서, Θ 는 헤비사이드 함수(Heaviside function)이고, ε 은 지정된 역치값(threshold)을 의미하는 것으로, CR_{ij} 는 실질적으로 x_i 와 y_j 사이의 거리를 의미할 수 있다.

[0079] 다음, 상기 색상표현단계(S46)에서는 행렬변환단계(S44)를 통해 얻어진 거리행렬의 강도에 따라 픽셀들에 기설정된 색상을 부여함으로써 최종적인 CRP 이미지를 생성할 수 있다.

[0080] 상기과 같이 생성된 CRP 이미지는 실시간 음성 인식을 위한 CNN 훈련의 학습 데이터로 사용될 수 있는데, 기존의 CNN 학습 데이터로 사용되는 STFT를 이용하여 생성된 스펙트로그램 이미지와 같은 $m \times n \times 3$ 행렬(여기서, $m \times n$ 은 이미지의 너비(width)와 높이(height)이고, 3은 색상 채널 수임)로 나타낼 수 있으므로 기존의 CNN 학습 모델에 사용될 수 있다.

[0081] 즉, 상기 CRP 이미지는 컨볼루션(convolution), 풀링(pooling), 드롭아웃(drop out)을 포함하는 이미지를 이용한 음성인식 CNN 학습의 입력데이터로 사용되어 음향신호의 특징이 추출될 수 있고, 추출된 특징들은 소프트맥스 함수(Softmax function)를 통해 최종적으로 분류될 수 있는데, 이와 같은 음성 인식 방법은 공지된 구성으로, 본 발명에서 권리로서 청구하고자 하는 바가 아니므로 이에 대한 보다 상세한 설명은 생략하기로 한다.

[0082] 따라서, 전술한 바와 같은 본 발명에 따른 음성 인식을 위한 학습 데이터 생성 시스템 및 방법에 의하면, 사람의 고막과 유사한 점탄성 특성을 갖는 다이어프램(112)의 음파에 의한 동적 거동 특성 및 이를 이용한 CRP 이미지 생성을 통해 STFT를 사용하지 않고서도 음성 인식을 위한 CNN 학습에 사용되는 입력 데이터(input data), 즉 학습 데이터를 생성할 수 있을 뿐만 아니라, 종래의 STFT 방식에 비해 CNN 학습 데이터 생성을 위해 필요한 연산에 대한 부담을 현저하게 줄일 수 있으며, 그에 따라 학습 데이터의 생성에 소요되는 시간을 대폭 단축시킬 수 있는 등의 다양한 장점을 갖는 것이다.

[0084] 전술한 실시예들은 본 발명의 가장 바람직한 예에 대하여 설명한 것이지만, 상기 실시예에만 한정되는 것은 아니며, 본 발명의 기술적 사상을 벗어나지 않는 범위 내에서 다양한 변형이 가능하다는 것은 당업자에게 있어서 명백한 것이다.

산업상 이용가능성

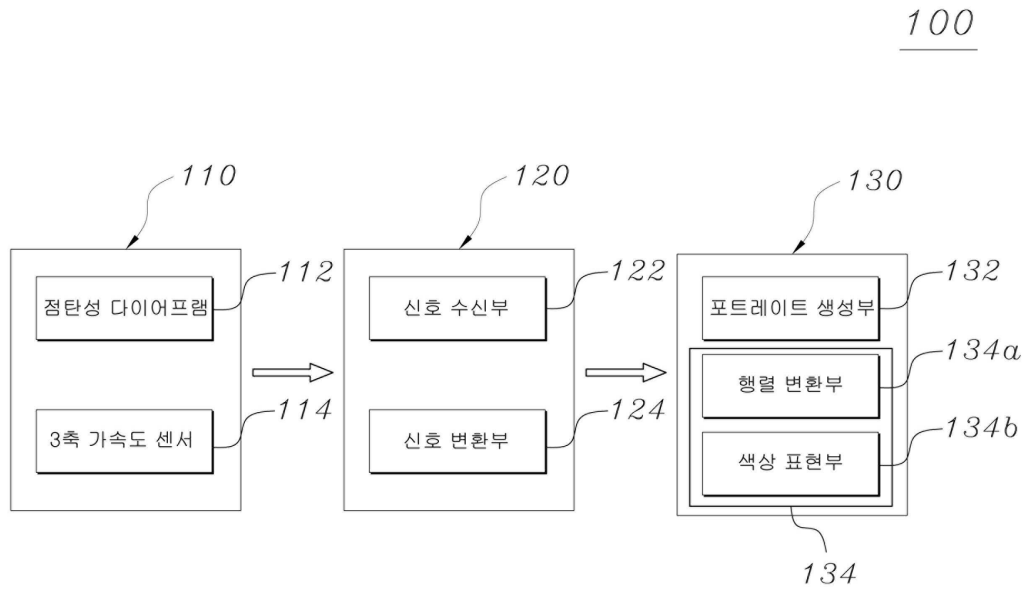
[0086] 본 발명은 음성 인식을 위한 학습 데이터 생성 시스템 및 방법에 관한 것으로, 보다 상세하게는 사람의 고막과 유사한 점탄성 특성을 갖는 음향 센서를 이용하여 수신한 음향 신호를 이용하여 음성 인식을 위한 CNN 학습에 입력 데이터로 사용될 수 있는 시각화된 이미지 형태의 학습 데이터를 생성할 수 있도록 하는 음성 인식을 위한 학습 데이터 생성 시스템 및 방법에 관한 것이다.

부호의 설명

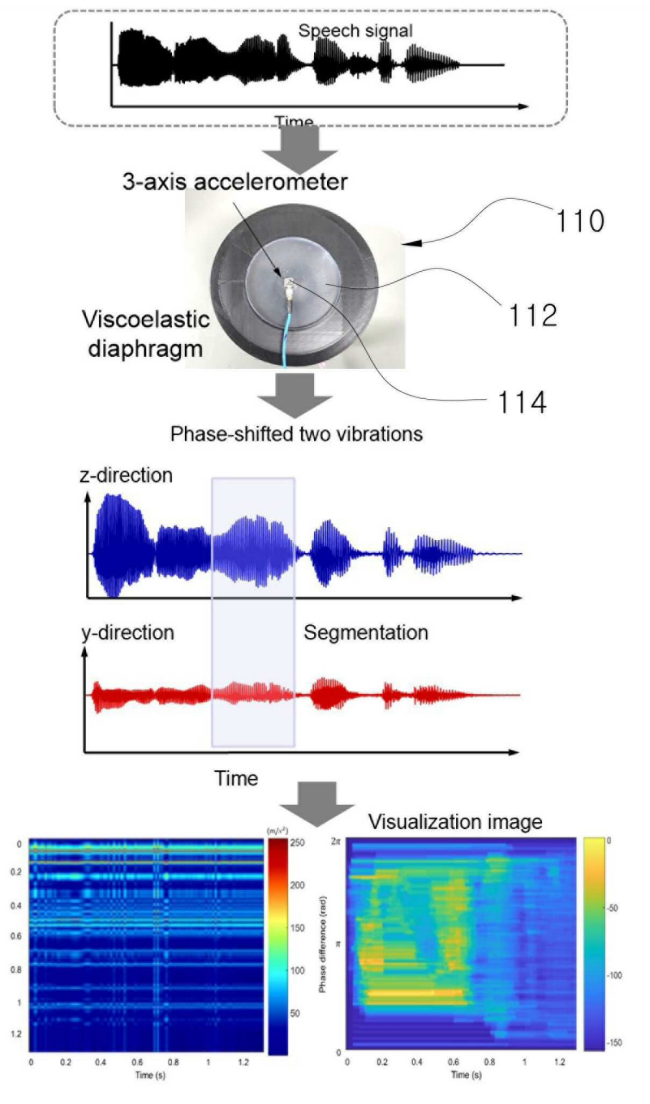
[0088] 100 : 학습 데이터 생성 시스템 110 : 음향 센서
 112 : 점탄성 다이어프램 114 : 3축 가속도 센서
 120 : 신호변환장치 122 : 신호수신부
 124 : 신호변환부 130 : 시각화장치
 132 : 포트레이트 생성부 134 : 이미지 생성부
 134a : 행렬변환부 134b : 색상표현부
 S10 : 음향신호 수신단계 S20 : 가속도 측정단계
 S30 : 신호변환단계 S40 : 이미지 생성단계
 S42 : 포트레이트 생성단계 S44 : 행렬변환단계
 S46 : 색상표현단계

도면

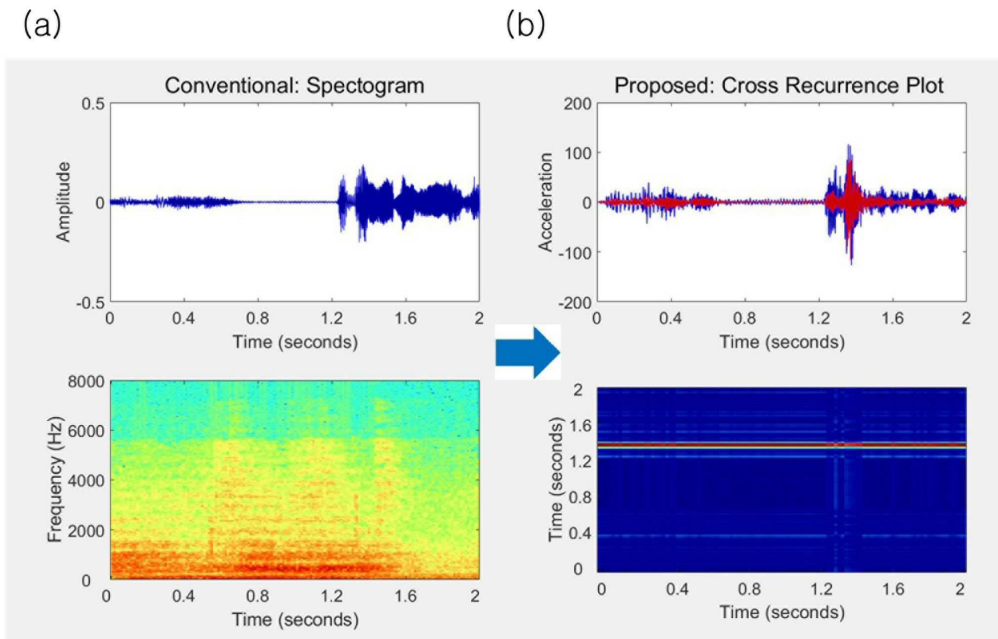
도면1



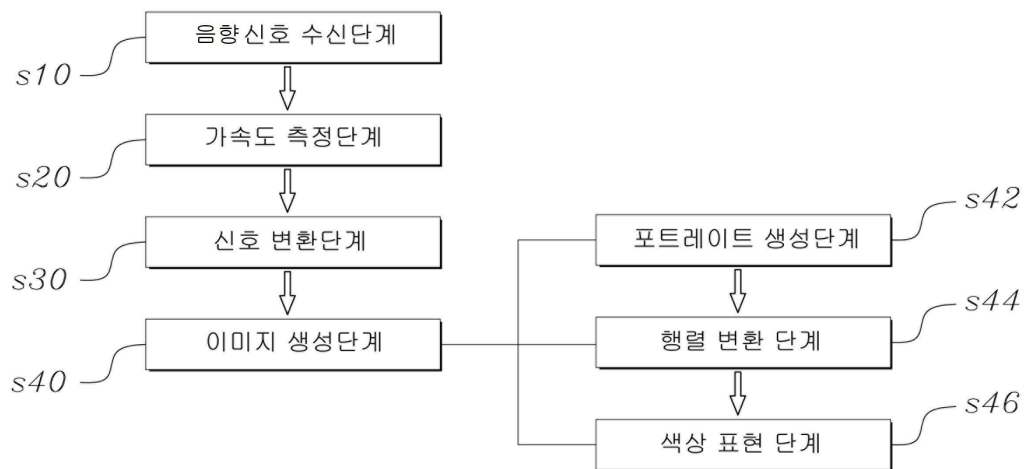
도면2



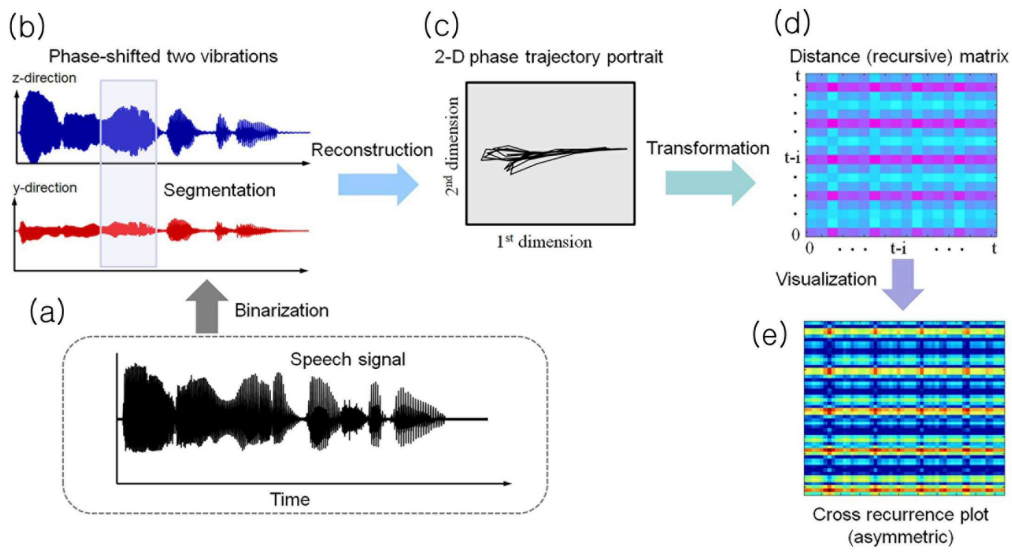
도면3



도면4



도면5



도면6

