



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년09월12일
(11) 등록번호 10-1777302
(24) 등록일자 2017년09월05일

(51) 국제특허분류(Int. Cl.)
G06F 17/14 (2006.01) G06F 15/18 (2017.01)
G06N 3/08 (2006.01) G06N 99/00 (2010.01)
(52) CPC특허분류
G06F 17/141 (2013.01)
G06F 15/18 (2013.01)
(21) 출원번호 10-2016-0094221
(22) 출원일자 2016년07월25일
심사청구일자 2016년07월25일
(30) 우선권주장
1020160047085 2016년04월18일 대한민국(KR)
(56) 선행기술조사문헌
KR1019960011843 A
KR1020060096684 A

(73) 특허권자
충남대학교산학협력단
대전광역시 유성구 대학로 99 (궁동, 충남대학교)
(72) 발명자
안강현
대전광역시 유성구 어은로 57, 101동 304호 (어은동, 한빛아파트)
(74) 대리인
유미특허법인

전체 청구항 수 : 총 18 항

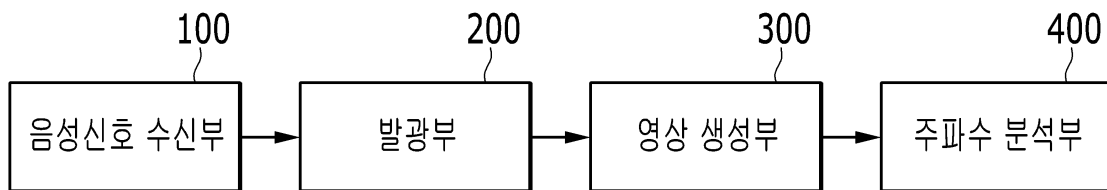
심사관 : 홍경아

(54) 발명의 명칭 음성 주파수 분석 시스템 및 음성 주파수 분석 방법과 이를 이용한 음성 인식 시스템 및 음성 인식 방법

(57) 요약

본 발명은 음성 주파수 분석 시스템 및 음성 주파수 분석 방법과 이를 이용한 음성 인식 시스템 및 음성 인식 방법에 관한 것으로, 본 발명의 실시예에 따른 음성 주파수 분석 시스템은 음성신호를 수신하고 수신된 음성신호에 대응하는 진행파를 생성하는 음성신호 수신부, 음성신호 수신부로 광을 조사하는 발광부, 발광부를 통해 조사된 광에 기초하여 음성신호 수신부를 통해 투과되는 광을 측정하여 영상 패턴을 생성하는 영상 생성부, 그리고 생성된 영상 패턴을 머신러닝을 이용하여 분석하는 주파수 분석부를 포함한다.

대표도 - 도1



(52) CPC특허분류

G06N 3/08 (2013.01)

G06N 99/005 (2013.01)

명세서

청구범위

청구항 1

음성신호를 수신하고 수신된 음성신호에 대응하는 진행파를 생성하는 음성신호 수신부,
 상기 음성신호 수신부로 광을 조사하는 발광부,
 상기 발광부를 통해 조사된 광에 기초하여 상기 음성신호 수신부를 통해 투과되는 광을 측정하여 영상 패턴을 생성하는 영상 생성부, 그리고
 상기 생성된 영상 패턴을 머신러닝을 이용하여 분석하는 주파수 분석부
 를 포함하는 음성 주파수 분석 시스템.

청구항 2

제1항에서,
 상기 음성신호 수신부는 기저막을 포함하는 파브리페로의 간섭계(Fabry-Perot interferometer)인 음성 주파수 분석 시스템.

청구항 3

제1항에서,
 상기 음성신호 수신부는,
 제1 방향으로 수신되는 상기 음성신호에 기초하여 제2 방향의 상기 진행파를 생성하는 제1 레이어,
 상기 제1 레이어 하단에 위치하며 상기 제1 레이어로부터 반사되는 광을 투과시키는 제2 레이어, 그리고
 상기 제1 레이어와 상기 제2 레이어 사이에 위치하며 상기 발광부를 통해 조사된 광이 반사되는 내부 공간이 형성되어 있는 지지부
 를 포함하는 음성 주파수 분석 시스템.

청구항 4

제3항에서,
 상기 제1 방향과 상기 제2 방향은 수직인 음성 주파수 분석 시스템.

청구항 5

제3항에서,
 상기 제1 레이어는 상기 진행파에 대응하는 강성을 갖는 음성 주파수 분석 시스템.

청구항 6

제3항에서,
 상기 제1 레이어는 다양한 두께 또는 길이를 포함하는 음성 주파수 분석 시스템.

청구항 7

제3항에서,
 상기 제2 레이어는 투명한 재료를 포함하며 일정한 두께를 갖는 음성 주파수 분석 시스템.

청구항 8

제1항에서,

상기 발광부는 LED(Light Emitting Diode), LD(Laser Diode), 또는 반도체 레이저(Semiconductor Laser)인 음성 주파수 분석 시스템.

청구항 9

제1항에서,

상기 영상 생성부는 CMOS(Complementary Metal-Oxide Semiconductor) 이미지 센서 또는 CCD(Charge-Coupled Device) 이미지 센서인 음성 주파수 분석 시스템.

청구항 10

제1항에서,

상기 주파수 분석부는 피드포워드(feedforward) 기반의 인공 신경망 학습과 역전파(back propagation) 학습을 통해 상기 영상 패턴에 기초하여 상기 수신된 음성신호의 주파수를 식별하는 음성 주파수 분석 시스템.

청구항 11

음성신호를 수신하고 수신된 음성신호에 대응하는 진행파를 생성하는 음성신호 수신부,

상기 음성신호 수신부로 광을 조사하는 발광부,

상기 발광부를 통해 조사된 광에 기초하여 상기 음성신호 수신부를 통해 투과되는 광을 측정하여 영상 패턴을 생성하는 영상 생성부, 그리고

상기 영상 패턴에 기초하여 음성인식을 수행하는 음성 인식부

를 포함하는 음성 인식 시스템.

청구항 12

제11항에서,

상기 영상 패턴을 머신러닝을 이용하여 분석하는 주파수 분석부를 더 포함하며,

상기 음성 인식부는 상기 분석결과에 기초하여 음성인식을 수행하는 음성 인식 시스템.

청구항 13

음성신호 수신부를 통해 외부 음성신호를 수신하고 수신된 음성신호에 대응하는 진행파를 생성하는 단계,

발광부를 통해 상기 음성신호 수신부로 광을 조사하는 단계,

영상 생성부를 통해 상기 조사된 광에 의해 상기 음성신호 수신부를 통해 투과되는 광을 측정하여 영상 패턴을 생성하는 단계, 그리고

주파수 분석부를 통해 상기 생성된 영상패턴을 분석하여 상기 수신된 음성신호의 주파수를 식별하는 단계

를 포함하는 음성 주파수 분석 방법.

청구항 14

제13항에서,

상기 진행파 생성 단계는 제1 방향의 상기 음성신호를 수신하고 제2 방향으로 상기 진행파를 생성하는 음성 주파수 분석 방법.

청구항 15

제13항에서,

상기 영상 패턴 생성 단계는 상기 수신된 음성신호의 주파수별 영상 패턴을 생성하는 음성 주파수 분석 방법.

청구항 16

제13항에서,

상기 주파수 식별 단계는 피드포워드(feedforward) 기반의 인공 신경망 학습과 역전파(back propagation) 학습을 통해 상기 수신된 음성신호의 주파수를 식별하는 음성 주파수 분석 방법.

청구항 17

음성신호 수신부를 통해 외부 음성신호를 수신하고 수신된 음성신호에 대응하는 진행파를 생성하는 단계,

발광부를 통해 상기 음성신호 수신부로 광을 조사하는 단계,

영상 생성부를 통해 상기 조사된 광에 의해 상기 음성신호 수신부를 통해 투과되는 광을 측정하여 영상 패턴을 생성하는 단계, 그리고

음성 인식부를 통해 상기 영상 패턴을 분석하여 음성인식을 수행하는 단계

를 포함하는 음성 인식 방법.

청구항 18

제17항에서,

주파수 분석부를 통해 상기 영상패턴을 분석하여 상기 수신된 음성신호의 주파수를 식별하는 단계를 더 포함하며,

상기 음성인식단계는 상기 주파수 식별결과에 기초하여 음성인식을 수행하는 음성 인식 방법.

발명의 설명

기술 분야

[0001] 본 발명은 음성 주파수 분석 시스템 및 음성 주파수 분석 방법과 이를 이용한 음성 인식 시스템 및 음성 인식 방법에 관한 것으로, 인공 기저막을 통해 수신되는 음성 신호를 시각적 신호로 변환하고 머신러닝을 이용하여 음성의 주파수를 분석하고 음성인식을 수행하는 시스템 및 방법에 관한 것이다.

배경 기술

[0002] 종래 음성인식을 위한 특징 추출 방법으로는 LPC 필터를 이용하여 특징을 추출하는 LPC(Linear Prediction coding)와 고속 푸리에 변환(FFT)에 기초하여 특징을 추출하는 MFCC(Mel-Frequency Cepstral Coefficients)가 있다. MFCC는 인간의 귀의 특성을 고려하는 점에서 LPC 대비 잡음에 강하며 인식성능이 뛰어나 널리 활용되고 있으나, 마이크로폰을 통해 수신된 음성신호를 저장한 후 고속 푸리에 변환(FFT) 및 MFCC를 수행하는 데에 많은 시간이 소요된다.

[0003] 음성은 자음과 모음이 결합하여 소리의 단위를 이루므로 음성인식에서 자음과 모음을 추출하기 위해서는 음성신호를 시간 영역에서 주파수 영역으로 전환하는 푸리에 변환을 수행해야 한다. 이때, 신호 처리 과정에서 신호 처리 지연 시간이 10ms 이상이 되는 경우 사람이 시간 지연을 감지할 수 있으므로, 이를 방지하기 위해 대부분의 푸리에 변환에서는 약 10ms의 프레임을 사용한다. 이로 인해, 발음시간이 약 150ms인 모음은 10개 이상의 프레임을 이용하여 인식할 수 있으나, 발음시간이 약 17ms인 자음은 2개의 프레임을 이용하여 인식하므로 자음은 모음 대비 인식율이 낮다. 또한, 자음은 모음 대비 주파수 성분이 불규칙적이며 진폭이 작아 외부 노이즈로 인해 왜곡되기 쉽다는 점에서 인식 정확도가 낮을 수 있다.

[0004] 종래 음성인식시스템에서는 인식된 음성을 분석하여 자음 후보군을 추출하고 문장 전체를 다 듣고 난 후 문장에 적합한 하나의 자음을 선택하는 방법을 사용하고 있으나, 문장을 다 듣고 난 후 문장에 적합한 자음을 선택하여 문장을 인식하므로 실시간 음성인식이 어렵고 인식 정확도가 낮다.

발명의 내용

해결하려는 과제

- [0005] 본 발명의 하나의 실시예가 해결하려는 과제는 인공 기저막을 통해 음성 신호를 수신하고 수신된 음성 신호를 시각적 신호로 변환하여 음성의 주파수를 분석하는 시스템 및 방법을 제공하는 것이다.
- [0006] 본 발명의 하나의 실시예가 해결하려는 과제는 인공 기저막을 통해 음성 신호를 수신하여 영상 패턴을 생성하고 생성된 영상 패턴을 분석하여 음성 인식을 수행하는 시스템 및 방법을 제공하는 것이다.
- [0007] 본 발명의 하나의 실시예가 해결하려는 과제는 머신러닝을 통해 인공 기저막의 움직임에 대응하는 영상 패턴을 분석하여 주파수를 식별하고 음성인식을 수행하는 시스템 및 방법을 제공하는 것이다.
- [0008] 상기 과제 이외에도 구체적으로 언급되지 않은 다른 과제를 달성하는 데 본 발명에 따른 실시예가 사용될 수 있다.

과제의 해결 수단

- [0009] 상기 과제를 해결하기 위해 본 발명의 하나의 실시예는, 음성신호를 수신하고 수신된 음성신호에 대응하는 진행파를 생성하는 음성신호 수신부, 음성신호 수신부로 광을 조사하는 발광부, 발광부를 통해 조사된 광에 기초하여 음성신호 수신부를 통해 투과되는 광을 측정하여 영상 패턴을 생성하는 영상 생성부, 그리고 생성된 영상 패턴을 머신러닝을 이용하여 분석하는 주파수 분석부를 포함하는 음성 주파수 분석 시스템을 제안한다.
- [0010] 여기서, 음성신호 수신부는 기저막을 포함하는 파브리페로의 간섭계(Fabry-Perot interferometer)일 수 있다.
- [0011] 또한, 음성신호 수신부는, 제1 방향으로 수신되는 음성신호에 기초하여 제2 방향의 진행파를 생성하는 제1 레이어, 제1 레이어 하단에 위치하며 제1 레이어로부터 반사되는 광을 투과시키는 제2 레이어, 그리고 제1 레이어와 제2 레이어 사이에 위치하며 발광부를 통해 조사된 광이 반사되는 내부 공간이 형성되어 있는 지지부를 포함할 수 있다.
- [0012] 또한, 제1 방향과 제2 방향은 수직일 수 있다.
- [0013] 또한, 제1 레이어는 진행파에 대응하는 강성을 가질 수 있다.
- [0014] 또한, 제1 레이어는 다양한 두께 또는 길이를 포함할 수 있다.
- [0015] 또한, 제2 레이어는 투명한 재료를 포함하며 일정한 두께를 가질 수 있다.
- [0016] 또한, 발광부는 LED(Light Emitting Diode), LD(Laser Diode), 또는 반도체 레이저(Semiconductor Laser)일 수 있다.
- [0017] 또한, 영상 생성부는 CMOS(Complementary Metal-Oxide Semiconductor) 이미지 센서 또는 CCD(Charge-Coupled Device) 이미지 센서일 수 있다.
- [0018] 또한, 주파수 분석부는 피드포워드(feedforward) 기반의 인공 신경망 학습과 역전파(back propagation) 학습을 통해 영상 패턴에 기초하여 수신된 음성신호의 주파수를 식별할 수 있다.
- [0019] 상기 과제를 해결하기 위해 본 발명의 하나의 실시예는, 음성신호를 수신하고 수신된 음성신호에 대응하는 진행파를 생성하는 음성신호 수신부, 음성신호 수신부로 광을 조사하는 발광부, 발광부를 통해 조사된 광에 기초하여 음성신호 수신부를 통해 투과되는 광을 측정하여 영상 패턴을 생성하는 영상 생성부, 그리고 영상 패턴을 분석하여 음성인식을 수행하는 음성 인식부를 포함하는 음성 인식 시스템을 제안한다.
- [0020] 여기서, 영상 패턴을 머신러닝을 이용하여 분석하는 주파수 분석부를 더 포함하며, 음성 인식부는 분석결과에 기초하여 음성인식을 수행할 수 있다.
- [0021] 상기 과제를 해결하기 위해 본 발명의 하나의 실시예는, 음성신호 수신부를 통해 외부 음성신호를 수신하고 수신된 음성신호에 대응하는 진행파를 생성하는 단계, 발광부를 통해 음성신호 수신부로 광을 조사하는 단계, 영상 생성부를 통해 조사된 광에 의해 음성신호 수신부를 통해 투과되는 광을 측정하여 영상 패턴을 생성하는 단계, 그리고 주파수 분석부를 통해 생성된 영상패턴을 분석하여 수신된 음성신호의 주파수를 식별하는 단계를 포함하는 음성 주파수 분석 방법을 제안한다.
- [0022] 여기서, 진행파 생성 단계는 제1 방향의 음성신호를 수신하고 제2 방향으로 진행파를 생성할 수 있다.

- [0023] 또한, 영상 패턴 생성 단계는 수신된 음성신호의 주파수별 영상 패턴을 생성할 수 있다.
- [0024] 또한, 주파수 식별 단계는 피드포워드(feedforward) 기반의 인공 신경망 학습과 역전파(back propagation) 학습을 통해 수신된 음성신호의 주파수를 식별할 수 있다.
- [0025] 상기 과제를 해결하기 위해 본 발명의 하나의 실시예는, 음성신호 수신부를 통해 외부 음성신호를 수신하고 수신된 음성신호에 대응하는 진행파를 생성하는 단계, 발광부를 통해 음성신호 수신부로 광을 조사하는 단계, 영상 생성부를 통해 조사된 광에 의해 음성신호 수신부를 통해 투과되는 광을 측정하여 영상 패턴을 생성하는 단계, 그리고 음성 인식부를 통해 영상 패턴을 분석하여 음성인식을 수행하는 단계를 포함하는 음성 인식 방법을 제안한다.
- [0026] 여기서, 주파수 분석부를 통해 영상패턴을 분석하여 수신된 음성신호의 주파수를 식별하는 단계를 더 포함하며, 음성인식단계는 주파수 식별결과에 기초하여 음성인식을 수행할 수 있다.

발명의 효과

- [0027] 본 발명의 하나의 실시예에 따르면 음성의 주파수 분석 및 음성 인식을 실시간으로 수행할 수 있다. 또한, 음성의 주파수 분석 및 음성 인식에 소요되는 시간 및 소비전력을 감소시킬 수 있다.

도면의 간단한 설명

- [0028] 도 1은 본 발명의 하나의 실시예에 따른 음성 주파수 분석 시스템의 구성을 나타낸다.
- 도 2는 본 발명의 하나의 실시예에 따른 음성신호 수신부의 구조를 나타낸다.
- 도 3은 도 2의 음성신호 수신부의 제조 방법을 나타낸다.
- 도 4는 본 발명의 하나의 실시예에 따른 음성신호 수신부, 발광부 및 영상 생성부의 구현 예시이다.
- 도 5는 본 발명의 하나의 실시예에 따른 음성신호 수신부의 기울기 강성과 진동의 관계를 나타내는 모델이다.
- 도 6은 하나의 진동 사이클 시간에 생성되는 피크 포인트와 파장 포락선을 나타낸다.
- 도 7은 본 발명의 하나의 실시예에 따른 기저막 변위의 위치 의존적 분산(position-dependent variance)을 나타낸다.
- 도 8은 도 4의 영상 생성부를 통해 생성된 영상의 예시이다.
- 도 9는 신경망을 도식화하여 나타낸다.
- 도 10은 본 발명의 실시예에 따른 역전파 방법을 이용하는 훈련방법을 나타낸다.
- 도 11은 종래 훈련 방법을 이용한 인공 신경망 모델의 시뮬레이션 결과를 나타낸다.
- 도 12는 본 발명의 하나의 실시예에 따른 훈련 방법을 이용한 인공 신경망 모델의 시뮬레이션 결과를 나타낸다.
- 도 13은 도 1의 음성 주파수 분석 시스템을 이용하는 음성 주파수 분석 방법을 나타낸다.
- 도 14는 본 발명의 하나의 실시예에 따른 음성 인식 시스템의 구성을 나타낸다.
- 도 15는 본 발명의 실시예에 따른 제1 레이어의 움직임 시뮬레이션 결과이다.
- 도 16은 도 14의 음성 인식 시스템을 이용한 음성 인식 방법을 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0029] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 대해 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며 명세서 전체에서 동일 또는 유사한 구성요소에 대해서는 동일한 도면부호가 사용되었다. 또한, 널리 알려진 공지기술의 경우 그 구체적인 설명은 생략한다.
- [0030] 본 명세서에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된

"...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

- [0031] 본 명세서에서, "주파수 분해능(frequency resolution)"은 음성신호 자극시 시간에 따른 주파수의 변화를 판단하는 능력을 의미하며, "음성신호"는 복수의 주파수를 포함하는 인간의 음성을 의미한다.
- [0032] 도 1은 본 발명의 하나의 실시예에 따른 음성 주파수 분석 시스템의 구성을 나타낸다.
- [0033] 도 1의 음성 주파수 분석 시스템은 음성신호 수신부(100), 발광부(200), 영상 생성부(300), 그리고 주파수 분석부(400)를 포함한다.
- [0034] 음성신호 수신부(100)는 외부에서 전달되는 음성신호를 수신한다. 본 발명의 실시예에 따른 음성신호 수신부(100)는 인공 기저막을 포함하는 파브리페로의 간섭계(Fabry-Perot interferometer)로 구현된다. 여기서, 인공 기저막은 인간의 달팽이관 내부의 기저막(Basilar Membrane)을 모방한 형태이며, 수신된 음성신호의 주파수에 기초하여 서로 다른 위치에서 진동을 발생시킨다.
- [0035] 아래에서는 도 2와 도 3을 참조하여 음성신호 수신부에 대해 상세히 설명한다.
- [0036] 도 2는 본 발명의 하나의 실시예에 따른 음성신호 수신부의 구조를 나타낸다.
- [0037] 도 2의 (a)는 음성신호 수신부의 측면도이며, (b)는 음성신호 수신부의 평면도이다.
- [0038] 도 2의 (a)에서 보면, 음성신호 수신부(110)는 음성신호를 수신하며 수신된 음성신호에 기초하여 진동을 발생시키는 제1 레이어(111), 제1 레이어(111)에서 발생하는 진동에 대응하는 빛을 투과시키는 제2 레이어(112), 그리고 제1 레이어(111)와 제2 레이어(112) 사이에 위치하여 제1 레이어(111)를 지지하는 지지부(113)를 포함한다.
- [0039] 제1 레이어(111)는 수신되는 음성신호에 대응하는 강성을 갖는 기저막으로 구현되며 감광성 고분자(photosensitive polymer)를 포함한다. 이때, 기저막의 강성은 질량 또는 길이에 의해 결정된다. 이에 따라, 기저막은 (a)와 같이 다양한 두께(D, D', D'')(질량)로 구현되거나, (b)와 같이 윗변(E)과 아랫변(E')의 길이가 상이한 형태로 구현되거나, 또는 다양한 질량 및 길이를 포함하는 형태로 구현될 수 있다. 예를 들어, 기저막은 (b)에서 보듯이 윗변(E)이 아랫변(E')보다 큰 사다리꼴 형상일 수 있다.
- [0040] 제2 레이어(112)는 투명한 재료를 포함하며 일정한 두께를 갖도록 구현된다. 예를 들어, 유리(glass), 폴리에틸렌(polyethylene), 폴리카보네이트(polycarbonate) 및 메타크릴수지(methacrylic resin) 중 하나를 포함할 수 있다.
- [0041] 지지부(113)는 제1 레이어(111)와 제2 레이어(112)의 양단을 지지하며 제1 레이어(111)와 제2 레이어(112)로 형성되는 내부 공간(S)을 포함한다. 여기서, 지지부(113)는 실리콘을 포함할 수 있다.
- [0042] 도 3은 도 2의 음성신호 수신부의 제조 방법을 나타낸다.
- [0043] 먼저, 실리콘 웨이퍼(114) 위에 감광성 고분자(115)를 코팅하고(S10), 그레이-스케일 포토리소그래피(gray-scale photolithography)를 통해 다양한 두께를 갖는 제1 레이어(111)를 형성한다(S20). 이후, 딥 반응성 이온 에칭(deep reactive ion etching)을 통해 지지부(113)를 형성한다(S30). 이후, S30 단계에서 형성된 지지부(113) 아래에 유리를 접합하여 제2 레이어(112)를 형성한다(S40).
- [0044] 다시 도 1의 설명으로 돌아가서, 발광부(200)는 음성신호 수신부(100)의 제1 레이어(111)에서 발생하는 진동을 측정하기 위한 광을 제공한다. 발광부(200)는 LED(Light Emitting Diode), LD(Laser Diode) 및 반도체 레이저(Semiconductor Laser) 중 하나로 구현될 수 있다.
- [0045] 영상 생성부(300)는 발광부(200)에서 제공되는 광을 이용하여 제1 레이어(111)에서 발생하는 진동에 대응하는 영상 패턴을 생성한다. 영상 생성부(300)는 CMOS(Complementary Metal-Oxide Semiconductor) 이미지 센서 또는 CCD(Charge-Coupled Device) 이미지 센서로 구현될 수 있다.
- [0046] 도 4는 본 발명의 하나의 실시예에 따른 음성신호 수신부, 발광부 및 영상 생성부의 구현 예시이다.
- [0047] 도 4에서 보면, 음성신호 수신부(100)에 음성신호가 수신되는 경우 발광부(200)를 통해 진동이 발생된 제1 레이어(111)로 광을 조사하고, 영상 생성부(300)를 통해 제2 레이어(112)를 통해 투과되는 광을 수신하여 영상 패턴을 생성한다. 본 발명의 실시예에 따른 발광부(200)는 저전력 LED로 구현하고, 영상 생성부(300)는 저전력 CMOS 이미지 센서로 구현할 수 있다.

[0048] 도 4에서 제1 레이어(111)에 발생하는 진동은 제1 레이어(111)와 수직방향으로 진행하는 진행파(traveling wave)이며, 진행파는 공진이 거의 발생하지 않는다는 점에서 정상파(standing wave)보다 주파수 분해능이 좋다. 또한, 종래 달팽이관의 기저막을 모방하는 음성신호 센싱 장치에서는 귓속뼈(Stapes)에 대응하는 구성에 가해진 음성신호 자극을 기저막 전체 영역으로 전달하기 위해 유체를 사용하나, 본 발명의 실시예에 따른 음성신호 수신부에서는 기저막(제1 레이어)을 연속되는 다양한 두께를 갖도록 구현하고 기저막에 음성신호 자극을 직접 인가하여 유체 없이 진행파를 생성하고 광학적으로 측정함으로써 감도를 향상시키고 전력소비를 감소시킬 수 있다.

[0049] 아래에서는 도 5 내지 도 7을 참조하여 음성신호 수신부에서 생성되는 진행파의 특징에 대해 설명한다. 진행파 에너지는 아래 수학식 1과 같다.

수학식 1

$$m_i \ddot{u}_i = \frac{k'_i}{2a^2} (u_{i+1} - u_i)^3 - \frac{k'_{i-1}}{2a^2} (u_i - u_{i-1})^3 - k_i u_i - m_i \gamma \dot{u}_i + f(t)$$

[0050]

[0051] 여기서, m_i 는 질량(mass), u_i 는 수직 변위(vertical displacement), k'_i 는 inter-mass coupling stiffness, k_i 는 국부 강성(local stiffness), 그리고 γ 는 기저막(제1 레이어)의 i 번째 유한 요소의 손실 계수(dissipation constant)이다.

[0052] 도 5는 본 발명의 하나의 실시예에 따른 음성신호 수신부의 기울기 강성과 진동의 관계를 나타내는 모델이며, 도 6은 하나의 진동 사이클 시간에 생성되는 피크 포인트와 파장 포락선을 나타낸다.

[0053] 도 5에서 보면, 달팽이관에 유체가 존재하는 것과 같이 동일한 자극 $f(t)$ 를 각각의 위치에 적용하는 경우 왼쪽에서 오른쪽으로 질량이 증가하는 동안 강성은 감소함을 알 수 있다. 이로 인해, 도 6에서 보듯이, 높은 주파수 진행파의 피크 포지션일수록 왼쪽에 위치함을 알 수 있다.

[0054] $f_a < f < f_b$ 의 주파수 윈도우(표준 파장)에서 스펙트럼 분석시 기저막에 진행파를 생성하기 위해 필요한 최소 시간(T_{membrane})은 아래의 수학식 2와 같다.

수학식 2

$$T_{\text{membrane}} = \frac{1}{f_a}$$

[0055]

[0056] 이는 요구되는 시간(T_{membrane})이 주파수 분해능에 의존적이지 않음을 나타낸다.

[0057] 반면, 수치해석 푸리에 변환을 포함하는 종래 마이크로폰 분석에서는 주어진 시간 동안 데이터를 수신하고, 아래의 수학식 3과 같이 데이터 수신 시간은 주파수 분해능(δf)의 역에 의해 산출한다. 이로 인해, 최소 시간($T_{\text{microphone}}$)은 주파수 분해능(δf)에 의해 결정된다.

수학식 3

$$T_{\text{microphone}} = \frac{1}{\delta f}$$

[0058]

[0059] 따라서, 가장 낮은 주파수 영역(f_a)이 주파수 분해능(δf)보다 큰 경우 기저막을 이용하는 푸리에 분석에 소요되

는 시간(T_{membrane})은 마이크로폰을 이용하는 푸리에 분석에 소요되는 시간($T_{\text{microphone}}$)보다 짧다.

[0060] 도 7은 본 발명의 하나의 실시예에 따른 기저막 변위의 위치 의존적 분산(position-dependent variance)을 나타낸다.

[0061] 도 7은 광학적 측정을 통해 아래의 수학적식 4와 같이 평균화된 변위의 제곱을 산출한 결과를 나타낸다.

수학적식 4

$$[0062] \quad \langle u_i^2 \rangle = \frac{1}{T_{\text{mem}}} \int_t^{t+T_{\text{mem}}} u_i^2 dt$$

[0063] 도 7에서는 70개의 주파수($f_1, f_2, f_3, \dots, f_{70}$)에 대응하는 평균화된 변위의 제곱 값이 서로 다르게 나타나며, 생물학적 기저막과 유사한 피크 구조를 보이는 것을 확인할 수 있다. 구체적으로, 본 발명의 실시예에 따른 기저막(제1 레이어)을 이용하는 푸리에 분석에 소요되는 시간(T_{membrane})은 저주파수 영역에 의해 결정되므로 도 7에서 시간(T_{membrane})은 $1/(f_{30})$ 이나, 마이크로폰을 이용하는 푸리에 분석에 소요되는 시간($T_{\text{microphone}}$)은 $1/f_1$ 이 되어 $1/(f_{30})$ 보다 크다.

[0064] 따라서, 신호 처리 시간을 감소시킴으로써 자음 인식 성능 및 정확도를 향상시킬 수 있다.

[0065] 도 8은 도 4의 영상 생성부를 통해 생성된 영상의 예시이다.

[0066] 도 8의 (a)는 음성신호가 수신되지 않는 경우 영상 생성부(300)를 통해 생성된 영상이며, (b), (c), (d), (e) 및 (f)는 각각 100Hz, 150Hz, 200Hz, 250Hz 및 300Hz의 주파수를 갖는 음성신호가 수신되는 경우 영상 생성부(300)를 통해 생성된 영상이다.

[0067] 다시 도 1의 설명으로 돌아가서, 주파수 분석부(400)는 머신러닝(Machine Learning)을 이용하여 영상 생성부(300)에서 생성된 영상 패턴을 분석한다. 본 발명의 실시예에 따른 주파수 분석부(400)는 피드포워드 신경망에 기초하여 음성신호 수신부(100)에서 수신된 음성신호의 주파수를 분석할 수 있다.

[0068] 아래에서는 본 발명의 실시예에 따른 피드포워드 기반의 인공 신경망 학습방법에 대해 상세히 설명한다.

[0069] 인공 신경망 모델에서 뉴런은 입력과 연결 강도의 가중치 합을 산출하여 활성화 함수를 통해 출력하며, 1번째 층의 j번째 뉴런의 활성화 함수(a_j^1)는 아래의 수학적식 5와 같다. 이때, 활성화 함수는 음성신호 자극이 주어지는 경우 음성신호를 활성화할 것인지 비활성화할 것인지 결정하는데 사용된다.

수학적식 5

$$[0070] \quad a_j^l = \sigma(\sum_k w_{jk}^l a_k^{l-1} + b_j^l)$$

[0071] 여기서, $\sigma(x)$ 는 시그모이드 함수(sigmoid function)이고, w_{jk}^1 은 1번째 층의 k번째 뉴런과 1번째 층의 j번째 뉴런 사이의 연결을 위한 가중치 요소이며, b_j^1 은 1번째 층의 j번째 뉴런을 위한 파라미터 제어 임계값이다.

[0072] 인공 신경망 훈련(분류화 또는 최적화라고도 함) 방법은 오차를 최소화하는 것으로 오차는 아래 수학적식 6과 같다.

수학식 6

$$C = \sum_x |y(x) - a^L(x)|^2$$

[0073]

[0074]

여기서, $y(x) = (y_1(x), y_2(x), \dots)$ 는 입력 데이터 세트 x 를 위한 원하는 출력이고, L 은 출력층의 인덱스이다.

[0075]

종래에는 오차 C 를 최소화하기 위해 아래 수학식 7의 최대 경사법(steepest gradient method)를 사용하였으나 (여기서, η 는 학습률 파라미터(learning rate parameter)이다.), 본 발명의 실시예에서는 신경망이 많은 층을 포함할수록 최대 경사법을 이용하여 오차 함수를 최소화하는 과정이 복잡해지는 문제를 해결하기 위해 아래 수학식 8을 통해 오차 C 를 최소화한다.

수학식 7

$$\begin{cases} w_{jk}^{l'} = w_{jk}^l - \eta \frac{\partial C}{\partial w_{jk}^l} \\ b_j^{l'} = b_j^l - \eta \frac{\partial C}{\partial b_j^l} \end{cases}$$

[0076]

수학식 8

$$\begin{cases} \delta_j^L = \frac{\partial C}{\partial a_j^L} \sigma'(z_j^L) \\ \delta_j^l = \frac{\partial C}{\partial z_j^l} = \sum_k \delta_j^{l+1} w_{jk}^{l+1} \sigma'(z_j^l) \\ \frac{\partial C}{\partial b_j^l} = \delta_j^l \\ \frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l \end{cases}$$

[0077]

[0078]

여기서, z_j^l 은 역전파 방법(back propagation method)으로 아래 수학식 9와 같다.

수학식 9

$$z_j^l = \sum_k w_{jk}^l a_k^{l-1} + b_j^l$$

[0079]

[0080]

다양한 주파수에 대응하는 기저막의 응답을 분류하기 위한 신경망의 입력 데이터는 기저막의 시간 평균화된 로컬 진폭이다. 이때, 신경망의 가중치 요소와 임계값은 훈련을 통해 획득할 수 있다.

[0081]

도 9는 신경망을 도식화하여 나타내며, 도 10은 역전파 방법을 이용하는 훈련방법을 나타낸다.

[0082]

도 10에서 보면, 훈련 결과에 따라 오답인 로컬 트랩에 도달하는 경우 로컬 트랩의 전구체(precursor)를 인지하고 한 단계 후진으로 탈출한 후 랜덤하게 전진하여 오류 최소화 반복을 수행한다.

- [0083] 도 11은 종래 훈련 방법을 이용한 인공 신경망 모델의 시뮬레이션 결과를 나타내며, 도 12는 본 발명의 하나의 실시예에 따른 훈련 방법을 이용한 인공 신경망 모델의 시뮬레이션 결과를 나타낸다.
- [0084] 도 11과 도 12에서 입력 데이터는 70개의 서로 다른 주파수에 대응하는 기저막의 시간 평균화된 진폭의 위치(40개)이다(도 8).
- [0085] 도 11은 수학적 7을 통해 오차를 최소화한 결과를 나타낸 것으로 오랜 훈련에도 100% 최소화를 달성하지 못한 것을 확인할 수 있으며, 도 12는 수학적 8 및 9를 통해 오차를 최소화한 결과를 나타낸 것으로 단시간에 오차를 최소화한 결과를 확인할 수 있다.
- [0086] 이로 인해, 본 발명의 실시예에서는 시간 영역의 데이터 개수와 주파수 영역의 푸리에 성분의 개수가 동일한 종래 푸리에 변환방법과 달리 40개의 입력을 이용하여 70개의 다른 주파수에 대응하는 기저막의 응답을 분류할 수 있음을 확인할 수 있다. 따라서, 단시간에 주파수를 식별할 수 있다.
- [0087] 도 13은 본 발명의 하나의 실시예에 따른 음성 주파수 분석 방법을 나타낸다.
- [0088] 먼저, 음성신호 수신부(100)를 통해 외부 음성신호를 수신하고 기저막(제1 레이어)의 진행과 발생을 감지한다(S100).
- [0089] 이후, S100 단계에서 진행과가 감지되는 경우 발광부(200)를 통해 음성신호 수신부(100)에 광을 조사한다(S200).
- [0090] 이후, 영상 생성부(300)를 통해 음성신호 수신부(100)로부터 투과되는 광을 측정하여 영상 패턴을 생성한다(S300).
- [0091] 이후, 주파수 분석부(400)를 통해 S300 단계에서 생성된 영상 패턴을 머신러닝을 이용하여 분석하여 주파수를 식별한다(S400).
- [0092] 도 14는 본 발명의 하나의 실시예에 따른 음성 인식 시스템의 구성을 나타낸다.
- [0093] 도 14의 음성 인식 시스템은 음성신호 수신부(500), 발광부(600), 영상 생성부(700), 주파수 분석부(800), 그리고 음성 인식부(900)를 포함한다. 여기서, 음성신호 수신부(500), 발광부(600), 영상 생성부(700) 그리고 주파수 분석부(800)는 도 1의 음성신호 수신부(100), 발광부(200), 영상 생성부(300), 그리고 주파수 분석부(400)와 동일한 기능을 수행하므로 중복되는 설명은 생략한다.
- [0094] 도 14의 음성 인식부(900)는 영상 생성부(700)를 통해 생성된 영상 패턴을 분석하여 음성인식을 수행한다.
- [0095] 도 15는 본 발명의 실시예에 따른 제1 레이어의 움직임 시뮬레이션 결과이다.
- [0096] 도 15의 (A)는 음성신호 '가'의 자음 부분의 신호에 대한 압력에 대응하는 제1 레이어의 움직임을 시뮬레이션한 결과이며, (B)는 음성신호 '바'의 자음 부분의 신호에 대한 압력에 대응하는 제1 레이어의 움직임을 시뮬레이션한 결과이고, (C)는 음성신호 '다'의 자음 부분의 신호에 대한 압력에 대응하는 제1 레이어의 움직임을 시뮬레이션한 결과이다.
- [0097] 도 15에서 보면, '가', '바', 그리고 '다'의 자음 신호에 대응하는 제1 레이어의 움직임이 서로 다른 것을 확인할 수 있으며, 음성 인식부(900)는 이를 통해 주파수 분석 과정을 거치지 않고 제1 레이어의 움직임에 대응하는 영상 패턴을 분석하여 자음을 인식할 수 있다.
- [0098] 다시 도 14의 설명으로 돌아가서, 본 발명의 실시예에 따른 음성 인식부(900)는 주파수 분석부(800)의 주파수 분석 결과에 기초하여 음성인식을 수행한다. 예를 들어, 주파수 분석 결과에 따라 특징을 추출하고 추출된 특징을 미리 저장되어 있는 음성모델 데이터베이스와 비교하여 음성 인식 결과를 출력할 수 있다.
- [0099] 도 16은 도 14의 음성 인식 시스템을 이용한 음성 인식 방법을 나타낸다.
- [0100] 도 16의 음성 인식 방법에서 S500 단계 내지 S800 단계는 도 14의 S100 단계 내지 S400 단계와 동일하므로 중복되는 설명은 생략한다.
- [0101] 도 16에서는 음성 인식부(900)를 통해 S800 단계의 주파수 분석 결과에 기초하여 음소를 인식하고(S910), S910 단계의 음소 인식 결과를 바탕으로 S500 단계에서 수신된 음성신호에 대응하는 단어 또는 문장을 인식하여 출력한다(S920). 또한, 음성 인식부(900)를 통해 S700 단계에서 생성된 영상 패턴을 분석하여 음소를 인식하고 단어 또는 문장을 인식하여 출력할 수 있다.

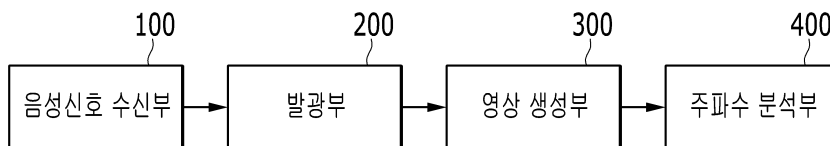
- [0102] 본 발명의 실시예에 따르면 주파수 분해 과정을 거치지 않더라도 음절 인식이 가능하므로 주파수 분해 과정이 필요한 종래 기술 대비 짧은 시간 내에 자음을 인식할 수 있으며, 자음 인식 성능을 향상시킬 수 있다.
- [0103] 본 발명의 실시예에 따른 음성 주파수 분석방법은 장치에 설치되거나 사용자에게 의해 직접 설치된 프로그램으로 구현되어 컴퓨터에 의해 읽힐 수 있는 기록매체에 기록될 수 있다. 여기서, 컴퓨터는 데스크탑, 노트북, 스마트폰, 태블릿 PC, PDA(개인정보 단말기), 이동통신장치 등을 포함할 수 있다. 또한, 기록매체는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 미디어 저장장치 등을 포함할 수 있다.
- [0104] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였으나, 본 발명의 권리범위가 이에 한정되는 것은 아니며 본 발명이 속하는 분야에서 통상의 지식을 가진 자가 여러 가지로 변형 및 개량한 형태 또한 본 발명의 권리범위에 속한다.

부호의 설명

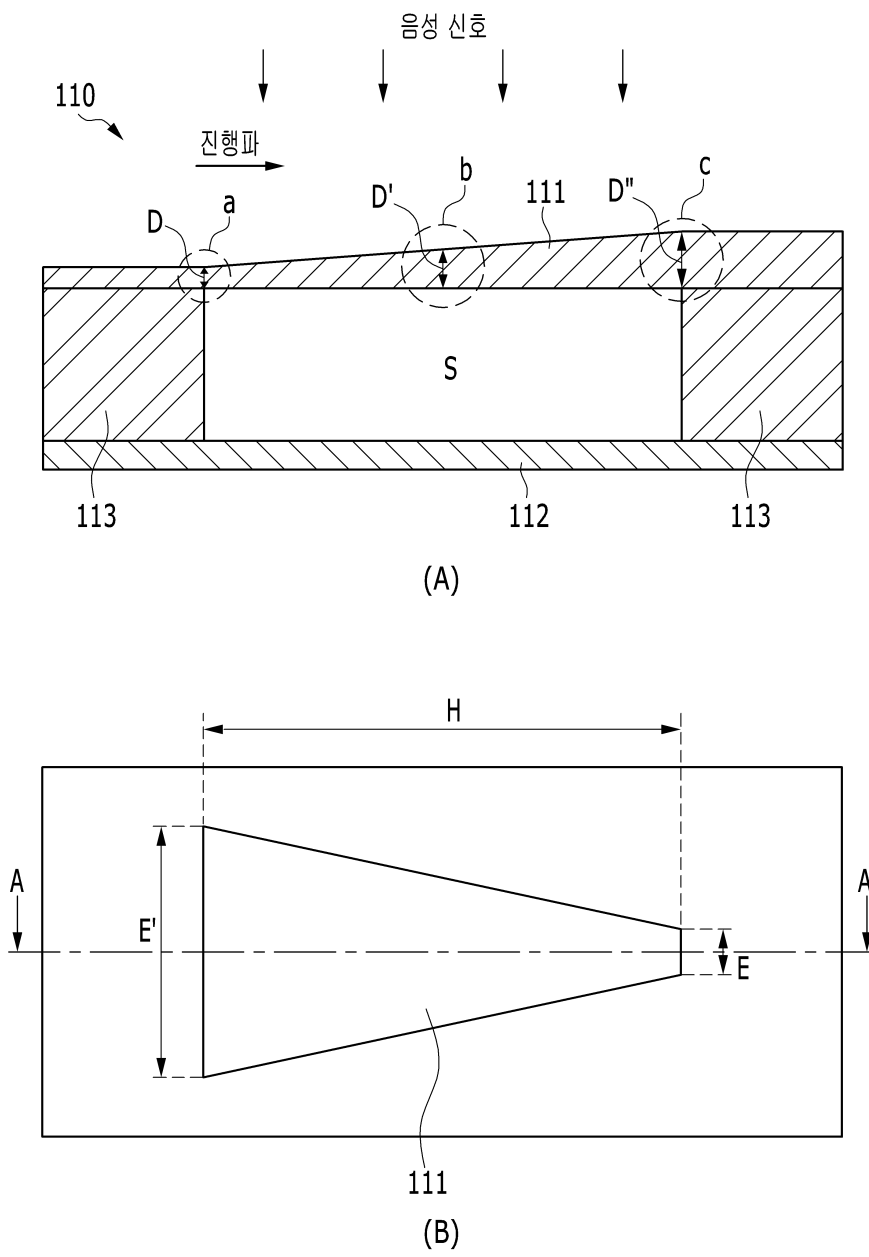
- [0105] 100, 500 : 음성신호 수신부
111 : 제1 레이어
112 : 제2 레이어
113 : 지지부
200, 600 : 발광부
300, 700 : 영상 생성부
400, 800 : 주파수 분석부
900 : 음성 인식부

도면

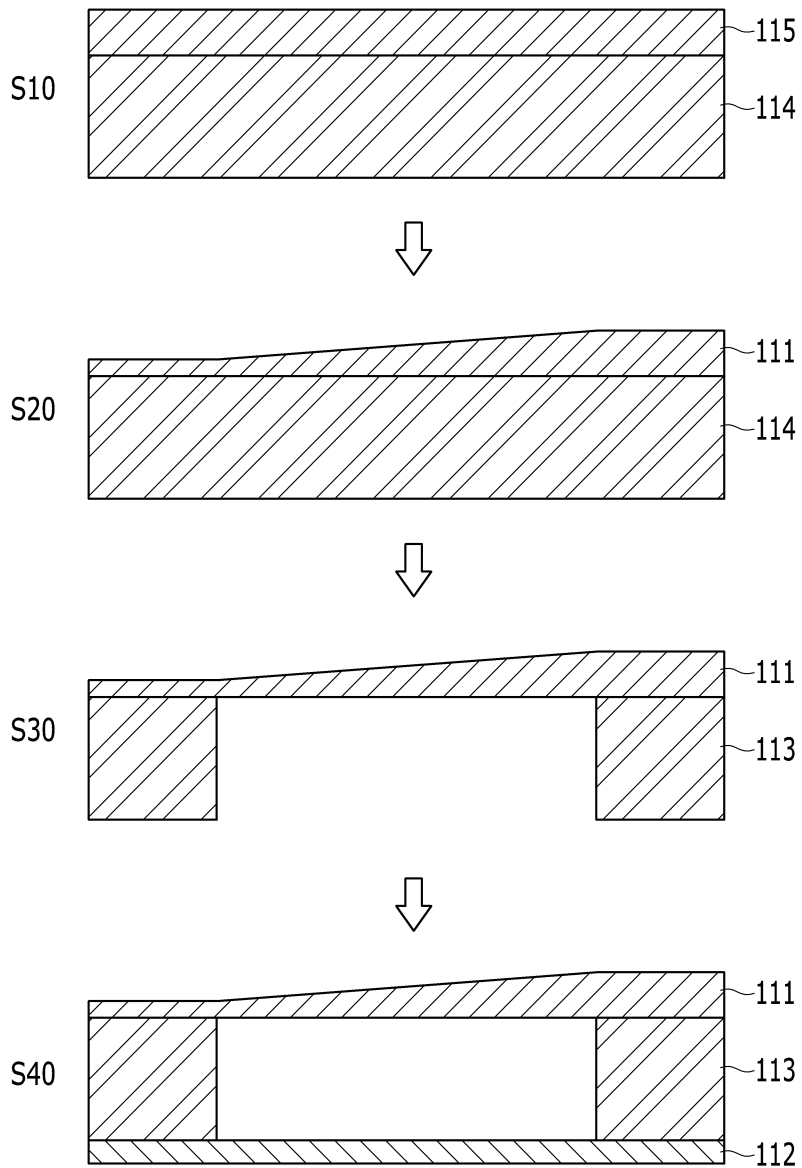
도면1



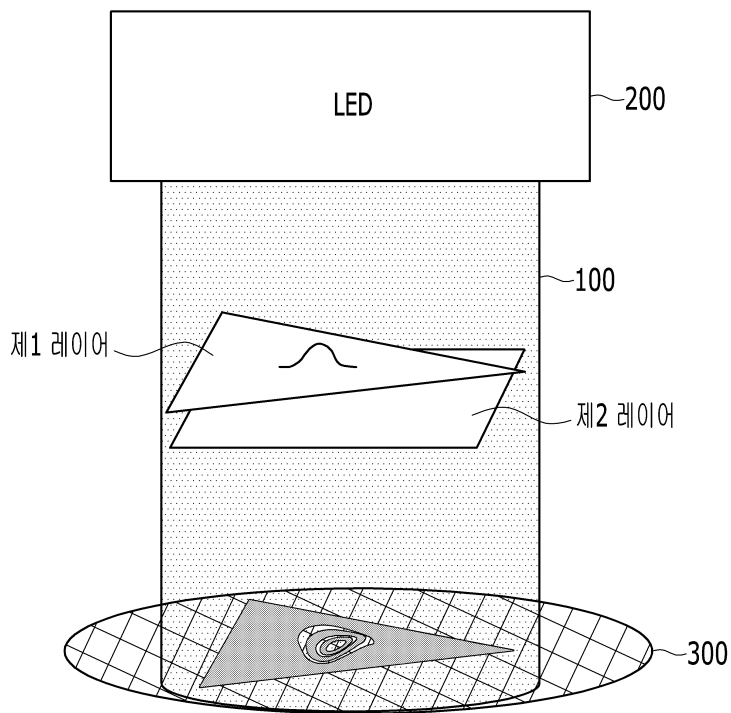
도면2



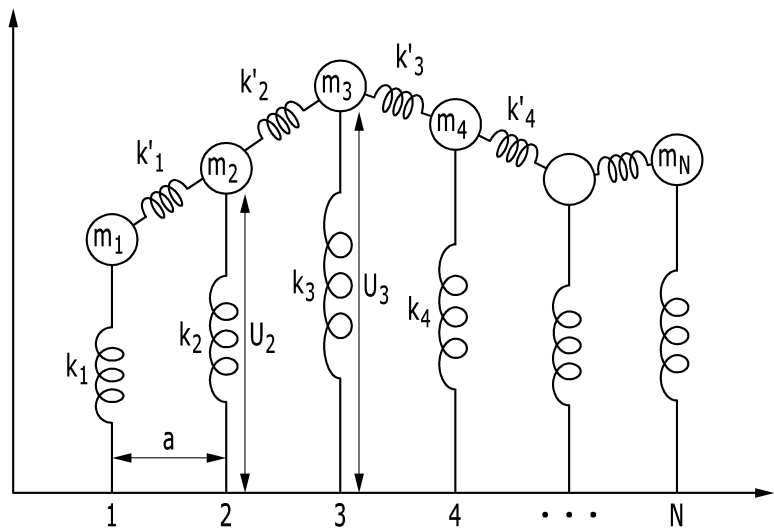
도면3



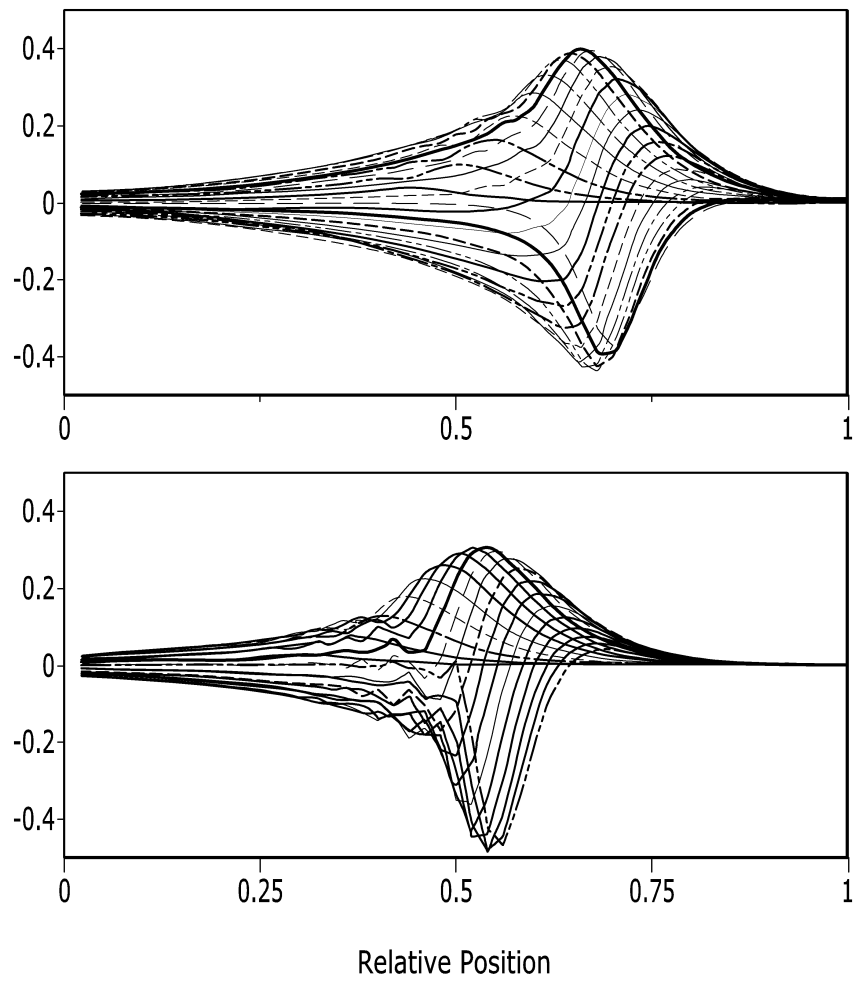
도면4



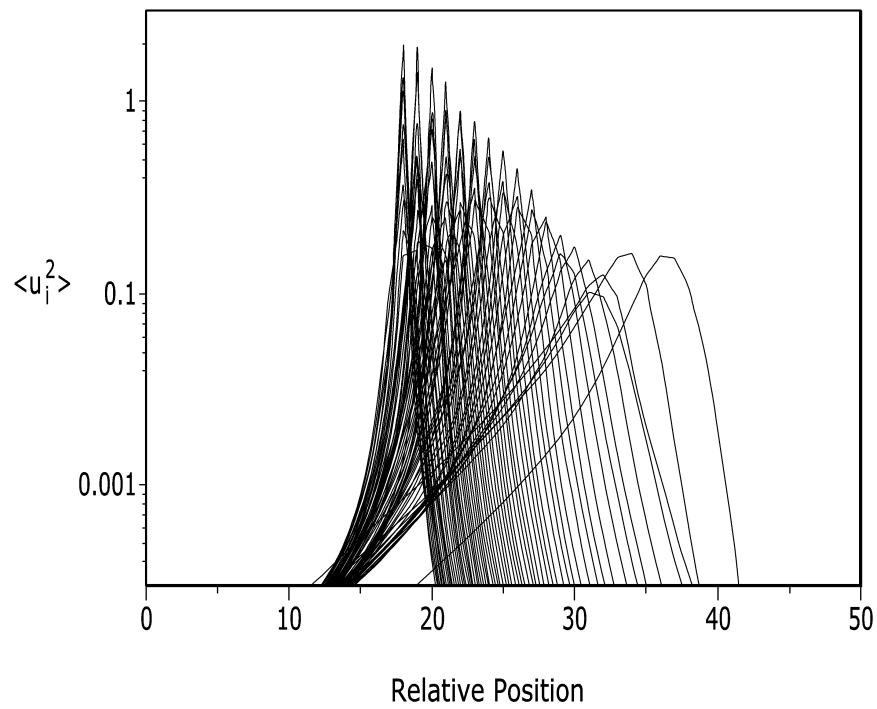
도면5



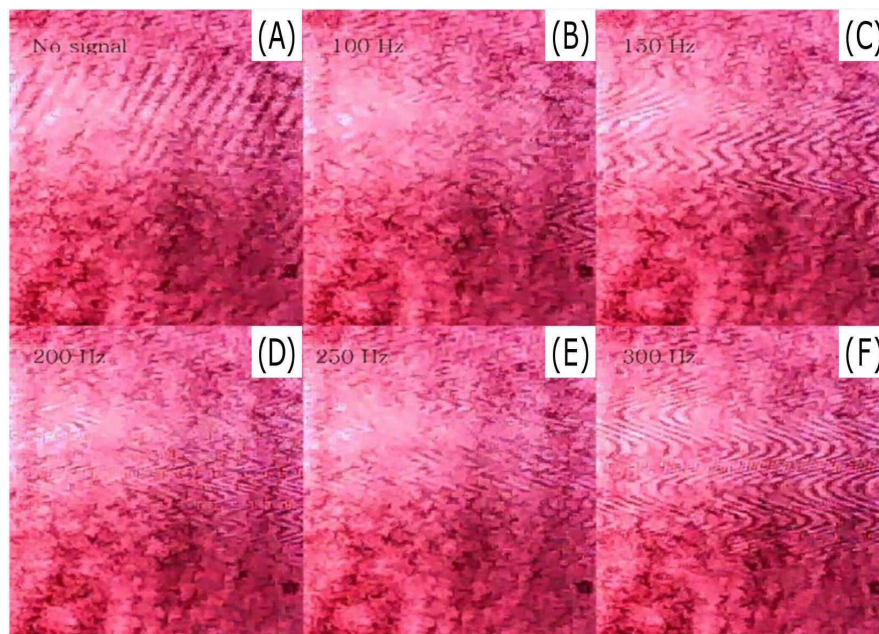
도면6



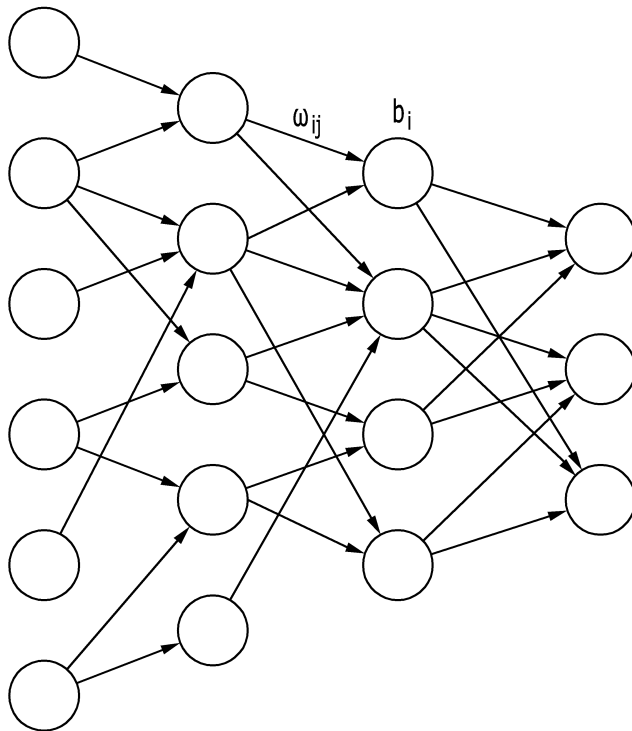
도면7



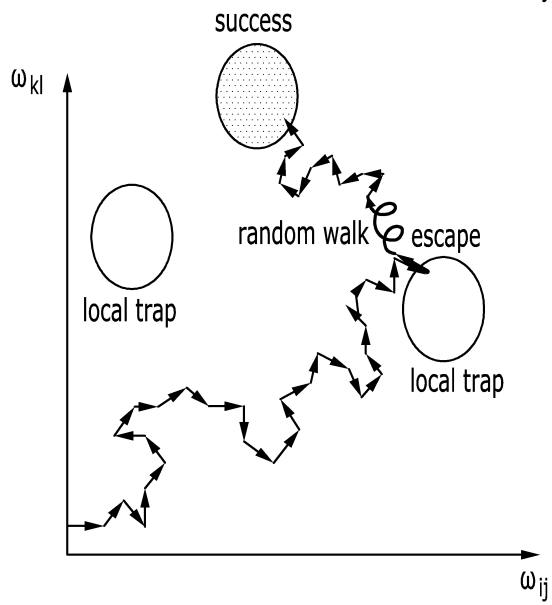
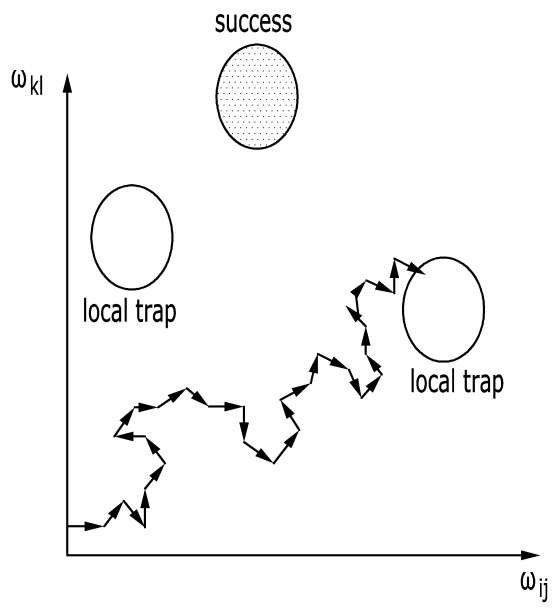
도면8



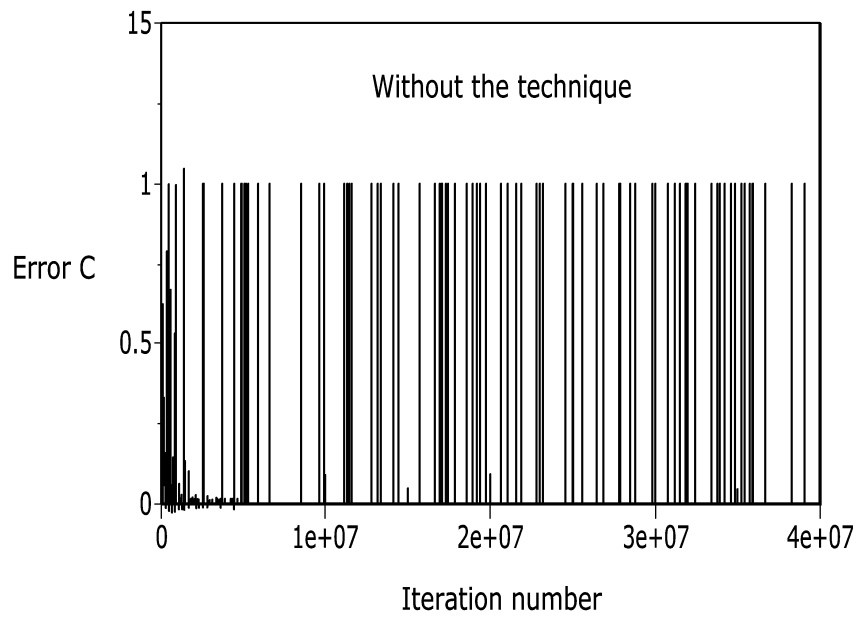
도면9



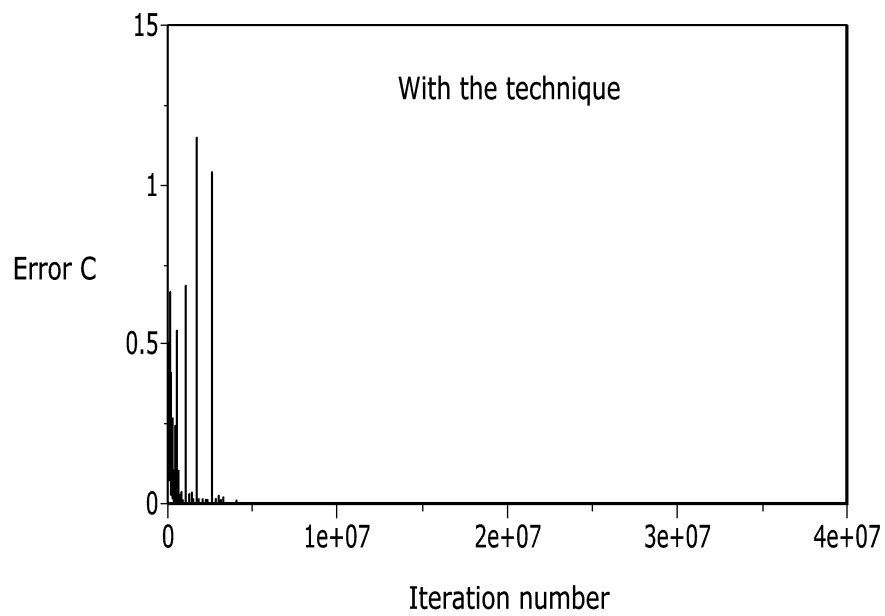
도면10



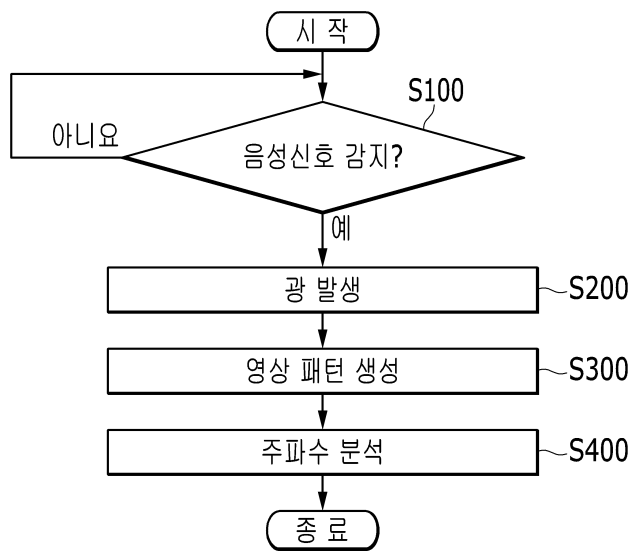
도면11



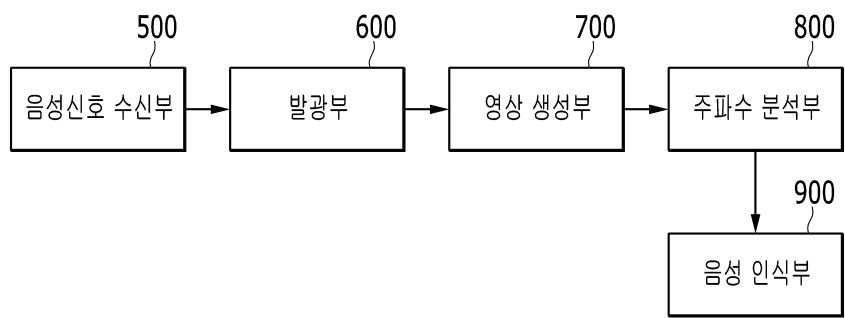
도면12



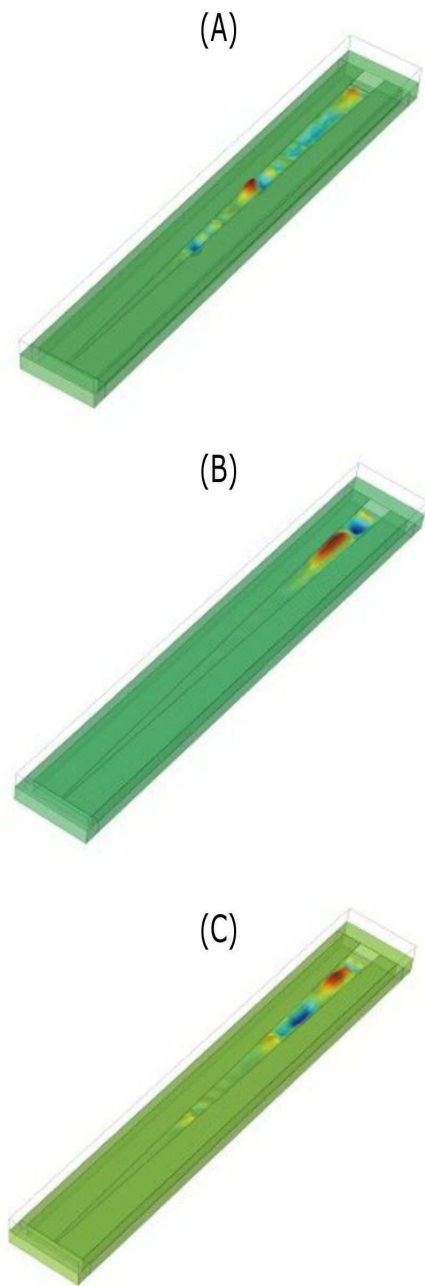
도면13



도면14



도면15



도면16

