



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0128075  
(43) 공개일자 2017년11월22일

(51) 국제특허분류(Int. Cl.)  
G06F 17/30 (2006.01) G06N 3/04 (2006.01)  
G06N 3/08 (2006.01)  
(52) CPC특허분류  
G06F 17/30743 (2013.01)  
G06N 3/04 (2013.01)  
(21) 출원번호 10-2017-0025908  
(22) 출원일자 2017년02월28일  
심사청구일자 없음

(71) 출원인  
반병현  
경상북도 안동시 안기1길 39, 102동403호(안기동, 안기동대원아파트)  
(72) 발명자  
반병현  
경상북도 안동시 안기1길 39, 102동403호(안기동, 안기동대원아파트)  
(74) 대리인  
특허법인(유한)아이시스

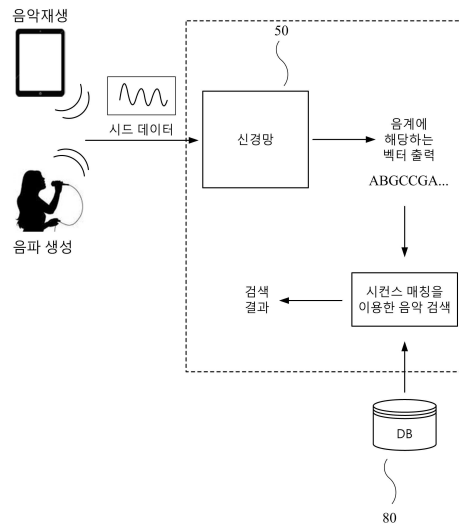
전체 청구항 수 : 총 5 항

(54) 발명의 명칭 신경망을 이용한 음악 검색 방법

### (57) 요약

신경망을 이용한 음악 검색 방법은 컴퓨터 장치가 입력받은 시드 데이터를 주파수 영역에서 일정한 구간 단위로 크로마그램(chromagram)으로 변환하는 단계, 상기 컴퓨터 장치가 크로마그램으로 변환된 데이터를 사전에 학습된 신경망에 입력하여 음계를 나타내는 복수의 벡터 정보를 출력하는 단계 및 상기 컴퓨터 장치가 상기 복수의 벡터 정보가 나타내는 일련의 멜로디와 악보 데이터베이스에 저장된 멜로디에 대한 시퀀스 매칭을 수행하여 특정 음악을 검색하는 단계를 포함한다. 신경망은 훈련 데이터인 음악의 음파를 입력받아 멜로디를 출력하고, 출력된 멜로디와 상기 신경망에 입력된 음파에 대응하는 악보를 비교하는 방식으로 사전에 강화학습을 수행한다.

대표도 - 도1



(52) CPC특허분류  
*G06N 3/084* (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

컴퓨터 장치가 입력받은 시드 데이터를 주파수 영역에서 일정한 구간 단위로 크로마그램(chromagram)으로 변환하는 단계;

상기 컴퓨터 장치가 크로마그램으로 변환된 데이터를 사전에 학습된 신경망에 입력하여 음계를 나타내는 복수의 벡터 정보를 출력하는 단계; 및

상기 컴퓨터 장치가 상기 복수의 벡터 정보가 나타내는 일련의 멜로디와 악보 데이터베이스에 저장된 멜로디에 대한 시퀀스 매칭을 수행하여 특정 음악을 검색하는 단계를 포함하되,

상기 신경망은 훈련 데이터인 음악의 음파를 입력받아 멜로디를 출력하고, 출력된 멜로디와 상기 신경망에 입력된 음파에 대응하는 악보를 비교하는 방식으로 사전에 강화학습을 수행하는 신경망을 이용한 음악 검색 방법.

#### 청구항 2

제1항에 있어서,

상기 신경망은 CNN(Convolution Neural Network), FFN(Feed-forward neural networks), RNN(Recurrent Neural Network) 및 LSTM(Long Short-Term Memory Unit)을 포함하는 신경망 그룹 중 어느 하나의 신경망인 신경망을 이용한 음악 검색 방법.

#### 청구항 3

제1항에 있어서,

상기 신경망은 상기 변환된 데이터를 한 개 이상의 컨볼루션(convolution) 계층을 사용하여 특징을 압축하고, 압축된 특징을 기준으로 상기 벡터 정보를 출력하는 신경망을 이용한 음악 검색 방법.

#### 청구항 4

제1항에 있어서,

상기 벡터 정보는 복수의 차원에 대한 불리언 벡터(Boolean vector)인 신경망을 이용한 음악 검색 방법.

#### 청구항 5

제1항에 있어서,

상기 컴퓨터 장치는 BLAST(Basic Local Alignment Search Tool)를 이용하여 상기 시퀀스 매칭을 수행하는 신경망을 이용한 음악 검색 방법.

## 발명의 설명

### 기술 분야

[0001] 이하 설명하는 기술은 신경망을 이용한 음악 검색 기법에 관한 것이다.

### 배경 기술

[0002] 종래 음악 분석 방법은 크게 심볼 분석 기법과 신호 스펙트럼 분석 기법으로 구분된다. 심볼 분석 기법은 음악을 구성하는 악보상의 음표 정보를 분석하는 것으로, 화성학이나 고전 음악에 대한 전문적 지식이 요구된다. 따라서 IT 업계에서는 주로 신호 스펙트럼 분석 기법이 사용된다. 신호 스펙트럼 분석 기법은 음악을 로우 레벨의 웨이브 형태로 분석한다. 예컨대, 인터넷 포털 기업 등에서 제공되는 음악 검색 서비스는 파형의 일치 정도를 기준으로 입력되는 음원과 DB에 저장된 음악을 비교하여 검색을 수행한다.

## 선행기술문헌

### 특허문헌

[0003] (특허문헌 0001) 한국등록특허 제10-1051803호

## 발명의 내용

### 해결하려는 과제

[0004] 현재 제공되는 음악 검색 서비스는 음악의 파형이 완전하게 일치하는 것을 전제로 한다. 따라서 현재 음악 검색 기법은 특정한 곡을 다른 사람이 연주하거나, 동일한 멜로디를 허밍(humming)으로 재생하면 해당 곡을 검색하지 못한다.

[0005] 이하 설명하는 기술은 음원의 파형을 입력받아 멜로디에 해당하는 정보를 출력하는 신경망을 이용하여 음악을 검색하는 서비스를 제공하고자 한다.

### 과제의 해결 수단

[0006] 신경망을 이용한 음악 검색 방법은 컴퓨터 장치가 입력받은 시드 데이터를 주파수 영역에서 일정한 구간 단위로 크로마그램(chromagram)으로 변환하는 단계, 상기 컴퓨터 장치가 크로마그램으로 변환된 데이터를 사전에 학습된 신경망에 입력하여 음계를 나타내는 복수의 벡터 정보를 출력하는 단계 및 상기 컴퓨터 장치가 상기 복수의 벡터 정보가 나타내는 일련의 멜로디와 악보 데이터베이스에 저장된 멜로디에 대한 시퀀스 매칭을 수행하여 특정 음악을 검색하는 단계를 포함한다. 상기 신경망은 훈련 데이터인 음악의 음파를 입력받아 멜로디를 출력하고, 출력된 멜로디와 상기 신경망에 입력된 음파에 대응하는 악보를 비교하는 방식으로 사전에 강화학습을 수행한다.

### 발명의 효과

[0007] 이하 설명하는 기술은 음악의 음파가 동일하지 않더라도 음악적 구성 요소가 동일하다면 동일한 곡을 검색할 수 있다. 나아가 이하 설명하는 기술은 음악적 구성 요소가 유사 내지 동일한 곡을 검색하여 저작권 침해 여부를 판단할 수 있다.

### 도면의 간단한 설명

[0008] 도 1은 신경망을 이용한 음악 검색 방법에 대한 과정을 도시한 예이다.

도 2는 전처리한 신호를 입력받는 신경망 구조에 관한 예이다.

### 발명을 실시하기 위한 구체적인 내용

[0009] 이하 설명하는 기술은 다양한 변경을 가할 수 있고 여러 가지 실시례를 가질 수 있는 바, 특정 실시례들을 도면에 예시하고 상세하게 설명하고자 한다. 그러나, 이는 이하 설명하는 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 이하 설명하는 기술의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0010] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 해당 구성요소들은 상기 용어들에 의해 한정되지는 않으며, 단지 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 이하 설명하는 기술의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0011] 본 명세서에서 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함한다" 등의 용어는 실시된 특징, 개수, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 의미하는 것이지, 하나 또는 그 이상의 다른 특징들이나 개수, 단계 동작 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 배제하지 않는 것으로 이해되어야 한다.

- [0012] 도면에 대한 상세한 설명을 하기에 앞서, 본 명세서에서의 구성부들에 대한 구분은 각 구성부가 담당하는 주기능 별로 구분한 것에 불과함을 명확히 하고자 한다. 즉, 이하에서 설명할 2개 이상의 구성부가 하나의 구성부로 합쳐지거나 또는 하나의 구성부가 보다 세분화된 기능별로 2개 이상으로 분화되어 구비될 수도 있다. 그리고 이하에서 설명할 구성부 각각은 자신이 담당하는 주기능 이외에도 다른 구성부가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성부 각각이 담당하는 주기능 중 일부 기능이 다른 구성부에 의해 전담되어 수행될 수도 있음은 물론이다.
- [0013] 또, 방법 또는 동작 방법을 수행함에 있어서, 상기 방법을 이루는 각 과정들은 문맥상 명백하게 특정 순서를 기재하지 않은 이상 명기된 순서와 다르게 일어날 수 있다. 즉, 각 과정들은 명기된 순서와 동일하게 일어날 수도 있고 실질적으로 동시에 수행될 수도 있으며 반대의 순서대로 수행될 수도 있다.
- [0015] 이하 설명하는 기술은 소위 딥 러닝(deep learning)이라는 기법을 이용하여 음악을 검색하는 기법에 관한 것이다. 이하 신경망을 이용하여 음악을 검색하는 장치를 컴퓨터 장치라고 명명한다. 컴퓨터 장치는 PC, 스마트 기기, 서버 등일 수 있다.
- [0017] 도 1은 신경망을 이용한 음악 검색 방법에 대한 과정을 도시한 예이다. 도 1에서 우측에 점선으로 도시한 구성 내지 과정이 컴퓨터 장치에서 수행하는 것이다. 컴퓨터 장치는 음악 검색을 위한 시드 데이터를 입력받는다. 시드 데이터는 사용자가 검색하고자 하는 음악에 대한 데이터이다. 시드 데이터는 아날로그 음악 파형에 해당한다. 예컨대, 시드 데이터는 별도의 음악 재생 장치가 특정한 음악 파일을 재생하여 스피커로 출력하는 음파일 수 있다. 또는 시드 데이터는 사용자가 음성이나 허밍 등으로 출력하는 음파일 수도 있다.
- [0018] 컴퓨터 장치는 입력된 시드 데이터를 주파수 영역으로 변환하고, 크로마그램(chromagram)과 같은 형태의 데이터로 변환한다. 컴퓨터 장치는 크로마그램 형태의 데이터를 신경망(50)에 입력한다. 신경망(50)은 음계에 해당하는 일정한 벡터를 출력한다. 복수의 벡터는 멜로디와 같은 음표를 나타낸다. 도 1에서 신경망(50)의 출력을 ABGCGA..와 같은 멜로디로 표현하였다. 이제 컴퓨터 장치는 음악 정보를 저장한 DB(80)에서 신경망(50)의 출력물과 동일 내지 유사한 정보가 있는지 검색한다. DB(80)는 특정 음악들에 대한 멜로디, 악보 등과 같은 정보를 저장할 수 있다. DB(80)는 컴퓨터 장치에 포함된 구성일 수 있다. 또는 DB(80)는 네트워크에 위치하는 서버와 같은 장치일 수도 있다. 후자의 경우 컴퓨터 장치는 통신 인터페이스를 통해 DB(80)와 통신한다.
- [0019] 컴퓨터 장치는 최종적으로 검색된 결과를 출력한다. 컴퓨터 장치는 검색 결과를 디스플레이 화면에 표시할 수 있다. 나아가 컴퓨터 장치는 검색된 음악을 재생할 수도 있다.
- [0021] 도 2는 전처리한 신호를 입력받는 신경망 구조에 관한 예이다. 컴퓨터 장치는 전술한 시드 데이터를 일정하게 전처리할 수 있다. 컴퓨터 장치는 전처리부(30)를 포함할 수 있다. 전처리부(30)는 음파에 해당하는 시드 데이터를 STFT(Short-Time Fourier Transform) 변환한다. STFT는 시간 및 위상으로 표현되는 2차원 데이터인 음파를 시간, Hz 및 강도를 갖는 3차원 데이터로 변환한다. 컴퓨터 장치는 STFT 변환된 데이터를 크로마그램 형태로 변환한다. 크로마그램을 사용하는 경우 일반적으로 에너지가 높게 표현된 구간의 음이 사용되었을 것으로 추측하고 이를 악곡 정보로 변환하는 접근이 일반적이다. 다만 음악을 구성하는 악기와 그 음색 또는 이들의 파동이 간섭함에 따라 실제로 연주되지 않은 구간의 음에도 에너지분포가 존재할 수 있다. 결과적으로 음악 파동은 (시간, 음계, 강도)로 구분되는 3차원 데이터로 변환된다. 한편 도 2에 도시하지 않았지만 크로마그램의 형태를 갖는 데이터를 일정한 시간 간격으로 신경망(50)에 전달하는 스플라이서(splicer)를 포함할 수 있다.
- [0022] 신경망(50)은 압축부(51), 학습부(52) 및 출력부(53)를 포함할 수 있다.
- [0023] 압축부(51)는 크로마그램 형태의 데이터에서 보다 뚜렷한 특징을 추출하는 구성이다. 압축부(51)는 다양한 형태로 구현될 수 있다. 예컨대, 압축부(51)는 한 층 이상의 컨볼루션 계층(convolution layer)을 활용하여 음악 정보를 압축할 수 있다. 여기서 컨볼루션 계층은 CNN(Convolution Neural Network)의 일부일 수 있다. 이때, 컨볼루션 계층은 x축에 평행한 특징들이 부각될 수 있도록 구성한다. 예를 들면, 3 x 3컨볼루션을 하며 [[0, 0, 0], [1, 1, 1], [0, 0, 0]] 계층을 적용할 수 있다. 입력된 데이터는 가로축을 따라 압축되며 동일한 음이 일정 시간 이상 지속될 경우 지속되는 부분이 더욱 부각된다. 컨볼루션 계층의 가로 축 차원이 증가할 수록 이러한 경향은 강해진다.

- [0024] 학습부(52)는 하나의 계층 이상의 은닉 계층(Hidden Layer)으로 구성된 심층신경망을 포함한다. 학습부는 CNN(Convolution Neural Network), FFN(Feed-forward neural networks), RNN(Recurrent Neural Network) 및 LSTM(Long Short-Term Memory Unit) 등 다양한 형태의 신경망 중 어느 하나일 수 있다. RNN은 이전까지의 시퀀스의 영향을 받게되므로 동일한 음표가 반복되는 경우 이를 다른 음표로 인식하는 오류를 최소화할 수 있다. RNN 중에서도 LSTM을 이용하면 우수한 성능을 부여할 수 있다.
- [0025] 출력부(530)는 학습부(52)로부터 출력받은 자료를 기반으로 음계를 구성하는 불리언 벡터(Boolean Vector)를 출력한다. 불리언 벡터는 0 또는 1만을 구성 요소로 가지는 다차원 벡터이다. 불리언 벡터는 일정한 멜로디를 나타낸다. 또는 불리언 벡터는 악보에 포함되는 정보를 의미할 수도 있다. 불리언 벡터는 음악 장르에 따라서 그 크기를 조절할 수 있다. 예컨대, 피아노로 표현 가능한 모든 음계를 악보형태로 사상시키는 인공지능을 구축하려면 88차원 벡터를 이용한다. 이 경우 눌린 건반에 해당하는 인덱스는 1, 눌리지 않은 건반에 해당하는 인덱스는 0으로 표현할 수 있다.
- [0027] 신경망(50)은 사전에 학습되어야 한다. 학습 데이터는 악보 데이터와 악보 데이터를 연주한 음악의 음파 정보를 포함할 수 있다. 예컨대, 악보 데이터는 Midi 포맷과 같은 데이터일 수 있다. 음파 정보는 Midi 파일을 곧바로 wav 포맷과 같은 파일로 디코딩한 데이터일 수 있다.
- [0028] 컴퓨터 장치는 전술한 전처리 과정을 통해 크로마그램 형태의 데이터로 학습 데이터를 마련할 수 있다. 신경망(50)은 학습 데이터 중 음파 정보를 입력받아 일련의 음계에 해당하는 불리언 벡터를 출력한다. 컴퓨터 장치는 출력된 불리언 벡터와 레이블을 비교하여 성능을 평가한다. 레이블은 최초 학습 데이터에 포함된 악보 데이터이다. 컴퓨터 장치는 불리언 벡터와 레이블을 비교하여 정확하게 일치하는 음표의 개수에 따라 점수를 가산하고, 잘못된 음표나 출력되지 않은 음표의 개수에 따라 점수를 감산하는 방식으로 강화학습한다.
- [0030] 컴퓨터 장치는 사전에 학습된 신경망을 이용하여 음악을 검색한다. 최종적인 검색 과정을 설명한다. 컴퓨터 장치는 신경망에서 출력한 불리언 벡터와 사전에 DB(80)에 보유하고 있는 음악 정보를 비교한다. DB(80)에 저장된 정보는 멜로디, 악보 등과 같은 정보에 해당한다. 일련의 불리언 벡터의 악보 내지 멜로디 라인에 해당한다. 따라서 불리언 벡터는 특정한 데이터를 갖는 시퀀스에 해당한다.
- [0031] 컴퓨터 장치는 시퀀스 매칭 알고리즘을 사용하여 불리언 벡터와 DB(80)에 저장된 정보를 비교한다. 예컨대, 컴퓨터 장치는 BLAST(Basic Local Alignment Search Tool)와 같은 시퀀스 매칭 알고리즘을 사용하여 DB(80)에 저장된 곡들과 유사한 곡들을 검색할 수 있다.
- [0032] 검색시 고려해야 할 요소로는 질의 커버리지(Query Coverage)와 호몰로지(Homology)가 있다. 질의 커버리지는 구성요소간의 포함관계를 의미하는 것으로, 입력값과 검색값이 서로 포함관계를 이루는 범위가 클수록 높은 값을 가진다. 예를 들어, 서로 간에 포함관계가 전혀 성립하지 않으면 0%이며, 서로 간에 완벽한 포함관계가 성립하면 100%가 된다. (포함 범위) / (포함 범위 + 불포함 범위)로 계산한다. 호몰로지는 구성요소간의 일치관계를 의미하는 것으로, 입력값과 검색값이 서로 포함관계는 적으나 동일한 구간이 있는 경우 유의미한 값을 출력한다. 두 대상 간 전혀 동일성이 발견되지 않는 경우 Homology는 0%가 되며 두 대상 가운데 완전히 일치하는 구간이 존재하는 경우 Homology는 100%가 된다. 두 악곡의 시점을 Slicing해가며 비교하며, 유사한 구간을 기준으로 (구간 내 일치 범위) / (구간 내 일치 범위 + 구간 내 비일치 범위)로 계산한다.
- [0033] 컴퓨터 장치는 특정한 음악을 검색하는 경우 질의 커버리지가 가장 높은 곡을 선택할 수 있다. 경우에 따라서 컴퓨터 장치는 호몰로지를 기준으로 상위에 랭크된 하나 이상의 곡들을 선택할 수도 있다. 예컨대, 특정 음악과 음악적 요소가 동일 내지 유사하여 저작권 침해가 문제되는 음악을 검색하기 위한 경우, 컴퓨터 장치는 호몰로지를 기준으로 하나 이상의 음악을 선택할 수 있다.
- [0035] 본 실시례 및 본 명세서에 첨부된 도면은 전술한 기술에 포함되는 기술적 사상의 일부를 명확하게 나타내고 있는 것에 불과하며, 전술한 기술의 명세서 및 도면에 포함된 기술적 사상의 범위 내에서 당업자가 용이하게 유추할 수 있는 변형 예와 구체적인 실시례는 모두 전술한 기술의 권리범위에 포함되는 것이 자명하다고 할 것이다.

## 부호의 설명

[0036]

30 : 전처리부

50 : 신경망

51 : 압축부

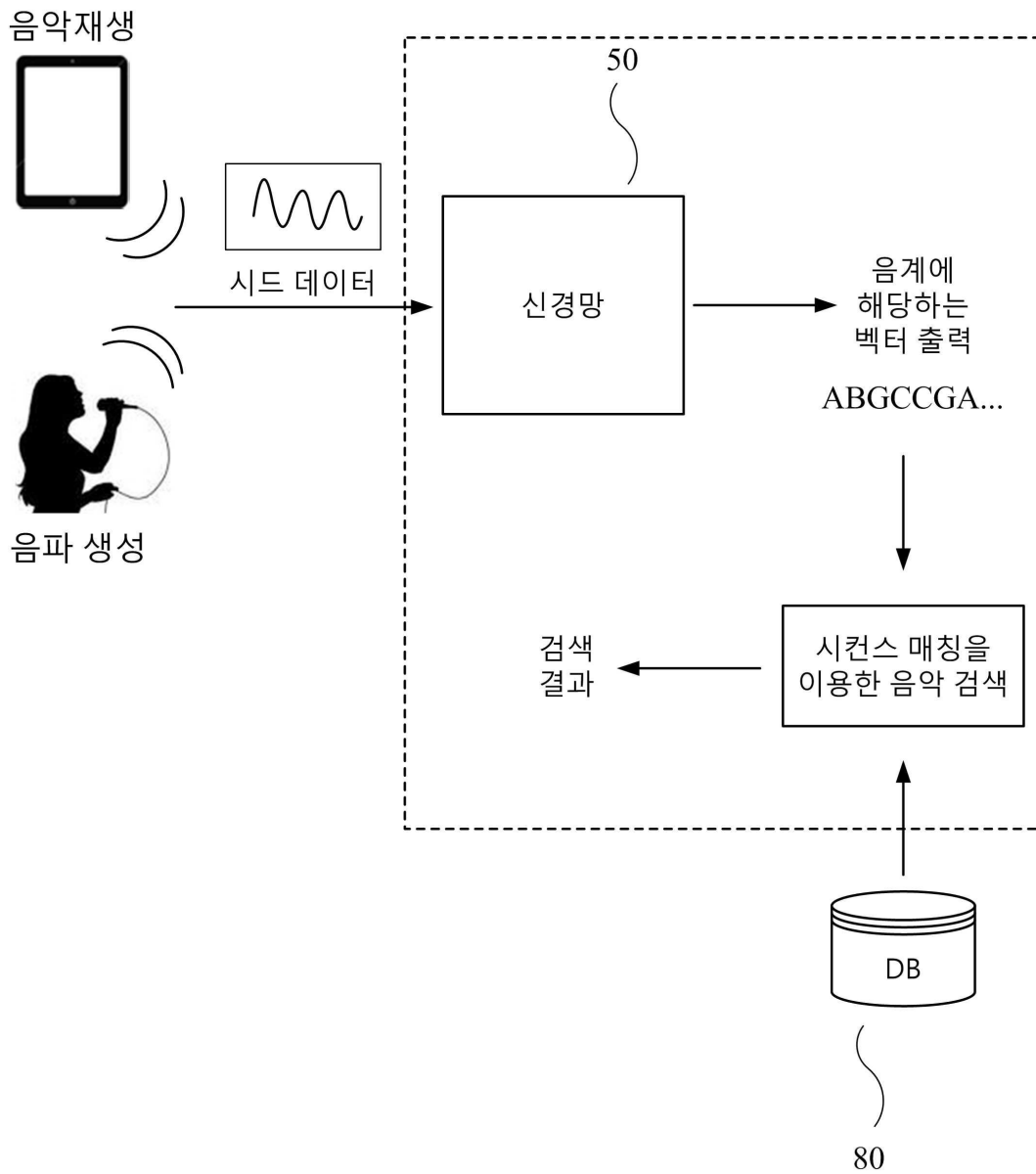
52 : 학습부

53 : 출력부

80 : DB

도면

도면1



도면2

