(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0128060
(43) 공개일자 2017년11월22일

(51) 국제특허분류(Int. Cl.)
G10H 1/00 (2006.01) *G06N 99/00* (2010.01)
(52) CPC특허분류
G10H 1/0008 (2013.01)
G06N 99/005 (2013.01)
(21) 출원번호 10-2016-0169802
(22) 출원일자 2016년12월13일
심사청구일자 없음

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

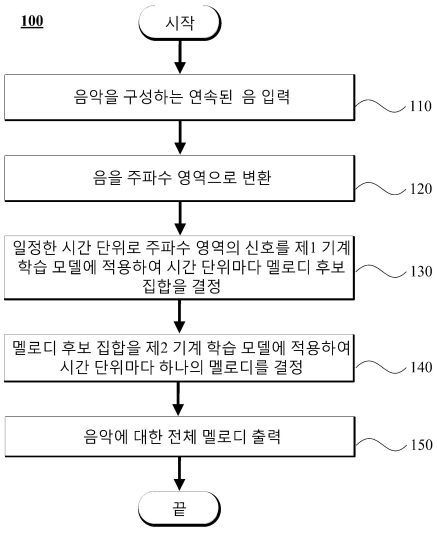
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 **재생되는 음악을 분석하여 멜로디를 추출하는 방법**

(57) 요 약

재생되는 음악을 분석하여 멜로디를 추출하는 방법은 컴퓨터 장치가 음악을 구성하는 연속된 음을 입력받는 단계, 상기 컴퓨터 장치가 상기 음을 주파수 영역으로 변환하는 단계, 상기 컴퓨터 장치가 일정한 시간 단위로 상기 주파수 영역의 음을 제1 기계 학습 모델에 적용하여 상기 시간 단위마다 복수의 멜로디를 포함하는 멜로디 후보 집합을 결정하는 단계 및 상기 컴퓨터 장치가 상기 멜로디 후보 집합을 사전에 음원에 대한 멜로디 집합으로 학습한 제2 기계 학습 모델에 적용하여 상기 시간 단위마다 하나의 멜로디를 결정하는 단계를 포함한다.

대 표 도 - 도3



(52) CPC특허분류
G10H 2210/036 (2013.01)
G10H 2210/071 (2013.01)
G10H 2210/076 (2013.01)
G10H 2240/121 (2013.01)

명 세 서

청구범위

청구항 1

컴퓨터 장치가 음악을 구성하는 연속된 음을 입력받는 단계;

상기 컴퓨터 장치가 상기 음을 주파수 영역으로 변환하는 단계;

상기 컴퓨터 장치가 일정한 시간 단위로 상기 주파수 영역의 음을 제1 기계 학습 모델에 적용하여 상기 시간 단위마다 복수의 멜로디를 포함하는 멜로디 후보 집합을 결정하는 단계; 및

상기 컴퓨터 장치가 상기 멜로디 후보 집합을 사전에 음원에 대한 멜로디 집합으로 학습한 제2 기계 학습 모델에 적용하여 상기 시간 단위마다 하나의 멜로디를 결정하는 단계를 포함하는 재생되는 음악을 분석하여 멜로디를 추출하는 방법.

청구항 2

제1항에 있어서,

상기 제1 기계 학습 모델은 KNN(K Nearest Neighbor) 알고리즘, DNN(Deep Neural Network) 또는 CNN(Convolutional Neural Network)에 기반한 모델인 재생되는 음악을 분석하여 멜로디를 추출하는 방법.

청구항 3

제1항에 있어서,

상기 제2 기계 학습 모델은 DNN(Deep Neural Network) 또는 CNN(Convolutional Neural Network)에 기반한 모델인 재생되는 음악을 분석하여 멜로디를 추출하는 방법.

청구항 4

제1항에 있어서,

상기 제1 기계 학습 모델은 복수의 상용 음원을 재생하여 발생하는 음을 주파수 영역으로 변환한 데이터를 이용하여 사전에 구축되는 재생되는 음악을 분석하여 멜로디를 추출하는 방법.

청구항 5

제1항에 있어서,

상기 컴퓨터 장치는 현재 입력되는 시간 단위에 대한 멜로디 후보 집합 및 이전에 입력된 시간 단위에 대한 멜로디 후보 집합을 상용 음원에 대한 멜로디 라인을 이용하여 화성을 학습한 상기 제2 기계 학습 모델에 적용하여 상기 현재 입력되는 시간 단위에 대한 상기 하나의 멜로디를 결정하는 재생되는 음악을 분석하여 멜로디를 추출하는 방법.

청구항 6

제1항에 있어서,

상기 멜로디 후보 집합은 음의 높이를 나타내는 정보로 구성되고, 상기 컴퓨터 장치는 상기 음에 대하여 연속하여 상기 시간 단위로 상기 정보로 표현되는 상기 하나의 멜로디를 결정하는 재생되는 음악을 분석하여 멜로디를 추출하는 방법.

청구항 7

제1항에 있어서,

상기 제1 기계 학습 모델은 CNN(Convolutional Neural Network)에 기반하고, 상기 CNN은 주파수 영역의 음을

기준으로 생성되며, 1×4 풀링(pooling) 레이어를 갖는 재생되는 음악을 분석하여 멜로디를 추출하는 방법.

발명의 설명

기 술 분 야

[0001] 이하 설명하는 기술은 재생되는 음악으로부터 멜로디 라인을 추출하는 기법에 관한 것이다.

배 경 기 술

[0002] 아날로그 신호인 음악을 분석하여 일정한 서비스를 제공하는 기술이 등장하였다. 예컨대, 스마트 기기를 이용하여 현재 재생되는 음악에 대한 정보를 찾거나, 재생되는 음악을 일정하게 분류하는 서비스 등이 있다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) 한국공개특허 제10-2011-0080554호

발명의 내용

해결하려는 과제

[0004] 이하 설명하는 기술은 재생되는 음악을 분석하여 자동으로 멜로디 라인을 추출하는 방법을 제공하고자 한다.

과제의 해결 수단

[0005] 재생되는 음악을 분석하여 멜로디를 추출하는 방법은 컴퓨터 장치가 음악을 구성하는 연속된 음을 입력받는 단계, 상기 컴퓨터 장치가 상기 음을 주파수 영역으로 변환하는 단계, 상기 컴퓨터 장치가 일정한 시간 단위로 상기 주파수 영역의 음을 제1 기계 학습 모델에 적용하여 상기 시간 단위마다 복수의 멜로디를 포함하는 멜로디 후보 집합을 결정하는 단계 및 상기 컴퓨터 장치가 상기 멜로디 후보 집합을 사전에 음원에 대한 멜로디 집합으로 학습한 제2 기계 학습 모델에 적용하여 상기 시간 단위마다 하나의 멜로디를 결정하는 단계를 포함한다.

발명의 효과

[0006] 이하 설명하는 기술은 두 단계의 기계 학습 모델을 이용하여 낮은 복잡도와 비용으로 재생되는 음악에 대한 멜로디 라인을 결정한다.

도면의 간단한 설명

[0007] 도 1은 재생되는 음원을 분석하여 멜로디를 추출하는 장치에 대한 구성을 도시한 예이다.

도 2는 재생되는 음원을 분석하여 멜로디를 추출하는 장치에 대한 구성을 도시한 다른 예이다.

도 3은 재생되는 음원을 분석하여 멜로디를 추출하는 방법에 대한 순서도의 예이다.

도 4는 재생되는 음원을 분석하여 멜로디를 추출하는 과정에 대한 예이다.

도 5는 재생되는 음원을 분석 과정에 사용되는 CNN에 대한 예이다.

발명을 실시하기 위한 구체적인 내용

[0008] 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.

[0009] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 해당 구성요소들은 상기 용어

들에 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.

[0010] 본 명세서에서 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함한다" 등의 용어는 설시된 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.

[0011] 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.

[0012] 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.

[0014] CD, MP3 등을 재생 장치로 재생하면 재생 장치는 스피커를 통해 아날로그 신호를 출력한다. 또는 현장에서 연주자가 직접 악기를 연주하거나, 가수가 노래를 해도 연속되는 음이 아날로그 신호로 청자(listener)에게 전달된다. 기본적으로 인간은 아날로그 신호를 귀로 인식하여 특정한 음을 식별한다. 이하 설명하는 기술은 재생되는 음악과 같은 아날로그 신호를 기준으로 해당 음악에 대한 멜로디를 추정하는 기법에 관한 것이다.

[0015] 음악과 관련된 용어를 간략하게 설명한다. 음 높이(pitch)는 발생하는 음의 주파수에 따라 달라진다. 음의 높이는 A, B, C, D, E, F 및 G와 같이 7도음을 주로 사용한다. 음의 높이는 음계에서 표현이 되는데 음표(note)의 위치에 따라 같은 음이라도 높이가 달라진다. 예컨대, A4 음표는 약 440Hz의 크기이고, A5는 880Hz의 크기이다.

[0016] 멜로디(melody)는 시간의 흐름에 따른 음의 높이의 배열에 해당한다. 또한 각 음의 높이는 일정한 발생 시간(duration)을 갖는다. 일정한 발생 시간을 갖는 음의 높이의 배열은 멜로디 리듬에 해당한다. 리듬(rhythm)은 어떤 음이 발생할 때 음이 얼마나 지속되는가에 따라 결정된다. 이하 설명하는 기술은 재생되는 음악에 대한 멜로디를 추출하는 기법이다.

[0017] 조성(tonality)는 음의 높이 또는 화음(chord)이 안정감(stability)과 끌림(attraction)이 느껴지도록 배열되는 체계를 의미한다. 일반적으로 음의 높이 또는 음표가 안정감이 있다면 조성이 조화(consonant)로운 것이고, 음의 높이 또는 음표가 안정적이지 않다면 조성이 조화롭지 않은(dissonant) 것이다. 불협화음은 안정감에 반하는 긴장감(tension)을 포함한다. 조화 과정(resolution)은 화음 또는 음표가 부조화에서 조화로 변경되는 과정을 의미한다. 조성학은 조화로운 조성과 조화롭지 않은 조성을 구별하는 정보를 포함한다. 후술하겠지만 기계 학습 모델은 조성학에 대한 정보를 학습하여 멜로디를 추정하는데 사용한다.

[0019] 도 1은 재생되는 음원을 분석하여 멜로디를 추출하는 장치에 대한 구성을 도시한 예이다.

[0020] 도 1(a)는 스마트폰과 같은 스마트 기기(50)를 이용하여 멜로디를 추출하는 예이다. 스마트 기기(50)는 내장된 마이크로 아날로그 음악 신호를 획득한다. 스마트 기기(50)는 아날로그 신호를 주파수 영역으로 변환하고, 주파수 영역의 신호를 기계 학습 모델에 적용하여 입력된 음악에 대한 멜로디를 결정한다. 기계 학습 모델에 대해서는 후술한다. 한편 스마트 기기(50)는 음원을 스피커로 재생하면서 동시에 마이크로 신호를 입력받아 멜로디를 결정할 수도 있다.

[0021] 도 1(a)에서 스마트 기기(50)는 마아크(51), 저장 장치(52), 연산 장치(53) 및 출력 장치(54)를 포함한다. 마이크(51)는 아날로그 음악 신호를 획득한다. 연산 장치(53)는 아날로그 음악 신호를 기계 학습 모델에 적용하여 멜로디를 결정한다. 저장 장치(52)는 기계 학습 모델에 대한 데이터, 기계 학습을 위한 훈련 데이터 등을 저장할 수 있다. 저장 장치(52)는 획득한 아날로그 음악 신호를 임시로 저장할 수 있다. 저장 장치(52)는 아날로그 음악 신호에 대응하는 멜로디에 대한 정보를 저장할 수 있다. 출력 장치(54)는 결정한 멜로디에 대한 정보를 출력할 수 있다.

[0022] 도 1(b)는 PC와 같은 장치를 이용하여 멜로디를 추출하는 예이다. 사용자는 컴퓨터(85)에 연결된 마이크(81)로 아날로그 음악 신호를 획득한다. 컴퓨터(85)는 아날로그 신호를 주파수 영역으로 변환하고, 주파수 영역의 신호를 기계 학습 모델에 적용하여 입력된 음악에 대한 멜로디를 결정한다.

[0023] 도 1(c)는 원격지에 있는 서버(98)가 멜로디를 결정하는 예이다. 컴퓨터(85)는 마이크(81)를 통해 아날로그 음악 신호를 획득한다. 컴퓨터(85)는 아날로그 음악 신호를 서버(98)에 전달할 수 있다. 이 경우 서버(98)는 컴퓨터(95)가 전달하는 아날로그 신호를 주파수 영역으로 변환하고, 주파수 영역의 신호를 기계 학습 모델에 적용하여 입력된 음악에 대한 멜로디를 결정한다.

[0024] 또는 컴퓨터(85)가 획득한 아날로그 음악 신호를 주파수 영역 신호로 변환하고, 주파수 영역의 신호를 서버(98)에 전달할 수도 있다. 이 경우 서버(98)는 수신한 주파수 영역의 신호를 기계 학습 모델에 적용하여 입력된 음악에 대한 멜로디를 결정한다.

[0026] 도 2는 재생되는 음원을 분석하여 멜로디를 추출하는 장치에 대한 구성을 도시한 다른 예이다. 도 2는 도 1(c)에서 설명한 장치를 통해 멜로디를 추출하는 과정에 대한 예이다. 컴퓨터(95)는 전술한 바와 같이 아날로그 음악 신호 또는 주파수 영역으로 변환한 신호를 네트워크를 통해 서버(98)에 전달한다. 서버(98)는 주파수 영역의 신호를 기계 학습 모델에 적용하여 일정한 멜로디를 추출한다. 모델 DB(99)는 기계 학습 모델에 대한 데이터 내지 학습을 위한 데이터를 저장할 수 있다. 서버(98)는 음악에 대응하는 멜로디 정보를 컴퓨터(95)에 전달할 수 있다. 멜로디 정보는 다양한 방식으로 표현할 수 있지만 이하 설명에서는 음의 높이를 나타내는 문자(text)로 표현한다고 가정한다. 도 2는 옥타브에 관계 없이 7도음을 기준으로 음의 높이를 나타내는 문자를 전송하는 예(1번) 및 7도음에 옥타브까지 표현한 문자를 전송하는 예(2번)를 도시한다. 2번 예에서 옥타브는 숫자로 표시하였다. 도 2에 도시하지 않았지만, 컴퓨터(95)는 수신한 멜로디 정보를 화면에 출력하거나, 수신한 멜로디 정보를 저장할 수 있다.

[0028] 설명의 편의를 위해 이하 컴퓨터 장치가 멜로디를 추출 내지 결정한다고 설명한다. 컴퓨터 장치는 전술한 스마트 기기(50), 컴퓨터(85), 서버(98) 등이 될 수 있다.

[0030] 도 3은 재생되는 음원을 분석하여 멜로디를 추출하는 방법(100)에 대한 순서도의 예이다. 컴퓨터 장치는 2 단계의 기계 학습 모델을 사용하여 재생되는 음악(음원)에 대한 멜로디를 추울한다. 먼저 컴퓨터 장치가 음악을 구성하는 연속된 음을 입력받는다(110). 컴퓨터 장치는 입력받은 음을 주파수 영역으로 변한한다(120). 예컨대, 컴퓨터 장치는 아날로그 신호에 STFT(Short-Time Fourier Transform)를 적용하여 주파수 영역으로 변환할 수 있다. 또는 컴퓨터 장치는 아날로그 신호를 Mel-spectrogram과 같은 주파수 영역의 신호로 변환할 수도 있다. 나아가 컴퓨터 장치는 다른 변환 기법을 사용하여 아날로그 신호를 주파수 영역의 신호로 변활할 수도 있다. 아날로그 신호를 주파수 영역의 신호로 변환하는 구체적인 과정에 대해서는 설명을 생략한다. 컴퓨터 장치는 연속된 음을 일정한 시간 단위로 처리한다. 일정한 시간 단위는 예컨대, 몇 초, 몇 십초와 같은 시간이 될 수 있다. 이를 위해 컴퓨터 장치가 STFT와 같은 변환을 사용할 수 있다.

[0031] 컴퓨터 장치가 일정한 시간 단위로 주파수 영역의 음(신호)을 제1 기계 학습 모델에 적용하여 시간 단위마다 가능한 모든 멜로디를 결정한다(130). 제1 기계 학습 모델은 각 시간 단위마다 가능한 멜로디의 후보를 결정한다. 따라서 컴퓨터 장치는 제1 기계 학습 모델을 통해 시간 단위마다 복수(또는 적어도 하나)의 멜로디 라인을 포함하는 집합을 얻는다. 제1 기계 학습 모델을 통해 획득하는 멜로디 집합을 이하 멜로디 후보 집합이라고 명명한다.

[0032] 컴퓨터 장치가 멜로디 후보 집합을 사전에 음원에 대한 멜로디 집합으로 학습한 제2 기계 학습 모델에 적용하여 시간 단위마다 하나의 멜로디를 결정한다(140). 컴퓨터 장치는 멜로디 후보 집합을 입력으로 사용하여 각 시간 단위마다 하나의 멜로디 라인을 추출하는 것이다. 최종적으로는 컴퓨터 장치는 시간 단위로 추출한 멜로디를 결합하여 전체 멜로디를 추출할 수 있다. 한편 컴퓨터 장치는 추출한 멜로디를 화면에 출력할 수 있다(150). 컴퓨터 장치는 추출한 멜로디에 대한 텍스트 정보를 저장할 수 있다. 컴퓨터 장치는 추출한 멜로디에 대한 텍스트 정보를 네트워크를 통해 특정 클라이언트 장치에 전달할 수도 있다.

[0034] 이하 전술한 기계 학습 모델을 중심으로 멜로디 추출 과정을 구체적으로 설명한다. 도 4는 재생되는 음원을 분석하여 멜로디를 추출하는 과정에 대한 예이다. 도 4에 도시한 과정은 모두 컴퓨터 장치에서 수행되는 것이다.

[0035] 전술한 바와 같이 컴퓨터 장치는 일정한 시간 단위로 멜로디를 추출한다. 도 4에 도시하지 않았지만 컴퓨터 장치는 아날로그 음악 신호를 일정한 시간 단위로 파싱(parsing)하거나, 주파수 영역으로 변환된 신호를 일정한 시간 단위로 파싱한다. 이후 모든 과정은 일정한 시간 단위로 수행된다. 예컨대, 컴퓨터 장치는 제1 시간 단위

에 속하는 신호를 제1 기계 학습 모델에 적용하여 제1 멜로디 후보 집합을 결정하고, 제1 멜로디 후보 집합을 제2 기계 학습 모델에 적용하여 하나의 제1 멜로디를 결정한다. 이후 컴퓨터 장치는 다른 시간 단위에 대해서도 동일한 과정을 반복한다. 컴퓨터 장치는 연속되는 시간 단위 순서로 멜로디를 추정할 수 있다. 또는 컴퓨터 장치는 병렬적으로 복수의 시간 단위에 대한 작업을 수행할 수도 있다. 최종적으로 컴퓨터 장치는 시간 흐름에 따라 각 시간 단위에 대한 멜로디 라인을 결합하여 전체 멜로디 라인을 추출한다.

[0036] 도 4에서는 전술한 제1 기계 학습 모델을 제1 AI 모델(210)이라고 표시하였고, 제2 기계 학습 모델을 제2 AI 모델(220)이라고 표시하였다. 컴퓨터 장치가 인공지능(AI)를 이용하여 아날로그 음악 신호에서 멜로디를 추출한다는 의미이다.

[0037] 제1 AI 모델(210)은 주파수 영역 신호를 입력받아 멜로디 후보 집합을 생성한다. 제1 훈련 데이터(215)는 제1 AI 모델(210)이 멜로디 후보 집합을 결정하기 위한 학습에 사용된다.

[0038] 주파수 영역으로 변환된 음파는 Hz축 및 Db 축 상에서 표현될 수 있다. 시간까지 고려하면 음악은 3차원 데이터가 된다. 제1 AI 모델(210)은 3차원 데이터를 압축하여 시간 단위로 가능한 화성의 집합으로 분류하는 일종의 분류기(classifier)에 해당한다.

[0039] 제1 AI 모델(210)은 기계 학습 모델을 이용하여 마련된다. 예컨대, 기계 학습 모델은 KNN(K Nearest Neighbor) 알고리즘, DNN(Deep Neural Network) 또는 CNN(Convolutional Neural Network) 등이 사용될 수 있다. 각 모델을 간략하게 설명한다.

[0040] 훈련 데이터는 화성학적으로 조화로운 음악을 이용하는 것이 바람직하다. 훈련 데이터인 음악에는 해당 음악에 대한 멜로디가 태깅된다. 멜로디는 코드를 나타내는 기호(C, D, C7 등) 및 옥타브를 나타내는 정보(숫자 등)로 표현될 수 있다. 훈련 데이터는 하나 이상의 악기로 연주한 음파를 포함할 수 있다. 몇 가지 알고리즘에 대해 설명한다. 물론 제1 AI 모델(210)은 다른 알고리즘에 기반하여 생성될 수도 있다.

[0042] KNN에 기반한 모델

[0044] KNN은 입력이 특징 공간 내 k개의 가장 가까운 훈련 데이터로 구성되어 있다. KNN은 k개의 최근접 이웃 사이에서 가장 공통적인 항목에 할당되는 항목으로 최종 결과를 도출한다. 한편 서포트 벡터 머신, 인공 신경망, 나이브 베이즈 분류방법 등 레이블이 부착된 대량의 데이터를 학습하여 분류 인공지능을 구축할 수 있는 알고리즘은 다양하다.

[0045] 제1 AI 모델(210)은 총 n종류의 레이블(lable)이 부착된 데이터들을 학습하여 n개의 클러스터를 분류한다. 이후 제1 AI 모델(210)은 입력된 데이터가 사전에 분류한 클러스터 중 어느 클러스터에 배정되면 좋은지 결정한다. 결정/예측하는 방법론은 기계학습 영역 안에서도 지도학습이라는 영역에 속한다. 지도학습과정을 통하여 훈련 데이터를 학습한 제1 AI 모델(210)은 추가로 입력된 데이터가 어떤 종류의 레이블에 부착될 수 있는지 표식하는 분류(Classification)작업을 수행할 수 있다.

[0046] KNN은 새로운 샘플이 들어오면 그 샘플과 가장 거리가 가까운 샘플들을 찾고, 그 샘플들 중에서 가장 우세한(dominant) 샘플을 기준으로 그룹을 만드는 비모수(Non-Parametric) 기법이다. 비모수 방법은 지도학습과 결합하면 비선형적인 개별 데이터들을 처리하는 데 유용하게 사용할 수 있다. KNN은 클러스터링을 한 뒤, 비모수 방법으로 유추된 클러스터들을 실제 부착된 레이블을 기준으로 평가하여 실루엣(silhouette) 수치를 계산한다. 이때 각 클러스터들이 레이블을 기준으로 잘 군집화되었다면 실루엣 값이 1에 가깝게 평가될 것이다. 제1 AI 모델(210)은 KNN을 실행하고, 실루엣 값과 1 사이의 차이를 비용 함수로 둔다.

[0047] 대량의 음악 데이터는 고차원 데이터이다. 데이터 압축을 위해 전처리 과정으로 주성분 분석(Principal Component Analysis: PCA)을 하여 차원 축소를 할 수 있다. 이 경우 제1 AI 모델(210)은 어떤 주성분(PC)의 조합을 기준으로 클러스터링 한 경우 비용 함수가 최소가 되는지 결정하고, 비용 함수가 최소가 되는 주성분 조합을 저장한다. 이후 입력 데이터가 들어올 경우 제1 AI 모델(210)은 해당 주성분의 조합에 대응하도록 입력 데이터를 선형변환하고, 변환된 데이터를 기존 클러스터에 대응하여 유클리드 거리를 기준으로 분류한다. 제1 AI 모델(210)은 입력된 테스트 데이터가 배정될 가능성이 높은 클러스터들을 추출하여, p-value 0.05 이하의 유의미한 클러스터 목록을 출력한다.

[0049] DNN에 기반한 모델

[0050] DNN은 입력층과 출력층 사이에 여러 개의 은닉층들로 이뤄진 인공신경망이다. DNN은 일반적인 인공신경망과 마

찬가지로 복잡한 비선형 관계(non-linear relationship)들을 모델링할 수 있다.

[0051] DNN을 사용하는 경우 부착된 레이블 중 옥타브 정보는 무시하며, 코드 정보만을 취하여 원 핫 벡터(one-hot vector)로 구성할 수 있다. 여기서 코드 정보는 멜로디에서 옥타브를 배제하고 음의 높이만을 나타내는 정보를 의미한다. DNN의 1층은 입력 데이터의 차원 수와 동일하게 구현하며, 마지막 층의 차원은 레이블의 종류인 n과 같도록 구성한다. 역전파 방법으로는 경사하강법을 사용하여 구현한다.

[0053] CNN에 기반한 모델

[0054] CNN은 최소한의 전처리를 사용하도록 설계된 다계층 퍼셉트론(multilayer perceptrons)의 한 종류이다. CNN은 하나 또는 여러개의 합성곱 계층과 그 위에 올려진 일반적인 인공 신경망 계층들로 이루어져 있으며, 가중치와 통합 계층(pooling layer)들을 추가로 활용한다. 이러한 구조 덕분에 CNN은 2차원 구조의 입력 데이터를 충분히 활용할 수 있다.

[0055] CNN은 옥타브 정보를 무시하는 것이 오버피팅 문제를 회피하기 위하여 꼭 필요한 작업이다. 레이블은 n차원 벡터로 사상되며, 이 벡터는 구성요소의 총 합이 1이다. 입력데이터가 하나의 레이블로 사상될 경우 해당 클러스터에 해당하는 파라미터 값이 1이 되며, 동시에 여러 개의 레이블에 사상될 경우 학습데이터 내부에서의 누적 회수를 각각 레이블에 해당하는 파라미터에 할당하며, 할당이 끝나면 모든 레이블의 총 합이 1이 되도록 평준화한다.

[0056] 예컨대, 입력 데이터가 625차원인 경우 컨볼루션 계층(convolution layer)와 풀링 계층(pooling layer)를 3번씩 겹쳐 배치하고, 나온 결과물을 완전 연결 레이어(fully connected neural network)에 연결하여 최종적으로 n개의 뉴런을 통하여 출력값을 취한다. 이 경우 풀링 레이어는 입력 자료를 정사각형 형태가 아니라 좌우로 뻗은 직사각형 형태로 취한다. 그 이유는 음파 데이터는 그 형상이나 모양 또는 색채나 그의 결합이 중요한 영상자료와 달리 푸리에변환을 통하여 주파수 영역으로 옮겨간 가상의 값이며, hz 축을 따라 증가하는 db값의 비교가 음파 자료의 분석에 있어 핵심적이기 때문이다. 따라서 주파수 영역의 신호에서 멜로디를 추출하기 위하여 1 × 4 풀링 레이어를 배치한다. 참고로 일반적인 영상자료에서 1/4로 차원을 풀링할 때 마다 2x2 풀링 레이어를 배치한다.

[0057] 도 5는 재생되는 음원을 분석 과정에 사용되는 CNN에 대한 예이다. 아날로그 음악 신호를 주파수 영역의 신호로 변환한 신호를 입력값을 갖는다. 도 5에서는 모두 N개의 컨볼루션 계층과 N개의 풀링 계층을 도시한다. 한편 전술한 바와 같이 풀링 계층은 1 × 4 형태를 갖는다.

[0059] 도 4에 대한 설명으로 돌아간다. 제1 AI 모델(210)은 단위 시간 동안의 주파수 영역 신호를 입력으로 멜로디 후보 집합을 결정한다. 도 4에서는 모두 3개의 멜로디 후보를 도출한 예를 도시하였다.

[0060] 제2 AI 모델(220)은 멜로디 후보 집합을 입력 받아 하나의 멜로리를 추출한다. 제2 훈련 데이터(225)는 제2 AI 모델(220)이 멜로디 후보 집합을 결정하기 위한 학습에 사용된다. 제2 훈련 데이터(225)는 음악 파동이 일정 시간 단위로 분할된 자료다. 도 4에서 제2 AI 모델(210)은 최종적으로 해당 단위 시간에 대해 하나의 멜로디를 결정한다.

[0061] 제2 AI 모델(212)도 제1 AI 모델(210)과 같이 다양한 알고리즘으로 마련될 수 있다. 설명의 편의를 위해 은닉 마르코프 모델(HMM)을 기준으로 설명한다. 즉 CNN과 같은 모델이라고 가정한다. 제2 AI 모델(212)은 훈련 데이터를 통해 화성학에 대한 학습을 수행한 모델이다. 제2 AI 모델(212)은 멜로디에 해당하는 문자열을 입력으로 받는다. 제2 AI 모델(212)은 사용자의 용도에 따라 구성된 랭크 n개의 문자를 기준으로 그 다음에 올 문자의 종류와 그 확률을 학습하는 인공지능이다. 랭크 n은 기본값으로 1 이상의 정수를 가지며, 매 반복마다 n을 1씩 증가시키며 비용 함수가 가장 최소가 되는 n을 찾아가는 과정을 거친다. 비용 함수는 해당 모델이 기존 트레이닝 데이터의 도입부를 테스트 데이터로 입력받은 뒤 제작한 결과물과 기존 데이터와의 차이로 정의된다. 음악의 장르에 따라 별개의 2차 인공지능을 구축하면 보다 예측 정확도가 가장 높아진다.

[0062] 멜로디 후보 집합은 제2 AI 모델(212)에 입력되어, 매 시간 단위마다 존재하는 다양한 코드들 중 가장 2차 인공지능 내의 확률적 오토마타와 부합하는 선택지를 순차적으로 하나씩 선택한다. 제2 AI 모델(212)은 최종적으로 시간의 흐름에 따라 나열된 단 하나의 코드 진행 문자열을 출력한다.

[0064] 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

[0065] 50 : 스마트 기기

51 : 마이크

52 : 저장 장치

53 : 연산 장치

54 : 출력 장치

81 : 마이크

85 : 컴퓨터

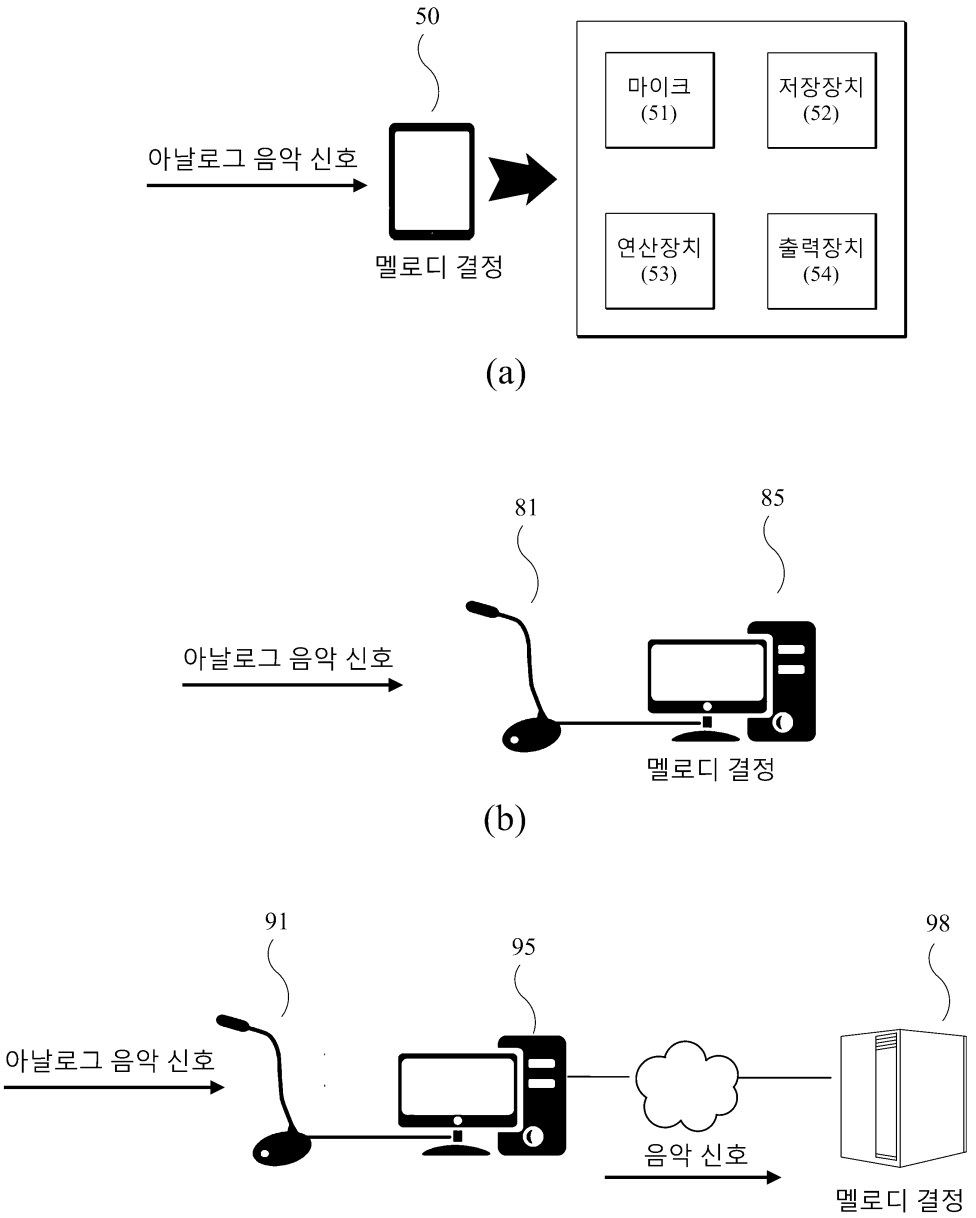
91 : 마이크

95 : 컴퓨터

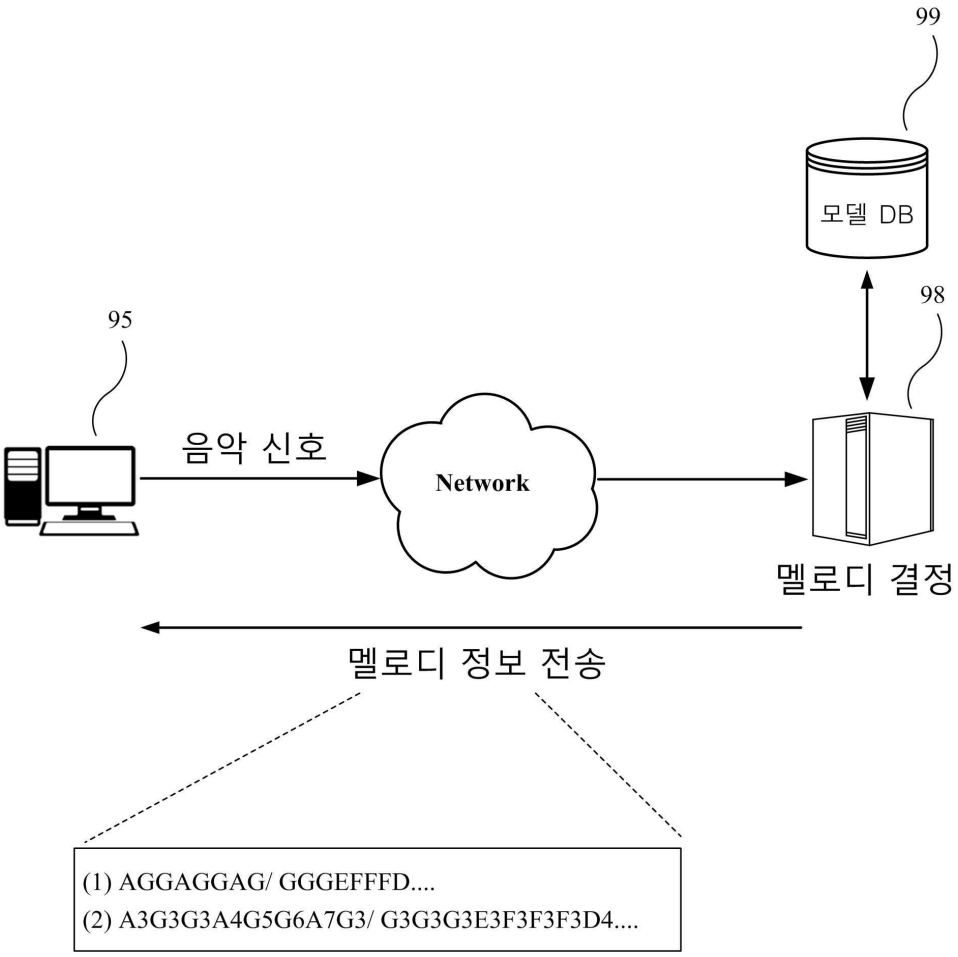
98 : 서버

도면

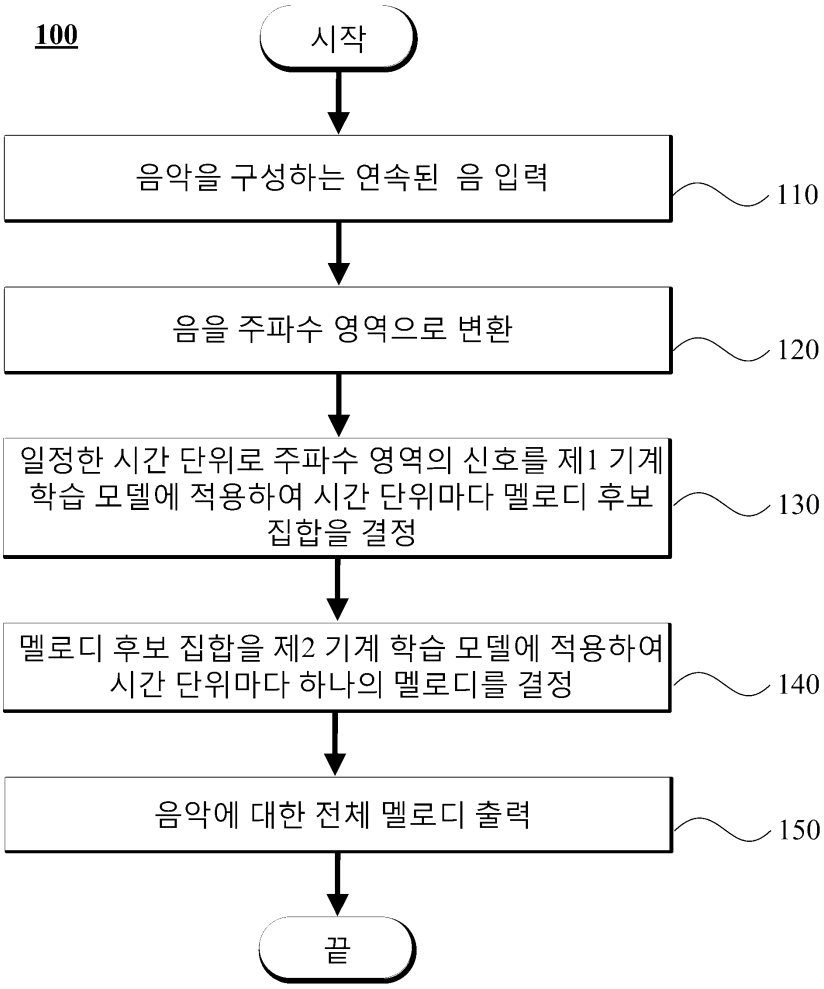
도면1


50
아날로그 음악 신호
멜로디 결정
마이크
(51)
저장장치
(52)
연산장치
(53)
출력장치
(54)
(a)
81
85
아날로그 음악 신호
멜로디 결정
(b)
91
95
98
아날로그 음악 신호
음악 신호
멜로디 결정
(c)

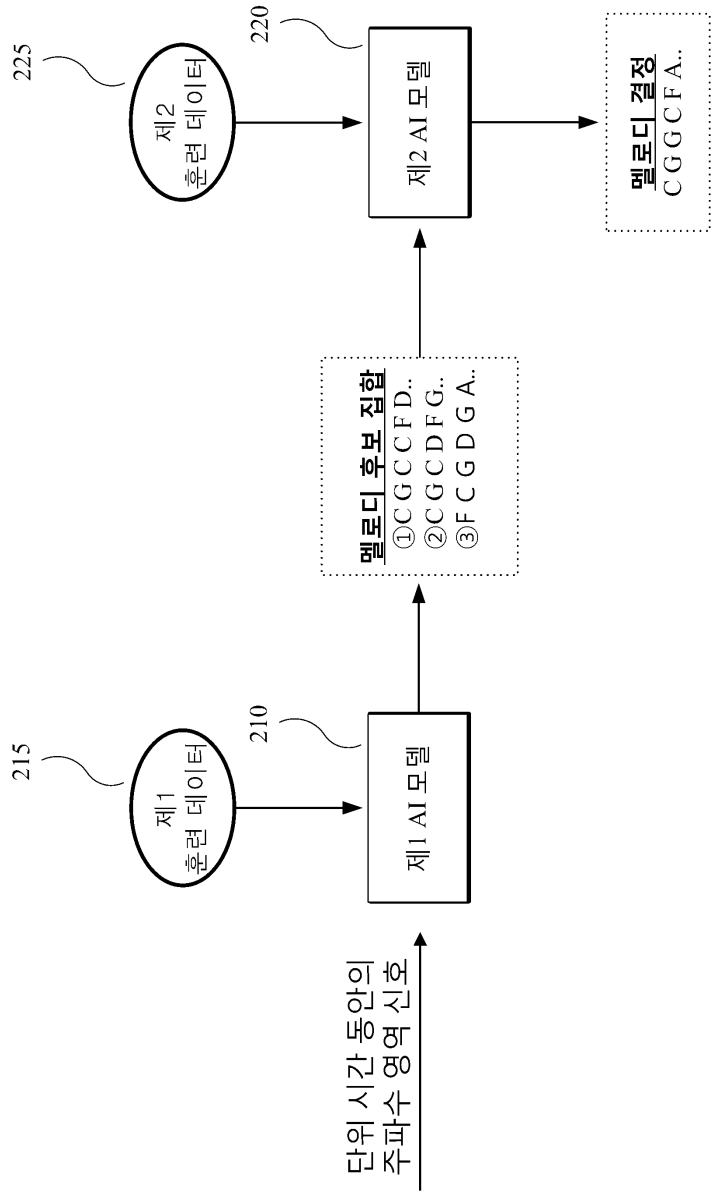
도면2


99
모델 DB
98
95
음악 신호
Network
멜로디 결정
멜로디 정보 전송
(1) AGGAGGAG/ GGGEFFFD....
(2) A3G3G3A4G5G6A7G3/ G3G3G3E3F3F3F3D4....

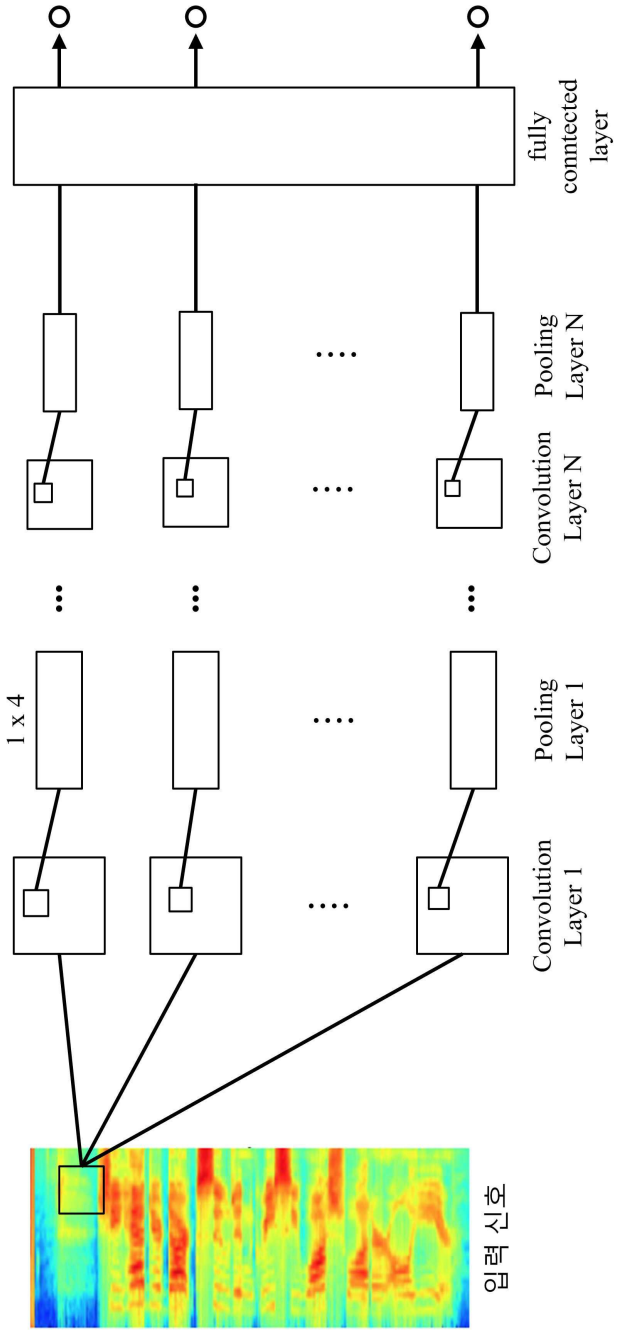
도면3


100
시작
음악을 구성하는 연속된 음 입력
110
음을 주파수 영역으로 변환
120
일정한 시간 단위로 주파수 영역의 신호를 제1 기계 학습 모델에 적용하여 시간 단위마다 멜로디 후보 집합을 결정
130
멜로디 후보 집합을 제2 기계 학습 모델에 적용하여 시간 단위마다 하나의 멜로디를 결정
140
음악에 대한 전체 멜로디 출력
150
끝

도면4


단위 시간 동안의
주파수 영역 신호
제1 AI 모델
210
제1
훈련 데이터
215
멜로디 후보 집합
①C G C C F D..
②C G C D F G..
③F C G D G A..
제2 AI 모델
220
제2
훈련 데이터
225
멜로디 결정
C G G C F A..

도면5


fully
conntected
layer
Pooling
Layer N
Convolution
Layer N
Pooling
Layer 1
Convolution
Layer 1
1 x 4
입력 신호