



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0107451
(43) 공개일자 2013년10월02일

(51) 국제특허분류(Int. Cl.)

G10K 15/00 (2006.01) G11B 27/02 (2006.01)

G11B 20/10 (2006.01) G11B 31/02 (2006.01)

(21) 출원번호 10-2012-0029237

(22) 출원일자 2012년03월22일

심사청구일자 2012년03월22일

(71) 출원인

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

(72) 발명자

박형민

경기도 화성시 반송동 나루마을 월드반도아파트 638동 902호

정일영

경기도 용인시 수지구 고기동 244번지

(74) 대리인

이지연

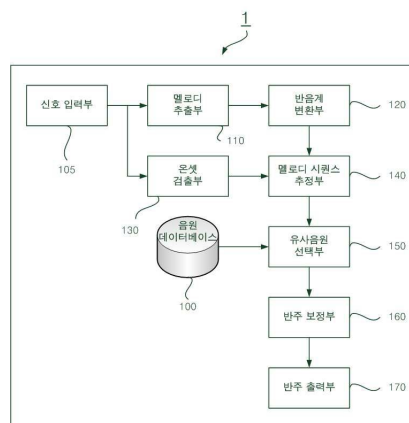
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 사운드 신호를 기반으로 한 반주제공방법 및 이를 이용한 반주제공장치

(57) 요약

본 발명은 외부의 사운드 신호를 기반으로 한 반주 제공 장치에 관한 것이다. 상기 반주 제공 장치는 외부의 입력 장치로부터 사운드 신호가 입력되는 신호 입력부; 다수 개의 음원들에 대한 정보가 저장된 음원 데이터베이스; 상기 신호 입력부로부터 사운드 신호가 제공되면, 사전 설정된 시간 간격의 프레임으로 분할하고 각 프레임에 대한 멜로디를 추출하는 멜로디 추출부; 상기 멜로디 추출부에 의해 추출된 각 프레임에 대한 멜로디를 반음계로 변환시키는 반음계 변환부; 상기 신호 입력부로부터 사운드 신호가 제공되면, 각 프레임에 대한 온셋 정보를 검출하여 제공하는 온셋 검출부; 상기 반음계 변환부로부터 각 프레임에 대한 반음 정보를 수신하고, 상기 온셋 검출부로부터 각 프레임에 대한 온셋 정보를 수신하며, 상기 온셋 정보를 이용하여 사운드 신호의 시작 프레임을 검출하고, 상기 사운드 신호의 시작 프레임부터 일정시간동안 기본 주파수에 해당하는 멜로디 시퀀스를 추정하여 제공하는 멜로디 시퀀스 추정부; 상기 멜로디 시퀀스 추정부에 의해 추정된 멜로디 시퀀스와 상기 음원 데이터베이스의 음원들의 멜로디 시퀀스를 비교하여 상기 사운드 신호의 멜로디 시퀀스와 가장 유사한 멜로디 시퀀스를 갖는 유사 음원을 검색하는 음원 선택부; 상기 사운드 신호의 음정과 박자에 따라 유사 음원에 대한 반주를 보정하여 출력하는 반주 보정부와 반주 출력부;를 구비한다.

대표도 - 도1



특허청구의 범위

청구항 1

외부의 입력 장치로부터 사운드 신호가 입력되는 신호 입력부;

다수 개의 음원들에 대한 정보가 저장된 음원 데이터베이스;

상기 신호 입력부로부터 사운드 신호가 제공되면, 사전 설정된 시간 간격의 프레임으로 분할하고 각 프레임에 대한 멜로디를 추출하는 멜로디 추출부;

상기 멜로디 추출부에 의해 추출된 각 프레임에 대한 멜로디를 반음계로 변환시키는 반음계 변환부;

상기 신호 입력부로부터 사운드 신호가 제공되면, 각 프레임에 대한 온셋 정보를 검출하여 제공하는 온셋 검출부;

상기 반음계 변환부로부터 각 프레임에 대한 반음 정보를 수신하고, 상기 온셋 검출부로부터 각 프레임에 대한 온셋 정보를 수신하며, 상기 온셋 정보를 이용하여 사운드 신호의 시작 프레임을 검출하고, 상기 사운드 신호의 시작 프레임부터 일정시간동안 기본 주파수에 해당하는 멜로디 시퀀스를 추정하여 제공하는 멜로디 시퀀스 추정부;

상기 멜로디 시퀀스 추정부에 의해 추정된 멜로디 시퀀스와 상기 음원 데이터베이스의 음원들의 멜로디 시퀀스를 비교하여 상기 사운드 신호의 멜로디 시퀀스와 가장 유사한 멜로디 시퀀스를 갖는 유사 음원을 검색하는 음원 선택부;

상기 유사 음원에 대한 반주를 상기 음원 데이터베이스로부터 검색하여 출력하는 반주 출력부;

를 구비하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 장치.

청구항 2

제1항에 있어서, 상기 반주 제공 장치는 상기 유사 음원에 대한 반주의 음정을 보정하는 반주 보정부를 더 구비하고,

상기 반주 보정부는 사운드 신호의 멜로디 시퀀스의 평균과 상기 유사 음원의 멜로디 시퀀스의 평균의 차이를 이용하여 상기 유사 음원에 대한 반주의 음정을 보정하는 것을 특징으로 하며,

상기 반주 출력부는 상기 반주 보정부에 의해 보정된 음정을 갖는 반주를 출력하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 장치.

청구항 3

제1항에 있어서, 상기 반주 제공 장치는 상기 유사 음원에 대한 반주의 빠르기를 보정하는 반주 보정부를 더 구비하고,

상기 반주 보정부는 상기 음원 선택부가 수행한 동적 타임 워핑(Dynamic Time Warping)에서 획득한 경로의 기울기 정보를 이용하여 상기 유사 음원에 대한 반주의 빠르기를 보정하는 것을 특징으로 하며,

상기 반주 출력부는 상기 반주 보정부에 의해 보정된 빠르기를 갖는 반주를 출력하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 장치.

청구항 4

제1항에 있어서, 상기 음원 데이터베이스는 다수 개의 음원들에 대한 멜로디 시퀀스 및 반주 정보를 구비하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 장치.

청구항 5

제1항에 있어서, 상기 멜로디 추출부는 각 프레임의 신호들에 대한 자기 상관 계수($R(\tau)$)를 구하여 제공하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 장치.

청구항 6

제1항에 있어서, 상기 반응계 변환부는 상기 멜로디 추출부로부터 각 프레임의 신호들에 대한 자기상관계수들을 수신하고, 상기 자기상관계수들을 양자화하고 반응계의 에너지값으로 변환하여 제공하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 장치.

청구항 7

제1항에 있어서, 상기 온셋 검출부는 상기 신호 입력부로부터 입력된 각 프레임에 대하여 이전과 현재의 다수개의 프레임의 에너지들의 가중치를 둔 합을 계산하고, 각 프레임에 대해 상기 계산된 합이 사전 설정된 문턱값보다 큰 경우 온셋으로 판단하고, 온셋으로 판단된 시간 정보 또는 프레임 정보를 멜로디 시퀀스 추정부로 제공하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 장치.

청구항 8

제1항에 있어서, 상기 멜로디 시퀀스 추정부는 상기 온셋 검출부로부터 제공된 첫번째 온셋으로 판단된 시점으로부터 사전 설정된 일정 시간 동안의 사운드 신호에 대해서 멜로디 시퀀스를 추정하여 제공하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 장치.

청구항 9

제1항에 있어서, 상기 멜로디 시퀀스 추정부는 동적 계획법(dynamic programming)을 수행하여 멜로디 시퀀스를 추정하여 제공하며,

상기 동적 계획법은 상기 반응계 변환부로부터 각 프레임의 각 반응에 대한 양자화된 에너지값들을 수신하고, 상기 수신된 에너지값들을 이용하여 각 프레임의 모든 반응에 대하여 각각 현재 프레임 및 이전 프레임의 각 반응의 에너지 및 반응의 차이를 고려한 스코어(score)를 계산하고, 마지막 프레임에서 가장 큰 스코어를 갖는 반응을 기본 주파수로 결정하고, 상기 마지막 프레임의 가장 큰 스코어를 갖는 반응까지 오는 프레임간의 경로를 역추적하여 멜로디 시퀀스를 결정하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 장치.

청구항 10

제9항에 있어서, 상기 멜로디 시퀀스 추정부는 상기 온셋 검출부로부터 제공되는 온셋으로 판단된 시간 정보 또는 프레임 정보에 따라 온셋 구간을 나누고, 각 단위 온셋 구간별로 동적 계획법(dynamic programming)을 수행하여 멜로디 시퀀스를 추정하여 제공하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 장치.

청구항 11

제1항에 있어서, 상기 음원 선택부는 상기 멜로디 시퀀스 추정부로부터 제공된 멜로디 시퀀스를 정규화시킨 후, 상기 정규화된 멜로디 시퀀스와 상기 음원 데이터베이스의 각 음원들의 멜로디 시퀀스에 대하여 동적 타임 워핑(Dynamic Time Warping)을 수행하여 사운드 신호의 프레임에 대응하는 상기 음원들의 프레임을 추정하고, 서로 대응되는 사운드 신호의 프레임들과 음원의 프레임들의 멜로디 차이를 검출하고, 상기 멜로디 차이가 가장 작은 음원을 유사 음원으로 선택하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 장치.

청구항 12

다수 개의 음원들에 대한 멜로디 시퀀스 및 반주 정보가 저장된 음원 데이터베이스를 이용하여, 외부의 입력장치로부터 입력된 사운드 신호를 기반으로 하여 반주를 제공하는 방법에 있어서,

- (a) 외부의 입력 장치로부터 사운드 신호를 입력받는 단계;
- (b) 상기 신호 입력부로부터 사운드 신호가 제공되면, 사전 설정된 시간 간격의 프레임으로 분할하고 각 프레임에 대한 멜로디를 추출하는 단계;
- (c) 상기 추출된 각 프레임에 대한 멜로디를 반응계로 변환시켜 각 프레임의 각 반응에 대한 에너지를 구하는 단계;
- (d) 상기 신호 입력부로부터 사운드 신호가 제공되면, 각 프레임에 대한 온셋 정보를 검출하는 단계;

(e) 상기 검출된 온셋 정보를 이용하여 사운드 신호의 시작 프레임을 검출하고, 각 프레임의 각 반응에 대한 에너지들로부터 기본 주파수를 검출하고, 상기 사운드 신호의 시작 프레임부터 일정시간동안 기본 주파수에 해당하는 멜로디 시퀀스를 추정하는 단계;

(f) 상기 추정된 멜로디 시퀀스와 상기 음원 데이터베이스의 음원들의 멜로디 시퀀스를 비교하여 상기 사운드 신호의 멜로디 시퀀스와 가장 유사한 멜로디 시퀀스를 갖는 음원을 검색하여 유사 음원을 선택하는 단계;

(g) 상기 유사 음원에 대한 반주를 상기 음원 데이터베이스로부터 검색하여 출력하는 단계;

를 구비하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 방법.

청구항 13

제12항에 있어서, 상기 (g) 단계는 상기 유사 음원에 대한 반주를 상기 음원 데이터베이스로부터 검색한 후, 사운드 신호의 멜로디 시퀀스의 평균과 상기 유사 음원의 멜로디 시퀀스의 평균의 차이를 이용하여 상기 유사 음원에 대한 반주의 음정을 보정하고, 상기 보정된 음정을 갖는 반주를 출력하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 방법.

청구항 14

제12항에 있어서, 상기 (g) 단계는 상기 유사 음원에 대한 반주를 상기 음원 데이터베이스로부터 검색한 후, 상기 음원 선택부가 수행한 동적 타임 워핑(Dynamic Time Warping)에서 획득한 경로의 기울기 정보를 이용하여 상기 검색된 음원에 대한 반주의 빠르기를 보정하고, 상기 보정된 빠르기를 갖는 반주를 출력하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 방법.

청구항 15

제12항에 있어서, 상기 (b)단계에서 각 프레임의 신호들에 대한 멜로디는 각 프레임의 신호들에 대한 자기 상관 계수($R(\tau)$)로 제공되는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 방법.

청구항 16

제12항에 있어서, 상기 (c)단계는 각 프레임의 신호들에 대한 자기상관계수들을 반응계의 에너지로 변환하여 양자화하여 제공하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 방법.

청구항 17

제12항에 있어서, 상기 (d) 단계는 상기 사운드 신호의 각 프레임에 대하여 이전과 현재의 다수개의 프레임의 에너지들의 가중치를 둔 합을 계산하고, 각 프레임에 대해 상기 계산된 합이 사전 설정된 문턱값보다 큰 경우 온셋으로 판단하고, 온셋으로 판단된 시간 정보 또는 프레임 정보를 제공하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 방법.

청구항 18

제12항에 있어서, 상기 (e) 단계는 동적 계획법(dynamic programming)을 수행하여 멜로디 시퀀스를 추정하여 제공하며,

상기 동적 계획법은 각 프레임의 각 반응에 대한 양자화된 에너지값들을 수신하고, 상기 수신된 에너지값들을 이용하여 각 프레임의 모든 반응에 대하여 각각 현재 프레임 및 이전 프레임의 각 반응의 에너지 및 반응의 차이를 고려한 스코어(score)를 계산하고, 마지막 프레임에서 가장 큰 스코어를 갖는 반응을 기본 주파수로 결정하고, 상기 마지막 프레임의 가장 큰 스코어를 갖는 반응까지 오는 프레임간의 경로를 역추적하여 멜로디 시퀀스를 결정하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 방법.

청구항 19

제18항에 있어서, 상기 (e) 단계는 온셋으로 판단된 시간 정보 또는 프레임 정보에 따라 온셋 구간을 나누고, 각 단위 온셋 구간별로 동적 계획법(dynamic programming)을 수행하여 멜로디 시퀀스를 추정하여 제공하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 방법.

청구항 20

제12항에 있어서, 상기 (f) 단계는 상기 추정된 멜로디 시퀀스를 정규화시킨 후, 상기 정규화된 멜로디 시퀀스와 상기 음원 데이터베이스의 각 음원들의 멜로디 시퀀스에 대하여 동적 타임 워핑(Dynamic Time Warping)을 수행하여 사운드 신호의 프레임에 대응하는 상기 음원들의 프레임을 추정하고, 서로 대응되는 사운드 신호의 프레임들과 음원의 프레임들의 멜로디 차이를 검출하고, 상기 멜로디 차이가 가장 작은 음원을 유사 음원으로 선택하는 것을 특징으로 하는 외부의 사운드 신호를 기반으로 한 반주 제공 방법.

명세서

기술분야

- [0001] 본 발명은 사용자로부터 입력된 사운드를 기반으로 한 반주제공방법 및 이를 이용한 반주제공장치에 관한 것으로서, 더욱 구체적으로는 사용자로부터 입력된 사운드로부터 멜로디의 특징 정보를 추출하고 이를 음원데이터베이스의 음원들과 비교하여 가장 유사한 음원을 추출하고, 이 음원의 반주를 제공하는 것을 특징으로 하는 반주 제공방법 및 이를 이용한 반주제공장치에 관한 것이다.

배경 기술

- [0002] 최근 노래방 등에 사용되는 노래 연습 장치 또는 노래 반주 장치들의 성능이 다양하게 발전하고 있다. 이러한 노래 연습 장치나 노래 반주 장치들에서 선곡하기 위하여, 사용자는 가사집 등을 통해 특정 곡에 대한 식별 번호를 파악하고, 상기 식별 번호를 리모트 컨트롤러나 키패드를 이용하여 노래 연습 장치나 노래 반주 장치로 입력하여야 한다. 하지만, 사용자가 원하는 곡에 대한 제목을 모르고 특정 소절만을 알거나 멜로디만을 알고 있는 경우에는 선곡을 할 수 없는 불편함이 있다.
- [0003] 따라서, 노래 반주 장치나 노래 연습 장치에 있어서, 노래의 제목을 모르더라도 특정 소절이나 멜로디만을 입력하여 해당 곡을 자동으로 선곡할 수 있도록 하는 기능의 필요성이 증대되고 있다.

선행기술문헌

특허문헌

- [0004] (특허문헌 0001) (1) 한국공개특허공보 제10-1997-22955호
(특허문헌 0002) (2) 한국공개특허공보 제10-1998-67955호
(특허문헌 0003) (3) 한국공개특허공보 제10-2000-25944호

발명의 내용

해결하려는 과제

- [0005] 전술한 문제점을 해결하기 위한 본 발명의 목적은 사용자로부터 입력되는 사운드 신호로부터 멜로디 시퀀스를 추출하고, 이를 이용하여 유사 음원에 대한 반주를 제공하는 반주 제공 장치 및 그 방법을 제공하는 것이다.
- [0006] 본 발명의 다른 목적은 사용자로부터 입력되는 사운드 신호로부터 멜로디 시퀀스를 추출하고, 상기 멜로디 시퀀스로부터 사용자가 의도하지 않은 오차, 예컨대 잡음이나 바이브레이션, 잘못된 음정이나 박자 등을 검출하고 보정한 후, 이에 대한 반주를 제공할 수 있는 반주 제공 장치 및 그 방법을 제공하는 것이다.
- [0007] 본 발명의 다른 목적은 사용자로부터 입력되는 사운드 신호로부터 멜로디 시퀀스를 추출하고, 이를 이용하여 유사 음원에 대한 반주를 데이터베이스로부터 판독하고, 사용자의 사운드 신호로부터 추출된 음정과 빠르기를 고려하여 유사 음원의 반주를 보정하여 제공함으로써, 사용자에게 맞는 반주를 제공할 수 있도록 하는 반주 제공 장치 및 그 방법을 제공하는 것이다.

과제의 해결 수단

- [0008] 전술한 기술적 과제를 달성하기 위한 본 발명의 제1 특징은 외부의 사운드 신호를 기반으로 한 반주 제공 장치에 관한 것으로서, 상기 반주 제공 장치는 외부의 입력 장치로부터 사운드 신호가 입력되는 신호 입력부; 다수 개의 음원들에 대한 정보가 저장된 음원 데이터베이스; 상기 신호 입력부로부터 사운드 신호가 제공되면, 사전 설정된 시간 간격의 프레임으로 분할하고 각 프레임에 대한 멜로디를 추출하는 멜로디 추출부; 상기 멜로디 추출부에 의해 추출된 각 프레임에 대한 멜로디를 반음계로 변환시키는 반음계 변환부; 상기 신호 입력부로부터 사운드 신호가 제공되면, 각 프레임에 대한 온셋 정보를 검출하여 제공하는 온셋 검출부; 상기 반음계 변환부로부터 각 프레임에 대한 반음 정보를 수신하고, 상기 온셋 검출부로부터 각 프레임에 대한 온셋 정보를 수신하며, 상기 온셋 정보를 이용하여 사운드 신호의 시작 프레임을 검출하고, 상기 사운드 신호의 시작 프레임부터 일정 시간동안 기본 주파수에 해당하는 멜로디 시퀀스를 추정하여 제공하는 멜로디 시퀀스 추정부; 상기 멜로디 시퀀스 추정부에 의해 추정된 멜로디 시퀀스와 상기 음원 데이터베이스의 음원들의 멜로디 시퀀스를 비교하여 상기 사운드 신호의 멜로디 시퀀스와 가장 유사한 멜로디 시퀀스를 갖는 유사 음원을 검색하는 음원 선택부; 상기 유사 음원에 대한 반주를 상기 음원 데이터베이스로부터 검색하여 출력하는 반주 출력부;를 구비한다.
- [0009] 전술한 제1 특징에 따른 반주 보정 장치에 있어서, 상기 반주 제공 장치는 상기 유사 음원에 대한 반주의 음정을 보정하는 반주 보정부를 더 구비하고, 상기 반주 보정부는 사운드 신호의 멜로디 시퀀스의 평균과 상기 유사 음원의 멜로디 시퀀스의 평균의 차이를 이용하여 상기 유사 음원에 대한 반주의 음정을 보정하는 것을 특징으로 하며, 상기 반주 출력부는 상기 반주 보정부에 의해 보정된 음정을 갖는 반주를 출력한다.
- [0010] 전술한 제1 특징에 따른 반주 보정 장치에 있어서, 상기 반주 제공 장치는 상기 유사 음원에 대한 반주의 빠르기를 보정하는 반주 보정부를 더 구비하고, 상기 반주 보정부는 상기 음원 선택부가 수행한 동적 타임 워핑(Dynamic Time Warping)에서 획득한 경로의 기술기 정보를 이용하여 상기 유사 음원에 대한 반주의 빠르기를 보정하는 것을 특징으로 하며, 상기 반주 출력부는 상기 반주 보정부에 의해 보정된 빠르기를 갖는 반주를 출력한다.
- [0011] 전술한 제1 특징에 따른 반주 보정 장치에 있어서, 상기 음원 데이터베이스는 다수 개의 음원들에 대한 멜로디 시퀀스 및 반주 정보를 구비하는 것이 바람직하다.
- [0012] 전술한 제1 특징에 따른 반주 보정 장치에 있어서, 상기 멜로디 추출부는 각 프레임의 신호들에 대한 자기 상관계수($R(\tau)$)를 구하여 제공하는 것이 바람직하다.
- [0013] 전술한 제1 특징에 따른 반주 보정 장치에 있어서, 상기 반음계 변환부는 상기 멜로디 추출부로부터 각 프레임의 신호들에 대한 자기상관계수들을 수신하고, 상기 자기상관계수들을 양자화하고 반음계의 에너지값으로 변환하여 제공하는 것이 바람직하다.
- [0014] 전술한 제1 특징에 따른 반주 보정 장치에 있어서, 상기 온셋 검출부는 상기 신호 입력부로부터 입력된 각 프레임에 대하여 이전과 현재의 다수개의 프레임의 에너지들의 가중치를 둔 합을 계산하고, 각 프레임에 대해 상기 계산된 합이 사전 설정된 문턱값보다 큰 경우 온셋으로 판단하고, 온셋으로 판단된 시간 정보 또는 프레임 정보를 멜로디 시퀀스 추정부로 제공하는 것이 바람직하다.
- [0015] 전술한 제1 특징에 따른 반주 보정 장치에 있어서, 상기 멜로디 시퀀스 추정부는 상기 온셋 검출부로부터 제공된 첫번째 온셋으로 판단된 시점으로부터 사전 설정된 일정 시간 동안의 사운드 신호에 대해서 멜로디 시퀀스를 추정하여 제공하는 것이 바람직하다.
- [0016] 전술한 제1 특징에 따른 반주 보정 장치에 있어서, 상기 멜로디 시퀀스 추정부는 동적 계획법(dynamic programming)을 수행하여 멜로디 시퀀스를 추정하여 제공하며,
- [0017] 상기 동적 계획법은 상기 반음계 변환부로부터 각 프레임의 각 반음에 대한 양자화된 에너지값들을 수신하고, 상기 수신된 에너지값들을 이용하여 각 프레임의 모든 반음에 대하여 각각 현재 프레임 및 이전 프레임의 각 반음의 에너지 및 반음의 차이를 고려한 스코어(score)를 계산하고, 마지막 프레임에서 가장 큰 스코어를 갖는 반음을 기본 주파수로 결정하고, 상기 마지막 프레임의 가장 큰 스코어를 갖는 반음까지 오는 프레임간의 경로를 역추적하여 멜로디 시퀀스를 결정하는 것이 바람직하며,
- [0018] 특히 상기 멜로디 시퀀스 추정부는 상기 온셋 검출부로부터 제공되는 온셋으로 판단된 시간 정보 또는 프레임 정보에 따라 온셋 구간을 나누고, 각 단위 온셋 구간별로 동적 계획법(dynamic programming)을 수행하여 멜로디

시퀀스를 추정하여 제공하는 것이 더욱 바람직하다.

[0019] 전술한 제1 특징에 따른 반주 보정 장치에 있어서, 상기 음원 선택부는 상기 멜로디 시퀀스 추정부로부터 제공된 멜로디 시퀀스를 정규화시킨 후, 상기 정규화된 멜로디 시퀀스와 상기 음원 데이터베이스의 각 음원들의 멜로디 시퀀스에 대하여 동적 타임 워핑(Dynamic Time Warping)을 수행하여 사운드 신호의 프레임에 대응하는 상기 음원들의 프레임을 추정하고, 서로 대응되는 사운드 신호의 프레임들과 음원의 프레임들의 멜로디 차이를 검출하고, 상기 멜로디 차이가 가장 작은 음원을 유사 음원으로 선택하는 것이 바람직하다.

[0020] 본 발명의 제2 특징은 다수 개의 음원들에 대한 멜로디 시퀀스 및 반주 정보가 저장된 음원 데이터베이스를 이용하여, 외부의 입력장치로부터 입력된 사운드 신호를 기반으로 하여 반주를 제공하는 방법에 있어서, (a) 외부의 입력 장치로부터 사운드 신호를 입력받는 단계; (b) 상기 신호 입력부로부터 사운드 신호가 제공되면, 사전 설정된 시간 간격의 프레임으로 분할하고 각 프레임에 대한 멜로디를 추출하는 단계; (c) 상기 추출된 각 프레임에 대한 멜로디를 반음계로 변환시켜 각 프레임의 각 반음에 대한 에너지를 구하는 단계; (d) 상기 신호 입력부로부터 사운드 신호가 제공되면, 각 프레임에 대한 온셋 정보를 검출하는 단계; (e) 상기 검출된 온셋 정보를 이용하여 사운드 신호의 시작 프레임을 검출하고, 각 프레임의 각 반음에 대한 에너지들로부터 기본 주파수를 검출하고, 상기 사운드 신호의 시작 프레임부터 일정시간동안 기본 주파수에 해당하는 멜로디 시퀀스를 추정하는 단계; (f) 상기 추정된 멜로디 시퀀스와 상기 음원 데이터베이스의 음원들의 멜로디 시퀀스를 비교하여 상기 사운드 신호의 멜로디 시퀀스와 가장 유사한 멜로디 시퀀스를 갖는 음원을 검색하여 유사 음원을 선택하는 단계; (g) 상기 유사 음원에 대한 반주를 상기 음원 데이터베이스로부터 검색하여 출력하는 단계; 를 구비한다.

[0021] 전술한 제2 특징에 따른 반주 보정 방법에 있어서, 상기 (g) 단계는 상기 유사 음원에 대한 반주를 상기 음원 데이터베이스로부터 검색한 후, 사운드 신호의 멜로디 시퀀스의 평균과 상기 유사 음원의 멜로디 시퀀스의 평균의 차이를 이용하여 상기 유사 음원에 대한 반주의 음정을 보정하고, 상기 보정된 음정을 갖는 반주를 출력하는 것이 바람직하다.

발명의 효과

[0022] 본 발명에 따른 반주제공방법 및 장치에 의하여, 사용자로부터 짧은 시간 동안의 사운드가 입력되면, 입력된 사운드로부터 멜로디를 추출하고 이를 이용하여 가장 유사한 음원을 검색하고, 사용자로부터 입력된 사운드로부터 추출된 멜로디의 음정과 빠르기를 이용하여 상기 검색된 음원의 반주를 보정하여 출력함으로써, 사용자의 음정과 박자에 맞는 반주를 제공할 수 있게 된다.

[0023] 본 발명에 따른 반주 제공 장치는 연산 속도를 최소화시켜 빠른 시간내에 유사 음원을 찾을 수 있게 된다.

[0024] 또한, 본 발명에 따른 반주 제공 장치는 사용자가 의도하지 않은 잡음, 바이브레이션 등과 같은 오차들을 검출하고 보정하여 가장 정확한 유사 음원을 검색할 수 있게 된다.

도면의 간단한 설명

[0025] 도 1은 본 발명에 따른 반주제공장치(1)의 구성을 개략적으로 도시한 블록도이다.

도 2는 본 발명의 바람직한 실시예에 사용한 Optimal Filter의 파형을 예시적으로 도시한 것이다.

도 3은 본 발명에 따른 최적 필터를 통과한 후의 에너지 함수 $F(t)$ 를 도시한 그래프이다.

도 4는 본 발명에 따른 DTW 연산 결과를 도시한 그래프이다.

도 5는 수학적 9를 간단하게 표현한 그래프이다.

도 6은 Dynamic Time Warping을 수행함에 있어서, 사운드 신호가 유사음원의 1/2배~2배의 빠르기를 가진다고 가정했을 때, 고려해야하는 경로의 범위를 도시한 그래프이다.

도 7은 본 발명에 따른 멜로디 추출부에 의해 사운드 신호를 Autocorrelation function을 이용한 FFT 한 후 각 프레임마다 기본 주파수들로 추정되는 값들을 출력한 그래프이다.

도 8은 본 발명에 따른 반주 제공 장치의 반음계 변환부에 의한 결과를 도시한 그래프이다.

도 9는 본 발명에 따른 반주 제공 장치의 멜로디 시퀀스 추정부의 결과를 도시한 그래프이다.

발명을 실시하기 위한 구체적인 내용

- [0026] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예에 따른 외부로부터 입력된 사운드를 기반으로 한 반주제공방법 및 이를 이용한 반주제공장치에 대하여 구체적으로 설명한다. 본 발명에 따른 반주제공장치는 사용자가 부른 약 7초 동안의 노래 또는 허밍으로부터 멜로디를 추출하고, 추출된 멜로디와 데이터베이스에 저장된 원곡 멜로디들을 비교하여 가장 유사한 원곡 멜로디를 갖는 음원에 대한 반주를 재생하는 것을 특징으로 한다. 이때 외부 사운드에서 추출된 멜로디의 음정과 박자에 따라 음원의 반주를 보정하여 재생하는 것을 특징으로 한다.
- [0027] 도 1은 본 발명에 따른 반주제공장치(1)의 구성을 개략적으로 도시한 블록도이다. 도 1을 참조하면, 본 발명에 따른 반주제공장치(1)는 다수개의 음원들에 대한 멜로디 시퀀스 정보 및 반주 정보가 저장된 음원데이터베이스(100), 외부로부터 사운드 신호가 입력되는 신호 입력부(105), 멜로디 추출부(110), 반음계 변환부(120), 온셋 검출부(130), 멜로디 시퀀스 추정부(140), 유사 음원 선택부(150), 반주 보정부(160) 및 반주 출력부(170)를 구비한다. 이하, 전술한 각 구성요소들의 동작에 대하여 구체적으로 설명한다.
- [0028] 상기 신호 입력부(105)는 마이크 등과 같은 입력 장치를 통해 외부로부터 입력된 사운드 신호를 멜로디 추출부(110) 및 온셋 검출부(130)로 동시에 전송한다. 상기 입력 장치는 비교적 주변 잡음에 강인한 지향성 마이크를 사용하는 것이 바람직하며, 상기 지향성 마이크를 통해 기본 주파수(fundamental frequency f_0)가 뚜렷하게 나타나는 신호를 입력받기 때문에 본 발명에 따른 방법을 사용할 경우 성능 및 연산 속도에서 모두 이득을 얻을 수 있다.
- [0029] 상기 멜로디 추출부(110)는 입력된 사운드 신호에 대하여 자기 상관 함수를 사용하여 기본 주파수를 추정함으로써, 상기 입력된 사운드 신호의 전체로부터 각 시점의 멜로디 정보를 추출한다. 본 발명에 따른 자기 상관 함수를 이용하여 멜로디 정보를 추출하는 방법은 연산 속도가 빠르고 수식이 단순하다는 장점이 있는 반면, 주변 잡음에 상대적으로 취약하다는 단점이 있다. 하지만, 입력 장치로서 지향성 마이크를 사용하는 경우 비교적 주변 잡음에 강인하고 음성의 기본 주파수가 뚜렷하게 드러나는 신호를 입력받기 때문에, 자기 상관 함수를 사용하는 것이 성능 및 속도면에서 가장 바람직하다. 이하, 상기 멜로디 추출부의 동작에 대하여 구체적으로 설명한다.
- [0030] 상기 멜로디 추출부는 각 시점에 대한 멜로디 정보를 추출하기 위하여, 먼저 사운드 신호를 일정 시간 단위의 프레임(frame)으로 분할한 후, 각 프레임에 대한 주파수 정보를 구한다. 이 때, 예를 들면 상기 프레임은 60ms의 길이를 가지며 480 샘플들로 구성되며, 20ms의 이동(shift)구간을 가질 수 있다. 하지만, 프레임의 길이, 샘플들 개수 및 이동 구간은 시스템의 설계에 따라 적절하게 설정될 수 있다.
- [0031] 상기 멜로디 추출부(110)는 각 프레임에 대한 주파수 정보를 포함하는 자기 상관 계수 $R(\tau)$ 를 구하기 위하여, 각 프레임의 신호들($x(t)$)에 대하여 고속 푸리에 변환(Fast Fourier Transform: 'FFT')한 후 자기상관함수(Autocorrelation Function) 처리한다. 각 프레임의 신호들에 대하여 FFT 한 후 자기상관함수로 처리한 자기상관계수 $R(t, \tau)$ 는 수학식 1과 같이 나타낼 수 있다.

수학식 1

$$R(t, \tau) = IFFT (|FFT (x_t(n))|^2)$$

- [0032]
- [0033] 여기서, $x_t(n)$ 는 상기 신호 입력부로부터 입력된 사운드 신호 중 t 번째 프레임의 입력 신호를 나타내며, $R(t, \tau)$ 는 $x_t(n)$ 의 자기상관계수이다. 특정 시간지연값(τ)에서의 자기상관계수가 매우 크다면 해당 입력신호는 τ 에 해당하는 시간차의 주기로 반복되는 형태를 가지고 있다고 볼 수 있으며, 따라서 이 시간지연값의 역수에 해당하는 주파수를 기본 주파수로 가지고 있다고 볼 수 있다. 추가적으로 음성 신호는 기본주파수 외에도 높은 세기의 배음(harmonics) 성분을 함께 가지고 있기 때문에, 이러한 배음 성분으로 인해 신호의 주기성이 잘못 나타나는 현상을 최소화하기 위하여 사전에 사람의 음역에 해당하는 주파수 대역인 55Hz ~ 1000 Hz 외의 주파수 성분을 필터링하는 것이 바람직하다.
- [0034] 도 7은 본 발명에 따른 멜로디 추출부에 의해 사운드 신호를 Autocorrelation function을 이용한 FFT 한 후 각 프레임마다 기본 주파수들로 추정되는 값들을 출력한 그래프이다. 도 7을 참조하면, 일정 구간은 멜로디 라인이 잘 찾아지는 것을 확인할 수 있으나, 중간 중간 엄청난 기본 주파수의 값이 측정되는 것을 볼수 있는데, 이는

묵음 구간으로 사용자의 음성이 아니라 노이즈로 인해 발생한 신호의 미세한 power 때문에 발생한 것으로 볼 수 있다.

[0035] 상기 반응계 변환부(120)는 멜로디 시퀀스 추정부(140)에서의 연산 속도를 줄여주기 위한 것으로서, 상기 멜로디 추출부로부터 각 프레임의 신호들에 대한 자기상관계수들($R(\tau)$)을 수신하고, 수신된 자기상관계수들을 수학식 2를 이용하여 양자화된 반응계의 에너지(Energy at semitones of Interest; 이하 'ESI'라 한다)로 변환한다.

수학식 2

$$Y(t, m) = \max_{\tau \in T_m} (R(t, \tau))$$

[0037] 일반적인 음악은 서양의 12음계를 따르는데, 이는 한 옥타브 내에는 12개의 반응으로 이루어져 있다는 것을 의미한다. 따라서, 이 이상으로 정확하게 기본 주파수를 찾는 것은 무의미할 뿐만 아니라 전체 연산 속도를 느리게 하는 원인이 될 수 있다. 따라서, $\tau = 1/f$ 에 대한 자기상관계수를 각 반응에 해당하도록 수학식 2를 이용하여 양자화하게 된다. 수학식 2에 있어서, T_m 은 m 번째 반응에 해당하는 시간 지연값(τ)들의 집합이다. $Y(t, m)$ 은 자기상관계수들($R(\tau)$)을 양자화하여 구한 t 번째 프레임의 m 번째 반응의 자기상관계수를 의미한다.

[0038] 사람의 음역대를 고려했을 때, 가장 낮은 음인 $mel(1)$ 을 55Hz 라고 볼 수 있으며, 그 외의 m 번째 반응들에 해당하는 주파수는 수학식 3으로 나타낼 수 있다.

수학식 3

$$mel(m) = 55 \times 2^{\frac{m-1}{12}}$$

[0040] 따라서, 연속하는 두 반응의 기하평균에 해당하는 주파수를 경계로 하여 m 번째 반응의 주파수 대역을 정할 수 있다. 이를 수학식 4와 같이 나타낼 수 있다. F_m 은 m 번째 반응의 주파수 대역을 의미하며, $mel(m)$ 은 m 번째 반응의 주파수를 의미한다.

수학식 4

$$F_m = \{ f \mid \sqrt{mel(m-1) \times mel(m)} \leq f < \sqrt{mel(m) \times mel(m+1)} \}$$

[0042] 자기상관함수의 자기상관계수들(τ)과 주파수(f)는 서로 역수관계이기 때문에, T_m 은 수학식 5와 같이 정의할 수 있다.

수학식 5

$$T_m = \left\{ \tau \mid \frac{1}{\tau} \in F_m \right\}$$

[0044] 여기서, $mel(1)$ 을 55Hz로 하였으며, 5옥타브의 음역을 고려하여 전체 반응의 수는 60개로 정한다. 이러한 과정을 통해 각 프레임당 약 2000 개의 주파수 데이터를 약 60개로 줄일 수 있게 된다.

[0045] 도 8은 본 발명에 따른 반주 제공 장치의 반응계 변환부에 의한 결과를 도시한 그래프이다. 도 8을 참조하면, ESI 과정을 거친 후의 결과로 이전 FFT의 결과에서의 y축(주파수)은 0~2000의 인덱스를 가지는데, ESI 과정을 거치면서 60 개의 인덱스로 감소됨을 알 수 있다.

- [0046] 상기 온셋(Onset) 검출부(130)는 입력된 사운드 신호로부터 묵음 구간을 제외하고 사운드가 입력되는 프레임을 검출하기 위한 것이다. 이하, 상기 온셋 검출부의 동작에 대하여 구체적으로 설명한다.
- [0047] 상기 온셋 검출부는 최적 필터를 이용하여 상기 신호 입력부로부터 입력된 각 프레임에 대하여 이전과 현재의 다수개의 프레임의 에너지들의 가중치를 둔 합을 계산하고, 이렇게 계산된 값이 사전 설정된 문턱값보다 큰 경우 온셋으로 판단하고, 온셋으로 판단된 시간 정보 또는 프레임 정보를 멜로디 시퀀스 추정부로 제공한다.
- [0048] 온셋 검출부에 의해 검출된 온셋 지점 정보를 이용하여, 상기 신호입력부로부터 입력된 각 프레임에 대하여 입력된 신호의 시작 이전의 음성 신호가 존재하지 않는 묵음 구간을 제거하고, 신뢰할 수 있는 구간과 신뢰할 수 없는 구간을 구분하고, 신뢰성있는 구간에 대해서만 연산함으로써, 멜로디 추정부(140)에서의 연산량을 최소화시켜 연산 속도를 줄일 수 있도록 한다. 상기 묵음 구간은 기본 주파수의 신호가 존재하지 않거나 잡음의 기본 주파수의 신호가 가장 우세한 구간을 의미한다.
- [0049] 따라서, 온셋 검출부에 의하여 음성 신호의 시작 전의 묵음 구간을 제거하여 데이터베이스의 원곡의 멜로디들과 비교하는 기준점을 확실하게 설정하게 된다. 온셋 검출부는 수학적 식 6을 사용하여 설계된 최적 필터(optimal filter) h 를 이용하여 온셋을 판단한다.

수학적 식 6

$$h(i) = \{-f(-W \leq i \leq 0), f(1 \leq i \leq W)\}$$

$$f(x) = e^{Ax} [K_1 \sin(Ax) + K_2 \cos(Ax)] + e^{-Ax} [K_3 \sin(Ax) + K_4 \cos(Ax)] + K_5 + K_6 e^{8x}$$

[0050]

- [0051] 여기서, $h(i)$ 는 최적 필터이며, W 는 최적 필터의 크기(Size)를 결정짓는 변수이다.
- [0052] 전술한 최적 필터는 f 의 전부가 아닌 일부분만을 사용하여 만들었으며, 모양은 도 2와 같다. 도 2는 본 발명의 바람직한 실시예에 사용한 Optimal Filter의 파형을 예시적으로 도시한 것이다.
- [0053] 기본적으로 이전 프레임에 비해 현재 프레임의 에너지가 급격하게 높으면, 즉 현재 프레임과 이전 프레임의 차이가 일정 이상 클 경우 온셋(Onset)이라고 판단할 수 있다. 하지만, 이전과 현재 프레임만을 가지고 판단하게 되면 순간적인 잡음 및 유저의 호흡 등에 매우 민감해지게 된다. 따라서, 전술한 최적 필터를 사용하여 이전의 다수개의 프레임과 현재의 다수개의 프레임의 에너지들의 가중치를 둔 합을 기준으로 온셋(Onset)을 판단할 경우, 보다 강인한 성능을 보장할 수 있다. 전술한 바와 같이, 수학적 식 7은 최적 필터를 사용하여 온셋 검출하기 위한 수학적식으로서, 이전과 현재의 다수개의 프레임의 에너지들의 가중치를 둔 합을 이용하여 온셋을 판단한다.

수학적 식 7

$$F(t) = \sum_{i=-W}^W h(i)g(t+i)$$

[0054]

- [0055] 여기서, $g(t)$ 는 t 번째 프레임의 에너지(Energy)를 의미한다.
- [0056] 도 3은 본 발명에 따른 최적 필터를 통과한 후의 에너지 함수 $F(t)$ 를 도시한 그래프이다. 도 3을 참조하면, A 지점에서 $F(t)$ 가 사전 설정된 문턱값을 넘게 되고, 이 지점부터 입력 구간으로 설정하게 된다. 최초의 음성 구간 이전의 묵음 구간은 입력이 시작되기 전으로 판단하여 이후의 연산 과정에서 제거하고 처음으로 온셋이 검출된 시점부터 일정 시간동안의 신호를 녹음하여 전체 연산 과정에 사용한다. 본 발명에서는 온셋이 처음 검출된 시점으로부터 7초 동안의 신호를 녹음하여 진행하였다.
- [0057] 또한, 음성이 처음 시작된 이후에 나타나는 묵음 구간으로 인해 발생하는 문제들을 해결하기 위하여, 각 온셋 구간별로 동적계획법(dynamic Programming)을 수행하여 최적의 멜로디 시퀀스를 추정하게 된다.
- [0058] 상기 멜로디 시퀀스 추정부(140)는 동적 계획법(Dynamic Programming)을 사용하여 외부로부터 입력된 음성 신호로부터 최적의 멜로디 시퀀스를 추정하는 것으로서, 원 신호와 잡음이 섞인 입력 신호에서 피치(pitch)를 검출

한 결과에서 하모닉(harmonic) 성분이 잡음(noise)과 합성되어 기본주파수(f_0)에 해당하는 성분의 값보다 크게 측정되어 제대로 된 기본주파수 성분을 검출하지 못하는 것을 방지하기 위한 것이다. 입력 신호가 음성, 특히 노래 신호임을 고려했을 때 멜로디가 20ms 의 프레임의 시간차 사이에 급격하게 변화하지 않을 것으로 가정할 수 있다. 이러한 가정에 근거하여, 본 발명에 따른 멜로디 시퀀스 추정부는 전후 프레임의 기본 주파수를 고려한 주파수 추출 방법을 사용하며, 이를 위해 수학적 식 8에 따른 동적 계획법을 사용한다.

수학적 식 8

$$D(1,m) = Y(1,m)$$

$$D(t,m) = Y(t,m) + \max_{k \in [1,n]} \{D(t-1,k) - \theta \times |k-m|\}$$

[0059]

[0060]

여기서, $Y(t,m)$ 은 t 번째 프레임의 m 번째 반응의 에너지를 의미하며, $D(t,m)$ 은 현재의 에너지뿐만 아니라 이전 프레임의 각 반응의 에너지 및 음 차이를 고려한 스코어(score)를 의미한다. 또한, n은 전체 반응의 수를 나타내는 것으로서, 본 발명에서는 5옥타브의 음역을 고려하여 전체 반응의 수를 60 으로 정하였으므로 n은 60이 될 수 있다.

[0061]

수학적 식 8을 살펴 보면, 첫번째 프레임은 이전 프레임이 존재하지 않기 때문에 $D(1,m)$ 은 현재의 $Y(t,m)$ 이 그대로 스코어로 반영된다. 두번째 프레임에 있어서, m 번째 반응들은 자기 자신들의 에너지와 첫번째 프레임의 k 번째 반응의 스코어, 즉 에너지를 더한 값을 스코어로 가지는데, 이 때 k는 $D(1,k)$ 의 값이 크면서 $|k-m|$ 이 작은, 즉 m과 가능한 비슷한 반응에 해당한다. $|k-m|$ 의 값을 고려함으로써 현재 프레임의 기본 주파수가 직전 프레임에 비해 급격히 달라지지는 않을 것을 제약 조건으로 고려하게 되는 것이다. θ 는 결과가 정확하게 도출될 수 있도록 하기 위하여 설계자가 설정하는 가중치이다. θ 의 값이 작을수록 k는 $D(1,m)$ 의 크기에 영향을 받으며, 클수록 m과 k의 비슷한 정도에 영향을 받는다. θ 의 크기가 적절하다면, 상대적으로 프레임의 에너지가 큰 음성 구간에서는 프레임 자체의 기본 주파수를 그대로 따라갈 것이며 프레임의 에너지가 작은 묵음 구간에서는 $|k-m|$ 의 값을 가능한 줄일 수 있도록 이전 프레임의 기본 주파수를 유지할 것으로 기대할 수 있다. 마지막 프레임의 스코어가 계산되고 나면, 마지막 프레임에서 가장 큰 스코어를 가지는 반응을 해당 프레임의 기본 주파수로 결정하고 해당 반응까지 오게 되는 프레임간의 경로를 역추적하여 최종 멜로디 시퀀스를 결정한다.

[0062]

동적계획법의 특성상, 현재 프레임에서 추정되는 기본 주파수는 이전 프레임뿐만 아니라 이후 프레임의 기본 주파수에도 영향을 받게 되는데, 이것은 문제가 될 여지가 있다. 두 음성 구간 사이의 묵음 구간에 대해 생각해 보았을 때, 이 구간의 기본 주파수가 원래는 이전 음성구간의 기본 주파수와 동일하게 나타났어야 하나 사용자의 호흡 등의 문제로 묵음으로 처리되었을 것으로 판단하는 것은 적절한 것으로 생각되지만, 이후 음성구간의 기본주파수가 미리 나타났을 것으로 판단하는 것은 적절치 못하기 때문이다. 따라서 이와 같은 문제를 해결하기 위해, 온셋 검출부에 의해 검출된 온셋 정보를 이용한다. 온셋 검출부에 의해 각 온셋의 위치들이 검출되면, 하나의 온셋으로부터 다음 온셋까지의 구간을 하나의 단위 음성구간으로 설정한다. 이때, 단위 음성구간의 앞부분에는 실제 음성 구간이 존재할 것이며, 뒷부분에는 호흡 등으로 인한 묵음 구간이 존재하게 된다. 이렇게 나누어진 단위 음성구간에 대해 독립적으로 동적계획법을 적용하여 각 단위 음성 구간별로 기본 주파수를 추출함으로써, 묵음 구간에서 추정되는 기본 주파수가 이후의 음성 구간에 영향을 미치는 것을 막을 수 있다.

[0063]

도 9는 본 발명에 따른 반주 제공 장치의 멜로디 시퀀스 추정부의 결과를 도시한 그래프이다. 도 9를 참조하면, 멜로디 시퀀스 추정부에 의해 Dynamic Programming 과정을 거친 것으로서 묵음 구간의 잘못된 기본 주파수 정보를 대부분 제거되었고, 온셋 검출부에 의해 검출된 스피치 구간의 시작 지점을 적용하여 스피치 구간 이전의 값은 모두 제거하여 ESI 결과의 첫 부분의 지저분한 파형이 정리되었음을 쉽게 확인할 수 있다.

[0064]

상기 유사음원 선택부(150)는 상기 멜로디 시퀀스 추정부로부터 제공된 추정된 멜로디 시퀀스와 데이터베이스에 저장된 다수 개의 원곡들에 대한 멜로디 시퀀스를 비교하여 상기 추정된 멜로디 시퀀스와 가장 유사한 멜로디 시퀀스를 갖는 원곡인 유사음원을 검출한다. 이하, 본 발명에 따른 반주 제공 장치의 유사음원 선택부(150)가 상기 멜로디 시퀀스 추정부에 의해 추정된 멜로디 시퀀스와 가장 유사한 멜로디 시퀀스를 갖는 유사음원을 검색하는 과정을 보다 구체적으로 설명한다.

[0065]

상기 멜로디 시퀀스 추정부에 의해 추정된 멜로디 시퀀스는 각 사용자마다 기본적인 음역대가 모두 다르며 전체

적인 음정도 시간에 따라 조금씩 변화할 뿐만 아니라, 상황에 따라 추정된 멜로디 시퀀스를 완전히 신뢰할 수 없는 경우가 발생하므로, 상기 추정된 멜로디 시퀀스를 그대로 데이터베이스의 멜로디 시퀀스와 매칭시키는 것은 바람직하지 못하다. 또한, 추정된 멜로디의 빠르기도 데이터베이스의 멜로디의 빠르기와 다를 수 있으며, 이 또한 시간에 따라 달라질 수 있다. 이러한 문제점을 해결하기 위해, 본 발명에 따른 유사음원 선택부(150)는 상기 추정된 멜로디 시퀀스를 local-mean normalization을 사용하여 정규화시킨 후, Dynamic Time warping(DTW)을 사용해 음정과 빠르기에 무관하게 매칭이 가능하도록 한다.

[0066] 상기 Local-mean normalization은 기본적으로 멜로디 시퀀스의 평균이 0이 되도록 정규화하는 방법이다. 이러한 방법을 사용했을 경우 음역대만 달라진 동일한 노래의 경우 같은 결과를 출력하게 된다. 그러나 사용자의 음역은 시간의 변화에 따라 조금씩 변화할 수 있기 때문에, 이를 무시하고 전체 구간에 대해 정규화하는 것은 적절하지 못하다. 따라서 여기에서는 1초, 즉 50프레임 내의 멜로디의 평균이 모두 0이 될 수 있도록 지역적인(local) 정규화 방법을 사용하였다.

[0067] 상기 DTW는 기본적으로 멜로디 시퀀스 추정부에서 사용한 동적계획법과 동일한 알고리즘이지만, 멜로디를 추정하기 위한 동적계획법의 결과가 프레임마다 하나의 반음을 거치는 경로로 나타나는 반면 DTW의 결과는 왼쪽 아래의 끝점, 즉 신호가 시작되는 지점으로부터 오른쪽 위의 끝점 즉 신호가 끝나는 지점으로 이어지는 경로로 나타난다는 점에서 차이가 있다. 도 4는 본 발명에 따른 DTW 연산 결과를 도시한 그래프이다. 도 4에 있어서, x축과 y축은 각각 입력 신호와 데이터베이스의 원곡 신호에 대한 프레임 번호들을 의미하여, (a,b) 위치의 값 즉 밝기는 입력 신호의 a번째 프레임의 멜로디와 DB 신호의 b번째 프레임이 멜로디의 차이를 의미한다. 여기서, 값이 작을수록 입력 신호와 DB 신호가 비슷한 멜로디를 가지고 있다는 것을 의미하기 때문에, 입력신호의 시작 부분부터 끝부분까지 작은 값들을 따라 경로를 추적해 가게 되면 입력신호의 각각의 프레임에 해당하는 DB의 프레임을 추정할 수 있다.

[0068] 도 4를 참조하면, 사용자는 노래가 시작하는 순간에는 원곡에 비해 매우 빠르게 노래를 부르다가 곧 원곡보다 느리게 부르기 시작했으며, 다시 차츰 박자가 빨라져 최종적으로는 350프레임, 즉 7초에 해당하는 시간동안 원곡의 400프레임, 즉 8초만큼에 해당하는 노래를 불렀다는 것을 알 수 있다. 수학식 9는 이를 수식적으로 나타낸 것이다.

수학식 9

$$D(i_x, i_y) = \min \left\{ \begin{matrix} D(i_x-1, i_y)g(k) \\ D(i_x-1, i_y-1) \\ D(i_x-1, i_y-2) \end{matrix} \right\} \text{ with } g(k) = \begin{cases} 1 & \Phi(k-1) \neq \Phi_y(k-2) \\ \infty & \Phi(k-1) = \Phi_y(k-2) \end{cases}$$

[0069]

[0070] 도 5는 수학식 9를 간단하게 표현한 것이다. 동적계획법과 마찬가지로 이전 프레임의 D 값을 고려하는데, 이때 DB의 신호가 원곡과 극단적으로 속도가 빠르거나 느리지 않을 것이며, 시간을 거스르지 않을 것이며 또한 연속적인 경로로 나타날 것을 전제로 했을 때 도 5와 같이 입력 신호의 한 프레임에 해당하는 시간 즉 20ms가 지나가는 동안 DB는 시간이 흐르지 않거나 20ms 또는 40ms정도가 흘렀을 것으로 기대할 수 있다. 또한 입력신호가 두 프레임 이상 지나가는 동안 DB가 한 프레임도 지나가지 않는 상황도 없을 것으로 예상할 수 있기 때문에 해당 경로에 대해서만 DTW를 수행하여 연산 속도를 최소화 할 수 있다. 또한 전체적으로 입력 신호는 DB의 1/2배 ~ 2배의 빠르기를 가진다고 가정하면 도 6과 같은 범위 내의 경로에 대해서만 고려해도 충분하다고 판단할 수 있다. 마지막으로, 에너지가 클수록, 즉 D의 값이 클수록 기본 주파수일 가능성이 높아지는 동적계획법의 경우와 달리, 입력 신호와 DB 신호의 멜로디 차이가 작을수록, 즉 D의 값이 작을수록 경로에 해당할 가능성이 높아지기 때문에 최소값을 찾아 경로를 추정한다.

[0071] 상기 음원 보정부(160)는 상기 유사 음원 선택부에 의해 선택된 원본 반주신호의 음정 높낮이와 재생속도를 사용자가 부르는 멜로디에 맞추도록 보정한다. 이하, 본 발명에 따른 반주 제공 장치의 음원 보정부(160)의 동작을 구체적으로 설명한다.

[0072] 일반적으로 카세트테이프에 녹음된 음성신호를 재생할 때 테이프의 진행속도를 결정하는 모터의 회전속도를 바꾸거나, 레코드판에 기록되어있는 음성신호를 재생할 때 레코드판의 회전속도를 빠르게 또는 느리게 바꾸면, 음정의 높낮이와 신호가 재생되는 속도가 함께 바뀌게 된다. 이는 임의의 기록된 음성신호의 경우 간단한 방법으

로는 음정의 높낮이와 빠르기를 따로따로 변형시키는 것이 불가능하다는 것을 의미한다.

[0073] 보통 빠르기를 증가시키면 음정의 높낮이도 함께 높아지고, 반대의 경우 음정의 높낮이가 낮아지는데, 이러한 현상은 디지털방식으로 녹음된 신호에서도 기록된 샘플링 주파수와 재생장치의 샘플링주파수를 다르게 설정함으로써 연출할 수 있다. 예를 들어 44.1kHz 샘플링주파수로 녹음된 반주신호를 1.2배 빠르고 음정이 높아지게 재생하려면, 재생장치의 샘플링주파수를 $44.1\text{kHz} \times 1.2 = 52.92\text{kHz}$ 로 설정하면 된다. 또한 위와 반대로 재생장치의 샘플링주파수는 44.1kHz로 고정하고 기록된 신호의 샘플링주파수를 $44.1\text{kHz} / 1.2 = 36.75\text{kHz}$ 로 변경하여 다시 기록된 것을 재생하면 똑같은 효과를 거둘 수 있다.

[0074] 본 발명에 따른 반주 제공 장치의 음원 보정부는 a 배로 음정을 높이기 위해 반주신호를 $1/a$ 배 Resampling하여 우선적으로 반주의 음정 높낮이를 사용자가 부르고 있는 수준으로 보정한 후, 덩달아 a 배 변한 빠르기를 보정하는 $1/a$ 비율과 사용자가 부르고 있는 빠르기에 맞추기 위한 비율 β 를 함께 고려한 β/a 배 만큼을 빠르기만 조절할 수 있는 알고리즘으로 조작하여 사용자에게 적응적인 반주를 출력한다.

[0075] 먼저 샘플링주파수 변경에는 결과물의 음질왜곡과 연산량 사이에 적절한 타협을 이루는 선형보간법을 사용하는 데, 시간영역의 원본 반주신호 $s(n)$ 에 대하여 수학적 10과 같이 변경된 샘플링주파수의 신호 $y(\tilde{n})$ 을 구한다.

수학적 10

$$y(\tilde{n}) = s(n_1) + \frac{\tilde{n}/\tilde{F}_s - n_1/F_s}{n_2/F_s - n_1/F_s} \{s(n_2) - s(n_1)\}, \quad \frac{n_1}{F_s} \leq \frac{\tilde{n}}{\tilde{F}_s} \leq \frac{n_2}{F_s}$$

[0077] 여기서, $\tilde{n}$ 은 출력신호의 샘플링주파수에 따른 이산적인 시간 인덱스를 나타내고, n_1 과 n_2 는 원본신호의 시간 축에서 $\tilde{n}$ 와 이웃한 두 시간 인덱스를 나타낸다.

[0078] 음정의 높낮이의 변화 없이 빠르기만을 조작할 수 있는 알고리즘은 잘 알려진 몇 가지 방법 중에서도 음악신호와 같은 폴리포닉 사운드에 잘 맞는 phase vocoder를 사용하였다. 이 알고리즘은 먼저 STFT를 취해 주파수 정보를 구하고, 복소수로 이루어진 신호의 주파수표현들을 크기와 위상으로 나눈 뒤, 크기 신호는 선형보간법으로 늘어나거나 줄어든 프레임 개수에 맞게 신호의 길이를 팽창시키거나 압축시키고, 위상 신호는 갑작스러운 위상의 도약을 방지하기 위해 신호 자체의 인위적인 위상변화만을 최근방이웃 보간법으로 선택하여 새로운 프레임간격으로 누적해나가는 방법이다.

[0079] 먼저 크기신호들은 각각의 주파수밴드 ω 에서 프레임단위의 시퀀스에 대하여 위의 샘플링주파수 변경과 같은 방식의 선형보간법으로 늘어나거나 줄어든 갯수의 출력 프레임들의 크기신호를 구한다. 그리고 위상신호의 경우에는 수학적 11과 같이 원본신호의 프레임 간 위상변화에서 각 밴드의 중심주파수에 해당하는 정현파의 위상변화량을 빼 것을 유의미한 위상정보로 가정한다.

수학적 11

$$\Delta\phi(\omega, \tau) = \angle S(\omega, \tau) - \angle S(\omega, \tau-1) - \angle \exp(-j\omega\Delta\tau)$$

[0081] 여기서 $\Delta\phi(\omega, \tau)$ 는 ω 주파수밴드의 τ 번째 프레임에서의 유의미한 위상정보를 나타내고, $\Delta\tau$ 는 프레임간의 시간차이를 의미한다. 또한 $\angle S(\omega, \tau)$ 는 원본 반주신호 $s(n)$ 의 시간주파수 표현에서의 위상신호를 의미한다. 이렇게 $\Delta\phi(\omega, \tau)$ 를 구하면, 수학적 12와 같이 출력신호의 각 프레임마다 원본신호의 최근방이웃 프레임을 골라서 위상정보를 취해 출력프레임 $\tilde{\tau}$ 에 대한 위상정보로 삼는다.

수학식 12

[0082]
$$\Delta \tilde{\phi}(\omega, \tilde{\tau}) = \Delta \phi(\omega, \text{round}(\frac{\alpha}{\beta} \tau))$$

[0083] 이를 수학식 13과 같이 시작프레임부터 누적하는 방법으로 연속성이 보장되는 새로운 위상신호를 생성한다.

수학식 13

[0084]
$$\tilde{\phi}(\omega, \tilde{\tau}) = \tilde{\phi}(\omega, \tilde{\tau} - 1) + \angle \exp(-j\omega \Delta \tau) + \Delta \tilde{\phi}(\omega, \tilde{\tau})$$

[0085] 이렇게 구하여진 크기신호와 위상신호를 곱하여 출력신호의 시간-주파수 표현을 완성한 뒤, inverse STFT를 수행하면 음정의 높낮이 변화 없이 빠르기만 변화된 신호를 얻을 수 있다.

[0086] 사용자의 음역대 및 빠르기를 구하는 방법은 다음과 같다. 먼저, 음역대 α 는 매칭된 DB의 원곡과 사용자로부터 입력된 음성 신호의 멜로디 시퀀스의 평균의 차이를 통해 구할 수 있다. 예를 들어, 사용자의 멜로디 시퀀스의 평균값이 매칭된 원곡의 멜로디 시퀀스의 평균보다 1이 작다면, 사용자가 원곡보다 반음 낮게 부른 것으로 추정하는 것이다. 사용자의 음정은 시간에 따라 조금씩 변할 수 있기 때문에, 전체 입력신호를 비교하지 않고 마지막 2~3초 가량의 비교를 통해 구하는 것이 바람직하다.

[0087] 사용자의 빠르기 β 는 DTW를 통해 구한 경로의 전체적인 기울기를 통해 구할 수 있다. 예를 들어 사용자의 멜로디의 7초에 해당하는 부분이 원곡의 14초에 해당하는 부분과 매칭된다면, 이 경우 사용자가 원곡보다 두 배 빠르게 불렀다고 추정한다. 또한 사용자가 입력한 끝점과 매칭되는 원곡의 시점을 찾아 그 부분부터 반주를 재생시켜 주게 되면 유저가 계속해서 노래를 부를 수 있게 할 수 있으며, 이 때 입력이 모두 들어온 시점부터 최종 phase vocoder 연산까지 끝나는 대략적인 시간을 고려하여 약간 뒷부분부터 재생시켜주면 보다 정확한 성능을 기대할 수 있다.

산업상 이용가능성

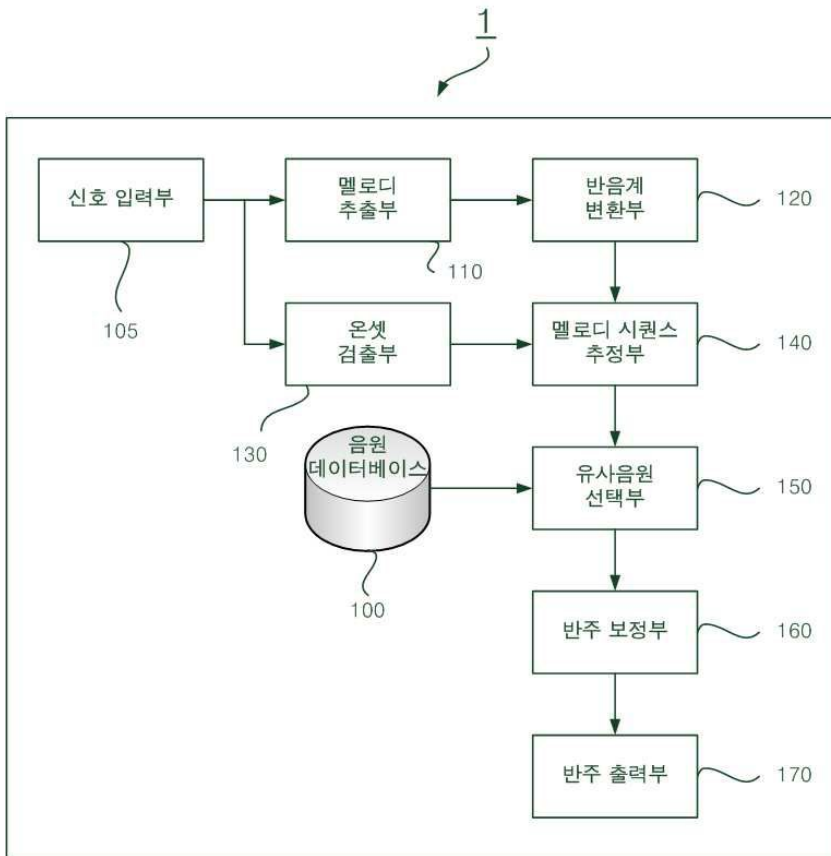
[0088] 본 발명에 따른 반주 제공 장치 및 반주 제공 방법은 노래방용 반주제공장치 등에 널리 사용될 수 있다

부호의 설명

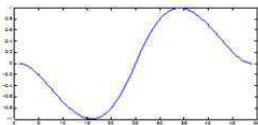
- [0089]
- 1 : 반주제공장치
 - 100 : 음원데이터베이스
 - 105 : 신호 입력부
 - 110 : 멜로디 추출부
 - 120 : 반음계 변환부
 - 130 : 온셋 검출부
 - 140 : 멜로디 시퀀스 추정부
 - 150 : 유사 음원 선택부
 - 160 : 반주 보정부
 - 170 : 반주 출력부

도면

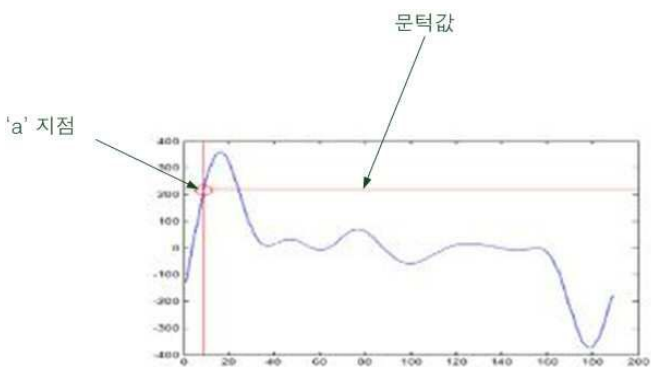
도면1



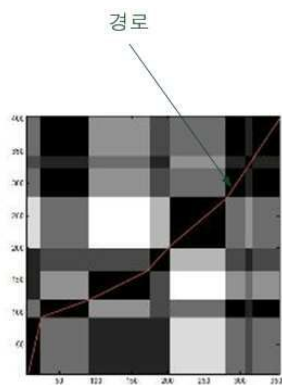
도면2



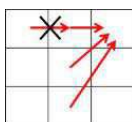
도면3



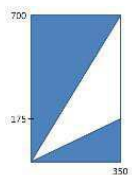
도면4



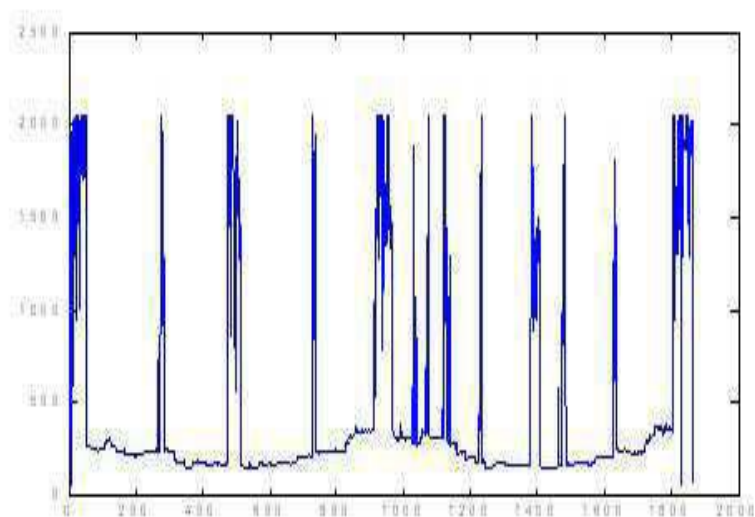
도면5



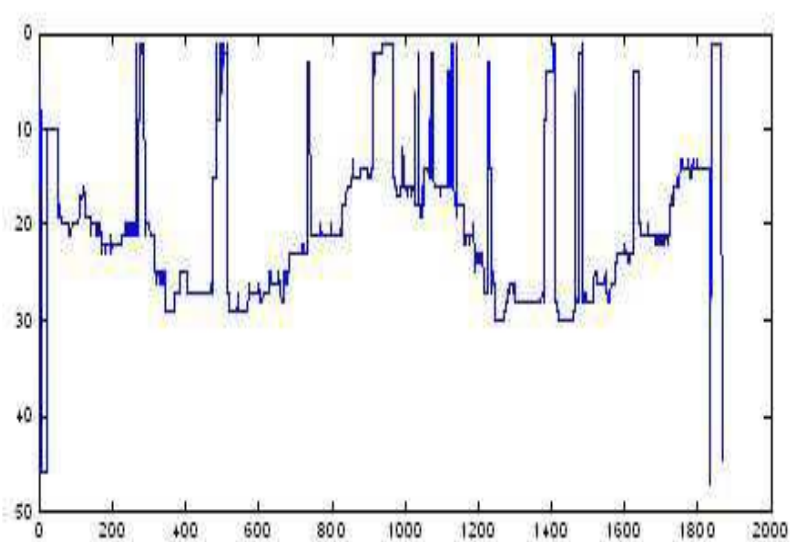
도면6



도면7



도면8



도면9

