(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2008-0071746
(43) 공개일자 2008년08월05일

(51) Int. Cl.
G10L 13/02 (2006.01) *G10L 15/04* (2006.01)
G10L 13/00 (2006.01) *G10L 21/06* (2006.01)
(21) 출원번호 10-2007-0010008
(22) 출원일자 2007년01월31일
심사청구일자 2007년01월31일

(71) 출원인
연세대학교 산학협력단
서울 서대문구 신촌동 134 연세대학교
(72) 발명자
이인권
서울 양천구 목6동 한신청구아파트 107동 401호
노창환
서울 서초구 서초1동 1427-7 한승미메이드아파트 가-401
유민준
경기 성남시 중원구 중동 삼창아파트 7동 202호
(74) 대리인
백남훈, 이학수

전체 청구항 수 : 총 5 항

(54) EM 최적화 방법을 이용한 오디오 텍스처 합성 방법

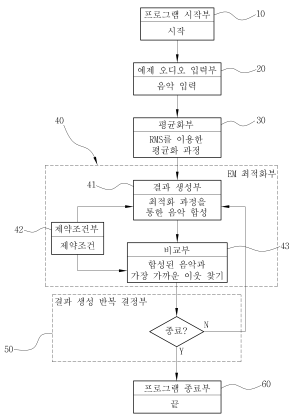
(57) 요 약

본 발명은 오디오 텍스처 합성 방법에 관한 것으로서, 더욱 상세하게는 EM 최적화 방법을 이용하여 최적화된 결과물을 만들어 낼 수 있는 오디오 텍스처 합성 방법에 관한 것이다.

이러한 본 발명의 오디오 텍스처 합성 방법은, 예제 오디오 입력부에 예제 오디오 텍스처가 입력되면 평균화부가 예제 오디오 텍스처의 진폭값 데이터에 대해 RMS를 통한 평균화를 수행하여 제공하는 단계와; EM 최적화부의 결과 생성부가 평균화 후 제공된 진폭값 데이터를 일정 단위로 나누고 이렇게 나눠진 부분들을 일정 길이만큼 서로 겹쳐지게 합성하여 임의의 길이의 최적화된 결과 오디오 텍스처를 만드는 E 단계와; 상기 E 단계에서 만들어진 결과 오디오 텍스처를 EM 최적화부의 비교부가 예제 오디오 텍스처와 부분 부분 비교하여 확장된 결과 오디오 텍스처와 유사한 부분을 예제 오디오 텍스처에서 찾는 M 단계와; 상기 E, M 단계를 반복하여 최적화된 최종의 결과 오디오 텍스처를 생성하는 단계;를 포함하여 구성된다.

상기와 같은 본 발명의 오디오 텍스처 합성 방법에 의하면, 최적화 방법을 이용하여 예제 오디오를 직접 합성함으로써 반복성을 줄이면서도 원하는 길이를 갖는 오디오를 생성할 수 있게 된다. 특히, 본 발명에서는 EM 최적화를 이용함으로써 정확한 합성이 이루어지도록 하면서 원본과 유사한 느낌을 갖게 하는 최적화된 결과 오디오를 생성할 수 있게 된다.

대표도 - 도1



특허청구의 범위

청구항 1

주어진 짧은 예제 오디오 텍스처를 합성하여 원하는 길이의 오디오 텍스처를 생성하는 오디오 텍스처 합성 방법에 있어서,

예제 오디오 입력부에 예제 오디오 텍스처가 입력되면 평균화부가 예제 오디오 텍스처의 진폭값 데이터에 대해 RMS를 통한 평균화를 수행하여 제공하는 단계와;

EM 최적화부의 결과 생성부가 평균화 후 제공된 진폭값 데이터를 일정 단위로 나누고 이렇게 나눠진 부분들을 일정 길이만큼 서로 겹쳐지게 합성하여 임의의 길이의 최적화된 결과 오디오 텍스처를 만드는 E 단계와;

상기 E 단계에서 만들어진 결과 오디오 텍스처를 EM 최적화부의 비교부가 예제 오디오 텍스처와 부분 부분 비교하여 확장된 결과 오디오 텍스처와 유사한 부분을 예제 오디오 텍스처에서 찾는 M 단계와;

상기 E, M 단계를 반복하여 최적화된 최종의 결과 오디오 텍스처를 생성하는 단계;

를 포함하여 구성되는 것을 특징으로 하는 EM 최적화 방법을 이용한 오디오 텍스처 합성 방법.

청구항 2

청구항 1에 있어서,

상기 RMS를 통한 평균화를 수행하는 단계에서, 상기 예제 오디오 텍스처의 진폭값 데이터를 일정 시간 혹은 일정 개수 단위로 평균화하는 것을 특징으로 하는 EM 최적화 방법을 이용한 오디오 텍스처 합성 방법.

청구항 3

청구항 1에 있어서,

상기 E 단계는,

평균화 후 제공된 진폭값 데이터를 여러 개의 부분벡터 z_p 로 나눈 뒤 나누어진 부분벡터 z_p를 임의의 순서로 일정 길이만큼 겹쳐지게 나열하는 단계와;

에너지 E를 정의한 하기 식(E1)의 선형 방정식을 이용하여 에너지 E가 최소화되는 부분벡터 x_p 및 합성된 결과 오디오 텍스처 X 를 구한 뒤 상기 M 단계로 제공하는 단계;

를 포함하는 것을 특징으로 하는 EM 최적화 방법을 이용한 오디오 텍스처 합성 방법.

식(E1) :

$$E_t(x;\{z_p\}) = \sum_{p \in X^\dagger} \left\| x_p - z_p \right\|^2$$

(여기서, X 를 합성된 결과 오디오 텍스처로, Z 를 예제 오디오 텍스처로, x와 z을 각각 X와 Z의 전체 샘플의 진폭값으로, N_p를 p 를 중심으로 w만큼의 너비 안에 포함된 주변 진폭값들을 나타내는 이웃 하나로 각각 정의할 때, 상기 x_p 는 N_p에 대응하는 x 의 부분벡터(진폭값)들이고, 상기 z_p는 N_p에 대응하는 z의 부분벡터들이며, 상기 $X^\dagger$ 는 X 에 대한 부분집합을 의미하는 것임.)

청구항 4

청구항 3에 있어서,

상기 M 단계에서는 상기 E 단계로부터 구해진 x_p 를 상기 z_p 와 비교하여 x_p와 가장 유사한 z_p의 인덱스 리스트를 구하며,

이후 E, M 단계를 반복하여 최적화된 최종의 결과 오디오 텍스처를 생성하는 단계에서는, 나누어진 부분벡터 z_p 를 상기 인덱스 리스트의 순으로 겹쳐지게 나열하여 상기 식(E1)을 이용하는 동일한 방식으로 x_p 및 합성된 결과 오디오 텍스처 X 를 새로이 구하는 E 단계와, 새로이 구해진 x_p 로부터 인덱스 리스트를 새로이 구하는 상기 M 단계를 반복하되, 새로이 생성된 인덱스 리스트가 이전 인덱스 리스트와 차이가 없는 수렴상태가 된 경우, 혹은 설정된 횟수로 E, M 단계를 반복한 경우, 최종적으로 구해진 결과 오디오 텍스처 X 를 최종의 결과물로 하여 과정을 종료하는 것을 특징으로 하는 EM 최적화 방법을 이용한 오디오 텍스처 합성 방법.

청구항 5

청구항 1 또는 청구항 4에 있어서,

상기 E, M 단계에서, 손실된 음악을 복구할 경우에 손실 부분을 배제한 상태에서 나머지 부분의 데이터를 이웃으로 가져와야 하는 제약조건과, 손실된 부분과 그 주위의 임의의 부분을 제외한 나머지 부분은 고정시켜야 하는 제약조건을 적용하는 것을 특징으로 하는 EM 최적화 방법을 이용한 오디오 텍스처 합성 방법.

명 세 서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

<12> 본 발명은 오디오 텍스처 합성 방법에 관한 것으로서, 더욱 상세하게는 EM 최적화 방법을 이용하여 최적화된 결과물을 만들어 낼 수 있는 오디오 텍스처 합성 방법에 관한 것이다.

<13> 오늘날 오디오 미디어는 게임, 애니메이션 혹은 영화의 배경음악과 특수음향을 비롯한 거의 대부분의 멀티미디어 매체에서 중요한 역할을 하는 요소 중에 하나라고 할 수 있다. 주로 비디오와 같은 시각적인 효과에 맞물려져 재생되거나 혹은 멀티미디어 매체 자체 내에서 중심적인 역할을 담당한다.

<14> 예를 들어, 게임의 경우 특수효과와 같은 시각적인 요소도 중요하지만 배경음악이 덧붙여진다면 이를 이용하여 주인공이 위치하고 있는 장소에 대한 정보를 전달하여 주거나 분위기에 맞는 오디오 미디어를 제공하여 줌으로써 사용자에게 효과적인 역할을 수행할 수 있다. 영화에서의 역할도 마찬가지이다. 특히, 사용자들에게 청각적인 요소만이 받아들여질 수 있는 상황이라면 더더욱 중요한 역할을 담당할 수 있게 된다.

<15> 이러한 멀티미디어 매체 내 요소들의 중요한 결합 인자는 각각 요소들의 시간의 동기화이다. 즉, 각각의 멀티미디어들이 적절히 동기화된 길이를 갖고 서로 연관된 시간에 재생이 되어야만 사용자의 몰입도가 증가하게 된다.

<16> 게임이나 애니메이션, 영화에서의 배경음악은 사용자의 몰입도를 높이는데 있어서 빠질 수 없는 중요한 요소이며, 앞서 설명한 바와 같이 분위기를 만들어 주고 현재의 환경을 시각적으로 전달하여 주는 효과가 있다.

<17> 하지만, 거의 대부분의 경우에 배경음악은 단순히 재생/멈춤 방식으로 되어 있다. 즉, 미리 정의된 여러 음악들이 단순히 재생이 되고 재생이 끝나면 다음 곡으로 넘어가거나 혹은 반복이 되는 형태로 되어 있으며, 따라서 배경음악은 때때로 단조로운 느낌을 주게 되면서 매우 지루해지게 된다.

<18> 이를 해결하기 위한 가장 간단한 방법은 오디오의 길이를 변화시키는 것이다. 이전에 제안된 오디오의 길이를 변화시키는 방법은 원본 오디오를 여러 개의 마디로 나누고 이들 간에 이동할 수 있는 확률 그래프를 만든 후 그래프 간의 연결을 통하여 임의의 길이를 가지는 오디오를 만드는 방법이다.

<19> 이러한 방법은 짧은 원본 오디오를 가지고도 임의의 길이를 가진 오디오를 만들어 낼 수 있지만, 역시 기존의 방법인 복사와 붙이기 방법에서 크게 벗어나지 못하기 때문에 반복적인 느낌을 준다.

<20> 그 밖에 관련 특허 또는 문헌을 살펴보면, 공개특허 제2003-83655호(2003.10.30)는 휴대용 이동통신단말기의 벨소리 파일과 음악 파일의 합성 및 변환 방법에 관한 것으로서, 음악에 대한 여러 가지 정보를 조절하여 들려주는 기술을 개시하고 있으며, 이를 위하여 제공자가 수동적으로 음에 관한 여러 가지 정보를 수록하도록 하고 있다.

<21> 그리고, 공개특허 제2005-57372호(2005.06.16)는 기존의 합성 방법을 개선하여 음성 혹은 음악을 합성하는 방법을 제안하고 있다.

<22> 또한 Lie Lu 등은 주어진 예제 오디오를 조각으로 나누어 이것들을 조합하여 음악을 합성하는 기술을 제안한 바 있으나[Audio Textures / Lie Lu, Liu Wenyin, and Hong-Jiang Zhang / IEEE Transactions on Speech and Audio Processing, Vol.15 No.2 pp. 156-167, 2004.], 이 또한 반복감을 느끼게 하는 문제점을 개선하지 못하였으며, 음의 흐름을 조절할 수는 없다.

발명이 이루고자 하는 기술적 과제

<23> 따라서, 본 발명은 상기와 같은 문제점을 해결하기 위하여 발명한 것으로서, EM 최적화 방법을 이용하여 예제 오디오를 직접 합성함으로써 반복성을 줄이면서도 임의의 길이로 오디오를 합성할 수 있고 원본과 유사한 느낌을 갖게 하는 결과 오디오를 생성할 수 있는 방법, 이를 응용하여 손상된 오디오 예제를 복구하는 방법, 및 음의 전체적인 흐름을 조절할 수 있는 방법을 제공하는데 그 목적이 있다.

발명의 구성 및 작용

<24> 이하, 첨부한 도면을 참조하여 본 발명을 상세히 설명하면 다음과 같다.

<25> 상기한 목적을 달성하기 위해, 본 발명은, 주어진 짧은 예제 오디오 텍스처를 합성하여 원하는 길이의 오디오 텍스처를 생성하는 오디오 텍스처 합성 방법에 있어서, 예제 오디오 입력부에 예제 오디오 텍스처가 입력되면 평균화부가 예제 오디오 텍스처의 진폭값 데이터에 대해 RMS를 통한 평균화를 수행하여 제공하는 단계와; EM 최적화부의 결과 생성부가 평균화 후 제공된 진폭값 데이터를 일정 단위로 나누고 이렇게 나눠진 부분들을 일정 길이만큼 서로 겹쳐지게 합성하여 임의의 길이의 최적화된 결과 오디오 텍스처를 만드는 E 단계와; 상기 E 단계에서 만들어진 결과 오디오 텍스처를 EM 최적화부의 비교부가 예제 오디오 텍스처와 부분 부분 비교하여 확장된 결과 오디오 텍스처와 유사한 부분을 예제 오디오 텍스처에서 찾는 M 단계와; 상기 E, M 단계를 반복하여 최적화된 최종의 결과 오디오 텍스처를 생성하는 단계;를 포함하여 구성되는 것을 특징으로 하는 EM 최적화 방법을 이용한 오디오 텍스처 합성 방법을 제공한다.

<26> 바람직하게는, 상기 RMS를 통한 평균화를 수행하는 단계에서, 상기 예제 오디오 텍스처의 진폭값 데이터를 일정 시간 혹은 일정 개수 단위로 평균화하는 것을 특징으로 한다.

<27> 또한 상기 E 단계는, 평균화 후 제공된 진폭값 데이터를 여러 개의 부분벡터 z_p로 나눈 뒤 나누어진 부분벡터 z_p를 임의의 순서로 일정 길이만큼 겹쳐지게 나열하는 단계와; 에너지 E를 정의한 하기 식(E1)의 선형 방정식을 이용하여 에너지 E가 최소화되는 부분벡터 x_p 및 합성된 결과 오디오 텍스처 X를 구한 뒤 상기 M 단계로 제공하는 단계;를 포함하는 것을 특징으로 한다.

<28> 식(E1) :

$$E_t(x;\{z_p\}) = \sum_{p\in X^\dagger} \left\| x_p - z_p \right\|^2$$

<29> (여기서, X를 합성된 결과 오디오 텍스처로, Z를 예제 오디오 텍스처로, x와 z을 각각 X와 Z의 전체 샘플의 진폭값으로, N_p를 p를 중심으로 w만큼의 너비 안에 포함된 주변 진폭값들을 나타내는 이웃 하나로 각각 정의할 때, 상기 x_p는 N_p에 대응하는 x의 부분벡터(진폭값)들이고, 상기 z_p는 N_p에 대응하는 z의 부분벡터들이며, 상기 $X^\dagger$는 X에 대한 부분집합을 의미하는 것임.)

<30> 그리고, 상기 M 단계에서는 상기 E 단계로부터 구해진 x_p를 상기 z_p와 비교하여 x_p와 가장 유사한 z_p의 인덱스 리스트를 구하며; 이후 E, M 단계를 반복하여 최적화된 최종의 결과 오디오 텍스처를 생성하는 단계에서는, 나누어진 부분벡터 z_p를 상기 인덱스 리스트의 순으로 겹쳐지게 나열하여 상기 식(E1)을 이용하는

동일한 방식으로 x_p 및 합성된 결과 오디오 텍스처 X를 새로이 구하는 E 단계와, 새로이 구해진 x_p로부터 인덱스 리스트를 새로이 구하는 상기 M 단계를 반복하되, 새로이 생성된 인덱스 리스트가 이전 인덱스 리스트와 차이가 없는 수렴상태가 된 경우, 혹은 설정된 횟수로 E, M 단계를 반복한 경우, 최종적으로 구해진 결과 오디오 텍스처 X를 최종의 결과물로 하여 과정을 종료하는 것을 특징으로 한다.

<31> 그리고, 상기 E, M 단계에서, 손실된 음악을 복구할 경우에 손실 부분을 배제한 상태에서 나머지 부분의 데이터를 이웃으로 가져와야 하는 제약조건과, 손실된 부분과 그 주위의 임의의 부분을 제외한 나머지 부분은 고정시켜야 하는 제약조건을 적용하는 것을 특징으로 한다.

<32> 이하, 첨부한 도면을 참조하여 본 발명에 대해 더욱 상세히 설명하면 다음과 같다.

<33> 알려진 바와 같이, 주어진 짧은 예제 오디오 클립(텍스처)으로부터 임의의 길이를 갖는 새로운 오디오 클립을 생성하는 방법을 오디오 텍스처 합성이라 한다.

<34> 본 발명에서는 최적화 방법을 이용하여 예제 오디오를 직접 합성함으로써 반복성을 줄이면서도 원하는 길이를 갖는 오디오를 생성할 수 있는 방법을 제시한다. 특히, 본 발명에서는 EM 최적화를 이용함으로써 정확한 합성이 이루어지도록 하면서 원본과 유사한 느낌을 갖게 하는 최적화된 결과 오디오를 생성할 수 있는 방법을 제시한다.

<35> 상기 EM 최적화는 최적화에서 이용되는 데이터와 최적화의 변수를 모두 알지 못할 때 두 가지 단계의 최적화를 연속적으로 행함으로써 최종의 최적해를 찾는 방법이다. 이와 같이 최적화를 이용한 합성의 가장 큰 특징 중에 하나는 합성시에 제약조건을 주어 결과를 조정할 수 있다는 것이며, 따라서 본 발명에서는 기존의 방법에 비해 더욱 간단하게 사용자가 원하는 오디오 결과를 유도 및 생성할 수 있게 된다.

<36> 첨부한 도 1은 본 발명에 따른 합성 과정의 순서도로서, 본 발명에 따른 합성 과정을 수행하는 소프트웨어의 구성부와 각 구성부가 수행하는 과정을 보여주고 있다.

<37> 본 발명의 오디오 텍스처 합성 과정은 컴퓨터상에서 실행되는 소프트웨어에 의해 수행될 수 있는데, 이때 소프트웨어는 크게 프로그램 시작부(10), 예제 오디오 입력부(20), RMS를 이용한 평균화부(30), EM 최적화부(40), 결과 생성 반복 결정부(50), 프로그램 종료부(60)로 구성되며, 상기 EM 최적화부(40)는 최적화 과정에서 결과가 생성되는 결과 생성부(41), 제약조건부(42), 생성된 결과와 원본 오디오를 비교하는 비교부(43)를 포함한다.

<38> 도 1을 참조하면, 프로그램 시작부(10)에서 프로그램이 시작된다. 프로그램이 실행된 후 예제 오디오 입력부(20)에서 사용자가 합성하길 원하는 예제 오디오 클립(텍스처)을 입력하면 RMS를 이용한 평균화부(30)에서 평균화된다. 이후 EM 최적화부(40)에서 평균화된 값들을 기반으로 하여 최적화된 값이 얻어진다. 그리고, EM 최적화부(40)에 의해 최적화된 값에서 다시 좀 더 최적화된 값을 얻거나 혹은 수렴되어 종료시켜야 하는가의 여부를 결과 생성 반복 결정부(50)가 결정한다. 프로그램 종료는 프로그램 종료부(60)에서 이루어진다.

<39> 여기서, 평균화 과정에 대하여 우선 설명하면 다음과 같다.

<40> 오디오 시그널은 매우 짧은 시간 단위를 기준으로 하여 단위 시간마다 진폭값이 기록되는 형태로 되어 있다. 본 발명에서는 합성하고자 하는 예제 오디오 텍스처를 사용자가 예제 오디오 입력부(20)에 입력하게 되면 그 예제 오디오 텍스처에서 얻은 데이터를 합성 및 최적화하여 원하는 길이의 결과 오디오를 생성하게 되는데, 최초 입력되는 원본 오디오, 즉 상기 예제 오디오 텍스처의 오디오 시그널에서 진폭값이 기록되므로, 본 발명의 과정에서 사용되는 데이터는 예제 오디오 텍스처에서 얻은 진폭값 데이터가 된다. 이와 같이 사용자가 입력한 예제 오디오 텍스처의 진폭값 데이터를 사용함에 있어서, 진폭값이 기록되는 단위 시간의 간격은 굉장히 짧기 때문에, 이렇게 얻어지는 모든 진폭값 데이터를 그대로 사용하게 되면 많은 계산 비용이 필요하게 된다.

<41> 따라서, 최적화 과정의 계산 이전에 평균화 과정을 통하여 진폭값 데이터들을 일정 시간 혹은 일정 개수 단위로 평균화하여 데이터들의 수를 줄인다. 즉, 선 실시되는 평균화 과정을 통해 소정의 샘플링 비율로 평균화하여 얻은 진폭값 데이터만을 후 실시되는 최적화 과정에서 사용하는 것이다. 여기서, 평균화 과정을 통해 얻어지는 진폭값 데이터는 일정 시간 혹은 일정 개수 단위로 얻어지는 대표값으로서 설정된 단위 구간의 평균값이다. 이와 같이 이후 실시되는 최적화 과정에서 사용될 진폭값 데이터를, 평균화 과정을 통해 단위 구간의 대표값으로 얻기 위하여, 본 발명에서는 RMS를 이용한 평균화부(30)에서 오디오 0.1초의 간격으로 RMS(Root Mean Square)를 적용한 값을 구한 뒤 이 값을 직접 최적화 과정에 사용한다.

<42> z_p는 예제 오디오 텍스처의 진폭값이며, 최적화 과정의 계산시에는 이의 RMS 버전, 즉 평균화 과정을 통해 일정 시간 혹은 일정 개수 단위로 얻어지는 평균값인 z'_p를 사용하고, 이는 아래 식(1)과 같이 얻어진다. 실제 최종적으로 얻어지는 결과 오디오는 최종적으로 얻어진 인덱스 리스트에 따라 평균화되지 않은 원본 데이터(사용자가 입력한 예제 오디오 텍스처)를 직접 합성한 결과가 된다.

<43>
$$z'_p = \sqrt{\frac{1}{N}\sum_{i=0}^{N} z_p} \qquad (1)$$

<44> 이하, 상기와 같이 RMS를 이용한 평균화 과정을 통하여 얻어진 진폭값 데이터(z'_p)들을 사용하여 EM 최적화를 통한 음악 합성이 이루어지는 과정에 대하여 설명하면 다음과 같다.

<45> 이미지에서는 각 픽셀 정보 하나하나가 색이라는 정보로 인식되는 반면 사운드에서는 어느 정도 수 이상의 샘플링 값이 소리 정보로 인식되게 된다. 따라서, 본 발명에서는 어느 정도로 인지될 수 있는 음을 들려줄 수 있을 만큼의 시간 단위를 사용한다. 하나의 이웃에 대한 에너지는 결과 오디오 텍스처의 이웃과 예제 오디오 텍스처의 이웃과의 거리로 정의하며, 결국 전체 에너지는 이런 지역적 이웃들의 에너지의 합이라고 볼 수 있다.

<46> 첨부한 도 2는 본 발명에서 EM 합성 과정(E 단계)에 대한 그래프를 나타낸 도면으로서, 최상측의 도면은 예제 오디오 텍스처에서 평균화 과정을 통해 얻어진 진폭값 데이터(source)로서 상기 평균화부(30)에서 제공된 소스 데이터를 나타낸 것이며, 최하측의 도면은 본 발명의 EM 합성 및 최적화 과정을 거쳐서 원하는 길이로 생성된 최종 결과 오디오의 진폭값 데이터를 나타낸다.

<47> EM 합성 과정에서는 평균화 과정을 통해 얻어진 진폭값 데이터를 소스 데이터로 사용하여 이를 일정 단위의 균일한 길이로 나누고, 나누어진 각 세그먼트 데이터를 연결 및 합성(overlapping)한 뒤 그 결과를 최적화하여 최종의 해를 구하게 되며, 최종적으로 자연스럽게 연결된 결과 오디오가 생성되게 된다.

<48> 도 2를 참조하면, 마디마디로 나뉜 지역적 이웃의 값이 최적화 과정을 통해 적절히 섞여 전체 에너지의 합을 구한다는 것을 볼 수 있다. 이 에너지를 구하는 방법은 다음과 같다.

<49>
$$E_t(x;\{z_p\}) = \sum_{p\in X^\dagger} \left\| x_p - z_p \right\|^2 \qquad (2)$$

<50> X는 사용자가 원하는 결과, 즉 합성되는 오디오 텍스처를 의미하고, Z은 예제 오디오 텍스처를 의미하다고 할 때, x와 z은 각각 X와 Z를 벡터화한 것이다. 즉, X와 Z의 전체 샘플의 진폭값을 의미한다. N_p는 p를 중심으로 w만큼의 너비 안에 포함된 주변 진폭값들을 나타내는 이웃 하나를 의미한다. N_p에 대응하는 x의 부분벡터(진폭값)들은 x_p로 나타낸다. z_p는 마찬가지로 N_p에 대응하는 z의 부분벡터들을 나타낸다. 이렇게 정의한 후 상기 식(2)을 통해 에너지 E를 얻게 된다. $X^\dagger$는 X에 대한 부분집합을 의미한다. 이는 모든 진폭값에 대해 이웃을 구하는 것보다 부분집합에 대한 이웃을 가지고 결과를 구함으로써 중복성을 줄이고 계산시간을 줄이게 된다.

<51> EM 최적화 과정은 크게 두 단계로 나눌 수 있다.

<52> 우선, 첫 번째 E 단계에서는 최적화 과정의 결과를 생성하는 결과 생성부(41)가 최적화 과정을 통하여 음악 합성을 수행하며, 여기서 최적화된 결과 오디오 텍스처 X를 구하게 된다. 도 2에서 위로부터 첫 번째 도면은 RMS를 이용한 평균화부(30)에서 제공된 데이터를 나타낸 것으로, 이 평균화된 데이터를 두 번째 도면과 같이 일정 단위로 여러 개의 부분벡터 z_p로 나눈다. 그리고, 나누어진 부분벡터 z_p를 일정 길이만큼 겹쳐지게 적절히 나열시킨다. 처음 이 단계를 거칠 경우 임의의 순서로 나열시킨다. X를 구하기 위하여 적절히 나열된 값들

을 가지고 선형 방정식, 즉 상기 식(2)를 이용하여 x_p를 구한다. 이것이 의미하는 바는 E를 최소화하는 형태의 x를 구하는 것인데, 이것은 좀 더 원본 오디오 텍스처 z에 가까운 형태로 만들겠다는 것을 의미한다.

<53> 다음으로, 두 번째 M 단계, 즉 생성된 결과와 원본 오디오를 비교하는 단계에서 비교부(43)는 x_p와 유사한 부분을 원본에서 찾기 위하여 z_p와 비교하여 x_p가 z_p의 어느 부분과 가장 유사한가에 대한 인덱스 리스트를 구하게 된다. 이 단계는 결과 오디오 텍스처의 이웃들과 가장 가까운 이웃을 예제 오디오 텍스처에서 찾는 문제라고 볼 수 있다. 여기서, E 단계를 통해 최적화된 X에 따라 대응되는 z_p는 조금씩 변하게 되고, 에너지 E는 줄어들게 된다.

<54> 가장 가까운 이웃들의 인덱스 리스트라고 할 수 있는 z_p의 인덱스 리스트를 구하고 나면 다시 E 단계로 돌아가 임의의 순서가 아닌 상기 인덱스 리스트의 순으로 또다시 적절히 겹쳐지게 나열하여 결과 오디오 텍스처 X를 새로이 만들어낸다. 그 결과를 얻어내면 역시 마찬가지로 M 단계를 적용하여 새로운 z_p의 인덱스 리스트를 만들어낸다. 그리고, 다시 E 단계를 거쳐 z_p에 대응되는 최적화된 X를 생성하고, 또다시 X는 바뀌게 되므로 M 단계에서 생성되는 z_p의 인덱스 리스트는 계속 새로운 값을 가지게 된다.

<55> 상기 과정에서 E, M 단계가 차례로 한 번씩 실행될 때마다 M 단계 이후 결과 생성 반복 결정부(50)로 넘어가는데, 여기서 새로이 생성된 인덱스 리스트가 이전 인덱스 리스트와 차이가 없는 수렴상태가 되었다고 판단되거나 혹은 설정 횟수로 E, M 단계를 반복하였다고 판단되면 프로그램 종료부(60)로 이동하여 프로그램을 종료하게 된다.

<56> 첨부한 도 4는 예제 오디오와 위의 과정을 통해 얻어진 결과물을 보여준다.

<57> 추가로 가우시안 폴-오프 평션(Gaussian fall-off function)을 통해서 얻어진 확률을 하나의 이웃 내에서 각각의 진폭값에 대한 가중치로 적용하였다. 이와 같이 가중치를 적용한 이유는 이웃의 중심으로 갈수록 보다 데이터가 결정되는데 중요한 역할을 하기 때문이다. 이 방법은 보통 두 개의 오디오 클립을 연결시킬 때 쓰이는 교차 페이드(Cross Fade) 효과를 적용한 결과와 유사한 결과를 기대할 수 있다. 또한 이를 응용하여 전혀 다른 이웃들 간의 결합도 가능하게 할 수 있다.

<58> 본 발명에서는 EM 최적화부(40)가 첨부한 도 3에서와 같이 다단계 합성 과정을 수행하도록 구성된다. 먼저, 예제 오디오 클립을 낮은 샘플링 비율에서 합성을 한 후 이 결과를 보간을 통해 업샘플링하여 샘플링 비율을 높인다. 이후 업샘플링된 결과를 가지고 다시 EM 최적화부(40)를 거치는 식으로 계속 이 과정을 반복한다. 또한 샘플링 비율 증가하는데 반비례하여 이웃의 크기를 축소함으로써 이웃 내의 데이터의 개수를 조절해 나간다.

<59> 그리고, EM 최적화부(40)에 제약조건부(42)를 도입하여 예제 오디오 클립의 손실된 부분을 복구하거나 혹은 오디오의 특정 위치에서 사용자가 원하는 부분의 음악을 재생할 수 있도록 하는데, 이로써 특정 흐름을 만들어낼 수 있게 된다. 제약조건은 도 1에서와 같이 EM 최적화부(40)에서 이루어지는 모든 과정에 항상 적용된다.

<60> 손실된 음악을 복구할 경우에서 제약조건부(42)에는 제약조건으로 두 가지가 부여된다. 첫 번째는 손실된 음악을 입력으로 줄 때 손실된 부분은 이웃으로 가져오면 안 된다는 제약조건을 주어야 한다. 이는 가장 가까운 이웃 z_p들을 탐색할 때 전체 탐색을 통해서 가져오기 때문에 만약 손실 부분에 대한 배제 없이 이웃들을 탐색할 시에는 손실된 부분에 가장 잘 매칭이 잘 되는 부분이 원래 손실된 부분이 되는 문제가 생길 수 있다. 그렇기 때문에 손실 부분을 배제한 상태에서 나머지 부분의 데이터를 이웃으로 가져와야 한다.

<61> 다음 제약조건은 손실된 부분과 그 주위의 임의의 부분을 제외한 나머지 부분들을 고정시켜야 한다는 것이다. 손실된 부분을 동시에 복원될 결과로 가정해야 하는데, 손실 부분을 제외한 나머지 부분은 고정시켜야 한다. 고정시키지 않는다면 합성할 시에 역시 전체 탐색을 통해 원본 데이터의 구조가 바뀌는 현상이 있을 수 있기 때문이다. 또한 추가적으로 손실 부분의 주변 값들은 고정시키지 말고 겹쳐질 수 있는 정도의 크기로 놔둔다. 겹쳐지는 부분이 없을 경우에는 피치가 끊어지는 문제가 발생하여 비정상적인 음을 들려줄 수 있다. 그렇기 때문에

일정 부분이 겹쳐져야 보다 자연스러운 결과를 들려 줄 수 있다.

<62> 이런 제약조건을 주게 되면 결국 손실되지 않은 부분에서 가장 적합한 이음새를 지닌 부분이 추출되고, 그 부분을 메워 주게 되어 손실 부분을 복구해준다. 초기에 손실 부분에는 임의의 값들을 넣어 줌으로써 보다 자연스럽게 합성되게 할 수 있다. 첨부한 도 5는 원본 오디오, 일부가 손실된 오디오 그리고 복구된 오디오의 그래프의 예를 보여준다.

<63> 본 발명에 따른 합성 방법을 활용하면 예제 오디오 텍스처를 이용하여 사용자가 원하는 흐름의 형태로 오디오를 재생할 수 있게 된다. 예를 들어 예제 오디오 텍스처 안에 비가 천천히 내리거나 혹은 많이 내릴 때의 소리가 포함되어 있다면 이를 이용하여 빗소리가 조금씩 비가 내리다가 점점 많이 오게 하는 비의 흐름을 유도할 수 있다. 반대로도 역시 가능하고 점점 천천히 혹은 조금 내렸다 다시 많이 내리고 다시 조금 내리거나 하는 식의 복잡하고 다양한 형태로도 만들 수 있다.

<64> 이런 접근은 제약조건부(42)에 일부를 고정시키는 제약조건을 줌으로써 유도해 낼 수 있다. 가령 천천히 비가 내리다 점점 많이 내리는 예제 오디오 클립이 있다고 가정하고 합성된 결과는 예제 오디오 클립과 역으로 비가 많이 내리는 소리를 점점 천천히 비가 내리는 소리로 만들고 싶다면 사용자는 예제의 비가 많이 내리는 부분을 결과의 초기값 앞부분에 두고 고정시키고 조금씩 내리는 빗소리는 뒷부분에 배치하여 고정시키며 나머지 부분에 대해서는 앞서 제시한 알고리즘을 그대로 적용시켜 합성을 하게 되면 거친 빗소리에서 천천히 내리는 빗소리를 들려줄 수 있는 데이터가 생성된다. 물론, 고정시키는 부분들의 거리가 멀게 되면 고정시키는 부분들 사이에 있는 부분들은 랜덤 혹은 적절한 값으로 바뀌게 된다. 또한 피치가 끊기는 문제로 인해 고정시키는 부분들을 너무 가까이 두면 듣기 좋지 않은 결과를 초래할 수 있다.

발명의 효과

<65> 이상에서 설명한 바와 같이, 본 발명에 따른 EM 최적화 방법을 이용한 오디오 텍스처 합성 방법에 의하면, 최적화 방법을 이용하여 예제 오디오를 직접 합성함으로써 반복성을 줄이면서도 원하는 길이를 갖는 오디오를 생성할 수 있게 된다. 게임과 같은 인터랙티브한 환경에서 사용자의 몰입성을 증가시키는데 효과적으로 사용될 수 있다.

<66> 특히, 본 발명에서는 EM 최적화를 이용함으로써 정확한 합성이 이루어지도록 하면서 원본과 유사한 느낌을 갖게 하는 최적화된 결과 오디오를 생성할 수 있게 된다. 기존에 미리 정의되어 있는 음악 소스를 단순히 반복, 재생하여 들려주는 것이 아니라, 원본 오디오를 합성 과정을 통해 원본과 유사한 느낌의 배경음악을 생성하여 무한한 길이의 배경음악을 제공함으로써 더욱 높은 수준의 디지털 콘텐츠 생성이 가능해진다. 정해진 재생시간에 구애받지 않는 게임뿐만 아니라 영화,애니메이션 등에서 보다 긴 재생시간의 배경음악이 필요한 모든 콘텐츠의 음악을 생성할 때도 유용하게 사용될 수 있다.

<67> 또한 최적화에 제약조건을 부여함으로써 손실된 음악의 복구가 가능하고 사용자가 음악의 흐름을 조절할 수 있게 된다.

도면의 간단한 설명

<1> 도 1은 본 발명에 따른 합성 과정의 순서도로서, 본 발명에 따른 합성 과정을 수행하는 소프트웨어의 구성부와 각 구성부가 수행하는 과정을 나타낸 도면,

<2> 도 2는 본 발명에서 EM 합성 과정의 두 단계에 대한 그래프를 나타낸 도면,

<3> 도 3은 본 발명에서 합성 결과의 향상을 위한 다단계 합성 과정의 그래프를 나타낸 도면,

<4> 도 4는 본 발명의 과정에서 원본과 합성 비교 그래프를 나타낸 도면,

<5> 도 5는 본 발명의 과정에서 손실된 오디오 복구 그래프를 나타낸 도면.

<6> <도면의 주요 부분에 대한 부호의 설명>

<7> 10 : 프로그램 시작부 20 ; 예제 오디오 입력부

<8> 30 : RMS를 이용한 평균화부 40 : EM 최적화부

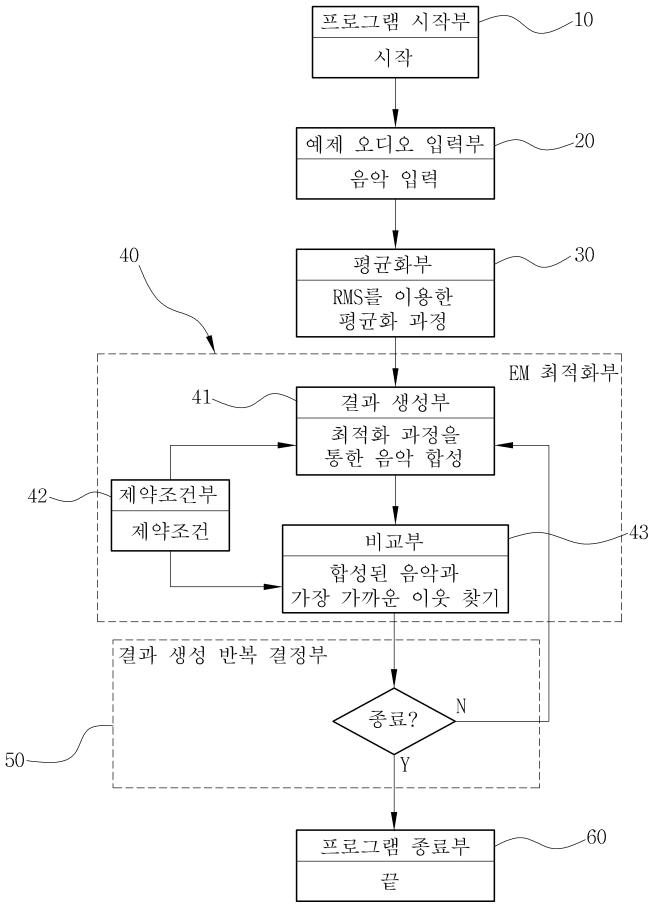
<9> 41 : 결과 생성부 42 : 제약조건부

<10> 43 : 비교부 50 : 결과 생성 반복 결정부

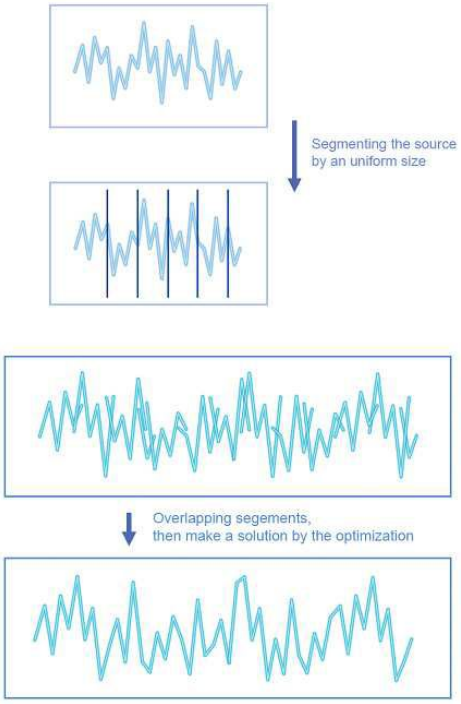
<11> 60 : 프로그램 종료부

도면

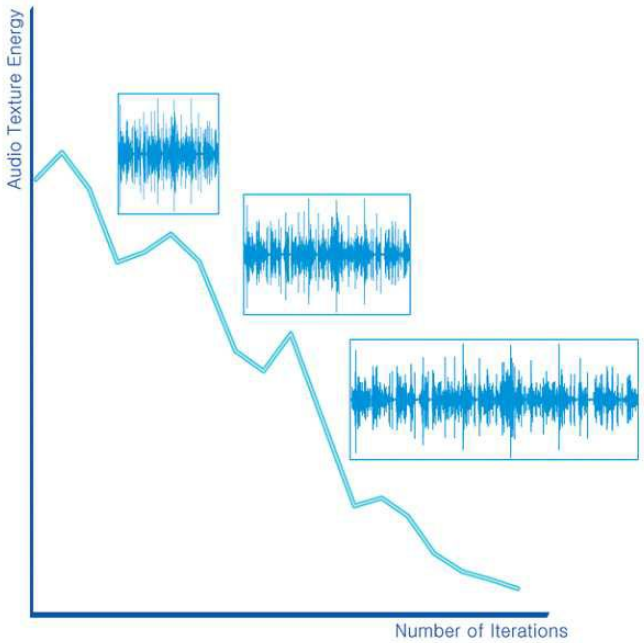
도면1



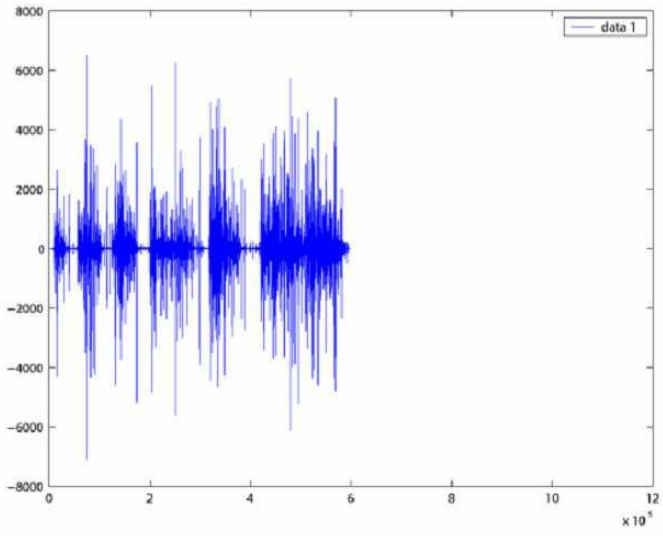
도면2


Segmenting the source
by an uniform size
Overlapping segements,
then make a solution by the optimization

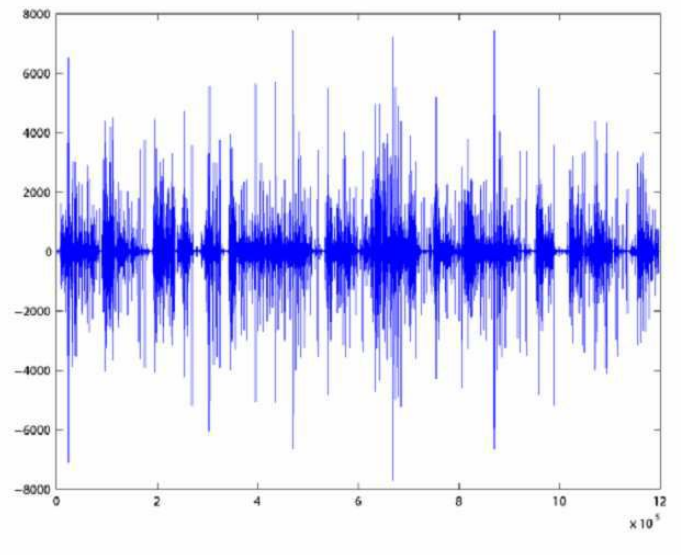
도면3


Audio Texture Energy
Number of Iterations

도면4

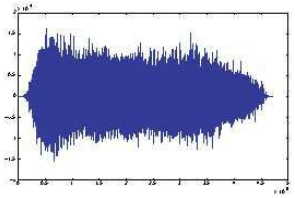


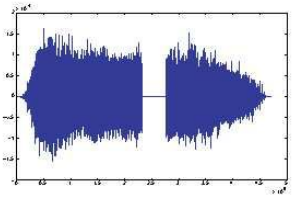
(a) 원본 오디오 텍스처



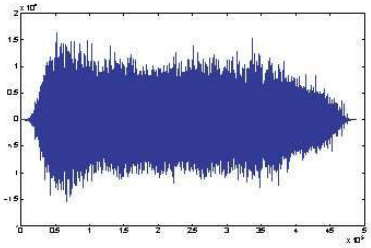
(b) 결과 오디오 텍스처

도면5





(a) 원본 오디오 텍스처 b) 일부가 손실된 오디오 텍스처



(c) 복구된 오디오 텍스처