



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2011년01월17일
(11) 등록번호 10-1008250
(24) 등록일자 2011년01월07일

(51) Int. Cl.

G10L 21/02 (2006.01) G10L 15/20 (2006.01)

(21) 출원번호 10-2005-7021034
(22) 출원일자(국제출원일자) 2004년05월26일
심사청구일자 2007년07월09일
(85) 번역문제출일자 2005년11월04일
(65) 공개번호 10-2006-0034637
(43) 공개일자 2006년04월24일
(86) 국제출원번호 PCT/JP2004/007587
(87) 국제공개번호 WO 2004/107319
국제공개일자 2004년12월09일

(30) 우선권주장

JP-P-2003-00154964 2003년05월30일 일본(JP)
JP-P-2003-00167118 2003년06월11일 일본(JP)

(56) 선행기술조사문헌

JP07199990 A*
US05204969 A1*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

도꾸리쯔교세이호진 상교기쥬쯔 소고겐쥬쥬
일본 도쿄도 치요다쿠 카스미가세키 1-3-1

(72) 발명자

고토 마사타카

일본국 이바라키켄 츠쿠바시 우메조노 1-1-1 쥬오
다이2도꾸리쯔교세이호진 상교기쥬쯔 소고겐쥬쥬
나이

(74) 대리인

하영욱, 하상구

전체 청구항 수 : 총 18 항

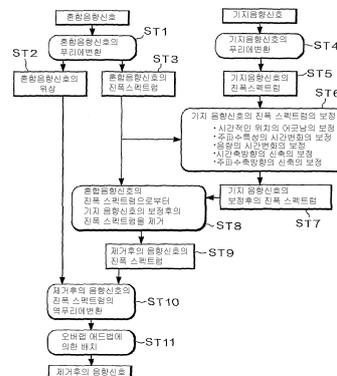
심사관 : 정성윤

(54) 기지 음향신호 제거방법 및 장치

(57) 요약

복수의 음향신호가 혼합된 음향신호를 입력으로 하고, 그 중 하나의 음향신호에 유사한 기지 음향신호가 부여되었을 때, 그 기지 음향신호를 제거하는 것을 가능하게 하는 기지 음향신호 제거장치가 제공된다. 이 기지 음향신호 제거장치는 입력된 혼합 음향신호(m(t))와 기지 음향신호(b'(t))를 각각 시간 주파수 영역에서의 진폭 스펙트럼(M(ω,t), B'(ω,t))으로 변환하고, M(ω,t)중의 B'(ω,t)에 대응하는 성분을 감산해서 제거함으로써, 제거후의 진폭 스펙트럼(S(ω,t))을 얻는다. 그 때, M(ω,t)중의 B'(ω,t)에 대응하는 성분은 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화 등의 요인에 의해 변형되어 있으므로, 이들을 보정한 B(ω,t)를 감산한다. 마지막으로, m(t)의 위상과, S(ω,t)를 이용해서 시간영역으로 역변환하여 원하는 제거후의 음향신호(s(t))를 얻는다.

대표도 - 도2



특허청구의 범위

청구항 1

복수의 음향신호가 혼합된 혼합 음향신호로부터 기지의 음향신호의 성분을 제거하는 기지 음향신호 제거방법으로서,

상기 혼합 음향신호를 시간주파수 표현으로 변환해서 상기 혼합 음향신호의 진폭 스펙트럼과 상기 혼합 음향신호의 위상을 구하는 혼합 음향신호 변환스텝;

상기 혼합 음향신호중에 포함되어 있는 기지의 음향신호에 해당하는 기지 음향신호를 시간주파수 표현으로 변환해서 상기 기지 음향신호의 진폭 스펙트럼을 구하는 기지 음향신호 변환스텝;

상기 혼합 음향신호의 진폭 스펙트럼을 입력으로 하여, 상기 혼합 음향신호의 진폭 스펙트럼에 대한 상기 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수방향의 신축을 보정한 상기 기지 음향신호의 보정 진폭 스펙트럼을 구하는 보정스텝;

상기 혼합 음향신호의 진폭 스펙트럼으로부터 상기 기지 음향신호의 보정 진폭 스펙트럼을 제거하는 제거스텝;

상기 제거스텝에 의해 얻은 제거후 진폭 스펙트럼과 상기 혼합 음향신호의 위상에 기초하여 시간표현으로 역변환을 행해서 단위파형을 구하는 역변환스텝; 및

상기 단위파형을 합성해서 상기 기지 음향신호의 성분을 제거한 음향신호를 얻는 합성스텝으로 이루어지고,

상기 보정스텝에서는,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간적인 위치를 추정하고, 추정된 상기 시간적인 위치에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간적인 위치의 어긋남을 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수특성의 시간변화를 추정하고, 추정된 상기 주파수특성의 시간변화에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수특성의 시간변화를 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 음량의 시간변화를 추정하고, 추정된 상기 음량의 시간변화에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 음량의 시간변화를 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간축방향의 신축을 추정하고, 추정된 상기 시간축방향의 신축에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간축방향의 신축을 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수축방향의 신축을 추정하고, 추정된 상기 주파수축방향의 신축에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수축방향의 신축을 보정하는 것을 특징으로 하는 기지 음향신호 제거방법.

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

삭제

청구항 6

삭제

청구항 7

제1항에 있어서, 상기 혼합 음향신호의 진폭 스펙트럼과 상기 기지 음향신호의 진폭 스펙트럼을 시각에 의해 대비할 수 있도록 화상표시하는 화상표시시스템; 및

상기 혼합 음향신호, 상기 기지 음향신호 및 상기 합성스텝의 출력신호를 음으로서 음향재생하는 음향재생스텝을 더 구비하고;

상기 화상표시와 상기 음향재생에 기초하여 상기 혼합 음향신호중에 있어서의 상기 기지의 음향신호가 포함되어 있는 구간이 정해지고;

상기 구간에 대해서 상기 보정스텝, 상기 제거스텝, 상기 역변환스텝 및 상기 합성스텝을 실행하는 것을 특징으로 하는 기지 음향신호 제거방법.

청구항 8

제1항에 있어서, 상기 혼합 음향신호의 진폭 스펙트럼에 기초하여 상기 혼합 음향신호중에 있어서의 상기 기지의 음향신호가 포함되어 있는 구간을 자동추정하고,

상기 구간에 대해서 상기 보정스텝, 상기 제거스텝, 상기 역변환스텝 및 상기 합성스텝을 실행하는 것을 특징으로 하는 기지 음향신호 제거방법.

청구항 9

제1항에 있어서, 상기 혼합 음향신호중에 포함되어 있는 상기 기지의 음향신호에 상당하는 복수의 상기 기지 음향신호가 존재하는 경우에,

상기 복수의 기지 음향신호의 전부에 관해서 상기 기지 음향신호 변환스텝 및 상기 보정스텝을 실행하고,

상기 혼합 음향신호의 진폭 스펙트럼으로부터 상기 복수의 기지 음향신호의 보정 진폭 스펙트럼을 전부 제거하는 제거스텝에 의해 얻은 제거후 진폭 스펙트럼을 이용하여, 상기 역변환스텝 및 상기 합성스텝을 실행하는 것을 특징으로 하는 기지 음향신호 제거방법.

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

청구항 13

삭제

청구항 14

복수의 음향신호가 혼합된 혼합 음향신호로부터 기지의 음향신호의 성분을 제거하는 기지 음향신호 제거장치로서,

상기 혼합 음향신호를 시간주파수 표현으로 변환해서 상기 혼합 음향신호의 진폭 스펙트럼과 상기 혼합 음향신호의 위상을 구하는 혼합 음향신호 변환수단;

상기 혼합 음향신호중에 포함되어 있는 기지의 음향신호에 상당하는 기지 음향신호를 시간주파수 표현으로 변환

해서 상기 기지 음향신호의 진폭 스펙트럼을 구하는 기지 음향신호 변환수단;

상기 혼합 음향신호의 진폭 스펙트럼을 입력으로 하여, 상기 혼합 음향신호의 진폭 스펙트럼에 대한 상기 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수방향의 신축을 보정한 상기 기지 음향신호의 보정 진폭 스펙트럼을 구하는 보정수단;

상기 혼합 음향신호의 진폭 스펙트럼으로부터 상기 기지 음향신호의 보정 진폭 스펙트럼을 제거하는 제거수단;

상기 제거수단에 의해 얻은 제거후 진폭 스펙트럼과 상기 혼합 음향신호의 위상에 기초하여 시간표현으로 역변환을 행해서 단위파형을 구하는 역변환수단; 및

상기 단위파형을 합성해서 상기 기지 음향신호의 성분을 제거한 음향신호를 얻는 합성수단으로 이루어지고,

상기 보정수단에서는,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간적인 위치를 추정하고, 추정된 상기 시간적인 위치에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간적인 위치의 어긋남을 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수특성의 시간변화를 추정하고, 추정된 상기 주파수특성의 시간변화에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수특성의 시간변화를 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 음량의 시간변화를 추정하고, 추정된 상기 음량의 시간변화에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 음량의 시간변화를 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간축방향의 신축을 추정하고, 추정된 상기 시간축방향의 신축에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간축방향의 신축을 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수축방향의 신축을 추정하고, 추정된 상기 주파수축방향의 신축에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수축방향의 신축을 보정하는 것을 특징으로 하는 기지 음향신호 제거장치.

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

복수의 음향신호가 혼합된 혼합 음향신호로부터 기지의 음향신호의 성분을 제거하는 처리를 컴퓨터에 의해 실행하기 위한 프로그램을 기록한 기록매체로서,

상기 혼합 음향신호를 시간주파수 표현으로 변환해서 상기 혼합 음향신호의 진폭 스펙트럼과 상기 혼합 음향신호의 위상을 구하는 혼합 음향신호 변환스텝;

상기 혼합 음향신호중에 포함되어 있는 기지의 음향신호에 상당하는 기지 음향신호를 시간주파수 표현으로 변환해서 상기 기지 음향신호의 진폭 스펙트럼을 구하는 기지 음향신호 변환스텝;

상기 혼합 음향신호의 진폭 스펙트럼을 입력으로 하여, 상기 혼합 음향신호의 진폭 스펙트럼에 대한 상기 기지

음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수방향의 신축을 보정한 상기 기지 음향신호의 보정 진폭 스펙트럼을 구하는 보정스텝;

상기 혼합 음향신호의 진폭 스펙트럼으로부터 상기 기지 음향신호의 보정 진폭 스펙트럼을 제거하는 제거스텝;

상기 제거스텝에 의해 얻은 제거후 진폭 스펙트럼과 상기 혼합 음향신호의 위상에 기초하여 시간표현으로 역변환을 행해서 단위파형을 구하는 역변환스텝; 및

상기 단위파형을 합성해서 상기 기지 음향신호의 성분을 제거한 음향신호를 얻는 합성스텝의 처리를 컴퓨터에 의해 실행시키고,

상기 보정스텝에서는,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간적인 위치를 추정하고, 추정된 상기 시간적인 위치에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간적인 위치의 어긋남을 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수특성의 시간변화를 추정하고, 추정된 상기 주파수특성의 시간변화에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수특성의 시간변화를 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 음량의 시간변화를 추정하고, 추정된 상기 음량의 시간변화에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 음량의 시간변화를 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간축방향의 신축을 추정하고, 추정된 상기 시간축방향의 신축에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 시간축방향의 신축을 보정하고,

상기 혼합 음향신호의 상기 진폭 스펙트럼에 포함되는 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수축방향의 신축을 추정하고, 추정된 상기 주파수축방향의 신축에 기초하여 상기 기지 음향신호의 상기 진폭 스펙트럼의 주파수축방향의 신축을 보정하는 것을 특징으로 하는 기지 음향신호 제거장치용 프로그램을 기록한 기록매체.

청구항 20

삭제

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

제19항에 있어서, 상기 혼합 음향신호의 상기 진폭 스펙트럼과 상기 기지 음향신호의 상기 진폭 스펙트럼을 시각에 의해 대비할 수 있도록 화상표시하는 화상표시시스템을 또한 컴퓨터에 의해 실행시키는 것을 특징으로 하는 기지 음향신호 제거장치용 프로그램을 기록한 기록매체.

청구항 26

제19항에 있어서, 상기 혼합 음향신호, 상기 기지 음향신호 및 상기 합성스텝의 출력신호를 음향으로서 재생하는 음향재생스텝을 또한 컴퓨터에 의해 실행시키는 것을 특징으로 하는 기지 음향신호 제거장치용 프로그램을 기록한 기록매체.

청구항 27

제19항에 있어서, 상기 혼합 음향신호의 상기 진폭 스펙트럼에 기초해서 상기 혼합 음향신호중에 있어서의 상기 기지의 음향신호가 포함되어 있는 구간을 자동추정하는 스텝의 처리를 컴퓨터에 의해 실행시키고,

상기 구간에 대해서 상기 보정스텝, 상기 제거스텝, 상기 역변환스텝 및 상기 합성스텝의 처리를 컴퓨터에 의해 실행시키는 것을 특징으로 하는 기지 음향신호 제거장치용 프로그램을 기록한 기록매체.

청구항 28

제19항에 있어서, 상기 혼합 음향신호중에 포함되어 있는 상기 기지의 음향신호에 상당하는 복수의 상기 기지 음향신호가 존재하는 경우에,

상기 복수의 기지 음향신호의 전부에 관해서 상기 기지 음향신호 변환스텝 및 상기 보정스텝을 상기 컴퓨터에 의해 실행시키고,

상기 혼합 음향신호의 상기 진폭 스펙트럼으로부터 상기 복수의 기지 음향신호의 보정 진폭 스펙트럼을 전부 제거하는 제거스텝에 의해 얻은 제거후 진폭 스펙트럼 을 이용하여, 상기 역변환스텝 및 상기 합성스텝을 상기 컴퓨터에 의해 실행시키는 것을 특징으로 하는 기지 음향신호 제거장치용 프로그램을 기록한 기록매체.

청구항 29

삭제

청구항 30

제1항에 있어서, 상기 보정스텝을 실행할 때, 상기 시간적인 위치의 어긋남, 상기 주파수특성의 시간변화, 상기 음량의 시간변화, 상기 시간축방향의 신축 및 상기 주파수축방향의 신축의 보정을 지정하는 것을 가능하게 하는 조작패널을 구비하는 인터페이스를 사용하는 것을 특징으로 하는 기지 음향신호 제거방법.

청구항 31

제30항에 있어서, 상기 인터페이스는 상기 혼합 음향신호의 진폭 스펙트럼과 상기 기지 음향신호의 진폭 스펙트럼을 시각에 의해 대비할 수 있도록 화상표시하는 화상표시부를 구비하고 있는 것을 특징으로 하는 기지 음향신호 제거방법.

청구항 32

제30항에 있어서, 상기 인터페이스는 상기 혼합 음향신호, 상기 기지 음향신호 및 상기 합성스텝의 출력신호를 음향으로서 재생하는 음향재생부를 구비하고 있는 것을 특징으로 하는 기지 음향신호 제거방법.

청구항 33

제30항에 있어서, 상기 인터페이스는, 상기 혼합 음향신호의 진폭 스펙트럼과 상기 기지 음향신호의 진폭 스펙트럼을 시각에 의해 대비할 수 있도록 화상표시하는 화상표시부; 및 상기 혼합 음향신호, 상기 기지 음향신호 및 상기 합성스텝의 출력신호를 음향으로서 재생하는 음향재생부를 구비하고 있는 것을 특징으로 하는 기지 음향신호 제거방법.

청구항 34

삭제

청구항 35

제14항에 있어서, 상기 보정수단은 상기 시간적인 위치의 어긋남, 상기 주파수특성의 시간변화, 상기 음량의 시간변화, 상기 시간축방향의 신축 및 상기 주파수축방향의 신축의 보정을 지정하는 것을 가능하게 하는 조작패널

을 구비하는 인터페이스를 구비하고 있는 것을 특징으로 하는 기지 음향신호 제거장치.

청구항 36

제35항에 있어서, 상기 인터페이스는, 상기 혼합 음향신호의 상기 진폭 스펙트럼과 상기 기지 음향신호의 상기 진폭 스펙트럼을 시각에 의해 대비할 수 있도록 화상표시하는 화상표시부; 및 상기 혼합 음향신호, 상기 기지 음향신호 및 상기 합성수단의 출력신호를 음향으로서 재생하는 음향재생부를 구비하고;

상기 화상표시부에 표시된 상기 혼합 음향신호의 상기 진폭 스펙트럼과 상기 기지 음향신호의 상기 진폭 스펙트럼과, 상기 음향재생부로부터의 재생음에 기초하여 상기 혼합 음향신호중에 포함되어 있는 상기 기지 음향신호의 구간의 지정과, 상기 기지 음향신호의 진폭 스펙트럼의 상기 시간적인 위치의 어긋남, 상기 주파수특성의 시간변화, 상기 음량의 시간변화, 상기 시간축방향의 신축 및 상기 주파수축방향의 신축의 보정의 지정을 행하기 위한 조작패널을 구비하고 있는 것을 특징으로 하는 기지 음향신호 제거장치.

청구항 37

제36항에 있어서, 상기 화상표시부는 상기 기지의 음향신호가 포함되어 있는 상기 혼합 음향신호중의 구간의 상기 진폭 스펙트럼; 및, 상기 혼합 음향신호중에 포함되어 있는 상기 기지 음향신호의 대응구간의 상기 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축을 보정한 보정 진폭 스펙트럼을 시간축상에서 위치를 맞춰서 표시할 수 있는 구성인 것을 특징으로 하는 기지 음향신호 제거장치.

청구항 38

제36항 또는 제37항에 있어서, 상기 화상표시부는 상기 혼합 음향신호의 상기 진폭 스펙트럼으로부터 상기 보정 진폭 스펙트럼을 제거한 음향신호의 진폭 스펙트럼을 화상표시할 수 있는 구성인 것을 특징으로 하는 기지 음향신호 제거장치.

청구항 39

삭제

명세서

기술분야

[0001] 본 발명은 복수의 음향신호가 혼합된 혼합 음향신호 중에서, 기지(known)의 음향신호의 성분을 제거하는 기지 음향신호 제거방법 및 기지 음향신호 제거장치에 관한 것이다.

배경기술

[0002] 종래부터, 음향신호처리로서, 스펙트럼 서브트랙션법(비특허문헌1)이라 불리는 방법이 알려져 있다. 종래의 스펙트럼 서브트랙션법은 정상잡음(스펙트럼이 시간적으로 변화되지 않고, 주파수특성이나 음량 등이 거의 일정한 잡음)과 원하는 음(타겟음)이 혼합된 음향신호(혼합음)로부터 정상잡음을 제거해서 타겟음을 얻는 방법이다. 이 방법에서는, 사전에 정상적인 스펙트럼의 평균을 구하는 등의 간단한 방법으로 정상잡음의 스펙트럼을 학습해 두고, 입력된 혼합음의 스펙트럼으로부터 정상잡음의 스펙트럼을 빼는 처리를 행한다. 즉, 잡음의 평균을 빼는 처리를 행한다.

[0003] 일반적으로, 음향신호 제거에 관해서는, 복수의 마이크로폰으로부터의 입력을 사용하는 방법이 다수 제안되고 있다. 또, 스펙트럼 서브트랙션법에는, 특허문헌1~7에 개시되어 있듯이, 여러가지 개량이 이루어져 있다.

[0004] [비특허문헌1]

[0005] Steven Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, No.2, April 1979.

[0006] [특허문헌1] 일본 특허공개 2002-175099호 공보

[0007] [특허문헌2] 일본 특허공개 2002-014694호 공보

- [0008] [특허문헌3] 일본 특허공개 2001-228892호 공보
- [0009] [특허문헌4] 일본 특허공개 2001-215992호 공보
- [0010] [특허문헌5] 일본 특허공개 평11-003094호 공보
- [0011] [특허문헌6] 일본 특허공개 평10-240294호 공보
- [0012] [특허문헌7] 일본 특허공개 평08-221092호 공보

발명의 상세한 설명

- [0013] 종래에 있어서의 스펙트럼 서브트랙션법은 정상잡음을 전제로 하고 있으며, 비정상잡음(스펙트럼이 시간적으로 크게 변화하고, 주파수특성이나 음량 등도 변화하는 잡음)에는 적용할 수 없었다. 예를 들면, 백그라운드 뮤직(BGM)으로서 사용되고 있는 음악과 같은 시간적으로 크게 변화되는 비정상잡음을 제거하는 것은 불가능했다. 이것은 비정상잡음의 스펙트럼의 변화가 지나치게 커서 학습을 할 수 없기 때문이다.
- [0014] 또, 가령, 종래의 방법에 의해 비정상잡음이 사전에 부여된 조건을 취급하려고 해도, 비정상잡음의 주파수특성, 음량, 진폭 스펙트럼의 시간축방향의 신축 및 주파수축방향의 신축 등의 변화의 영향에 의해, 빼는 처리를 적절하게 행할 수는 없었다. 복수의 마이크로폰으로부터의 입력을 사용하는 방법에서는, 모노럴(monaural) 음향신호에는 적용할 수 없었다. 개량된 종래의 스펙트럼 서브트랙션법의 어느 방법이나, 주로 음성인식의 전처리를 목적으로 하고 있기 때문에, 비정상잡음이 사전에 부여되어, 그 비정상잡음을 제거하는 용도에는 이용할 수 없었다.
- [0015] 따라서, 본 발명의 목적은, 복수의 음향신호가 혼합된 혼합 음향신호 중에서, 기지의 음향신호(비정상이어도 정상이어도 좋음)의 성분을, 그것에 대응하는 원음원으로부터의 기지 음향신호를 이용하여 제거할 수 있는 기지 음향신호 제거방법 및 기지 음향신호 제거장치 및 그 장치에 사용하는 프로그램이 기록된 기록매체를 제공하는 것에 있다.
- [0016] 또, 본 발명의 다른 목적은, 예를 들면 기지의 음향신호가 음악이며, 그 음악 음향신호가 인간의 음성이나 소리에 대한 백그라운드 뮤직(BGM)으로서 사용되고 있는 혼합음으로부터, 기지의 음향신호에 대응하는 원음원인 기지 음향신호(예를 들면 CD나 레코드 등으로부터 동일음악의 음향신호를 별도로 입수한 것)를 이용하여 BGM을 제거할 수 있는 기지 음향신호 제거방법 및 장치 및 그 장치에 사용하는 프로그램이 기록된 기록매체를 제공하는 것에 있다.
- [0017] 본 발명의 또 다른 목적은, 복수의 음향신호가 혼합된 음향신호(혼합음) 중에서, 기지 음향신호의 성분을 제거할 때, 혼합음중에서의 기지 음향신호의 정확한 위치를 자동추정하여, 그 위치의 기지의 음향신호를 제거할 수 있는 기지 음향신호 제거방법 및 장치 및 그 장치에 사용하는 프로그램이 기록된 기록매체를 제공하는 것에 있다.
- [0018] 본 발명의 또 다른 목적은, 복수의 음향신호가 혼합된 음향신호(혼합음) 중에서, 기지 음향신호의 성분을 제거할 때, 혼합음중에서의 기지 음향신호의 정확한 위치를 인간이 지정할 수 있는 인터페이스를 구비한 기지 음향신호 제거장치를 제공하는 것에 있다.
- [0019] 본 발명의 또 다른 목적은, 복수의 음향신호가 혼합된 음향신호(혼합음) 중에서, 기지 음향신호의 성분을 제거할 때, 혼합음중에서 기지 음향신호의 주파수특성이나 음량이 시간적으로 변화되고 있을 때, 이들 변화를 자동추정해서 보정하면서 제거할 수 있는 기지 음향신호 제거방법 및 장치 및 그 장치에 사용하는 프로그램이 기록된 기록매체를 제공하는 것에 있다.
- [0020] 본 발명의 또 다른 목적은, 복수의 음향신호가 혼합된 음향신호(혼합음) 중에서, 기지 음향신호의 성분을 제거할 때, 혼합음중에서 기지 음향신호의 주파수특성이나 음량이 시간적으로 변화되고 있을 때, 이들 변화를 인간이 지정할 수 있는 인터페이스를 구비한 기지 음향신호 제거장치를 제공하는 것에 있다.
- [0021] 본 발명의 또 다른 목적은, 복수의 음향신호가 혼합된 음향신호(혼합음) 중에서, 기지 음향신호의 성분을 제거할 때, 혼합음중에서 기지의 음향신호가 시간축 또는 주파수축방향으로 신축하고 있을 때, 이들 신축을 자동추정해서 보정하면서 제거할 수 있는 기지 음향신호 제거방법 및 장치 및 그 장치에 사용하는 프로그램이 기록된 기록매체를 제공하는 것에 있다.
- [0022] 본 발명의 또 다른 목적은, 복수의 음향신호가 혼합된 음향신호(혼합음) 중에서, 기지 음향신호의 성분을 제거

할 때, 혼합음중에서 기지의 음향신호가 시간축 또는 주파수축방향으로 신축하고 있을 때, 이들 신축을 인간이 지정할 수 있는 인터페이스를 구비한 기지 음향신호 제거장치를 제공하는 것에 있다.

- [0023] 본 발명의 또 다른 목적은, 복수의 음향신호가 혼합된 음향신호 중에서, 복수의 기지 음향신호의 성분을 제거할 때, 기지의 음향신호를 하나씩 반복해서 제거할 수 있도록 한 기지 음향신호 제거방법 및 장치 및 그 장치에 사용하는 프로그램이 기록된 기록매체를 제공하는 것에 있다.
- [0024] 본 발명에 의한 기지 음향신호 제거방법에 있어서는, 복수의 음향신호가 혼합된 혼합 음향신호로부터, 기지의 음향신호(비정상이어도 정상이어도 좋음)의 성분을, 그것에 대응하는 원음원으로부터의 기지 음향신호를 이용하여 제거한다.
- [0025] 이 때문에, 본 발명의 기지 음향신호 제거방법에서는, 우선, 혼합 음향신호를 시간주파수 표현으로 변환해서 혼합 음향신호의 진폭 스펙트럼과 혼합 음향신호의 위상을 구한다(혼합 음향신호 변환스텝). 여기에서의 음향신호를 시간주파수 표현으로 변환하는 방법으로서, 푸리에 변환이나 웨이브렛 변환 등 공지의 변환방법을 사용한다.
- [0026] 다음에, 혼합 음향신호중에 포함되어 있는 기지의 음향신호에 대응(유사)하고 있는 기지 음향신호(CD나 레코드 등으로부터 동일음악의 음향신호를 별도 입수한 것)를 시간주파수 표현으로 변환해서 기지 음향신호의 진폭 스펙트럼을 구한다(기지 음향신호 변환스텝).
- [0027] 계속해서, 구한 혼합 음향신호의 진폭 스펙트럼에 기초하여, 혼합 음향신호의 진폭 스펙트럼에 대한 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 적어도 하나를 보정한 상기 기지 음향신호의 보정 진폭 스펙트럼을 구한다(보정스텝).
- [0028] 다음에, 혼합 음향신호의 진폭 스펙트럼으로부터 기지 음향신호의 보정 진폭 스펙트럼을 제거한다(제거스텝). 이 제거스텝에 의해 얻은 제거후 진폭 스펙트럼과 혼합 음향신호의 위상에 기초하여 시간표현으로 역변환을 행해서 단위파형을 구한다(역변환스텝).
- [0029] 마지막으로, 단위파형을 오버랩 애드(Overlap Add)법 등의 합성방법을 이용하여 합성해서 기지 음향신호의 성분을 제거한 음향신호를 얻는다(합성스텝).
- [0030] 또, 본 발명의 기지 음향신호 제거방법에 있어서는, 다음과 같은 보정스텝을 실행함으로써, 혼합 음향신호의 진폭 스펙트럼에 대한 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 적어도 하나를 보정한 기지 음향신호의 보정 진폭 스펙트럼을 구하고, 이 보정 진폭 스펙트럼을 혼합 음향신호의 진폭 스펙트럼으로부터 제거한다. 이 때문에, 혼합 음향신호중에 비정상잡음으로서 포함되어 있는 기지 음향신호를 높은 정밀도로 제거할 수 있다.
- [0031] 원칙적으로는, 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축 중에서, 실제로 혼합 음향신호중에서 그 현상 또는 변화가 일어나고 있었던 것을 모두 보정하는 것이 바람직하다.
- [0032] 그러나, 아무것도 보정하지 않은 경우보다, 실제로 혼합 음향신호중에서 그 현상 또는 변화가 일어나고 있는 것의 하나라도 보정하면, 기지 음향신호의 제거 정밀도를 높일 수 있으므로, 보정의 전부를 행하지 않아도 좋다. 물론 필요한 보정의 전부를 행해도 좋다.
- [0033] 보정스텝의 실행에서는, 예를 들면 혼합 음향신호에 포함되는 기지 음향신호의 시간적인 위치를 추정하고, 추정된 시간적인 위치에 기초하여 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남을 보정한다. 추정방법은, 예를 들면 혼합 음향신호의 진폭 스펙트럼의 소정의 구간과 기지 음향신호의 진폭 스펙트럼의 소정의 구간의 거리(유사도)를 구하여, 거리가 가장 가까운 구간을 혼합 음향신호에 포함되는 기지 음향신호의 시간적인 위치로 추정한다.
- [0034] 또, 보정스텝의 실행에서는, 예를 들면 혼합 음향신호에 포함되는 기지 음향신호의 주파수특성의 변화를 추정하고, 추정된 주파수특성의 시간변화에 기초하여 기지 음향신호의 진폭 스펙트럼의 주파수특성의 시간변화를 보정한다. 주파수특성의 변화의 추정은, 예를 들면 혼합 음향신호중의 기지의 음향신호만이 포함되어 있는 구간을 특정하고, 이 구간의 주파수특성과 이 구간에 대응하는 기지 음향신호의 주파수특성의 대비로부터, 혼합 음향신호에 포함되는 기지 음향신호의 주파수특성의 변화를 추정한다.

- [0035] 또, 보정스텝의 실행에서는, 예를 들면 혼합 음향신호에 포함되는 기지 음향신호의 음량의 시간변화를 추정하고, 추정한 음량의 시간변화에 기초하여 기지 음향신호의 진폭 스펙트럼의 음량의 시간변화를 보정한다. 음량의 시간변화의 추정은 주파수특성의 보정을 행한 후에, 예를 들면 혼합 음향신호에 포함되는 기지 음향신호에 상당하는 진폭을 갖는 주파수대역을 각 시각에 있어서 특정하고, 그 주파수대역에 있어서의 혼합 음향신호의 진폭과 기지 음향신호의 진폭의 대비로부터 추정한다.
- [0036] 또, 보정스텝의 실행에서는, 예를 들면 혼합 음향신호에 포함되는 기지 음향신호의 시간축방향의 신축을 추정하고, 추정한 시간축방향의 신축에 기초하여 기지 음향신호의 진폭 스펙트럼의 시간축방향의 신축을 보정한다. 시간축방향의 신축의 추정에는, 예를 들면 혼합 음향신호중의 기지의 음향신호만이 포함되어 있는 구간을 특정하고, 이 구간에 대응하는 기지 음향신호의 구간과의 시간축의 대비에 의해 시간축방향의 신축을 추정한다. 또는, 시간축을 짧은 구간으로 분할한 전구간의 대비에 의해 추정한다.
- [0037] 또, 보정스텝의 실행에서는, 예를 들면 혼합 음향신호에 포함되는 기지 음향신호의 주파수축방향의 신축을 추정하고, 추정한 주파수축방향의 신축에 기초하여 기지 음향신호의 진폭 스펙트럼의 주파수축방향의 신축을 보정한다. 주파수축방향의 신축의 추정에는, 예를 들면 혼합 음향신호중의 기지의 음향신호만이 포함되어 있는 구간을 특정하고, 이 구간에 대응하는 기지 음향신호의 구간과의 주파수축의 대비에 의해 주파수축방향의 신축을 추정한다.
- [0038] 또한 본 발명의 기지 음향신호 제거방법에 있어서는, 혼합 음향신호의 진폭 스펙트럼과 기지 음향신호의 진폭 스펙트럼을 시각에 의해 인식할 수 있도록 화상표시하는 화상표시시스템을 더 실행하도록 해도 좋다. 이 경우에는, 화상표시에 기초하여 인간이 혼합 음향신호중에 있어서의 기지의 음향신호가 포함되어 있는 구간을 정하고, 이 구간에 대해서 보정스텝, 제거스텝, 역변환스텝 또는 합성스텝을 실행한다.
- [0039] 또, 본 발명의 기지 음향신호 제거방법에 있어서는, 혼합 음향신호, 기지 음향신호 및 합성스텝의 출력신호를 음향으로서 재생하는 음향재생스텝을 더 실행하도록 해도 좋다. 이 경우에는, 음향재생스텝으로부터의 재생음에 기초하여 인간이 혼합 음향신호중에 있어서의 기지의 음향신호가 포함되어 있는 구간을 정하고, 이 구간에 대해서 보정스텝, 제거스텝, 역변환스텝 및 합성스텝을 실행한다.
- [0040] 또, 본 발명의 기지 음향신호 제거방법에 있어서는, 혼합 음향신호의 진폭 스펙트럼에 기초하여 혼합 음향신호중에 있어서의 기지의 음향신호가 포함되어 있는 구간을 자동추정하고, 이 구간에 대해서 보정스텝, 제거스텝, 역변환스텝 및 합성스텝을 실행하도록 해도 좋다. 혼합 음향신호중에 비교적 확실하게 기지의 음향신호가 포함되어 있는 경우(예를 들면 혼합 음향신호중에서 기지의 음향신호가 단독으로 울리고 있는 구간이 있는 경우)에는, 자동추정에 의해 구간을 특정할 수 있고, 자동추정을 이용함으로써, 기지 음향신호의 제거작업을 빠르게 실시할 수 있다. 또, 혼합 음향신호중에 포함되는 기지 음향신호의 존재가 그다지 확실하지 않은 경우에 있어서는 인간이 구간을 지정한다.
- [0041] 또한, 본 발명의 기지 음향신호 제거방법에 있어서는, 혼합 음향신호중에 포함되어 있는 음향신호에 상당하는 기지 음향신호가 복수종류 존재하는 경우에는, 이들 복수의 기지 음향신호의 전부에 관해서, 기지 음향신호 변환스텝 및 보정스텝을 실행하고, 혼합 음향신호의 진폭 스펙트럼으로부터 복수의 기지 음향신호의 보정 진폭 스펙트럼을 전부 제거하는 제거스텝을 실행해서 얻은 제거후 진폭 스펙트럼을 이용하여 역변환스텝 및 합성스텝을 실행한다. 이것에 의해, 혼합 음향신호중에서 복수종류의 모든 기지 음향신호를 제거할 수 있다.
- [0042] 또, 보정스텝을 실행할 때, 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축 중 적어도 하나의 보정을 인간이 수작업으로 지정할 수 있는 인터페이스 처리를 행하는 그래픽 유저 인터페이스(GUI)를 사용한다.
- [0043] 인터페이스 처리를 행하는 처리모듈은 복수의 음향신호가 혼합된 혼합 음향신호 중에서, 기지 음향신호의 성분을 제거할 때, 혼합 음향신호중에서의 기지 음향신호의 정확한 위치를 인간이 지정할 수 있도록 구성된다.
- [0044] 또, 인터페이스 처리를 행하는 처리모듈은 혼합 음향신호중에서 기지 음향신호의 주파수특성이 시간적으로 변화되고 있을 때, 이들 변화를 인간이 지정할 수 있도록 구성된다. 또, 인터페이스 처리를 행하는 처리모듈은 혼합 음향신호중에서 기지 음향신호의 음량이 시간적으로 변화되고 있을 때, 이들 변화를 인간이 지정할 수 있도록 구성된다.
- [0045] 또한, 인터페이스 처리를 행하는 처리모듈은 혼합 음향신호중에서 기지의 음향신호가 시간축 또는 주파수축방향으로 신축하고 있을 때, 이들 신축을 인간이 지정할 수 있도록 구성된다.

- [0046] 또, 인터페이스 처리를 행하는 처리모듈은 혼합 음향신호와 기지 음향신호가 대응하는 구간을 인간이 지정할 수 있도록 구성된다.
- [0047] 또, 본 발명에 의한 기지 음향신호 제거장치는 혼합 음향신호를 시간주파수 표현으로 변환해서 혼합 음향신호의 진폭 스펙트럼과 혼합 음향신호의 위상을 구하는 혼합 음향신호 변환수단과, 혼합 음향신호중에 포함되어 있는 음향신호에 상당하는 기지 음향신호를 시간주파수 표현으로 변환해서 기지 음향신호의 진폭 스펙트럼을 구하는 기지 음향신호 변환수단과, 혼합 음향신호의 진폭 스펙트럼에 기초하여 혼합 음향신호의 진폭 스펙트럼에 대한 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 적어도 하나를 보정한 기지 음향신호의 보정 진폭 스펙트럼을 구하는 보정수단과, 혼합 음향신호의 진폭 스펙트럼으로부터 기지 음향신호의 보정 진폭 스펙트럼을 제거하는 제거수단과, 제거수단에 의해 얻은 제거후 진폭 스펙트럼과 혼합 음향신호의 위상에 기초하여 시간표현으로 역변환을 행해서 단위파형을 구하는 역변환수단과, 단위파형을 합성해서 기지 음향신호의 성분을 제거한 음향신호를 얻는 합성수단으로 구성된다.
- [0048] 여기에서의 보정수단에는, 혼합 음향신호의 진폭 스펙트럼에 대한 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 적어도 하나의 보정의 지정을 인간이 수작업으로 행할 수 있는 인터페이스 처리를 행하는 처리모듈을 설치한다.
- [0049] 인터페이스 처리를 행하는 처리모듈은 혼합 음향신호의 진폭 스펙트럼과 기지 음향신호의 진폭 스펙트럼을 시각에 의해 대비할 수 있도록 화상표시하는 화상표시부와, 혼합 음향신호, 기지 음향신호 및 합성수단의 출력신호를 음향으로서 재생하는 음향재생부를 구비한다.
- [0050] 인터페이스 처리를 행하는 처리모듈을 사용하면, 화상표시부에 표시된 혼합 음향신호의 진폭 스펙트럼 및 기지 음향신호의 진폭 스펙트럼의 화상표시 및/또는 음향재생부로부터의 재생음에 기초하여 혼합 음향신호중에 포함되어 있는 기지 음향신호의 구간을 인간이 지정할 수 있을 뿐만 아니라, 이 구간에 대해서 인간이 수작업으로 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 적어도 하나의 보정을 지정할 수 있다. 그 결과, 혼합 음향신호중에 포함되어 있는 기지 음향신호의 형태가 다소 복잡해도 높은 제거 정밀도로 기지 음향신호를 제거할 수 있다.
- [0051] 또, 화상표시부는 기지의 음향신호가 포함되어 있는 혼합 음향신호중의 구간의 진폭 스펙트럼과, 기지 음향신호의 대응구간의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 적어도 하나를 보정한 보정 진폭 스펙트럼을 시간축상에서 위치를 맞춰서 표시할 수 있도록 구성되어 있는 것이 바람직하다.
- [0052] 이러한 구성으로 하면, 보정 진폭 스펙트럼의 상태를 시각으로 확인할 수 있으므로, 보정 스펙트럼을 어떻게 하면 제거 정밀도를 높일 수 있는지를 화상을 보면서 추측할 수 있으므로 제거작업이 빨라진다.
- [0053] 또, 화상표시부는 상기 혼합 음향신호의 상기 진폭 스펙트럼으로부터 상기 보정 진폭 스펙트럼을 제거한 음향신호의 진폭 스펙트럼을 화상표시할 수 있도록 구성하는 것이 바람직하다. 이러한 구성으로 하면, 보정의 효과를 화상으로 확인할 수 있으므로, 컷앤티라이 방식으로 보정을 행하면서, 혼합 음향신호중에서 기지 음향신호를 최대한 제거할 수 있다.
- [0054] 또, 본 발명에 의한 기지 음향신호 제거 프로그램이 기록된 기록매체는 혼합 음향신호를 시간주파수 표현으로 변환해서 혼합 음향신호의 진폭 스펙트럼과 혼합 음향신호의 위상을 구하는 혼합 음향신호 변환스텝과, 혼합 음향신호중에 포함되어 있는 음향신호에 상당하는 기지 음향신호를 시간주파수 표현으로 변환해서 기지 음향신호의 진폭 스펙트럼을 구하는 기지 음향신호 변환스텝과, 혼합 음향신호의 진폭 스펙트럼을 입력함으로써 혼합 음향신호의 진폭 스펙트럼에 대한 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 적어도 하나를 보정한 상기 기지 음향신호의 보정 진폭 스펙트럼을 구하는 보정스텝과, 혼합 음향신호의 진폭 스펙트럼으로부터 기지 음향신호의 보정 진폭 스펙트럼을 제거하는 제거스텝과, 제거스텝에 의해 얻은 제거후 진폭 스펙트럼과 혼합 음향신호의 위상에 기초하여 시간표현으로 역변환을 행해서 단위파형을 구하는 역변환스텝과, 단위파형을 합성해서 기지 음향신호의 성분을 제거한 음향신호를 얻는 합성스텝의 처리를 컴퓨터에 의해 실행시키도록 구성되어 있다.
- [0055] 본 발명의 기지 음향신호 제거방법에 의하면, 보정스텝에 의해 혼합 음향신호의 진폭 스펙트럼에 대한 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 적어도 하나를 보정한 기지 음향신호의 보정 진폭 스펙트럼을 구하고, 이 보정 진

폭 스펙트럼을 혼합 음향신호의 진폭 스펙트럼으로부터 제거하므로, 혼합 음향신호중에 비정상적인 잡음으로서 포함되어 있는 기지 음향신호를 높은 정밀도로 제거할 수 있는 이점이 얻어진다.

[0056] 또, 본 발명의 기지 음향신호 제거방법에 의하면, 예를 들면 인간의 음성이나 소리의 배경으로 BGM이 울리고 있는 텔레비전 프로그램이나 영화 등의 음향신호를 입력으로 하면, 별도 준비한 BGM의 음악 음향신호를 이용하여 텔레비전 프로그램중의 BGM을 제거하고, 인간의 음성이나 소리만의 음향신호를 얻는 것이 가능해진다. 또한, BGM 제거후의 음향신호에, 별도의 음악을 BGM으로서 부여함으로써, 텔레비전 프로그램이나 영화 등의 음악을 바꾼 재이용이 가능해진다.

[0057] 여기에서의 기지 음향신호는 임의의 음향신호이면 되므로, 음악의 장르에 상관없이, 보컬의 유무에 상관없이, 반주의 유무에 상관없이 적용할 수 있다. 또, 음악에 한정되지 않고, 정상잡음 및 비정상잡음을 포함한 임의의 기지의 잡음에 적용할 수 있다.

[0058] 또, 기지 음향신호 제거장치에 있어서의 유저 인터페이스를 사용해서 인간이 수작업으로 수정함으로써 실무의 현장에서 보다 고품질의 제거작업을 실현할 수 있다.

실시예

[0071] 이하, 도면을 참조해서 본 발명의 실시형태의 일례를 상세하게 설명한다. 도1은, 본 발명의 기지 음향신호 제거 방법을 실시하는 기지 음향신호 제거장치의 일 실시형태의 구성을 나타내는 블록이다.

[0072] 기지 음향신호 제거장치는 시스템 구성으로서는 혼합 음향신호 변환수단(1)과, 기지 음향신호 변환수단(2)과, 보정수단(3)과, 인터페이스(4)와, 제거수단(5)과, 역변환수단(6)과, 합성수단(7)으로 구성된다.

[0073] 혼합 음향신호 변환수단(1)은 원하는 음성이나 소리 등의 음향신호($s(t)$)(t 는 시간축)에, BGM 등의 음향신호($b(t)$)가 혼합된 혼합 음향신호($m(t)$)를 (이 시점에서는 $s(t)$ 와 $b(t)$ 는 미지이며 $m(t)$ 만이 입력된다) 시간주파수 표현으로 변환해서 혼합 음향신호의 진폭 스펙트럼($M(\omega, t)$)과 혼합 음향신호의 위상($\theta_m(\omega, t)$)을 구한다.

[0074] 또, 기지 음향신호 변환수단(2)은 제거해야 할 음향신호($b(t)$)의 근원이 되는 음원의 기지 음향신호($b'(t)$)를 시간주파수 표현으로 변환해서 기지 음향신호의 진폭 스펙트럼($B'(\omega, t)$)을 구한다.

[0075] 그리고, 보정수단(3)은 혼합 음향신호의 진폭 스펙트럼($M(\omega, t)$)에 기초하여 혼합 음향신호의 진폭 스펙트럼($M(\omega, t)$)에 대한 기지 음향신호의 진폭 스펙트럼($B'(\omega, t)$)의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축을 보정한 기지 음향신호의 보정 진폭 스펙트럼($B(\omega, t)$)을 구한다. 자동화를 위해서는, 자동으로 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 전부를 자동으로 추정해서 보정하도록 보정수단(3)을 구성할 수 있다.

[0076] 이 실시형태에서는 보정수단(3)은 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 모든 보정을, 인터페이스(4)를 이용하여 인간이 수작업으로 지정할 수 있도록 구성되어 있다.

[0077] 이 인터페이스(4)는 나중에 상세하게 설명하는 바와 같이, 혼합 음향신호의 진폭 스펙트럼과 기지 음향신호의 진폭 스펙트럼을 시각에 의해 대비할 수 있도록 화상표시하는 화상표시부를 구비하고 있으며, 그래픽 유저 인터페이스(GUI)에 의해 인터페이스 처리를 행하는 처리모듈이다.

[0078] 인터페이스(4)는 화면표시된 입력부에 의해 혼합 음향신호의 진폭 스펙트럼과 기지 음향신호의 진폭 스펙트럼에 기초하여 혼합 음향신호중에 포함되어 있는 기지 음향신호의 구간을 인간이 지정할 수 있으며 또한 상술한 보정을 지정할 수 있도록 구성되어 있다.

[0079] 제거수단(5)은 혼합 음향신호의 진폭 스펙트럼($M(\omega, t)$)으로부터 기지 음향신호의 보정 진폭 스펙트럼($B(\omega, t)$)을 제거한다. 그리고, 역변환수단(6)은 제거수단(5)에 의해 얻은 제거후 진폭 스펙트럼($S(\omega, t)$)과 혼합 음향신호의 위상($\theta_m(\omega, t)$)에 기초하여 시간표현으로 역변환을 행해서 단위파형($s'(t)$)을 구한다.

[0080] 마지막으로, 합성수단(7)은 역변환수단(6)으로부터 출력되는 단위파형($s'(t)$)을 합성해서 기지 음향신호의 성분을 제거한 음향신호($s(t)$)를 얻는다. 인터페이스(4)는 제거수단(5)으로부터 출력된 제거후 진폭 스펙트럼($S(\omega, t)$)을 화상표시부(도6참조)에 표시한다. 또, 인터페이스(4)는 음향재생부를 내장하고 있으며, 혼합 음향신호, 기지 음향신호 및 합성수단(7)으로부터 출력된 합성된 음향신호를 재생한다.

- [0081] 이 구성에 의하면, 보정의 효과를 화상표시부에서 시각에 의해 확인하고, 또 내장된 음향재생부에서 청각에 의해서도 확인할 수 있으므로, 컷앤티라이 방식으로 보정을 행하면서, 인터페이스(4)의 화상표시부의 화면표시를 보면서, 인간이 필요한 보정을 지정함으로써, 혼합 음향신호중에서 기지 음향신호를 최대한 제거할 수 있다.
- [0082] 다음에, 도2 및 도3을 이용하여, 본 발명의 기지 음향신호 제거장치의 상세한 실시형태의 일례를 설명한다. 도2는, 본 발명의 기지 음향신호 제거방법을 실시하는 경우의 스텝을 나타내는 블록도이며, 도3은 본 발명의 기지 음향신호 제거장치의 주요부를 컴퓨터를 이용하여 실현하는 경우에 사용하는 프로그램의 알고리즘의 일례를 나타내는 플로차트이다.
- [0083] 도4는, 도3의 스텝ST103내의 상세한 처리를 나타내는 플로차트이다. 또, 도5는 인간이 관여하는 추정과 자동추정의 어느 것이라도 추정처리를 실행하는 경우의 스텝의 상세를 나타내는 플로차트이다. 이하, 이들 도1 내지 도5를 참조하면서, 본 발명의 기지 음향신호 제거방법 및 장치에 있어서의 기지 음향신호 제거의 동작을 설명한다.
- [0084] 우선, 이하의 설명에서는, 원하는 음성이나 소리 등의 음향신호(s(t))(t는 시간축)에 제거하는 기지 음향신호인 BGM 등의 음향신호(b(t))가 혼합된 혼합 음향신호(m(t))가 관측되는 것으로 한다.
- [0085] $m(t)=s(t)+b(t)\cdots(1)$
- [0086] 여기에서는, b(t)의 근원이 되는 음원의 음향신호(b'(t))가 기지라는 조건하에서, m(t)이 부여되었을 때, 미지의 s(t)를 구하는 문제를 푼다. 예를 들면 인간의 음성이나 소리와 함께 BGM이 울리고 있는 텔레비전 프로그램 등의 음향신호(m(t))를 입력으로 하고, 그 BGM의 악곡이 기지이며, 그 음향신호(b'(t))를 별도 준비할 수 있을 때, 그 BGM의 음악 음향신호를 이용하여 텔레비전 프로그램중의 BGM을 제거하여, 인간의 음성이나 소리만의 음향신호(s(t))를 얻는 처리를 실현한다. 여기에서, b(t)와 b'(t)는 완전히 일치하지 않으므로,
- [0087] $s(t)=m(t)-b(t)\cdots(2)$
- [0088] 의 감산에 상당하는 처리에서는, b'(t)로부터 b(t)에 상당하는 성분을 추정하여 s(t)를 구할 필요가 있다. 구체적으로는, 기지 음향신호(b'(t))는 혼합음(m(t))중에서는 이하와 같은 변형을 따르는 일이 많으므로, 보정함으로써 b(t)에 상당하는 성분을 추정한다. 보정의 대상은 주로 이하에 설명하는 바와 같이, 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축 또는 주파수축방향의 신축이다.
- [0089] (시간적인 위치의 어긋남)
- [0090] 혼합음(m(t))중에서 기지 음향신호(b'(t))가 울리고 있는 위치는 반드시 선두로부터라고는 할 수 없다. 그래서, 기지 음향신호(b'(t))를 시간축방향으로 어긋나게 해서, 양자의 상대위치를 맞춰서 혼합음으로부터 기지 음향신호를 감산할 필요가 있다.
- [0091] (주파수특성의 시간변화)
- [0092] 혼합음(m(t))중에서 기지 음향신호(b'(t))가 울릴 때는, 그래픽 이퀄라이저 등의 영향으로 주파수특성이 변화되는 일이 많다. 예를 들면, 저역이나 고역이 강조·감쇠되는 일이 있다. 그래서, b'(t)의 주파수특성을 마찬가지로 변화시켜서 보정하여 혼합음으로부터 기지 음향신호를 감산할 필요가 있다.
- [0093] (음량의 시간변화)
- [0094] 혼합음(m(t))중에서 기지 음향신호(b'(t))가 울릴 때는, 혼합음 작성시의 믹서의 페이더 등의 조작으로 혼합비율이 변경되어 음량이 시간변화되는 일이 많다. 그래서, b'(t)의 음량을 마찬가지로 시간변화시켜서 보정하여 혼합음으로부터 기지 음향신호를 감산할 필요가 있다.
- [0095] (시간축 또는 주파수축방향의 신축)
- [0096] 혼합음(m(t))중에서 기지 음향신호(b'(t))가 울릴 때는, 레코드 등의 회전수의 차이에 의해 시간축 또는 주파수축방향으로 신축되는 일이 있다. 그래서, b'(t)를 시간축 또는 주파수축방향으로 신축해서 보정하여 혼합음으로부터 음향신호를 감산할 필요가 있다.
- [0097] 본 발명의 기지 음향신호 제거방법에 있어서는, 기본적인 처리로서, 도2에 나타내듯이, 스텝ST1에 있어서, 우선, 혼합 음향신호를 푸리에 변환하여, 혼합 음향신호의 위상(스텝ST2)과 혼합 음향신호의 진폭 스펙트럼(스텝ST3)을 구함(혼합 음향신호 변환스텝)과 아울러, 스텝ST4에서 혼합 음향신호중에 포함되어 있는 음향신호에 상당하는 기지 음향신호를 푸리에 변환하여 기지 음향신호의 진폭 스펙트럼(스텝ST5)을 구한다(기지 음향신호

변환스텝).

[0098] 그리고, 스텝ST6에 의해, 혼합 음향신호의 진폭 스펙트럼에 기초하여 혼합 음향신호의 진폭 스펙트럼에 대한 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 적어도 하나를 보정한 기지 음향신호의 보정 진폭 스펙트럼(스텝ST7)을 구한다(보정스텝). 다음에, 스텝ST8에서, 혼합 음향신호의 진폭 스펙트럼으로부터 기지 음향신호의 보정 진폭 스펙트럼을 제거해서 제거후 진폭 스펙트럼(스텝ST9)을 구하고(제거스텝), 다음의 스텝ST10에 의해, 제거스텝에 의해 얻은 제거후 진폭 스펙트럼과 혼합 음향신호의 위상에 기초하여 역푸리에 변환을 행해서 단위과형을 구한다(역변환스텝). 마지막으로, 스텝ST11에서, 단위과형을 오버랩 애드법에 의해 합성해서 기지 음향신호의 성분을 제거한 음향신호를 얻는다(합성스텝).

[0099] 이들 처리를 컴퓨터를 이용하여 실현하는 경우에 사용하는 프로그램의 알고리즘에서는, 도3의 플로차트에 나타내듯이, 우선, 스텝ST101에서, 혼합 음향신호를 푸리에 변환해서 혼합 음향신호의 진폭 스펙트럼과 혼합 음향신호의 위상을 구한다. 다음에, 스텝ST102에서, 혼합 음향신호중에 포함되어 있는 음향신호에 상당하는 기지 음향신호를 푸리에 변환해서 기지 음향신호의 진폭 스펙트럼을 구한다.

[0100] 다음의 스텝ST103에서는, 혼합 음향신호의 진폭 스펙트럼에 기초하여 혼합 음향신호의 진폭 스펙트럼에 대한 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 적어도 하나를 보정한 기지 음향신호의 보정 진폭 스펙트럼을 구한다.

[0101] 그 후에, 스텝ST104에서, 혼합 음향신호의 진폭 스펙트럼으로부터 기지 음향신호의 보정 진폭 스펙트럼을 제거해서 제거후 진폭 스펙트럼을 구한다. 다음에 스텝ST105에서, 스텝ST104에서 얻은 제거후 진폭 스펙트럼과 혼합 음향신호의 위상에 기초하여 역푸리에 변환을 행해서 단위과형을 구하고, 스텝ST106에서 단위과형을 오버랩 애드법에 의해 합성해서 기지 음향신호의 성분을 제거한 음향신호를 얻는다.

[0102] 그 후에 스텝ST107에서, 제거후의 음향신호를 유저가 만족했다고 평가한 것인지 아닌지의 판정이 가해져서, 판정결과가 불만족이면 스텝ST103으로 되돌아와서 다시 보정된다. 유저가 만족할 때까지는, 스텝ST103으로부터 스텝ST107이 반복된다.

[0103] 이하, 또한 각 스텝에서 실행되는 내용을 상세하게 설명한다. 본 발명의 실시형태의 방법에서는, 시간영역에서 과형을 감산처리를 하지 않고, 시간주파수영역에서의 진폭 스펙트럼상에서 감산처리를 행한다.

[0104] 예를 들면, 음향신호(m(t), b'(t))에 대한 창함수(h(t))를 사용한 시각(t)에 있어서의 단시간 푸리에 변환(STFT) X_m(ω, t), X_{b'}(ω, t)이

$$X_m(\omega, t) = \int_{-\infty}^{\infty} m(\tau)h(\tau - t)e^{-j\omega\tau} d\tau \quad \dots\dots\dots(3)$$

$$= \alpha_m(\omega, t) + j\beta_m(\omega, t) \quad \dots\dots\dots(4)$$

$$X_{b'}(\omega, t) = \int_{-\infty}^{\infty} b'(\tau)h(\tau - t)e^{-j\omega\tau} d\tau \quad \dots\dots\dots(5)$$

$$= \alpha_{b'}(\omega, t) + j\beta_{b'}(\omega, t) \quad \dots\dots\dots(6)$$

[0105] 로 정의될 때, 이들의 진폭 스펙트럼(M(ω, t), B'(ω, t))은

$$M(\omega, t) = |X_m(\omega, t)| \quad \dots\dots\dots(7)$$

$$= \sqrt{\alpha_m^2(\omega, t) + \beta_m^2(\omega, t)} \quad \dots\dots\dots(8)$$

[0107]

$$B'(\omega, t) = |X_{b'}(\omega, t)| \quad \dots\dots\dots (9)$$

$$= \sqrt{\alpha_{b'}^2(\omega, t) + \beta_{b'}^2(\omega, t)} \quad \dots\dots\dots (10)$$

[0108]

으로 구해진다.

[0109]

현재의 실장에서는, 음향신호를 표본화 주파수 44.1kHz, 양자화 비트수 16bit로 A/D변환하고, 창함수(h(t))로서 창폭 8192점의 해닝(hanning)창을 사용한 단시간 푸리에 변환(STFT)을 고속 푸리에 변환(FFT)에 의해 계산한다. 그 때, 고속 푸리에 변환(FFT)의 프레임을 441점씩 시프트하므로, 프레임 시프트 시간(1프레임 시프트)은 10ms가 된다. 이 프레임 시프트를 처리의 시간단위로 한다.

[0110]

기지 음향신호 제거후의 원하는 음향신호(s(t))의 진폭 스펙트럼(S(ω, t))은 진폭 스펙트럼(M(ω, t), B'(ω, t))으로부터 이하의 식에 의해 구한다. 여기에서, B(ω, t)는 B'(ω, t)를 보정한 후의 진폭 스펙트럼이다.

[0111]

$$S(\omega, t) = \begin{cases} c(\omega, t) (M(\omega, t) - B(\omega, t)) & \text{if } M(\omega, t) > B(\omega, t) \\ 0 & \text{otherwise} \end{cases} \quad \dots\dots (11)$$

$$B(\omega, t) = a(t) g(\omega, t) B'(p(\omega), q(t) + r(t)) \quad \dots\dots\dots (12)$$

[0112]

상기 식에 있어서의 각종 파라미터 함수 a(t), g(ω, t), p(ω), q(t), r(t), c(ω, t)를 순차적으로 설명한다.

[0113]

여기에서의 a(t)는 혼합음의 진폭 스펙트럼으로부터 기지 음향신호의 진폭 스펙트럼에 상당하는 성분을 감산하는 분량을 최종적으로 조정하기 위한 임의의 형상의 함수이며, 통상, a(t)≥1로 한다. 이것이 클 수록, 감산량이 크게 된다.

[0114]

g(ω, t)는 주파수특성의 시간변화와 음량의 시간변화를 보정하기 위한 함수이며,

[0115]

$$g(\omega, t) = g_{\omega}(\omega, t) g_t(t) + g_r(t) \dots\dots (13)$$

[0116]

과 같이 정의한다. 여기에서, g_ω(ω, t)는 주파수특성의 시간변화를 나타내고, 주파수특성의 변화가 없을 때는 g_ω(ω, t)=1이 된다. 한편, g_t(t)는 음량의 시간변화를 나타내고, 음량의 변화가 없을 때는 정수가 된다. M(ω, t)과 B'(ω, t)의 음량차는 기본적으로 g_t(t)에 의해 보정된다. g_r(t)는 주로 g(ω, t)의 값을 전체적으로 올리기 위한 함수이며, 보정시의 미조정에 사용된다. 사용하지 않은 경우에는 g_r(t)=0으로 한다.

[0117]

p(ω)는 주파수축방향의 신축을 보정하기 위한 함수이며, 진폭 스펙트럼(B'(ω, t))의 주파수축(ω)을 변환함으로써 주파수축방향의 선형·비선형의 신축을 가능하게 한다. 또, B'(ω, t)는 본래의 ω의 정의역 밖에서는 0을 취하고, 이산화(離散化)해서 실장할 때는 적절히 보간하는 것으로 한다.

[0118]

q(t)는 시간축방향의 신축을 보정하기 위한 함수이며, 진폭 스펙트럼(B'(ω, t))의 시간축(t)을 변환함으로써 시간축방향의 선형·비선형의 신축을 가능하게 한다. 또, B'(ω, t)는 본래의 t의 정의역 밖에서는 0을 취하고, 이산화해서 설치할 때는 적절히 보간하는 것으로 한다.

[0119]

r(t)는 시간적인 위치의 어긋남을 보정하기 위한 함수이며, 통상은 정수를 설정함으로써 일정한 어긋남폭을 보정한다. 어긋남폭이 시간변화될 때에는, 각 시각에서의 폭을 보정하는 함수를 설정한다. 또, B'(ω, t)는 본래의 t의 정의역 밖에서는 0을 취하고, 이산화해서 실장할 때는 적절히 보간하는 것으로 한다. q(t)와 r(t)를 통합한 하나의 함수로 표현하는 것도 가능하지만, 여기에서는, q(t)는 연속적인 신축을 나타내는 목적으로 설정하고, r(t)은 불연속적인 위치의 어긋남을 나타내는 목적으로 설정하는 것으로 한다.

[0120]

c(ω, t)는 진폭 스펙트럼에 대한 이퀄라이징 처리 및 페이더 조작처리를 위한 임의의 형상의 함수이다. ω방향의 형상에 의해, 그래픽 이퀄라이저와 같이, 기지 음향신호 제거후의 주파수특성을 조정할 수 있다. 또, t방향의 형상에 의해, 믹서의 볼륨 페이더 조작과 같이, 기지 음향신호 제거후의 음량변화를 조정할 수 있다. 사용하지 않은 경우에는 c(ω, t)=1로 한다.

[0121]

[0122] 이렇게 해서 구한 진폭 스펙트럼($S(\omega, t)$)과, 혼합음($m(t)$)의 위상($\theta_m(\omega, t)$)을 이용하여 $X_s(\omega, t)$ 를 구하고, 그것을 역푸리에 변환(IFFT)함으로써, 단위파형($s'(t)$)을 얻는다.

$$\theta_m(\omega, t) = \arctan \left(\frac{\beta_m(\omega, t)}{\alpha_m(\omega, t)} \right) \dots\dots\dots (14)$$

$$X_s(\omega, t) = S(\omega, t)e^{j\theta_m(\omega, t)} \dots\dots\dots (15)$$

$$s'(t) = \frac{h(t)}{2\pi} \int_{-\infty}^{\infty} X_s(\omega, t)e^{j\omega t} d\omega \dots\dots\dots (16)$$

[0123]

[0124] 이 단위파형($s'(t)$)을 오버랩 애드(Overlap Add)법에 의해 배치함으로써, 기지 음향신호 제거후의 원하는 음향신호($s(t)$)를 합성한다.

[0125] 이상에서는, 혼합 음향신호($m(t)$) 중에, 기지 음향신호($b'(t)$)가 1종류 포함되어 있는 경우를 설명했지만, $b'_1(t)$, $b'_2(t)$, ..., $b'_N(t)$ 와 같이 복수 포함되어 있는 경우에는, 이들 진폭 스펙트럼($B'_1(\omega, t), B'_2(\omega, t), \dots, B'_N(\omega, t)$)으로부터 각각에 따른 파라미터 함수의 설정에서 식(12)에 의해 각각 구한 $B_1(\omega, t)$, $B_2(\omega, t), \dots, B_N(\omega, t)$ 를 이용하여,

$$S(\omega, t) = \begin{cases} c(\omega, t) (M(\omega, t) - \sum_{n=1}^N B_n(\omega, t)) & \text{if } M(\omega, t) > \sum_{n=1}^N B_n(\omega, t) \\ 0 & \text{otherwise} \end{cases} \dots\dots\dots (17)$$

[0126]

[0127] 과 같이 $S(\omega, t)$ 를 구하는 처리로 확장할 수 있다. 그 때는, $B_n(\omega, t)$ 의 각종 파라미터 함수를 순서대로 설정하거나, 전체의 밸런스를 취하면서, 복수의 $B_n(\omega, t)$ 의 각종 파라미터 함수를 평행하게 설정한다.

[0128] 또 이상에서는, 모노럴 신호를 대상으로 설명했지만, 스테레오 신호는 좌우를 혼합해서 모노럴 신호로 변환해서 적용해도 좋고, 스테레오 신호의 좌우의 각 신호에 대해서 적용해도 좋다. 또, 스테레오 신호중의 음원방향을 이용해서 적용해도 좋다.

[0129] 상기 각종 파라미터 함수의 설정에 대해서 설명한다. 본 발명의 방법을 적용할 때, 식(11), 식(12), 식(13)의 각종 파라미터 함수 $a(t)$, $g(\omega, t)(g_\omega(\omega, t), g_t(t), g_r(t))$, $p(\omega)$, $q(t)$, $r(t)$, $c(\omega, t)$ 의 형상은 자동추정해도 좋고, 인간이 수작업으로 설정해도 좋다. 또는, 자동추정후에 인간이 수정해도 좋다. 이하에서는, 구체적인 자동 추정방법과, 인간의 수작업에 의한 수정을 가능하게 하는 기지 음향신호 제거장치에 있어서의 인터페이스(4)를 사용하는 경우에 대해서 설명한다.

[0130] 처음에, 식(11), 식(12), 식(13)의 각종 파라미터 함수 $g(\omega, t)(g_\omega(\omega, t), g_t(t)), p(\omega), q(t), r(t)$ 의 형상을 추정하는 방법을 도4를 이용하여 이하에 설명한다. 우선, 스텝ST201에서 BGM 구간(ψ)의 집합(Ψ)의 지정·자동추정을 행하고, 스텝ST202에서 $p(\omega)$, $q(t)$ 의 자동추정을 행하고, 스텝ST203에서 $g_\omega(\omega, t)$, $g_t(t)$, $r(t)$ 의 자동추정을 행한다. 그리고 추정결과와 파라미터 함수가 수속될 때까지 이들 스텝이 계속된다(스텝ST204). 스텝ST205 이후에서는 보정동작이 인터페이스(4)를 이용하여 실행된다.

[0131] $g(\omega, t)$ 의 추정에서는 우선 주파수특성의 시간변화($g_\omega(\omega, t)$)를 추정하고, 다음에 음량의 시간변화($g_t(t)$)를 추정한다. 단, $g(\omega, t)$ 의 추정에 앞서 $p(\omega)$, $q(t)$, $r(t)$ 은 결정되어 있을 필요가 있다. 여기에서는, 편의상, $B'(p(\omega), q(t)+r(t))$ 를 $B'(\omega, t)$ 이라고 기술한다.

[0132] 주파수특성의 시간변화($g_\omega(\omega, t)$)의 추정에서는, 원칙적으로, 인간의 음성이나 소리만의 음향신호($s(t)$)가 거의 포함되어 있지 않은 구간(이하, BGM 구간이라고 함)을 사용한다. BGM 구간은 복수 사용해도 좋다. BGM 구간에서는 혼합음($m(t)$)의 진폭 스펙트럼($M(\omega, t)$)은 기지 음향신호($b'(t)$)에 의한 BGM에 상당하는 진폭 스펙트럼($B'(\omega, t)$)에 유래된 성분이 대부분이 된다. 그래서, 주파수특성이 시간변화되지 않고 정상, 즉,

$g_{\omega}(\omega, t) = g'_{\omega}(\omega)$ 라고 가정할 수 있을 때에는, $g'_{\omega}(\omega)$ 를

$$g'_{\omega}(\omega) = \frac{\int_{\psi \in \Psi} M(\omega, t) dt}{\int_{\psi \in \Psi} B'(\omega, t) dt} \quad \dots\dots\dots (18)$$

[0133]

에 의해 추정한다. 단, ψ 는 하나의 BGM 구간(시간축상의 영역)을 나타내고, Ψ 는 ψ 의 집합으로 한다. 한편, 주파수특성이 시간변화되어 갈 때에는, $g_{\omega}(\omega, t)$ 의 시각(t)에 가까운 BGM 구간(ψ)으로부터

$$\frac{\int_{\psi} M(\omega, t) dt}{\int_{\psi} B'(\omega, t) dt} \quad \dots\dots\dots (19)$$

[0135]

을 구하고, 보간(내삽 또는 외삽)함으로써 $g_{\omega}(\omega, t)$ 를 추정한다(양측에 BGM 구간이 있을 때에는 양측으로부터 내삽한다). 마지막으로, $g_{\omega}(\omega, t)$ 를 주파수축방향으로 평활화한다. 또, 평활화폭은 임의로 설정할 수 있고, 평활화를 하지 않아도 좋다.

[0136]

음량의 시간변화($g_i(t)$)의 추정에서는 진폭 스펙트럼($M(\omega, t)$)과, 주파수특성 보정후의 $g_{\omega}(\omega, t)$, $B'(\omega, t)$ 의 각 시각에 있어서의 진폭을 비교한다. 그러나, 진폭 스펙트럼($M(\omega, t)$)에는, $B'(\omega, t)$ 에 유래된 성분 이외에, 음향신호($s(t)$)에 유래된 성분도 포함된다. 그래서, 주파수축(ω)을 복수의 주파수대역(Φ)으로 분할하고, 각각의 대역(ϕ)($\phi \in \Phi$)마다

[0137]

$$g'_i(\phi, t) = \frac{\int_{\phi} M(\omega, t) d\omega}{\int_{\phi} g_{\omega}(\omega, t) B'(\omega, t) d\omega} \quad (\phi \in \Phi) \quad \dots\dots\dots (20)$$

[0138]

을 구한다(Φ 는 ϕ 의 집합을 나타낸다). Φ 로서 임의의 분할을 적용할 수 있지만, 예를 들면 음악에서 사용하는 평균율의 1옥타브마다 분할(쌍주파수축상에서 등간격으로 분할)하면 좋다. 그리고, $g_i(t)$ 는 $\min(g'_i(\phi, t))$ 또는

[0139]

$$\frac{\sum_{\phi \in \Phi} g'_i(\phi, t)}{\sum_{\phi \in \Phi} 1} \quad \dots\dots\dots (21)$$

[0140]

에 의해 추정한다. $\min(g'_i(\phi, t))$ 의 경우에는, $M(\omega, t)$ 과 $g_{\omega}(\omega, t)B'(\omega, t)$ 가 가장 가까운 주파수대역에 있어서 진폭이 비교되게 된다. 마지막으로, $g_i(t)$ 를 시간축방향으로 평활화한다. 또, 평활화폭은 임의로 설정할 수 있고, 평활화를 하지 않아도 좋다.

[0141]

$p(\omega)$, $q(t)$ 의 추정에서는 $M(\omega, t)$ 과 $B(\omega, t)$ 의 거리(예를 들면 대수 스펙트럼 거리 등)가 최소가 되도록 $p(\omega)$ 와 $q(t)$ 를 변경한다. 그 때, $B(\omega, t) = a(t)g(\omega, t)B'(p(\omega), q(t) + r(t))$ 의 우변 중, $a(t) = 1$ 로 하고,

[0142]

1.(추정 도중의) $p(\omega)$ 와 $q(t)$ 를 임시로 고정한 후에, $g(\omega, t)$ 와 $r(t)$ 을 추정하고,

[0143]

2.(추정 도중의) $g(\omega, t)$ 와 $r(t)$ 을 임시로 고정한 후에, $p(\omega)$ 와 $q(t)$ 를 추정한다라는 2가지의 추정을 반복적으로 되풀이해서 적절한 $p(\omega), q(t)$ 를 추정한다. 이것은, 음향신호의 전구간에 대하여 한번에 실행하지 않고, 시간축을 분할해서 구분적으로 행하면 좋다. 초기값은 전후의 구간의 연속성을 고려해서 정한다. 또, BGM 구간(ψ)의 집합(Ψ)을 이용하여, 이들 복수의 구간에 있어서의 $M(\omega, t)$ 과 $B(\omega, t)$ 의 대응관계의 시간축을 맞추도록 $p(\omega), q(t)$ 를 추정하면 좋다.

[0144]

$r(t)$ 의 추정에서는 원칙적으로, BGM 구간(ψ)의 집합(Ψ)을 이용하여, 이들 구간에 있어서의 $M(\omega, t)$ 와 $B(\omega, t)$ 의 대응관계의 시간축을 맞추도록 $r(t)$ 을 구한다. $r(t)$ 은 정수인 것이 대부분이지만, 가지 음향신호($b'(t)$)의 일부구간이 사용되지 않고, 띄엄띄엄 사용되면서 혼합되어 있었을 때 등에는 그 구간을 건너뛰도록 $r(t)$ 이 불연속 함수로 된다.

[0145]

상기의 $g(\omega, t)$ 나 $r(t)$ 등의 추정에서는 BGM 구간(ψ)의 집합(Ψ)을 사용하고 있었다. 이것은 인간이 수작업으

[0146]

로 지정해도 좋다. 또는, 수작업으로 지정한 BGM 구간의 집합에 자동추정에 의해 추가해도 좋다. 도5는 인간이 수작업으로 지정하는 경우와 자동추정하는 경우의 어느 것이라도 대응하는 프로그램의 소프트웨어의 알고리즘을 나타내는 플로차트이다. 자동추정하는 경우에는 도5의 스텝ST302~ST313을 실행한다. Ψ 의 자동추정에서는 기본적으로 어느 1개소의 BGM 구간(ψ_1)을 단서로 해서 나머지의 BGM 구간의 집합을 구한다. 우선, 최초의 ψ_1 은 인간이 수작업으로 지정하거나, 음향신호의 시간축을 세세하게 분할해서, 이들 짧은 분할구간의 대응관계를 판정해서 구한다. 인간이 수작업으로 지정하지 않은 경우, $B(\omega, t)$ 를 임시로 계산하고(스텝ST302), $M(\omega, t)$ 과 $B(\omega, t)$ 를 세세하게 분할한 시간창의 진폭 스펙트럼간의 거리(유사도에 상당)를 계산한다(스텝ST303).

[0147] 그리고, 그 최소거리의 시간창의 대응관계를 조사해서(스텝ST304), 그 결과를 포함하는 구간을 ψ_1 로 설정해서 초기의 Ψ 로 한다(스텝ST305). 다음에, ψ_1 을 포함하는 Ψ 에 기초하여 $B(\omega, t)$ 의 각종 파라미터 함수를 추정하여(스텝ST306 내지 스텝ST309), $B(\omega, t)$ 를 계산한다(스텝ST310). 각 파라미터의 추정값이 수속되어 있는지를 조사하고, 수속되어 있지 않은 경우에는, Ψ 의 전구간에 대해서 $M(\omega, t)$ 과 $B(\omega, t)$ 의 진폭 스펙트럼간의 거리(유사도에 상당)를 구한다. 여기에서, 그 최대값(또는 평균값)의 정수배를 BGM 구간 판정용 임계값으로 한다(스텝ST312). 그리고, BGM 구간 판정용 임계값이하의 거리를 갖는 구간을 검출하여, 새롭게 Ψ 에 추가한다(스텝ST313). 단, 추가에는 상한을 둘 수도 있다. 이 추정과 추가를 반복함으로써, Ψ 가 갱신되어 각종 파라미터 함수가 적절하게 구해져 간다. 여기에서, $M(\omega, t)$ 과 $B(\omega, t)$ 의 거리로서는, 예를 들면 제곱평균대수 스펙트럼 거리

$$\int (\log M(\omega, t) - \log B(\omega, t))^2 d\omega \quad \dots\dots\dots (22)$$

[0148] 가 유효하다.
[0149]

[0150] 다음에, 기지 음향신호 제거 에디터상의 인터페이스에 의한 각종 파라미터 함수의 조정에 대해서 설명한다.

[0151] 식(11)~식(13)의 모든 파라미터 함수 $a(t)$, $g(\omega, t)$ ($g_{\omega}(\omega, t)$, $g_t(t)$, $g_r(t)$), $p(\omega)$, $q(t)$, $r(t)$, $c(\omega, t)$ 의 형상을 인간이 수작업으로 설정하기 위한 기지 음향신호 제거장치의 유저 인터페이스인 에디터를 이하에 설명한다. 에디터의 유저는 처음부터 임의의 함수형상을 그려서 지정해도 좋고, 처음에는 우선 자동추정을 해서 그 결과를 수정해도 좋다.

[0152] 에디터의 화면구성을 도6에 나타낸다. 이 에디터는 크게 나누어서 혼합 음향신호($m(t)$) 조작용의 서브 윈도우(W1), 기지 음향신호($b'(t)$) 조작용의 서브 윈도우(W2), 기지 음향신호 제거후의 원하는 음향신호($s(t)$) 조작용의 서브 윈도우(W3)의 3개의 서브 윈도우로 구성되어 있다. 기지 음향신호($b'(t)$)가 복수 종류인 경우에는, 전환 스위치(W2S)에 의해, 서브 윈도우(W2)에서 조작하는 기지 음향신호($b'(t)$)를 전환할 수 있다. 이 인터페이스에서는 도4에 나타낸 스텝ST205로부터 스텝ST219가 실행된다.

[0153] 우선, 전체 서브 윈도우에 공통의 기능을 서술한다. 조작범위 슬라이더(P1)는 음향신호중의 어디를 현재 표시하고 있는지를 나타낸다. 커서(P2)는 현재의 조작 대상의 시간축상의 위치를 나타낸다. 아이콘화(절첩) 버튼(P3)은 이것을 누르면 일시적으로 그 버튼이 속하는 서브 윈도우가 절첩되어 작아진다. 현재 조작대상 이외의 미사용의 서브 윈도우를 숨겨서 좁은 화면을 유효하게 활용할 수 있다. 플로트화(확대) 버튼(P4)은 이것을 누르면 일시적으로 그 버튼이 속하는 서브 윈도우가 메인 윈도우로부터 분리되고(플로트화), 더욱 확대되어서 조작·편집이 용이하게 된다. 플로트화(확대) 버튼(P4)밖에 그려져 있지 않은 경우에는, 이 버튼을 누르면 그것에 관련된 서브 윈도우가 플로트화되어서 새롭게 출현한다.

[0154] 서브 윈도우(W1)에는 혼합 음향신호($m(t)$)의 파워의 그래프(E1)와 그 진폭 스펙트럼($M(\omega, t)$)의 그래프(E2)가 표시되어 있다. 서브 윈도우(W2)에는 기지 음향신호($b'(t)$)의 파워의 그래프(E3)와 그 진폭 스펙트럼($B'(\omega, t)$)의 그래프(E4)가 표시되어 있다. 서브 윈도우(W3)에는 기지 음향신호 제거후의 음향신호($s(t)$)의 파워의 그래프(E5)와 그 진폭 스펙트럼($S(\omega, t)$)의 그래프(E6)가 표시되어 있다. 각 진폭 스펙트럼의 그래프(E1, E2, E3)에서는 좌측에 농담으로 진폭이 그려지고(가로축이 시간축, 세로축이 주파수축), 우측에 커서위치에서의 진폭이 그려져 있다(가로축이 파워, 세로축이 주파수축).

[0155] 또, 재생제어 조작패널(P51)에는 인간이 듣고 확인하기 위해서, 혼합 음향신호의 재생, 정지, 패스트 포워드(fast-forward), 패스트 리와인드(fast rewind)가 가능한 버튼군이 배열되어 있다. 재생제어 조작패널(P51)의 조작에 의해, 인터페이스(4)는 내장된 음향재생부에 의해 혼합 음향신호를 재생한다.

- [0156] 기지 음향신호($b'(t)$) 조작용의 서브 윈도우(W2)가 조작의 중심이 되는 윈도우이며, 식(12), 식(13)의 모든 파라미터 함수 $a(t)$, $g(\omega, t)$ ($g_{\omega}(\omega, t)$, $g_t(t)$, $g_r(t)$), $p(\omega)$, $q(t)$, $r(t)$ 의 형상을 자유롭게 설정할 수 있다. 이하, 각 조작패널의 설명을 서술한다.
- [0157] 1. 주파수특성의 시간변화의 보정용 조작패널(C1)(E7의 우측)
- [0158] $g_{\omega}(\omega, t)$ 를 표시·조작하기 위한 패널이며, 커서위치의 시각(t)에서의 $g_{\omega}(\omega, t)$ 가 그려져 있다(가로축이 크기, 세로축이 주파수축). 설정조작결과는 $g(\omega, t)$ 의 표시패널(E7)에 바로 반영된다(스텝ST205, ST206). E7에는, 농담으로 $g(\omega, t)$ 의 값의 크기가 그려져 있다(가로축이 시간축, 세로축이 주파수축).
- [0159] 2. 음량의 시간변화의 보정용 조작패널(C2)(E7의 하측)
- [0160] $g_t(t)$ 를 표시·조작하기 위한 패널이며, 설정조작결과는 $g(\omega, t)$ 의 표시패널(E7)에 바로 반영된다(스텝ST207, ST208).
- [0161] 3. $g(\omega, t)$ 의 값을 전체적으로 올리기를 위한 조작패널(C3)(E7의 하측)
- [0162] $g_r(t)$ 를 표시·조작하기 위한 패널이며, 설정조작결과는 $g(\omega, t)$ 의 표시패널(E7)에 바로 반영된다(스텝ST209, ST210).
- [0163] 4. 혼합음의 진폭 스펙트럼으로부터 기지 음향신호의 진폭 스펙트럼에 해당하는 성분을 감산하는 분량을 최종적으로 조정하기 위한 조작패널(C4)
- [0164] $a(t)$ 를 표시·조작하기 위한 패널이다. 이 패널을 조작하면 $a(t)$ 의 변경이 바로 표시에 반영된다(스텝ST211, ST212).
- [0165] 5. 주파수축방향의 신축을 보정하기 위한 조작패널(C5)
- [0166] $p(\omega)$ 를 표시·조작하기 위한 패널이다. 이 패널을 조작하면 $p(t)$ 의 변경이 바로 표시에 반영된다(스텝ST213, ST214).
- [0167] 6. 시간축방향의 신축을 보정하기 위한 조작패널(C6)
- [0168] $q(t)$ 를 표시·조작하기 위한 패널이다. 이 패널을 조작하면 $q(t)$ 의 변경이 바로 표시에 반영된다(스텝ST215, ST216).
- [0169] 7. 시간적인 위치의 어긋남을 보정하기 위한 조작패널(C7)
- [0170] $r(t)$ 를 표시·조작하기 위한 패널이다. 이 패널을 조작하면 $r(t)$ 의 변경이 바로 표시에 반영된다(스텝ST217, ST218).
- [0171] 또, 재생제어 조작패널(P52)에는 인간이 듣고 확인하기 위해서, 기지 음향신호의 재생, 정지, 패스트 포워드, 패스트 리와인드가 가능한 버튼군이 배열되어 있다. 재생제어 조작패널(P52)의 조작에 의해, 인터페이스(4)는 내장하는 음향 재생부에 의해 기지 음향신호를 재생한다.
- [0172] 다음에, 기지 음향신호 제거후의 음향신호($s(t)$) 조작용의 서브 윈도우(W3)에서는, 식(11)의 파라미터 함수 $c(\omega, t)$ 의 형상을 자유롭게 설정할 수 있다. 이하, 각 조작패널을 설명한다.
- [0173] 1. 그래픽 이퀄라이저(GEQ) 조작패널(C8)(E8의 우측)
- [0174] $c(\omega, t)$ 의 ω 방향의 형상을 표시·조작하기 위한 패널이며, 커서위치의 시각(t)에서의 $c(\omega, t)$ 가 그려져 있다(가로축이 크기, 세로축이 주파수축). 설정조작결과는 $c(\omega, t)$ 의 표시패널(E8)에 바로 반영된다. E8에는, 농담으로 $c(\omega, t)$ 의 값의 크기가 그려져 있다(가로축이 시간축, 세로축이 주파수축).
- [0175] 2. 볼륨 페이더 조작패널(C9)(E8의 하측)
- [0176] $c(\omega, t)$ 의 t 방향의 형상을 표시·조작하기 위한 패널이며, 설정조작결과는 $c(\omega, t)$ 의 표시패널(E8)에 바로 반영된다.
- [0177] 또, 재생제어 조작패널(P53)에는, 인간이 듣고 확인하기 위해서, 합성한 음향신호(합성수단(7)의 출력)의 재생, 정지, 패스트 포워드, 패스트 리와인드가 가능한 버튼군이 배열되어 있다. 재생제어 조작패널(P53)의 조작에 의

해, 인터페이스(4)는 내장된 음향재생부에 의해 합성된 음향신호를 재생한다.

- [0178] 다음에 본 실시형태의 실장에 대해서 설명한다. 우선, 음성이나 소리 등의 음향신호(s(t))에 BGM 등의 음향신호(b(t))가 혼합되어 있는 혼합 음향신호(m(t))가 관측되었을 때, b(t)의 근원이 되는 음원의 음향신호(b'(t))가 기지라는 조건하에서, 미지의 s(t)를 구하는 것이 가능한 프로그램은, 각종 오퍼레이팅 시스템(Linux2.4, SGI IRIX6.5, Microsoft windows XP:등록상표)상에 실장했다. 본 프로그램에 m(t)과 b'(t)가 수록된 오디오 파일을 부여하면, s(t)의 오디오 파일을 얻을 수 있다.
- [0179] 인간의 음성이나 소리에 백그라운드 뮤직(BGM)이 혼합된 여러가지 혼합음에 대해서 실험한 결과, 그 BGM의 원곡의 음향신호를 이용하여, 혼합음중의 BGM을 제거하여 인간의 음성이나 소리가 얻어지는 것을 확인했다. 드럼이 울리고 있는 곡이나 울리고 있지 않은 곡, 포퓰러뮤직이나 클래식음악 등의 여러가지 장르의 곡이 BGM으로서 포함되어 있어도 제거가 가능했다.
- [0180] 실험결과와 예로서, 두사람의 남녀의 대화의 BGM에 클래식 음악이 울리고 있는 혼합음을 실제로 처리한 결과를, 도7~도12에 나타낸다. 도7, 도8에 나타내는 혼합 음향신호(m(t))를 입력으로 하고, 도9, 도10에 나타내는 원음원의 기지 음향신호(b'(t))를 이용하여 BGM성분을 제거한 결과가 도11, 도12에 나타내는 기지 음향신호 제거후의 음향신호(s(t))로 된다. 이 처리결과와 예의 혼합음은 「RWCP 음성대화 데이터베이스」로부터 발췌한 두사람의 남녀의 대화의 음향신호에, 「RWC 연구용 음악 데이터베이스」로부터 발췌한 클래식 음악의 음향신호가 혼합된 것이다.
- [0181] 이상에서 설명한 바와 같이, 본 발명에 의하면, 특히, 보정스텝에 의해 혼합 음향신호의 진폭 스펙트럼에 대한 기지 음향신호의 진폭 스펙트럼의 시간적인 위치의 어긋남, 주파수특성의 시간변화, 음량의 시간변화, 시간축방향의 신축 및 주파수축방향의 신축의 적어도 하나를 보정한 기지 음향신호의 보정 진폭 스펙트럼을 구하고, 이 보정 진폭 스펙트럼을 혼합 음향신호의 진폭 스펙트럼으로부터 제거하므로 혼합 음향신호중에 비정상적인 잡음으로서 포함되어 있는 기지 음향신호를 높은 정밀도로 제거할 수 있는 이점이 얻어진다.
- [0182] 또, 인간의 음성이나 소리의 배경에 BGM이 울리고 있는 텔레비전 프로그램이나 영화 등의 음향신호를 입력으로 하면, 별도 준비한 BGM의 음악 음향신호를 이용하여 텔레비전 프로그램중의 BGM을 제거하여, 인간의 음성이나 소리만의 음향신호를 얻는 것이 가능해진다.
- [0183] 또한, BGM 제거후의 음향신호에, 별도의 음악을 BGM으로서 부여함으로써 텔레비전 프로그램이나 영화 등의 음악을 바꾼 재이용이 가능해진다.
- [0184] 기지 음향신호는 임의의 음향신호면 되므로, 음악의 타입에 상관없이, 보컬의 유무에 상관없이, 반주의 유무에 상관없이 적용할 수 있다. 또, 음악에 한정되지 않고, 정상잡음 및 비정상잡음을 포함한 임의의 기지의 잡음에 적용할 수 있다.

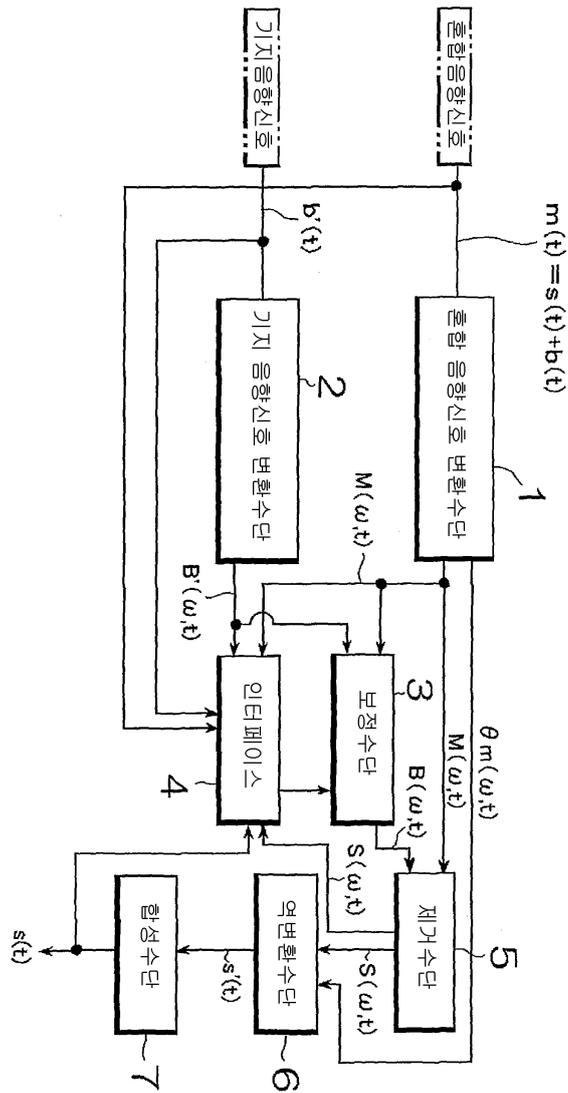
도면의 간단한 설명

- [0059] 도1은, 본 발명의 기지 음향신호 제거장치의 실시형태의 일례의 구성을 나타내는 블럭도이다.
- [0060] 도2는, 본 발명의 기지 음향신호 제거방법을 실시하는 경우의 스텝을 나타내는 블럭도이다.
- [0061] 도3은, 본 발명의 기지 음향신호 제거장치의 주요부를 컴퓨터를 이용하여 실현하는 경우에 사용하는 프로그램의 알고리즘의 일례를 나타내는 플로차트이다.
- [0062] 도4는, 도3의 스텝ST103내의 상세한 처리를 나타내는 플로차트이다.
- [0063] 도5는, 인간이 관여하는 추정과 자동추정의 어느 것이라도 추정동작을 하는 경우의 스텝의 상세를 나타내는 플로차트이다.
- [0064] 도6는, 에디터의 인터페이스의 화면구성을 나타내는 도면이다.
- [0065] 도7은, 혼합 음향신호의 파워의 시간변화를 나타내는 도면이다.
- [0066] 도8은, 혼합 음향신호의 진폭 스펙트럼의 시간변화를 나타내는 도면이다.
- [0067] 도9는, BGM의 근원이 되는 음원의 기지 음향신호의 파워의 시간변화를 나타내는 도면이다.

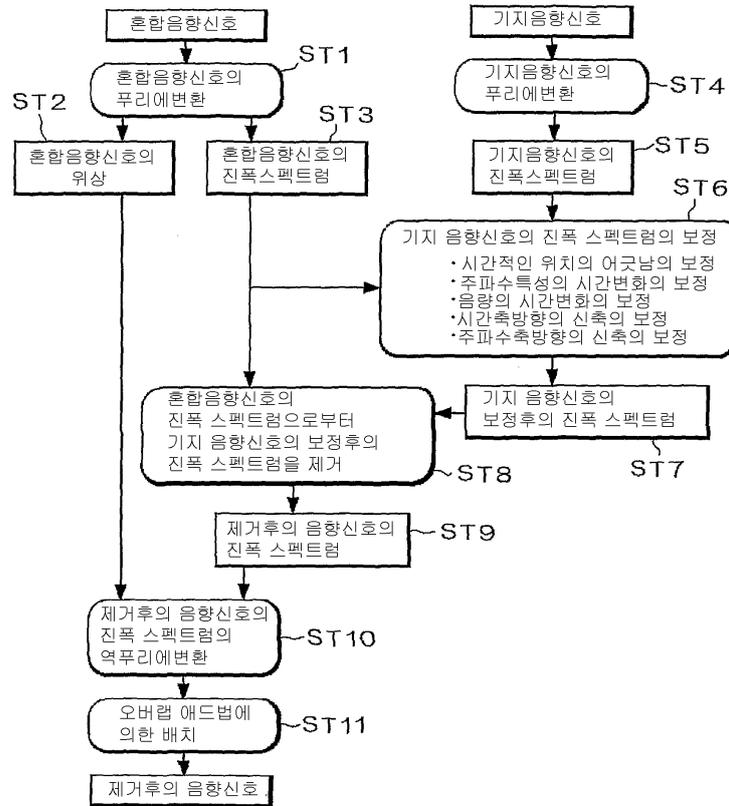
- [0068] 도10은, BGM의 근원이 되는 음원의 기저 음향신호의 진폭 스펙트럼의 시간변화를 나타내는 도면이다.
- [0069] 도11은, 기저 음향신호 제거후의 원하는 음향신호의 파워의 시간변화를 나타내는 도면이다.
- [0070] 도12는, 기저 음향신호 제거후의 원하는 음향신호의 진폭 스펙트럼의 시간변화를 나타내는 도면이다.

도면

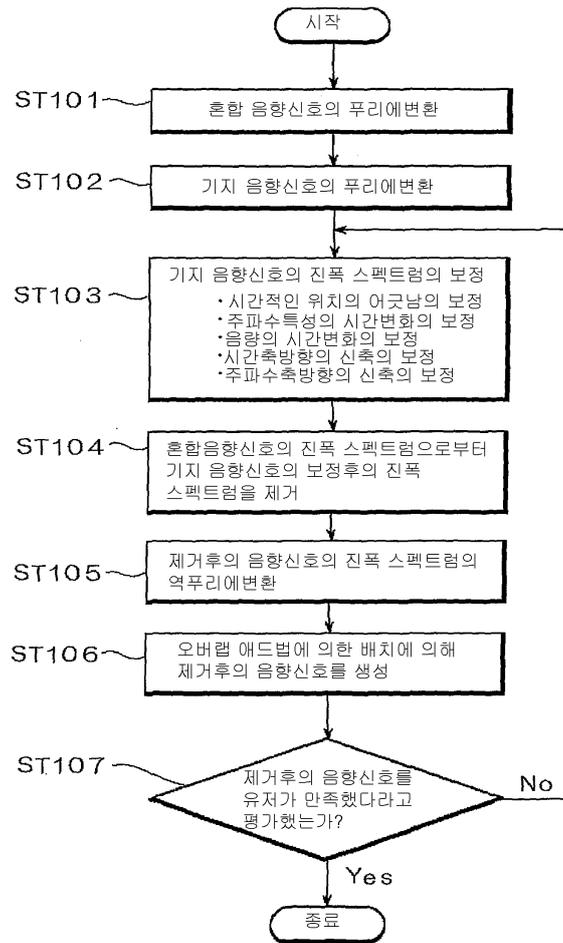
도면1



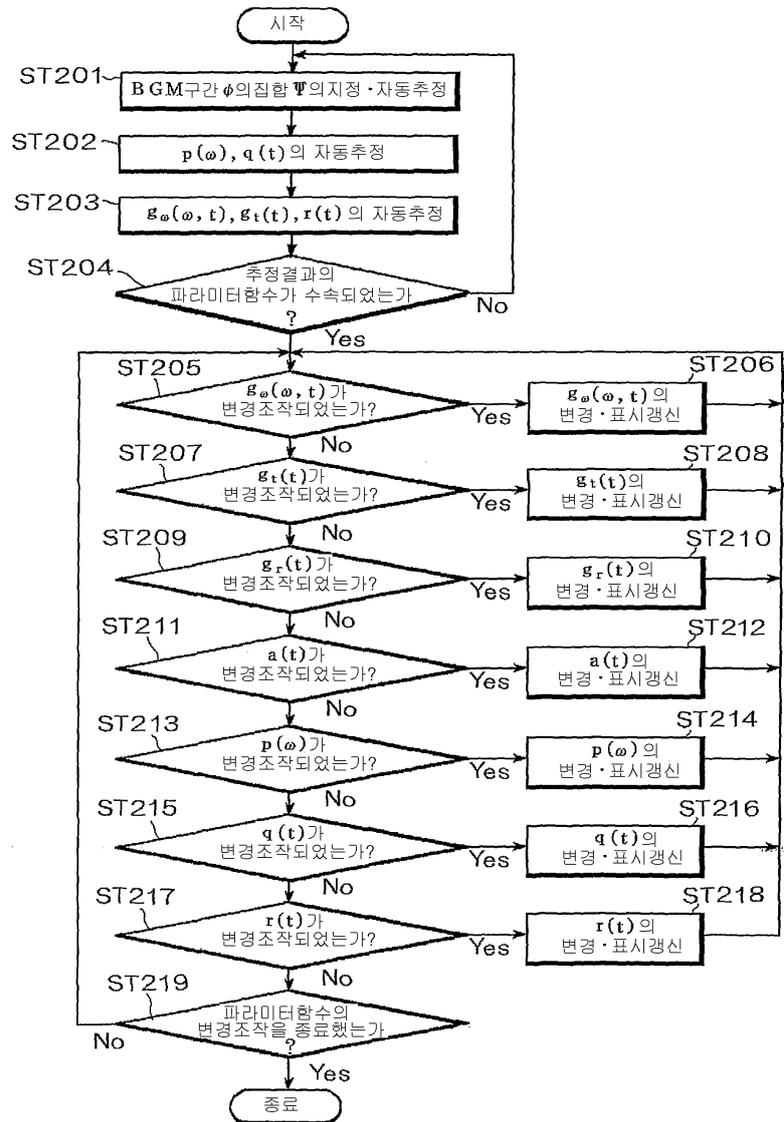
도면2



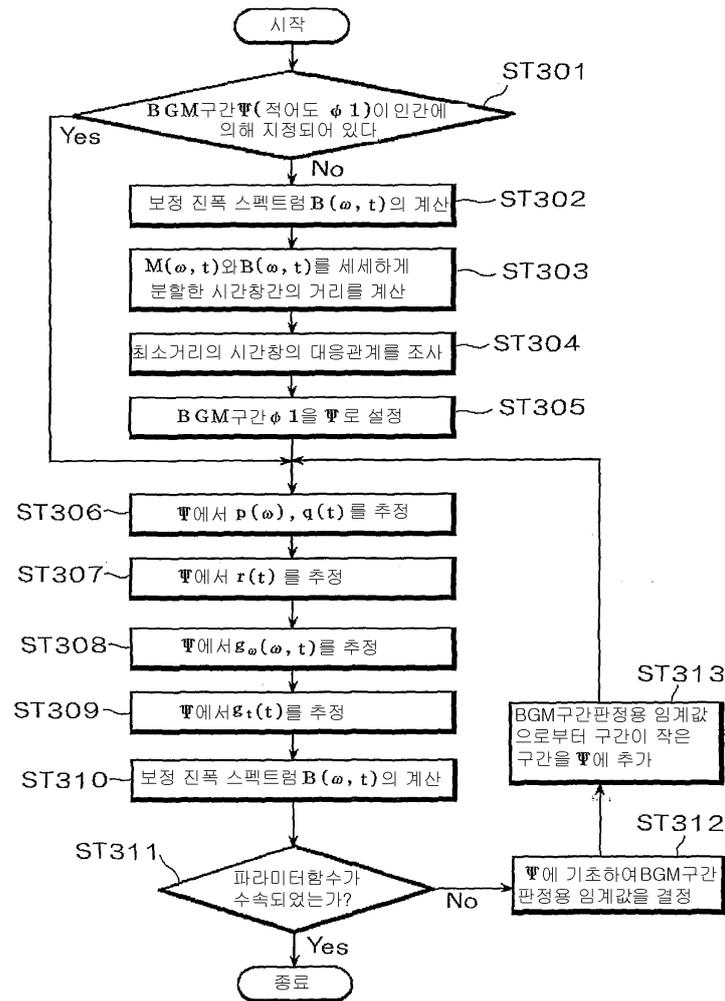
도면3



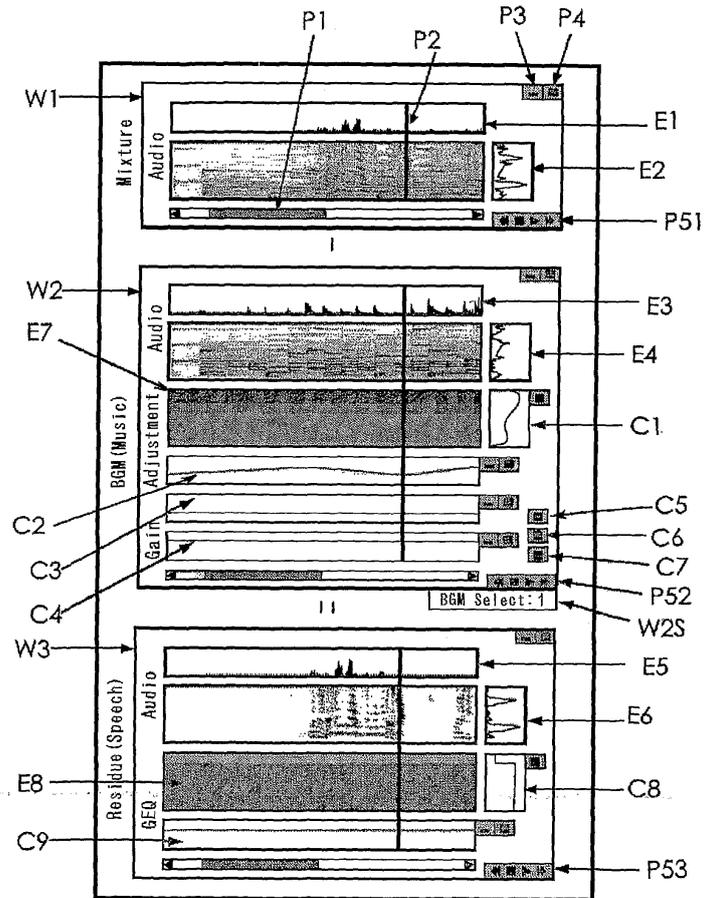
도면4



도면5



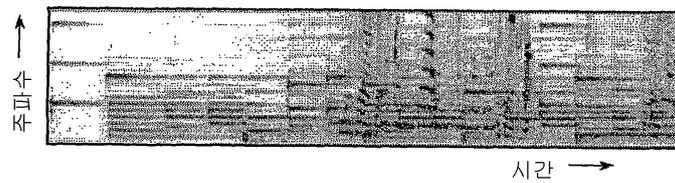
도면6



도면7



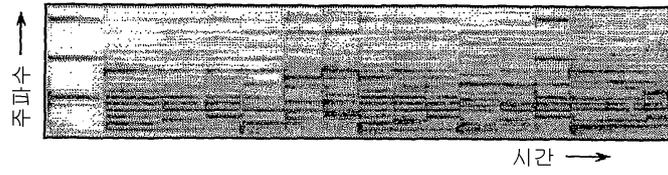
도면8



도면9



도면10



도면11



도면12

