



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0068409
(43) 공개일자 2021년06월09일

- (51) 국제특허분류(Int. Cl.)
H04S 7/00 (2006.01) H04R 5/027 (2006.01)
H04S 1/00 (2006.01)
- (52) CPC특허분류
H04S 7/302 (2013.01)
H04R 5/027 (2013.01)
- (21) 출원번호 10-2021-7008094
- (22) 출원일자(국제) 2019년10월03일
심사청구일자 없음
- (85) 번역문제출일자 2021년03월18일
- (86) 국제출원번호 PCT/JP2019/039103
- (87) 국제공개번호 WO 2020/075622
국제공개일자 2020년04월16일
- (30) 우선권주장
JP-P-2018-191513 2018년10월10일 일본(JP)

- (71) 출원인
소니그룹주식회사
일본국 도쿄도 미나토쿠 코난 1-7-1
- (72) 발명자
후쿠다 가즈미
일본 1080075 도쿄도 미나토쿠 코난 1-7-1 소니
주식회사 내
- 마가리아치 테츠
일본 1080075 도쿄도 미나토쿠 코난 1-7-1 소니
주식회사 내
- (74) 대리인
장수길, 이중희

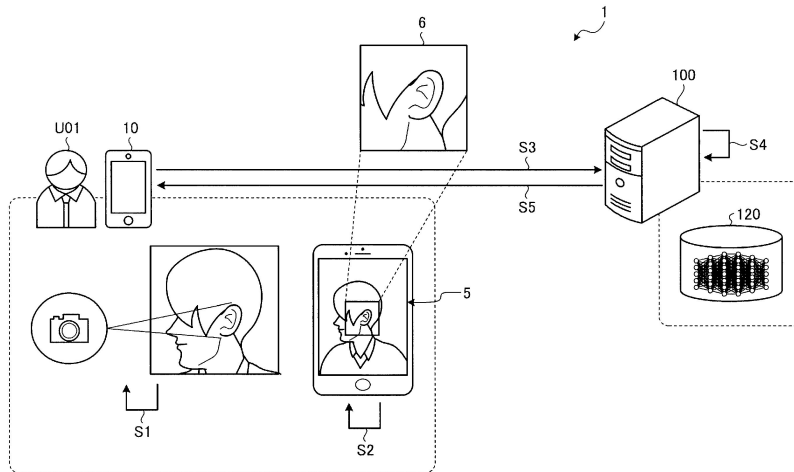
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 정보 처리 장치, 정보 처리 방법 및 정보 처리 프로그램

(57) 요약

본 개시에 따른 정보 처리 장치(100)는, 유저의 귀의 영상을 포함하는 제1 화상을 취득하는 취득부(141)와, 귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 머리 전달 함수를 출력하도록 학습된 학습 완료 모델을 이용하여, 취득부(141)에 의해 취득된 제1 화상에 기초하여, 유저에 대응하는 머리 전달 함수를 산출하는 산출부(142)를 구비한다.

대표도



(52) CPC특허분류

H04S 1/007 (2013.01)

H04S 2420/01 (2013.01)

H04S 2420/13 (2013.01)

명세서

청구범위

청구항 1

유저의 귀의 영상을 포함하는 제1 화상을 취득하는 취득부와,

귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 머리 전달 함수를 출력하도록 학습된 학습 완료 모델을 이용하여, 상기 취득부에 의해 취득된 제1 화상에 기초하여, 상기 유저에 대응하는 머리 전달 함수를 산출하는 산출부

를 구비하는 정보 처리 장치.

청구항 2

제1항에 있어서,

상기 취득부는,

상기 제1 화상에 포함되는 귀의 특징을 나타내는 변수인 귀 파라미터를 취득하고,

상기 산출부는,

상기 귀 파라미터를 상기 학습 완료 모델에 입력함으로써, 상기 유저에 대응하는 머리 전달 함수를 산출하는 정보 처리 장치.

청구항 3

제2항에 있어서,

상기 취득부는,

귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 귀 파라미터를 출력하도록 학습된 귀 파라미터 추정 모델을 이용하여, 상기 제1 화상에 포함되는 귀의 귀 파라미터를 취득하는

정보 처리 장치.

청구항 4

제3항에 있어서,

귀의 영상을 포함하는 화상과 당해 귀의 귀 파라미터의 관계성을 학습함으로써, 상기 귀 파라미터 추정 모델을 생성하는 제1 학습부

를 더 구비하는 정보 처리 장치.

청구항 5

제4항에 있어서,

상기 제1 학습부는,

상기 귀 파라미터에 기초하여 생성된 귀의 3차원 데이터를 렌더링함으로써 얻어지는 귀 화상과 당해 귀 파라미터의 관계성을 학습함으로써, 상기 귀 파라미터 추정 모델을 생성하는

정보 처리 장치.

청구항 6

제5항에 있어서,

상기 제1 학습부는,

상기 귀 혹은 머리부의 3차원 데이터의 텍스처, 렌더링에서의 카메라 각도 또는 렌더링에서의 명도를 변경한 복수의 귀 화상과 당해 복수의 귀 화상에 공통되는 귀 파라미터의 관계성을 학습함으로써, 상기 귀 파라미터 추정 모델을 생성하는

정보 처리 장치.

청구항 7

제4항에 있어서,

귀의 영상을 포함하는 화상과 당해 귀에 대응하는 머리 전달 함수의 관계성을 학습함으로써, 상기 학습 완료 모델을 생성하는 제2 학습부

를 더 구비하는 정보 처리 장치.

청구항 8

제7항에 있어서,

상기 제2 학습부는,

상기 귀 파라미터에 기초하여 생성된 귀의 3차원 데이터와 머리부의 3차원 데이터를 합성한 3차원 데이터에 대하여 음향 시뮬레이션을 행하여, 당해 음향 시뮬레이션에 의해 얻어진 머리 전달 함수와 당해 귀 파라미터의 관계성을 학습함으로써, 상기 학습 완료 모델을 생성하는

정보 처리 장치.

청구항 9

제8항에 있어서,

상기 제2 학습부는,

상기 음향 시뮬레이션에 의해 얻어진 머리 전달 함수의 정보량을 압축하고, 압축한 머리 전달 함수와 상기 귀 파라미터의 관계성을 학습함으로써, 상기 학습 완료 모델을 생성하는

정보 처리 장치.

청구항 10

제8항에 있어서,

상기 제2 학습부는,

상기 귀 파라미터에 기초하여 생성된 귀의 3차원 데이터의 청취점을 설정하고, 설정한 청취점을 사용하여 상기 음향 시뮬레이션을 행하는

정보 처리 장치.

청구항 11

제1항에 있어서,

상기 유저의 머리부 전체의 영상을 포함하는 제2 화상으로부터 당해 유저의 귀의 영상을 특정하고, 특정한 범위를 상기 제1 화상으로서 검출하는 전처리부를 더 구비하고,

상기 취득부는,

상기 전처리부에 의해 검출된 상기 제1 화상을 취득하는

정보 처리 장치.

청구항 12

제11항에 있어서,

상기 전처리부는,

상기 제2 화상에 포함되는 유저의 머리부의 특징점과 당해 유저의 자세의 관계성에 기초하여, 상기 범위를 특정하는

정보 처리 장치.

청구항 13

제12항에 있어서,

상기 전처리부는,

상기 제2 화상에 포함되는 유저의 머리부의 특징점과 당해 유저의 자세의 관계성에 기초하여 상기 범위를 특정할 수 없는 경우, 당해 제2 화상과는 다른 화상이며, 상기 유저의 머리부 전체의 영상을 포함하는 화상의 취득을 새롭게 요구하는

정보 처리 장치.

청구항 14

제11항에 있어서,

상기 전처리부는,

상기 제2 화상에 포함되는 보정 정보에 기초하여 당해 제2 화상의 회전을 보정함으로써 상기 유저의 귀의 영상을 특정하고, 특정한 범위를 상기 제1 화상으로서 검출하는

정보 처리 장치.

청구항 15

컴퓨터가,

유저의 귀의 영상을 포함하는 제1 화상을 취득하고,

귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 머리 전달 함수를 출력하도록 학습된 학습 완료 모델을 이용하여, 취득된 상기 제1 화상에 기초하여, 상기 유저에 대응하는 머리 전달 함수를 산출하는

정보 처리 방법.

청구항 16

컴퓨터를,

유저의 귀의 영상을 포함하는 제1 화상을 취득하는 취득부와,

귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 머리 전달 함수를 출력하도록 학습된 학습 완료 모델을 이용하여, 상기 취득부에 의해 취득된 제1 화상에 기초하여, 상기 유저에 대응하는 머리 전달 함수를 산출하는 산출부

로서 기능시키기 위한 정보 처리 프로그램.

발명의 설명

기술 분야

[0001] 본 개시는, 정보 처리 장치, 정보 처리 방법 및 정보 처리 프로그램에 관한 것이다. 상세하게는, 머리 전달 함수의 산출 처리에 관한 것이다.

배경 기술

[0002] 음원으로부터 귀에 대한 소리의 도달 방법을 수학적으로 나타내는 머리 전달 함수(이하, HRTF(Head-Related Transfer Function)라 칭하는 경우가 있음)를 사용함으로써 헤드폰 등에서의 음상(音像)을 입체적으로 재현하는

기술이 이용되고 있다.

[0003] 머리 전달 함수는 개인차가 크다는 점에서, 그 이용 시에는 개인마다 생성된 머리 전달 함수를 사용하는 것이 바람직하다. 예를 들어, 유저의 이개를 촬영한 화상에 기초하여 머리부의 3차원 디지털 모델(이하, 「3D 모델」이라고 표기함)을 생성하고, 생성된 3D 모델로부터 당해 유저의 머리 전달 함수를 산출하는 기술이 알려져 있다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) 미국 특허 제9544706호 명세서

발명의 내용

해결하려는 과제

[0005] 종래 기술에 의하면, 각각의 유저에 맞추어 개별로 산출된 머리 전달 함수를 정보 처리에 사용할 수 있기 때문에, 음상의 정위감을 높일 수 있다.

[0006] 그러나, 상기한 종래 기술은, 유저가 촬영한 화상에 기초하여 3D 디지털 모델을 생성하고, 생성된 모델로부터 머리 전달 함수를 산출하기 때문에, 계산의 처리 부하가 비교적 커진다. 즉, 상기한 종래 기술에서는, 화상을 송신한 유저에 대하여 머리 전달 함수를 제공하기까지 오랜 시간이 걸릴 것으로 추측되기 때문에, 편리성이 높다고 말하기는 어렵다.

[0007] 그래서, 본 개시에서는, 머리 전달 함수에 관한 처리에 있어서 유저의 편리성을 향상시킬 수 있는 정보 처리 장치, 정보 처리 방법 및 정보 처리 프로그램을 제안한다.

과제의 해결 수단

[0008] 상기 과제를 해결하기 위해서, 본 개시에 따른 일 형태의 정보 처리 장치는, 유저의 귀의 영상을 포함하는 제1 화상을 취득하는 취득부와, 귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 머리 전달 함수를 출력하도록 학습된 학습 완료 모델을 이용하여, 상기 취득부에 의해 취득된 제1 화상에 기초하여, 상기 유저에 대응하는 머리 전달 함수를 산출하는 산출부를 구비한다.

도면의 간단한 설명

[0009] 도 1은 본 개시의 제1 실시 형태에 따른 정보 처리의 개요를 나타내는 도면이다.

도 2는 본 개시에 따른 정보 처리의 전체의 흐름을 나타낸 개념도이다.

도 3은 본 개시의 제1 실시 형태에 따른 정보 처리 장치의 구성예를 나타내는 도면이다.

도 4는 본 개시에 따른 귀 화상 기억부의 일례를 나타내는 도면이다.

도 5는 본 개시에 따른 귀 모델에 관한 학습 처리의 일례를 나타내는 도면이다.

도 6은 본 개시에 따른 귀 파라미터 추정 모델에 관한 학습 처리의 일례를 나타내는 도면이다.

도 7은 본 개시에 따른 귀 화상의 생성 처리의 일례를 나타내는 도면이다.

도 8은 본 개시에 따른 귀 파라미터 추정 모델을 설명하기 위한 도면이다.

도 9는 HRTF에 관한 모델의 생성 처리의 흐름의 개요를 나타내는 도면이다.

도 10은 본 개시에 따른 3D 모델의 재구성을 설명하기 위한 도면이다.

도 11은 HRTF에 관한 모델의 생성 처리의 상세를 설명하는 도면이다.

도 12는 본 개시에 따른 HRTF의 압축 및 복원을 설명하기 위한 도면이다.

도 13은 본 개시에 따른 HRTF의 추정 처리의 흐름을 나타내는 도면이다.

도 14는 본 개시의 제1 실시 형태에 따른 유저 단말기의 구성예를 나타내는 도면이다.

도 15는 본 개시에 따른 검출 처리의 흐름을 나타내는 도면이다.

도 16은 본 개시의 제2 실시 형태에 따른 정보 처리 장치의 구성예를 나타내는 도면이다.

도 17은 정보 처리 장치의 기능을 실현하는 컴퓨터의 일례를 나타내는 하드웨어 구성도이다.

발명을 실시하기 위한 구체적인 내용

[0010] 이하에, 본 개시의 실시 형태에 대하여 도면에 기초하여 상세히 설명한다. 또한, 이하의 각 실시 형태에 있어서, 동일한 부위에는 동일한 번호를 부여함으로써 중복되는 설명을 생략한다.

[0011] (1. 제1 실시 형태)

[0012] [1-1. 제1 실시 형태에 따른 정보 처리의 개요]

[0013] 우선, 도 1을 이용하여, 본 개시에 따른 정보 처리 시스템(1)의 구성, 및 정보 처리 시스템(1)에 의해 실행되는 정보 처리의 개요를 설명한다. 도 1은, 본 개시의 제1 실시 형태에 따른 정보 처리의 개요를 나타내는 도면이다. 본 개시의 제1 실시 형태에 따른 정보 처리는, 도 1에 도시한 정보 처리 시스템(1)에 의해 실현된다. 정보 처리 시스템(1)은, 정보 처리 장치(100)와, 유저 단말기(10)를 포함한다. 정보 처리 장치(100)와 유저 단말기(10)는, 도시하지 않은 유선 또는 무선 네트워크를 사용하여 서로 통신을 행한다. 또한, 정보 처리 시스템(1)을 구성하는 각 장치의 대수는, 도시한 것에 한정되지는 않는다.

[0014] 정보 처리 장치(100)는, 본 개시에 따른 정보 처리 장치의 일례이며, 각 유저에 대응하는 HRTF(머리 전달 함수)를 산출하고, 산출된 HRTF를 제공한다. 정보 처리 장치(100)는, 예를 들어 서버 장치 등에 의해 실현된다.

[0015] 유저 단말기(10)는, HRTF의 제공을 받으려고 하는 유저에 의해 이용되는 정보 처리 단말기이다. 유저 단말기(10)는, 예를 들어 촬영 기능을 가진 스마트폰 등에 의해 실현된다. 도 1의 예에서는, 유저 단말기(10)는, 유저의 일례인 유저 U01에 의해 이용된다.

[0016] HRTF는, 인간의 이개(귓바퀴)나 머리부의 형상 등을 포함하는 주변물에 의해 발생하는 소리의 변화를 전달 함수로서 표현하는 것이다. 일반적으로, HRTF를 구하기 위한 측정 데이터는, 인간이 이개 내에 장착한 마이크로폰이나 더미 헤드 마이크로폰 등을 이용하여 측정용 음향 신호를 측정함으로써 취득된다.

[0017] 예를 들어 3D 음향 등의 기술에서 이용되는 HRTF는, 더미 헤드 마이크로폰 등으로 취득된 측정 데이터나, 다수의 인간에게서 취득된 측정 데이터의 평균값 등을 이용하여 산출되는 경우가 많다. 그러나, HRTF는 개인차가 크다는 점에서, 보다 효과적인 음향 연출 효과를 실현하기 위해서는, 유저 자신의 HRTF를 사용하는 것이 바람직하다. 즉, 일반적인 HRTF를 유저 자신의 HRTF로 치환함으로써, 보다 임장감이 있는 음향 체감을 유저에게 제공할 수 있다.

[0018] 그러나, 유저의 HRTF를 개별로 측정하기 위해서는, 여러가지 문제가 있다. 예를 들어, 우수한 음향 효과를 초래하는 HRTF를 얻기 위해서는, 비교적 고밀도의 측정 데이터가 필요해진다. 고밀도의 측정 데이터를 취득하기 위해서는, 유저를 둘러싼 다양한 각도로부터 유저에 대하여 출력된 음향 신호의 측정 데이터가 필요해진다. 이와 같은 측정은 장시간을 요하기 때문에, 유저의 신체적인 부담이 커지게 된다. 또한, 정확한 측정에는 무향실 등에서의 측정이 필요하기 때문에, 비용적인 부담도 크다. 이 때문에, HRTF의 산출에서는, 유저의 부담을 경감하는 것과, 측정 비용을 경감시키는 것 등이 과제가 된다.

[0019] 상기 과제에 관하여, 유저의 귀나 머리부를 3D 모델로 표현하고, 3D 모델 상에서 음향 시뮬레이션을 행함으로써, 의사적인 측정을 행하는 기술이 존재한다. 이러한 기술에 의하면, 유저는, 머리부의 스캔 데이터나 머리부를 촬영한 화상을 제공함으로써, 측정실에서의 실측을 행하지 않고 HRTF를 산출시킬 수 있다.

[0020] 그러나, 3D 모델의 생성 처리나, 3D 모델에 대한 음향 시뮬레이션은, 계산의 처리 부하가 매우 크다. 이 때문에, 상기 기술을 이용하여, 예를 들어 3D 음향을 이용하는 소프트웨어 등에 유저 고유의 HRTF를 편성하려고 해도, 수 십분이나 수 시간의 타임 래그가 발생할 우려가 있다. 이러한 점은, 유저에게 있어서 편리성이 높다고는 말하기 어렵다. 즉, HRTF를 유저에게 활용시킨다는 관점에서 보면, HRTF의 산출에 있어서는, 처리 속도를 향상시킨다는 과제도 존재한다.

- [0021] 이상과 같이, 유저 개인에 대응한 HRTF를 얻기 위해서는, 여러가지 과제가 존재한다. 본 개시에 따른 정보 처리 장치(100)는, 본 개시의 정보 처리에 의해, 상기 과제를 해결한다.
- [0022] 구체적으로는, 정보 처리 장치(100)는, 귀의 영상을 포함하는 화상이 입력된 경우에, 당해 귀에 대응하는 HRTF를 출력하도록 학습된 학습 완료 모델(이하, 단순히 「모델」이라고 표기함)을 사용하여, 유저에 대응하는 HRTF를 산출한다. 예를 들어, 정보 처리 장치(100)는, 유저 U01의 귀의 영상을 포함하는 화상을 유저 단말기(10)로부터 취득한 경우, 당해 화상을 모델에 입력하여, 유저 U01 고유의 HRTF를 산출한다. 즉, 정보 처리 장치(100)는, 유저 U01의 화상에 기초하여 3D 모델을 생성하여 음향 시뮬레이션을 행한다고 하는 과정을 거치지 않고, HRTF를 산출한다.
- [0023] 이에 의해, 정보 처리 장치(100)는, 음향 시뮬레이션을 실행한 경우와 비교하여, 극히 단시간에 HRTF의 산출을 행할 수 있다. 이하, 본 개시에 따른 정보 처리 장치(100)가 실행하는 정보 처리의 개요에 대하여, 도 1을 이용하여 흐름을 따라 설명한다.
- [0024] 도 1에 도시한 바와 같이, 유저 U01은, 자신의 귀의 영상을 포함하는 화상을 취득하기 위해서, 머리부 횡측에서 자신을 촬영한다(스텝 S1). 예를 들어, 유저 U01은, 유저 단말기(10)가 구비하는 카메라를 이용하여, 자신의 머리부를 촬영한다. 또한, 본 개시에 있어서, 귀의 화상이란, 일반적으로 유저 단말기(10) 등으로 촬영 가능한 2차원 컬러 화상에 한하지 않고, 모노크롬 화상이나, 심도 정보를 포함하는 맵스 화상이나, 그것들을 임의로 조합한 것이어도 된다. 또한, 본 개시의 처리에 사용되는 화상은, 1장이 아니라, 복수의 화상이어도 된다.
- [0025] 유저 단말기(10)는, 스텝 S1에 의해 얻어진 화상(5)에 대하여, 정보 처리 장치(100)에 송신하기 위한 전처리를 실행한다(스텝 S2). 구체적으로는, 유저 단말기(10)는, 전처리로서, 화상(5)에 포함되는 유저 U01의 귀의 영상을 검출하고, 검출된 귀의 영상의 범위를 화상(5)으로부터 잘라내는 처리를 행한다. 또한, 귀의 검출 등의 전처리에 대하여, 상세는 후술한다.
- [0026] 유저 단말기(10)는, 전처리에 의해, 유저 U01의 귀의 영상을 포함하는 화상(6)을 작성한다. 그리고, 유저 단말기(10)는, 화상(6)을 정보 처리 장치(100)에 송신한다(스텝 S3). 또한, 촬영에 의해 얻어진 화상(5)으로부터 화상(6)을 작성하는 처리나, 화상(6)을 송신하는 처리 등의 일련의 처리는, 예를 들어 정보 처리 장치(100)로부터 제공되는 프로그램(예를 들어 스마트폰용 애플리케이션)에 의해 실행된다. 이와 같이, 본 개시에 따른 정보 처리에서는, 화상(5)으로부터 귀의 영상만을 잘라낸 화상(6)만을 네트워크 상에 송신하고, 개인이 특정될 우려가 있는 화상(5)을 송신하지 않기 때문에, 안전성이 우수한 처리를 실현할 수 있다. 또한, 정보 처리 장치(100)는, 화상(5)을 취득하지 않고 화상(6)만을 취득함으로써, 개인 정보를 취급하는 리스크를 피할 수 있다. 또한, 정보 처리 장치(100)는, 유저 단말기(10)로부터 화상(5)을 취득하고, 정보 처리 장치(100)측에서 화상(6)을 작성하는 처리를 행해도 된다. 이 구성에 대해서는, 제2 실시 형태로서 후술한다.
- [0027] 정보 처리 장치(100)는, 유저 단말기(10)로부터 송신된 화상(6)을 취득한다. 그리고, 정보 처리 장치(100)는, 기억부(120) 내에 저장된 모델에 화상(6)을 입력한다(스텝 S4). 이 모델은, 귀의 영상을 포함하는 2차원의 화상이 입력된 경우에, 당해 귀에 대응하는 HRTF를 출력하도록 학습된 모델이다. 즉, 정보 처리 장치(100)는, 화상(6)을 모델에 입력함으로써, 화상(6)에 포함되는 귀(바뀌 말하면 유저 U01)에 대응한 HRTF를 산출한다.
- [0028] 그리고, 정보 처리 장치(100)는, 산출된 HRTF를 유저 단말기(10)에 제공한다(스텝 S5). 이와 같이, 유저 U01은, 자신의 옆 얼굴을 촬영하여 화상(5)만 준비한다면, 실제의 측정 처리나 3D 모델의 음향 시뮬레이션 등을 거치지 않고, 자신에게 고유한 HRTF를 얻을 수 있다. 즉, 정보 처리 장치(100)는, 유저 U01에게 측정의 부담을 주지 않고, 극히 단시간에 HRTF를 유저 U01에게 제공할 수 있다. 결과적으로, 정보 처리 장치(100)는, HRTF에 관한 처리에 있어서 유저의 편리성을 향상시킬 수 있다.
- [0029] 상술한 바와 같이, 본 개시에 따른 정보 처리에서는, 학습 처리를 거쳐 작성한 모델을 이용하여 HRTF를 산출함으로써, 처리의 고속화를 도모한다. 도 1에서는 본 개시에 따른 정보 처리 중, 유저 U01에게 HRTF를 제공하는 처리의 개요를 나타내었지만, 도 2 이하에서는, 모델의 학습 처리를 포함한, 정보 처리 장치(100)에 의한 일련의 정보 처리를 상세히 설명한다. 또한, 상세는 도 2 이하에서 설명하지만, 도 1에서 나타난 「모델」은, 반드시 하나의 모델을 나타내는 것은 아니며, 다양한 수치를 출력하는 복수의 모델을 조합한 것이어도 된다.
- [0030] [1-2. 본 개시에 관한 정보 처리의 전체의 흐름]
- [0031] 정보 처리 장치(100)의 구성 등의 상세한 설명에 앞서, 본 개시에 따른 정보 처리 장치(100)가 실행하는 정보 처리의 전체의 흐름을 도 2에 나타낸다. 도 2는, 본 개시에 따른 정보 처리의 전체의 흐름을 나타낸 개념도이

다.

- [0032] 우선, 정보 처리 장치(100)는, 복수의 인물의 귀 모양에 관한 데이터를 채취하고, 채취한 귀 모양에 기초하여, 귀 모델을 작성한다(스텝 S11). 또한, 귀 모양은, 반드시 석고 등으로 인물의 귀를 본뜬 것에 한하지 않고, 인물의 귀의 형상을 나타내는 정보라면, 어떠한 정보여도 된다. 또한, 본 개시에 있어서, 귀 모델은, 귀의 특징을 나타낸 파라미터(이하, 「귀 파라미터」라고 칭함)가 입력된 경우에, 대응하는 귀의 형상을 출력하는 모델이다. 귀 파라미터는, 예를 들어 귀의 형상에 관한 데이터(예를 들어, 채취한 귀 모양을 CT(Computed Tomography) 스캔에 의해 데이터화한 것)에 기초하여, 귀의 형상을 주성분 분석함으로써 얻어진다. 이에 의해, 정보 처리 장치(100)는, 귀 파라미터가 얻어지면, 당해 귀 파라미터에 대응하는 귀의 형상의 데이터(바꿔 말하면, 귀를 본뜬 3D 모델)가 얻어지게 된다.
- [0033] 그 후, 정보 처리 장치(100)는, 귀 모델에 기초하여, 귀 파라미터 추정 모델을 생성한다(스텝 S12). 정보 처리 장치(100)는, 스텝 S11에서 생성된 귀 모델에 귀 파라미터를 입력함으로써, 다수의 귀 화상을 생성 가능하다. 귀 파라미터는, 랜덤으로 입력되어도 되고, 임의의 규칙(예를 들어, 특정한 인종마다 귀의 형상에 특정한 경향이 있다는 사실이 발견되면, 당해 사실에 기초하여 규칙을 도출해도 됨)을 따라 귀 파라미터를 자동적으로 생성하고, 생성된 값이 입력되어도 된다. 이 때문에, 정보 처리 장치(100)는, 생성된 귀 화상과 생성원이 된 귀 파라미터의 관계성을 학습함으로써, 귀를 포함하는 화상이 입력된 경우에, 당해 귀에 대응하는 귀 파라미터를 출력하는 모델을 생성할 수 있다. 이러한 모델이, 귀 파라미터 추정 모델이다. 이에 의해, 정보 처리 장치(100)는, 귀의 영상을 포함하는 2차원의 화상이 얻어지면, 당해 귀에 대응하는 귀 파라미터가 얻어지게 된다. 그리고, 정보 처리 장치(100)는, 귀 파라미터가 얻어지면, 스텝 S11에서 생성된 귀 모델을 이용하여, 화상에 포함되는 귀의 3D 모델이 얻어지게 된다. 또한, 상기 학습에 대해서는, 귀의 형상이 데이터화된 본인의 귀를 촬영한 화상과 데이터화된 귀를 귀 파라미터로 변환한 것의 관계성을 학습해도 된다. 이 경우, CG(Computer Graphics) 화상이 아니라, 실제의 촬영 화상을 사용한 학습이 행해지기 때문에, 생성되는 귀 파라미터 추정 모델의 정밀도를 보다 높일 수 있다고 상정된다.
- [0034] 정보 처리 장치(100)는, 귀 파라미터 추정 모델을 사용하여 생성된 3D 모델에 대하여 음향 시뮬레이션을 행하여, 당해 3D 모델에 대응하는 고유의 HRTF(이하, 이와 같은 개개의 귀 형상에 대응하여 생성된 HRTF를 「개인화 HRTF」라고 칭함)를 산출한다(스텝 S13). 즉, 스텝 S11 내지 스텝 S13의 과정을 거침으로써, 정보 처리 장치(100)는, 귀를 포함하는 화상으로부터, 음향 시뮬레이션에 의해 개인화 HRTF를 산출하기 위한 일련의 처리가 실현 가능하게 된다.
- [0035] 또한, 정보 처리 장치(100)는, 랜덤 혹은 규칙적으로 생성된 귀 파라미터로부터 다수의 3D 모델을 생성하고, 생성된 3D 모델에 대하여 음향 시뮬레이션을 행하는 처리를 반복함으로써, 귀 파라미터와 개인화 HRTF의 관계성을 학습한다. 즉, 정보 처리 장치(100)는, 산출한 개인화 HRTF에 기초하여, HRTF 학습 모델을 생성한다(스텝 S14).
- [0036] 본 개시에 있어서, HRTF 학습 모델이란, 귀 파라미터가 입력된 경우에, 당해 귀 파라미터에 대응하는 개인화 HRTF를 출력하는 모델이다. 이에 의해, 정보 처리 장치(100)는, 귀 파라미터가 얻어지면, 당해 귀 파라미터에 대응하는 개인화 HRTF가 얻어지게 된다.
- [0037] 이 후, 정보 처리 장치(100)는, 유저로부터 화상을 취득한 경우에는, 당해 화상(보다 정확하게는, 화상에 포함되는 귀의 귀 파라미터)을 HRTF 학습 모델에 입력함으로써, 유저의 개인화 HRTF를 산출한다(스텝 S15). 스텝 S15에 나타낸 처리는, 도 1에서 나타낸 일련의 처리에 대응한다.
- [0038] 상기한 바와 같이 정보 처리 장치(100)는, 복수의 모델을 생성하고, 생성된 모델을 사용한 정보 처리를 행함으로써, 유저로부터 취득한 화상으로부터 개인화 HRTF를 산출할 수 있다. 또한, 도 2에서 나타낸 처리는, 반드시 스텝 S11 내지 S15의 순으로 실행될 것을 요하지는 않는다. 예를 들어, 정보 처리 장치(100)는, 처리에 요하는 각 모델을 생성한 후에는, 스텝 S15의 처리를 행하기만 하면 된다.
- [0039] 이상, 본 개시에 따른 정보 처리의 전체의 흐름의 개요를 설명하였다. 도 3이하에서는, 정보 처리 장치(100) 및 유저 단말기(10)의 구성을 설명함과 함께, 다양한 학습 처리의 상세를 순서대로 설명한다.
- [0040] [1-3. 제1 실시 형태에 따른 정보 처리 장치의 구성]
- [0041] 도 3을 이용하여, 제1 실시 형태에 따른 정보 처리 장치(100)의 구성에 대하여 설명한다. 도 3은, 본 개시의 제1 실시 형태에 따른 정보 처리 장치(100)의 구성예를 나타내는 도면이다.

- [0042] 도 3에 도시한 바와 같이, 정보 처리 장치(100)는, 통신부(110)와, 기억부(120)와, 제어부(130)를 갖는다. 또한, 정보 처리 장치(100)는, 정보 처리 장치(100)를 관리하는 관리자 등으로부터 각종 조작을 접수하는 입력부(예를 들어, 키보드나 마우스 등)나, 각종 정보를 표시하기 위한 표시부(예를 들어, 액정 디스플레이 등)를 가져도 된다.
- [0043] 통신부(110)는, 예를 들어 NIC(Network Interface Card) 등에 의해 실현된다. 통신부(110)는, 네트워크 N(인터넷 등)과 유선 또는 무선으로 접속되고, 네트워크 N을 통해 유저 단말기(10) 등과의 사이에서 정보의 송수신을 행한다.
- [0044] 기억부(120)는, 예를 들어 RAM(Random Access Memory), 플래시 메모리(Flash Memory) 등의 반도체 메모리 소자, 또는 하드 디스크, 광 디스크 등의 기억 장치에 의해 실현된다. 기억부(120)는, 학습 처리에 사용하는 각종 데이터나, 학습 처리에 의해 생성된 모델 등을 기억한다.
- [0045] 도 3에 도시한 바와 같이, 기억부(120)는, 귀 모양 정보 기억부(121)와, 귀 모델 기억부(122)와, 귀 화상 기억부(123)와, 귀 파라미터 추정 모델 기억부(124)와, HRTF 처리 모델 기억부(125)와, HRTF 학습 데이터 기억부(126)와, HRTF 학습 모델 기억부(127)를 갖는다.
- [0046] 귀 모양 정보 기억부(121)는, 실제로 인체로부터 채취된 귀 모양을 3D 모델 데이터화한 정보(즉, 귀의 형상에 관한 정보)를 기억한다. 구체적으로는, 귀 모양 정보 기억부(121)는, 채취된 귀 모양을 CT 스캔함으로써 얻어진, 3차원 형상을 나타내는 데이터(3D 폴리곤 등)를 기억한다.
- [0047] 귀 모델 기억부(122)는, 본 개시에 따른 귀 모델을 기억한다. 귀 모델은, 귀의 특징을 나타낸 귀 파라미터가 입력된 경우에, 대응하는 귀의 형상을 출력하는 모델이다.
- [0048] 귀 파라미터는, 귀 모양 정보 기억부(121)에 기억된 귀 모양의 형상을 나타내는 데이터를 주성분 분석함으로써 얻어진다. 즉, 귀 파라미터는, 귀의 3D 폴리곤에 대하여 통계적으로 분석(주성분 분석)을 행함으로써, 귀 중에서 변화가 큰(귀의 형상을 특징짓는) 부위를 수치화한 것이다. 본 개시에 따른 귀 파라미터는, 예를 들어 10의 수치의 조합으로 나타내어지고, 또한 각각의 수치는, 예를 들어 -10에서 +10까지의 수치로 나타내어진다. 예를 들어, 모든 수치가 「0」인 귀 파라미터는, 학습 데이터(채취된 귀 모양)의 평균적인 형상을 갖는 귀에 대응한다. 또한, 정보 처리 장치(100)는, 주성분 분석에 의해 귀의 형상을 나타내는 모델의 생성 처리에 대하여, 예를 들어 인물의 얼굴의 생성 처리 등에서 사용되는 기지의 기술을 적절히 적용해도 된다. 또한, 정보 처리 장치(100)는, 주성분 분석에 한하지 않고, 독립 성분 분석이나 그 밖의 비선형 모델 등, 기지의 분석 방법을 적절히 이용하여, 귀 파라미터를 생성해도 된다. 여기서, 귀 파라미터는, 귀 중에서 변화가 큰 부위를 수치화한 것에 한하지 않고, 예를 들어 HRTF에 대한 영향이 커지는, 귀의 형상에 관한 특징을 파라미터화한 것이어도 된다.
- [0049] 귀 화상 기억부(123)는, 귀의 영상을 포함하는 화상을 기억한다. 예를 들어, 귀 화상 기억부(123)는, 귀 화상으로서, 귀 모델에 의해 생성된 귀의 형상(귀의 3D 모델)을 렌더링하여 얻어진 CG 화상을 기억한다. 또한, 귀 화상 기억부(123)는, 귀 화상으로서, 유저로부터 송신된 귀의 영상을 포함하는 화상을 기억해도 된다.
- [0050] 여기서, 도 4에, 본 개시에 따른 귀 화상 기억부(123)의 일례를 나타낸다. 도 4는, 본 개시의 귀 화상 기억부(123)의 일례를 나타내는 도면이다. 도 4에 도시한 예에서는, 귀 화상 기억부(123)는, 「귀 파라미터」, 「귀 3D 모델 데이터」, 「머리부 3D 모델 데이터」, 「귀 화상 ID」, 「화상 생성 파라미터」와 같은 항목을 갖는다. 또한, 「화상 생성 파라미터」는, 「텍스처」, 「카메라 각도」, 「해상도」, 「명도」와 같은 소항목을 갖는다.
- [0051] 「귀 파라미터」는, 귀의 형상의 특징을 나타내는 파라미터이다. 예를 들어, 귀 파라미터는, 10차원의 수치 등에 의해 표기된다. 「귀 3D 모델 데이터」는, 귀 파라미터에 기초하여 재구성된 귀의 3차원 형상을 나타내는 데이터이다. 「머리부 3D 모델 데이터」는, 인물의 3D 모델의 재구성 시에, 귀 3D 모델 데이터가 합성되는 머리부의 3차원 형상을 나타내는 데이터이다.
- [0052] 「귀 화상 ID」는, 3D 모델을 렌더링하여 얻어지는 귀 화상을 식별하는 식별 정보를 나타낸다. 도 4에 도시한 바와 같이, 귀 화상은, 렌더링 시에 설정되는 파라미터(화상 생성 파라미터)를 다양하게 변경함으로써, 하나의 3D 모델로부터 복수가 생성된다.
- [0053] 「화상 생성 파라미터」는, 화상을 생성하기 위한 렌더링에서의 설정 파라미터를 나타낸다. 「텍스처」는, CG의 텍스처의 설정을 나타낸다. 「카메라 각도」는, 3D 모델을 렌더링하여 2차원의 화상을 얻을 때의, 의사적인 카메라의 촬영 각도를 나타낸다. 「해상도」는, 렌더링 시의 해상도를 나타낸다. 「명도」는, 렌더링 시의 명도를

나타낸다. 명도의 항목에는, 렌더링에서의 광(입사광)의 각도 등의 설정 데이터 등이 포함되어도 된다.

- [0054] 또한, 도 4에서는, 각 항목의 데이터를 「A01」이나 「B01」과 같이 개념적으로 기재하고 있지만, 실제로는, 각 항목의 데이터에는, 각 항목에 대응한 구체적인 데이터가 기억된다. 예를 들어, 「귀 파라미터」의 항목에는, 구체적인 10종류의 수치의 나열이 기억된다. 다른 항목도 마찬가지로, 각 항목에 대응한 각종 수치나 정보가 기억된다.
- [0055] 즉, 도 4에 도시한 예에서는, 귀 파라미터 「A01」에 의해 생성된 귀 3D 모델 데이터는 「B01」이며, 또한, 당해 귀 3D 모델 데이터와 조합하여, 인물의 3D 모델을 구성하는 머리부 3D 모델 데이터는 「C01」임을 나타내고 있다. 또한, 생성된 인물의 3D 모델로부터 얻어진 귀 화상은, 귀 화상 ID 「D01」, 「D02」, 「D03」 등으로 식별되는 복수의 귀 화상임을 나타내고 있다. 또한, 귀 화상 ID 「D01」로 식별되는 귀 화상은, 렌더링 시의 화상 생성 파라미터로서, 텍스처가 「E01」이고, 카메라 각도가 「F01」이고, 해상도가 「G01」이며, 명도가 「H01」임을 나타내고 있다.
- [0056] 도 3으로 되돌아가서, 설명을 계속한다. 귀 파라미터 추정 모델 기억부(124)는, 귀 파라미터 추정 모델을 기억한다. 귀 파라미터 추정 모델은, 귀의 영상을 포함하는 2차원 화상이 입력된 경우에, 당해 귀에 대응하는 귀 파라미터를 출력하는 모델이다.
- [0057] HRTF 처리 모델 기억부(125)는, HRTF 처리 모델을 기억한다. 상세는 후술하지만, HRTF 처리 모델은, 음향 시뮬레이션 등에 의해 산출된 HRTF의 정보량을 압축하는 처리를 행한다. 또한, 이하의 설명에서는, HRTF 처리 모델에 의해 압축된 HRTF를, HRTF 파라미터라고 칭하는 경우가 있다.
- [0058] HRTF 학습 데이터 기억부(126)는, 귀의 영상을 포함하는 화상으로부터 HRTF를 산출하기 위한 모델(후술하는 HRTF 학습 모델)을 생성하기 위한 학습 데이터를 기억한다. 구체적으로는, HRTF 학습 데이터 기억부(126)는, 학습 데이터로서, 귀의 형상을 나타내는 귀 파라미터와, 당해 귀 파라미터에 기초하여 특정되는 귀의 형상에 대응한 HRTF가 조합된 데이터를 기억한다.
- [0059] HRTF 학습 모델 기억부(127)는 HRTF 학습 모델을 기억한다. HRTF 학습 모델은, 귀의 영상을 포함하는 화상이 입력된 경우에, 당해 귀에 대응하는 HRTF를 출력하는 모델이다. 예를 들어, HRTF 학습 모델은, 귀의 영상을 포함하는 화상을 취득한 경우에, 귀 파라미터 추정 모델을 사용하여 당해 귀에 대응하는 귀 파라미터를 출력시키고, 추가로, 당해 귀 파라미터에 대응하는 HRTF를 출력한다.
- [0060] 제어부(130)는, 예를 들어 CPU(Central Processing Unit)나 MPU(Micro Processing Unit) 등에 의해, 정보 처리 장치(100) 내부에 기억된 프로그램(예를 들어, 본 개시에 따른 정보 처리 프로그램)이 RAM(Random Access Memory) 등을 작업 영역으로 하여 실행됨으로써 실현된다. 또한, 제어부(130)는, 컨트롤러(controller)이며, 예를 들어 ASIC(Application Specific Integrated Circuit)나 FPGA(Field Programmable Gate Array) 등의 집적 회로에 의해 실현되어도 된다.
- [0061] 도 3에 도시한 바와 같이, 제어부(130)는, 학습부(131)와, 추정부(140)를 갖는다. 또한, 학습부(131)는, 접수부(132)와, 귀 모델 학습부(133)와, 화상 생성부(134)와, 귀 파라미터 학습부(135)와, HRTF 학습부(136)를 갖고, 이하에 설명하는 정보 처리의 기능이나 작용을 실현 또는 실행한다. 또한, 추정부(140)는, 취득부(141)와, 산출부(142)와, 제공부(143)를 갖고, 이하에 설명하는 정보 처리의 기능이나 작용을 실현 또는 실행한다. 또한, 제어부(130)의 내부 구성은, 도 3에 도시한 구성에 한하지 않고, 후술하는 정보 처리를 행하는 구성이라면 다른 구성이어도 된다.
- [0062] 우선, 학습부(131)에 대하여 설명한다. 학습부(131)는, 각종 데이터에 관한 학습 처리를 행하여, 추정부(140)가 이용하는 각종 모델을 생성한다.
- [0063] 또한, 학습부(131)는, 다양한 데이터에 기초하여 모델을 생성하기 위한 학습을 행하지만, 하기에서 설명하는 학습 처리는 일레이며, 학습부(131)가 실행하는 학습 처리의 종별은 어떠한 종별에 특정되지는 않는다. 예를 들어, 학습부(131)는, 뉴럴 네트워크, 서포트 벡터 머신(support vector machine), 클러스터링, 강화 학습 등의 각종 학습 알고리즘을 이용하여 모델을 생성해도 된다.
- [0064] 접수부(132)는, 각종 정보를 접수한다. 예를 들어, 접수부(132)는, 인체로부터 채취된 귀 모양의 CT 스캔 데이터를 접수한다. 접수부(132)는, 접수한 데이터를 귀 모양 정보 기억부(121)에 저장한다.
- [0065] 귀 모델 학습부(133)는, 귀 모델에 관한 학습 처리를 행하여, 귀 모델을 생성한다. 귀 모델 학습부(133)는, 생

성된 귀 모델을 귀 모델 기억부(122)에 저장한다.

- [0066] 여기서, 도 5를 이용하여, 접수부(132) 및 귀 모델 학습부(133)가 실행하는 학습 처리의 예에 대하여 설명한다. 도 5는, 본 개시에 따른 귀 모델에 관한 학습 처리의 일례를 나타내는 도면이다.
- [0067] 도 5에 도시한 바와 같이, 접수부(132)는, 귀 모양으로부터 채취되어, 스캔된 데이터를 접수하고, 접수된 데이터를 귀 모양 정보 기억부(121)에 저장한다. 또한, 접수부(132)는, 접수한 데이터를 귀 모델 학습부(133)에 보낸다(스텝 S16).
- [0068] 귀 모델 학습부(133)는, 취득한 귀 모양 데이터를 상동화하고, 귀 모양의 상동 데이터를 생성한다(스텝 S17). 여기서, 상동화란, 3D 모델의 정점 수나 폴리곤의 구성을, 기준이 되는 3D 모델과 동일해지도록 통일하는 것을 말한다. 이 경우, 상동화 전후에 형상이 변화하지 않도록 유의한다. 또한, 귀 모델 학습부(133)는, 상동 데이터에 대하여 주성분 분석을 행한다(스텝 S18). 이에 의해, 귀 모델 학습부(133)는, 귀의 형상으로부터, 당해 귀의 형상을 나타내는 귀 파라미터를 산출하는 모델(귀 모델)을 생성한다. 귀 모델 학습부(133)는, 생성된 귀 모델을 귀 모델 기억부(122)에 저장한다.
- [0069] 도 3으로 되돌아가서, 설명을 계속한다. 화상 생성부(134)는, 귀의 영상을 포함하는 화상을 생성한다. 예를 들어, 화상 생성부(134)는, 귀 파라미터를 랜덤으로 생성하고, 생성된 귀 파라미터를 귀 모델에 입력하여, 귀의 3D 모델을 생성한다. 또한, 화상 생성부(134)는, 생성된 3D 모델의 텍스처(예를 들어 피부의 색)나, 렌더링 품위(화질 등), CG 렌더링에서의 카메라 각도 등의 파라미터(이하, 「화상 생성 파라미터」라고 칭함)를 랜덤으로 생성한다. 그리고, 화상 생성부(134)는, 생성된 3D 모델과, 복수의 화상 생성 파라미터를 적절히 조합하여 렌더링을 행하고, 귀의 형상이나 피부의 색 등이 다양하게 변화하는 CG 화상을 생성한다.
- [0070] 후술하는 추정 처리에서는 유저로부터 송신되는 화상을 처리에 사용하지만, 유저로부터 송신되는 화상에서는, 유저의 피부의 색이나, 촬영 시의 귀의 각도 등이 다양하게 다를 가능성이 높다고 상정된다. 이 때문에, 이러한 처리에서는, 유저로부터 송신되는 모든 화상에 있어서 귀의 영상을 정확하게 화상 인식한다는 과제가 있다. 화상 생성부(134)는, 상기와 같이 여러가지 상황에 대응한 화상을 대량으로 생성함으로써, 화상 인식의 정밀도를 향상시켜, 상기 과제를 해결한다.
- [0071] 귀 파라미터 학습부(135)는, 귀의 영상을 포함하는 화상과 귀 파라미터의 관계성을 학습함으로써, 귀 파라미터 추정 모델을 생성한다. 귀 파라미터 학습부(135)는, 본 개시에 따른 제1 학습부에 대응한다. 귀의 영상을 포함하는 화상은, 실제로 인물의 귀를 촬영한 화상이어도 되고, 후술하는 바와 같이, 귀 파라미터에 기초하여 생성되는 CG 화상이어도 된다.
- [0072] 예를 들어, 귀 파라미터 학습부(135)는, 귀 파라미터에 기초하여 생성된 귀의 3차원 데이터와, 머리부의 3차원 데이터를 합성한 3차원 데이터를 렌더링함으로써 얻어지는 귀 화상과 당해 귀 파라미터의 관계성을 학습함으로써, 귀 파라미터 추정 모델을 생성한다. 구체적으로는, 귀 파라미터 학습부(135)는, 화상 생성부(134)에 의해 생성된 CG 화상과 귀 파라미터의 관계성을 학습한다. 상술한 바와 같이, 화상 생성부(134)는, 랜덤 혹은 규칙적으로 설정한 귀 파라미터에 기초하여 CG 화상을 생성하고 있기 때문에, 당해 CG 화상에 대하여 귀 파라미터는 일의적으로 정해진다. 이 때문에, 귀 파라미터 학습부(135)는, 입력된 CG 화상과 귀 파라미터의 관계성을 학습함으로써, 어떤 화상이 입력된 경우에, 당해 화상에 포함되는 귀의 영상에 대응한 귀 파라미터를 출력하는 모델을 생성할 수 있다. 또한, 귀 파라미터 학습부(135)는, 반드시 머리부의 합성된 3차원 데이터를 렌더링함으로써 얻어지는 귀 화상을 학습에 사용하지는 않아도 된다. 즉, 귀 파라미터 학습부(135)는, 귀 파라미터에 기초하여 생성된 귀의 3차원 데이터만을 렌더링하여 얻어진 귀 화상과 귀 파라미터의 관계성을 학습함으로써, 귀 파라미터 추정 모델을 생성해도 된다.
- [0073] 또한, 귀 파라미터 학습부(135)는, 귀 혹은 머리부의 3차원 데이터의 텍스처, 렌더링에서의 카메라 각도 또는 렌더링에서의 명도 등을 변경한 복수의 귀 화상과 복수의 귀 화상에 공통되는 귀 파라미터의 관계성을 학습함으로써, 귀 파라미터 추정 모델을 생성한다. 이와 같이, 귀 파라미터 학습부(135)는, 다양한 양태의 귀 화상을 사용하여 학습을 행함으로써, 어떤 화상이 입력되어도 정확하게 귀 파라미터를 출력할 수 있는(예를 들어, 입력된 화상에 포함되는 귀 파라미터 이외의 모든 변화에 대하여 안정적인 추정을 행할 수 있는), 안정감이 있는 로버스트(robust)한 모델을 생성할 수 있다.
- [0074] 여기서, 도 6을 이용하여, 화상 생성부(134) 및 귀 파라미터 학습부(135)가 실행하는 학습 처리의 예에 대하여 설명한다. 도 6은, 본 개시에 따른 귀 파라미터 추정 모델에 관한 학습 처리의 일례를 나타내는 도면이다.
- [0075] 도 6에 도시한 바와 같이, 화상 생성부(134)는, 귀 모델 기억부(122)를 참조하여(스텝 S21), 귀 모델을 취득한

다. 또한, 화상 생성부(134)는, 귀 파라미터에 대응하는 난수나, CG의 텍스처나 렌더링의 카메라 각도 등에 대응하는 난수를 생성한다(스텝 S22). 즉, 화상 생성부(134)는, 귀 화상을 생성하기 위한 각종 파라미터(화상 생성 파라미터)를 생성한다.

[0076] 그리고, 화상 생성부(134)는, 화상 생성 파라미터 중 귀 파라미터를 취득하고(스텝 S23), 취득한 귀 파라미터를 귀 모델에 입력하여, 귀의 3D 모델을 재구성한다(스텝 S24).

[0077] 계속해서, 화상 생성부(134)는, 화상 생성 파라미터 중 CG 텍스처 등의 파라미터를 취득하고(스텝 S25), 취득한 파라미터를 입력하여, 3D 모델에 3DCG 렌더링을 실시한다(스텝 S26). 또한, 렌더링에 사용할 때의 머리부에 대해서는, 예를 들어 복수의 인물(예를 들어, 귀 모양을 채취한 복수의 인물)의 평균적인 머리부나, 샘플로서 이용되는 머리부의 3D 모델 등을 사용한다. 여기서, 머리부의 3D 모델은, 귀 모델과 마찬가지로, 복수의 인물의 머리부를 3D 스캔하여 얻어지는 3D 데이터를 상동화하여 생성해도 된다. 이 경우, 화상 생성부(134)는, 난수에 의한 파라미터 생성을 행함으로써, 머리부 3D 모델을 랜덤으로 생성시킬 수 있다. 또한, 화상 생성부(134)는, 텍스처에 대해서도, 마찬가지로의 방법에 의해 생성된 텍스처 모델을 작성함으로써, 난수에 의해 다양한 텍스처를 생성할 수 있다.

[0078] 이에 의해, 화상 생성부(134)는, 귀의 영상을 포함하는 화상(귀 화상)을 생성한다. 또한, 화상 생성부(134)는, 텍스처 등의 파라미터를 다양하게 변화시킴으로써, 1개의 3D 모델로부터, 복수의 귀 화상을 생성하는 것이 가능하다.

[0079] 여기서, 도 7을 이용하여, 화상 생성부(134)가 생성하는 귀 화상의 일례를 나타낸다. 도 7은, 본 개시에 따른 귀 화상의 생성 처리의 일례를 나타내는 도면이다.

[0080] 화상 생성부(134)는, 랜덤으로 생성된 귀 파라미터를 사용하여, 귀의 3차원 형상을 나타내는 3D 모델을 생성한다(스텝 S41). 또한, 화상 생성부(134)는, 복수의 인물의 평균적인 머리부의 3차원 형상을 나타내는 데이터에 기초하여 생성된 머리부 3D 모델을 취득한다(스텝 S42). 그리고, 화상 생성부(134)는, 스텝 S41에서 생성된 귀와, 스텝 S42에서 취득한 머리부 3D 모델을 합성하여, 의사적인 인물의 3D 모델을 생성한다.

[0081] 계속해서, 화상 생성부(134)는, 생성된 3D 모델에 대하여 의사적인 촬영을 행하여, 3D 모델로부터 2차원의 화상을 생성하는 처리(렌더링)를 행한다. 예를 들어, 화상 생성부(134)는, 3D 모델의 귀의 정면의 각도를 의사적인 촬영 각도로 설정하여, 귀의 영상이 대략 중앙이 되는 화상을 생성한다.

[0082] 여기서, 화상 생성부(134)는, 3D 모델에 화상 생성 파라미터를 랜덤으로 입력함으로써, CG의 텍스처(구체적으로는 피부의 색 등)나, 렌더링의 품위(해상도 등)나, 머리부에 대한 귀의 위치 등을 다양하게 변경한다. 이에 의해, 화상 생성부(134)는, 피부의 색 등이 각각 다른 다수의 화상을 생성할 수 있다(스텝 S43).

[0083] 화상군(20)은, 화상 생성부(134)에 의해 생성된 복수의 귀 화상을 나타낸다. 이와 같이, 화상 생성부(134)는, 다양한 귀 화상을 다수 생성함으로써, 후술하는 귀 화상의 인식 정밀도를 향상시킬 수 있다.

[0084] 도 6으로 되돌아가서, 설명을 계속한다. 화상 생성부(134)는, 생성된 귀 화상을 귀 화상 기억부(123)에 기억한다(스텝 S27). 또한, 화상 생성부(134)는, 생성된 귀 화상에 대응지어, 각각의 화상을 생성했을 때의 화상 생성 파라미터를 귀 화상 기억부(123)에 기억한다(스텝 S28). 이에 의해, 화상 생성부(134)는, 학습 데이터로서, 귀 파라미터가 대응지어진 귀 화상을 대량으로 보유할 수 있다. 예를 들어, 화상 생성부(134)는, 학습 데이터로서, 귀 파라미터가 대응지어진 귀 화상을 대량으로 보유할 수 있다.

[0085] 계속해서, 귀 파라미터 학습부(135)는, 귀 화상 기억부(123)를 참조하여(스텝 S29), 귀 화상과 귀 파라미터를 취득한다. 그리고, 귀 파라미터 학습부(135)는, 귀 화상과 귀 파라미터의 관계성을 학습하여, 귀 파라미터 추정 모델을 생성한다. 귀 파라미터 학습부(135)는, 생성된 귀 파라미터 추정 모델을 귀 파라미터 추정 모델 기억부(124)에 저장한다(스텝 S30).

[0086] 귀 파라미터 추정 모델은, 예를 들어 화상의 특징량을 추출하기 위해서 유용한 컨벌루션 뉴럴 네트워크(Convolutional Neural Network) 등을 이용하여 생성된다. 또한, 학습에서의 비용 식(비용 함수)은, 예를 들어 하기 식 (1)로 나타내어진다.

[0087]
$$L(\alpha_{\text{true}}, \alpha_{\text{est}}) = \|A_{\text{car}} \alpha_{\text{true}} - A_{\text{car}} \alpha_{\text{est}}\|_2^2 \quad \dots (1)$$

[0088] 식 (1)에 있어서, 「 α_{true} 」는, 귀 파라미터의 참값을 나타낸다. 또한, 「 α_{est} 」는, 귀 파라미터의 추정값을 나

타낸다. 또한, 「 A_{ear} 」은, 주성분 분석에 의한 귀 모델을 나타낸다. 또한, 우변의 거리 함수는, L2놈(유클리드 거리)을 나타낸다. 또한, 귀 모델 파라미터의 참값은, 예를 들어 귀 모양의 채취 시에 실측된 인물의 귀를 나타내는 파라미터 등을 사용할 수 있다. 즉, 귀 화상을 생성할 때 사용한 귀 파라미터가 참값이며, 학습 중의 귀 파라미터 추정 모델에, 당해 귀 화상을 입력했을 때 출력하는 값이 추정값으로 된다. 정보 처리 장치(100)는, 학습 처리로서, 현재의 추정값에 대한 비용 식의 값을 최소화하도록 계수를 갱신한다.

- [0089] 여기서, 도 8을 이용하여, 귀 파라미터 학습부(135)가 생성된 귀 파라미터 추정 모델에 대하여 설명한다. 도 8은, 본 개시에 따른 귀 파라미터 추정 모델을 설명하기 위한 도면이다.
- [0090] 정보 처리 장치(100)는, 귀 화상(30)을 취득한 경우, 취득한 귀 화상(30)을 귀 파라미터 추정 모델에 입력한다. 귀 파라미터 추정 모델은 예를 들어 컨벌루션 뉴럴 네트워크의 구조를 갖고, 입력된 귀 화상(30)에서의 수 픽셀 마다의 직사각형부로 분할하면서, 귀 화상(30)을 나타내는 특징량을 구한다. 최종적으로, 귀 파라미터 추정 모델은, 귀 화상(30)을 나타내는 특징량으로서, 귀 화상(30)에 포함되는 귀의 영상에 대응하는 귀 파라미터를 출력한다(스텝 S45).
- [0091] 또한, 정보 처리 장치(100)는, 출력된 귀 파라미터를 귀 모델에 입력함으로써, 귀 화상(30)에 포함되는 귀에 대응한 귀 모양(3D 모델)을 재구성할 수 있다. 도 8에 나타난 3D 모델(40)은, 귀 파라미터에 기초하여, 귀 화상(30)에 포함되는 귀를 CG로 재구성한 3D 모델을 나타낸다.
- [0092] 도 3으로 되돌아가서, 설명을 계속한다. HRTF 학습부(136)는, 귀의 형상에 관한 정보와 HRTF의 관계성을 학습함으로써, HRTF에 관한 각종 모델을 생성한다. 예를 들어, HRTF 학습부(136)는, 귀의 영상을 포함하는 화상과 귀에 대응하는 HRTF의 관계성을 학습함으로써, HRTF를 산출하는 학습 완료 모델을 생성한다. HRTF 학습부(136)는, 본 개시에 따른 제2 학습부에 대응한다.
- [0093] 예를 들어, HRTF 학습부(136)는, 귀 파라미터에 기초하여 생성된 귀의 3차원 데이터와 머리부의 3차원 데이터를 합성한 3차원 데이터에 대하여 음향 시뮬레이션을 행하여, 음향 시뮬레이션에 의해 얻어진 HRTF와 귀 파라미터의 관계성을 학습함으로써, 학습 완료 모델을 생성한다.
- [0094] 또한, HRTF 학습부(136)는, 음향 시뮬레이션에 의해 얻어진 HRTF의 정보량을 압축하고, 압축된 HRTF와 귀 파라미터의 관계성을 학습함으로써, 학습 완료 모델을 생성해도 된다.
- [0095] 또한, HRTF 학습부(136)는, 귀 파라미터에 기초하여 생성된 귀의 3차원 데이터의 청취점을 설정하고, 설정된 청취점을 사용하여 음향 시뮬레이션을 행해도 된다. 청취점이란, 인간이 소리를 청취할 것으로 상정되는 위치를 가상적으로 설정한 것이다. 예를 들어, 청취점의 위치는, 더미 헤드 마이크로폰에 있어서 마이크가 설치되는 위치(더미 헤드에서의 외이도의 입구 등)에 대응한다.
- [0096] 상술한 HRTF 학습부(136)의 각 처리에 관하여, 도 9에, HRTF에 관한 각종 모델의 생성 처리의 흐름을 나타낸다. 도 9는, HRTF에 관한 각종 모델의 생성 처리의 흐름의 개요를 나타내는 도면이다.
- [0097] 도 9에서는, HRTF 학습부(136)가 유저로부터 송신된 화상에 기초하여, 소정의 학습 처리를 행하는 예를 나타내고 있다. 이 경우, 유저는, 유저 단말기(10)를 이용하여 자신의 귀(정확하게는, 귀를 포함하는 머리부)를 촬영한다(스텝 S51). 그 후, 유저 단말기(10)는, 촬영 사진으로부터 귀의 영상이 포함되는 범위를 특정하고, 특정한 범위를 잘라내어, 귀 화상을 취득한다고 하는 전처리를 행한다(스텝 S52).
- [0098] 그 후, HRTF 학습부(136)는, 귀 파라미터 추정 모델을 이용하여, 유저로부터 송신된 귀 화상에 포함되는 귀의 귀 파라미터를 산출한다(스텝 S53). 또한, HRTF 학습부(136)는, 귀 파라미터에 기초하여 귀의 3D 모델을 재구성하고, 또한 재구성한 귀에 머리부 3D 모델을 조합하여, 인물의 3D 모델을 생성한다(스텝 S54).
- [0099] 계속해서, HRTF 학습부(136)는, 생성된 3D 모델에 대하여 음향 시뮬레이션을 행하여, 당해 3D 모델의 개인화 HRTF(스텝 S55)를 구한다. 이에 의해, HRTF 학습부(136)는, 유저로부터 송신되는 귀 화상에 포함되는 귀와, 개인화 HRTF가 대응지어진 학습 데이터를 얻을 수 있다.
- [0100] 또한, 도 9의 예에서는, 음향 시뮬레이션에 의해 얻어진 개인화 HRTF와 귀의 데이터가 대응지어진 학습 데이터를 생성하는 예를 나타내었지만, HRTF 학습부(136)는, 반드시 음향 시뮬레이션에 의해 개인화 HRTF를 얻을 필요는 없는 경우도 있다. 예를 들어, 귀 모양을 채취한 인물의 개인화 HRTF(무향실 등에서 측정 장치를 사용하여 얻어진 HRTF)가 얻어진 경우에는, HRTF 학습부(136)는, 실측된 개인화 HRTF와, 인물의 귀 모양(귀 파라미터)이 대응지어진 학습 데이터를 취득해도 된다.

- [0101] HRTF 학습부(136)는, 음향 시뮬레이션 시에, 인물의 3D 모델에서의 청취점을 자동적으로 설정한다. 이 점에 대하여, 도 10을 이용하여 설명한다. 도 10은, 본 개시에 따른 3D 모델의 재구성을 설명하기 위한 도면이다.
- [0102] HRTF 학습부(136)는, 랜덤으로 생성된 귀 파라미터로부터 귀 3D 모델을 재구성한다(스텝 S71). 계속해서, HRTF 학습부(136)는, 귀 3D 모델을 머리부 3D 모델에 조합해서 인물의 3D 모델을 생성한다(스텝 S72). 또한, HRTF 학습부(136)는, 3D 모델에서의 귀의 형상에 기초하여, 음원의 청취점(60)을 설정한다(스텝 S73). 예를 들어, HRTF 학습부(136)는, 사전에 귀의 형상과 음원의 청취점의 위치를 대응지은 정보를 학습함으로써, 청취점을 자동적으로 설정할 수 있다. 예를 들어, HRTF 학습부(136)는, 3D 모델을 생성한 시점에서, 당해 3D 모델에서의 청취점의 위치를 추정하고, 청취점을 자동적으로 설정한다. 청취점은, 예를 들어 귀의 외이도에 대응하고, 일반적으로, 귀의 형상으로부터 위치를 추정할 수 있다.
- [0103] 그 후, HRTF 학습부(136)는, 음향 시뮬레이션에서의 3D 모델의 계산상의 제약을 충족하도록, 생성된 3D 모델을 리메싱한다(스텝 S74). 이것은, 3D 모델의 시뮬레이션에 있어서, 예를 들어 폴리곤의 총 수에 대한 상한이나, 정점을 연결하는 에지의 길이 등이 시뮬레이션의 조건에 의해 정해지는 경우가 있기 때문이다. 즉, HRTF 학습부(136)는, 생성된 3D 모델을 시뮬레이션에 적용하기 전에, 계산상의 제약을 충족하도록 3D 모델을 적절히 리메싱하여, 시뮬레이션을 적절하게 행할 수 있도록 처리한다. 그리고, HRTF 학습부(136)는, 생성된 3D 모델 및 설정한 청취점(60)에 대하여 음향 시뮬레이션을 행하여, 개인화 HRTF를 산출한다(스텝 S75).
- [0104] 다음으로, 도 11을 이용하여, HRTF에 관한 모델의 생성 처리에 관한 상세한 흐름을 설명한다. 도 11은, HRTF에 관한 모델의 생성 처리의 상세를 설명하는 도면이다.
- [0105] 도 10에서 도시한 바와 같이, HRTF 학습부(136)는, 머리부 합성을 행한 후에 (스텝 S81), 음향 시뮬레이션을 행한다(스텝 S82). HRTF 학습부(136)는, 음향 시뮬레이션에서 얻어진 측정 데이터를 분석하여(HRTF 후처리), 개인화 HRTF를 나타내는 수치를 산출한다(스텝 S83). 또한, HRTF 후처리는, 예를 들어 음향 시뮬레이션에 의해 얻어진 HRIF(Head-Related Impulse Response)를 푸리에 변환하여 HRTF를 얻는 것 등을 의미한다.
- [0106] 여기서, HRTF 학습부(136)는, HRTF 처리 모델 기억부(125)를 참조하여(스텝 S84), 산출된 HRTF를 처리하기 위한 모델(HRTF 처리 모델)에 HRTF를 입력한다. 이에 의해, HRTF 학습부(136)는, 차원이 삭감된 HRTF를 얻는다(스텝 S85). 즉, HRTF 학습부(136)는, HRTF 처리 모델로부터, 차원이 삭감된 HRTF인 HRTF 파라미터를 출력한다(스텝 S86).
- [0107] 이와 같이, HRTF 학습부(136)는, 음향 시뮬레이션에서 얻어진 HRTF를 그대로 처리에 사용하는 것이 아니라, 차원을 삭감시킨 HRTF 파라미터를 사용하여 처리를 행한다. 이것은, HRTF가 매우 차원수가 많은 함수이며, 그대로 모델의 생성 처리나 산출 처리를 행할 경우, 계산 처리의 부하가 커지기 때문이다.
- [0108] HRTF 학습부(136)는, 음향 시뮬레이션을 행한 머리부에 관한 데이터(머리부 합성의 기초가 되는 데이터이며, 예를 들어 귀 파라미터 등)와, 산출된 HRTF 파라미터를 대응지어, HRTF 학습 데이터 기억부(126)에 저장한다(스텝 S87).
- [0109] 그 후, HRTF 학습부(136)는, 다른 귀 파라미터를 새롭게 랜덤으로 생성하고(스텝 S88), 새롭게 생성된 귀 파라미터를 사용하여 머리부 합성을 행한다(스텝 S89). 스텝 S81부터 스텝 S89를 반복함으로써, HRTF 학습부(136)는, 학습에 요하는 학습 데이터를 수집한다.
- [0110] 그 후, 충분한 학습 데이터가 축적된 경우, HRTF 학습부(136)는, HRTF 학습 데이터 기억부(126)를 참조하여(스텝 S90), 귀 파라미터와 HRTF(정확하게는 HRTF 파라미터)의 관계성을 학습한다(스텝 S91). 이러한 학습에 의해, HRTF 학습부(136)는, 귀 파라미터로부터 직접적으로 HRTF를 얻기 위한 HRTF 학습 모델을 생성하고, 생성된 HRTF 학습 모델을 HRTF 학습 모델 기억부(127)에 저장한다.
- [0111] 다음으로, 도 12를 이용하여, HRTF와 HRTF 파라미터의 관계에 대하여 설명한다. 도 12는, 본 개시에 따른 HRTF의 압축 및 복원을 설명하기 위한 도면이다.
- [0112] 도 12에 도시한 바와 같이, HRTF 학습부(136)는, 음향 시뮬레이션에 의해 얻어진 HRIFHRTF(도 12의 예에서는, 가령 1000방향×500탭으로 함)에 FFT(Fast Fourier Transform)를 행한다(스텝 S101). 이러한 처리에 의해, HRTF 학습부(136)는, 진폭 특성을 추출하고(스텝 S102), 예를 들어 청감 상의 감도가 낮은 주파수 성분 등을 쉰링하는 처리를 행한다(스텝 S103). 구체적으로는, HRTF는, 각도(θ , ϕ 라 함)와, 주파수(f 라 함)에 관한 함수 $HRTF(\theta, \phi, f)$ 로 표현된다. 이때, 주파수의 빈(bin) 수를 k 라 하면, 함수에 입력되는 주파수 f 는, $f_k=f_0, f_1,$

f_2, \dots, f_{k-1} 로 표현된다. 즉, HRTF는, 하나의 방향 및 하나의 귀에 대하여, 복소 k 차원을 갖게 된다. 여기서, 나이키스트 주파수($f_{k/2}$) 이후의 HRTF는, 주파수 $f_{k/2}$ 이전의 복소 공역의 폴딩이기 때문에, 정보 처리에 있어서는, 주파수 빈으로서, $f_0=0$ 부터 나이키스트 주파수($f_{k/2}$)까지의 $(k/2)+1$ 개만을 사용할 수 있다. 또한, 적어도 하나 이상의 주파수 빈에 대하여, 절댓값을 이용할 수 있다. 예를 들어, f_0 부터 $f_{k/2}$ 까지의 모든 주파수를 절댓값으로 변환하면, 변환 후의 함수 H_2 는, 이하의 식 (2)로 표현된다.

[0113]
$$H_2(\theta, \phi, k) = |HRTF(\theta, \phi, f_k)| (f_k = f_0, f_1, \dots, f_{k/2}) \quad \dots (2)$$

[0114] 즉, HRTF 학습부(136)는, 원래의 HRTF의 차원을 실수 $(k/2)+1$ 차원으로 압축할 수 있다. 또한, HRTF 학습부(136)는, 상기 식 (2)의 H_2 에 대하여 주파수 압축을 행하고, 더욱 $(k/2)+1$ 보다 적은 차원으로 삭감하는 것이 가능하다. 차원의 압축에 관해서는 여러가지 기지의 방법이 존재한다. 예를 들어, HRTF 학습부(136)는, 함수에 대하여 케프스트럼 변환을 행하여, 1개 이상, $(k/2)+1$ 미만의 주파수 빈만을 취득하는 등의 방법을 이용한다. 일례로서, HRTF 학습부(136)는, 복수의 주파수 빈의 평균값을 구하고, 평균값에 기초하여 차원을 삭감한다. 예를 들어, 주파수 빈을 하기 식 (3)으로 나타낸 경우(a_1, L, l 은, 각각 0 이상의 정수), $f_{a_1} \leq f'_{l_1} < f_{a_1+1}$ 을 충족하는 l 을 이용하여, 새로운 함수 H_3 은, 하기 식 (4)로 표현된다.

[0115]
$$0 \leq a_0 < a_1 < a_2 \dots < a_{L-1} \leq K/2, L < K/2+1 \quad \dots (3)$$

[0116]
$$H_3(\theta, \phi, l) = \frac{1}{a_{l+1} - a_l} \sum_{k=a_l}^{a_{l+1}-1} H_2(k) \quad \dots (4)$$

[0117] 이에 의해, HRTF 학습부(136)는, $(K/2)+1$ 차원으로 표시되어 있던 함수 H_2 를 L 차원으로 삭감할 수 있다. 또한, 평균값을 구하는 방법은, 상기에 한하지 않고, 예를 들어 제곱 평균 평방근이나 가중치 부여 평균 등을 이용하여 구해도 된다. 이에 의해, HRTF는, 예를 들어 1000방향 \times 50차원 정도로 삭감된다. 또한, 함수 H_3 으로 삭감된 차원을 되돌리는 경우(후술하는 스텝 S110 등)에는, 선형 보간이나 스플라인 보간 등, HRTF 학습부(136)는, 다양한 방법을 이용하여 복원하는 것이 가능하다. 차원을 되돌린 함수 H'_2 는, 함수 H_2 에 대하여 보다 매끄러운 특성이 될 것이 예상되지만, a_1 의 선택하는 방법을 고안함으로써, 보다 청감 상 영향이 적어지는 $H'_2(\theta, \phi, k)$ 가 얻어진다. 일례로서, 높은 주파수일수록, f_{a_1} 과 f_{a_1+1} 의 주파수 간격이 커지도록 a_1 을 선택한다고 하는 고안이 있을 수 있다.

[0118] HRTF 학습부(136)는, 차원이 삭감된 HRTF에, 추가로 구면 조화 피팅의 처리를 행함으로써, 50계수 \times 50차원 정도로 정보량이 압축된다(스텝 S104). 여기서, 구면 조화 피팅이란, 구면 조화 함수를 사용하여, 압축된 주파수마다 공간 방향의 피팅을 행하는 것을 말한다. HRTF와 구면 조화 함수의 관계는, 하기 식 (5)로 나타내어진다.

[0119]
$$H(\theta, \phi, f) = \sum_{n=0}^{\infty} \sum_{m=-n}^n h_{nm}(f) Y_n^m(\theta, \phi) \quad \dots (5)$$

[0120] 상기 식 (5)에 나타낸 바와 같이, 구면 조화 함수 Y 는, 계수 $h_{nm}(f)$ 로 표현된다. 이때의 차원수 n 을, 어떤 유한한 N 으로 잘라버림으로써, 계수 $h_{nm}(f)$ 의 차원을, 원래의 HRTF의 차원수(방향수)보다도 작게 할 수 있다. 이러한 점은, 인간의 지각에 있어서 불필요한, 공간적으로 너무 미세한 진폭을 무시하고, 매끄러운 형상만을 취득하는 것을 의미한다. 또한, 계수 h_{nm} 의 벡터 $h=(h_{00}, h_{1-1}, \dots)^T$ 를 구하기 위해서는, 예를 들어 최소 제곱법 등이 이용된다.

[0121]
$$E = \|H - Yh\|^2 + \lambda \|h\|^2 \quad \dots (6)$$

[0122] 즉, 상기 식 (6)에서의 Y 를 구면 조화 함수의 행렬이라 하고, H 를 구면 조화 함수의 행렬이라 한 경우에, 좌변의 E 를 최소화하는 h 를 구한다. 또한, 상기 식 (6) 우변의 제2항은, 정규화항이기 때문에, 임의의 값 λ 를 선택해도 된다(예를 들어, $\lambda=0$ 이어도 됨). 그렇게 하면, 상기 h 는, 하기 식 (7)로 표현된다.

[0123]

$$\mathbf{h} = (\mathbf{Y}^T \mathbf{Y} + \lambda \mathbf{I})^{-1} \mathbf{Y}^T \mathbf{H} \quad \dots (7)$$

[0124]

상기 식 (7)을 이용함으로써, HRTF 학습부(136)는, 필요한 주파수에 대응하는 각각의 \mathbf{h} 를 구할 수 있다. 또한, HRTF 학습부(136)는, 주성분 분석에 의한 차원 압축을 행함으로써, HRTF의 정보량을 약 수백 차원에서 표현 가능하도록 압축한다(스텝 S105). 이러한 정보가, HRTF 파라미터로 된다(스텝 S106).

[0125]

또한, 주파수 씨닝을 한 후에 구면 조화 피팅이 행해지는 경우, 상기 f 의 값은, 씨닝 후의 대표 주파수로 된다. 또한, HRTF 학습부(136)는, 구면 조화 피팅의 후에, 주파수 씨닝을 행해도 된다. 또한, 공간적으로 차원을 압축하는 방법은, 구면 조화 함수나 주성분 분석과 같은 선형 결합에 한하지 않고, 어떠한 방법을 이용해도 무방하다. 예를 들어, HRTF 학습부(136)는, 커널 주성분 분석과 같이 비선형인 방법을 이용해도 된다. 또한, HRTF 학습부(136)는, 주파수 f 에 따라서, 구면 조화 함수의 중단 차수 N 을 변화시켜, $N(f)$ 라는 값을 이용해도 된다. 또한, 0부터 N 까지의 차원수나 자릿수에서 사용하지 않는 계수 h_{nm} 이 있어도 된다. 또한, HRTF 학습부(136)는, 좌우의 HRTF에 대하여 각각 구해도 되고, 좌우의 합과 차 등으로 변환하고 나서 각각 구해도 된다. 또한, 피팅 대상의 HRTF는, 진폭의 절댓값이나 그 대수 표현 등 다양한 변환을 행한 것이어도 무방하다.

[0126]

계속해서, HRTF 학습부(136)는, 스텝 S101 내지 스텝 S106은 역의 흐름으로 처리를 행함으로써, HRTF를 복호하는 것이 가능하다. 우선, HRTF 학습부(136)는, HRTF 파라미터를 취득하고(스텝 S107), 주성분 분석에 의한 차원 압축의 복원을 행한다(스텝 S108). 또한, HRTF 학습부(136)는, 구면 조화 재구성 처리를 행함과 함께(스텝 S109), 주파수의 보간을 행한다(스텝 S110). 또한, HRTF 학습부(136)는, 진폭 특성을 얻어서(스텝 S111), 최소 위상 복원을 행한다(스텝 S112). 최소 위상 복원에 대해서는, 여러가지 기지의 방법이 이용되어도 된다. 예를 들어, HRTF 학습부(136)는, 상기 함수 H^2 의 나이키스트 주파수 이후를 폴딩하여 복원한 함수인 함수 $H^1(\theta, \phi, k)$ 의 대수를 역 푸리에 변환(IFFT(Inverse Fast Fourier Transform)) 하고, 그 실부를 취한다. 또한, 이 영역에서 적절한 윈드 처리를 행하여, 그 지수 함수를 역 푸리에 변환하고, 실부를 취함으로써, 최소 위상 복원을 행할 수 있다. 예를 들어, 이하와 같은 관계식 (8)이 각각 성립한다.

$$\begin{aligned} H'_1(\theta, \phi, k) &= \begin{cases} H'_2(\theta, \phi, k) & (0 \leq k \leq K/2 \text{일 때}) \\ H'_2(\theta, \phi, K-k) & (K/2 < k \leq K-1 \text{일 때}) \end{cases} \\ H_c(\theta, \phi) &= \text{Re}\{F^{-1}[H'_1(\theta, \phi)]\} \\ H_w(\theta, \phi, k) &= \begin{cases} H_c(\theta, \phi, k) & (K=0, K/2 \text{일 때}) \\ 2H_c(\theta, \phi, k) & (1 \leq k \leq K/2-1 \text{일 때}) \\ 0 & (K/2 \leq k \leq K-1 \text{일 때}) \end{cases} \quad \dots (8) \\ H_m(\theta, \phi) &= \text{Re}\{F^{-1}[\exp(H_w(\theta, \phi))]\} \end{aligned}$$

[0127]

[0128]

또한, HRTF 학습부(136)는, 최소 위상 복원된 좌우의 HRIR(h_m)에 대하여, 추정된 ITD(양쪽 귀 사이 시간차(Interaural Time Difference)), 혹은 사전에 준비된 ITD를 부가해도 된다. 또한, ITD는, 좌우의 HRIR의 균지연의 차분에 기초하여, 예를 들어 하기 식 (9) 및 식 (10)에 의해 구해진다.

[0129]

$$\text{ITD}_g(\theta, \phi, f) = \frac{1}{2\pi} \left(\frac{d\phi_L}{df} - \frac{d\phi_R}{df} \right) \quad \dots (9)$$

[0130]

$$\phi_L = \arctan \left(\frac{\text{Im}(\text{HRTF}_L(f))}{\text{Re}(\text{HRTF}_L(f))} \right) \quad \dots (10)$$

[0131]

또는, ITD는, 좌우에서의 시간축 상의 상호 상관을 구하고, 상관 계수가 최대가 되는 시간을 ITD라고 정의함으로써 산출되어도 된다. 이 경우, ITD는, 예를 들어 하기 식 (11) 및 식 (12)에 의해 구해진다.

[0132]

$$\tau = a \frac{\int_{-\infty}^{+\infty} h_L(t+\tau)h_R(t)dt}{\sqrt{\int_{-\infty}^{+\infty} h_L^2(t)dt \cdot \int_{-\infty}^{+\infty} h_R^2(t)dt}} \quad \dots (11)$$

$$ITD_{\text{corr}}(\theta, \phi) = \tau_{\text{max}} \dots (12)$$

[0133]

[0134]

예를 들어, HRTF 학습부(136)는, 좌우의 HRIR에 대하여, 좌측을 우측보다 d 샘플만큼 지연시키는 경우, 하기 식 (13)과 같은 관계식을 이용한다.

$$h_L(i) = \begin{cases} 0 & i < d \text{ 일 때} \\ h_{m,L}(i-d) & i \geq d \text{ 일 때} \end{cases} \dots (13)$$

[0135]

[0136]

이때, 상기 식 (13)에서의 h_L 은, $h_{m,L}$ 보다 d만큼 긴 임펄스 응답으로 되지만, 길이를 $h_{m,L}$ 과 동일하게 하기 위해서는, 상기 식 (13)의 후반 부분을 삭제한다. 이때, HRTF 학습부(136)는, 임의의 창, 직사각형 창, 해닝 창 등의 처리를 행해도 된다. 또한, HRTF 학습부(136)는, 방향마다 상대적으로 ITD를 부가할뿐만 아니라, 공간 전체에서, 방향 간의 상대적인 시간차까지 포함해서 지연을 부가해도 된다. 그 때는, HRTF 학습부(136)는, ITD뿐만 아니라 방향 간의 상대적인 시간차를 나타내는 정보를 취득한다. 또한, HRTF 학습부(136)는, ITD가 주파수의 함수가 되어 있는 경우에는, 주파수 영역에서 ITD를 부가해도 되고, 대푯값 또는 평균값을 구해서 ITD를 부가해도 된다. 그리고, HRTF 학습부(136)는, 원래의 형식의 HRIR을 얻은 후, 역 푸리에 변환을 행함으로써 HRTF를 얻는다.

[0137]

이와 같이, HRTF 학습부(136)는, 원래의 HRTF보다도 정보량이 적은 HRTF 파라미터로 압축을 행하여, 압축된 형식으로 HRTF 학습 모델의 생성 처리나, 후술하는 HRTF의 산출 처리를 행해도 된다. 또한, 상기한 바와 같이 HRTF의 압축은, 예를 들어 인간의 청감에서는 위상의 변화에 민감하지 않음을 이용하거나, 청감에 영향을 주기 어려운 주파수를 우선적으로 씌닝하거나 하는 등, 청각 특성을 이용한 차원 삭감을 행한다. 이에 의해, HRTF 학습부(136)는, HRTF의 특성인 청감 상의 정위감을 손상시키지 않고, 또한, 정보 처리를 고속화시킬 수 있다.

[0138]

도 3으로 되돌아가서, 설명을 계속한다. 추정부(140)는, 유저로부터 송신된 화상에 기초하여, 당해 유저에 대응하는 HRTF의 추정 처리를 행한다.

[0139]

취득부(141)는, 유저의 귀의 영상을 포함하는 화상을 취득한다. 예를 들어, 취득부(141)는, 유저 단말기(10)에 의해 촬영된 화상으로부터, 유저의 귀 주변만이 절취된 귀 화상을 취득한다.

[0140]

또한, 취득부(141)는, 취득한 귀 화상을 귀 파라미터 추정 모델에 입력함으로써, 당해 화상에 포함되는 귀의 특징을 나타내는 귀 파라미터를 취득해도 된다.

[0141]

산출부(142)는, 귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 HRTF를 출력하도록 학습된 학습 완료 모델(HRTF 학습 모델)을 이용하여, 취득부(141)에 의해 취득된 화상에 기초하여, 유저에 대응하는 HRTF(개인화 HRTF)를 산출한다.

[0142]

구체적으로는, 산출부(142)는, 취득부(141)에 의해 취득된 귀 파라미터를 HRTF 학습 모델에 입력함으로써, 유저에 대응하는 개인화 HRTF를 산출한다.

[0143]

또한, 산출부(142)는, 개인화 HRTF를 산출하는 경우, 우선은 HRTF 파라미터를 산출하고, 산출된 HRTF 파라미터를 복호함으로써, HRTF를 산출해도 된다. 이와 같이, HRTF의 정보량이 압축된 상태에서 일련의 처리를 행함으로써, 산출부(142)는, 처리를 고속화할 수 있다. 또한, 산출부(142)는, HRTF 삭감 모델에 표현되지 않는 기이한 HRTF를 출력하는 것을 방지할 수 있기 때문에, 안정적인 출력을 행할 수 있다.

[0144]

제공부(143)는, 네트워크 N을 통해 산출부(142)에 의해 산출된 HRTF를 유저에게 제공한다.

[0145]

여기서, 도 13을 이용하여, 화상으로부터 HRTF를 추정하는 처리의 흐름에 대하여 설명한다. 도 13은, 본 개시에 따른 HRTF의 추정 처리의 흐름을 나타내는 도면이다.

[0146]

도 13에서는, 추정부(140)가 유저로부터 송신된 화상에 기초하여, 화상에 포함되는 귀에 대응하는 HRTF의 추정 처리를 행하는 예를 나타내고 있다. 이 경우, 유저는, 유저 단말기(10)를 이용하여 자신의 귀(정확하게는, 귀를 포함하는 머리부)를 촬영한다(스텝 S131). 그 후, 유저 단말기(10)는, 촬영 사진으로부터 귀의 영상이 포함되는 범위를 특정하고, 특정한 범위를 잘라내어, 귀 화상을 취득한다고 하는 전처리를 행한다(스텝 S132).

[0147]

취득부(141)는, 유저로부터 송신된 귀 화상을 취득한 경우, 취득한 귀 화상을 학습 완료 모델에 입력한다. 구체적으로는, 취득부(141)는, 귀 화상을 귀 파라미터 추정 모델에 입력한다. 귀 파라미터 추정 모델은, 귀 화상

을 나타내는 특징량으로서, 귀 화상에 포함되는 귀의 영상에 대응하는 귀 파라미터를 출력한다. 이에 의해, 취득부(141)는, 화상에 대응하는 귀 파라미터를 취득한다(스텝 S133).

- [0148] 산출부(142)는, 취득된 귀 파라미터를 HRTF 학습 모델에 입력하여, 귀 화상에 대응하는 개인화 HRTF를 산출한다(스텝 S133). 제공부(143)는, 산출된 개인화 HRTF를, 화상의 송신원인 유저 단말기(10)에 제공(송신)한다(스텝 S134).
- [0149] 이와 같이, 정보 처리 장치(100)는, 학습부(131)에 의해 다양한 모델이 생성된 경우, 귀 화상의 취득으로부터 개인화 HRTF의 제공까지를, 일련의 처리로서 행할 수 있다. 이에 의해, 정보 처리 장치(100)는, HRTF의 제공에 관한 유저의 편리성을 향상시킬 수 있다.
- [0150] 또한, 도 13의 예에서는, 학습 완료 모델의 예로서, 귀 파라미터 추정 모델과 HRTF 학습 모델의 조합을 나타내고 있지만, 학습 완료 모델의 조합은 이 예에 한정되지는 않는다. 또한, 학습 완료 모델은, 귀 파라미터 추정 모델과 HRTF 학습 모델이 개별로 조합된 것이어도 되고, 귀 파라미터 추정 모델과 HRTF 학습 모델에 대응하는 처리를 행하는, 하나의 모델로서 구성되어도 된다.
- [0151] [1-4. 제1 실시 형태에 따른 유저 단말기의 구성]
- [0152] 도 13에서 나타낸 바와 같이, 제1 실시 형태에 있어서는, 유저의 옆 얼굴의 촬영 및 귀 화상의 생성 등이 유저 단말기(10)에 의해 행해진다. 여기서, 제1 실시 형태에 따른 유저 단말기(10)의 구성에 대하여 설명한다. 도 14는, 본 개시의 제1 실시 형태에 따른 유저 단말기(10)의 구성예를 나타내는 도면이다. 도 14에 도시한 바와 같이, 유저 단말기(10)는, 통신부(11)와, 입력부(12)와, 표시부(13)와, 검지부(14)와, 기억부(15)와, 제어부(16)를 갖는다.
- [0153] 통신부(11)는, 예를 들어 NIC 등에 의해 실현된다. 이러한 통신부(11)는, 네트워크 N과 유선 또는 무선으로 접속되고, 네트워크 N을 통해 정보 처리 장치(100) 등과의 사이에서 정보의 송수신을 행한다.
- [0154] 입력부(12)는, 유저로부터 각종 조작을 접수하는 입력 장치이다. 예를 들어, 입력부(12)는, 유저 단말기(10)에 구비된 조작 키 등에 의해 실현된다. 표시부(13)는, 각종 정보를 표시하기 위한 표시 장치이다. 예를 들어, 표시부(13)는, 액정 디스플레이 등에 의해 실현된다. 또한, 유저 단말기(10)에 터치 패널이 채용 되는 경우에는, 입력부(12)의 일부와 표시부(13)는 일체화된다.
- [0155] 검지부(14)는, 각종 센서의 총칭이며, 유저 단말기(10)에 관한 각종 정보를 검지한다. 구체적으로는, 검지부(14)는, 유저 단말기(10)에 대한 유저의 조작이나, 유저 단말기(10)의 소재하는 위치 정보나, 유저 단말기(10)와 접속되어 있는 기기에 관한 정보나, 유저 단말기(10)에서의 환경 등을 검지한다.
- [0156] 또한, 검지부(14)는, 센서의 일례로서, 촬영을 행하기 위한 렌즈 및 이미지 센서를 갖는다. 즉, 검지부(14)는, 예를 들어 유저가 촬영 기능을 동작시키는 애플리케이션을 기동시킨 경우, 카메라로서의 기능을 발휘한다.
- [0157] 기억부(15)는, 각종 정보를 기억한다. 기억부(15)는, 예를 들어 RAM, 플래시 메모리 등의 반도체 메모리 소자 또는, 하드 디스크, 광 디스크 등의 기억 장치에 의해 실현된다. 기억부(15)는, 예를 들어 유저에 의해 촬영된 화상 등을 기억한다.
- [0158] 제어부(16)는, 컨트롤러이며, 예를 들어 CPU나 MPU 등에 의해, 유저 단말기(10) 내부의 기억 장치에 기억되어 있는 각종 프로그램이 RAM을 작업 영역으로 하여 실행됨으로써 실현된다. 또한, 제어부(16)는, 컨트롤러이며, 예를 들어 ASIC나 FPGA 등의 집적 회로에 의해 실현된다.
- [0159] 도 14에 도시한 바와 같이, 제어부(16)는, 취득부(161)와, 전처리부(162)와, 송신부(164)와, 수신부(165)를 갖고, 이하에 설명하는 정보 처리의 기능이나 작용을 실현 또는 실행한다. 또한, 전처리부(162)는, 자세 검출부(163A)와, 귀 검출부(163B)를 갖는다. 또한, 제어부(16)의 내부 구성은, 도 14에 도시한 구성에 한정되지는 않고, 후술하는 정보 처리를 행하는 구성이라면 다른 구성이어도 된다.
- [0160] 취득부(161)는, 각종 정보를 취득한다. 예를 들어, 취득부(161)는, 검지부(14)에 의해 촬영된 화상을 취득한다.
- [0161] 자세 검출부(163A)는, 취득부(161)에 의해 취득된 화상을 읽어들이고 함께, 당해 화상에 포함되는 유저의 자세를 검출한다.
- [0162] 귀 검출부(163B)는, 자세 검출부(163A)에 의해 검출된 유저의 자세에 기초하여, 화상에 포함되는 유저의 귀가

포함되는 범위(귀의 영상)를 검출한다. 구체적으로는, 귀 검출부(163B)는, 유저의 머리부 전체의 영상을 포함하는 화상으로부터, 유저의 귀의 영상을 특정하고, 특정한 범위를 귀 화상으로서 검출한다.

[0163] 예를 들어, 귀 검출부(163B)는, 전체 화상에 포함되는 유저의 머리부의 특징점과 유저의 자세의 관계성에 기초하여, 귀의 영상이 포함되는 범위를 특정한다.

[0164] 또한, 자세 검출부(163A) 또는 귀 검출부(163B)는, 전체 화상에 포함되는 유저의 머리부의 특징점과 유저의 자세의 관계성에 기초하여 귀의 영상이 포함되는 범위를 특정할 수 없는 경우, 전체 화상과는 다른 화상이며, 유저의 머리부 전체의 영상을 포함하는 화상의 취득을 새롭게 유저에게 요구해도 된다. 구체적으로는, 자세 검출부(163A) 또는 귀 검출부(163B)는, 유저가 촬영한 옆 얼굴의 화상으로는, 본 개시에 따른 정보 처리를 적절하게 행하지 못할 가능성이 있음을 나타내는 메시지를 표시부(13)에 표시하여, 유저에게 촬영의 재시도를 재촉한다. 또한, 자세 검출부(163A) 또는 귀 검출부(163B)는, 귀의 영상이 포함되는 범위를 특정할 수 없는 경우에 한하지 않고, 예를 들어 귀 파라미터 추정 모델의 학습 시에 사용된 카메라 각도와 유저의 자세가 일정한 임계값을 초과해서 이격되어 있는 경우 등에도, 유저에게 촬영의 재시도를 재촉해도 된다. 또한, 자세 검출부(163A) 또는 귀 검출부(163B)는, 전처리로서 유저의 귀 화상을 검출하는 것이 아니라, 화상 내의 유저의 자세나 위치를 보정하는 보정 정보를 생성해도 된다. 보정 정보는, 예를 들어 유저의 머리부의 특징점의 기울기나 회전에 따라서, 귀의 영상이 포함되는 범위를 회전시킬 양 등을 지시하는 정보이다. 이러한 정보는, 후술하는 바와 같이, 유저의 자세나, 유저의 옆 얼굴과 검출된 귀의 위치의 위치 관계 등에 기초하여 생성된다. 이 경우, 자세 검출부(163A) 또는 귀 검출부(163B)는, 보정 정보에 기초하여 전체 화상의 회전을 보정함으로써 유저의 귀의 영상을 특정하고, 특정한 범위를 귀 화상으로서 검출해도 된다. 또한, 자세 검출부(163A) 또는 귀 검출부(163B)는, 생성된 보정 정보와 함께 전체 화상을 정보 처리 장치(100)에 송신해도 된다. 이 경우, 정보 처리 장치(100)는, 전체 화상과 함께 송신된 보정 정보에 기초하여 전체 화상의 회전을 보정함으로써 유저의 귀의 영상을 특정하고, 특정한 범위를 귀 화상으로서 검출하는 전처리를 자장치에서 행한다.

[0165] 여기서, 도 15에, 전처리부(162)(자세 검출부(163A) 및 귀 검출부(163B))가 실행하는 전처리의 흐름에 대하여 설명한다. 도 15는, 본 개시에 따른 검출 처리의 흐름을 나타내는 도면이다.

[0166] 도 15에 도시한 바와 같이, 취득부(161)는, 유저에 의해 유저의 옆 얼굴이 촬영된 경우, 전체 화상(50)을 취득한다(스텝 S141).

[0167] 자세 검출부(163A)는, 취득된 전체 화상(50)에 있어서, 유저의 옆 얼굴을 검출한다(스텝 S142). 예를 들어, 자세 검출부(163A)는, 인물의 얼굴 검출 처리 등의 기지의 기술을 이용하여, 전체 화상(50) 중, 유저의 옆 얼굴 영상이 포함되는 범위를 특정한다.

[0168] 여기서, 자세 검출부(163A)는, 화상(51)에 나타난 바와 같이, 유저의 옆 얼굴에 포함되는 특징점을 검출한다. 예를 들어, 자세 검출부(163A)는, 옆 얼굴 중, 수평 방향으로 돌출되어 있는 개소(구체적으로는, 유저의 코의 정점 등)나, 머리부의 정점이나, 입의 위치나, 턱의 위치 등의 특징점을 검출한다. 또한, 자세 검출부(163A)는, 머리카락과 피부의 경계의 정보 등으로부터, 유저의 귀나 구레나룻의 위치 등을 검출한다. 또한, 자세 검출부(163A)는, 옆 얼굴의 영상의 색 정보로부터, 유저의 눈의 위치 등을 검출한다.

[0169] 그리고, 자세 검출부(163A)는, 검출된 특징점에 기초하여, 유저의 자세를 검출한다(스텝 S143). 예를 들어, 자세 검출부(163A)는, 화상(54)에 나타난 바와 같은 특징점의 3차원적인 배치로부터 유저의 머리부의 자세를 검출한다.

[0170] 이와 같은 자세의 검출 처리는, 유저가 송신하는 귀 화상에서의 자세가, 학습 시에 사용한 3D 모델의 자세와 크게 어긋남이 생기지 않도록 하기 위한 처리이다. 즉, 3D 모델과 크게 자세가 다른 화상이 유저 단말기(10)로부터 송신된 경우, 학습 데이터와 송신된 귀 화상의 어긋남이 나지 않도록, 정보 처리 장치(100)가 귀의 화상 인식을 적절하게 행하지 못할 가능성이 있기 때문이다.

[0171] 이 때문에, 자세 검출부(163A)는, 학습에 사용된 머리부 3D 모델(55)에서의 렌더링 시의 각도의 평균값과, 화상(54)에서 얻어진 각도가 소정의 임계값 이내인지를 판정하고, 유저가 적절하게 촬영을 행할 수 있었는지의 합격 여부를 판정한다(스텝 S144). 예를 들어, 귀 파라미터 추정 모델의 학습 시에는, 머리부 3D 모델(55)의 렌더링 시의 카메라의 방향과, 머리부 정점과 귀의 소정 위치(예를 들어, 외이도의 입구 등)를 연결하는 선분이 이루는 각도 ϕ 가 소정의 수치 이내였다고 상정된다. 마찬가지로, 귀 파라미터 추정 모델의 학습 시에는, 카메라의 방향과, 코의 정점과 귀의 소정 위치를 연결하는 선분이 이루는 각도 θ 가 소정의 수치 이내였다고 상정된다. 이것은, 화상 인식 정밀도를 높이기 위해서, 학습에 사용하는 귀 화상이, 인간의 옆 얼굴을 나타내는 화상으로부터

터 크게 벗어난 것으로 되지 않기 때문이다. 즉, 자세 검출부(163A)는, 학습 시의 화상과 마찬가지로, 유저로부터 송신된 화상이, 인간의 옆 얼굴을 나타내는 화상이라고 인식 가능한 정도의 각도를 유지하는 화상인지 여부를 판정한다.

[0172] 자세 검출부(163A)는, 유저가 적절하게 촬영을 행할 수 없다고 판정한 경우 (예를 들어, 유저의 얼굴에 있어서 소정 임계값을 초과해 코끝이 아래를 향하고 있는 경우 등)에는, 촬영의 재시도를 명하는 메시지를 표시하는 등의 처리를 행하고, 새롭게 촬영된 화상을 취득한다(스텝 S145).

[0173] 한편, 유저가 적절하게 촬영을 행할 수 있다고 판정한 경우에는(스텝 S146), 귀 검출부(163B)는, 화상(56)으로부터 귀의 영상이 포함되는 범위(57)를 특정하고, 범위(57)를 잘라낸다(스텝 S147). 이에 의해, 귀 검출부(163B)는, 귀 화상(58)을 취득한다.

[0174] 도 15에 도시한 검출 처리를 행함으로써, 정보 처리 장치(100)는, 촬영 상태가 나쁘기 때문에 귀가 기울어 있는 것인지, 혹은 실제로 그 유저의 귀의 각도가 기울어 있는 것인지를 판별하여, HRTF의 산출을 행할 수 있다.

[0175] 또한, 유저 단말기(10)는, 상기한 바와 같이 옆 얼굴의 전체 화상으로부터 귀 화상을 잘라냄으로써, 유저의 얼굴이 포함되는 전체 화상이 아니라, 귀 화상만을 송신시켜 처리를 행하게 할 수 있다. 이에 의해, 유저 단말기(10)는, 개인 정보의 유출을 방지할 수 있어, 정보 처리의 안전성을 높일 수 있다. 또한, 유저 단말기(10)는, 상기 검출 방법에 한하지 않고, 기계 학습에 의한 화상 인식 기술 등을 이용하여, 화상에 포함되는 유저의 귀를 검출함으로써, 옆 얼굴의 전체 화상으로부터 귀 화상을 잘라내는 처리를 행해도 된다.

[0176] 도 14로 되돌아가서, 설명을 계속한다. 송신부(164)는, 귀 검출부(163B)에 의해 검출된 범위에 기초하여 생성된 귀 화상을 정보 처리 장치(100)에 송신한다.

[0177] 수신부(165)는, 정보 처리 장치(100)로부터 제공되는 개인화 HRTF를 수신한다. 예를 들어, 수신부(165)는, 수신한 개인화 HRTF를 음성 재생 애플리케이션 등에서 음악이나 음성에 인터럽트함으로써, 유저 개인에 최적화된 3D 음향 등을 실현할 수 있다.

[0178] (2. 제2 실시 형태)

[0179] 다음으로, 제2 실시 형태에 대하여 설명한다. 상기 제1 실시 형태에서는, 유저 단말기(10)가, 유저가 촬영한 화상으로부터 귀의 영상만을 잘라내어 귀 화상을 생성하는 예를 나타내었다. 제2 실시 형태에 따른 정보 처리 장치(100A)는, 유저 단말기(10) 대신에, 자장치에서 귀의 영상만을 잘라내는 처리를 행한다.

[0180] 도 16을 이용하여, 제2 실시 형태에 따른 정보 처리 장치(100A)의 구성에 대하여 설명한다. 도 16은, 본 개시의 제2 실시 형태에 따른 정보 처리 장치(100A)의 구성예를 나타내는 도면이다. 도 16에 도시한 바와 같이, 정보 처리 장치(100A)는, 제1 실시 형태와 비교하여, 전처리부(144)(자세 검출부(145A)와, 귀 검출부(145B))를 더 구비한다.

[0181] 자세 검출부(145A)는, 제1 실시 형태에 따른 자세 검출부(163A)와 마찬가지로의 처리를 행한다. 또한, 귀 검출부(145B)는, 제1 실시 형태에 따른 귀 검출부(163B)와 마찬가지로의 처리를 행한다. 즉, 제2 실시 형태에 따른 정보 처리 장치(100A)는, 제1 실시 형태에 따른 유저 단말기(10)가 실행한 전처리를 자장치에서 실행한다.

[0182] 제2 실시 형태에서는, 취득부(141)는, 유저가 옆 얼굴을 촬영한 전체 화상을 유저 단말기(10)로부터 취득한다. 그리고, 자세 검출부(145A) 및 귀 검출부(145B)는, 도 15를 이용하여 설명한 처리와 마찬가지로의 처리를 행하여, 전체 화상으로부터 귀 화상을 생성한다. 산출부(142)는, 자세 검출부(145A) 및 귀 검출부(145B)에 의해 생성된 귀 화상으로부터, 개인화 HRTF를 산출한다.

[0183] 이와 같이, 제2 실시 형태에 따른 정보 처리 장치(100A)에 의하면, 유저는, 화상을 촬영하여 송신하는 것만으로, 개인화 HRTF의 제공을 받을 수 있다. 또한, 제2 실시 형태의 구성에 의하면, 유저 단말기(10)에 있어서 전처리를 실행할 것을 요하지 않기 때문에, 예를 들어 유저 단말기(10)의 처리 부하를 경감시킬 수 있다. 또한, 일반적으로, 유저 단말기(10)보다도 서버 장치(정보 처리 장치(100)) 쪽이 처리는 고속이라고 상정되기 때문에, 제2 실시 형태의 구성에 의하면, 본 개시에 따른 정보 처리의 전체의 속도를 향상시킬 수 있다. 또한, 자세 검출부(145A) 및 귀 검출부(145B)는, 전체 화상과 함께 보정 정보가 송신된 경우, 전체 화상에 포함되는 보정 정보에 기초하여 전체 화상의 회전을 보정함으로써 유저의 귀의 영상을 특정하고, 특정한 범위를 귀 화상으로서 검출해도 된다.

[0184] (3. 기타 실시 형태)

- [0185] 상술한 각 실시 형태에 따른 처리는, 상기 각 실시 형태 이외에도 다양한 다른 형태로 실시되어도 된다.
- [0186] 또한, 상기 각 실시 형태에 있어서 설명한 각 처리 중, 자동적으로 행해지는 것으로서 설명한 처리의 전부 또는 일부를 수동적으로 행할 수도 있으며, 혹은 수동적으로 행해지는 것으로서 설명한 처리의 전부 또는 일부를 공지의 방법으로 자동적으로 행할 수도 있다. 이밖에, 상기 문서 중이나 도면 중에서 나타난 처리 수순, 구체적인 명칭, 각종 데이터나 파라미터를 포함하는 정보에 대해서는, 특별히 기재하는 경우를 제외하고 임의로 변경할 수 있다. 예를 들어, 각 도면에 나타난 각종 정보는, 도시한 정보에 한정되지는 않는다.
- [0187] 또한, 도시한 각 장치의 각 구성 요소는 기능 개념적인 것이며, 반드시 물리적으로 도시한 것처럼 구성되어 있을 것을 요하지는 않는다. 즉, 각 장치의 분산·통합의 구체적인 형태는 도시한 것에 한하지는 않고, 그 전부 또는 일부를, 각종 부하나 사용 상황 등에 따라서 임의의 단위에서 기능적 또는 물리적으로 분산·통합하여 구성할 수 있다.
- [0188] 또한, 상술해 온 각 실시 형태 및 변형예는, 처리 내용을 모순시키지 않는 범위에서 적절하게 조합하는 것이 가능하다.
- [0189] 또한, 본 명세서에 기재된 효과는 어디까지나 예시이지 한정되는 것은 아니며, 다른 효과가 있어도 된다.
- [0190] (4. 하드웨어 구성)
- [0191] 상술해 온 각 실시 형태에 따른 정보 처리 장치(100)나 유저 단말기(10) 등의 정보 기기는, 예를 들어 도 17에 도시한 바와 같은 구성의 컴퓨터(1000)에 의해 실현된다. 이하, 제1 실시 형태에 따른 정보 처리 장치(100)를 예로 들어 설명한다. 도 17은, 정보 처리 장치(100)의 기능을 실현하는 컴퓨터(1000)의 일례를 나타내는 하드웨어 구성도이다. 컴퓨터(1000)는, CPU(1100), RAM(1200), ROM(Read Only Memory)(1300), HDD(Hard Disk Drive)(1400), 통신 인터페이스(1500), 및 입출력 인터페이스(1600)를 갖는다. 컴퓨터(1000)의 각 부는, 버스(1050)에 의해 접속된다.
- [0192] CPU(1100)는, ROM(1300) 또는 HDD(1400)에 저장된 프로그램에 기초하여 동작하고, 각 부의 제어를 행한다. 예를 들어, CPU(1100)는, ROM(1300) 또는 HDD(1400)에 저장된 프로그램을 RAM(1200)에 전개하여, 각종 프로그램에 대응한 처리를 실행한다.
- [0193] ROM(1300)은, 컴퓨터(1000)의 기동 시에 CPU(1100)에 의해 실행되는 BIOS(Basic Input Output System) 등의 부트 프로그램이나, 컴퓨터(1000)의 하드웨어에 의존하는 프로그램 등을 저장한다.
- [0194] HDD(1400)는, CPU(1100)에 의해 실행되는 프로그램, 및 이러한 프로그램에 의해 사용되는 데이터 등을 비밀시적으로 기록하는, 컴퓨터가 판독 가능한 기록 매체다. 구체적으로는, HDD(1400)는, 프로그램 데이터(1450)의 일례인 본 개시에 따른 정보 처리 프로그램을 기록하는 기록 매체이다.
- [0195] 통신 인터페이스(1500)는, 컴퓨터(1000)가 외부 네트워크(1550)(예를 들어 인터넷)와 접속하기 위한 인터페이스이다. 예를 들어, CPU(1100)는, 통신 인터페이스(1500)를 통해 다른 기기로부터 데이터를 수신하거나, CPU(1100)가 생성된 데이터를 다른 기기에 송신하거나 한다.
- [0196] 입출력 인터페이스(1600)는, 입출력 디바이스(1650)와 컴퓨터(1000)를 접속하기 위한 인터페이스이다. 예를 들어, CPU(1100)는, 입출력 인터페이스(1600)를 통해 키보드나 마우스 등의 입력 디바이스로부터 데이터를 수신한다. 또한, CPU(1100)는, 입출력 인터페이스(1600)를 통해 디스플레이나 스피커나 프린터 등의 출력 디바이스에 데이터를 송신한다. 또한, 입출력 인터페이스(1600)는, 소정의 기록 매체(미디어)에 기록된 프로그램 등을 판독하는 미디어 인터페이스로서 기능해도 된다. 미디어는, 예를 들어 DVD(Digital Versatile Disc), PD(Phase change rewritable Disk) 등의 광학 기록 매체, MO(Magneto-Optical disk) 등의 광자기 기록 매체, 테이프 매체, 자기 기록 매체 또는 반도체 메모리 등이다.
- [0197] 예를 들어, 컴퓨터(1000)가 제1 실시 형태에 따른 정보 처리 장치(100)로서 기능하는 경우, 컴퓨터(1000)의 CPU(1100)는, RAM(1200)상에 로드된 정보 처리 프로그램을 실행함으로써, 제어부(130) 등의 기능을 실현한다. 또한, HDD(1400)에는, 본 개시에 따른 정보 처리 프로그램이나, 기억부(120) 내의 데이터가 저장된다. 또한, CPU(1100)는, 프로그램 데이터(1450)를 HDD(1400)로부터 판독하여 실행하지만, 다른 예로서, 외부 네트워크(1550)를 통해 다른 장치로부터 이들 프로그램을 취득해도 된다.
- [0198] 또한, 본 기술은 이하와 같은 구성도 취할 수 있다.

- [0199] (1)
- [0200] 유저의 귀의 영상을 포함하는 제1 화상을 취득하는 취득부와,
- [0201] 귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 머리 전달 함수를 출력하도록 학습된 학습 완료 모델을 이용하여, 상기 취득부에 의해 취득된 제1 화상에 기초하여, 상기 유저에 대응하는 머리 전달 함수를 산출하는 산출부
- [0202] 를 구비하는 정보 처리 장치.
- [0203] (2)
- [0204] 상기 취득부는,
- [0205] 상기 제1 화상에 포함되는 귀의 특징을 나타내는 변수인 귀 파라미터를 취득하고,
- [0206] 상기 산출부는,
- [0207] 상기 귀 파라미터를 상기 학습 완료 모델에 입력함으로써, 상기 유저에 대응하는 머리 전달 함수를 산출하는
- [0208] 상기 (1)에 기재된 정보 처리 장치.
- [0209] (3)
- [0210] 상기 취득부는,
- [0211] 귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 귀 파라미터를 출력하도록 학습된 귀 파라미터 추정 모델을 이용하여, 상기 제1 화상에 포함되는 귀의 귀 파라미터를 취득하는
- [0212] 상기 (2)에 기재된 정보 처리 장치.
- [0213] (4)
- [0214] 귀의 영상을 포함하는 화상과 당해 귀의 귀 파라미터의 관계성을 학습함으로써, 상기 귀 파라미터 추정 모델을 생성하는 제1 학습부
- [0215] 를 더 구비하는 상기 (3)에 기재된 정보 처리 장치.
- [0216] (5)
- [0217] 상기 제1 학습부는,
- [0218] 상기 귀 파라미터에 기초하여 생성된 귀의 3차원 데이터를 렌더링함으로써 얻어지는 귀 화상과 당해 귀 파라미터의 관계성을 학습함으로써, 상기 귀 파라미터 추정 모델을 생성하는
- [0219] 상기 (4)에 기재된 정보 처리 장치.
- [0220] (6)
- [0221] 상기 제1 학습부는,
- [0222] 상기 귀 혹은 머리부의 3차원 데이터의 텍스처, 렌더링에서의 카메라 각도 또는 렌더링에서의 명도를 변경한 복수의 귀 화상과 당해 복수의 귀 화상에 공통되는 귀 파라미터의 관계성을 학습함으로써, 상기 귀 파라미터 추정 모델을 생성하는
- [0223] 상기 (5)에 기재된 정보 처리 장치.
- [0224] (7)
- [0225] 귀의 영상을 포함하는 화상과 당해 귀에 대응하는 머리 전달 함수의 관계성을 학습함으로써, 상기 학습 완료 모델을 생성하는 제2 학습부
- [0226] 를 더 구비하는 상기 (1) 내지 (6) 중 어느 하나에 기재된 정보 처리 장치.
- [0227] (8)
- [0228] 상기 제2 학습부는,

- [0229] 상기 귀 파라미터에 기초하여 생성된 귀의 3차원 데이터와 머리부의 3차원 데이터를 합성한 3차원 데이터에 대하여 음향 시뮬레이션을 행하여, 당해 음향 시뮬레이션에 의해 얻어진 머리 전달 함수와 당해 귀 파라미터의 관계성을 학습함으로써, 상기 학습 완료 모델을 생성하는
- [0230] 상기 (7)에 기재된 정보 처리 장치.
- [0231] (9)
- [0232] 상기 제2 학습부는,
- [0233] 상기 음향 시뮬레이션에 의해 얻어진 머리 전달 함수의 정보량을 압축하고, 압축된 머리 전달 함수와 상기 귀 파라미터의 관계성을 학습함으로써, 상기 학습 완료 모델을 생성하는
- [0234] 상기 (8)에 기재된 정보 처리 장치.
- [0235] (10)
- [0236] 상기 제2 학습부는,
- [0237] 상기 귀 파라미터에 기초하여 생성된 귀의 3차원 데이터의 청취점을 설정하고, 설정된 청취점을 사용하여 상기 음향 시뮬레이션을 행하는
- [0238] 상기 (8) 또는 (9)에 기재된 정보 처리 장치.
- [0239] (11)
- [0240] 상기 유저의 머리부 전체의 영상을 포함하는 제2 화상으로부터 당해 유저의 귀의 영상을 특정하고, 특정한 범위를 상기 제1 화상으로서 검출하는 전처리부를 더 구비하고,
- [0241] 상기 취득부는,
- [0242] 상기 전처리부에 의해 검출된 상기 제1 화상을 취득하는
- [0243] 상기 (1) 내지 (10) 중 어느 하나에 기재된 정보 처리 장치.
- [0244] (12)
- [0245] 상기 전처리부는,
- [0246] 상기 제2 화상에 포함되는 유저의 머리부의 특징점과 당해 유저의 자세의 관계성에 기초하여, 상기 범위를 특정하는
- [0247] 상기 (11)에 기재된 정보 처리 장치.
- [0248] (13)
- [0249] 상기 전처리부는,
- [0250] 상기 제2 화상에 포함되는 유저의 머리부의 특징점과 당해 유저의 자세의 관계성에 기초하여 상기 범위를 특정할 수 없는 경우, 당해 제2 화상과는 다른 화상이며, 상기 유저의 머리부 전체의 영상을 포함하는 화상의 취득을 새롭게 요구하는
- [0251] 상기 (12)에 기재된 정보 처리 장치.
- [0252] (14)
- [0253] 상기 전처리부는,
- [0254] 상기 제2 화상에 포함되는 보정 정보에 기초하여 당해 제2 화상의 회전을 보정함으로써 상기 유저의 귀의 영상을 특정하고, 특정한 범위를 상기 제1 화상으로서 검출하는
- [0255] 상기 (11) 내지 (13) 중 어느 하나에 기재된 정보 처리 장치.
- [0256] (15)
- [0257] 컴퓨터가,

- [0258] 유저의 귀의 영상을 포함하는 제1 화상을 취득하고,
- [0259] 귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 머리 전달 함수를 출력하도록 학습된 학습 완료 모델을 이용하여, 취득된 상기 제1 화상에 기초하여, 상기 유저에 대응하는 머리 전달 함수를 산출하는
- [0260] 정보 처리 방법.
- [0261] (16)
- [0262] 컴퓨터를,
- [0263] 유저의 귀의 영상을 포함하는 제1 화상을 취득하는 취득부와,
- [0264] 귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 머리 전달 함수를 출력하도록 학습된 학습 완료 모델을 이용하여, 상기 취득부에 의해 취득된 제1 화상에 기초하여, 상기 유저에 대응하는 머리 전달 함수를 산출하는 산출부
- [0265] 로서 기능시키기 위한 정보 처리 프로그램.
- [0266] (17)
- [0267] 정보 처리 장치와 유저 단말기를 포함하는 정보 처리 시스템이며,
- [0268] 상기 유저 단말기는,
- [0269] 유저의 머리부 전체의 영상을 포함하는 제2 화상으로부터 당해 유저의 귀의 영상을 특정하고, 특정한 범위를 제1 화상으로서 검출하는 전처리부와,
- [0270] 상기 전처리부에 의해 검출된 제1 화상을 상기 정보 처리 장치에 송신하는 송신부
- [0271] 를 구비하고,
- [0272] 상기 처리 장치는,
- [0273] 상기 유저의 귀의 영상을 포함하는 상기 제1 화상을 취득하는 취득부와,
- [0274] 귀의 영상을 포함하는 화상이 입력된 경우에 당해 귀에 대응하는 머리 전달 함수를 출력하도록 학습된 학습 완료 모델을 이용하여, 상기 취득부에 의해 취득된 제1 화상에 기초하여, 상기 유저에 대응하는 머리 전달 함수를 산출하는 산출부
- [0275] 를 구비하는 정보 처리 시스템.

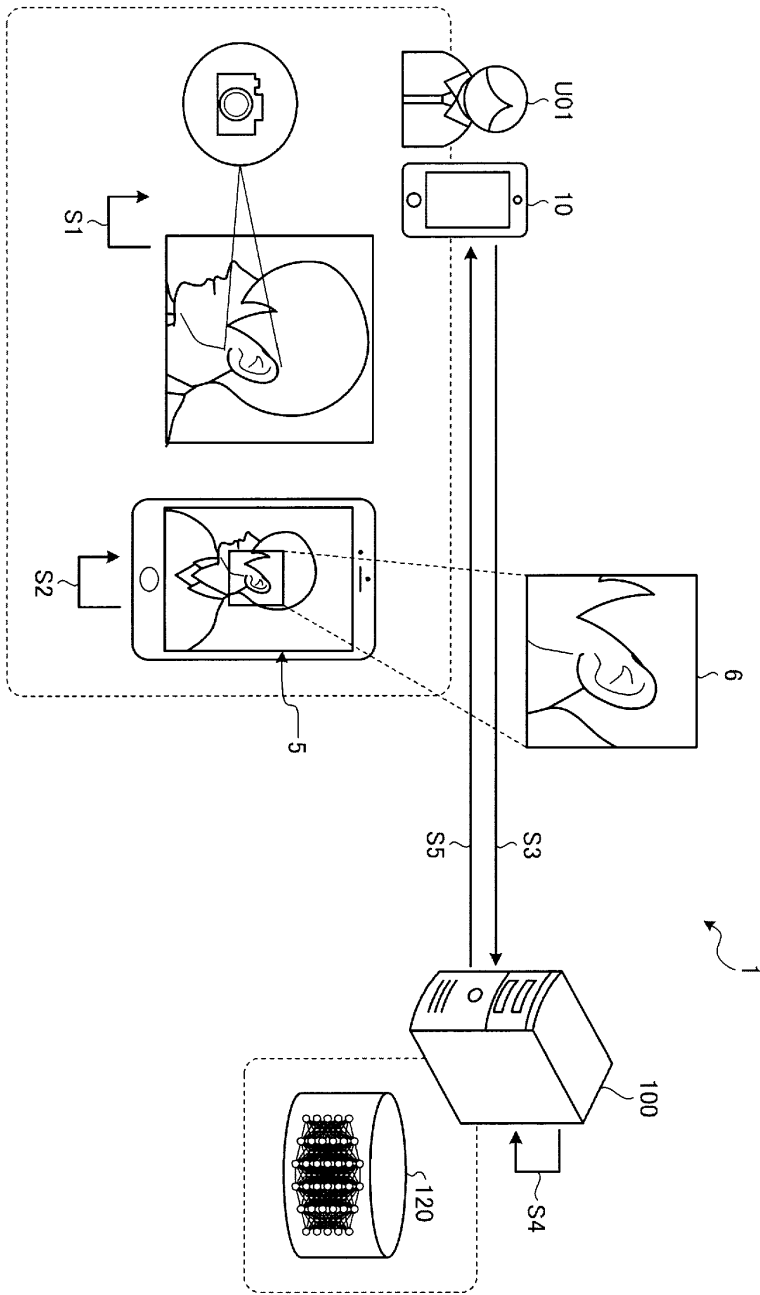
부호의 설명

- [0276] 1: 정보 처리 시스템
- 10: 유저 단말기
- 100: 정보 처리 장치
- 110: 통신부
- 120: 기억부
- 130: 제어부
- 131: 학습부
- 132: 접수부
- 133: 귀 모델 학습부
- 134: 화상 생성부
- 135: 귀 파라미터 학습부
- 136: HRTF 학습부

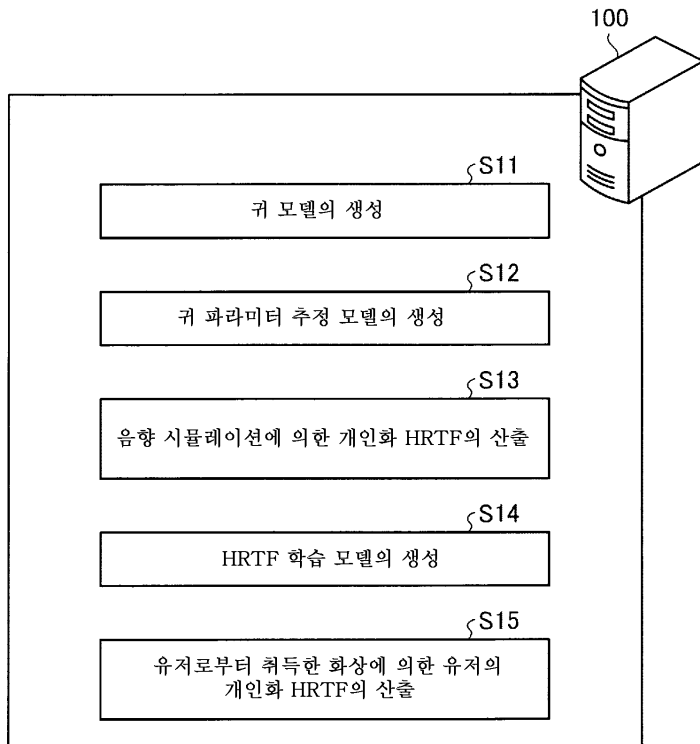
- 140: 추정부
- 141: 취득부
- 142: 산출부
- 143: 제공부
- 144: 전처리부
- 145A: 자세 검출부
- 145B: 귀 검출부
- 161: 취득부
- 162: 전처리부
- 163A: 자세 검출부
- 163B: 귀 검출부
- 164: 송신부
- 165: 수신부

도면

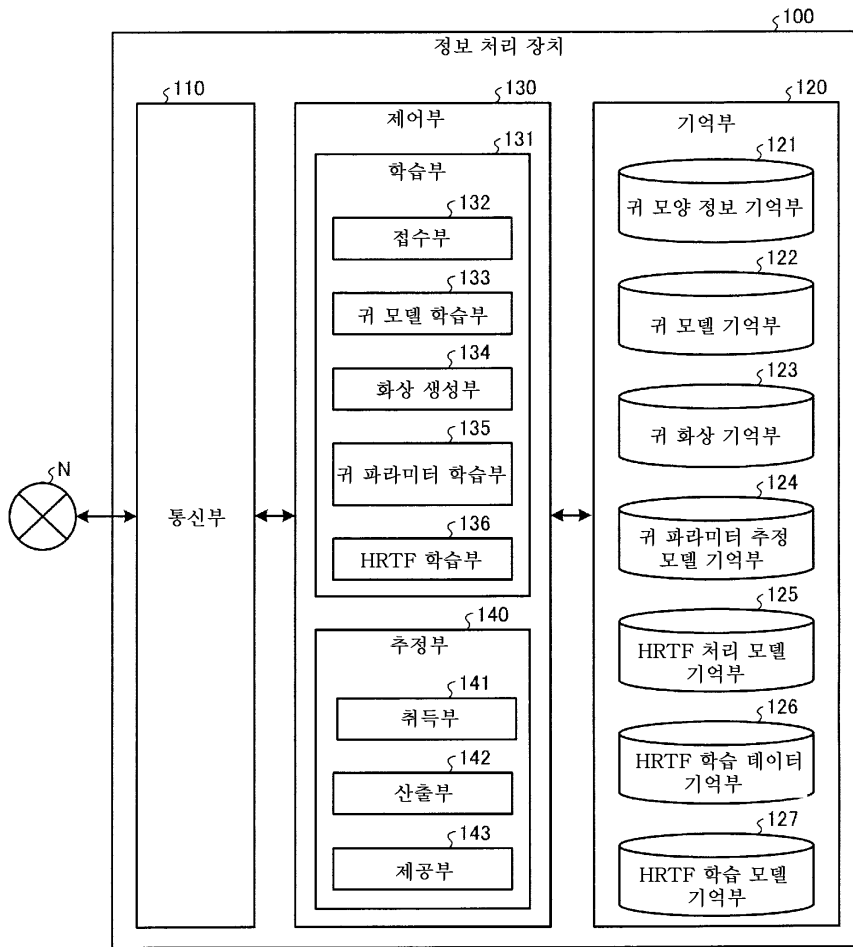
도면1



도면2



도면3

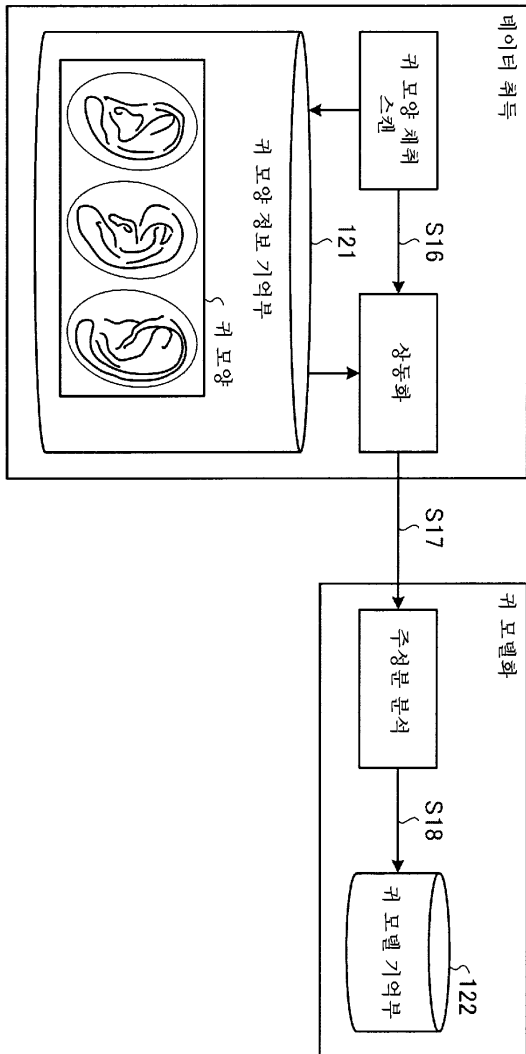


도면4

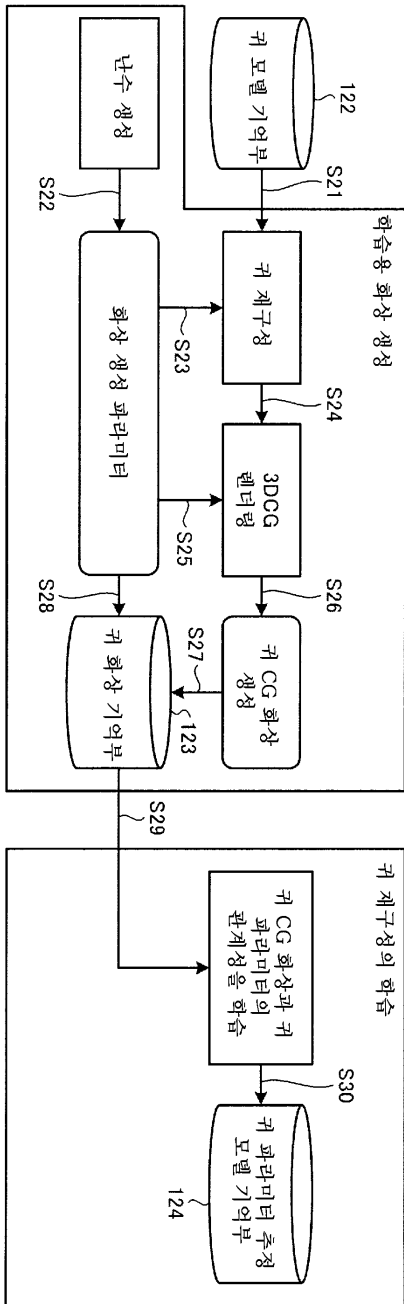
귀 파라미터	귀 3D 모델 데이터	머리부 3D 모델 데이터	귀 확장 ID	확장 생상 파라미터						...
				텍스처	카메라 각도	확장도	명도	
...
A01	B01	C01	D01	E01	F01	G01	H01	
			D02	E02	F02	G02	H02	...		
			D03	E03	F03	G03	H03	...		
...

123 ↶

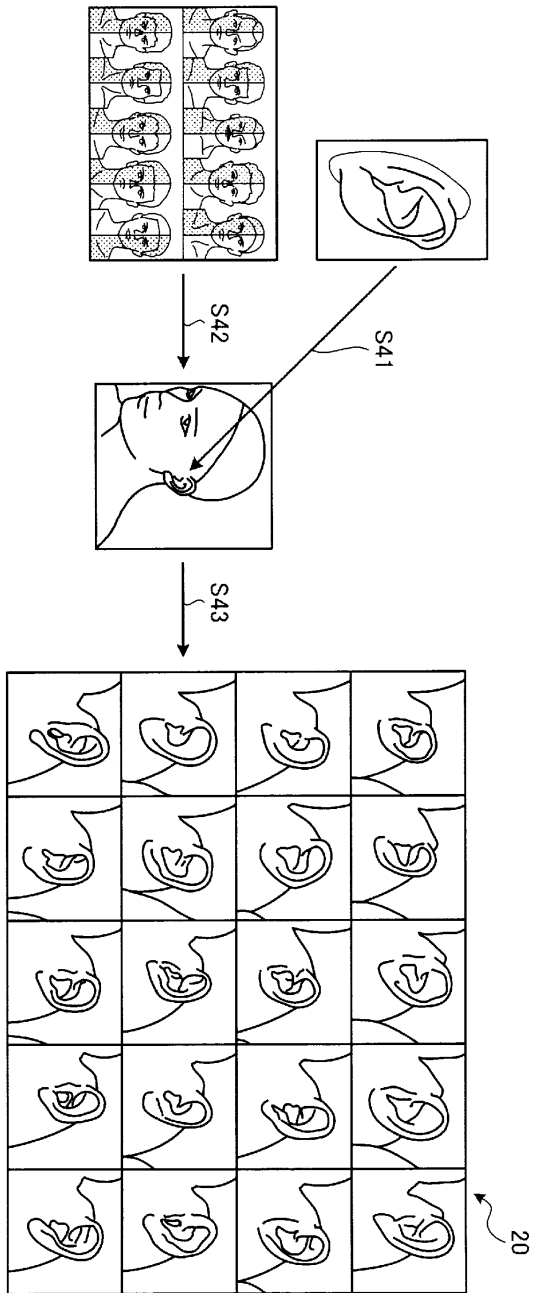
도면5



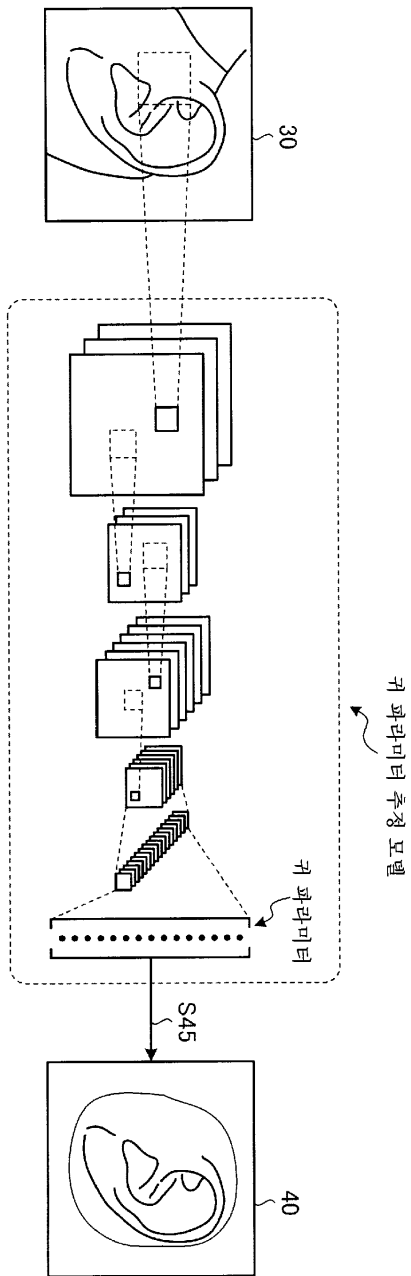
도면6



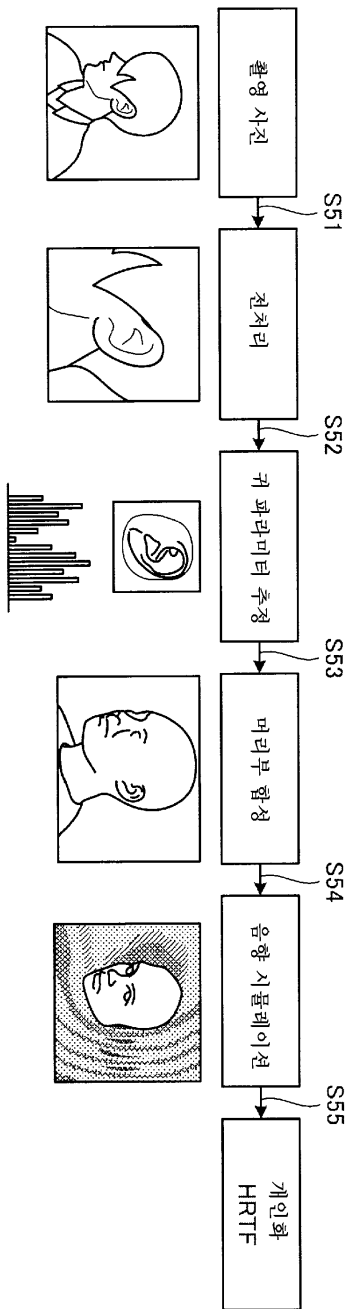
도면7



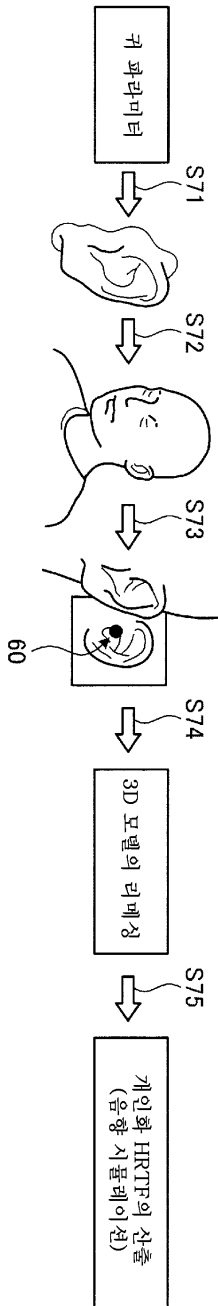
도면8



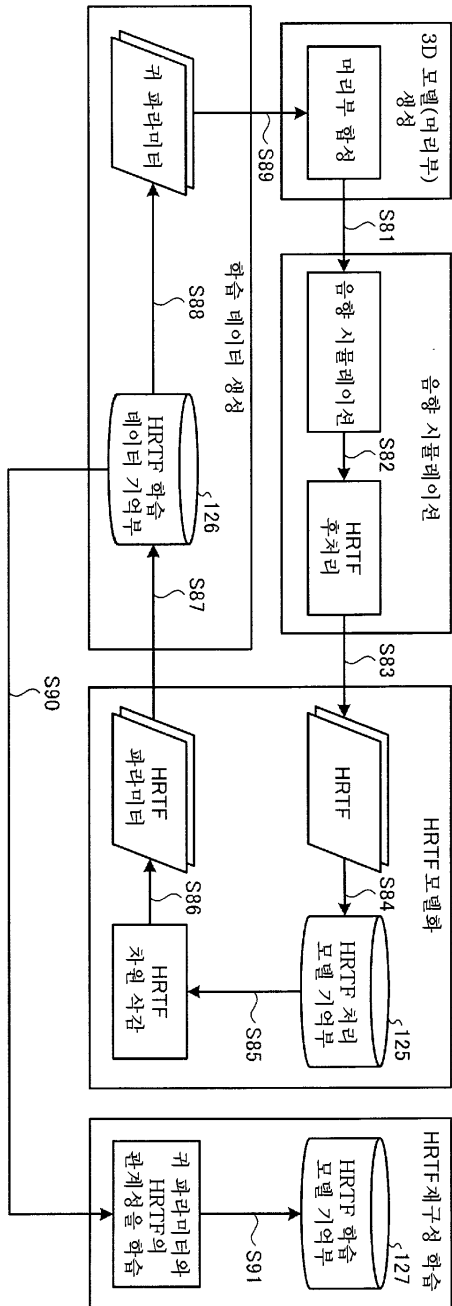
도면9



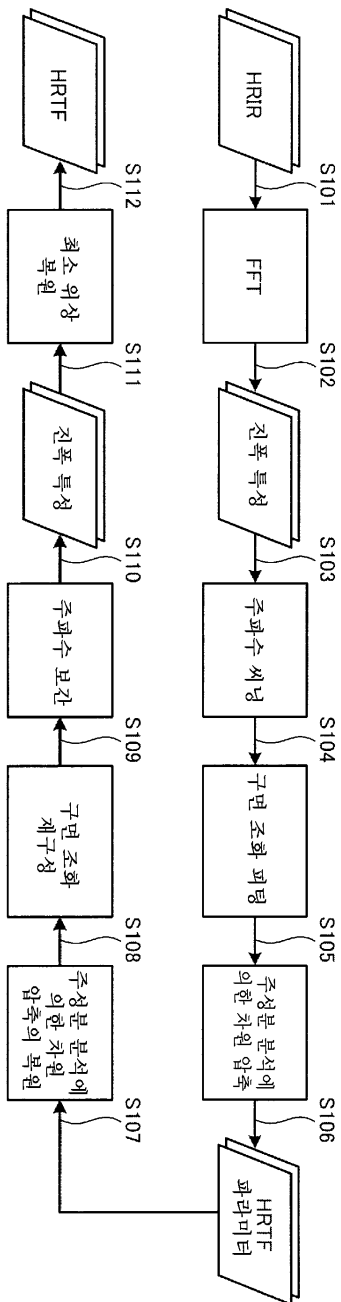
도면10



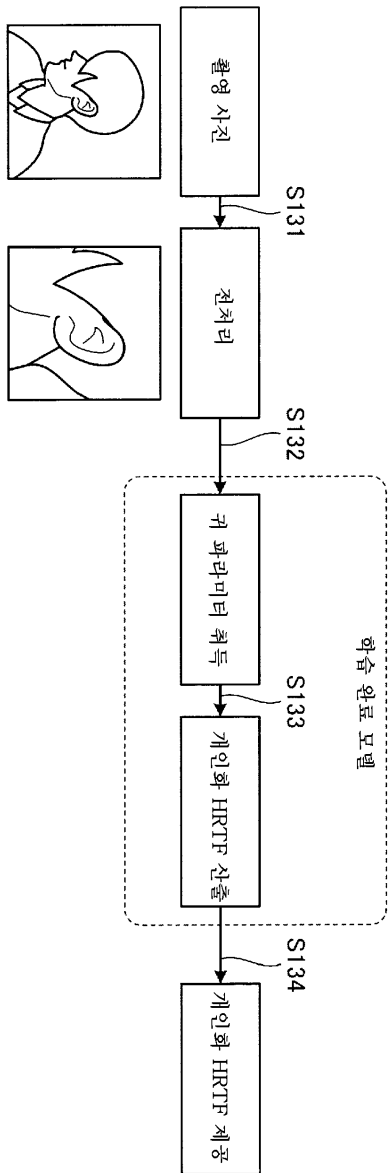
도면11



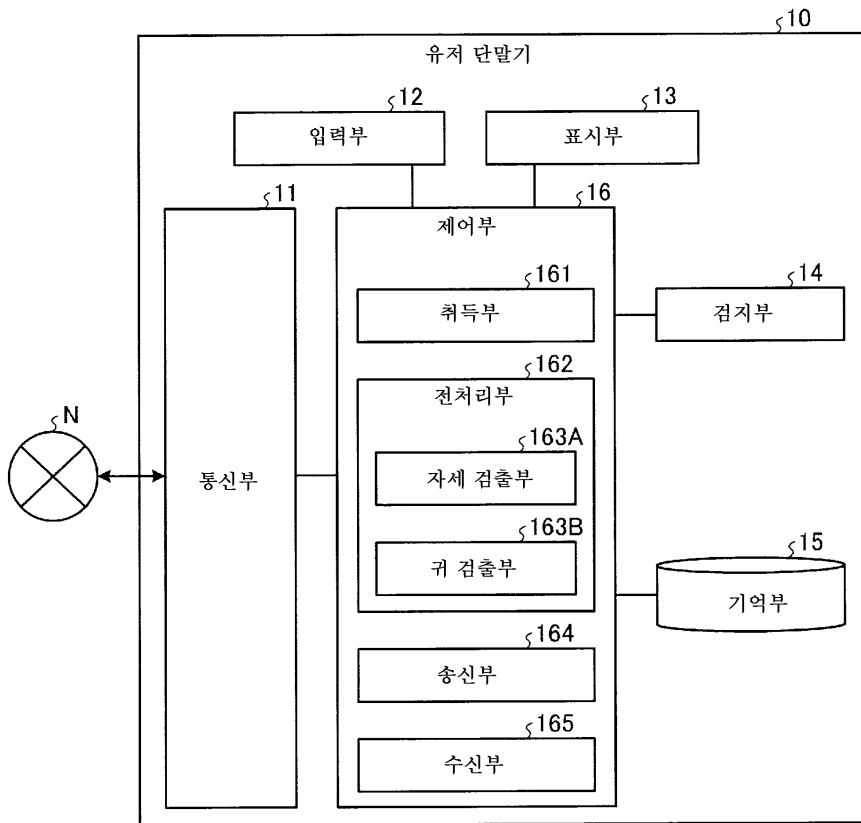
도면12



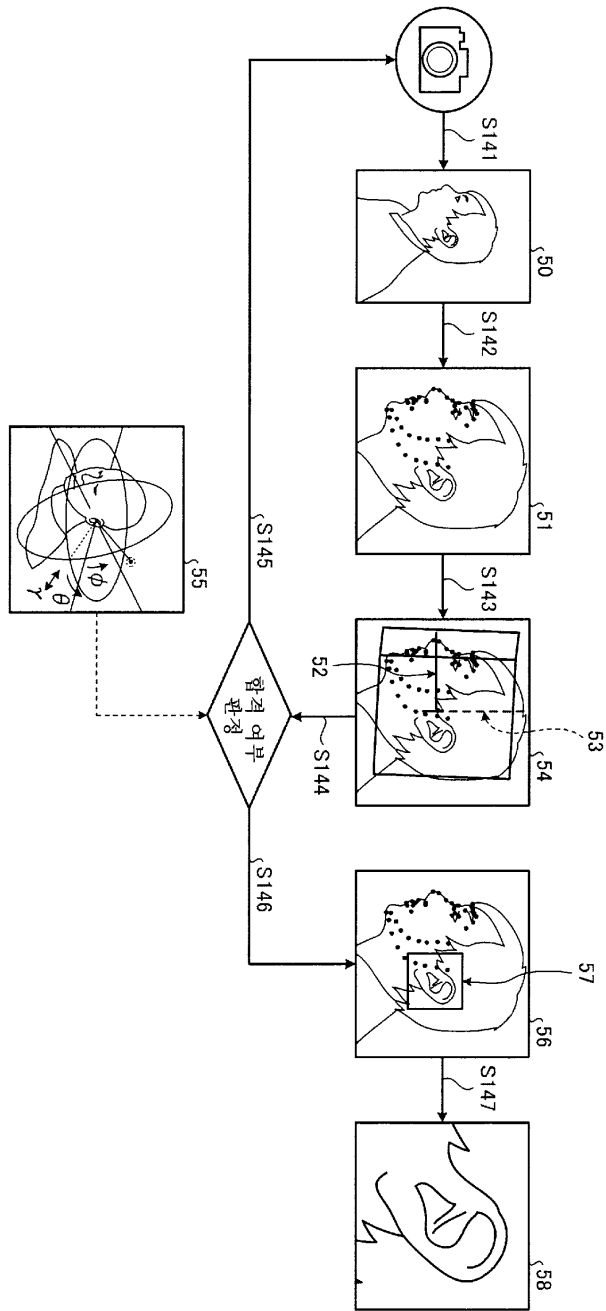
도면13



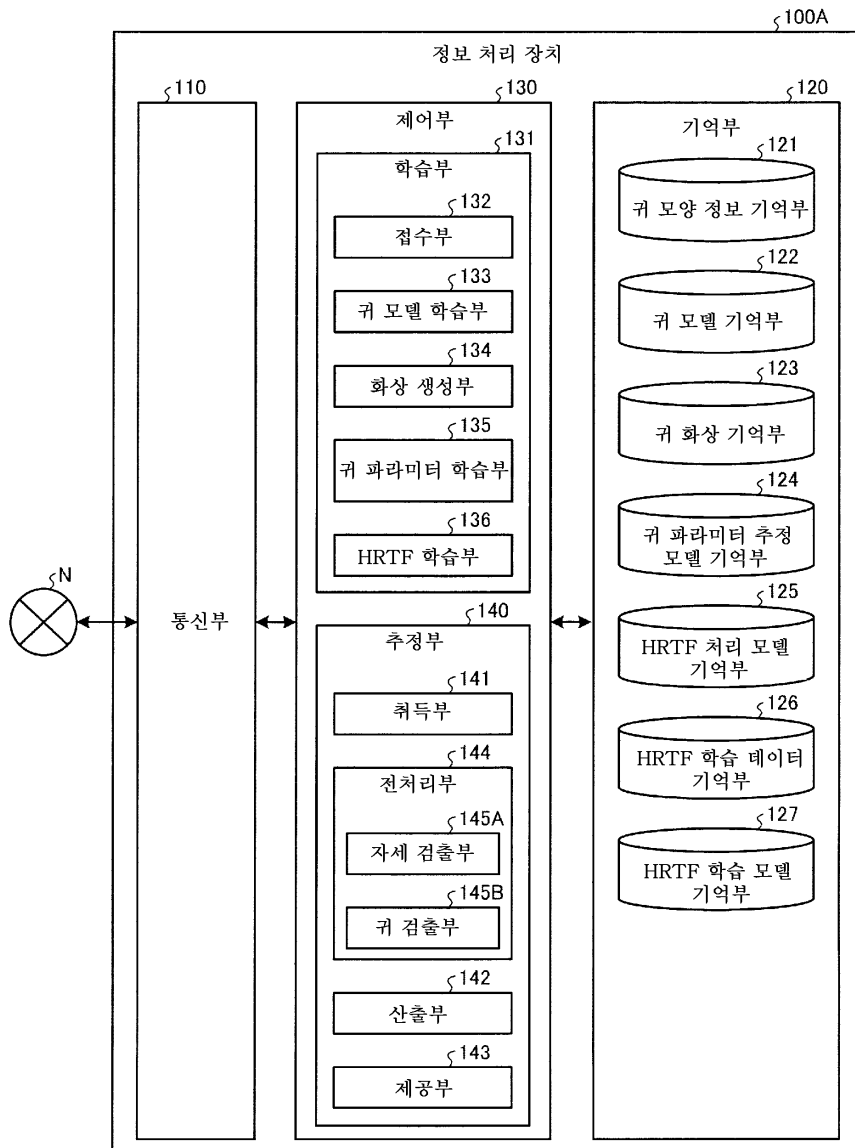
도면14



도면15



도면16



도면17

