



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2010년10월18일
 (11) 등록번호 10-0988708
 (24) 등록일자 2010년10월12일

- (51) Int. Cl.
G10L 15/22 (2006.01) *G10L 15/28* (2006.01)
G10L 15/20 (2006.01) *G06T 7/00* (2006.01)
- (21) 출원번호 10-2003-7014405
- (22) 출원일자(국제출원일자) 2003년03월05일
 심사청구일자 2008년03월03일
- (85) 번역문제출일자 2003년11월05일
- (65) 공개번호 10-2004-0094289
- (43) 공개일자 2004년11월09일
- (86) 국제출원번호 PCT/JP2003/002560
- (87) 국제공개번호 WO 2003/075261
 국제공개일자 2003년09월11일
- (30) 우선권주장
 JP-P-2002-00060425 2002년03월06일 일본(JP)
- (56) 선행기술조사문헌
 JP11175132 A*
 *는 심사관에 의하여 인용된 문헌

- (73) 특허권자
 소니 주식회사
 일본국 도쿄도 미나토쿠 코난 1-7-1
- (72) 발명자
 시모무라, 히데끼
 일본 141-0001 도쿄도 시나가와꾸 기따시나가와
 6쵸메 7-35 소니 가부시키 가이샤 내
 아오야마, 가즈미
 일본 141-0001 도쿄도 시나가와꾸 기따시나가와
 6쵸메 7-35 소니 가부시키 가이샤 내
 (뒷면에 계속)
- (74) 대리인
 이중희, 장수길, 구영창

전체 청구항 수 : 총 15 항

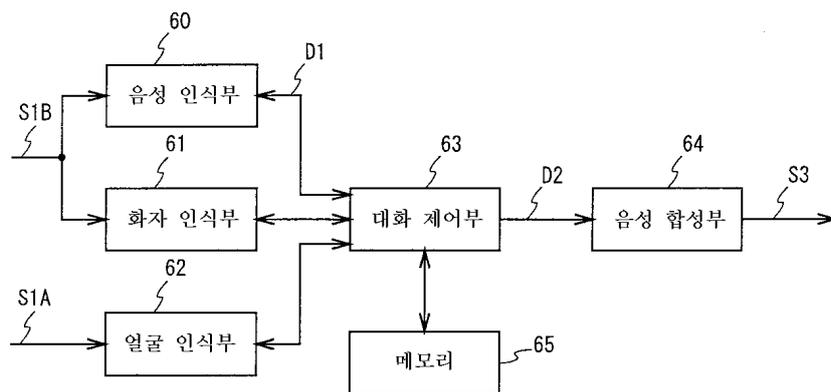
심사관 : 장호근

(54) 학습 장치, 학습 방법 및 로봇 장치

(57) 요약

종래의 로봇 장치 등에서는, 자연스럽게 이름 학습을 할 수 없었다. 사람과의 대화를 통하여 대상으로 하는 물체의 이름을 취득하고, 그 이름을 그 대상으로 하는 물체에 대하여 검출한 다른 복수의 각 특징의 데이터와 관련시켜 기억함과 함께, 이들 기억한 각 데이터 및 관련된 정보에 기초하여, 신규의 물체의 등장을 인식하고, 그 신규의 사람의 이름 및 각 특징의 데이터를 취득하고, 이들 관련된 정보를 기억하도록 하여, 물체의 이름을 순차 학습하도록 하였다.

대표도 - 도6



(72) 발명자

야마다, 게이이찌

일본 141-0001 도쿄도 시나가와쑤 기따시나가와 6
쫈메 7-35 소니 가부시키 가이샤 내

야사노, 야스하루

일본 141-0001 도쿄도 시나가와쑤 기따시나가와 6
쫈메 7-35 소니 가부시키 가이샤 내

오쿠보, 아쯔시

일본 141-0001 도쿄도 시나가와쑤 기따시나가와 6
쫈메 7-35 소니 가부시키 가이샤 내

특허청구의 범위

청구항 1

대화를 통하여 대상으로 하는 물체의 이름을 취득하는 대화 수단과,

상기 대상으로 하는 물체의 복수의 특징 데이터를 검출하고, 그 검출 결과 및 기지의 물체의 대응하는 특징의 데이터에 기초하여, 상기 대상으로 하는 물체를 인식하는 복수의 인식 수단과,

상기 기지의 물체의 이름에 대한 각 상기 인식 수단의 인식 결과를 관련시킨 관련 정보를 기억하는 기억 수단과,

상기 대화 수단이 취득한 상기 대상으로 하는 물체의 이름, 상기 대상으로 하는 물체에 대한 각 상기 인식 수단의 인식 결과, 및 상기 기억 수단이 기억하는 상기 관련 정보에 기초하여, 상기 대상으로 하는 물체가 신규의 물체인지의 여부를 판단하는 판단 수단과,

상기 판단 수단이 상기 대상으로 하는 물체를 신규의 물체라고 판단했을 때에, 상기 복수의 인식 수단에 의해 검출된 상기 대상으로 하는 물체에 대응하는 상기 복수의 특징 데이터와, 그 대상으로 하는 물체에 대한 관련 정보를 상기 기억 수단에 기억시키는 제어 수단을 갖는 학습 장치.

청구항 2

제1항에 있어서,

상기 제어 수단은,

상기 판단 수단이 상기 대상으로 하는 물체를 상기 기지의 물체라고 판단했을 때에, 그 대상으로 하는 물체를 정확하게 인식한 상기 인식 수단을 추가 학습하도록 제어하는 학습 장치.

청구항 3

제1항에 있어서,

상기 제어 수단은,

상기 판단 수단이 상기 대상으로 하는 물체를 상기 기지의 물체라고 판단했을 때에, 그 대상으로 하는 물체를 정확하게 인식할 수 없던 상기 인식 수단을, 정정 학습하도록 제어하는 학습 장치.

청구항 4

제1항에 있어서,

상기 판단 수단은,

상기 기억 수단이 기억하는 상기 관련 정보를 참조하여, 상기 대화 수단이 취득한 상기 대상으로 하는 물체의 이름 및 그 물체에 대한 각 상기 인식 수단의 인식 결과의 다수결에 의해, 상기 대상으로 하는 물체가 신규의 물체인지의 여부를 판단하는 학습 장치.

청구항 5

제1항에 있어서,

상기 제어 수단은,

상기 복수의 인식 수단이 상기 대상으로 하는 물체의 상기 복수의 특징 데이터를 충분히 수집할 때까지 대화를 연장시키도록 대화 수단을 제어하는 학습 장치.

청구항 6

대화를 통하여 대상으로 하는 물체의 이름을 취득하는 대화 단계와,

상기 대상으로 하는 물체의 복수의 특징 데이터를 검출하고, 그 검출 결과 및 기지의 물체의 대응하는 특징의 데이터에 기초하여, 상기 대상으로 하는 물체를 인식하는 복수의 인식 단계와,

상기 기지의 물체의 이름에 대한 각 상기 인식 단계의 인식 결과를 관련시킨 관련 정보를 기억하는 기억 단계와,

상기 대화 단계에서 취득된 상기 대상으로 하는 물체의 이름, 상기 대상으로 하는 물체에 대한 각 상기 인식 단계의 인식 결과, 및 상기 기억 단계에서 기억된 상기 관련 정보에 기초하여, 상기 대상으로 하는 물체가 신규의 물체인지의 여부를 판단하는 판단 단계와,

상기 판단 단계에서 상기 대상으로 하는 물체를 신규의 물체라고 판단했을 때에, 상기 복수의 인식 단계에 의해 검출된 상기 대상으로 하는 물체에 대응하는 상기 복수의 특징 데이터와, 그 대상으로 하는 물체에 대한 관련 정보를 상기 기억 단계에서 기억시키는 제어 단계

를 포함하는 학습 방법.

청구항 7

제6항에 있어서,

상기 제어 단계는,

상기 대상으로 하는 물체를 상기 기지의 물체라고 판단했을 때에, 그 대상으로 하는 물체를 정확하게 인식할 수 있었던 상기 특징에 대하여 추가 학습하는 학습 방법.

청구항 8

제6항에 있어서,

상기 제어 단계는,

상기 대상으로 하는 물체를 상기 기지의 물체라고 판단했을 때에, 그 대상으로 하는 물체를 정확하게 인식할 수 없었던 상기 특징에 대하여 정정 학습하는 학습 방법.

청구항 9

제6항에 있어서,

상기 판단 단계는,

상기 관련 정보를 참조하여, 취득한 상기 대상으로 하는 물체의 이름 및 그 물체의 각 상기 특징에 각각 기초하는 각 인식 결과의 다수결에 의해, 상기 대상으로 하는 물체가 신규의 물체인지의 여부를 판단하는 학습 방법.

청구항 10

제6항에 있어서,

상기 대화 단계에서는,

상기 복수의 인식 단계에서 상기 대상으로 하는 물체의 상기 복수의 특징 데이터를 충분히 수집할 때까지 그 대화를 연장시키는 학습 방법.

청구항 11

대화를 통하여 대상으로 하는 물체의 이름을 취득하는 대화 수단과,

상기 대상으로 하는 물체의 복수의 특징 데이터를 검출하고, 그 검출 결과 및 기지의 물체의 대응하는 특징의 데이터에 기초하여, 상기 대상으로 하는 물체를 인식하는 복수의 인식 수단과,

상기 기지의 물체의 이름에 대한 각 상기 인식 수단의 인식 결과를 관련시킨 관련 정보를 기억하는 기억 수단과,

상기 대화 수단이 취득한 상기 대상으로 하는 물체의 이름, 상기 대상으로 하는 물체에 대한 각 상기 인식 수단의 인식 결과, 및 상기 기억 수단이 기억하는 상기 관련 정보에 기초하여, 상기 대상으로 하는 물체가 신규의 물체인지의 여부를 판단하는 판단 수단과,

상기 판단 수단이 상기 대상으로 하는 물체를 신규의 물체라고 판단했을 때에, 상기 복수의 인식 수단에 의해 검출된 상기 대상으로 하는 물체에 대응하는 상기 복수의 특징 데이터와, 그 대상으로 하는 물체에 대한 관련 정보를 상기 기억 수단에 기억시키는 제어 수단을 포함하는 로봇 장치.

청구항 12

제11항에 있어서,
상기 제어 수단은,
상기 판단 수단이 상기 대상으로 하는 물체를 상기 기지의 물체라고 판단했을 때에, 그 대상으로 하는 물체를 정확하게 인식한 상기 인식 수단을, 추가 학습하도록 제어하는 로봇 장치.

청구항 13

제11항에 있어서,
상기 제어 수단은,
상기 판단 수단이 상기 대상으로 하는 물체를 상기 기지의 물체라고 판단했을 때에, 그 대상으로 하는 물체를 정확하게 인식할 수 없었던 상기 인식 수단을 정정 학습하도록 제어하는 로봇 장치.

청구항 14

제11항에 있어서,
상기 판단 수단은,
상기 기억 수단이 기억하는 상기 관련 정보를 참조하여, 상기 대화 수단이 취득한 상기 대상으로 하는 물체의 이름 및 그 물체에 대한 각 상기 인식 수단의 인식 결과의 다수결에 의해, 상기 대상으로 하는 물체가 상기 신규의 물체인지의 여부를 판단하는 로봇 장치.

청구항 15

제11항에 있어서,
상기 제어 수단은,
상기 복수의 인식 수단이 상기 대상으로 하는 물체의 상기 복수의 특징 데이터를 충분히 수집할 때까지 대화를 연장시키도록 대화 수단을 제어하는 로봇 장치.

명세서

기술분야

[0001] 본 발명은 학습 장치, 학습 방법 및 로봇 장치에 관한 것으로, 예를 들면 엔터테인먼트 로봇에 적용하기에 적합한 것이다.

배경기술

[0002] 최근, 일반 가정용 엔터테인먼트 로봇이 많이 상품화되고 있다. 그리고 이러한 엔터테인먼트 로봇 중에는 CCD(Charge Coupled Device) 카메라나 마이크로폰 등의 각종 외부 센서가 탑재되어, 이들 외부 센서의 출력에 기초하여 외부 상황을 인식하고, 인식 결과에 기초하여 자율적으로 행동할 수 있도록 구성된 것도 있다.

[0003] 그런데, 이러한 엔터테인먼트 로봇에 있어서 신규의 물체(인물도 포함함. 이하, 동일)의 이름을 그 물체와 대응시켜 기억하도록 할 수 있으면, 사용자와의 커뮤니케이션을 보다 원활히 할 수 있으며, 또 사용자로부터 「볼을 차」 등과 같은, 미리 이름이 등록된 물체 이외의 물체를 대상으로 하는 여러가지 명령에도 유연하게 대응할 수 있게 한 것으로 생각된다. 또, 이하에는 상세히 기술된 바와 같이 물체의 이름을 그 물체와 대응시켜 기억하는 것을 『이름을 학습한다』라고 표현하고, 그와 같은 기능을 『이름 학습 기능』이라고 부르기로 한다.

- [0004] 또한 이러한 이름 학습 기능을 엔터테인먼트 로봇에 탑재함에 있어서, 사람이 평소 행하는 것과 같이, 엔터테인먼트 로봇이 통상의 사람과의 대화를 통하여 신규의 물체의 이름을 학습할 수 있도록 하면, 그 자연성 측면에서 가장 바람직하고, 엔터테인먼트 로봇으로서의 오락성을 보다 한층 향상시킬 수 있는 것으로 생각된다.
- [0005] 그런데, 종래 기술에서는 이름을 학습하여야 할 신규의 물체가 언제 눈앞에 나타날지를 엔터테인먼트 로봇에게 판단시키는 것이 어려운 문제가 있다.
- [0006] 그 때문에 종래에서는 사용자가 명시적인 음성 커맨드를 제공하고 또는 로봇에 배치된 특정한 터치 센서를 가압 조작하는 등으로 동작 모드를 등록 모드로 변경하고나서, 물체의 인식 및 그 이름의 등록을 행한다고 하는 방법이 많이 이용되고 있다. 그러나, 사용자와 엔터테인먼트 로봇과의 자연스러운 인터랙션(interaction)을 생각할 때, 이러한 명시적인 지시에 의한 이름 등록은 부자연스러운 문제가 있었다.
- [0007] <발명의 개시>
- [0008] 본 발명은 이상의 점을 고려하여 이루어진 것으로, 오락성을 특히 향상시킬 수 있는 학습 장치, 학습 방법 및 로봇 장치를 제안하고자 한다.
- [0009] 이러한 과제를 해결하기 위해서, 본 발명에 따른 학습 장치에 있어서, 사람과 대화하기 위한 기능을 갖고, 해당 대화를 통하여 대상으로 하는 물체의 이름을 사람으로부터 취득하는 대화 수단과, 각각 대상으로 하는 물체가 다른 소정의 특징을 검출함과 함께, 그 검출 결과 및 미리 기억하고 있는 기지의 물체의 대응하는 특징의 데이터에 기초하여, 그 대상으로 하는 물체를 인식하는 복수의 인식 수단과, 기지의 물체의 이름 및 그 물체에 대한 각 인식 수단의 인식 결과를 관련시킨 관련 정보를 기억하는 기억 수단과, 대화 수단이 취득한 대상으로 하는 물체의 이름, 그 대상으로 하는 물체에 대한 각 인식 수단의 인식 결과, 및 기억 수단이 기억하는 관련 정보에 기초하여, 대상으로 하는 물체가 신규의 물체인지의 여부를 판단하는 판단 수단과, 판단 수단이 대상으로 하는 물체를 신규의 물체라고 판단했을 때에, 복수의 인식 수단에 의해 검출된 대상으로 하는 물체의 대응하는 특징 데이터와, 그 대상으로 하는 물체에 대한 관련 정보를 기억 수단에 기억시키는 제어 수단을 포함하도록 하였다.
- [0010] 그 결과, 이 학습 장치는 음성 커맨드의 입력이나 터치 센서의 가압 조작 등의 사용자로부터의 명시적인 지시에 의한 이름 등록을 필요로 하지 않고, 사람이 평소 행하는 것과 같이, 통상의 사람과의 대화를 통하여 신규의 인물이나 물체 등의 이름을 자연스럽게 학습할 수 있다.
- [0011] 또한 본 발명에 따른 학습 방법에 있어서, 사람과 대화하고 그 대화를 통하여 대상으로 하는 물체의 이름을 사람으로부터 취득함과 함께, 대상으로 하는 물체의 복수의 다른 소정의 특징을 각각 검출하고, 그 검출 결과 및 미리 기억하고 있는 기지의 물체의 각 특징의 데이터에 기초하여, 그 대상으로 하는 물체를 인식하는 제1 단계와, 취득한 대상으로 하는 물체의 이름과, 그 대상으로 하는 물체의 각 특징에 각각 기초하는 각 인식 결과와, 미리 기억하고 있는 기지의 물체의 이름 및 그 물체에 대한 각 인식 수단의 인식 결과를 관련시킨 관련 정보에 기초하여, 대상으로 하는 물체가 신규의 물체인지의 여부를 판단하는 제3 단계와, 대상으로 하는 물체를 신규의 물체라고 판단했을 때에 그 대상으로 하는 물체의 각 특징의 데이터, 및 그 대상으로 하는 물체에 대한 관련 정보를 각각 기억하는 제4 단계를 포함하도록 하였다.
- [0012] 그 결과, 이 학습 방법에 따르면, 음성 커맨드의 입력이나 터치 센서의 가압 조작 등의 사용자로부터의 명시적인 지시에 의한 이름 등록을 필요로 하지 않고, 사람이 평소 행하는 것과 같이, 통상의 사람과의 대화를 통하여 신규의 인물이나 물체 등의 이름을 자연스럽게 학습할 수 있다.
- [0013] 또한 본 발명에 따른 로봇 장치에 있어서, 사람과 대화하기 위한 기능을 갖고, 그 대화를 통하여 대상으로 하는 물체의 이름을 사람으로부터 취득하는 대화 수단과, 각각 대상으로 하는 물체의 다른 소정의 특징을 검출함과 함께, 그 검출 결과 및 미리 기억하고 있는 기지의 물체의 대응하는 특징의 데이터에 기초하여, 그 대상으로 하는 물체를 인식하는 복수의 인식 수단과, 기지의 물체의 이름 및 그 물체에 대한 각 인식 수단의 인식 결과를 관련시킨 관련 정보를 기억하는 기억 수단과, 대화 수단이 취득한 대상으로 하는 물체의 이름, 그 대상으로 하는 물체에 대한 각 인식 수단의 인식 결과, 및 기억 수단이 기억하는 관련 정보에 기초하여, 대상으로 하는 물체가 신규의 물체인지의 여부를 판단하는 판단 수단과, 판단 수단이 대상으로 하는 물체를 신규의 물체라고 판단했을 때에, 복수의 인식 수단에 의해 검출된 대상으로 하는 물체에 대응하는 특징 데이터와, 그 대상으로 하는 물체에 대한 관련 정보를 기억 수단에 기억시키는 제어 수단을 포함하도록 하였다.
- [0014] 그 결과, 이 로봇 장치는 음성 커맨드의 입력이나 터치 센서의 가압 조작 등의 사용자로부터의 명시적인 지시에

의한 이름 등록을 필요로 하지 않고, 사람이 평소 행하는 것과 같이, 통상의 사람과의 대화를 통하여 신규의 인물이나 물체 등의 이름을 자연스럽게 학습할 수 있다.

산업상 이용 가능성

[0326] 본 발명은 엔터테인먼트 로봇 등의 각종 로봇이나 퍼스널 컴퓨터, 시큐리티 시스템 등에 적용된다.

도면의 간단한 설명

- [0015] 도 1은 본 실시 형태에 따른 로봇의 외관 구성을 도시하는 사시도.
- [0016] 도 2는 본 실시 형태에 따른 로봇의 외관 구성을 도시하는 사시도.
- [0017] 도 3은 본 실시 형태에 따른 로봇의 외관 구성의 설명을 돕는 개략선도.
- [0018] 도 4는 본 실시 형태에 따른 로봇의 내부 구성의 설명을 돕는 개략선도.
- [0019] 도 5는 본 실시 형태에 따른 로봇의 내부 구성의 설명을 돕는 개략선도.
- [0020] 도 6은 이름 학습 기능에 관한 메인 제어부(40)의 처리의 설명을 돕는 블록도.
- [0021] 도 7은 메모리에서의 FID 및 SID와 이름과 관련된 설명을 돕는 개념도.
- [0022] 도 8은 이름 학습 처리 수순을 도시하는 흐름도.
- [0023] 도 9는 이름 학습 처리 수순을 도시하는 흐름도.
- [0024] 도 10은 이름 학습 처리 시의 대화예를 나타내는 개략선도.
- [0025] 도 11은 이름 학습 처리 시의 대화예를 나타내는 개략선도.
- [0026] 도 12는 FID 및 SID와 이름과의 신규 등록의 설명을 돕는 개념도.
- [0027] 도 13은 이름 학습 시의 대화예를 나타내는 개략선도.
- [0028] 도 14는 이름 학습 처리 시의 대화예를 나타내는 개략선도.
- [0029] 도 15는 음성 인식부의 구성을 도시하는 블록도.
- [0030] 도 16은 단어 사전의 설명을 돕는 개념도.
- [0031] 도 17은 문법 규칙의 설명을 돕는 개념도.
- [0032] 도 18은 특징 벡터 버퍼의 기억 내용의 설명을 돕는 개념도.
- [0033] 도 19는 스코어 시트의 설명을 돕는 개념도.
- [0034] 도 20은 음성 인식 처리 수순을 도시하는 흐름도.
- [0035] 도 21은 미등록어 처리 수순을 도시하는 흐름도.
- [0036] 도 22는 클러스터 분할 처리 수순을 도시하는 흐름도.
- [0037] 도 23은 시뮬레이션 결과를 나타내는 개념도.
- [0038] 도 24는 학습 시의 얼굴 인식부의 구성을 도시하는 블록도.
- [0039] 도 25는 인식 시의 얼굴 인식부의 구성을 도시하는 블록도.
- [0040] <발명을 실시하기 위한 최량의 형태>
- [0041] 이하 도면에 대하여, 본 발명의 일 실시 형태를 상세히 기술한다.
- [0042] (1) 본 실시 형태에 따른 로봇의 구성
- [0043] 도 1 및 도 2에 있어서, 참조 부호 1은 본 실시 형태에 따른 2족 보행형 로봇 전체를 나타내고, 동체부 유닛(2)의 상부에 머리부 유닛(3)이 배치됨과 함께, 그 동체부 유닛(2)의 상부 좌우에 각각 동일한 구성의 아암부 유닛(4A, 4B)이 각각 배치되고, 또한 동체부 유닛(2)의 하부 좌우에 각각 동일한 구성의 다리부 유닛(5A, 5B)이

각각 소정 위치에 부착됨으로써 구성되어 있다.

- [0044] 동체부 유닛(2)에 있어서는 몸통 상부를 형성하는 프레임(10) 및 몸통 하부를 형성하는 허리 베이스(11)가 허리 관절 기구(12)를 통하여 연결함으로써 구성되어 있고, 몸통 하부의 허리 베이스(11)에 고정된 허리 관절 기구(12)의 각 액츄에이터 A_1 , A_2 를 각각 구동함으로써, 몸통 상부를 도 3에 도시하는 직교하는 물축(13) 및 피치축(14)을 중심으로 각각 독립적으로 회전시킬 수 있도록 이루어져 있다.
- [0045] 또한 머리부 유닛(3)은 프레임(10) 상단에 고정된 어깨 베이스(15)의 상면 중앙부에 목 관절 기구(16)를 통하여 부착되어 있고, 그 목 관절 기구(16)의 각 액츄에이터 A_3 , A_4 를 각각 구동함으로써, 도 3에 도시하는 직교하는 피치축(17) 및 요우축(18)을 중심으로 각각 독립적으로 회전시킬 수 있도록 이루어져 있다.
- [0046] 또한 아암부 유닛(4A, 4B)은 각각 어깨 관절 기구(19)를 통하여 어깨 베이스(15) 좌우에 부착되어 있고, 대응하는 어깨 관절 기구(19)의 각 액츄에이터 A_5 , A_6 를 각각 구동함으로써 도 3에 도시하는 직교하는 피치축(20) 및 물축(21)을 중심으로 각각 독립적으로 회전시킬 수 있도록 이루어져 있다.
- [0047] 이 경우, 각 아암부 유닛(4A, 4B)은 각각 상 아암부를 형성하는 액츄에이터 A_7 의 출력축에 팔꿈치 관절 기구(22)를 통하여 전 아암부를 형성하는 액츄에이터 A_8 이 연결되고, 그 전 아암부의 선단에 손부(23)가 부착됨으로써 구성되어 있다.
- [0048] 그리고 각 아암부 유닛(4A, 4B)에는, 액츄에이터 A_7 를 구동함으로써 전 아암부를 도 3에 도시하는 요우축(24)을 중심으로 회전시키고, 액츄에이터 A_8 를 구동함으로써 전 아암부를 도 3에 도시하는 피치축(25)을 중심으로 각각 회전시킬 수 있도록 이루어져 있다.
- [0049] 이에 대하여 각 다리부 유닛(5A, 5B)은 각각 고관절 기구(26)를 통하여 몸통 하부의 허리 베이스(11)에 각각 부착되어 있고, 각각 대응하는 고관절 기구(26)의 각 액츄에이터 $A_9 \sim A_{11}$ 을 각각 구동함으로써, 도 3에 도시하는 상호 직교하는 요우축(27), 물축(28) 및 피치축(29)을 중심으로 각각 독립적으로 회전시킬 수 있도록 이루어져 있다.
- [0050] 이 경우 각 다리부 유닛(5A, 5B)은 각각 대퇴부를 형성하는 프레임(30) 하단에 무릎 관절 기구(31)를 통하여 하퇴부를 형성하는 프레임(32)이 연결됨과 함께, 그 프레임(32) 하단에 발목 관절 기구(33)를 통하여 발부(34)가 연결됨으로써 구성되어 있다.
- [0051] 이에 따라 각 다리부 유닛(5A, 5B)에서는 무릎 관절 기구(31)를 형성하는 액츄에이터 A_{12} 를 구동함으로써, 하퇴부를 도 3에 도시하는 피치축(35)을 중심으로 회전시킬 수 있고, 또한 발목 관절 기구(33)의 액츄에이터 A_{13} , A_{14} 를 각각 구동함으로써, 발부(34)를 도 3에 도시하는 직교하는 피치축(36) 및 물축(37)을 중심으로 각각 독립적으로 회전시킬 수 있도록 이루어져 있다.
- [0052] 한편, 동체부 유닛(2)의 몸통 하부를 형성하는 허리 베이스(11)의 배면측에는 도 4에 도시한 바와 같이, 로봇(1) 전체의 동작 제어를 담당하는 메인 제어부(40)와, 전원 회로 및 통신 회로 등의 주변 회로(41)와, 배터리(45)(도 5) 등이 박스에 수납되어 이루어지는 제어 유닛(42)이 배치되어 있다.
- [0053] 그리고 이 제어 유닛(42)은 각 구성 유닛(동체부 유닛(2), 머리부 유닛(3), 각 아암부 유닛(4A, 4B) 및 각 다리부 유닛(5A, 5B)) 내에 각각 배치된 각 서브 제어부(43A~43D)와 접속되어 있고, 이들 서브 제어부(43A~43D)에 대하여 필요한 전원 전압을 공급하거나, 이들 서브 제어부(43A~43D)와 통신을 행하거나 할 수 있도록 이루어져 있다.
- [0054] 또한 각 서브 제어부(43A~43D)는 각각 대응하는 구성 유닛 내의 각 액츄에이터 $A_1 \sim A_{14}$ 와 접속되어 있고, 그 구성 유닛 내의 각 액츄에이터 $A_1 \sim A_{14}$ 를 메인 제어부(40)로부터 제공되는 각종 제어 커맨드에 기초하여 지정된 상태로 구동할 수 있도록 이루어져 있다.
- [0055] 또한 머리부 유닛(3)에는 도 5에 도시한 바와 같이, 이 로봇(1)의 「눈」으로 기능하는 CCD(Charge Coupled Device) 카메라(50) 및 「귀」로 기능하는 마이크로폰(51) 및 터치 센서(52) 등으로 이루어지는 외부 센서부(53)와, 「입」으로 기능하는 스피커(54) 등이 각각 소정 위치에 배치고, 제어 유닛(42) 내에는 배터리 센서(55) 및 가속도 센서(56) 등으로 이루어지는 내부 센서부(57)가 배치되어 있다.

- [0056] 그리고 외부 센서부(53)의 CCD 카메라(50)는 주위의 상황을 촬상하고, 얻어진 화상 신호 S1A를 메인 제어부에 송출하는 한편, 마이크로폰(51)은 사용자로부터 음성 입력으로서 주어지는 「걸어」, 「엎드려」 또는 「불을 쫓아가」 등의 각종 명령 음성을 집음하고, 그렇게 해서 얻어진 음성 신호 S1B를 메인 제어부(40)로 송출하도록 이루어져 있다.
- [0057] 또한 터치 센서(52)는 도 1 및 도 2에서 밝힌 바와 같이 머리부 유닛(3) 상부에 설치되어 있고, 사용자로부터의 「어루만진다」나 「때린다」 등과 같은 물리적인 동작에 의해 받은 압력을 검출하고, 검출 결과를 압력 검출 신호 S1C로서 메인 제어부(40)로 송출한다.
- [0058] 또한 내부 센서부(57)의 배터리 센서(55)는 배터리(45)의 에너지 잔량을 소정 주기로 검출하고, 검출 결과를 배터리 잔량 검출 신호 S2A로서 메인 제어부(40)로 송출하는 한편, 가속도 센서(56)는 3축 방향(x축, y축 및 z축)의 가속도를 소정 주기로 검출하고, 검출 결과를 가속도 검출 신호 S2B로서 메인 제어부(40)로 송출한다.
- [0059] 메인 제어부(40)는 외부 센서부(53)의 CCD 카메라(50), 마이크로폰(51) 및 터치 센서(52) 등으로부터 각각 공급되는 화상 신호 S1A, 음성 신호 S1B 및 압력 검출 신호 S1C 등(이하, 이들을 통합하여 외부 센서 신호 S1이라고 함)과, 내부 센서부(57)의 배터리 센서(55) 및 가속도 센서 등으로부터 각각 공급되는 배터리 잔량 검출 신호 S2A 및 가속도 검출 신호 S2B 등(이하, 이들을 통합하여 내부 센서 신호 S2라고 함)에 기초하여, 로봇(1) 주위 및 내부의 상황이나 사용자로부터의 명령, 사용자로부터의 동작의 유무 등을 판단한다.
- [0060] 그리고 메인 제어부(40)는 그 판단 결과와, 미리 내부 메모리(40A)에 저장되어 있는 제어 프로그램과, 그 때 장착되어 있는 외부 메모리(58)에 저장되어 있는 각종 제어 파라미터에 기초하여 계속되는 행동을 결정하고, 결정 결과에 기초한 제어 커맨드를 대응하는 서브 제어부(43A~43D)로 송출한다. 그 결과, 이 제어 커맨드에 기초하여, 그 서브 제어부(43A~43D)의 제어 하에, 대응하는 액츄에이터 A₁~A₁₄가 구동되고, 그렇게 해서 머리부 유닛(3)을 상하 좌우로 요동시키거나 아암부 유닛(4A, 4B)을 위로 올리거나 보행하는 등의 행동이 로봇(1)에 의해 발현되게 된다.
- [0061] 또한 이 때 메인 제어부(40)는 필요에 따라 소정의 음성 신호 S3을 스피커(54)에 제공함으로써 그 음성 신호 S3에 기초한 음성을 외부로 출력시키거나 외견 상 「눈」으로 기능하는 머리부 유닛(3)의 소정 위치에 설치된 LED에 구동 신호를 출력함으로써 이것을 점멸시킨다.
- [0062] 이와 같이 하여 이 로봇(1)에 있어서는, 주위 및 내부의 상황이나 사용자로부터의 명령 및 동작의 유무 등에 기초하여 자율적으로 행동할 수 있도록 이루어져 있다.
- [0063] (2) 이름 학습 기능에 관한 메인 제어부(40)의 처리
- [0064] 다음에 이 로봇(1)에 탑재된 이름 학습 기능에 대하여 설명한다.
- [0065] 이 로봇(1)에는 사람과의 대화를 통하여 그 사람의 이름을 취득하고, 그 이름을 마이크로폰(51)이나 CCD 카메라(50)의 출력에 기초하여 검출한 그 사람의 소리의 음향적 특징 및 얼굴의 형태적 특징의 각 데이터와 관련시켜 기억함과 함께, 이들 기억한 각 데이터에 기초하여, 이름을 취득하지 않은 신규의 사람의 등장을 인식하고, 그 신규의 사람의 이름이나 소리의 음향적 특징 및 얼굴의 형태적 특징을 상술한 바와 마찬가지로 하여 취득하고 기억하도록 하고, 사람의 이름을 그 사람과 대응시켜 취득하고(이하, 이것을 이름의 학습이라고 함) 학습해가는 이름 학습 기능이 탑재되어 있다. 또 이하에서는 그 사람의 소리의 음향적 특징 및 얼굴의 형태적 특징과 대응시켜 이름을 기억해 낸 사람을 「기지의 사람」이라고 하고, 기억해 내지 못한 사람을 「신규의 사람」이라고 하기로 한다.
- [0066] 그리고 이 이름 학습 기능은 메인 제어부(40)에서의 각종 처리에 의해 실현되고 있다.
- [0067] 여기서, 이러한 이름 학습 기능에 관한 메인 제어부(40)의 처리 내용을 기능적으로 분류하면, 도 6에 도시한 바와 같이, 사람이 발성한 말을 인식하는 음성 인식부(60)와, 사람의 소리의 음향적 특징을 검출함과 함께 그 검출한 음향적 특징에 기초하여 그 사람을 식별하여 인식하는 화자 인식부(61)와, 사람의 얼굴의 형태적 특징을 검출함과 함께 그 검출한 형태적 특징에 기초하여 그 사람을 식별하여 인식하는 얼굴 인식부(62)와, 사람과의 대화 제어를 포함하는 신규의 사람의 이름 학습을 위한 각종 제어나, 기지의 사람의 이름, 소리의 음향적 특징 및 얼굴의 형태적 특징의 기억 관리를 담당하는 대화 제어부(63)와, 대화 제어부(63)의 제어 하에 각종 대화용 음성 신호 S3을 생성하여 스피커(54)(도 5)로 송출하는 음성 합성부(64)로 나눌 수 있다.
- [0068] 이 경우, 음성 인식부(60)는, 마이크로폰(51)(도 5)으로부터의 음성 신호 S1B에 기초하여 소정의 음성 인식 처

리를 실행함으로써 그 음성 신호 S1B에 포함되는 말을 단어 단위로 인식하는 기능을 갖는 것으로, 인식한 이들 단어를 문자열 데이터 D1로서 대화 제어부(63)로 송출하도록 이루어져 있다.

- [0069] 또 화자 인식부(61)는 마이크로폰(51)으로부터 주어지는 음성 신호 S1B에 포함되는 사람의 소리의 음향적 특징을, 예를 들면 "Segregation of Speakers for Recognition and Speaker Identification(CH2977-7/91/0000~0873 S1.00 1991 IEEE)"에 기재된 방법 등을 이용한 소정의 신호 처리에 의해 검출하는 기능을 갖고 있다.
- [0070] 그리고 화자 인식부(61)는, 통상은 이 검출한 음향적 특징의 데이터를 그 때 기억하고 있는 모든 기지의 사람의 음향적 특징의 데이터와 순차 비교하여, 그 때 검출한 음향적 특징이 어느 기지의 사람의 음향적 특징과 일치한 경우에는 그 기지의 사람의 음향적 특징과 대응시킨 그 음향적 특징에 고유의 식별자(이하, 이것을 SID라고 함)를 대화 제어부(63)로 통지하는 한편, 검출한 음향적 특징이 어느 기지의 사람의 음향적 특징과도 일치하지 않은 경우에는, 인식 불능을 의미하는 SID(=-1)을 대화 제어부(63)로 통지하도록 이루어져 있다.
- [0071] 또한 화자 인식부(61)는, 대화 제어부(63)가 신규의 사람이라고 판단했을 때에 대화 제어부(63)로부터 제공되는 신규 학습의 개시 명령 및 학습 종료 명령에 기초하여, 그 동안 그 사람의 소리의 음향적 특징을 검출하고, 그 검출한 음향적 특징의 데이터를 새로운 고유의 SID와 대응시켜 기억함과 함께, 이 SID를 대화 제어부(63)로 통지하도록 이루어져 있다.
- [0072] 또 화자 인식부(61)는 대화 제어부(63)로부터의 추가 학습이나 정정 학습의 개시 명령 및 종료 명령에 따라, 그 사람의 소리의 음향적 특징의 데이터를 추가적으로 수집하는 추가 학습이나, 그 사람의 소리의 음향적 특징의 데이터를 그 사람을 정확하게 인식할 수 있도록 정정하는 정정 학습을 행할 수 있게 이루어져 있다.
- [0073] 얼굴 인식부(62)는 CCD 카메라(50)(도 5)로부터 제공되는 화상 신호 S1A를 항상 감시하고, 그 화상 신호 S1A에 기초한 화상 내에 포함되는 사람의 얼굴의 형태적 특징을 소정의 신호 처리에 의해 검출하는 기능을 갖고 있다.
- [0074] 그리고 얼굴 인식부(62)는 통상은, 이 검출한 형태적 특징의 데이터를 그 때 기억하고 있는 모든 기지의 사람의 얼굴의 형태적 특징의 데이터와 순차 비교하여, 그 때 검출한 형태적 특징이 어느 기지의 사람의 얼굴의 형태적 특징과 일치한 경우에는 해당 기지의 사람의 형태적 특징과 대응시킨 해당 형태적 특징에 고유의 식별자(이하, 이것을 FID라고 함)를 대화 제어부에 통지하는 한편, 검출한 형태적 특징 중 어느 기지의 사람의 얼굴의 형태적 특징과도 일치하지 않는 경우에는 인식 불능을 의미하는 FID(=-1)을 대화 제어부에 통지하도록 이루어져 있다.
- [0075] 또한 얼굴 인식부(62)는 대화 제어부(63)가 신규의 사람이라고 판단했을 때에 대화 제어부(63)로부터 제공되는 학습 개시 명령 및 학습 종료 명령에 기초하여, 그 동안에 CCD 카메라(50)로부터의 화상 신호 S1A에 기초한 화상 내에 포함되는 사람의 얼굴의 형태적 특징을 검출하고, 검출한 형태적 특징의 데이터를 새로운 고유의 FID와 대응시켜 기억함과 함께, 이 FID를 대화 제어부(63)로 통지하도록 이루어져 있다.
- [0076] 또 얼굴 인식부(62)는 대화 제어부(63)로부터의 추가 학습이나 정정 학습의 개시 명령 및 종료 명령에 따라, 사람의 얼굴의 형태적 특징의 데이터를 추가적으로 수집하는 추가 학습이나 사람의 얼굴의 형태적 특징의 데이터를 그 사람을 정확하게 인식할 수 있도록 정정하는 정정 학습을 행할 수 있도록 이루어져 있다.
- [0077] 음성 합성부(64)는 대화 제어부(63)로부터 제공되는 문자열 데이터 D2를 음성 신호 S3으로 변환하는 기능을 갖고, 그렇게 해서 얻어진 음성 신호 S3을 스피커(54)(도 5)로 송출하도록 이루어져 있다. 이에 따라 이 음성 신호 S3에 기초한 음성을 스피커(54)로부터 출력시킬 수 있도록 이루어져 있다.
- [0078] 대화 제어부(63)에서는 도 7에 도시한 바와 같이, 기지의 사람의 이름과, 화자 인식부(61)가 기억하고 있는 그 사람의 소리의 음향적 특징의 데이터에 대응된 SID와, 얼굴 인식부(62)가 기억하고 있는 그 사람의 얼굴의 형태적 특징의 데이터에 대응된 FID를 관련시켜 기억하는 메모리(65)(도 6)를 갖고 있다.
- [0079] 그리고 대화 제어부(63)는 소정의 타이밍에서 소정의 문자열 데이터 D2를 음성 합성부(64)에 제공함으로써, 말하는 상대의 사람에 대하여 이름을 질문하고 또는 이름을 확인하기 위한 음성 등을 스피커(54)로부터 출력시키는 한편, 이 때의 그 사람의 응답 등에 기초한 음성 인식부(60) 및 화자 인식부(61)의 각 인식 결과 및 그 사람에 대한 얼굴 인식부(62)의 인식 결과와, 메모리(65)에 저장된 상술한 기지의 사람의 이름, SID 및 FID의 관련 정보에 기초하여 그 사람이 신규의 사람인지의 여부를 판단하도록 이루어져 있다.
- [0080] 그리고 대화 제어부(63)는 그 사람이 신규의 사람이라고 판단했을 때에는 화자 인식부(61) 및 얼굴 인식부(62)에 대하여 신규 학습의 개시 명령 및 종료 명령을 공급함으로써, 이들 화자 인식부(61) 및 얼굴 인식부(62)에

그 신규의 사람의 소리의 음향적 특징이나 얼굴의 형태적 특징의 데이터를 수집 및 기억시킴과 함께, 그 결과로서 이들 화자 인식부(61) 및 얼굴 인식부(62)로부터 각각 제공되는 그 신규의 사람의 소리의 음향적 특징의 데이터나 얼굴의 형태적 특징의 데이터에 대응된 SID 및 FID를, 이러한 대화에 의해 얻어진 그 사람의 이름과 관련시켜 메모리(65)에 저장하도록 이루어져 있다.

[0081] 또한 대화 제어부(63)는 그 사람이 기지의 사람이라고 판단했을 때에는 필요에 따라 화자 인식부(61) 및 얼굴 인식부(62)에 추가 학습이나 정정 학습의 개시 명령을 공급함으로써 화자 인식부(61) 및 얼굴 인식부(62)에 추가 학습이나 정정 학습을 행하게 하는 한편, 이와 함께 음성 합성부(64)에 소정의 문자열 데이터 D2를 소정의 타이밍으로 순차 송출함으로써, 화자 인식부(61) 및 얼굴 인식부(62)가 추가 학습이나 정정 학습을 하는데 필요한 상당량의 데이터를 수집할 수 있을 때까지 그 사람과의 대화를 오래 끌게 하는 등의 대화 제어를 행하도록 이루어져 있다.

[0082] (3) 이름 학습 기능에 관한 대화 제어부(63)의 구체적 처리

[0083] 다음에, 이름 학습 기능에 관한 대화 제어부(63)의 구체적인 처리 내용에 대하여 설명한다.

[0084] 대화 제어부(63)는 외부 메모리(58)(도 5)에 저장된 제어 프로그램에 기초하여, 도 8 및 도 9에 도시하는 이름 학습 처리 수순 RT1에 따라 신규의 사람의 이름을 순차 학습하기 위한 각종 처리를 실행한다.

[0085] 즉 대화 제어부(63)는 CCD 카메라(50)로부터의 화상 신호 S1A에 기초하여 얼굴 인식부(62)가 사람의 얼굴을 인식함으로써 그 얼굴 인식부(62)로부터 FID가 제공되면 이름 학습 처리 수순 RT1을 단계 SP0에서 개시하고, 계속되는 단계 SP1에서, 메모리(65)에 저장된 기지의 사람의 이름과, 이에 대응하는 SID 및 이에 대응하는 FID를 관련시킨 정보(이하, 이것을 관련 정보라고 함)에 기초하여 그 FID로부터 대응하는 이름을 검색 가능한지의 여부(즉, FID가 인식 불능을 의미하는 「-1」 인지의 여부)를 판단한다.

[0086] 여기서 이 단계 SP1에서 긍정 결과를 얻는 것은, 그 사람이 얼굴 인식부(62)가 그 사람의 얼굴의 형태적 특징의 데이터를 기억하고 있고, 그 데이터와 대응된 FID가 그 사람의 이름과 관련되어 메모리(65)에 저장되어 있는 기지의 사람인 것을 의미한다. 단 이 경우에도, 얼굴 인식부(62)가 신규의 사람을 기지의 사람이라고 오인식한 것으로도 생각할 수 있다.

[0087] 그래서 대화 제어부(63)는 단계 SP1에서 긍정 결과를 얻은 경우에는 단계 SP2로 진행하여 소정의 문자열 데이터 D2를 음성 합성부(64)로 송출함으로써, 예를 들면 도 10에 도시한 바와 같이, 「○○씨군요.」 등과 같이 그 사람의 이름이 FID로부터 검색된 이름(상술한 ○○에 해당하는 이름)과 일치하는지의 여부를 확인하기 위한 질문의 음성을 스피커(54)로부터 출력시킨다.

[0088] 계속해서 대화 제어부(63)는 단계 SP3으로 진행하여, 이러한 질문에 대한 그 사람의 「예, 그렇습니다」나 「아니요, 틀렸습니다.」 등과 같이 응답의 음성 인식 결과가 음성 인식부(60)로부터 제공되는 것을 기다린다. 그리고 대화 제어부(63)는, 음성 인식부(63)로부터 이와 같은 음성 인식 결과가 주어지고, 또한 화자 인식부(61)로부터 그 때의 화자 인식 결과인 SID가 주어지면, 단계 SP4로 진행하여, 음성 인식부(63)로부터의 음성 인식 결과에 기초하여 그 사람의 응답이 긍정적인 것인지의 여부를 판단한다.

[0089] 여기서 이 단계 SP4에서 긍정 결과를 얻는 것은, 단계 SP1에서 얼굴 인식부(62)로부터 제공된 FID에 기초하여 검색된 이름이 그 사람의 이름과 일치하고, 따라서 그 사람은 대화 제어부(63)가 검색한 이름을 갖는 본인이라고 거의 단정할 수 있는 상태에 있는 것을 의미한다.

[0090] 그렇게 해서 이 때 대화 제어부(63)는 그 사람이 대화 제어부(63)가 검색한 이름을 갖는 본인이라고 단정하고, 단계 SP5로 진행하여 화자 인식부(61)에 대하여 추가 학습의 개시 명령을 공급한다. 또한 이와 함께 대화 제어부(63)는 최초로 화자 인식부(61)로부터 주어진 SID가, 이러한 이름으로부터 메모리(65)에 저장된 관련 정보에 기초하여 검색할 수 있는 SID와 일치하고 있는 경우에는 화자 인식부(61)에 대하여 추가 학습의 개시 명령을 공급하고, 이것에 대하여 일치하지 않은 경우에는 정정 학습의 개시 명령을 공급한다.

[0091] 그리고 대화 제어부(63)는 이 후 단계 SP6으로 진행하여 예를 들면 도 10과같이 「오늘 날씨가 좋군요.」 등과 같이, 그 사람과의 대화를 오래 끌게하기 위한 잡담을 하게 하기 위한 문자열 데이터 D2를 음성 합성부(64)에 순차 송출하고, 이 후 추가 학습 또는 정정 학습에 충분한 소정 시간이 경과하면, 단계 SP7로 진행하여 화자 인식부(61) 및 얼굴 인식부(62)에 대하여 추가 학습 또는 정정 학습의 종료 명령을 공급한 후, 단계 SP20으로 진행하여 그 사람에 대한 이름 학습 처리를 종료한다.

[0092] 한편, 단계 SP1에서 부정 결과를 얻는 것은 얼굴 인식부(62)에 의해 얼굴 인식된 사람이 신규의 사람이거나, 또

는 얼굴 인식부(62)가 기지의 사람을 신규의 사람이라고 오인식한 것을 의미한다. 또한 단계 SP4에서 부정 결과를 얻는 것은 최초로 얼굴 인식부(62)로부터 제공된 FID에서 검색된 이름이 그 사람의 이름과 일치하지 않은 것을 의미한다. 그리고, 이들 중 어느 하나의 경우에도, 대화 제어부(63)가 그 사람을 정확하게 파악하지 않은 상태에 있다고 할 수 있다.

- [0093] 그래서 대화 제어부(63)는 단계 SP1에서 부정 결과를 얻었을 때나 단계 SP4에서 부정 결과를 얻었을 때에는, 단계 SP8로 진행하여 음성 합성부(64)에 문자열 데이터 D2를 제공함으로써, 예를 들면 도 11에 도시한 바와 같이, 「저, 이름을 가르쳐 주세요..」 등과 같이, 그 사람의 이름을 묻기 위한 질문의 음성을 스피커(54)로부터 출력시킨다.
- [0094] 그리고 대화 제어부(63)는 이 후 단계 SP9로 진행하여, 이러한 질문에 대한 그 사람의 「○○입니다..」 등과 같은 응답의 음성 인식 결과(즉 이름)와, 해당 응답 시의 화자 인식부(61)의 화자 인식 결과(즉 SID)가 각각 음성 인식부(60) 및 화자 인식부(61)로부터 제공되는 것을 기다린다.
- [0095] 그리고 대화 제어부(63)는, 음성 인식부(60)로부터 음성 인식 결과가 제공되고, 화자 인식부(61)로부터 SID가 제공되면, 단계 SP10으로 진행하며, 이들 음성 인식 결과 및 SID 및 최초로 얼굴 인식부(62)로부터 제공된 FID에 기초하여, 그 사람이 신규의 사람인지의 여부를 판단한다.
- [0096] 여기서 이 실시 형태의 경우, 이러한 판단은 음성 인식부(60)의 음성 인식에 의해 얻어진 이름과, 화자 인식부(61)로부터의 SID와, 얼굴 인식부(62)로부터의 FID로 이루어지는 3개의 인식 결과의 다수결에 의해 행해진다.
- [0097] 예를 들면, 화자 인식부(61)로부터의 SID 및 얼굴 인식부(62)로부터의 FID가 모두 인식 불능을 의미하는 「-1」이고, 또한 단계 SP에서 음성 인식부(60)로부터의 음성 인식 결과에 기초하여 얻어진 그 사람의 이름이 메모리(65)에서 어떤 SID나 FID와도 관련되어 있지 않은 경우에는, 그 사람이 신규의 사람이라고 판단한다. 기지의 어떤 얼굴 또는 어떤 소리라도 닮지 않은 사람이 새로운 이름을 가지고 있다고 하는 상황이므로, 그와 같은 판단을 할 수 있다.
- [0098] 또한 대화 제어부(63)는 화자 인식부(61)로부터의 SID 및 얼굴 인식부(62)로부터의 FID가 메모리(65)에서 다른 이름과 관련되어 있거나 또는 그 한쪽이 인식 불능을 의미하는 「-1」이고, 또한 단계 SP9에서 음성 인식부(60)로부터의 음성 인식 결과에 기초하여 얻어진 그 사람의 이름이 메모리(65)에 저장되어 있지 않은 경우에도, 그 사람이 신규의 사람이라고 판단한다. 이것은 각종 인식 처리에 있어서, 신규 카테고리를 기지 카테고리 중 어느 하나라고 오인식하는 것은 발생하기 쉬우며, 또한 음성 인식된 이름이 등록되어 있지 않은 것을 생각하면, 꽤 높은 확신도를 갖고 신규인 사람이라고 판단할 수 있기 때문이다.
- [0099] 이에 대하여 대화 제어부(63)는 화자 인식부(61)로부터의 SID 및 얼굴 인식부(62)로부터의 FID가 메모리(65)에서 동일한 이름과 관련되어 있고, 또한 단계 SP9에서 음성 인식부(60)로부터의 음성 인식 결과에 기초하여 얻어진 그 사람의 이름이 그 SID 및 FID가 관련된 이름인 경우에는 그 사람이 기지의 사람이라고 판단한다.
- [0100] 또한 대화 제어부(63)는 화자 인식부(61)로부터의 SID 및 얼굴 인식부(62)로부터의 FID가 메모리(65)에서 다른 이름과 관련되어 있고, 또한 단계 SP9에서 음성 인식부(60)로부터의 음성 인식 결과에 기초하여 얻어진 그 사람의 이름이 이러한 SID 또는 FID 중 한쪽이 관련된 이름인 경우에는 그 사람이 기지의 사람이라고 판단한다. 이 경우에는 화자 인식부(61) 및 얼굴 인식부(62) 중 어느 한쪽의 인식 결과가 틀렸다고 생각되기 때문에, 이러한 다수결에 의해 그와 같이 판단한다.
- [0101] 한편, 대화 제어부(63)는 화자 인식부(61)로부터의 SID 및 얼굴 인식부(62)로부터의 FID가 메모리(65)에서 다른 이름과 관련되어 있고, 또한 단계 SP9에서 음성 인식부(60)로부터의 음성 인식 결과에 기초하여 얻어진 그 사람의 이름이 메모리(65)에서 이러한 SID 및 FID 중 어디에도 관련되어 있지 않은 이름인 경우에는 그 사람이 기지의 사람인지 또는 신규의 사람인지를 판단하지 않는다. 이 경우에는 음성 인식부(60), 화자 인식부(61) 및 얼굴 인식부(62) 중 하나 또는 전부의 인식이 틀렸다고도 생각할 수 있지만, 이 단계에서는 그것을 판정할 수 없다. 따라서 이 경우에는 이러한 판단을 보류한다.
- [0102] 그리고 대화 제어부(63)는 이러한 판단 처리에 의해, 단계 SP10에서 이러한 사람이 신규의 사람이라고 판단한 경우에는 단계 SP11로 진행하여 신규 학습의 개시 명령을 화자 인식부(61) 및 얼굴 인식부(62)에 제공하고, 이후 단계 SP12로 진행하여 예를 들면 도 11과 같이 「저는 로봇입니다. 잘 부탁드립니다..」 또는 「○○씨, 오늘 날씨가 좋군요..」 등과 같이 그 사람과의 대화를 오래 끌게 하는 잡담을 하기 위한 문자열 데이터 D2를 음성 합

성부(64)로 송출한다.

- [0103] 또한 대화 제어부(63)는 이 후 단계 SP13으로 진행하여 화자 인식부(61)에서의 음향적 특징의 데이터의 수집 및 얼굴 인식부(62)에서의 얼굴의 형태적 특징의 데이터의 수집이 모두 충분량에 도달했는지의 여부를 판단하고, 부정 결과를 얻으면 단계 SP12로 되돌아가, 이 후 단계 SP13에서 긍정 결과를 얻기까지 단계 SP12-SP13-SP12의 루프를 반복한다.
- [0104] 그리고 대화 제어부(63)는, 화자 인식부(61)에서의 음향적 특징의 데이터의 수집 및 얼굴 인식부(62)에서의 얼굴의 형태적 특징의 데이터의 수집이 모두 충분량에 도달함으로써 단계 SP13에서 긍정 결과를 얻으면, 단계 SP14로 진행하여, 이들 화자 인식부(61) 및 얼굴 인식부(62)에 신규 학습의 종료 명령을 공급한다. 그 결과, 화자 인식부(61)에서 그 음향적 특징의 데이터가 새로운 SID와 대응되어 기억되고, 얼굴 인식부(62)에서 그 형태적 특징의 데이터가 새로운 FID와 대응되어 기억된다.
- [0105] 또한 대화 제어부(63)는 이 후 단계 SP15로 진행하여, 화자 인식부(61) 및 얼굴 인식부(62)로부터 각각 이러한 SID 및 FID가 제공되는 것을 대기하고, 이들이 주어지면, 예를 들면 도 12에 도시한 바와 같이, 이들을 단계 SP9에서 음성 인식부(60)로부터의 음성 인식 결과에 기초하여 얻어진 그 사람의 이름과 관련시켜 메모리(65)에 등록한다. 그리고 대화 제어부(63)는 이 후 단계 SP20으로 진행하여 그 사람에 대한 이름 학습 처리를 종료한다.
- [0106] 이에 대하여 대화 제어부(63)는 단계 SP10에서, 이러한 사람이 기지의 사람이라고 판단한 경우에는 단계 SP16으로 진행하고, 화자 인식부(61) 및 얼굴 인식부(62)가 그 기지의 사람을 정확하게 인식하고 있는 경우(즉 화자 인식부(61)나 얼굴 인식부(62)가 관련 정보로서 메모리(65)에 저장된 그 기지의 사람에 대응하는 SID 또는 FID와 동일한 SID 또는 FID를 인식 결과로서 출력하고 있는 경우)에는, 그 화자 인식부(61) 또는 얼굴 인식부(62)에 대하여 추가 학습의 개시 명령을 공급하고, 화자 인식부(61) 및 얼굴 인식부(62)가 그 기지의 사람을 정확하게 인식할 수 없던 경우(즉 화자 인식부(61)나 얼굴 인식부(62)가 관련 정보로서 메모리(65)에 저장된 그 기지의 사람에 대응하는 SID 또는 FID와 동일한 SID 또는 FID를 인식 결과로서 출력하고 있는 경우)에는 그 화자 인식부(61) 또는 얼굴 인식부(62)에 대하여 정정 학습의 개시 명령을 공급한다.
- [0107] 구체적으로는, 대화 제어부(63)는 단계 SP9에서 얻어진 화자 인식부(61)로부터의 SID와, 최초로 얼굴 인식부(62)로부터 제공된 FID가 메모리(65)에서 동일한 이름과 관련되어 있고, 또한 단계 SP9에서 음성 인식부(60)로부터의 음성 인식 결과에 기초하여 얻어진 이름이 그 SID 및 FID가 관련된 이름인 것에 의해 단계 SP10에서 그 사람이 기지의 사람이라고 판단했을 때에는 화자 인식부(61) 및 얼굴 인식부(62)에 대하여 각각 추가 학습의 개시 명령을 공급한다.
- [0108] 또한 화자 인식부(63)는 단계 SP9에서 얻어진 화자 인식부(61)로부터의 SID와, 최초로 얼굴 인식부(62)로부터 제공된 FID가 메모리(65)에서 다른 이름과 관련되어 있고, 또한 단계 SP9에서 음성 인식부(60)로부터의 음성 인식 결과에 기초하여 얻어진 이름이 이러한 SID 또는 FID 한쪽이 관련된 이름인 것에 의해 단계 SP10에서 그 사람이 기지의 사람이라고 판단했을 때에는 음성 인식부(60)로부터의 음성 인식 결과에 기초하여 얻어진 이름과 관련된 SID 또는 FID를 출력한 한쪽의 화자 인식부(61) 또는 얼굴 인식부(62)에 대하여 추가 학습의 개시 명령을 공급하고, 음성 인식부(60)로부터의 음성 인식 결과에 기초하여 얻어진 이름과 관련되어 있지 않는 FID 또는 SID를 출력한 다른 쪽의 얼굴 인식부(62) 또는 화자 인식부(61)에 정정 학습의 개시 명령을 공급한다.
- [0109] 그리고 대화 제어부(63)는 이 후 단계 SP17로 진행하여, 예를 들면 도 13에 도시한 바와 같이, 「어머나 ○○씨 이군요. 생각났어요. 오늘 날씨가 좋군요.」, 「전에, 음... 언제 만났더라?」 등의 그 사람과의 대화를 오래 끌기 위한 잡담을 하기 하기 위한 문자열 데이터 D2를 음성 합성부(64)에 순차 송출하고, 이 후 추가 학습 또는 정정 학습에 충분한 소정 시간이 경과하면, 단계 SP18로 진행하여 화자 인식부(61) 및 얼굴 인식부(62)에 대하여 추가 학습 또는 정정 학습의 종료 명령을 공급한 후, 단계 SP20으로 진행하여 그 사람에 대한 이름 학습 처리를 종료한다.
- [0110] 한편, 대화 제어부(63)는 단계 SP10에서 이러한 사람이 기지의 사람인지 신규의 사람인지를 판정할 수 없다고 판단한 경우에는 단계 SP19로 진행하고, 예를 들면 도 14에 도시한 바와 같이, 「아, 그렇습니까? 안녕하셨어요?」 등의 잡담을 하게 하기 위한 문자열 데이터 D2를 음성 합성부(64)로 순차 송출한다.
- [0111] 그리고 이 경우에는 대화 제어부(63)는 신규 학습, 추가 학습 또는 정정 학습의 개시 명령 및 그 종료 명령을 화자 인식부(61) 및 얼굴 인식부(62)에 제공하지 않고(즉 신규 학습, 추가 학습 및 정정 학습 모두 화자 인식부(61) 및 얼굴 인식부(62)에 행하게 하지 않고), 소정 시간이 경과하면, 단계 SP20으로 진행하여 그 사람에 대한

이름 학습 처리를 종료한다.

- [0112] 이와 같이 하여 대화 제어부(63)는 음성 인식부(60), 화자 인식부(61) 및 얼굴 인식부(62)의 각 인식 결과에 기초하여, 사람과의 대화 제어나 화자 인식부(61) 및 얼굴 인식부(62)의 동작 제어를 행함으로써, 신규의 사람의 이름을 순차 학습할 수 있도록 이루어져 있다.
- [0113] (4) 음성 인식부(60) 및 얼굴 인식부(62)의 구체적 구성
- [0114] 다음에, 상술한 바와 같은 이름 학습 기능을 구현화하기 위한 음성 인식부(60) 및 얼굴 인식부(62)의 구체적 구성에 대하여 설명한다.
- [0115] (4-1) 음성 인식부(60)의 구체적 구성
- [0116] 도 15는 이러한 음성 인식부(60)의 구체적 구성을 도시하는 것이다.
- [0117] 이 음성 인식부(60)에서는 마이크로폰(51)으로부터의 음성 신호 S1B를 AD(Analog Digital) 변환부(70)에 입력한다. AD 변환부(70)는 공급되는 아날로그 신호인 음성 신호 S1B를 샘플링, 양자화하고, 디지털 신호인 음성 데이터로 A/D 변환한다. 이 음성 데이터는 특징 추출부(71)에 공급된다.
- [0118] 특징 추출부(71)는 입력되는 음성 데이터에 대하여, 적당한 프레임마다, 예를 들면 MFCC(Mel Frequency Cepstrum Coefficient) 분석을 행하여, 그 분석의 결과 얻어지는 MFCC를, 특징 벡터(특징 파라미터)로서 매칭부(72)와 미등록어 구간 처리부(76)로 출력한다. 또, 특징 추출부(71)에서는 그 후 예를 들면 선형 예측 계수, cepstrum 계수, 선 스펙트럼쌍, 소정의 주파수마다의 파워(필터 बैं크의 출력) 등을, 특징 벡터로서 추출하는 것이 가능하다.
- [0119] 매칭부(72)는 특징 추출부(71)로부터의 특징 벡터를 이용하여, 음향 모델 기억부(73), 사전 기억부(74) 및 문법 기억부(75)를 필요에 따라 참조하여, 마이크로폰(51)에 입력된 음성(입력 음성)을, 예를 들면 연속 분포 HMM(Hidden Markov Model)법에 기초하여 음성 인식한다.
- [0120] 즉 음향 모델 기억부(73)는 음성 인식하는 음성의 언어에서의 개개의 음소나, 음절, 음운 등의 서브 워드에 대하여 음향적인 특징을 나타내는 음향 모델(예를 들면, HMM 외에, DP(Dynamic Programming) 매칭에 이용되는 표준 패턴 등을 포함함)을 기억하고 있다. 또, 여기서는 연속 분포 HMM법에 기초하여 음성 인식을 행하는 것을 하고 있기 때문에, 음향 모델로서는 HMM(Hidden Markov Model)이 이용된다.
- [0121] 사전 기억부(74)는 인식 대상의 각 단위마다 클러스터링된, 그 단어의 발음에 관한 정보(음향 정보)와, 그 단어의 표제가 대응된 단어 사전을 인식하고 있다.
- [0122] 여기서, 도 16은 사전 기억부(74)에 기억된 단어 사전을 나타내고 있다.
- [0123] 도 16에 도시한 바와 같이, 단어 사전에서는 단어의 표제와 그 음운 계열이 대응되고 있고, 음운 계열은 대응하는 단어마다 클러스터링되어 있다. 도 16의 단어 사전에는 1개의 엔트리(도 16의 1행)가 1개의 클러스터에 상당한다.
- [0124] 또, 도 16에서, 표제는 로마자와 일본어(가나한자)로 나타내고 있고, 음운 계열은 로마자로 나타내고 있다. 단, 음운 계열에서의 「N」은 발음 「ん」을 나타낸다. 또한, 도 16에서는 1개의 엔트리에 1개의 음운 계열을 기술하고 있지만, 1개의 엔트리에 복수의 음운 계열을 기술하는 것도 가능하다.
- [0125] 도 4로 되돌아가, 문법 기억부(26)는 사전 기억부(25)의 단어 사전에 등록되어 있는 각 단어가 어떻게 연쇄되는지(연결되는지)를 기술한 문법 규칙을 기억하고 있다.
- [0126] 여기서, 도 17은 문법 기억부(75)에 기억된 문법 규칙을 나타내고 있다. 또, 도 17의 문법 규칙은 EBNF(Extended Backus Naur Form)로 기술되어 있다.
- [0127] 도 17에서는, 행두로부터 최초로 나타나는 「;」까지가 1개의 문법 규칙을 나타내고 있다. 또 선두에 「\$」가 첨부된 알파벳(열)은 변수를 나타내고, 「\$」가 첨부되어 있지 않은 알파벳(열)은 단어의 표제(도 16에 도시한 로마자에 의한 표제)를 나타낸다. 또한 []으로 둘러싸인 부분은 생략 가능한 것을 나타내고, 「|」은 그 전후에 배치된 표제의 단어(혹은 변수) 중 어느 한쪽을 선택하는 것을 나타낸다.
- [0128] 따라서, 도 17에서, 예를 들면 제1행(위에서 1행째)의 문법 규칙 「\$col=[Kono|sono] iro wa;」는, 변수 \$col이, 「이 색은」 또는 「그 색은」 등과 같은 단어 열인 것을 나타낸다.

- [0129] 또, 도 17에 도시한 문법 규칙에서는, 변수 \$sil과 \$garbage가 정의되어 있지 않지만, 변수 \$sil은 무음의 음향 모델(무음 모델)을 나타내고, 변수 \$garbage는 기본적으로는 음운끼리의 사이에서의 자유로운 천이를 허가한 가비지 모델을 나타낸다.
- [0130] 다시 도 15로 되돌아가, 매칭부(72)는 사전 기억부(74)의 단어 사전을 참조함으로써, 음향 모델 기억부(73)에 기억되어 있는 음향 모델을 접속함으로써, 단어의 음향 모델(단어 모델)을 구성한다. 또한 매칭부(72)는 몇몇 단어 모델을 문법 기억부(75)에 기억된 문법 규칙을 참조함으로써 접속하고, 그와 같이 하여 접속된 단어 모델을 이용하여, 특징 벡터에 기초하여, 연속 분포 HMM법에 의해, 마이크로폰(51)에 입력된 음성을 인식한다. 즉 매칭부(72)는 특징 추출부(71)가 출력하는 시계열의 특징 벡터가 관측되는 스코어(우도)가 가장 높은 단어 모델의 계열을 검출하고, 그 단어 모델의 계열에 대응하는 단어 열의 표제를 음성의 인식 결과로서 출력한다.
- [0131] 보다 구체적으로는, 매칭부(72)는 접속된 단어 모델에 대응하는 단어에 의해 접속하고, 그와 같이 하여 접속된 단어 모델을 이용하여 특징 벡터에 기초하여, 연속 분포 HMM법에 의해, 마이크로폰(51)에 입력된 음성을 인식한다. 즉 매칭부(72)는 특징 추출부(71)가 출력하는 시계열의 특징 벡터가 관측되는 스코어(우도)가 가장 높은 단어 모델의 계열을 검출하고, 그 단어 모델의 계열에 대응하는 단어 열의 표제를 음성 인식 결과로서 출력한다.
- [0132] 보다 구체적으로는, 매칭부(72)는 접속된 단어 모델에 대응하는 단어 열에 대하여, 각 특징 벡터의 출현 확률(출력 확률)을 누적하고, 그 누적치를 스코어로서, 그 스코어를 가장 높게 하는 단어 열의 표제를 음성 인식 결과로서 출력한다.
- [0133] 이상과 같이 하여 출력되는 마이크로폰(51)에 입력된 음성 인식 결과는, 문자열 데이터 D1로서 대화 제어부(63)에 출력된다.
- [0134] 여기서 도 17의 실시 형태에서는, 제9행(위에서 9행째)에 가비지 모델을 나타내는 변수 \$garbage를 이용한 문법 규칙(이하, 적절하게, 미등록어용 규칙이라고 함) 「\$pat1=\$color1 \$garbage \$color2;」가 있지만, 매칭부(72)는 이 견등록어용 규칙이 적용된 경우에는 변수 \$garbage에 대응하는 음성 구간을 미등록어의 음성 구간으로서 검출한다. 또한, 매칭부(72)는 미등록어용 규칙이 적용된 경우에서의 변수 \$garbage가 나타내는 가비지 모델에서의 음운의 천이로서의 음운 계열을 미등록어의 음운 계열로서 검출한다. 그리고 매칭부(72)는 미등록어용 규칙이 적용된 음성 인식 결과가 얻어진 경우에 검출되는 미등록어의 음성 구간과 음운 계열을 미등록어 구간 처리부(76)에 공급한다.
- [0135] 또 상술한 미등록어용 규칙 「\$pat1=\$color1 \$garbage \$color2;」에 따르면, 변수 #color1로 나타나는 단어 사전에 등록되어 있는 단어(열)의 음운 계열과, 변수 \$color2로 나타나는 단어 사전에 등록되어 있는 단어(열)의 음운 계열과의 사이에 있는 1개의 미등록어가 검출되지만, 이 실시 형태에서는, 발화에 복수의 미등록어가 포함되어 있는 경우나 미등록어가 단어 사전에 등록되어 있는 단어(열) 사이에 있지 않은 경우에도 적용 가능하다.
- [0136] 미등록어 구간 처리부(76)는 특징 추출부(71)로부터 공급되는 특징 벡터의 계열(특징 벡터 계열)을 일시 기억한다. 또한, 미등록어 구간 처리부(76)는 매칭부(72)로부터 미등록어의 음성 구간과 음운 계열을 수신하면, 그 음성 구간에서의 음성의 특징 벡터 계열을, 일시 기억하고 있는 특징 벡터 계열로부터 검출한다. 그리고 미등록어 구간 처리부(76)는 매칭부(72)로부터의 음운 계열(미등록어)에 고유한 ID(identification)를 붙이고, 미등록어의 음운 계열과, 그 음성 구간에서의 특징 벡터 계열과 함께, 특징 벡터 버퍼(77)에 공급한다.
- [0137] 특징 벡터 버퍼(77)는, 예를 들면 도 18에 도시한 바와 같이, 미등록어 구간 처리부(76)로부터 공급되는 미등록어의 ID, 음운 계열 및 특징 벡터 계열을 대응하여 일시 기억한다.
- [0138] 여기서 도 18에서는, 미등록어에 대하여 1부터의 순차적인 숫자가 ID로서 첨부되어 있다. 따라서, 예를 들면 현재, 특징 벡터 버퍼(77)에 있어서, N개의 미등록어의 ID, 음운 계열 및 특징 벡터 계열이 기억되어 있는 경우에, 매칭부(72)가 미등록어의 음성 구간과 음운 계열을 검출하면, 미등록어 구간 처리부(76)에서는 그 미등록어에 대하여 N+1이 ID로서 첨부되고, 특징 벡터 버퍼(77)에서는 도 18에 점선으로 나타낸 바와 같이, 그 미등록어의 ID, 음운 계열 및 특징 벡터 계열이 기억된다.
- [0139] 다시 도 15로 되돌아가, 클러스터링부(78)는 특징 벡터 버퍼(77)에 새롭게 기억된 미등록어(이하, 적절하게, 신미등록어라 함)에 대하여 특징 벡터 버퍼(77)에 이미 기억되어 있는 다른 미등록어(이하, 적절하게, 기기억 미등록어라 함) 각각에 대한 스코어를 계산한다.

- [0140] 즉 클러스터링부(78)는 신미등록어를 입력 음성으로 하고, 또한 기기억 미등록어를 단어 사전에 등록되어 있는 단어라고 간주하여, 매칭부(72)에서의 경우와 마찬가지로 하여, 신미등록어에 대하여 각 기기억 미등록어에 대한 스코어를 계산한다. 구체적으로는, 클러스터링부(78)는 특징 벡터 버퍼(77)를 참조함으로써 신미등록어의 특징 벡터 계열을 인식함과 함께, 기기억 미등록어의 음운 계열에 따라서 음향 모델을 접속하고, 그 접속된 음향 모델로부터 신미등록어의 특징 벡터 계열이 관측된 우도로서의 스코어를 계산한다.
- [0141] 또, 음향 모델은 음향 모델 기억부(73)에 기억되어 있는 것이 이용된다.
- [0142] 클러스터링부(78)는 마찬가지로 하여, 각 기기억 미등록어에 대하여 신미등록어에 대한 스코어도 계산하고, 그 스코어에 따라서 스코어 시트 기억부(79)에 기억된 스코어 시트를 갱신한다.
- [0143] 또한 클러스터링부(78)는 갱신한 스코어 시트를 참조함으로써, 이미 구해진 미등록어(기기억 미등록어)를 클러스터링한 클러스터 중에서, 신미등록어를 새로운 멤버로서 더하는 클러스터를 검출한다. 또한 클러스터링부(78)는 신미등록어를 검출한 클러스터의 새로운 멤버로 하여, 그 클러스터를 그 클러스터의 멤버에 기초하여 분할하고, 그 분할 결과에 기초하여 스코어 시트 기억부(79)에 기억되어 있는 스코어 시트를 갱신한다.
- [0144] 스코어 시트 기억부(79)는 신미등록어에 대한 기기억 미등록어에 대한 스코어나 기기억 미등록어에 대한 신미등록어에 대한 스코어 등이 등록된 스코어 시트를 기억한다.
- [0145] 여기서, 도 19는 스코어 시트를 나타내고 있다.
- [0146] 스코어 시트는 미등록어의 「ID」, 「음운 계열」, 「클러스터 번호」, 「대표 멤버 ID」 및 「스코어」가 기술된 엔트리로 구성된다.
- [0147] 미등록어의 「ID」와 「음운 계열」로서는, 특징 벡터 버퍼(77)에 기억된 것과 동일한 것이 클러스터링부(78)에 의해 등록된다. 「클러스터 번호」는, 그 엔트리의 미등록어가 멤버로 되어 있는 클러스터를 특정하기 위한 숫자로, 클러스터링부(78)에 의해 첨부되고, 스코어 시트에 등록된다. 「대표번호 ID」는 그 엔트리의 미등록어가 멤버로 되어 있는 클러스터를 대표하는 대표 멤버로서의 미등록어의 ID로, 이 대표 멤버 ID에 의해 미등록어가 멤버로 되어 있는 클러스터의 대표 멤버를 인식할 수 있다. 또, 클러스터의 대표 멤버는 클러스터링부(29)에 의해서 구해지고, 그 대표 멤버의 ID가 스코어 시트의 대표 멤버 ID에 등록된다. 「스코어」는 그 엔트리의 미등록어에 대한 다른 미등록어 각각에 대한 스코어로, 상술한 바와 같이 클러스터링부(78)에 의해 계산된다.
- [0148] 예를 들면, 현재 특징 벡터 버퍼(77)에서, N개의 미등록어의 ID, 음운 계열 및 특징 벡터 계열이 기억되어 있다고 하면, 스코어 시트에는 그 N개의 미등록어의 ID, 음운 계열, 클러스터 번호, 대표번호 ID 및 스코어가 등록되어 있다.
- [0149] 그리고 특징 벡터 버퍼(77)에, 신미등록어의 ID, 음운 계열, 및 특징 벡터 계열이 새롭게 기억되면, 클러스터링부(78)에서는 스코어 시트가 도 19에서 점선으로 나타낸 바와 같이 갱신된다.
- [0150] 즉 스코어 시트에는 신미등록어의 ID, 음운 계열, 클러스터 번호, 대표 멤버 ID, 신미등록어에 대한 기기억 미등록어 각각에 대한 스코어(도 19에서의 스코어 $s(N+1, 1)$, $s(2, N+1)$, ..., $s(N+1, N)$)이 추가된다. 또한 스코어 시트에는 기기억 미등록어 각각에 대한 신미등록어에 대한 스코어(도 19에서의 $s(N+1, 1)$, $s(2, N+1)$, ..., $s(N+1, N)$)가 추가된다. 또한 후술하는 바와 같이, 스코어 시트에서의 미등록어의 클러스터 번호와 대표 멤버 ID가 필요에 따라 변경된다.
- [0151] 또, 도 19의 실시 형태에서는, ID가 i인 미등록어(의 발화)에 대한, ID가 j인 미등록어(의 음운 계열)에 대한 스코어를, $s(i, j)$ 로서 나타내고 있다.
- [0152] 또한, 스코어 시트(도 19)에는 ID가 i인 미등록어(의 발화)에 대한, ID가 j인 미등록어(의 음운 계열)에 대한 스코어 $s(i, j)$ 도 등록된다. 단, 이 스코어 $s(i, j)$ 는 매칭부(72)에 있어서, 미등록어의 음운 계열이 검출될 때에 계산되기 때문에, 클러스터링부(78)로 계산할 필요는 없다.
- [0153] 다시 도 15로 되돌아가, 메인터너스부(80)는 스코어 시트에 기억부(79)에서의 갱신 후의 스코어 시트에 기초하여 사전 기억부(74)에 기억된 단어 사전을 갱신한다.
- [0154] 여기서, 클러스터의 대표 멤버는, 다음과 같이 결정된다. 즉, 예를 들면 클러스터의 멤버로 되어 있는 미등록어 중, 다른 미등록어 각각에 대한 스코어의 총합(그 외에, 예를 들면 총합을 다른 미등록어의 수로 계산한 평균값이어도 된다)을 최대로 하는 것이 그 클러스터의 대표 멤버가 된다. 따라서, 이 경우, 클러스터에 속하는

멤버의 멤버 ID를 k 로 나타내는 것으로 하면, 다음식

수학식 1

$$K = \max_k \{ \sum s(k^3, k) \}$$

- [0155]
- [0156] 로 나타내는 값 $k (\in k)$ 를 ID로 하는 멤버가 대표 멤버가 된다.
- [0157] 단, 수학식 1에서, $\max_k \{ \}$ 는, $\{ \}$ 내의 값을 최대로 하는 k 를 의미한다. 또한 k^3 은, k 와 마찬가지로, 클러스터에 속하는 멤버의 ID를 의미한다. 또한, \sum 는 k^3 을 클러스터에 속하는 멤버 모든 ID에 걸쳐서 변화시킨 총합을 의미한다.
- [0158] 또 상술된 바와 같이 대표 멤버를 결정하는 경우, 클러스터의 멤버가 1 또는 2개의 미등록어일 때에는 대표 멤버를 결정함에 있어서 스코어를 계산할 필요는 없다. 즉 클러스터의 멤버가 1개의 미등록어인 경우에는 그 1개의 미등록어가 대표 멤버가 되고, 클러스터의 멤버가 2개의 미등록어인 경우에는 그 2개의 미등록어 중 어느 하나를 대표 멤버로 해도 된다.
- [0159] 또한 대표 멤버의 결정 방법은 상술한 것에 한정되는 것은 아니고, 그 외에, 예를 들면 클러스터의 멤버로 되어 있는 미등록어 중 다른 미등록어 각각과의 특징 벡터 공간에서의 거리의 총합을 최소로 하는 것 등을 그 클러스터의 대표 멤버로 하는 것도 가능하다.
- [0160] 이상과 같이 구성되는 음성 인식부(60)에서는 마이크로폰(51)에 입력된 음성을 인식하는 음성 인식 처리와, 미등록어에 관한 미등록어 처리가 도 20에 도시하는 음성 인식 처리 수순 RT2에 따라 행해진다.
- [0161] 실제 상, 음성 인식부(60)에서는 사람이 발화를 행함으로써 얻어진 음성 신호 S1B가 마이크로폰(51)으로부터 AD 변환부(70)를 통하여 음성 데이터가 되어 특징 추출부(71)에 제공되면 이 음성 인식 처리 수순 RT2가 단계 SP30에서 개시된다.
- [0162] 그리고 계속되는 단계 SP31에서, 특징 추출부(71)가 그 음성 데이터를 소정의 프레임 단위로 음향 분석함으로써 특징 벡터를 추출하고, 그 특징 벡터의 계열을 매칭부(72) 및 미등록어 구간 처리부(76)에 공급한다.
- [0163] 매칭부(76)는 계속되는 단계 S32에서, 특징 추출부(71)로부터의 특별 주문 오벡터 계열에 대하여, 상술한 바와 같이 스코어 계산을 행하고, 그 후 단계 S33에서, 스코어 계산의 결과 얻어지는 스코어에 기초하여 음성 인식 결과가 되는 단어 열의 표제를 구하여 출력한다.
- [0164] 또한 매칭부(72)는 계속되는 단계 S34에서, 사용자의 음성에 미등록어가 포함되어 있는지의 여부를 판정한다.
- [0165] 여기서, 이 단계 S34에서, 사용자의 음성에 미등록어가 포함되어 있지 않다고 판정된 경우, 즉 상술한 미등록어용 규칙 「\$pat1=\$color1 \$garbage \$color2 ;」가 적용되지 않고 음성 인식 결과가 얻어진 경우, 단계 S35로 진행하여 처리가 종료한다.
- [0166] 이에 대하여 단계 S34에서, 사용자의 음성에 미등록어가 포함되어 있다고 판정된 경우, 즉 미등록어용 규칙 「\$pat1=\$color1 \$garbage \$color2;」가 적용되어 음성 인식 결과가 얻어진 경우, 매칭부(72)는 계속되는 단계 S35에서, 미등록어용 규칙의 변수 \$garbage에 대응하는 음성 구간을 미등록어의 음성 구간으로서 검출함과 함께, 그 변수 \$garbage가 나타내는 가비지 모델에서의 음운의 천이로서의 음운 계열을 미등록어의 음운 계열로서 검출하고, 그 미등록어의 음성 구간과 음운 계열을 미등록어 구간 처리부(76)에 공급하여 처리를 종료한다 (단계 SP36).
- [0167] 한편, 미등록어 기관 처리부(76)는 특징 추출부(71)로부터 공급되는 특징 벡터 계열을 일시 기억하고 있고, 매칭부(72)로부터 미등록어의 음성 구간과 음운 계열이 공급되면, 그 음성 구간에서의 음성의 특징 벡터 계열을 검출한다. 또한 미등록어 구간 처리부(76)는 매칭부(72)로부터의 미등록어(의 음운 계열)에 ID를 첨부하고, 미등록어의 음운 계열과, 그 음성 구간에서의 특징 벡터 계열과 함께, 특징 벡터 버퍼(77)에 공급한다.
- [0168] 이상과 같이 하여, 특징 벡터 버퍼(77)에 새로운 미등록어(신미등록어)의 ID, 음운 계열 및 특징 벡터 계열이 기억되면, 이 후 미등록어의 처리가 도 21에 도시하는 미등록어 처리 수순 RT3에 따라 행해진다.
- [0169] 즉 음성 인식부(60)에 있어서는, 상술된 바와 같이 특징 벡터 버퍼(77)에 새로운 미등록어(신미등록어)의 ID,

음운 계열 및 특징 벡터 계열이 기억되면 이 미등록어 처리 수순 RT3이 단계 SP40에서 개시되고, 우선 최초로 단계 S41에서 클러스터링부(78)가 특징 벡터 버퍼(77)로부터 신미등록어의 ID와 음운 계열을 판독한다.

- [0170] 계속해서 단계 S42에서, 클러스터링부(78)는 스코어 시트 기억부(30)의 스코어 시트를 참조함으로써, 이미 구해진(생성되어 있는) 클러스터가 존재하는지의 여부를 판정한다.
- [0171] 그리고 이 단계 S42에서, 이미 구해진 클러스터가 존재하지 않는다고 판정된 경우, 즉 신미등록어가 처음의 미등록어이고, 스코어 시트에 기 기억 미등록어의 엔트리가 존재하지 않는 경우에는 단계 S43으로 진행하여, 클러스터링부(78)는 그 신미등록어를 대표 멤버로 하는 클러스터를 새롭게 생성하고, 그 새로운 클러스터에 관한 정보와, 신미등록어에 관한 정보를 스코어 시트 기억부(79)의 스코어 시트에 등록함으로써, 스코어 시트를 갱신한다.
- [0172] 즉 클러스터링부(78)는 특징 벡터 버퍼(77)로부터 판독한 신미등록어의 ID 및 음운 계열을 스코어 시트(도 19)에 등록한다. 또한 클러스터링부(78)는 고유한 클러스터 번호를 생성하고, 신미등록어의 클러스터 번호로서 스코어 시트에 등록한다. 또한 클러스터링부(78)는 신미등록어의 ID를 그 신미등록어의 대표번호 ID로서, 스코어 시트에 등록한다. 따라서, 이 경우에는 신미등록어는 새로운 클러스터의 대표 멤버가 된다.
- [0173] 또, 현재의 경우, 신미등록어와의 스코어를 계산하는 기 기억 미등록어가 존재하지 않기 때문에, 스코어의 계산은 행해지지 않는다.
- [0174] 이러한 단계 S43의 처리 후에는 단계 S52로 진행하여, 메인터너스부(80)는 단계 S43에서 갱신된 스코어 시트에 기초하여, 사전 기억부(74)의 단어 사전을 갱신하고, 처리를 종료한다(단계 SP54).
- [0175] 즉, 현재의 경우, 새로운 클러스터가 생성되어 있기 때문에, 메인터너스부(31)는 스코어 시트에서의 클러스터 번호를 참조하여, 그 새롭게 생성된 클러스터를 인식한다. 그리고 메인터너스부(80)는 그 클러스터에 대응하는 엔트리를 사전 기억부(74)의 단어 사전에 추가하고, 그 엔트리의 음운 계열로서, 새로운 클러스터의 대표 멤버의 음운 계열, 즉 현재의 경우에는, 신미등록어의 음운 계열을 등록한다.
- [0176] 한편, 단계 S42에서 이미 구해진 클러스터가 존재한다고 판정된 경우, 즉 신미등록어가 처음의 미등록어가 아니고, 따라서 스코어 시트(도 19)에 기 기억 미등록어의 엔트리(행)가 존재하는 경우, 단계 S44로 진행하여, 클러스터링부(78)는 신미등록어에 대하여, 각 기 기억 미등록어 각각에 대한 스코어를 계산함과 함께, 각 기 기억 미등록어 각각에 대하여 신미등록어에 대한 스코어를 계산한다.
- [0177] 즉, 예를 들면 현재, ID가 1 내지 N개의 기 기억 미등록어가 존재하고, 신미등록어의 ID를 N+1로 하면, 클러스터링부(78)에서는 도 19에서 점선으로 나타난 부분의 신미등록어에 대한 N개의 기 기억 미등록어 각각에 대한 스코어 $s(N+1, 1)$, $s(N+1, 2)$, ..., $s(N, N+1)$ 과, N개의 기 기억 미등록어 각각에 대한 신미등록어에 대한 스코어 $s(1, N+1)$, $s(2, N+1)$..., $s(N, N+1)$ 이 계산된다. 또 클러스터링부(78)에 있어서, 이들 스코어를 계산함에 있어서는 신미등록어와 N개의 기 기억 미등록어 각각의 특징 벡터 계열이 필요하게 되지만, 이들 특징 벡터 계열은 특징 벡터 버퍼(28)를 참조함으로써 인식된다.
- [0178] 그리고 클러스터링부(78)는 계산한 스코어를 신미등록어의 ID 및 음운 계열과 함께 스코어 시트(도 19)에 추가하고 단계 S45로 진행한다.
- [0179] 단계 S45에서는 클러스터링부(78)는 스코어 시트(도 19)를 참조함으로써, 신미등록어에 대한 스코어 $s(N+1, i)$ ($i=1, 2, \dots, N$)을 가장 높게(크게) 하는 대표 멤버를 갖는 클러스터를 검출한다. 즉, 클러스터링부(78)는 스코어 시트의 대표 멤버 ID를 참조함으로써, 대표 멤버로 되어 있는 기 기억 미등록어를 인식하고, 또한 스코어 시트의 스코어를 참조함으로써, 신미등록어에 대한 스코어를 가장 높게 하는 대표 멤버로서의 기 기억 미등록어를 검출한다. 그리고 클러스터링부(78)는 그 검출한 대표 멤버로서의 기 기억 미등록어의 클러스터 번호의 클러스터를 검출한다.
- [0180] 그 후, 단계 S46으로 진행하여 클러스터링부(29)는 신미등록어를 단계 S45에서 검출한 클러스터(이하, 적절하게 검출 클러스터라고 함)의 멤버에 추가한다. 즉 클러스터링부(78)는 스코어 시트에서의 신미등록어의 클러스터 번호로서 검출 클러스터의 대표 멤버의 클러스터 번호를 기입한다.
- [0181] 그리고, 클러스터링부(78)는 단계 S47에서 검출 클러스터를 예를 들면 2개의 클러스터로 분할하는 클러스터 분할 처리를 행하고 단계 S48로 진행한다. 단계 S48에서는 클러스터링부(78)는 단계 S47의 클러스터 분할 처리에 의해서, 검출 클러스터를 2개의 클러스터로 분할할 수 있는지의 여부를 판정하고, 분할할 수 있다는 판정을 한 경우, 단계 S49로 진행한다. 단계 S49에서는, 클러스터링부(78)는 검출 클러스터의 분할에 의해 얻어지는 2개

의 클러스터(이 2개의 클러스터를, 이하, 적절하게, 제1 자(子) 클러스터와 제2 자 클러스터라고 함)끼리의 사이의 클러스터 간 거리를 구한다.

[0182] 여기서, 제1 및 제2 자 클러스터끼리 사이의 클러스터 간 거리란, 예를 들면 다음과 같이 정의된다.

[0183] 즉 제1 자 클러스터와 제2 자 클러스터의 양쪽의 임의의 멤버(미등록어)의 ID를 k로 나타냄과 함께, 제1과 제2 자 클러스터의 대표 멤버(미등록어)의 ID를 각각 k1 또는 k2로 나타내는 것으로 하면, 다음식

수학식 2

[0184]
$$D(k1, k2) = \max_{k} \left\{ \text{abs}(\log(s(k, k1)) - \log(s(k, k2))) \right\}$$

[0185] 로 나타나는 값 D(k1, k2)를 제1과 제2 자 클러스터의 사이의 클러스터 간 거리로 한다.

[0186] 단, 수학식 2에서 abs()는 () 내의 값의 절대값을 나타낸다. 또한, $\max_{k}\{\}$ 는, k를 바꾸어서 구해지는 {} 내의 값의 최대값을 나타낸다. 또한 log는 자연 대수 또는 상용 대수를 나타낸다.

[0187] 현재, ID가 i인 멤버를 멤버 #I로 나타내는 것으로 하면, 수학식 2에서의 스코어의 역수 $1/s(k, k1)$ 은, 멤버 #k와 대표 멤버 k1과의 거리에 상당하고, 스코어의 역수 $1/s(k, k2)$ 는 멤버 #k와 대표 멤버 k2와의 거리에 상당한다. 따라서, 수학식 2에 따르면, 제1과 제2 자 클러스터의 멤버 중, 제1 자 클러스터의 대표 멤버 #k1과의 거리와, 제2 자 클러스터의 대표 멤버 #k2와의 차의 최대값이 제1과 제2 자 클러스터끼리의 사이의 자 클러스터 간 거리로 된다.

[0188] 또 클러스터 간 거리는 상술한 것에 한정되지 않고, 그 외에, 예를 들면 제1 자 클러스터의 대표 멤버와, 제2 자 클러스터의 대표 멤버와의 DP 매칭을 행함으로써, 특징 벡터 공간에서의 거리의 적산치를 구하여, 그 거리의 적산치를 클러스터 간 거리로 하는 것도 가능하다.

[0189] 단계 S49의 처리 후에는, 단계 S50으로 진행하여, 클러스터링부(78)는 제1과 제2 자 클러스터끼리의 클러스터 간 거리가 소정의 임계값 ζ보다 큰지의 여부(혹은, 임계값 ζ 이상인지의 여부)를 판정한다.

[0190] 단계 S50에서, 클러스터 간 거리가 소정의 임계값 ζ보다 크다고 판정된 경우, 즉 검출 클러스터의 멤버로서의 복수의 미등록 후에는 그 음향적 특징으로부터, 2개의 클러스터에 클러스터링하여야 할 것이라고 생각되는 경우, 단계 S51로 진행하여, 클러스터링부(78)는 제1과 제2 자 클러스터를 스코어 시트 기억부(79)의 스코어 시트에 등록한다.

[0191] 즉 클러스터링부(78)는 제1과 제2 자 클러스터에 고유한 클러스터 번호를 할당하여, 검출 클러스터의 멤버 중, 제1 자 클러스터에 클러스터링된 것의 클러스터 번호를 제1 자 클러스터의 클러스터 번호로 함과 함께, 제2 자 클러스터에 클러스터링된 것의 클러스터 번호를 제2 자 클러스터의 클러스터 번호로 하도록, 스코어 시트를 갱신한다.

[0192] 또한 클러스터링부(78)는 제1 자 클러스터에 클러스터링된 멤버의 대표 멤버 ID를 제1 자 클러스터의 대표 멤버의 ID로 함과 함께, 제2 자 클러스터에 클러스터링된 멤버의 대표 멤버 ID를 제2 자 클러스터의 대표 멤버의 ID로 하도록, 스코어 시트를 갱신한다.

[0193] 또, 제1과 제2 자 클러스터 중 어느 한쪽에는 검출 클러스터의 클러스터 번호를 할당하도록 하는 것이 가능하다.

[0194] 클러스터링부(78)가 이상과 같이 하여 제1과 제2 자 클러스터를 스코어 시트에 등록하면, 단계 S51 내지 S52로 진행하여, 메인턴스부(80)는 스코어 시트에 기초하여, 사전 기억부(74)의 단어 사전을 갱신하고, 처리를 종료한다(단계 SP54).

[0195] 즉, 현재의 경우, 검출 클러스터가 제1과 제2 자 클러스터로 분할되었기 때문에, 메인턴스부(80)는 우선 단어 사전에서의 검출 클러스터에 대응하는 엔트리를 삭제한다. 또한 메인턴스부(80)는 제1과 제2 자 클러스터 각각에 대응하는 2개의 엔트리를 단어 사전에 추가하고, 제1 자 클러스터에 대응하는 엔트리의 음운 계열로서, 그 제1 자 클러스터의 대표 멤버의 음운 계열을 등록함과 함께, 제2 자 클러스터에 대응하는 엔트리의 음운 계열로서, 그 제2 자 클러스터의 대표 멤버의 음운 계열을 등록한다.

[0196] 한편, 단계 S48에서, 단계 S47의 클러스터 분할 처리에 의해, 검출 클러스터를 2개의 클러스터로 분할할 수 없

었다고 판정된 경우, 또는 단계 S50에서, 제1과 제2 자 클러스터의 클러스터 간 거리가 소정의 임계값 ζ 보다 크지 않다고 판정된 경우, 따라서 검출 클러스터의 멤버로서의 복수의 미등록 후의 음향적 특징이 제1과 제2 자 클러스터에 클러스터링할 정도로 유사하지 않은 것이 아닌 경우), 단계 S53으로 진행하여, 클러스터링부(78)는 검출 클러스터의 새로운 대표 멤버를 구하여 스코어 시트를 갱신한다.

[0197] 즉 클러스터링부(78)는 신미등록 후를 멤버로서 더한 검출 클러스터의 각 멤버에 대하여, 스코어 시트 기억부(79)의 스코어 시트를 참조함으로써, 수학적 식 1의 계산에 필요한 스코어 $s(k^3, k)$ 를 인식한다. 또한, 클러스터링(78)은 그 인식한 스코어 $s(k^3, k)$ 를 이용하여, 수학적 식 1에 기초하여 검출 클러스터의 새로운 대표 멤버가 되는 멤버의 ID를 구한다. 그리고 클러스터링부(78)는, 스코어 시트(도 19)에서의 검출 클러스터의 각 멤버의 대표 멤버 ID를, 검출 클러스터의 새로운 대표 멤버의 ID에 재기입한다.

[0198] 그 후, 단계 S52로 진행하여, 메인터너스부(80)는 스코어 시트에 기초하여 사전 기억부(74)의 단어 사전을 갱신하고, 처리를 종료한다(단계 SP54).

[0199] 즉, 현재의 경우, 메인터너스부(80)는 스코어 시트를 참조함으로써, 검출 클러스터의 새로운 대표 멤버를 인식하고 또한 그 대표 멤버의 음운 계열을 인식한다. 그리고 메인터너스부(80)는 단어 사전에서의 검출 클러스터에 대응하는 엔트리의 음운 계열을 검출 클러스터의 새로운 대표 멤버의 음운 계열로 변경한다.

[0200] 여기서, 도 21의 단계 SP47의 클러스터 분할 처리는 도 22에 도시하는 클러스터 분할 처리 수순 RT4에 따라 행해진다.

[0201] 즉 음성 인식 처리부(60)에서는, 도 22의 단계 SP46으로부터 단계 SP47로 진행하면 이 클러스터 분할 처리 수순 RT4를 단계 SP60에서 개시하고, 우선 최초로 단계 S61에서 클러스터링부(78)가 신미등록 후가 멤버로서 더해진 검출 클러스터로부터, 아직 선택되지 않은 임의의 2개의 멤버의 조합을 선택하여, 각각을 가상의 대표 멤버로 한다. 여기서, 이 2개의 가상의 대표 멤버를, 이하 적절하게 제1 가대표 멤버와 제2 가대표 멤버라고 한다.

[0202] 그리고, 계속되는 단계 S62에서 클러스터링부(78)는 제1 가대표 멤버 및 제2 가대표 멤버를 각각 대표 멤버로 할 수 있도록 검출 클러스터의 멤버를 2개의 클러스터로 분할할 수 있는지의 여부를 판정한다.

[0203] 여기서, 제1 또는 제2 가대표 멤버를 대표 멤버로 할 수 있는지의 여부는 수학적 식 1의 계산을 행할 필요가 있지만, 이 계산에 이용되는 스코어 $s(k', k)$ 는 스코어 시트를 참조함으로써 인식된다.

[0204] 단계 S62에서, 제1 가대표 멤버 및 제2 가대표 멤버를 각각 대표 멤버로 할 수 있도록, 검출 클러스터의 멤버를 2개의 클러스터로 분할할 수 없다고 판정된 경우, 단계 S62를 스킵하여 단계 S64로 진행한다.

[0205] 또한, 단계 S62에서 제1 가대표 멤버와, 제2 가대표 멤버를 각각 대표 멤버로 할 수 있도록, 검출 클러스터의 멤버를 2개의 클러스터로 분할할 수 있다고 판정된 경우, 단계 S63으로 진행하여 클러스터링부(78)는 제1 가대표 멤버와, 제2 가대표 멤버가 각각 대표 멤버가 되도록, 검출 클러스터의 멤버를 2개의 클러스터로 분할하고, 그 분할 후의 2개의 클러스터의 조를, 검출 클러스터의 분할 결과가 되는 제1 및 제2 자 클러스터의 후보(이하, 적절하게, 후보 클러스터의 조라고 함)로 하고, 단계 S64로 진행한다.

[0206] 단계 S64에서는, 클러스터링부(78)는 검출 클러스터의 멤버 중에서, 아직 제1과 제2 가대표 멤버의 조로서 선택되지 않은 2개의 멤버의 조가 있는지의 여부를 판정하여, 있다고 판정된 경우, 단계 S61로 되돌아가고, 아직 제1과 제2 가대표 멤버의 조로서 선택되지 않은 검출 클러스터의 2개의 멤버의 조가 선택되어, 이하 마찬가지로의 처리가 반복된다.

[0207] 또한 단계 S64에서, 제1과 제2 가대표 멤버의 조로서 선택하지 않은 검출 클러스터의 2개의 멤버의 조가 없다고 판정된 경우, 단계 S65로 진행하여, 클러스터링부(78)는 후보 클러스터의 조가 존재하는지의 여부를 판정한다.

[0208] 단계 S65에서 후보 클러스터의 조가 존재하지 않는다고 판정된 경우, 단계 S66을 스킵하여 리턴한다. 이 경우에는 도 21의 단계 S48에서, 검출 클러스터를 분할할 수 없었다고 판정된다.

[0209] 한편, 단계 S65에서, 후보 클러스터의 조가 존재한다고 판정된 경우, 단계 S66으로 진행하여, 클러스터링부(78)는 후보 클러스터의 조가 복수 존재할 때에는 각 후보 클러스터의 조의 2개의 클러스터끼리의 사이의 클러스터 간 거리를 구한다. 그리고, 클러스터링부(78)는 클러스터 간 거리가 최소인 후보 클러스터의 조를 구하고, 그 후보 클러스터의 조를 검출 클러스터의 분할 결과로 하여, 즉 제1과 제2 자 클러스터로 하고 리턴한다. 또,

후보 클러스터의 조가 1개만인 경우에는 그 후보 클러스터의 조가 그대로 제1과 제2 자 클러스터가 된다.

- [0210] 이 경우에는 도 21의 단계 S48에서, 검출 클러스터를 분할할 수 있었다고 판정된다.
- [0211] 이상과 같이, 클러스터링부(78)에서, 이미 구해진 미등록어를 클러스터링한 클러스터 중에서, 신미등록어를 새로운 멤버로서 더하는 클러스터(검출 클러스터)를 검출하여, 신미등록어를 그 검출 클러스터의 새로운 멤버로 하고, 검출 클러스터를 그 검출 클러스터의 멤버에 기초하여 분할하도록 하였기 때문에, 미등록어를 그 음향적 특징이 근사하고 있는 것끼리 용이하게 클러스터링할 수 있다.
- [0212] 또한 메인터너스부(80)에서, 그와 같은 클러스터링 결과에 기초하여 단어 사전을 갱신하도록 하였기 때문에, 단어 사전의 대규모화를 피하면서, 미등록어의 단어 사전으로의 등록을 용이하게 행할 수 있다.
- [0213] 또한, 예를 들면 가령 매칭부(72)에서, 미등록어의 음성 구간의 검출을 잘 못하였다고 해도, 그와 같은 미등록어는 검출 클러스터의 분할에 의해, 음성 구간이 정확하게 검출된 미등록어와는 별도의 클러스터에 클러스터링된다. 그리고, 이러한 클러스터에 대응하는 엔트리가 단어 사전에 등록되게 되지만, 이 엔트리의 음운 계열은 정확하게 검출되지 않은 음성 구간에 대응하는 것이 되므로, 그 후의 음성 인식에서 큰 스코어를 주지는 않는다. 따라서, 가령, 미등록어의 음성 구간의 검출을 잘못하였다고 해도, 그 오류는 그 후의 음성 인식에는 거의 영향을 주지 않는다.
- [0214] 여기서, 도 23은 미등록어의 발화를 행하여 얻어진 클러스터링 결과를 나타내고 있다. 또, 도 23에서는 각 엔트리(각 행)가 1개의 클러스터를 나타내고 있다. 또, 도 23의 좌측 란은 각 클러스터의 대표 멤버(미등록어)의 음운 계열을 나타내고 있고, 도 23의 우측 란은 각 클러스터의 멤버로 되어 있는 미등록어의 발화 내용과 수를 나타내고 있다.
- [0215] 즉 도 23에서, 예를 들면 제1행의 엔트리는, 미등록어 「風呂」의 1개의 발화만이 멤버로 되어 있는 클러스터를 나타내고 있고, 그 대표 멤버의 음운 계열은 「doroa:」(ドロア-)로 되어 있다. 또한, 예를 들면 제2행의 엔트리는 미등록어 「風呂」의 3개의 발화가 멤버로 되어 있는 클러스터를 나타내고 있고, 그 대표 멤버의 음운 계열은 「kuro」(クロ)로 되어 있다.
- [0216] 또한, 예를 들면 제7행의 엔트리는 미등록어 「本」의 4개의 발화가 멤버로 되어 있는 클러스터를 나타내고 있고, 그 대표 멤버의 음운 계열은, 「NhoNde:su」(ンホンテース)로 되어 있다. 또한, 예를 들면 제8행의 엔트리는 미등록어 「オレンジ」의 1개의 발화와, 미등록어 「本」의 19개의 발화가 멤버로 되어 있는 클러스터를 나타내고 있고, 그 대표 멤버의 음운 계열은 「ohoN」(オホン)으로 되어 있다. 다른 엔트리도 마찬가지로의 것을 나타내고 있다.
- [0217] 도 23에 따르면, 동일한 미등록어의 발화에 대하여, 양호하게 클러스터링되어 있는 것을 알 수 있다.
- [0218] 또, 도 23의 제8행의 엔트리에 있어서는, 미등록어 「オレンジ」의 1개의 발화와, 미등록어 「本」의 19개의 발화가 동일한 클러스터에 클러스터링되어 있다. 이 클러스터는 그 멤버로 되어 있는 발화로부터, 미등록어 「本」의 클러스터가 되어야 한다고 생각되지만, 미등록어 「オレンジ」의 발화도, 그 클러스터의 멤버로 되어 있다. 그러나 이 클러스터도, 그 후에 미등록어 「本」의 발화가 더 입력되면, 클러스터 분할되어, 미등록어 「本」의 발화만을 멤버로 하는 클러스터와, 미등록어 「オレンジ」의 발화만을 멤버로 하는 클러스터에 클러스터링된다고 생각된다.
- [0219] (4-2) 얼굴 인식부(62)의 구체적 구성
- [0220] 다음에, 얼굴 인식부(62)의 구체적 구성에 대하여 설명한다.
- [0221] 도 24 및 도 25에 도시한 바와 같이, 얼굴 인식부(62)는 동적으로 변화하는 환경 하에서 일정 시간 내에 응답할 수 있지만, CCD 카메라(50)(도 5)로부터 제공되는 화상 신호 S1A에 기초한 화상 내로부터 얼굴 패턴을 추출하는 얼굴 추출 처리부(90)와, 추출된 얼굴 패턴을 기초로 얼굴을 인식하는 얼굴 인식 처리부(91)로 구성된다. 본 실시 형태에서는, 얼굴 패턴을 추출하는 얼굴 추출 처리에 「가보어 필터링(Gabor Filtering)」을 채용하고, 또한 얼굴 패턴으로부터 얼굴을 인식하는 얼굴 인식 처리에는 「서포트 벡터 머신(Support Vector Machine: SVM)」을 채용하고 있다.
- [0222] 이 얼굴 인식부(62)는 얼굴 패턴을 얼굴 인식 처리부(91)가 학습하는 학습 단계와, 학습된 데이터를 기초로, 화상 신호 S1A로부터 추출된 얼굴 패턴을 인식하는 인식 단계를 갖는다.
- [0223] 도 24에는 얼굴 인식부(62)의 학습 단계의 구조를, 또한 도 25에는 얼굴 인식부(62)의 인식 단계의 구성을 각각

나타내고 있다.

- [0224] 학습 단계에서는 도 24에 도시한 바와 같이, CCD 카메라(50)(도 5)로부터 입력된 사용자의 촬상 화상을 가보어 필터로 이루어지는 얼굴 추출 처리부(90)로 얼굴 추출한 결과가 서포트 벡터 머신으로 이루어지는 얼굴 인식 처리부(91)에 투입된다. 얼굴 인식 처리부(91)에서는 외부로부터 공급되는 학습용 데이터 즉 교사 데이터를 이용하여 잠정적인 식별 함수를 얻는다.
- [0225] 또한, 식별 단계에서는 도 25에 도시한 바와 같이, CCD 카메라(50)로부터 공급되는 화상 신호 S1A에 기초한 화상 내의 사람의 얼굴을 얼굴 추출 처리부(90)에서 얼굴 추출한 결과가 얼굴 인식 처리부(91)에 투입된다. 얼굴 인식 처리부(91)에서는 잠정적으로 얻어진 식별 함수를 여러가지의 데이터베이스 상의 화상에 실시하여 얼굴의 검출을 행한다. 그리고, 검출에 성공한 것을 얼굴 데이터로서 출력한다. 또한 검출에 실패한 것을 비얼굴 데이터로서 학습 데이터에 추가하여, 또한 다시 학습한다.
- [0226] 이하, 얼굴 추출 처리부(90)에서의 가보어 필터링 처리와, 얼굴 인식 처리부(91)에서의 서포트 벡터 머신에 대하여, 각각 상세히 설명한다.
- [0227] (4-2-1) 가보어 필터링 처리
- [0228] 사람의 시각 세포에는, 어떤 특정한 방위에 대하여 선택성을 갖는 세포가 존재한다는 것을 이미 알고 있다. 이것은 수직의 선에 대하여 반응하는 세포와, 수평의 선에 반응하는 세포로 구성된다. 가보어 필터링은 이것과 마찬가지로, 방위 선택성을 갖는 복수의 필터로 구성되는 공간 필터이다.
- [0229] 가보어 필터는 가보어 함수에 의해 공간 표현된다. 가보어 함수 $g(x, y)$ 는, 다음식

수학식 3

[0230]
$$g(x, y) = s(x, y)w_r(x, y)$$

[0231] 에 나타낸 바와 같이, 코사인 성분으로 이루어지는 캐리어 $s(x, y)$ 와, 2차원가우스 분석형 엔벨로프 $w_r(x, y)$ 로 구성된다.

[0232] 캐리어 $s(x, y)$ 는 복수 함수를 이용하여, 아래 수학식 4와 같이 표현된다. 여기서, 좌표치(u_0, v_0)는 공간 주파수를 나타내고, 또한 P는 코사인 성분의 위상을 나타낸다.

[0233] 여기서, 다음식

수학식 4

[0234]
$$s(x, y) = \exp(j(2\pi(u_0x + v_0y) + P))$$

[0235] 로 나타내는 캐리어는, 다음식

수학식 5

$$\text{Re}(s(x, y)) = \cos(2\pi(u_0x + v_0y) + P).$$

[0236]
$$\text{Im}(s(x, y)) = \sin(2\pi(u_0x + v_0y) + P)$$

[0237] 로 나타낸 바와 같이, 실수 성분 $\text{Re}(s(x, y))$ 와 허수 성분 $\text{Im}(s(x, y))$ 로 분리할 수 있다.

[0238] 한편, 2차원 가우스 분포로 이루어지는 엔벨로프는, 다음식

수학식 6

[0239]
$$w_r(x, y) = K \exp(-\pi(a^2(x-x_0)^2 + b^2(y-y_0)^2))$$

[0240] 와 같이 표현된다.

[0241] 여기서, 좌표 축(x_0, y_0)은 이 함수의 피크이고, 상수 a 및 b는 가우스 분포의 스케일 파라미터이다. 또한, 첨

자 r은 다음식

수학식 7

$$(x-x_0) \ , \ = (x-x_0)\cos\theta + (y-y_0)\sin\theta$$

$$(y-y_0) \ , \ = -(x-x_0)\sin\theta + (y-y_0)\cos\theta$$

[0242]

로 나타낸 바와 같은 회전 조작을 의미한다.

[0243]

따라서, 상술한 수학식 4 및 수학식 6에 의해, 가보어 필터는 다음식

[0244]

수학식 8

$$g(x, y)=K \exp(-\pi(a^2(x-x_0) \ , \ ^2 + b^2(y-y_0) \ , \ ^2))$$

$$\exp(j(2\pi(u_0x+u_0y)+P))$$

[0245]

로 나타낸 바와 같은 공간 함수로서 표현된다.

[0246]

본 실시 형태에 따른 얼굴 추출 처리부(90)는, 8 종류의 방향과 3가지의 주파수를 채용하여, 합계 24개의 가보어 필터를 이용하여 얼굴 추출 처리를 행한다.

[0247]

가보어 필터의 응답은, G_i 를 i 번째의 가보어 필터로 하고, i 번째의 가보어의 결과(Gabor Jet)를 J_i 로 하고, 입력 이미지를 I 로 하고, 그렇게 하면, 다음식

[0248]

수학식 9

$$J \ , \ (x, y)=G \ , \ (x, y) \oplus I(x, y)$$

[0249]

로 나타낸다. 이 수학식 9의 연산은 실제로는 고속 푸리에 변환을 이용하여 고속화할 수 있다.

[0250]

작성한 가보어 필터의 성능을 조사하기 위해서는, 필터링하여 얻어진 화소를 재구축함으로써 행한다. 재구축된 이미지 H 는, 다음식

[0251]

수학식 10

$$H(x, y)=\sum_{i=1}^n a \ , \ J \ , \ (x, y)$$

[0252]

과 같이 표시된다.

[0253]

그리고, 입력 화상 I 와 재구축된 화상 H 와의 에러 E 는 다음식

[0254]

수학식 11

$$E=\frac{1}{2} \ ||I(x, y)-H(x, y)||^2=\frac{1}{2} \ \sum_{x, y} (I(x, y)-H(x, y))^2$$

[0255]

과 같이 표시된다.

[0256]

이 에러 E 를 최소로 하는 최적의 a 를 구함으로써 재구축할 수 있다.

[0257]

(4-2-2) 서포트 벡터 머신

[0258]

본 실시 형태에서는 얼굴 인식 처리부(91)에서의 얼굴 인식에 관하여, 패턴인식 분야에서 가장 학습 범화(汎化) 능력이 높다고 하는 서포트 벡터 머신(SVM)을 이용하여 해당하는 얼굴인지의 여부의 식별을 행한다.

[0259]

SVM 자체에 관해서는, 예를 들면 B.sholkopf 외 저술의 보고(B.Sholkopf, C.Burges, A.Smola, "Advance in Kernel Support Vector Learning", The MIT Press, 1999.)를 예로 들 수 있다. 본원 출원인이 행한 예비 실험의 결과로부터는 SVM에 의한 얼굴 인식 방법은, 주성분 분석(PCA)이나 뉴럴 네트워크를 이용하는 방법에

[0260]

비하여, 양호한 결과를 나타내는 것을 알 수 있다.

[0261] SVM은 식별 함수에 선형 식별기(Perceptron)를 이용한 학습 기계로, 커널 함수를 사용하는 것으로 비선형 공간으로 확장할 수 있다. 또 식별 함수의 학습에서는, 클래스 간 분리의 마진을 최대로 취하도록 행해지고, 그 해는 2차 수리 계획법을 푸는 것으로 얻어지기 때문에, 글로벌 해에 도달할 수 있는 것을 이론적으로 보증할 수 있다.

[0262] 통상, 패턴 인식의 문제는 테스트 샘플 $x=(x_1, x_2, \dots, x_n)$ 에 대하여, 다음식

수학식 12

$$f(x) = \sum_{j=1}^n w_j x_j + b$$

[0263]

으로 제공되는 식별 함수 $f(x)$ 를 구하는 것이다.

[0264]

여기서, SVM의 학습용 교사 라벨을 다음식

[0265]

수학식 13

$$y=(y_1, y_2, \dots, y_n)$$

[0266]

과 같이 둔다.

[0267]

그렇게 하면, SVM에서의 얼굴 패턴의 인식을 다음식

[0268]

수학식 14

$$y_i (w^T x_i + b) z_i$$

[0269]

에 나타내는 제약 조건 하에서의 웨이팅 인자 w 의 제곱을 최소화하는 문제로서 파악할 수 있다.

[0270]

이러한 제약이 따르는 문제는, 라그랑주의 미정 상수법을 이용하여 풀 수 있다. 즉, 다음식

[0271]

수학식 15

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_i a_i (y_i ((x_i^T w + b) - 1))$$

[0272]

에 나타내는 라그랑주를 우선 도입하고, 계속해서 다음식

[0273]

수학식 16

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial w} = 0$$

[0274]

에 나타낸 바와 같이, b, w 각각에 대하여 편미분한다.

[0275]

그 결과, SVM에서의 얼굴 패턴의 식별을

[0276]

수학식 17

$$\max \sum a_i - \frac{1}{2} \sum a_i a_j y_i y_j x_i^T x_j$$

[0277]

$$\text{제약 조건 : } a_i \geq 0, \sum a_i y_i = 0$$

에 나타내는 2차 계획 문제로서 파악할 수 있다.

[0278]

[0279] 특정 공간의 차원수가 훈련 샘플의 수보다도 적은 경우에는 스크래치 변수 $\zeta \geq 0$ 를 도입하여, 제약 조건을 다음 식

수학식 18

[0280] $y_i (w' x_i + b) \geq 1 - \xi_i,$

[0281] 과 같이 변경한다.

[0282] 최적화에 대해서는, 다음식

수학식 19

[0283] $\frac{1}{2} ||w||^2 + C \sum \xi_i,$

[0284] 의 목적 함수를 최소화한다.

[0285] 이 수학식 19에 있어서, C는 제약 조건을 어디까지 완화하는지를 지정하는 계수로, 실험적으로 값을 결정할 필요 있다.

[0286] 라그랑주 상수 a에 관한 문제는 다음식

수학식 20

[0287] $\max \sum a_i - \frac{1}{2} \sum a_i a_i y_i y_i' x_j$

[0287] 제약 조건 : $0 \leq a_i \leq C, \sum a_i y_i = 0$

[0288] 과 같이 변경된다.

[0289] 그러나, 이 수학식 20 자체로는, 비선형의 문제를 풀 수 없다. 그래서, 본 실시 형태에서는, 커널 함수 $K(x, x')$ 을 도입하여, 일단 고차원의 공간으로 사상하여(커널 트릭), 그 공간에서 선형 분리하는 것으로 한다. 따라서, 원래의 공간에서는 비선형 분리하고 있는 것과 동등해진다.

[0290] 커널 함수는 임의의 사상 Φ 을 이용하여 다음식

수학식 21

[0291] $K(x, y) = \Phi(x)' \Phi(x')$

[0292] 과 같이 나타낸다.

[0293] 또한, 수학식 12에 도시한 식별 함수도, 다음식

수학식 22

[0294] $f(\Phi(x)) = w' \Phi(x) + b$
 $= \sum a_i y_i K(x, x_i) + b$

[0295] 과 같이 나타낼 수 있다.

[0296] 또한 학습에 관해서도, 다음식

수학식 23

$$\max \sum a_i - \frac{1}{2} \sum a_i a_j y_i y_j K(x_i, x_j)$$

제약 조건 : $0 \leq a_i \leq C, \sum a_i y_i = 0$

[0297]

[0298]

[0299]

에 나타내는 2차 계획 문제로서 파악할 수 있다.

커널로서는, 다음식

수학식 24

$$K(x, x') = \exp \left(- \frac{\|x - x'\|^2}{\sigma^2} \right)$$

[0300]

[0301]

[0302]

[0303]

[0304]

[0305]

[0306]

[0307]

[0308]

[0309]

[0310]

에 나타내는 가우시안 커널(RBF(Radius Basic Function)) 등을 이용할 수 있다.

또, 가보어 필터링에 관해서는 인식 태스크에 따라 필터의 종류를 변경하도록 해도 된다.

저주파에서의 필터링에서는 필터링 후의 이미지 전부를 벡터로서 가지고 있는 것은 과잉이다. 그래서, 다운 샘플링하여, 벡터의 차원을 낮추도록 해도 된다. 다운 샘플링된 24 종류의 벡터를 일렬로 나열하여, 긴 벡터로 한다.

또한 본 실시 형태에 있어서 얼굴 패턴의 인식에 적용되는 SVM은 특징 공간을 2분하는 식별기이므로, 「사람 A이다」인지 「사람 A가 아니다」인지를 판별하도록 학습한다. 그 때문에, 데이터베이스의 화상 중에서 우선 사람 A의 얼굴 화상을 모아서, 가보어 필터링 후의 벡터에 「사람 A가 아니다」라는 라벨을 붙인다. 일반적으로, 모으는 얼굴 화상의 수는 특징 공간의 차원보다 많은 쪽이 바람직하다. 10명의 얼굴을 인식하고자 하는 경우에는 마찬가지로, 「사람 B이다」, 「사람 B가 아니다」...와 같이, 각각의 사람에 대하여 1개의 식별기를 구성한다.

이러한 학습에 의해, 예를 들면 「사람 A이다」인지 「사람 A가 아니다」인지를 알 수 있는 서포트 벡터가 구해진다. SVM은 특징 공간을 2개로 구획하는 식별기로, 새로운 얼굴 화상이 입력되어 왔을 때에, 역시 가보어 필터링의 벡터가 구한 서포트 벡터가 구성하는 경계면 중 어느 쪽에 있는지로 인식 결과를 출력한다. 그리고, 경계에 대하여, 「사람 A이다」의 영역에 있으면 「사람 A이다」라고 인식할 수 있다. 또, 「사람 A가 아니다」의 영역이면 「사람 A가 아니다」라고 인식된다.

CCD 카메라(50)로부터의 화상 신호 S1A에 기초한 화상으로부터 얼굴의 부분으로서 절취되는 영역은 일정하지 않다. 이 때문에 특징 공간에서 인식하고자 하는 카테고리과 떨어져 있는 점에 투영될 가능성이 있다. 따라서, 눈이나 코, 입 등과 같은 특징을 갖는 파트로 추정하여 아핀 변환에 의해 모핑함으로써 인식율이 향상될 가능성이 있다.

또한 인식 성능을 높이기 위해서, 부트스트랩 방법을 채용할 수 있다. 학습에 이용하는 화상과는 별도로 화상을 촬영하여, 부트스트랩에 이용한다. 이것은 학습한 식별기가 잘못된 인식 결과를 출력했을 때에, 그 입력 화상을 학습 세트에 투입하여 다시 학습하는 것을 의미한다.

또한 인식 성능을 높이기 위해서, 인식 결과의 시간 변화를 보는 방법도 있다. 가장 간단한 방법으로는, 10회 중 8회를 「사람 A이다」라고 인식되면 「사람 A이다」라고 판단하는 것이다. 그 외에, 칼만 필터를 이용한 예측법 등도 제안되어 있다.

(5) 본 실시 형태의 동작 및 효과

이상의 구성에 있어서, 이 로봇(1)에서는, 신규의 사람과의 대화를 통하여 그 사람의 이름을 취득하고, 해당 이름을 마이크로폰(51)이나 CCD 카메라(50)의 출력에 기초하여 검출한 그 사람의 소리의 음향적 특징 및 얼굴의 형태적 특징의 각 데이터와 관련시켜 기억함과 함께, 이들 기억한 각종 데이터에 기초하여, 이름을 취득하지 않은 신규의 사람의 등장을 인식하고, 그 신규의 사람의 이름이나 소리의 음향적 특징 및 얼굴의 형태적 특징을 상술과 마찬가지로 하여 취득하고 기억하도록 하여, 사람의 이름을 학습한다.

- [0311] 따라서, 이 로봇(1)은 음성 커맨드의 입력이나 터치 센서의 가압 조작 등의 사용자로부터의 명시적인 지시에 의한 이름 등록을 필요로 하지 않고, 사람이 평소 행하는 것과 같이, 통상의 사람과의 대화를 통하여 신규의 인물이나 물체 등의 이름을 자연스럽게 학습할 수 있다.
- [0312] 이상의 구성에 따르면, 신규의 사람과의 대화를 통하여 그 사람의 이름을 취득하고, 해당 이름을 마이크로폰(51)이나 CCD 카메라(50)의 출력에 기초하여 검출한 그 사람의 소리의 음향적 특징 및 얼굴의 형태적 특징의 각 데이터와 관련시켜 기억함과 함께, 이들 기억한 각 데이터에 기초하여, 이름을 취득하지 않은 또 다른 신규의 사람의 등장을 인식하고, 그 신규의 사람의 이름이나 소리의 음향적 특징 및 얼굴의 형태적 특징을 상술과 마찬가지로 하여 취득하고 기억하도록 하여, 사람의 이름을 학습하도록 함으로써, 통상의 사람과의 대화를 통하여 신규의 인물이나 물체 등의 이름을 자연스럽게 학습할 수 있도록 할 수 있어, 오락성을 특히 향상시킬 수 있는 로봇을 실현할 수 있다.
- [0313] (6) 다른 실시 형태
- [0314] 또 상술한 실시 형태에서는, 본 발명을 도 1과 같이 구성된 2족 보행형 로봇(1)에 적용하도록 한 경우에 대하여 설명하였지만, 본 발명은 이것에 한하지 않고, 이 외에 여러가지의 로봇 장치 및 로봇 장치 이외의 여러가지의 장치에 널리 적용할 수 있다.
- [0315] 또한 상술한 실시 형태에 있어서는, 사람과 대화하기 위한 기능을 갖고, 해당 대화를 통하여 대상으로 하는 물체의 이름을 사람으로부터 취득하는 대화 수단을, 음성 인식부(60), 대화 제어부(63) 및 음성 합성부(64)로 구성함으로써, 사람과의 음성 대화에 의해 사람의 이름을 취득하도록 한 경우에 대하여 설명하였지만, 본 발명은 이것에 한하지 않고, 예를 들면 키보드 입력 등에 의한 문자 대화에 의해 사람의 이름을 취득하도록 대화 수단을 구성하도록 해도 된다.
- [0316] 또한 상술한 실시 형태에서는, 이름 학습의 대상이 인물인 경우에 대하여 설명하였지만, 본 발명은 이것에 한하지 않고, 인물 대신에 또는 인물 외에 다른 여러가지의 물체를 이름 학습의 대상으로 하도록 해도 된다.
- [0317] 이 경우에, 상술한 실시 형태에서는 대상이 되는 인물의 소리의 음향적 특징 및 얼굴의 형태적 특징으로부터 그 인물을 각각 인식하고, 이들 인식 결과에 기초하여 그 인물이 신규의 인물인지의 여부를 판별하도록 한 경우에 대하여 설명하였지만, 본 발명은 이에 한하지 않고, 이 대신에 또는 이에 더하여, 이들 이외의 예를 들면 체형이나 냄새 등의 생물학적으로 개체를 식별 가능한 복수 종류의 각종 특징으로부터 그 인물을 각각 인식하고, 이들 인식 결과에 기초하여 그 인물이 신규의 사람인지의 여부를 판별하도록 해도 된다. 또한 이름 학습 대상이 인물 이외의 물체인 경우에는, 색이나 형상, 모양, 크기 등의 물체를 식별 가능한 복수 종류의 특징으로부터 각각 그 물체를 인식하고, 이들 인식 결과에 기초하여 그 물체가 신규의 물체인지의 여부를 판별하도록 해도 된다. 그리고 이들의 경우에는, 각각 물체의 다른 소정의 특징을 검출함과 함께, 해당 검출 결과 및 미리 기억하고 있는 기지의 물체의 대응하는 특징의 데이터에 기초하여, 해당 대상으로 하는 물체를 인식하는 복수의 인식 수단을 형성하도록 하면 된다.
- [0318] 또한 상술한 실시 형태에 있어서는, 기지의 물체의 이름 및 해당 물체에 대한 각 인식 수단(화자 인식부(61) 및 얼굴 인식부(62))의 인식 결과를 관련시킨 관련 정보를 기억하는 기억 수단을 메모리에 의해 구성하도록 한 경우에 대하여 설명하였지만, 본 발명은 이것에 한하지 않고, 정보를 기억할 수 있는 메모리 이외의 예를 들면 디스크형 기록 매체 등 이 외 여러가지의 기억 수단을 널리 적용할 수 있다.
- [0319] 또한 상술한 실시 형태에 있어서는, 화자 인식부(61) 및 얼굴 인식부(62)가 대상으로 하는 사람을 인식하는 인식 처리를 한번밖에 행하지 않도록 한 경우에 대하여 설명하였지만, 본 발명은 이것에 한하지 않고, 예를 들면 인식 불능(SID=-1)이던 경우에는 한번 더 인식 처리를 행하도록 하도록 해도 되고, 그 이외일 때라도 복수회의 인식 처리를 행하도록 해도 된다. 이와 같이 함으로써 인식 결과의 정밀도를 향상시킬 수 있다.
- [0320] 또한 상술한 실시 형태에 있어서는, 대화 제어부(63)가 복수의 인식 수단(음성 인식부(60), 화자 인식부(61), 얼굴 인식부(62))의 인식 결과의 다수결에 의해, 그 사람이 신규의 사람인지의 여부를 판단하도록 한 경우에 대하여 설명하였지만, 본 발명은 이에 한하지 않고, 다수결 이외의 방법에 의해 이들 복수의 인식 수단의 각 인식 결과에 기초하여 그 사람이 신규의 사람인지의 여부를 판단하도록 해도 된다.
- [0321] 이 경우에 있어서, 예를 들면 복수의 인식 수단의 각 인식 결과에 그 인식 수단의 인식 성능에 따라 웨이팅을 하고, 그 웨이팅한 각 인식 결과에 기초하여 대상으로 하는 물체가 신규의 것인지의 여부를 판단하는 방법이나, 가장 인식 성능이 높은 인식 수단과 다른 1개의 인식 수단의 인식 결과에 기초하여 신규의 사람이라고 판단할

수 있던 경우에는 다른 인식 수단의 인식 결과를 이용하지 않는 방법 등 여러가지의 방법을 널리 적용할 수 있다.

[0322] 또한 상술한 실시 형태에 있어서는, 화자 인식부(61)나 얼굴 인식부(62)가 대상으로 하는 사람을 정확하게 인식한 경우에 그 화자 인식부(61) 및 또는 얼굴 인식부(62)에 추가 학습시킴으로써, 통계적인 안정에 의해 인식 정밀도를 향상시키도록 한 경우에 대하여 설명하였지만, 본 발명은 이것에 한하지 않고, 메모리(65)에 저장되는 관련 정보에 대해서도, 마찬가지로 몇번이나 동일한 조합을 기억함으로써 그 관련 정보의 신뢰성을 향상시킬 수 있는 기능을 구비하도록 해도 된다. 구체적으로는, 이러한 기능의 구현화 방법으로서, 예를 들면 「일본 전자 정보 통신학회 논문지, D-II, Vol. J82-D-II, No6, pp.1072-1081」에 기재된 뉴럴 네트워크를 이용한 방법을 이용할 수 있다.

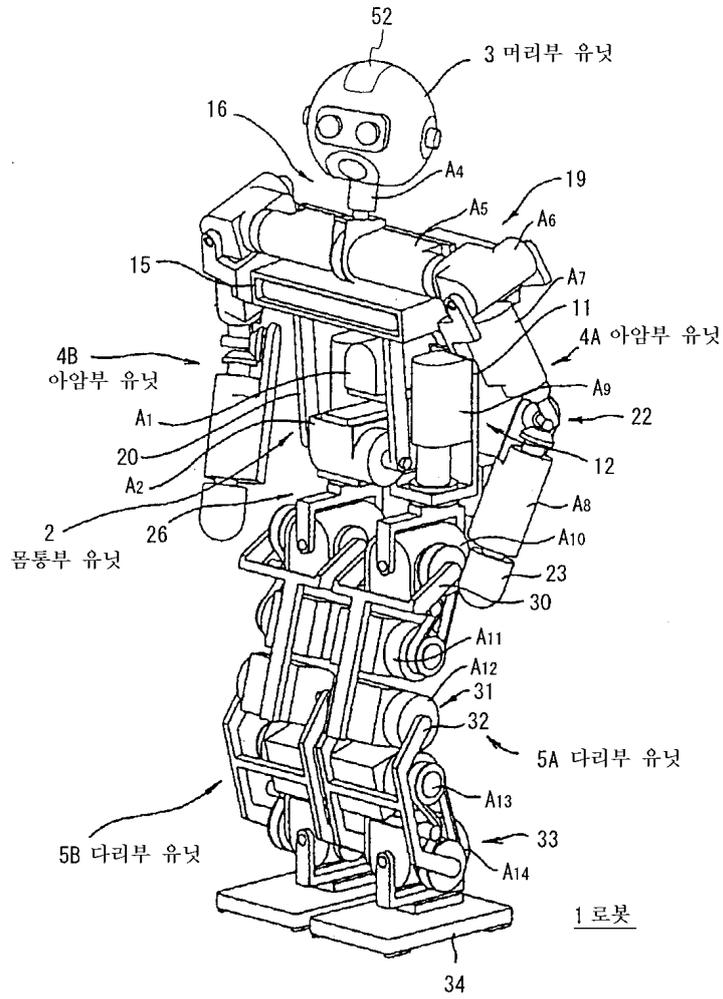
[0323] 이상과 같이 본 발명에 따르면, 학습 장치에서 사람과 대화하기 위한 기능을 갖고, 그 대화를 통하여 대상으로 하는 물체의 이름을 사람으로부터 취득하는 대화 수단과, 각각 대상으로 하는 물체의 다른 소정의 특징을 검출함과 함께, 그 검출 결과 및 미리 기억하고 있는 기지의 물체의 대응하는 특징의 데이터에 기초하여, 해당 대상으로 하는 물체를 인식하는 복수의 인식 수단과, 기지의 물체의 이름 및 그 물체에 대한 각 인식 수단의 인식 결과를 관련시킨 관련 정보를 기억하는 기억 수단과, 대화 수단이 취득한 대상으로 하는 물체의 이름, 그 대상으로 하는 물체에 대한 각 인식 수단의 인식 결과, 및 기억 수단이 기억하는 관련 정보에 기초하여, 대상으로 하는 물체가 신규의 물체인지의 여부를 판단하는 판단 수단과, 판단 수단이 대상으로 하는 물체를 신규의 물체라고 판단했을 때에, 그 대상으로 하는 물체의 대응하는 특징의 데이터를 각 인식 수단에 각각 기억시킴과 함께, 해당 대상으로 하는 물체에 대한 관련 정보를 기억 수단에 기억시키는 제어 수단을 형성하도록 함으로써, 사람이 평소 행하는 것과 같이, 통상의 사람과의 대화를 통하여 신규의 인물이나 물체 등의 이름을 자연스럽게 학습할 수 있어, 오락성을 특히 향상시킬 수 있는 학습 장치를 실현할 수 있다.

[0324] 또한 본 발명에 따르면, 학습 방법에 있어서, 사람과 대화하여, 그 대화를 통하여 대상으로 하는 물체의 이름을 사람으로부터 취득함과 함께, 대상으로 하는 물체의 복수의 다른 소정의 특징을 각각 검출하고, 그 검출 결과 및 미리 기억하고 있는 기지의 물체의 각 특징의 데이터에 기초하여, 해당 대상으로 하는 물체를 인식하는 제1 단계와, 취득한 대상으로 하는 물체의 이름과, 그 대상으로 하는 물체의 각 특징에 각각 기초하는 각 인식 결과와, 미리 기억하고 있는 기지의 물체의 이름 및 그 물체에 대한 각 인식 수단의 인식 결과를 관련시킨 관련 정보에 기초하여, 대상으로 하는 물체가 신규의 물체인지의 여부를 판단하는 제3 단계와, 대상으로 하는 물체를 신규의 물체라고 판단했을 때에, 그 대상으로 하는 물체의 각 특징의 데이터, 및 그 대상으로 하는 물체에 대한 관련 정보를 각각 기억하는 제4 단계를 포함하도록 함으로써, 사람이 평소 행하는 것과 같이, 통상의 사람과의 대화를 통하여 신규의 인물이나 물체 등의 이름을 자연스럽게 학습할 수 있어, 오락성을 특히 향상시킬 수 있는 학습 방법을 실현할 수 있다.

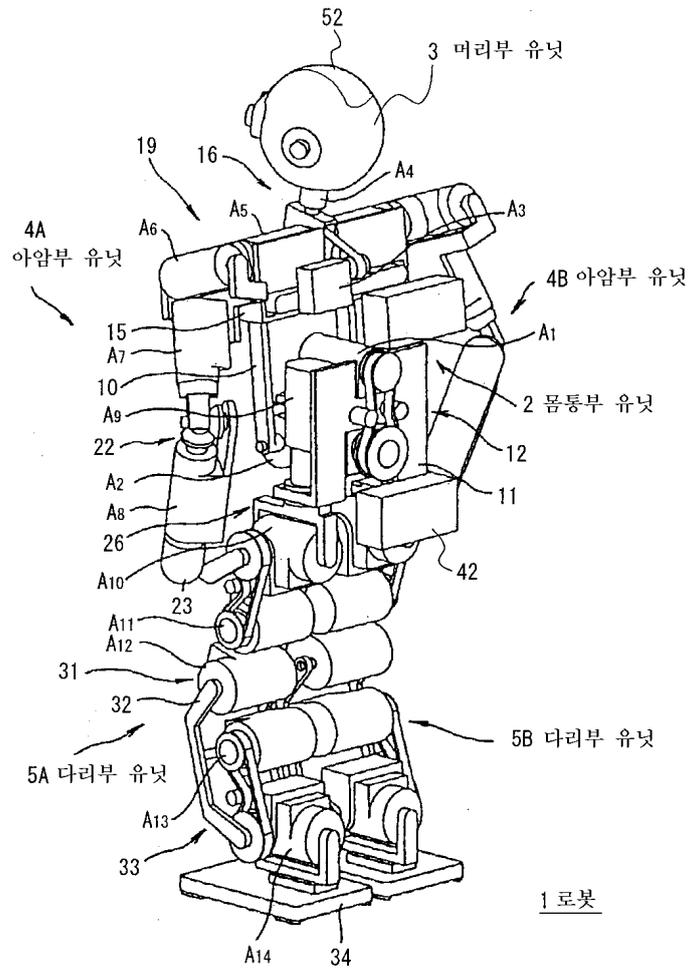
[0325] 또한 발명에 따르면, 로봇 장치에 있어서 사람과 대화하기 위한 기능을 갖고, 그 대화를 통하여 대상으로 하는 물체의 이름을 사람으로부터 취득하는 대화 수단과, 각각 대상으로 하는 물체의 다른 소정의 특징을 검출함과 함께, 그 검출 결과 및 미리 기억하고 있는 기지의 물체의 대응하는 특징의 데이터에 기초하여, 그 대상으로 하는 물체를 인식하는 복수의 인식 수단과, 기지의 물체의 이름 및 그 물체에 대한 각 인식 수단의 인식 결과를 관련시킨 관련 정보를 기억하는 기억 수단과, 대화 수단이 취득한 대상으로 하는 물체의 이름, 그 대상으로 하는 물체에 대한 각 인식 수단의 인식 결과, 및 기억 수단이 기억하는 관련 정보에 기초하여, 대상으로 하는 물체가 신규의 물체인지의 여부를 판단하는 판단 수단과, 판단 수단이 대상으로 하는 물체를 신규의 물체라고 판단했을 때에, 그 대상으로 하는 물체의 대응하는 특징의 데이터를 각 인식 수단에 각각 기억시킴과 함께, 그 대상으로 하는 물체에 대한 관련 정보를 기억 수단에 기억시키는 제어 수단을 포함하도록 함으로써, 사람이 평소 행하는 것과 같이, 통상의 사람과의 대화를 통하여 신규의 인물이나 물체 등의 이름을 자연스럽게 학습할 수 있어, 오락성을 특히 향상시킬 수 있는 로봇 장치를 실현할 수 있다.

도면

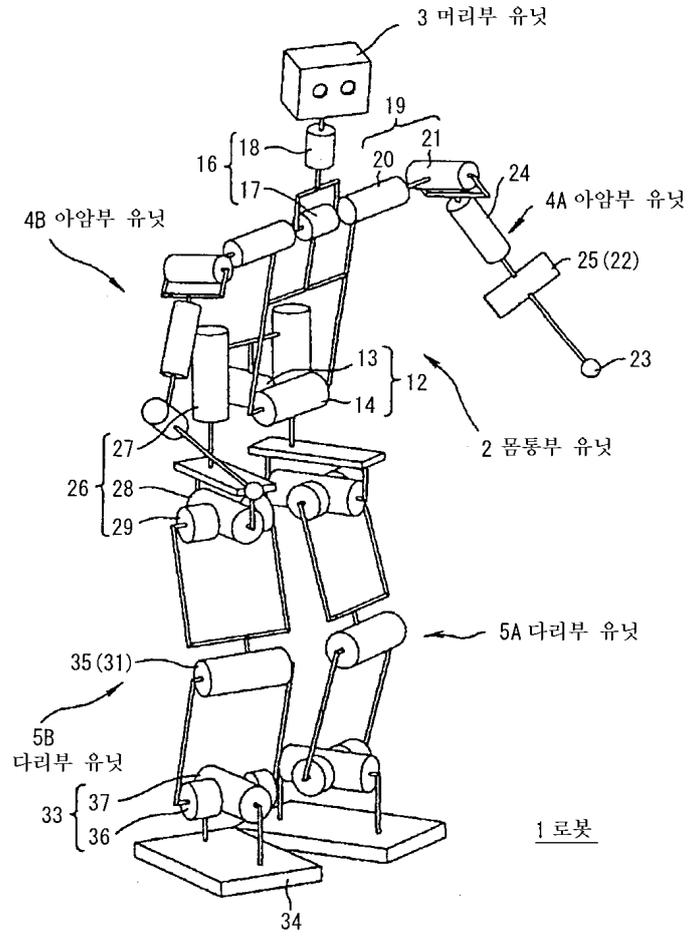
도면1



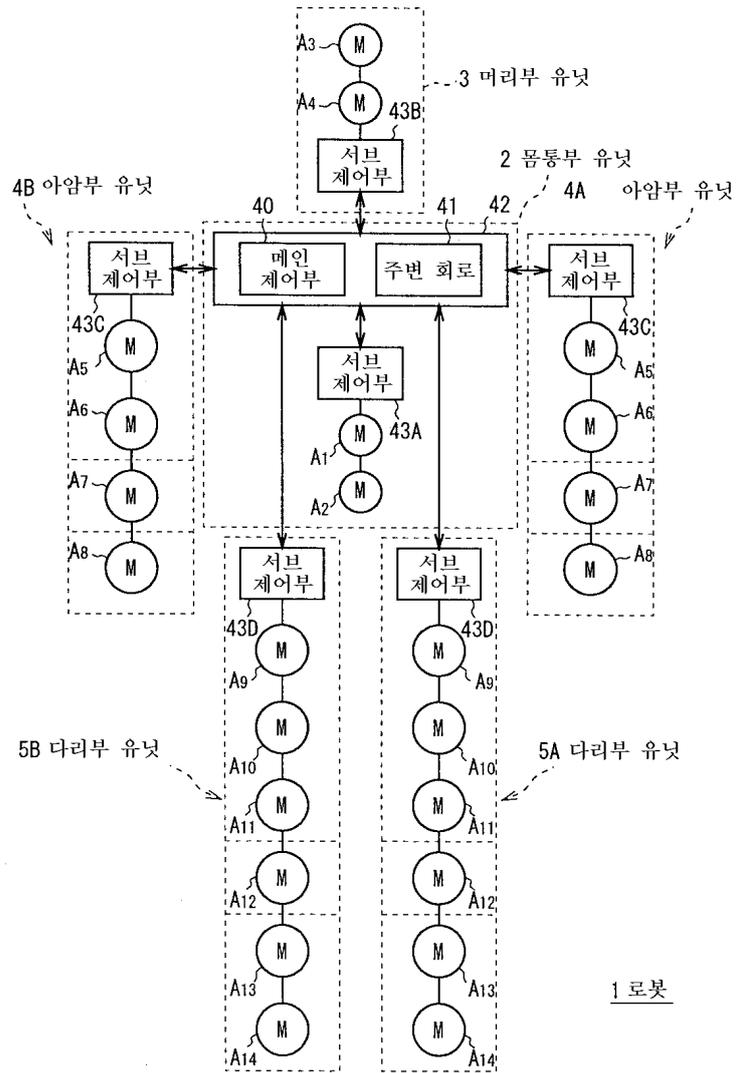
도면2



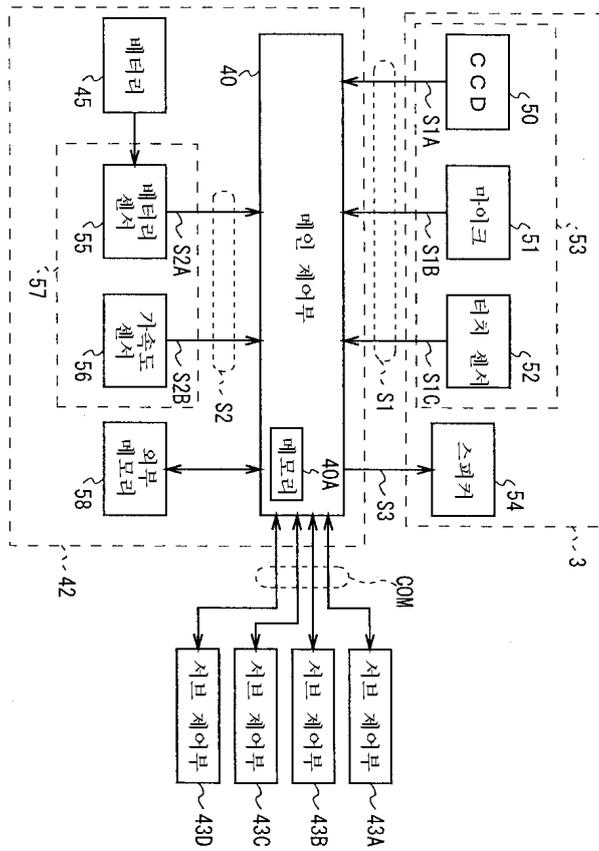
도면3



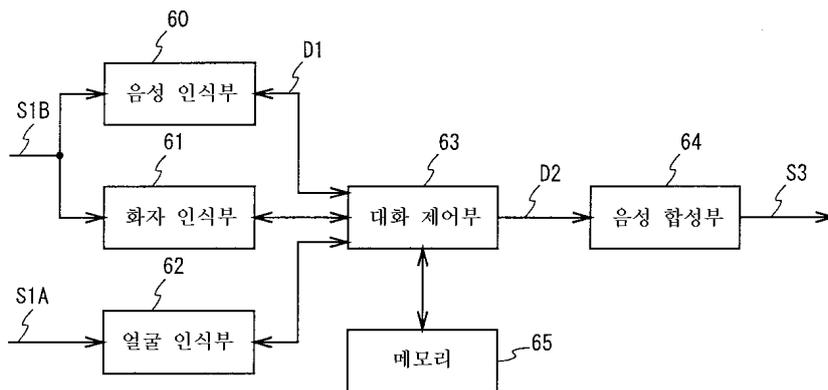
도면4



도면5



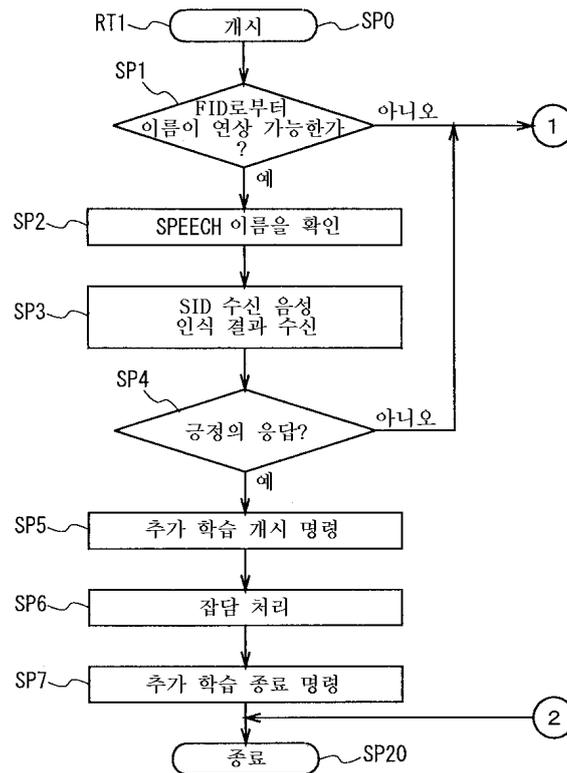
도면6



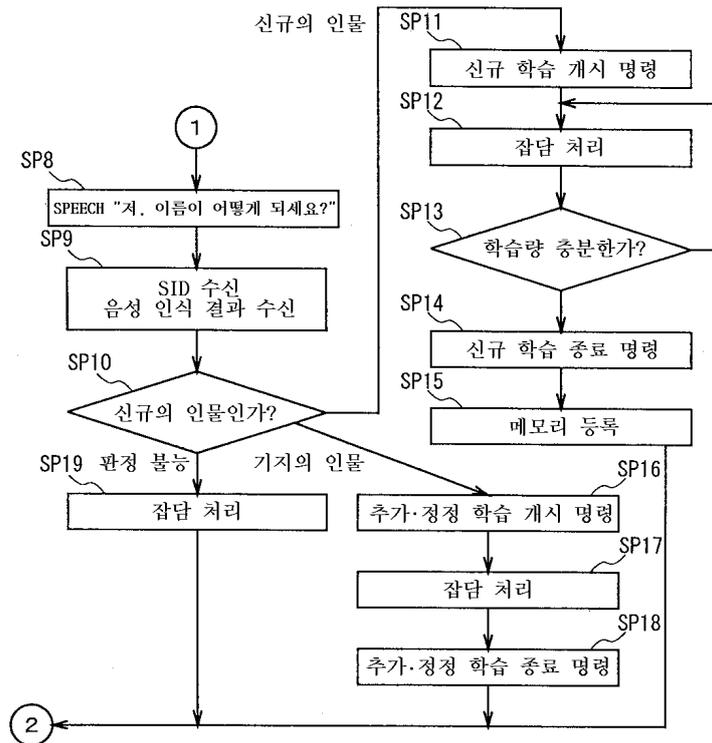
도면7

F I D	S I D	이름
1	2	후지타
2	5	요시다

도면8



도면9



도면10

R: 후지타씨군요.
 H: 예, 그렇습니다.
 R: 오늘 날씨가 좋군요.
 H: 그렇군요.
 .
 .
 .

도면11

(R: 후지타씨군요.
 H: 아니오, 틀렸습니다.)
 R: 저, 이름을 가르쳐 주세요.
 H: 야마모토입니다.
 R: 저는 로봇입니다. 잘 부탁드립니다.
 H: 잘 부탁드립니다.
 R: 야마모토씨, 오늘 날씨가 좋군요.
 H: 그렇군요.

도면12

F I D	S I D	이름
1	2	후지타
2	5	요시다
4	6	야마모토

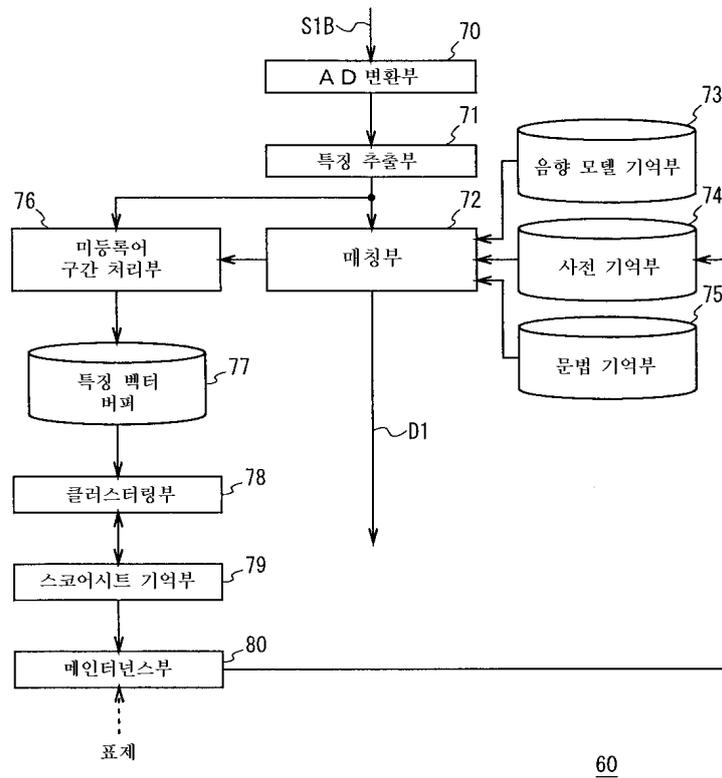
도면13

(R: 아, 후지타씨군요.
H: 아니요, 틀렸습니다.)
R: 저, 이름을 가르쳐 주세요.
H: 요시다입니다.
R: 아, 요시다씨군요. 생각났어요. 오늘 날씨가 좋군요.
H: 그렇군요.
R: 전에, 음... 언제 만났더라?
H: 음, 그저께였던것 같은데요...

도면14

(R: 아, 후지타씨군요.
H: 아니요, 틀렸습니다.)
R: 저, 이름을 가르쳐주세요.
H: 야마모토입니다.
R: 아 그렇습니까? 안녕하세요?
H: 예, 안녕하세요?

도면15



도면16

표제	음운계열	
boku	[僕]	boku
chigau	[違う]	chigau
doko	[どこ]	doko
genki	[元氣]	geNki
iro	[色]	iro
janai	[じゃない]	janai
kirai	[嫌い]	kirai
kudasai	[ください]	kudasai

도면17

```

$col = [kono | sono] iro wa;
$this = kore [(ga | wa | mo)];
$neg = (Chagau | iie) [$sil];
$null = $sil;
$des = (desu | da) [yo] | yo;
$not = janai [yo];
$color1 = $null | $neg | [$neg] $col | [$neg] $this;
$color2 = [iro] (desu | janai | da) [yo];
$pat1 = $color $garbage $color2;
    
```

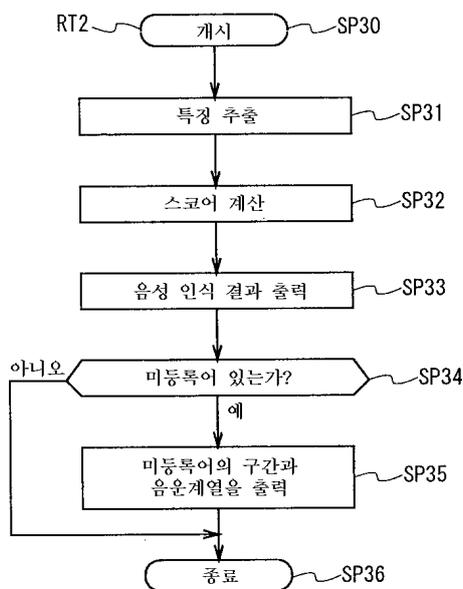
도면18

ID	음운계열	특징 벡터계열
1	---	---
2	---	---
⋮	⋮	⋮
⋮	⋮	⋮
N	---	---
N+1	---	---

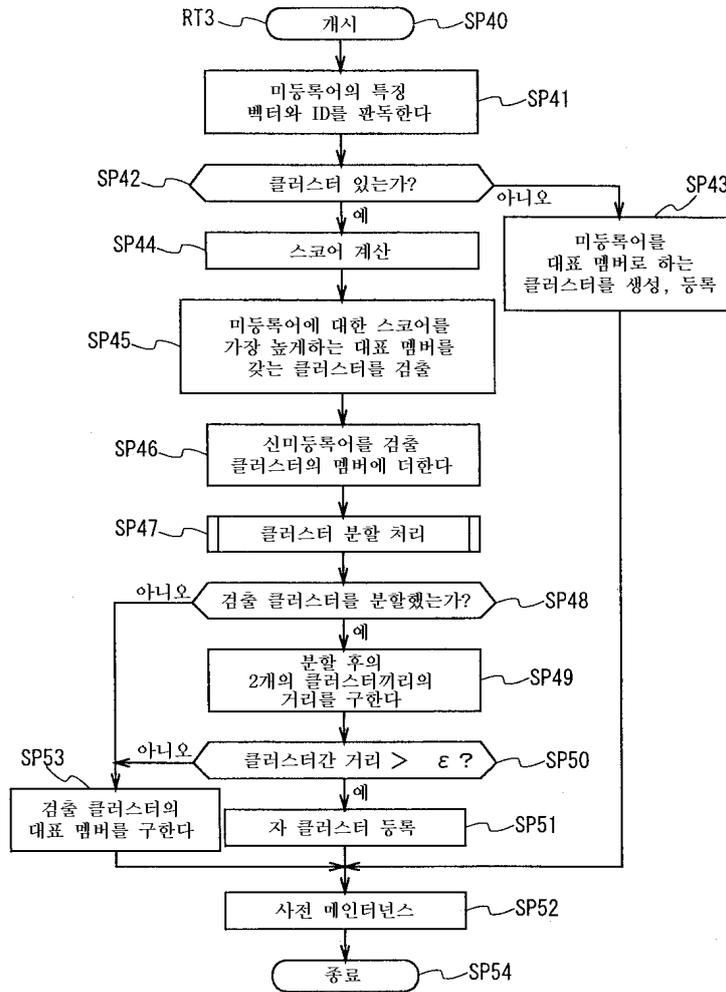
도면19

ID	음운계열	클러스터 넘버	대표 멤버ID	스코어 (거리)					N+1
				1	2	3	⋯	N	
1	⋯	1	1	s(1, 1)	s(1, 2)	s(1, 3)	⋯	s(1, N)	s(1, N+1)
2	⋯	2	2	s(2, 1)	s(2, 2)	s(2, 3)	⋯	s(2, N)	s(2, N+2)
3	⋯	1	1	s(3, 1)	s(3, 2)	s(3, 3)	⋯	s(3, N)	s(3, N+3)
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
N	⋯	1	1	s(N, 1)	s(N, 2)	s(N, 3)	⋯	s(N, N)	s(N, N+1)
N+1	⋯	2	2	s(N+1, 1)	s(N+1, 2)	s(N+1, 3)	⋯	s(N+1, N)	s(N+1, N+1)

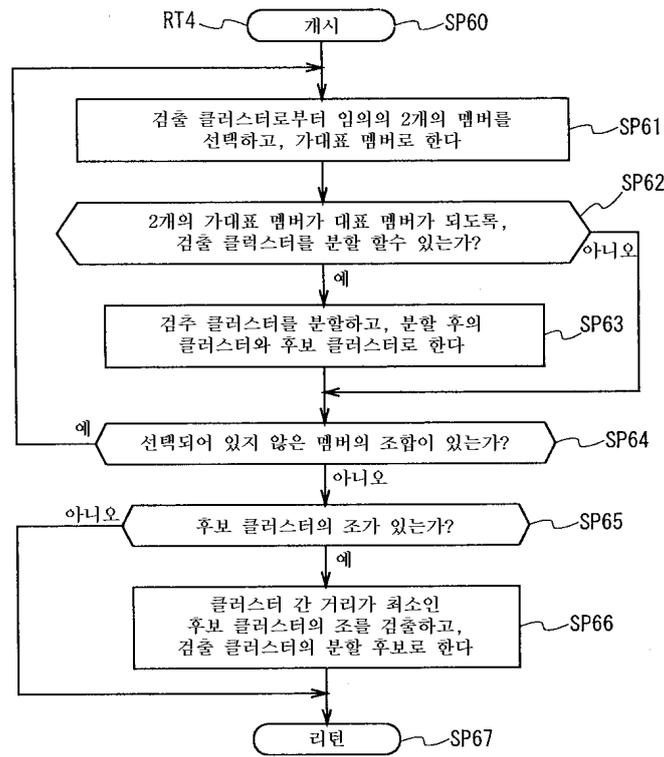
도면20



도면21



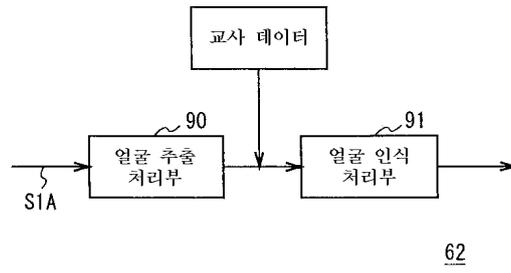
도면22



도면23

음운계열	발화 단어
doroa:	風呂×1;
kuro	風呂×3;
Nfuro	風呂×20;
NhoNn	本×18;
hoNN	本×6;
NhoNda	本×10;
NhoNde:su	本×4;
ohoN	オレンジ×1;本×19;
hoNgdawasoNre:a:	本×2;
a:modori:	綠色×11;
omidori:	綠色×10;
e:imidori:	綠色×3;
Nmidori:	綠色×5;
a:midori:iroiresu	綠色×4;
Nro:ka	廊下×10;
Nro:kaNa	廊下×10;

도면24



도면25

