



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0142884
(43) 공개일자 2024년10월02일

(51) 국제특허분류(Int. Cl.)

G10L 17/04 (2013.01) G06V 40/16 (2022.01)
G10L 15/22 (2006.01) G10L 17/02 (2013.01)
H04R 3/00 (2006.01)

(52) CPC특허분류

G10L 17/04 (2013.01)
G06V 40/161 (2022.01)

(21) 출원번호 10-2023-0037481

(22) 출원일자 2023년03월22일

심사청구일자 2023년03월22일

(71) 출원인

인하대학교 산학협력단

인천광역시 미추홀구 인하로 100(용현동, 인하대학교)

(72) 발명자

이보원

인천광역시 미추홀구 인하로 100, 하이테크센터 1018호(용현동)

김동현

경기도 고양시 일산서구 산현로 45, 1401동 1201호(탄현동, 탄현마을14단지아파트)

나중환

인천광역시 서구 건지로 404, 217동 702호(가좌동, 가좌한신휴플러스)

(74) 대리인

양성보

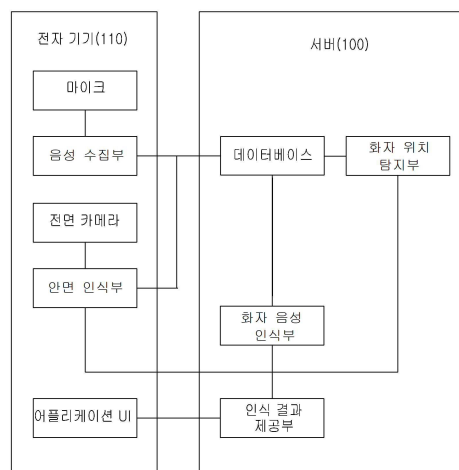
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 다중 모델 화자 인식 시스템

(57) 요약

다중 모델 화자 인식 시스템이 개시된다. 일 실시예에 따른 다중 모델 화자 인식 시스템에 의해 수행되는 다중 모델 화자 인식 방법은, 비학습 기반 화자 인식 기법에 기초하여 시청각 화자 인식을 통해 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보를 획득하는 단계 및 음원 위치 인식을 통해 화자 위치 정보를 포함하는 화자 인식 정보를 획득하는 단계; 상기 획득된 화자 인식 정보를 이용하여 화자 인식 모델을 학습시키는 단계; 상기 학습된 화자 인식 모델을 통해 화자의 발화 데이터로부터 화자 음성 정보를 인식하는 단계; 및 상기 획득된 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보와 상기 인식된 화자 음성 정보를 조합하여 화자 인식 결과를 제공하는 단계를 포함할 수 있다.

대표도 - 도1



(52) CPC특허분류

G10L 17/02 (2013.01)

H04R 3/005 (2013.01)

G10L 2015/221 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179347
과제번호	00155915
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신인재양성(교육운영수익)
연구과제명	[Ezbaro][정부+민간]인공지능융합혁신인재양성(2차년도)
기 여 율	1/1
과제수행기관명	인하대학교 산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

다중 모델 화자 인식 시스템에 의해 수행되는 다중 모델 화자 인식 방법에 있어서,

비학습 기반 화자 인식 기법에 기초하여 시청각 화자 인식을 통해 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보를 획득하고, 음원 위치 인식을 통해 화자 위치 정보를 포함하는 화자 인식 정보를 획득하는 단계;

상기 획득된 화자 인식 정보를 이용하여 화자 인식 모델을 학습시키는 단계;

상기 학습된 화자 인식 모델을 통해 화자의 발화 데이터로부터 화자 음성 정보를 인식하는 단계; 및

상기 획득된 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보와 상기 인식된 화자 음성 정보를 조합하여 화자 인식 결과를 제공하는 단계

를 포함하는 다중 모델 화자 인식 방법.

청구항 2

제1항에 있어서,

상기 획득하는 단계는,

전자 기기의 카메라를 통해 촬영되는 영상 정보 내에서 각 화자들의 안면 정보를 인식하고, 상기 인식된 각 화자들 중 발화 여부를 탐지하고, 상기 영상 정보 내에서 상기 탐지된 발화 중인 화자의 이동에 따른 안면 위치 변화를 추적하고, 상기 추적된 안면 위치 변화를 포함하는 발화 중인 화자 안면 정보를 인식하는 단계

를 포함하는 다중 모델 화자 인식 방법.

청구항 3

제2항에 있어서,

상기 획득하는 단계는,

서로 다른 문장의 모두에서 확인되는 발화 중인 화자가 존재하지 않을 경우 판단 불가, 서로 다른 문장 중 하나의 문장에서 확인되는 발화 중인 화자가 존재할 경우 화자 다름, 서로 다른 문장의 모두에서 확인되는 발화 중인 화자가 존재할 경우 화자의 일치 여부를 반환하는 단계

를 포함하는 다중 모델 화자 인식 방법.

청구항 4

제1항에 있어서,

상기 획득하는 단계는,

서로 다른 위치 또는 방향에 설치되어 유선 또는 무선으로 상호작용하는 복수의 마이크로로부터 수신된 음성 데이터로부터 음성의 휴지를 탐지하여 문장의 발화가 시작된 시점 및 종료된 시점을 결정하고, 상기 결정된 문장의 발화가 시작된 시점 및 종료된 시점에 기초하여 발화 음성을 문장 단위로 끊어 데이터베이스에 저장하는 단계

를 포함하는 다중 모델 화자 인식 방법.

청구항 5

제4항에 있어서,

상기 획득하는 단계는,

상기 데이터베이스를 통해 상기 복수의 마이크로로부터 녹음된 화자의 발화 데이터를 문장 별로 전달받아 각 문장을 발화한 화자의 위치를 계산하고, 상기 계산된 화자의 위치와 상기 결정된 문장의 발화가 시작점 시점 및 종료된 시점에 기초하여 탐지된 화자의 변화 여부를 포함하는 발화 중인 화자 위치 정보를 인식하는 단계를 포함하는 다중 모델 화자 인식 방법.

청구항 6

제5항에 있어서,

상기 획득하는 단계는,

서로 다른 문장의 모두를 발화 중인 화자의 위치가 확인 가능하거나 서로 다른 문장의 모두를 발화한 화자가 일정 거리 내에 위치할 경우 판단 불가, 서로 다른 문장의 모두를 발화 중인 화자의 위치가 확인 불가능하거나 서로 다른 문장의 모두를 발화한 화자가 일정 거리 내에 위치에 존재하지 않을 경우 화자 다름을 반환하는 단계를 포함하는 다중 모델 화자 인식 방법.

청구항 7

제1항에 있어서,

상기 학습시키는 단계는,

서로 다른 위치 또는 방향에 설치되어 유선 또는 무선으로 상호작용하는 복수의 마이크로로부터 생성된 음성 데이터를 데이터베이스에 저장하고, 상기 저장된 데이터베이스에 저장된 음성 데이터의 특성을 탐지하여 화자를 구별하도록 화자 인식 모델을 학습시키는 단계를 포함하는 다중 모델 화자 인식 방법.

청구항 8

제1항에 있어서,

상기 화자 인식 결과를 제공하는 단계는,

상기 획득된 화자 안면 정보를 이용하여 화자 일치 여부를 판단하고,

상기 획득된 화자 안면 정보를 이용하여 화자 일치 여부의 판단이 가능할 경우, 상기 획득된 화자 안면 정보를 화자 인식 결과로서 채택하고,

획득된 화자 안면 정보를 이용하여 화자 일치 여부의 판단이 불가능할 경우, 획득된 화자 위치 정보를 이용하여 화자 일치 여부를 판단하고,

상기 획득된 화자 위치 정보를 이용하여 화자 일치 여부의 판단이 가능할 경우, 상기 획득된 화자 위치 정보를 화자 인식 결과로서 채택하고,

상기 획득된 화자 위치 정보와 상기 획득된 화자 안면 정보를 이용하여 화자 일치 여부의 판단이 불가능할 경우, 상기 획득된 화자 위치 정보, 상기 획득된 화자 안면 정보 및 상기 인식된 화자 음성 정보를 조합하여 서로 다른 문장의 조합에 대해 화자 인식을 수행하여 각 문장의 화자를 인식 번호로 구별하는 단계를 포함하는 다중 모델 화자 인식 방법.

청구항 9

다중 모델 화자 인식 시스템에 있어서,

비학습 기반 화자 인식 기법에 기초하여 시청각 화자 인식을 통해 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보를 획득하는 안면 인식부;

음원 위치 인식을 통해 화자 위치 정보를 포함하는 화자 인식 정보를 획득하는 화자 위치 탐지부;

상기 획득된 화자 인식 정보를 이용하여 화자 인식 모델을 학습시키는 모델 학습부;

상기 학습된 화자 인식 모델을 통해 화자의 발화 데이터로부터 화자 음성 정보를 인식하는 화자 음성 인식부;

및

상기 획득된 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보와 상기 인식된 화자 음성 정보를 조합하여 화자 인식 결과를 제공하는 화자 인식 결과 제공부

를 포함하는 다중 모델 화자 인식 시스템.

발명의 설명

기술 분야

[0001] 아래의 설명은 복수의 수단을 이용하여 발화 중인 화자를 인식하는 기술에 관한 것이다.

배경 기술

[0003] 비대면 미팅 등의 상황에서 복수의 화자가 존재하는 경우 발화 중인 화자를 알아내는 것은 매우 중요하다. 종래의 기계 학습을 활용하는 기술은 화자의 발화 데이터를 수집한 뒤 특징을 모델링하고, 입력된 미상 화자의 발화와 가장 잘 맞는 모델을 찾음으로써 화자 인식을 수행하기 때문에 많은 발화 데이터를 요구하지만, 서비스의 개발 단계에서 처음부터 많은 발화 데이터를 가지는 것은 어렵다. 한편 기계 학습 이외의 화자 인식 기술은 많은 발화 데이터를 개발 단계부터 요구하지는 않지만, 대개 적용할 수 있는 범위나 그 성능이 제한적이다.

[0004] 기존의 기술들은 단일의 화자 인식 기법만을 사용하므로 기대할 수 있는 성능의 범위가 제한적이거나, 초기에 많은 사용자 데이터를 요구하는 한계점이 있다. 상용 서비스의 개발을 위해서는 적용 가능한 화자 인식 기법을 복합적으로 동원하여 서비스를 우선 구현한 뒤, 사용자의 데이터를 축적하고 이를 이용하여 점진적으로 성능 향상을 도모하는 것이 바람직하다. 따라서 우선 비 학습 기반 화자 인식 기법으로 서비스 제공을 개시하고, 축적된 사용자 데이터를 이용해 기계 학습 모델을 구축하여 추후에 성능 향상을 도모할 수 있는 방법의 개발이 요구된다.

[0005] 참고로, 대한민국공개특허 제10-2018-0073168호(공개일 2018.07.02)에 다자간 회의에서의 화자 인식 및 화자 추적 방법이 개시되어 있다.

발명의 내용

해결하려는 과제

[0007] 한 공간에서 다양한 화자가 발화하는 상황에서 발화의 내용을 화자를 구별하여 텍스트로 저장하는 것은 회의록 작성과 같은 상황에서 필수적이다. 화자를 구별하지 않고 발화의 내용을 텍스트로 변환하는 기술(Speech-to-Text; STT)은 현재 상당한 성능을 보여주고 있지만, 발화 중인 화자를 파악하는 기술(Speaker Recognition)은 상대적으로 제약 조건이 많고 연구가 미진하다. 화자 인식을 위해 기계 학습과 비 학습 분야에서 다양한 기법들이 존재한다. 비 학습 기법은 사용 가능한 상황이 제약적이지만 즉시 적용이 가능하고, 기계 학습 기법은 많은 사용자 데이터를 축적해야 하므로 즉시 적용이 불가능하지만 비 학습 기법들에 비해 보다 범용적인 상황에서 사용될 수 있다.

[0008] 본 발명은 복수의 화자 인식 기법을 이용하여, 우선 비 학습 기반 화자 인식 기법으로 일단 서비스 제공을 개시하고, 축적된 사용자 데이터를 이용해 딥 러닝 모델을 구축함으로써 추후의 성능 향상을 도모할 수 있는 화자 인식 시스템 및 방법을 제안하고자 한다.

[0009] 본 발명은 비 학습 기반 화자 인식 기법인 영상 기반으로 화자의 안면을 인식하여 화자를 인식하는 기법, 음원을 추적하여 화자를 인식하는 기법과 기계 학습 기반 화자 인식 기법인 화자의 발화에서 특징을 추출하여 화자를 인식하는 기법을 조합하여 화자에 대한 정보가 없는 상황에서도 즉시 화자 인식이 가능하면서 추후 축적된 사용자 데이터를 통해 화자의 특징을 학습하여 비 학습 기반 화자 인식이 불가능한 상황에서도 정확한 화자 인식이 가능하도록 하는 화자 인식 시스템 및 방법을 제공할 수 있다.

과제의 해결 수단

- [0011] 다중 모델 화자 인식 시스템에 의해 수행되는 다중 모델 화자 인식 방법은, 비학습 기반 화자 인식 기법에 기초하여 시청각 화자 인식을 통해 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보를 획득하고, 음원 위치 인식을 통해 화자 위치 정보를 포함하는 화자 인식 정보를 획득하는 단계; 상기 획득된 화자 인식 정보를 이용하여 화자 인식 모델을 학습시키는 단계; 상기 학습된 화자 인식 모델을 통해 화자의 발화 데이터로부터 화자 음성 정보를 인식하는 단계; 및 상기 획득된 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보와 상기 인식된 화자 음성 정보를 조합하여 화자 인식 결과를 제공하는 단계를 포함할 수 있다.
- [0012] 상기 획득하는 단계는, 전자 기기의 카메라를 통해 촬영되는 영상 정보 내에서 각 화자들의 안면 정보를 인식하고, 상기 인식된 각 화자들 중 발화 여부를 탐지하고, 상기 영상 정보 내에서 상기 탐지된 발화 중인 화자의 이동에 따른 안면 위치 변화를 추적하고, 상기 추적된 안면 위치 변화를 포함하는 발화 중인 화자 안면 정보를 인식하는 단계를 포함할 수 있다.
- [0013] 상기 획득하는 단계는, 서로 다른 문장의 모두에서 확인되는 발화 중인 화자가 존재하지 않을 경우 판단 불가, 서로 다른 문장 중 하나의 문장에서 확인되는 발화 중인 화자가 존재할 경우 화자 다름, 서로 다른 문장의 모두에서 확인되는 발화 중인 화자가 존재할 경우 화자의 일치 여부를 반환하는 단계를 포함할 수 있다.
- [0014] 상기 획득하는 단계는, 서로 다른 위치 또는 방향에 설치되어 유선 또는 무선으로 상호작용하는 복수의 마이크로로부터 수신된 음성 데이터로부터 음성의 휴지를 탐지하여 문장의 발화가 시작된 시점 및 종료된 시점을 결정하고, 상기 결정된 문장의 발화가 시작된 시점 및 종료된 시점에 기초하여 발화 음성을 문장 단위로 끊어 데이터베이스에 저장하는 단계를 포함할 수 있다.
- [0015] 상기 획득하는 단계는, 상기 데이터베이스를 통해 상기 복수의 마이크로로부터 녹음된 화자의 발화 데이터를 문장별로 전달받아 각 문장을 발화한 화자의 위치를 계산하고, 상기 계산된 화자의 위치와 상기 결정된 문장의 발화가 시작점 시점 및 종료된 시점에 기초하여 탐지된 화자의 변화 여부를 포함하는 발화 중인 화자 위치 정보를 인식하는 단계를 포함할 수 있다.
- [0016] 상기 획득하는 단계는, 서로 다른 문장의 모두를 발화 중인 화자의 위치가 확인 가능하거나 서로 다른 문장의 모두를 발화한 화자가 일정 거리 내에 위치할 경우 판단 불가, 서로 다른 문장의 모두를 발화 중인 화자의 위치가 확인 불가능하거나 서로 다른 문장의 모두를 발화한 화자가 일정 거리 내에 위치에 존재하지 않을 경우 화자 다름을 반환하는 단계를 포함할 수 있다.
- [0017] 상기 학습시키는 단계는, 서로 다른 위치 또는 방향에 설치되어 유선 또는 무선으로 상호작용하는 복수의 마이크로로부터 생성된 음성 데이터를 데이터베이스에 저장하고, 상기 데이터베이스에 저장된 음성 데이터의 특성을 탐지하여 화자를 구별하도록 화자 인식 모델을 학습시키는 단계를 포함할 수 있다.
- [0018] 상기 화자 인식 결과를 제공하는 단계는, 상기 획득된 화자 안면 정보를 이용하여 화자 일치 여부를 판단하고, 상기 획득된 화자 안면 정보를 이용하여 화자 일치 여부의 판단이 가능할 경우, 상기 획득된 화자 안면 정보를 화자 인식 결과로서 채택하고, 획득된 화자 안면 정보를 이용하여 화자 일치 여부의 판단이 불가능할 경우, 획득된 화자 위치 정보를 이용하여 화자 일치 여부를 판단하고, 상기 획득된 화자 위치 정보를 이용하여 화자 일치 여부의 판단이 가능할 경우, 상기 획득된 화자 위치 정보를 화자 인식 결과로서 채택하고, 상기 획득된 화자 위치 정보와 상기 획득된 화자 안면 정보를 이용하여 화자 일치 여부의 판단이 불가능할 경우, 상기 획득된 화자 위치 정보, 상기 획득된 화자 안면 정보 및 상기 인식된 화자 음성 정보를 조합하여 서로 다른 문장의 조합에 대해 화자 인식을 수행하여 각 문장의 화자를 인식 번호로 구별하는 단계를 포함할 수 있다.
- [0019] 다중 모델 화자 인식 시스템은, 마이크로로부터 녹음된 음성으로부터 발화를 인식하고 이를 분절하여 저장하는 음성 수집부; 비학습 기반 화자 인식 기법에 기초하여 시청각 화자 인식을 통해 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보를 획득하는 안면 인식부; 음원 위치 인식을 통해 화자 위치 정보를 포함하는 화자 인식 정보를 획득하는 화자 위치 탐지부; 상기 획득된 화자 인식 정보를 이용하여 화자 인식 모델을 학습시키는 모델 학습부; 상기 학습된 화자 인식 모델을 통해 화자의 발화 데이터로부터 화자 음성 정보를 인식하는 화자 음성 인식부; 및 상기 획득된 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보와 상기 인식된 화자 음성 정보를 조합하여 화자 인식 결과를 제공하는 화자 인식 결과 제공부를 포함할 수 있다.

발명의 효과

- [0021] 화자 및 발화의 특성에 대한 정보가 없는 상황에서도 즉시 화자 인식의 결과를 제공할 수 있으며, 축적되는 화자 및 발화의 정보를 이용해 점진적으로 화자 인식의 정확도를 개선할 수 있다.

도면의 간단한 설명

- [0023] 도 1은 일 실시예에 있어서, 다중 모델 화자 인식 동작을 설명하기 위한 도면이다.
 도 2는 일 실시예에 있어서, 다중 모델 화자 인식 방법을 설명하기 위한 흐름도이다.
 도 3은 일 실시예에 있어서, 화자 인식 결과를 제공하는 동작을 설명하기 위한 도면이다.
 도 4는 일 실시예에 있어서, 안면 인식 정보를 이용하여 화자 일치 여부를 판단하는 동작을 설명하기 위한 도면이다.
 도 5는 일 실시예에 있어서, 화자 위치 정보를 이용하여 화자 일치 여부를 판단하는 동작을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0024] 이하, 실시예를 첨부한 도면을 참조하여 상세히 설명한다.
- [0026] 실시예에서는 복수의 수단을 이용하여 화자 인식의 정확도를 향상시키는 동작에 대하여 설명하기로 한다. 구체적으로, 복수의 마이크를 기반으로 화자 위치 인식과 카메라를 사용한 화자 안면 인식을 사용하여 1차적 화자 인식을 수행하고, 복수의 마이크와 카메라를 통해 수집된 데이터를 이용해 화자 인식 모델을 학습시키며, 학습된 화자 인식 모델을 이용한 2차 화자 인식 결과와 전송한 1차 화자 인식 결과를 종합하여 최종적인 화자 인식 결과를 제공한다. 이에, 사전에 화자의 데이터가 존재하지 않으면서 화자가 중간에 참여하거나 이탈하더라도 즉시 높은 정확도의 화자 인식 결과를 제공할 수 있다.
- [0027] 도 1은 일 실시예에 있어서, 다중 모델 화자 인식 동작을 설명하기 위한 도면이다.
- [0028] 전자 기기(110)와 서버(100) 간의 동작을 통해 화자를 인식하는 동작에 대하여 설명하기로 한다. 일례로, 별도의 서버가 존재하는 경우, 회의에 참여하는 화자의 다수가 본인의 전자 기기(110)를 가지고 참여하고, 전자 기기(110)는 네트워크를 통해 서버(100)와 연결될 수 있다. 다른 예로서, 별도의 서버가 존재하지 않는 경우, 전자 기기(110) 중 하나를 로컬 서버로 지정하여 전자 기기 중 하나가 서버의 역할을 함께 하도록 할 수 있다.
- [0029] 전자 기기(110)의 위치는 와이파이 등을 이용한 위치 모니터링을 통해 판별될 수 있다. 전자 기기(110)는 서버로 네트워크를 통해 연결되어 있으며, 전자 기기(110)는 카메라 및 복수의 마이크를 포함할 수 있다. 전자 기기(110)는 전면 카메라 또는 후면 카메라를 통해 발화 중인 화자의 안면을 인식할 수 있으며, 복수의 마이크를 통해 화자의 음성을 녹음하고 문장을 구별하여 저장할 수 있다. 전자 기기(110)에서 카메라 및 복수의 마이크를 통해 수집된 데이터는 음성 수집부에 의해 처리될 수 있다. 전자 기기(110)의 음성 수집부가 처리한 데이터를 서버(100)로 전송할 수 있다. 예를 들면, 전자 기기(110)는 전면 카메라를 이용하여 주위를 촬영한 뒤 영상 정보를 서버(100)로 전송할 수 있다. 이때, 전자 기기(110)를 소유한 화자는 전자 기기(110)의 전면에 있을 것이므로, 영상 정보 내에는 전자 기기(110)를 소유한 화자가 전부 포함될 수 있으며, 전자 기기(110)를 소유하지 않은 화자는 영상 정보에 포함되지 않을 수 있다.
- [0030] 전자 기기(110)은 음성 수집부 및 안면 인식부를 포함할 수 있다.
- [0031] 서버(100)는 화자 위치 탐지부, 화자 음성 인식부 및 인식 결과 제공부를 포함할 수 있다. 또한, 서버(100)는 데이터베이스를 포함할 수 있다.
- [0032] 화자 안면 인식부는 전자 기기의 카메라를 통해 촬영되는 영상 정보 내에서 각 화자들의 안면 정보를 인식하고, 인식된 각 화자들 중 발화 여부를 탐지하고, 영상 정보 내에서 탐지된 발화 중인 화자의 이동에 따른 안면 위치

변화를 추적하고, 추적된 안면 위치 변화를 포함하는 발화 중인 화자 안면 정보를 인식할 수 있다.

[0033] 화자 위치 탐지부는 서로 다른 위치 또는 방향에 설치되어 유선 또는 무선으로 상호작용하는 복수의 마이크로로부터 수신된 음성 데이터로부터 음성의 휴지를 탐지하여 문장의 발화가 시작된 시점 및 종료된 시점을 결정하고, 결정된 문장의 발화가 시작된 시점 및 종료된 시점에 기초하여 발화 음성을 문장 단위로 끊어 데이터베이스에 저장할 수 있다. 화자 위치 인식부는 데이터베이스를 통해 복수의 마이크로로부터 녹음된 화자의 발화 데이터를 문장 별로 전달받아 각 문장을 발화한 화자의 위치를 계산하고, 계산된 화자의 위치와 결정된 문장의 발화가 시작점 시점 및 종료된 시점에 기초하여 탐지된 화자의 변화 여부를 포함하는 발화 중인 화자 위치 정보를 인식할 수 있다. 모든 전자 기기의 마이크에서 수집한 데이터를 네트워크를 통해 실시간으로 공유하기 때문에, 마이크에서 수집한 데이터를 이용하여 발화하는 화자의 위치를 파악할 수 있다.

[0034] 데이터베이스는 마이크로로부터 수집되어 문장 단위로 구분된 음성 데이터를 저장할 수 있다. 또한, 데이터베이스는 화자 안면 인식의 결과와 화자 위치 탐지의 결과를 저장할 수 있다. 이때, 화자 안면 인식의 결과와 화자 위치 탐지의 결과가 탐지 불가인 경우를 제외하고, 화자 안면 인식의 결과와 화자 위치 탐지의 결과를 저장할 수 있다.

[0035] 화자 음성 인식부는 획득된 화자 인식 정보를 이용하여 학습된 화자 인식 모델을 통해 화자의 발화 데이터로부터 화자 음성 정보를 인식할 수 있다.

[0036] 인식 결과 제공부는 획득된 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보와 인식된 화자 음성 정보를 조합하여 화자 인식 결과를 제공할 수 있다. 인식 결과 제공부는 두 문장의 조합에 대해 화자 안면 인식부, 화자 위치 탐지부, 화자 음성 인식부로부터 두 문장의 화자가 동일한지 여부를 확인하여, 화자 안면 인식부와 화자 위치 탐지부가 모두 판단 불가 결론을 내릴 때에만 화자 음성 인식부의 결과를 채택하고, 화자 안면 인식부만이 판단 불가 결론을 내릴 때에는 화자 위치 탐지부의 결과를 채택하며, 그렇지 않을 때에는 화자 안면 인식부의 결과를 채택하는 방법을 통해, 모든 두 문장의 조합에 대해 화자 인식을 수행하여 각 문장의 화자를 인식 번호로 구별할 수 있다. 이 경우 전자 기기(110)를 가지고 있는 화자는 화자 안면 인식부에 의해 높은 정확도로 인식이 가능하고, 전자 기기(110)를 가지고 있지 않지만 충분히 멀리 떨어져 있는 두 화자는 화자 위치 탐지부에 의해 높은 정확도로 인식이 가능하므로, 상대적으로 정확도가 낮은 화자 음성 인식부에 의해 화자가 구별되는 경우는 전자 기기(110)를 가지고 있지 않으며 가까이 붙어 있는 복수 화자 구별의 경우 밖에 없으므로 기계 학습 기반 화자 인식 시스템에서보다 일반적으로 높은 정확도를 기대할 수 있다.

[0037] 도 2는 일 실시예에 있어서, 다중 모델 화자 인식 방법을 설명하기 위한 흐름도이다.

[0038] 단계(210)에서 다중 모델 화자 인식 시스템은 비학습 기반 화자 인식 기법에 기초하여 시청각 화자 인식을 통해 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보를 획득할 수 있으며, 음원 위치 인식을 통해 화자 위치 정보를 포함하는 화자 인식 정보를 획득할 수 있다. 다중 모델 화자 인식 시스템은 전자 기기의 카메라를 통해 촬영되는 영상 정보 내에서 각 화자들의 안면 정보를 인식하고, 인식된 각 화자들 중 발화 여부를 탐지하고, 영상 정보 내에서 탐지된 발화 중인 화자의 이동에 따른 안면 위치 변화를 추적하고, 추적된 안면 위치 변화를 포함하는 발화 중인 화자 안면 정보를 인식할 수 있다. 도 4를 참고하면, 안면 인식 정보를 이용하여 화자 일치 여부를 판단하는 동작을 설명하기 위한 도면이다. 도 4에서는 서로 다른 두 문장을 이용하여 화자 일치 여부를 판단하는 것을 예를 들어 설명하기로 한다. 다중 모델 화자 인식 시스템은 서로 다른 문장의 모두에서 발화 중인 화자 모두 확인 가능한 지 여부를 판단할 수 있다(410). 다중 모델 화자 인식 시스템은 서로 다른 문장의 모두에서 확인되는 발화 중인 화자가 존재할 경우 두 화자의 일치 여부를 반환할 수 있다(450). 다중 모델 화자 인식 시스템은 서로 다른 문장의 모두에서 발화 중인 화자가 존재하지 않는 것으로 판단됨에 따라 어느 하나의 문장에서 발화 중인 화자가 확인 가능한지 여부를 판단할 수 있다(420). 다중 모델 화자 인식 시스템은 어느 하나의 문장에서 발화 중인 화자가 확인이 가능할 경우 판단 불가를 반환하고(430), 서로 다른 문장 중 하나의 문장에서 확인되는 발화 중인 화자가 존재할 경우 화자 다름을 반환할 수 있다(440).

[0039] 또한, 다중 모델 화자 인식 시스템은 서로 다른 위치 또는 방향에 설치되어 유선 또는 무선으로 상호작용하는 복수의 마이크로로부터 수신된 음성 데이터로부터 음성의 휴지를 탐지하여 문장의 발화가 시작된 시점 및 종료된 시점을 결정하고, 결정된 문장의 발화가 시작된 시점 및 종료된 시점에 기초하여 발화 음성을 문장 단위로 끊어 데이터베이스에 저장할 수 있다. 다중 모델 화자 인식 시스템은 데이터베이스를 통해 복수의 마이크로로부터 녹음된 화자의 발화 데이터를 문장 별로 전달받아 각 문장을 발화한 화자의 위치를 계산하고, 계산된 화자의 위치와 결정된 문장의 발화가 시작점 시점 및 종료된 시점에 기초하여 탐지된 화자의 변화 여부를 포함하는 발화 중인 화자 위치 정보를 인식할 수 있다. 도 5를 참고하면, 화자 위치 정보를 이용하여 화자 일치 여부를 판단하는

동작을 설명하기 위한 도면이다. 도 5에서는 서로 다른 두 문장을 이용하여 화자 일치 여부를 판단하는 것을 예를 들어 설명하기로 한다. 다중 모델 화자 인식 시스템은 서로 다른 문장의 모두에서 발화 중인 두 화자의 위치가 확인 가능한 지 여부를 판단할 수 있다(510). 다중 모델 화자 인식 시스템은 서로 다른 문장의 모두에서 발화 중인 두 화자의 위치가 확인 가능하고 서로 다른 문장의 모두를 발화한 화자가 일정 거리 내에 위치할 경우 판단 불가를 반환하고(530), 서로 다른 문장의 모두를 발화 중인 두 화자의 위치가 확인 불가능하거나 서로 다른 문장의 모두를 발화한 두 화자가 일정 거리 내에 위치에 존재하지 않을 경우 화자 다름을 반환할 수 있다(540).

[0040] 단계(220)에서 다중 모델 화자 인식 시스템은 획득된 화자 인식 정보를 이용하여 화자 인식 모델을 학습시킬 수 있다. 다중 모델 화자 인식 시스템은 서로 다른 위치 또는 방향에 설치되어 유선 또는 무선으로 상호작용하는 복수의 마이크로부터 생성된 음성 데이터를 음성 데이터베이스에 저장하고, 저장된 음성 데이터베이스에 저장된 음성 데이터의 특성을 탐지하여 화자를 구별하도록 화자 인식 모델을 학습시킬 수 있다. 예를 들면, 다중 모델 화자 인식 시스템은 음성 데이터를 학습 데이터 셋으로 이용하여 음의 높낮이, 억양, 음색, 음의 크기 등을 포함하는 특성을 추출하여 화자를 구별하도록 화자 인식 모델을 학습시킬 수 있다.

[0041] 단계(230)에서 다중 모델 화자 인식 시스템은 학습된 화자 인식 모델을 통해 화자의 발화 데이터로부터 화자 음성 정보를 인식할 수 있다. 다중 모델 화자 인식 시스템은 전자 기기에서 복수의 마이크를 통해 입력되는 새로운 화자의 발화 데이터를 학습된 화자 인식 모델의 입력 데이터로 입력할 수 있다.

[0042] 단계(240)에서 다중 모델 화자 인식 시스템은 획득된 화자 위치 정보 및 화자 안면 정보를 포함하는 화자 인식 정보와 인식된 화자 음성 정보를 조합하여 화자 인식 결과를 제공할 수 있다. 도 3을 참고하면, 화자 인식 결과를 제공하는 동작을 설명하기 위한 도면이다. 다중 모델 화자 인식 시스템은 획득된 화자 안면 정보를 이용하여 화자 일치 여부를 판단할 수 있다(310). 다중 모델 화자 인식 시스템은 획득된 화자 안면 정보를 이용하여 화자 일치 여부의 판단이 가능할 경우, 획득된 화자 안면 정보를 화자 인식 결과로서 채택할 수 있다(350). 다중 모델 화자 인식 시스템은 획득된 화자 안면 정보를 이용하여 화자 일치 여부의 판단이 불가능할 경우, 획득된 화자 위치 정보를 이용하여 화자 일치 여부를 판단할 수 있다(320). 다중 모델 화자 인식 시스템은 획득된 화자 위치 정보를 이용하여 화자 일치 여부의 판단이 가능할 경우, 획득된 화자 위치 정보를 화자 인식 결과로서 채택할 수 있다(340). 다중 모델 화자 인식 시스템은 획득된 화자 위치 정보와 획득된 화자 안면 정보를 이용하여 화자 일치 여부의 판단이 불가능할 경우, 획득된 화자 위치 정보, 획득된 화자 안면 정보 및 인식된 화자 음성 정보를 조합하여 서로 다른 문장의 조합에 대해 화자 인식을 수행하여 각 문장의 화자를 인식 번호로 구별할 수 있다(330).

[0043] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0044] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embodiment)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0045] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판

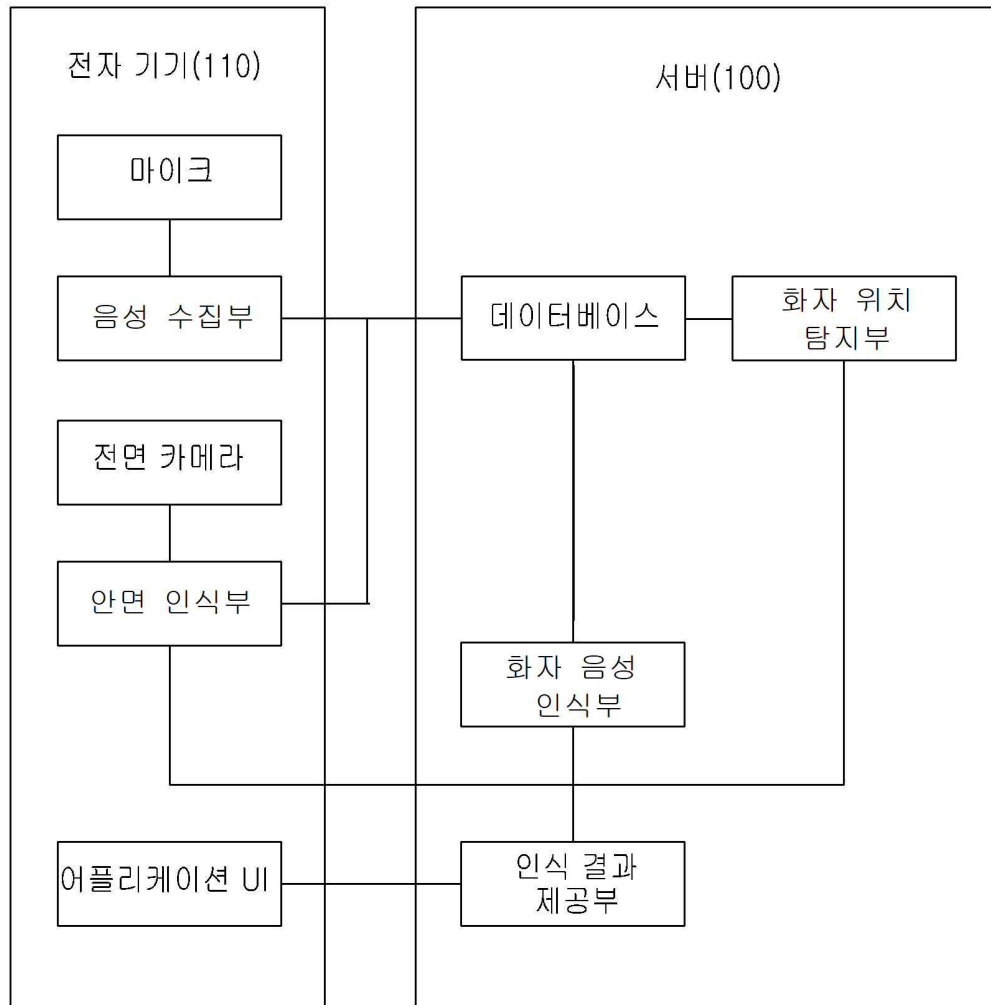
독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0046] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

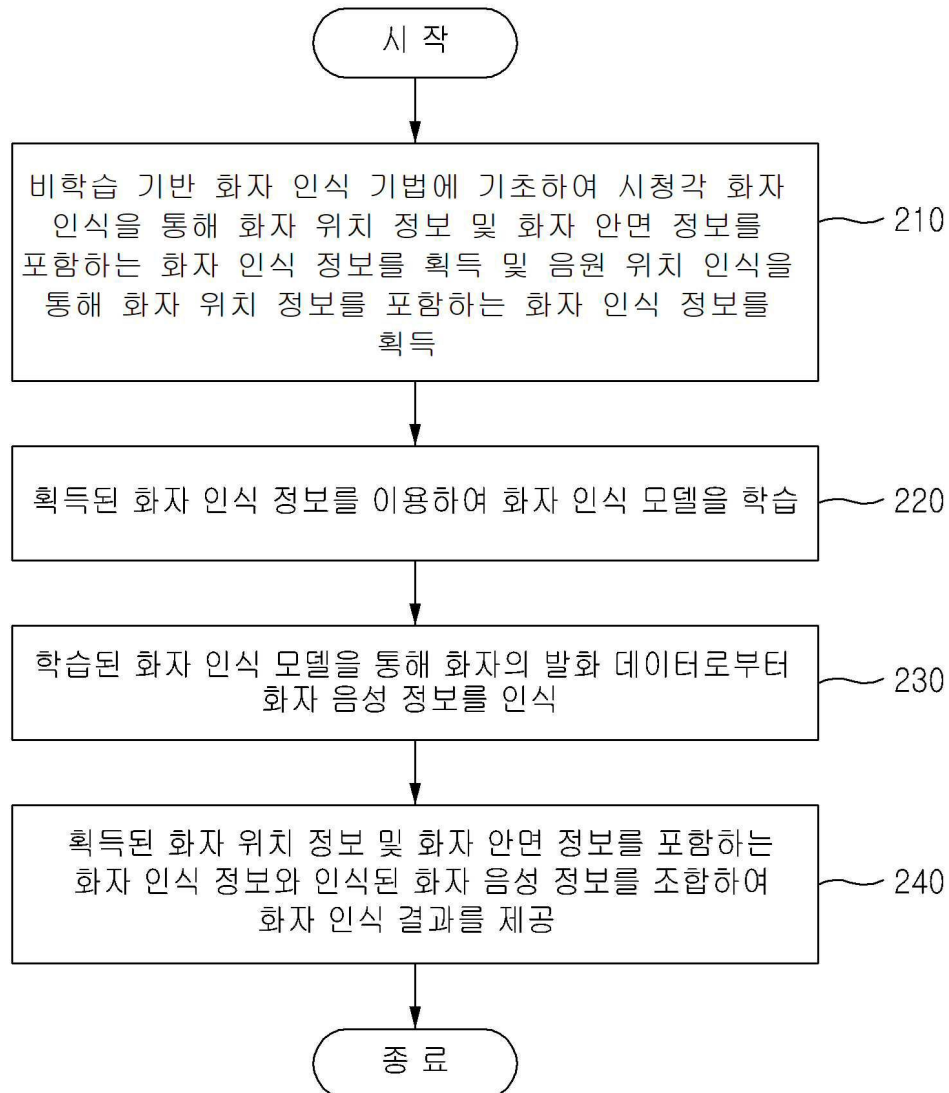
[0047] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

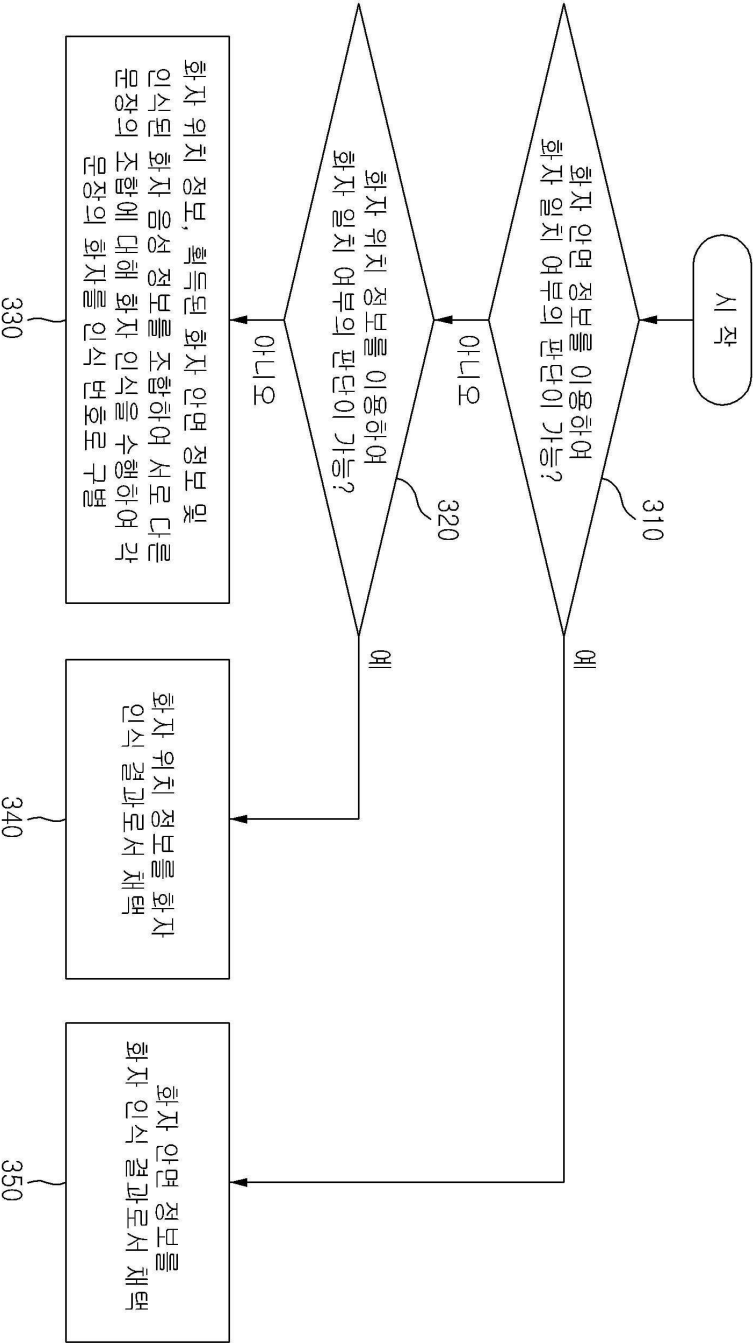
도면1



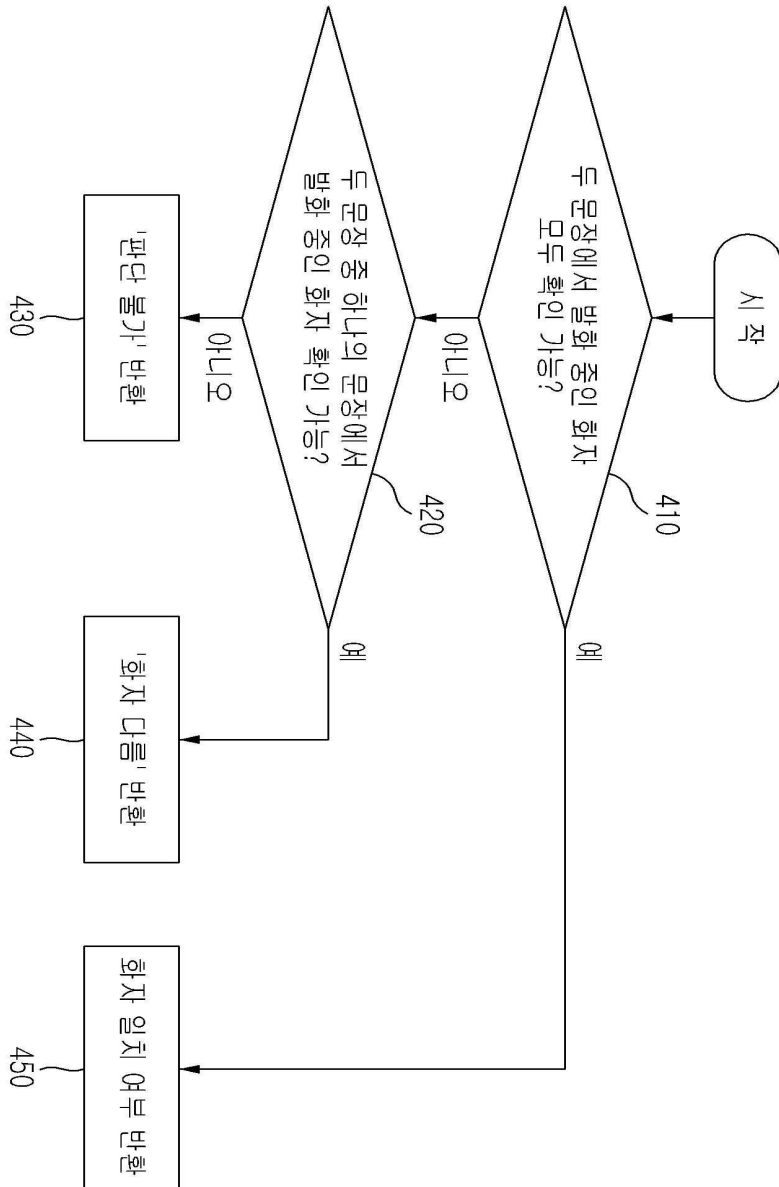
도면2



도면3



도면4



도면5

