(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0104389
(43) 공개일자 2024년07월05일

(51) 국제특허분류(Int. Cl.)
G10L 15/22 (2006.01) *G10L 15/02* (2006.01)
G10L 15/26 (2006.01) *G10L 21/055* (2013.01)
G10L 21/10 (2013.01) *G10L 25/63* (2013.01)
G10L 25/78 (2013.01) *G10L 25/93* (2013.01)
(52) CPC특허분류
G10L 15/22 (2013.01)
G10L 15/02 (2013.01)
(21) 출원번호 10-2022-0186600
(22) 출원일자 2022년12월28일
심사청구일자 2023년08월25일

(71) 출원인
한국전자통신연구원
대전광역시 유성구 가정로 218 (가정동)
(72) 발명자
윤승
대전광역시 유성구 가정로 218
김승희
대전광역시 유성구 가정로 218
(뒷면에 계속)
(74) 대리인
특허법인지명

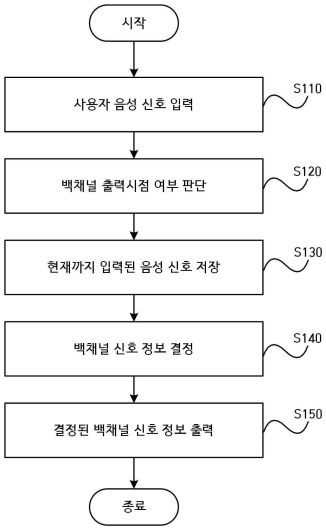
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 **교감형 백채널 신호 생성 방법 및 시스템**

(57) 요 약

교감형 백채널 신호 생성 방법이 제공된다. 상기 방법은 사용자로부터의 음성 신호를 입력받는 단계; 소정의 시점에서의 음성 신호가 입력됨에 따라 백채널 신호 출력시점인지 여부를 판단하는 단계; 상기 판단 결과 백채널 신호 출력시점에 해당하는 경우, 현재까지 입력된 음성 신호를 저장하는 단계; 상기 저장된 음성 신호에 기초하여 백채널 신호 정보를 결정하는 단계; 및 상기 결정된 백채널 신호 정보를 출력하는 단계를 포함한다.

대 표 도 - 도2



(52) CPC특허분류
G10L 15/26 (2013.01)
G10L 21/055 (2013.01)
G10L 21/10 (2013.01)
G10L 25/63 (2013.01)
G10L 25/78 (2013.01)
G10L 25/93 (2013.01)
G10L 2015/221 (2013.01)
G10L 2025/783 (2013.01)
(72) 발명자
김상훈
대전광역시 유성구 가정로 218
방정욱
대전광역시 유성구 가정로 218
이민규
대전광역시 유성구 가정로 218

이 발명을 지원한 국가연구개발사업

과제고유번호	1711160496
과제번호	2022-0-00608
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원(IITP)
연구사업명	정보통신 방송연구개발사업
연구과제명	인간과 교감하는 멀티모달 인터랙션 인공지능 기술
기 여 율	1/1
과제수행기관명	한국전자기술연구원
연구기간	2022.04.01 ~ 2022.12.31

명 세 서

청구범위

청구항 1

컴퓨터에 의해 수행되는 방법에 있어서,

사용자로부터의 음성 신호를 입력받는 단계;

소정의 시점에서의 음성 신호가 입력됨에 따라 백채널 신호 출력시점인지 여부를 판단하는 단계;

상기 판단 결과 백채널 신호 출력시점에 해당하는 경우, 현재까지 입력된 음성 신호를 저장하는 단계;

상기 저장된 음성 신호에 기초하여 백채널 신호 정보를 결정하는 단계; 및

상기 결정된 백채널 신호 정보를 출력하는 단계를 포함하는,

교감형 백채널 신호 생성 방법.

청구항 2

제1항에 있어서,

상기 결정된 백채널 신호 정보를 출력하는 단계는,

영상 기반 백채널 신호와 음성 기반 백채널 신호 중 적어도 하나를 기반으로 생성된 백채널 신호 정보를 출력하는 것인,

교감형 백채널 신호 생성 방법.

청구항 3

제1항에 있어서,

상기 소정의 시점에서의 음성 신호가 입력됨에 따라 백채널 신호 출력시점인지 여부를 판단하는 단계는,

미리 정해진 제1 임계치에 상응하는 길이의 음성 신호가 입력되었는지 여부를 판단하는 단계; 및

상기 제1 임계치에 상응하는 길이만큼 음성 신호가 입력된 경우, 해당 시점을 백채널 신호 출력시점으로 판단하는 단계를 포함하는,

교감형 백채널 신호 생성 방법.

청구항 4

제3항에 있어서,

상기 미리 정해진 제1 임계치에 상응하는 길이는 고정된 시간 길이로 결정되거나, 인공지능 캐릭터의 페르소나 또는 사용자의 반응 정보에 따라 가변된 시간 길이로 결정되는 것인,

교감형 백채널 신호 생성 방법.

청구항 5

제3항에 있어서,

상기 제1 임계치에 상응하는 길이만큼 음성 신호가 입력된 이후, 미리 정해진 제2 임계치에 상응하는 길이만큼 묵음 구간이 입력되었는지 여부를 판단하는 단계를 더 포함하고,

상기 제1 임계치에 상응하는 길이만큼 음성 신호가 입력된 경우, 해당 시점을 백채널 신호 출력시점으로 판단하는 단계는,

상기 제2 임계치에 상응하는 길이만큼 묵음 구간이 검출된 시점을 백채널 신호 출력시점으로 판단하는 것인,

교감형 백채널 신호 생성 방법.

청구항 6

제5항에 있어서,

상기 백채널 신호 출력시점은 영상 및 음성 기반의 백채널 신호 출력시점이되,

상기 판단 결과 백채널 신호 출력시점에 해당하는 경우, 현재까지 입력된 음성 신호를 저장하는 단계는,

상기 현재까지 입력된 음성 신호 및 영상 신호를 저장하는 것인,

교감형 백채널 신호 생성 방법.

청구항 7

제1항에 있어서,

상기 저장된 음성 신호에 기초하여 백채널 신호 정보를 결정하는 단계는,

상기 저장된 음성 신호를 대상으로 음성 인식 결과를 생성하는 단계;

상기 저장된 음성 신호 및 음성 인식 결과에 기초하여 사용자의 감정 상태 정보를 생성하는 단계; 및

상기 감정 상태 정보에 상응하는 백채널 신호 정보를 결정하는 단계를 포함하는,

교감형 백채널 신호 생성 방법.

청구항 8

제7항에 있어서,

상기 저장된 음성 신호에 기초하여 백채널 신호 정보를 결정하는 단계는,

상기 음성 인식 결과에 따른 텍스트가 절 경계 지점 또는 문장 종료 지점인지 여부를 판단하는 단계; 및

상기 절 경계 지점 또는 문장 종료 지점에 해당하는 경우, 상기 저장된 음성 신호에 대한 반응 정보를 표출하기 위한 소정의 제1 길이 이상의 백채널 신호 정보를 출력하는 단계를 더 포함하는,

교감형 백채널 신호 생성 방법.

청구항 9

제8항에 있어서,

상기 저장된 음성 신호에 기초하여 백채널 신호 정보를 결정하는 단계는,

상기 절 경계 지점 또는 문장 종료 지점이 아닌 경우, 사용자의 음성 신호를 계속 입력 받기 위한 상기 제1 길이보다 짧은 소정의 제2 길이 미만의 백채널 신호 정보를 출력하는 단계를 더 포함하는,

교감형 백채널 신호 생성 방법.

청구항 10

사용자로부터 음성 신호를 입력받는 입력부,

상기 음성 신호를 분석하여 교감형 백채널 신호를 생성하기 위한 프로그램이 저장된 메모리,

소정의 시점에서의 음성 신호가 입력됨에 따라 백채널 신호 출력시점인 것으로 판단시, 현재까지 입력된 음성 신호를 저장하고, 저장된 음성 신호에 기초해여 백채널 신호 정보를 결정하는 프로세서 및

상기 결정된 백채널 신호 정보를 출력하는 출력부를 포함하는,

교감형 백채널 신호 생성 시스템.

청구항 11

제10항에 있어서,

상기 프로세서는 미리 정해진 제1 임계치에 상응하는 길이의 음성 신호가 입력된 경우, 해당 시점을 영상 기반의 채널 신호 출력시점으로 판단하는 것인,

교감형 백채널 신호 생성 시스템.

청구항 12

제11항에 있어서,

상기 미리 정해진 제1 임계치에 상응하는 길이는 고정된 시간 길이로 결정되거나, 인공지능 캐릭터의 페르소나 또는 사용자의 반응 정보에 따라 가변된 시간 길이로 결정되는 것인,

교감형 백채널 신호 생성 시스템.

청구항 13

제11항에 있어서,

상기 프로세서는 상기 제1 임계치에 상응하는 길이만큼 음성 신호가 입력된 이후, 미리 정해진 제2 임계치에 상응하는 길이만큼 묵음 구간이 입력된 경우, 상기 제2 임계치에 상응하는 길이만큼 묵음 구간이 검출된 시점을 영상 및 음성 기반의 백채널 신호 출력시점으로 판단하는 것인,

교감형 백채널 신호 생성 시스템.

청구항 14

제10항에 있어서,

상기 프로세서는 상기 저장된 음성 신호를 대상으로 음성 인식 결과를 생성하고, 상기 저장된 음성 신호 및 음성 인식 결과에 기초하여 사용자의 감정 상태 정보를 생성한 후, 상기 감정 상태 정보에 상응하는 영상 기반의 백채널 신호 정보를 결정하는 것인,

교감형 백채널 신호 생성 시스템.

청구항 15

제14항에 있어서,

상기 프로세서는 상기 음성 인식 결과에 따른 텍스트가 절 경계 지점 또는 문장 종료 지점인지 여부를 판단하고,

상기 판단 결과 상기 절 경계 지점 또는 문장 종료 지점에 해당하는 경우, 상기 영상 기반의 백채널 신호와 함께, 상기 저장된 음성 신호에 대한 반응 정보를 표출하기 위한 소정의 제1 길이 이상의 음성 기반의 백채널 신호 정보를 출력하고,

교감형 백채널 신호 생성 시스템.

청구항 16

제15항에 있어서,

상기 프로세서는 상기 판단 결과 상기 절 경계 지점 또는 문장 종료 지점이 아닌 경우, 상기 영상 기반의 백채널 신호와 함께, 사용자의 음성 신호를 계속 입력 받기 위한 상기 제1 길이보다 짧은 소정의 제2 길이 미만의 음성 기반의 백채널 신호 정보를 출력하는 것인,

교감형 백채널 신호 생성 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 교감형 백채널 신호 생성 방법 및 시스템에 관한 것이다.

배 경 기 술

[0002] 백채널 신호란 청자(Listener)가 화자(Speaker)에게 주의를 기울이고 있음을 나타내거나, 또는 화자에게 계속 말할 것을 요청하기 위해, 청자가 사용하는 짧은 발성, 얼굴 표정, 눈짓, 머리 움직임 또는 이들의 조합을 말한다. 사람 간의 대화에서는 통상 백채널 신호를 청자의 스타일에 따라 주기적으로 상대 화자에게 전달한다.

[0003] 한편, 최근 인공지능 기술의 발달에 따라 디지털 휴먼, 지능형 로봇, 음성 아바타 챗봇과 관련된 기술이 널리 확산되고 있다. 이러한 디지털 휴먼, 지능형 로봇, 음성 아바타 챗봇 등은 단순히 사용자와 대화만을 주고받는 것이 아니라, 이들이 스크린을 통한 캐릭터 등 실체화된 모습으로 존재하기 때문에, 사용자가 발화시 실제 인간 청자와 같이 시스템이 청자로서 백채널 신호를 전달해 주어야만 자연스러운 대화를 진행할 수 있다. 그리고 이 때 화자의 감정 등에 교감한 상태로 백채널 신호를 생성하고 대화를 진행해야만 자연스러운 대화가 가능하다.

[0004] 하지만, 현재 대부분의 디지털 휴먼, 지능형 로봇, 음성 아바타 챗봇 등은 백채널 신호를 전달하지 못하고 해당 캐릭터가 고정된 모습 또는 단순 동작의 반복 등만을 행하는 채로 음성 합성을 통해 단순하게 대화만을 진행하는데 그치고 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 공개특허공보 제10-2022-0094008호 (2022.07.05)

발명의 내용

해결하려는 과제

[0006] 본 발명이 해결하고자 하는 과제는 사용자로부터 입력되는 영상 또는 음성 정보를 기반으로 백채널 신호를 생성하고 출력하는 시점을 결정하여, 사용자로 하여금 인공지능 대화 시스템과 보다 자연스러운 대화가 가능하도록

하는 교감형 백채널 신호 생성 방법 및 시스템을 제공하는 것이다.

[0007] 다만, 본 발명이 해결하고자 하는 과제는 상기된 바와 같은 과제로 한정되지 않으며, 또다른 과제들이 존재할 수 있다.

과제의 해결 수단

[0008] 상술한 과제를 해결하기 위한 본 발명의 제1 측면에 따른 교감형 백채널 신호 생성 방법은 사용자로부터의 음성 신호를 입력받는 단계; 소정의 시점에서의 음성 신호가 입력됨에 따라 백채널 신호 출력시점인지 여부를 판단하는 단계; 상기 판단 결과 백채널 신호 출력시점에 해당하는 경우, 현재까지 입력된 음성 신호를 저장하는 단계; 상기 저장된 음성 신호에 기초하여 백채널 신호 정보를 결정하는 단계; 및 상기 결정된 백채널 신호 정보를 출력하는 단계를 포함한다.

[0009] 또한, 본 발명의 제2 측면에 따른 교감형 백채널 신호 생성 시스템은 사용자로부터 음성 신호를 입력받는 입력부, 상기 음성 신호를 분석하여 교감형 백채널 신호를 생성하기 위한 프로그램이 저장된 메모리, 소정의 시점에서의 음성 신호가 입력됨에 따라 백채널 신호 출력시점인 것으로 판단시, 현재까지 입력된 음성 신호를 저장하고, 저장된 음성 신호에 기초해여 백채널 신호 정보를 결정하는 프로세서 및 상기 결정된 백채널 신호 정보를 출력하는 출력부를 포함한다.

[0010] 상술한 과제를 해결하기 위한 본 발명의 다른 면에 따른 컴퓨터 프로그램은, 하드웨어인 컴퓨터와 결합되어 교감형 백채널 신호 생성 방법을 실행하며, 컴퓨터 판독가능 기록매체에 저장된다.

[0011] 본 발명의 기타 구체적인 사항들은 상세한 설명 및 도면들에 포함되어 있다.

발명의 효과

[0012] 전술한 본 발명의 일 실시예에 의하면, 인공지능 대화 시스템을 이용한 대화시 단순히 응답 정보만 출력하는 종래와는 달리, 교감형 백채널 신호 정보를 함께 출력함으로써 대화 시스템의 자연성 및 몰입도를 극대화시킬 수 있다.

[0013] 또한, 영상 정보 활용이 가능할 경우, 입력된 음성 및 텍스트 정보에 추가적으로 영상 정보를 활용함으로써, 백채널 신호 출력시점 결정 및 백채널 신호 정보 결정 과정에서의 정확도를 높일 수 있다.

[0014] 뿐만 아니라, 영상 정보만을 통하여 백채널 신호 정보를 출력할 수도 있고, 음성 및 영상 정보를 동시에 백채널 신호 정보로 출력할 수도 있도록 구성함으로써 디바이스의 종속성을 최소화시키며, 대화 시스템의 오류 발생 가능성을 줄이고, 자연성을 개선할 수 있는 효과가 있다.

[0015] 이러한 교감형 백채널 신호 정보를 생성 및 제공하는 것을 통해 본 발명의 일 실시예는 인간과 대화하는 듯한 자연스러운 인공지능 대화 시스템의 구성이 가능하도록 한다.

[0016] 본 발명의 효과들은 이상에서 언급된 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0017] 도 1은 본 발명의 일 실시예에 따른 교감형 백채널 신호 생성 시스템의 블록도이다.

도 2는 본 발명의 일 실시예에 따른 교감형 백채널 신호 생성 방법의 순서도이다.

도 3은 본 발명의 일 실시예에서 영상 백채널 신호 출력시점을 판단하는 과정의 순서도이다.

도 4는 본 발명의 일 실시예에서 영상 및 음성 백채널 신호 출력시점을 판단하는 과정의 순서도이다.

도 5는 본 발명의 일 실시예에서의 영상 백채널 신호 정보를 결정하는 과정의 순서도이다.

도 6은 본 발명의 일 실시예에서의 영상 및 음성 백채널 신호 정보를 결정하는 과정의 순서도이다.

도 7은 본 발명의 일 실시예에서 백채널 신호 정보 출력 과정을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0018] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시

예들을 참조하면 명확해질 것이다. 그러나, 본 발명은 이하에서 개시되는 실시예들에 제한되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야의 통상의 기술자에게 본 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.

[0019] 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 "포함한다(comprises)" 및/또는 "포함하는(comprising)"은 언급된 구성요소 외에 하나 이상의 다른 구성요소의 존재 또는 추가를 배제하지 않는다. 명세서 전체에 걸쳐 동일한 도면 부호는 동일한 구성 요소를 지칭하며, "및/또는"은 언급된 구성요소들의 각각 및 하나 이상의 모든 조합을 포함한다. 비록 "제1", "제2" 등이 다양한 구성요소들을 서술하기 위해서 사용되나, 이들 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소를 다른 구성요소와 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제1 구성요소는 본 발명의 기술적 사상 내에서 제2 구성요소일 수도 있음은 물론이다.

[0020] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 발명이 속하는 기술 분야의 통상의 기술자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또한, 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.

[0021] 이하 첨부된 도면을 참조하여 본 발명의 실시예들을 상세하게 설명하도록 한다.

[0022] 도 1은 본 발명의 일 실시예에 따른 교감형 백채널 신호 생성 시스템(100)의 블록도이다.

[0023] 본 발명의 일 실시예에 따른 교감형 백채널 신호 생성 시스템(100)은 입력부(110), 통신부(120), 메모리(130), 프로세서(140) 및 출력부(150)를 포함한다.

[0024] 입력부(110)는 사용자로부터 음성 신호를 수신하는 마이크를 포함한다. 그밖에 입력부(110)는 교감형 백채널 신호 생성 시스템(100)의 사용자 입력에 대응하여 입력 데이터를 발생시킬 수 있다. 입력부(110)는 적어도 하나의 입력수단을 포함할 수 있으며, 키보드(key board), 키패드(key pad), 돔 스위치(dome switch), 터치패널(touch panel), 터치 키(touch key), 마우스(mouse), 메뉴 버튼(menu button) 등을 포함할 수 있다.

[0025] 통신부(120)는 교감형 백채널 신호 생성 시스템(100) 내부 장치 또는 사용자 단말 등과 같은 외부 장치와의 통신을 수행한다. 이와 같은 통신부(120)는 유선 통신 모듈 및 무선 통신 모듈을 모두 포함할 수 있다. 유선 통신 모듈은 전력선 통신 장치, 전화선 통신 장치, 케이블 홈(MoCA), 이더넷(Ethernet), IEEE1294, 통합 유선 홈 네트워크 및 RS-485 제어 장치로 구현될 수 있다. 또한, 무선 통신 모듈은 WLAN(wireless LAN), Bluetooth, HDR WPAN, UWB, ZigBee, Impulse Radio, 60GHz WPAN, Binary-CDMA, 무선 USB 기술 및 무선 HDMI 기술, 그밖에 5G(5th generation communication), LTE-A(long term evolution-advanced), LTE(long term evolution), Wi-Fi(wireless fidelity) 등의 기능을 구현하기 위한 모듈로 구성될 수 있다.

[0026] 메모리(130)는 음성 신호를 분석하여 교감형 백채널 신호를 생성하기 위한 프로그램이 저장된다. 여기에서, 메모리(130)는 전원이 공급되지 않아도 저장된 정보를 계속 유지하는 비휘발성 저장장치 및 휘발성 저장장치를 통칭하는 것이다. 예를 들어, 메모리(140)는 콤팩트 플래시(compact flash; CF) 카드, SD(secure digital) 카드, 메모리 스틱(memory stick), 솔리드 스테이트 드라이브(solid-state drive; SSD) 및 마이크로(micro) SD 카드 등과 같은 낸드 플래시 메모리(NAND flash memory), 하드 디스크 드라이브(hard disk drive; HDD) 등과 같은 마그네틱 컴퓨터 기억 장치 및 CD-ROM, DVD-ROM 등과 같은 광학 디스크 드라이브(optical disc drive) 등을 포함할 수 있다.

[0027] 프로세서(140)는 프로그램 등 소프트웨어를 실행하여 교감형 백채널 신호 생성 시스템(100)의 적어도 하나의 다른 구성요소(예: 하드웨어 또는 소프트웨어 구성요소)를 제어할 수 있고, 다양한 데이터 처리 또는 연산을 수행할 수 있다. 프로세서(140)는 백채널 신호 정보를 생성하거나 출력 시점을 결정하기 위하여 소정의 학습된 인공지능 알고리즘을 이용할 수도 있다.

[0028] 출력부(150)는 결정된 백채널 신호 정보를 출력한다. 출력부(150)는 영상 기반으로 백채널 신호 정보를 출력하거나 또는 영상 및 음성 기반으로 백채널 신호 정보를 출력할 수 있다. 물론 음성 기반으로 백채널 신호 정보도 출력할 수 있다.

[0029] 한편, 본 발명의 설명에서는 교감형 백채널 신호 생성 시스템(100)과 인공지능 대화 시스템(미도시)는 각각 독

립된 구성인 것으로 설명하고 있으나 반드시 이에 한정되는 것은 아니다. 즉, 하나의 서버 컴퓨터 내에 별개의 프로그램 형태로 구성되거나, 교감형 백채널 신호 생성 시스템(100)은 인공지능 대화 시스템의 일부로 구성되는 등 다양한 형태로 실시 가능하다.

[0030] 이하에서는 도 2 내지 도 7을 참조하여 본 발명의 일 실시예에 따른 교감형 백채널 신호 생성 시스템에 의해 수행되는 방법에 대해 상세히 설명하도록 한다.

[0031] 도 2는 본 발명의 일 실시예에 따른 교감형 백채널 신호 생성 방법의 순서도이다.

[0032] 본 발명의 일 실시예는 먼저 마이크를 통해 사용자로부터 음성 신호를 입력받는다(S110).

[0033] 다음으로, 소정의 시점에서의 음성 신호가 입력됨에 따라 백채널 신호 출력시점인지 여부를 판단한다(S120).

[0034] 다음으로, 판단 결과 백채널 출력시점에 해당하는 경우, 현재까지 입력된 음성 신호를 저장한다(S130).

[0035] 다음으로, 저장된 음성 신호에 기초하여 어떤 백채널 신호를 출력할 것인지 백채널 신호 정보를 결정하고(S140), 결정된 백채널 신호 정보를 출력한다(S150).

[0036] 한편, 본 발명의 일 실시예에서 대상으로 하는 백채널 신호는 영상 기반의 백채널 신호(이하, 영상 백채널 신호)와 음성 기반의 백채널 신호(이하, 음성 백채널 신호)를 포함할 수 있다.

[0037] 영상 백채널 신호는 디지털 휴먼, 지능형 로봇, 음성 아바타 챗봇 등에서 캐릭터의 얼굴 표정, 몸짓, 제스처 등을 출력하는 백채널 신호(대표적으로 고개 끄덕임)이고, 음성 백채널 신호는 해당 캐릭터의 음성 합성을 통해 출력하는 백채널 신호이다. 음성 백채널 신호의 일 예로는, 대표적으로 한국어에서 경어체의 경우 '네', '그랬어요', '맞아요', 평어체의 경우 '어', '응', '그래', '맞아' 등과 영어에서 'Um', 'Hmm' 'Yeah', 'Right' 등이 이에 해당할 수 있다.

[0038] 본 발명의 일 실시예는 음성 합성시에도 영상 백채널 신호를 동시에 출력하기 때문에, 영상 백채널 신호와, 음성 및 영상 백채널 신호인 두 가지 백채널 신호로 운영되는 것을 기준으로 설명하도록 한다.

[0039] 도 3은 본 발명의 일 실시예에서 영상 백채널 신호 출력시점을 판단하는 과정의 순서도이다.

[0040] 일 실시예로, 사용자의 음성 신호가 입력되고 나면(S110), 입력된 신호가 음성인지 여부를 판단하여 음성을 검출한다(S121). 그리고 미리 정해진 제1 임계치에 상응하는 길이의 음성 신호가 입력되었는지 여부를 판단한다(S122). 판단 결과 제1 임계치에 상응하는 길이만큼 음성 신호가 입력된 경우(S122-Y), 해당 시점을 영상 백채널 신호 출력시점으로 판단할 수 있으며(S123), 현 시점까지 입력된 음성 신호를 저장한다(S130). 이때 카메라를 통해 촬영한 영상 신호가 존재하는 경우 영상 신호도 함께 저장할 수 있다.

[0041] 여기에서 미리 정해진 제1 임계치에 상응하는 길이는 고정된 시간 길이(3초, 5초 등)로 결정되거나, 또는 인공지능 캐릭터의 페르소나 또는 사용자의 반응 정보에 따라 가변된 시간 길이로 결정될 수 있다.

[0042] 이처럼, 본 발명의 일 실시예는 영상 백채널 신호 출력시점을 판단하는 과정을 통해, 사람 간의 대화에서와 마찬가지로 주기적으로 백채널 신호를 보낼 수 있다.

[0043] 한편, 위 과정에서 마이크로부터 입력되는 신호가 음성인지 아닌지는 음성 검출 기술(Voice Activity Detection)을 사용하여 판단할 수 있다. 만약, 카메라를 사용하는 경우라면, 사용자의 입이 말을 하기 위해 움직이는지를 입 모양 움직임이 포함된 영상 신호를 활용하여 동시에 음성을 검출하면 음성 검출 성능을 더욱 높일 수 있다.

[0044] 그밖에, 본 발명의 일 실시예는 사용자의 음성 신호가 입력되는 속도에 기반하여 영상 백채널 출력시점을 판단하기 위한 제1 임계치가 결정될 수 있다. 일 예로, 초기 제1 임계치가 고정된 3초로 설정된 경우, 3초 후에 영상 백채널 출력시점이 출력된다. 이때, 사용자로부터 피드백 정보를 직접 입력받아 해당 시점이 적절한 영상 백채널 출력시점이었는지 여부를 평가하는 정보를 입력받거나 또는 사용자의 영상 및 음성을 분석한 결과를 기반으로 피드백 정보를 생성하여 평가할 수도 있다. 이러한 피드백 정보에 따라 초기 설정된 제1 임계치에 상응하는 길이가 증감될 수 있으며, 복수 회 반복 수행됨에 따라 최적화된 영상 백채널 출력시점을 결정할 수 있다.

[0045] 또한, 본 발명의 일 실시예는 소정 시간 구간 동안 저장된 음성 신호를 기반으로 제1 영상 백채널 출력시점으로 판단된 후, 해당 시점을 기준으로 다음 시간 구간 동안 저장된 음성 신호를 기반으로 제2 영상 백채널 출력시점을 결정한다. 이 과정에서 제1 영상 백채널 출력시점이 결정됨과 동시에 이전 누적 저장된 음성 신호를 기반으로 다음 시간 구간에서의 예상되는 제2 영상 백채널 출력 시점을 추정할 수 있다. 그리고 실제 제2 영상 백채널

출력시점과 추정된 제2 영상 백채널 출력시점을 비교하여 유사도 또는 일치도에 따라 영상 백채널 출력시점 판단을 위한 알고리즘에 가중치를 부여할 수 있다. 여기에서 가중치는 음성 신호를 입력받기 위한 시간 구간의 증감을 위한 가중치일 수 있다. 이러한 가중치 기반의 학습 과정이 복수회 수행됨에 따라 더욱 정확한 영상 백채널 출력시점 판단이 가능하다.

[0046] 도 4는 본 발명의 일 실시예에서 영상 및 음성 백채널 신호 출력시점을 판단하는 과정의 순서도이다.

[0047] 도 3의 실시예와 마찬가지로, 사용자의 음성 신호가 입력되고 나면(S110), 입력된 신호가 음성인지 여부를 판단하여 음성을 검출한다(S126). 그리고 미리 정해진 제1 임계치에 상응하는 길이의 음성 신호가 입력되었는지 여부를 판단한다(S127). 판단 결과 제1 임계치에 상응하는 길이만큼 음성 신호가 입력된 경우(S127-Y), 제1 임계치에 상응하는 길이만큼 음성 신호가 입력된 이후, 미리 정해진 제2 임계치에 상응하는 길이만큼 묵음 구간이 입력되었는지 여부를 판단한다(S128). 판단 결과 제2 임계치에 상응하는 묵음 구간이 입력된 경우(S128-Y), 제2 임계치에 상응하는 길이만큼 묵음 구간이 검출된 시점을 영상 및 음성 백채널 신호 출력시점으로 판단할 수 있으며(S129), 현 시점까지 입력된 음성 신호를 저장한다(S130). 이때 카메라를 통해 촬영한 영상 신호가 존재하는 경우 영상 신호도 함께 저장할 수 있다.

[0048] 이때, 도 4의 실시예의 경우 음성 신호 입력 이후에 묵음 구간을 이어서 검출해야만 영상 및 음성 백채널 신호 출력시점으로 결정할 수 있으므로, 도 3의 영상 백채널 신호 출력시점에서의 제1 임계치보다 더 짧은 제1 임계치를 적용할 수 있다.

[0049] 또한, 묵음 구간의 충분한 입력이 있는지 여부를 판단하기 위한 제2 임계치는 고정된 시간 길이(0.5초 등)로 결정되거나, 또는 인공지능 캐릭터의 페르소나 또는 사용자의 반응 정보에 따라 가변된 시간 길이로 결정될 수 있다.

[0050] 본 발명의 일 실시예에서 영상 및 음성 백채널 신호 출력시점을 결정함에 있어서 묵음 구간을 검출 및 이용하는 것은 영상 백채널 신호와 함께 음성 신호 역시 백채널 신호로 출력해야 하기 때문에, 백채널 출력 타이밍을 사용자가 말을 마치거나 쉬고 있는 묵음 구간으로 설정하기 위해서이다.

[0051] 도 4의 실시예도 마찬가지로 마이크로부터 입력되는 신호가 음성인지 아닌지는 음성 검출 기술(Voice Activity Detection)을 사용하여 판단할 수 있다. 만약, 카메라를 사용하는 경우라면, 사용자의 입이 말을 하기 위해 움직이는지를 입 모양 움직임이 포함된 영상 신호를 활용하여 동시에 음성을 검출하면 음성 검출 성능을 더욱 높일 수 있다.

[0052] 도 5는 본 발명의 일 실시예에서의 영상 백채널 신호 정보를 결정하는 과정의 순서도이다.

[0053] 일 실시예로, 영상 백채널 신호 정보를 결정하는 단계(S140)에서는, 우선 저장된 음성 신호를 대상으로(S130) 음성 인식을 수행하여 음성 인식 결과를 생성한다(S141).

[0054] 그 다음, 저장된 음성 신호와 음성 인식 결과(텍스트)에 기초하여 사용자의 감정 상태 정보를 생성한다(S142). 이때, 카메라가 구비된 경우라면 영상 신호를 분석하는 과정을 추가적으로 수행하여 감정 상태 정보를 생성할 수 있다. 이 경우 모든 모달리티에서 입력되는 감정 인식 결과가 동일하면 해당 감정 인식 결과를 감정 상태 정보로 그대로 적용 가능하다. 이와 달리, 각 모달리티에서 상반되는 감정 인식 결과가 존재하는 경우, 더 많은 모달리티에서 출력한 감정 인식 결과를 감정 상태 정보로 적용할 수 있다. 또는, 각 모달리티에서의 감정 인식 결과에 대한 신뢰도나 강도를 산출하고, 이를 종합하여 가장 신뢰도나 강도가 높은 모달리티의 감정 인식 결과를 감정 상태 정보로 적용할 수 있다. 한편, 모달리티의 개수는 소정의 단위 음성 신호, 단위 영상 신호의 개수일 수 있다.

[0055] 위와 같은 방법에도 불구하고 감정 상태 정보의 결정이 어려운 경우에는 '중립'적인 결과로 감정 상태 정보를 결정할 수도 있다.

[0056] 다음으로, 감정 상태 정보에 상응하는 영상 백채널 신호 정보를 결정한다(S143).

[0057] 감정 상태 정보의 일 예로는 '중립', '행복', '슬픔' 및 '분노' 등이 있으며 분석 방법론에 따라 그 분류를 확대 또는 변경할 수 있음은 물론이다. 사용자의 감정 상태 정보가 '중립'일 경우 '중립'에 해당하는 영상 백채널 신호로 결정한다. 또는, 사용자의 감정 상태 정보가 '행복'일 경우 '행복'이나 '즐거움'에 해당하는 영상 백채널 신호로 결정하며, '슬픔' 이거나 '분노'일 경우 '위로'에 해당하는 영상 백채널 신호를 출력할 수 있도록 결정할 수 있다.

[0058] 한편, 결정되는 영상 백채널 신호 정보는 먼저 입력될 수 있는 감정 인식 범위가 설정된 다음, 이에 대응하는 영상 백채널 신호 정보가 매핑되는 타입으로 구성될 수 있다. 실제로 매핑시에는 스크린 등 장치에서 출력할 수 있는 표정, 몸짓, 동작 등의 영상 신호를 고려해야 한다.

[0059] 도 6은 본 발명의 일 실시예에서의 영상 및 음성 백채널 신호 정보를 결정하는 과정의 순서도이다.

[0060] 일 실시예로, 영상 및 음성 백채널 신호 정보를 결정하는 단계(S140)에서는, 도 5의 실시예와 마찬가지로 우선 저장된 음성 신호를 대상으로(S130) 음성 인식을 수행하여 음성 인식 결과를 생성한다(S145).

[0061] 그 다음, 저장된 음성 신호와 음성 인식 결과(텍스트)에 기초하여 사용자의 감정 상태 정보를 생성한다(S146). 이때, 카메라가 구비된 경우라면 영상 신호를 분석하는 과정을 추가적으로 수행하여 감정 상태 정보를 생성할 수 있다. 이 경우 모든 모달리티에서 입력되는 감정 인식 결과가 동일하면 해당 감정 인식 결과를 감정 상태 정보로 그대로 적용 가능하다. 이와 달리, 각 모달리티에서 상반되는 감정 인식 결과가 존재하는 경우, 더 많은 모달리티에서 출력한 감정 인식 결과를 감정 상태 정보로 적용할 수 있다. 또는, 각 모달리티에서의 감정 인식 결과에 대한 신뢰도나 강도를 산출하고, 이를 종합하여 가장 신뢰도나 강도가 높은 모달리티의 감정 인식 결과를 감정 상태 정보로 적용할 수 있다. 한편, 모달리티의 개수는 소정의 단위 음성 신호, 단위 영상 신호의 개수일 수 있다.

[0062] 위와 같은 방법에도 불구하고 감정 상태 정보의 결정이 어려운 경우에는 '중립'적인 결과로 감정 상태 정보를 결정할 수도 있다.

[0063] 감정 상태 정보 생성이 완료되면, 다음으로 음성 인식 결과에 따른 텍스트가 절 경계 지점 또는 문장 종료 지점인지 여부를 판단한다(S147).

[0064] 판단 결과, 절 경계 또는 문장 종료 지점에 해당하는 경우(S147-Y), 저장된 음성 신호에 대한 반응 정보를 표출하기 위한 소정의 제1 길이 이상의 백채널 신호 정보를 출력할 수 있다(S148). 이는 문장의 끝이나 절 경계 등에서는 '이해(Understanding)' 또는 '동의(Agreement)'를 의미하는 긴 길이의 '네', '응', 또는 '그랬어요', '맞아요' 등과 같은 백채널 신호를 출력하도록 하기 위함이다.

[0065] 이와 달리, 판단 결과 절 경계 또는 문장 종료 지점에 해당하지 않는 경우(S147-N), 사용자의 음성 신호를 계속 입력 받기 위한 제1 길이보다 짧은 소정의 제2 길이 미만의 백채널 신호 정보를 출력할 수 있다(S149). 이 경우에 적용되는 백채널 신호 자체는 대화의 주도권을 가져오지 않은 상태로 사용자에게 계속 대화를 진행할 것을 독려하는 신호이므로, 음성 백채널 신호를 출력할 때 문장의 중간에서는 '계속(Continuer)'을 의미하는 길이가 짧은 '네', '어' 등과 같은 백채널 신호를 출력할 수 있다.

[0066] 한편, 본 발명의 일 실시예에서는 '계속' 백채널 신호 정보와 '이해' 백채널 신호 정보 두 가지 음성 백채널 신호를 결정하는 것을 예를 들어 설명하고 있으나, 반드시 이에 한정되는 것은 아니다. 즉, 음성 인식 결과에 대한 의미 분석 정보를 추출하고, 의미 분석 정보를 통해 '놀람', '확인' 등 결정할 수 있는 음성 백채널 신호의 정보를 추가할 수도 있다.

[0067] 또한, 인공지능 대화 시스템의 구성에 있어 백채널 신호 정보와 실제 대화 시스템의 응답이 동시에 출력될 수도 있는데, 이는 대화 시스템의 정책에 따라 어떤 신호를 우선시할지 또는 동시에 출력할지를 결정하도록 한다.

[0068] 또한, 본 발명의 일 실시예는 음성 백채널 신호를 결정할 때 감정 상태 정보를 활용하여 영상 백채널 신호도 함께 결정할 수 있다.

[0069] 도 5의 실시예와 마찬가지로 감정 상태 정보의 일 예로는 '중립', '행복', '슬픔' 및 '분노' 등이 있으며 분석 방법론에 따라 그 분류를 확대 또는 변경할 수 있음은 물론이다. 사용자의 감정 상태 정보가 '중립'일 경우 '중립'에 해당하는 영상 백채널 신호로 결정한다. 또는, 사용자의 감정 상태 정보가 '행복'일 경우 '행복'이나 '즐거움'에 해당하는 영상 백채널 신호로 결정하며, '슬픔' 이거나 '분노'일 경우 '위로'에 해당하는 영상 백채널 신호를 출력할 수 있도록 결정할 수 있다.

[0070] 한편, 결정되는 영상 백채널 신호 정보는 먼저 입력될 수 있는 감정 인식 범위가 설정된 다음, 이에 대응하는 영상 백채널 신호 정보가 매핑되는 타입으로 구성될 수 있다. 실제로 매핑시에는 스크린 등 장치에서 출력할 수 있는 표정, 몸짓, 동작 등의 영상 신호를 고려해야 한다.

[0071] 도 7은 본 발명의 일 실시예에서 백채널 신호 정보 출력 과정을 설명하기 위한 도면이다.

[0072] 일 실시예로, 백채널 신호 정보가 결정되고 나면, 결정된 백채널 신호 정보를 출력한다(S151, S152). 실제 백채

널 신호 정보를 출력할 때는 출력부(150)의 디스플레이 영역에 결정된 영상 백채널 신호 정보를 전달하여 기본적인 고개 끄덕임, 표정, 몸짓, 동작 등이 이에 대응하여 출력될 수 있도록 한다.

[0073] 예를 들어, '중립'에 해당하는 영상 백채널 신호 정보가 결정된 경우, 고개를 끄덕이며 캐릭터가 기본 상태의 표정, 몸짓, 눈짓 등을 하도록 영상 백채널 신호 정보를 출력한다.

[0074] 또는, '행복'에 해당하는 영상 백채널 신호 정보가 결정된 경우, 캐릭터가 고개를 끄덕이면서 행복한 표정, 몸짓, 눈짓 등을 하도록 영상 백채널 신호를 출력한다.

[0075] 또는, '위로'에 해당하는 영상 백채널 신호 정보가 결정된 경우, 캐릭터가 고개를 끄덕이면서 함께 슬퍼하거나 공감하는 표정, 몸짓, 눈짓 등을 하도록 영상 백채널 신호를 출력한다.

[0076] 이때, 본 발명의 일 실시예는 캐릭터의 형상, 또는 인공지능 대화 시스템의 정책에 따라 다양한 방식으로 캐릭터의 영상 백채널 신호를 출력할 수 있다.

[0077] 음성 및 영상 백채널 신호 정보를 출력할 경우, 영상 백채널 신호 정보는 동일한 방식으로 출력할 수 있다. 그리고 음성 백채널 신호 정보는 사용자의 음성 신호를 '계속' 입력받도록 유도하기 위한 백채널 신호 정보일 경우, 짧은 길이의 상승조 억양으로 '네', '응' 등이 출력되도록 한다. 또한, 사용자의 음성 신호에 대해 반응 정보(예를 들어, '이해')를 표출하기 위한 백채널 신호 정보일 경우, 비교적 긴 길이의 하강조 억양으로 '네', '응', '그래', '맞아' 등이 출력되도록 한다.

[0078] 또한, 본 발명의 일 실시예에 따른 교감형 백채널 신호 생성 시스템이 감정 표현이 가능한 음성 합성기를 더 포함하는 경우, 분석된 감정 상태 정보를 반영하여 음성 합성기를 통해 해당 감정을 반영한 음성이 합성될 수 있도록 할 수 있다. 음성 합성기를 통해 출력되는 음성 텍스트 내용 자체는 캐릭터의 스타일과 인공지능 대화 시스템의 성격을 반영해 랜덤하게 출력하거나 규칙에 기반하여 출력하도록 할 수 있으며, 데이터를 충분히 확보할 수 있다면 딥러닝 기반의 기계 학습을 이용하는 방식도 가능하다.

[0079] 한편, 상술한 설명에서, 단계 S110 내지 S160는 본 발명의 구현예에 따라서, 추가적인 단계들로 더 분할되거나, 더 적은 단계들로 조합될 수 있다. 또한, 일부 단계는 필요에 따라 생략될 수도 있고, 단계 간의 순서가 변경될 수도 있다. 아울러, 도 1의 내용과 도 2 내지 도 7의 내용은 상호 적용될 수 있다.

[0080] 이상에서 전술한 본 발명의 일 실시예는, 하드웨어인 컴퓨터와 결합되어 실행되기 위해 프로그램(또는 어플리케이션)으로 구현되어 매체에 저장될 수 있다.

[0081] 상기 전술한 프로그램은, 상기 컴퓨터가 프로그램을 읽어 들여 프로그램으로 구현된 상기 방법들을 실행시키기 위하여, 상기 컴퓨터의 프로세서(CPU)가 상기 컴퓨터의 장치 인터페이스를 통해 읽힐 수 있는 C, C++, JAVA, Ruby, 기계어 등의 컴퓨터 언어로 코드화된 코드(Code)를 포함할 수 있다. 이러한 코드는 상기 방법들을 실행하는 필요한 기능들을 정의한 함수 등과 관련된 기능적인 코드(Functional Code)를 포함할 수 있고, 상기 기능들을 상기 컴퓨터의 프로세서가 소정의 절차대로 실행시키는데 필요한 실행 절차 관련 제어 코드를 포함할 수 있다. 또한, 이러한 코드는 상기 기능들을 상기 컴퓨터의 프로세서가 실행시키는데 필요한 추가 정보나 미디어가 상기 컴퓨터의 내부 또는 외부 메모리의 어느 위치(주소 번지)에서 참조되어야 하는지에 대한 메모리 참조관련 코드를 더 포함할 수 있다. 또한, 상기 컴퓨터의 프로세서가 상기 기능들을 실행시키기 위하여 원격(Remote)에 있는 어떠한 다른 컴퓨터나 서버 등과 통신이 필요한 경우, 코드는 상기 컴퓨터의 통신 모듈을 이용하여 원격에 있는 어떠한 다른 컴퓨터나 서버 등과 어떻게 통신해야 하는지, 통신 시 어떠한 정보나 미디어를 송수신해야 하는지 등에 대한 통신 관련 코드를 더 포함할 수 있다.

[0082] 상기 저장되는 매체는, 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상기 저장되는 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있지만, 이에 제한되지 않는다. 즉, 상기 프로그램은 상기 컴퓨터가 접속할 수 있는 다양한 서버 상의 다양한 기록매체 또는 사용자의 상기 컴퓨터상의 다양한 기록매체에 저장될 수 있다. 또한, 상기 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장될 수 있다.

[0083] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가

지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

[0084] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

[0085] 100: 교감형 백채널 신호 생성 시스템

110: 입력부

120: 통신부

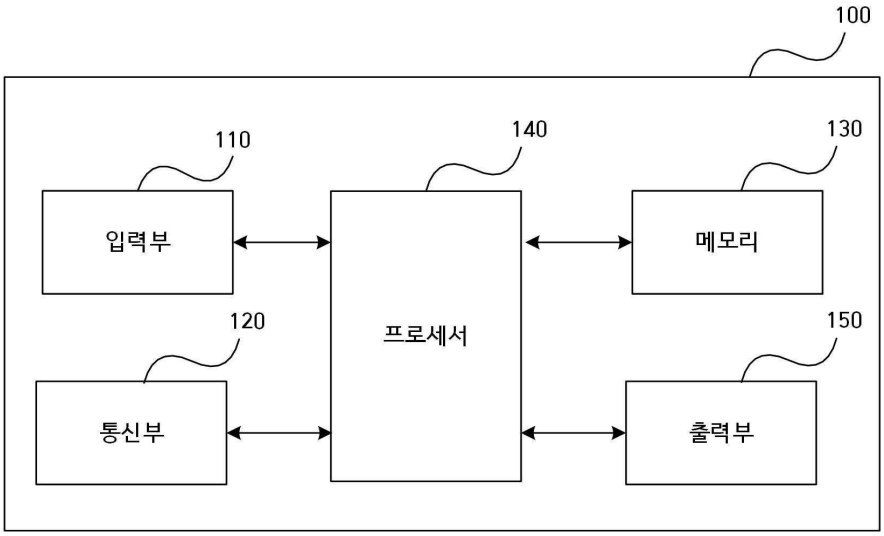
130: 메모리

140: 프로세서

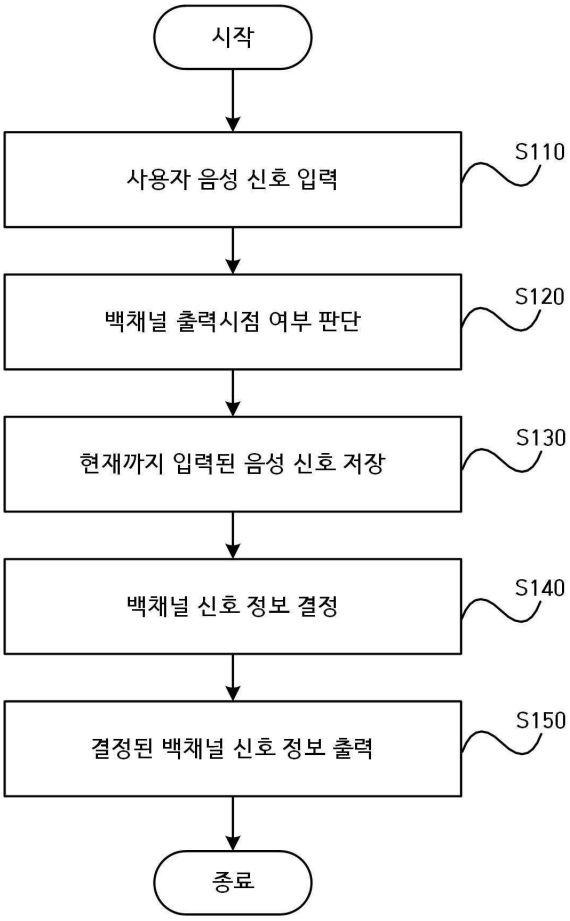
150: 출력부

도면

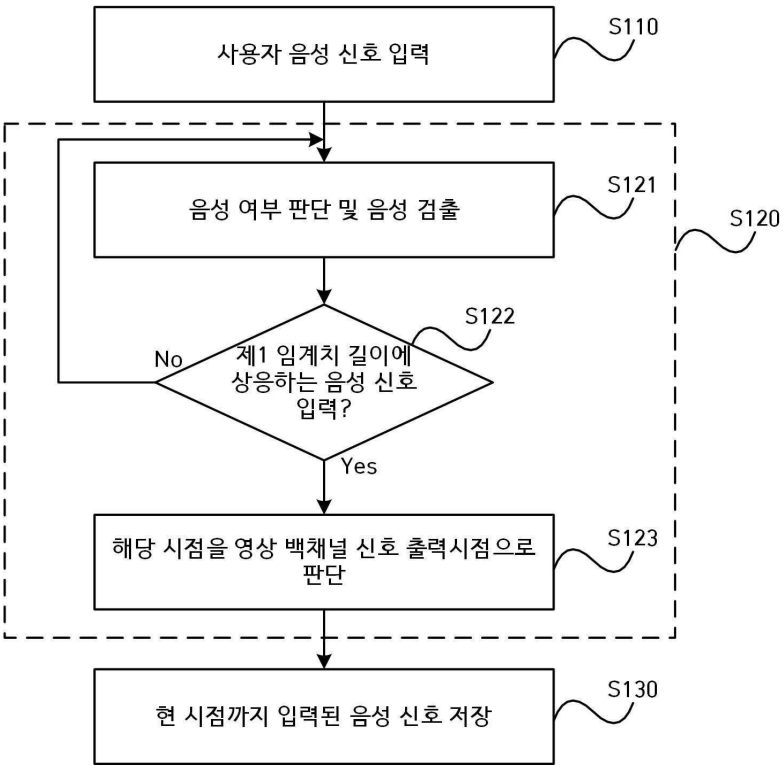
도면1



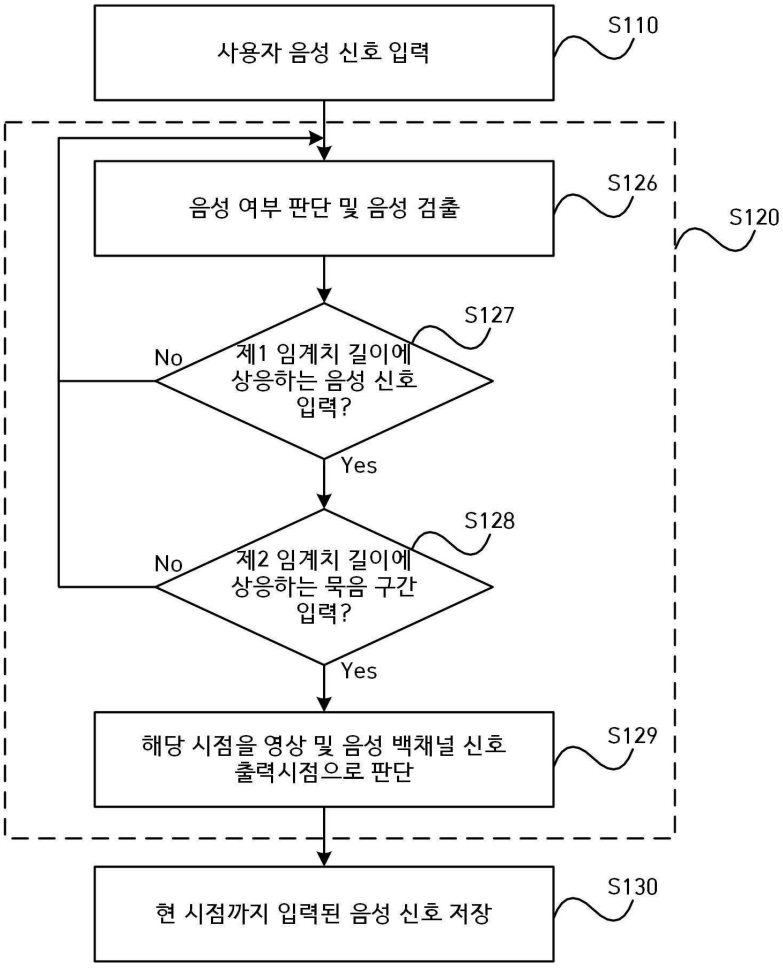
도면2


시작
사용자 음성 신호 입력
S110
백채널 출력시점 여부 판단
S120
현재까지 입력된 음성 신호 저장
S130
백채널 신호 정보 결정
S140
결정된 백채널 신호 정보 출력
S150
종료

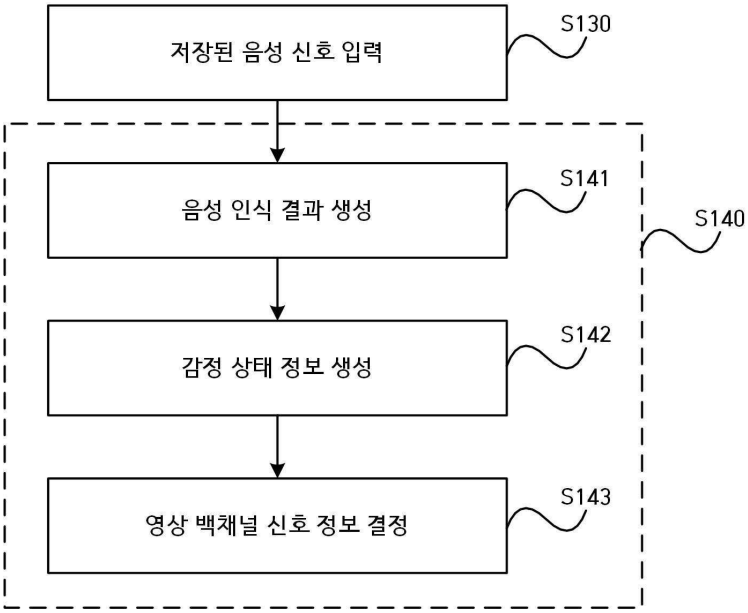
도면3


사용자 음성 신호 입력
S110
음성 여부 판단 및 음성 검출
S121
S120
S122
No
제1 임계치 길이에 상응하는 음성 신호 입력?
Yes
해당 시점을 영상 백채널 신호 출력시점으로 판단
S123
현 시점까지 입력된 음성 신호 저장
S130

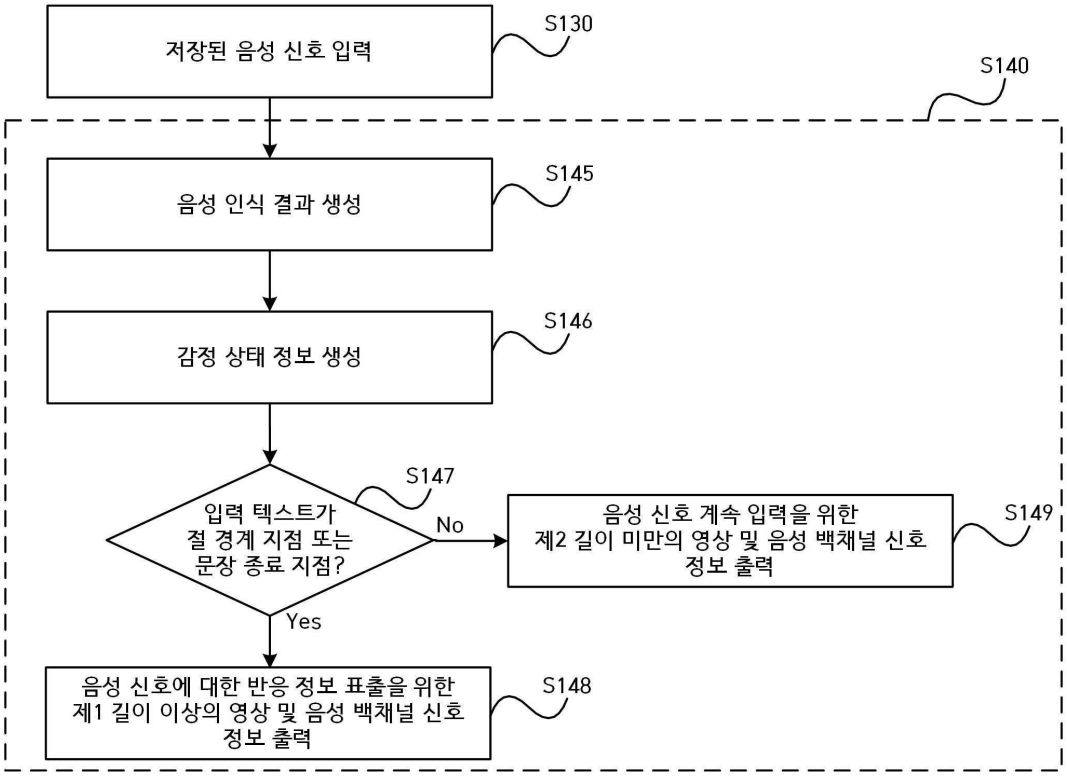
도면4


사용자 음성 신호 입력
S110
음성 여부 판단 및 음성 검출
S126
S120
S127
No
제1 임계치 길이에 상응하는 음성 신호 입력?
Yes
S128
No
제2 임계치 길이에 상응하는 묵음 구간 입력?
Yes
해당 시점을 영상 및 음성 백채널 신호 출력시점으로 판단
S129
현 시점까지 입력된 음성 신호 저장
S130

도면5


저장된 음성 신호 입력
S130
음성 인식 결과 생성
S141
S140
감정 상태 정보 생성
S142
영상 백채널 신호 정보 결정
S143

도면6


저장된 음성 신호 입력
S130
S140
음성 인식 결과 생성
S145
감정 상태 정보 생성
S146
입력 텍스트가
절 경계 지점 또는
문장 종료 지점?
S147
No
음성 신호 계속 입력을 위한
제2 길이 미만의 영상 및 음성 백채널 신호
정보 출력
S149
Yes
음성 신호에 대한 반응 정보 표출을 위한
제1 길이 이상의 영상 및 음성 백채널 신호
정보 출력
S148

도면7