



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2019년09월06일
(11) 등록번호 10-2019470
(24) 등록일자 2019년09월02일

(51) 국제특허분류(Int. Cl.)
G10L 25/63 (2013.01) G10L 15/02 (2006.01)
G10L 15/04 (2006.01)
(52) CPC특허분류
G10L 25/63 (2013.01)
G10L 15/02 (2013.01)
(21) 출원번호 10-2017-0170416
(22) 출원일자 2017년12월12일
심사청구일자 2017년12월12일
(65) 공개번호 10-2019-0069992
(43) 공개일자 2019년06월20일
(56) 선행기술조사문헌
JP2011237957 A*
(뒷면에 계속)

(73) 특허권자
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
권순일
서울특별시 강남구 압구정로 201, 77동 1402호 (압구정동, 현대아파트)
손귀영
서울특별시 관악구 장군봉7길 9, 202호 (봉천동)
안서연
부산광역시 해운대구 대천로 35, 108동 901호 (좌동, 코오롱아파트)
(74) 대리인
특허법인엠에이피에스

전체 청구항 수 : 총 10 항

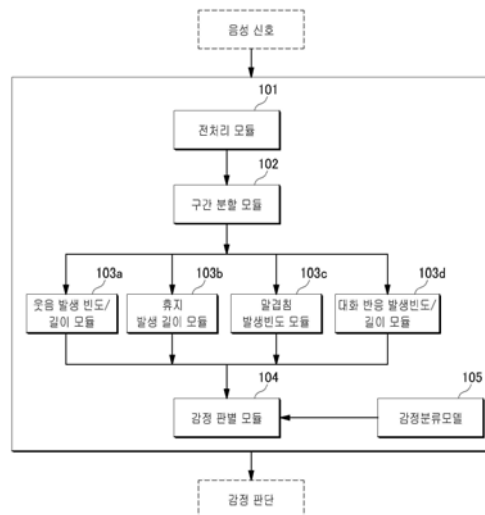
심사관 : 김응권

(54) 발명의 명칭 음성 발화 양식을 이용한 발화자 감정인식 방법 및 시스템

(57) 요약

본 발명의 일 실시예에 따른, 음성 발화 양식에 따른 발화자의 감정인식 방법에 있어서, (a) 제 1 사용자의 음성 신호, 제 1 사용자 및 제 1 사용자와 서로 다른 제 2 사용자 간 대화 음성 신호를 수신하는 단계; (b) 음성 신호에서 제 1 특징 요소를 추출하고, 대화 음성 신호에서 제 2 특징 요소를 추출하는 단계; 및 (c) 제 1 특징 요소 및 제 2 특징 요소에 기초하여 제 1 사용자의 감정을 판단하는 단계;를 포함하는 것인, 음성 발화 양식에 따른 발화자의 감정인식 방법일 수 있다.

대표도 - 도1



(52) CPC특허분류
G10L 15/04 (2013.01)

(56) 선행기술조사문헌
KR1020170083391 A*
KR1020090034052 A
KR1020100094182 A
KR1020110017559 A
*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호 1711056622

부처명 과학기술정보통신부

연구관리전문기관 정보통신기술진흥센터

연구사업명 방송통신산업기술개발

연구과제명 감성미디어 서비스를 위한 음성기반 감정인식 및 인텍싱 기술 개발

기 여 율 1/1

주관기관 세종대학교 산학협력단

연구기간 2017.04.01 ~ 2019.12.31

명세서

청구범위

청구항 1

음성 발화 양식에 따른 발화자의 감정인식 방법에 있어서,

(a) 제 1 사용자의 음성 신호, 상기 제 1 사용자 및 상기 제 1 사용자와 서로 다른 제 2 사용자 간 대화 음성 신호를 수신하는 단계;

(b) 상기 제 1 사용자의 음성 신호에서 제 1 특징 요소를 추출하고, 상기 대화 음성 신호에서 제 2 특징 요소를 추출하는 단계; 및

(c) 상기 제 1 특징 요소 및 제 2 특징 요소에 기초하여 상기 제 1 사용자의 감정을 판단하는 단계;

를 포함하고,

상기 제 1 특징 요소는 웃음 빈도수, 웃음 길이 및 휴지기를 포함하고,

상기 제 2 특징 요소는 상기 제 1 사용자 및 제 2 사용자의 말걸침 빈도수 및 길이, 대화 반응 횟수 및 대화 반응 길이를 포함하되,

상기 (c) 단계는

상기 특징 요소들에 기초하여 배일런스(valence)-어라우절(arousal) 감정 모델 상의 배일런스 값 및 어라우절 값을 결정하여 상기 제 1 사용자의 감정을 판단하는 것인, 음성 발화 양식에 따른 발화자의 감정인식 방법.

청구항 2

삭제

청구항 3

제 1 항에 있어서,

상기 (c) 단계는

상기 특징 요소들에 기초하여, 상기 제 1 사용자의 배일런스 값을 포지티브 배일런스(positive valence) 또는 네거티브 배일런스(negative valence) 중 어느 하나에 해당하는 값으로 결정하고, 상기 제 1 사용자의 어라우절 값을 하이 어라우절(high arousal) 또는 로우 어라우절(low arousal) 중 어느 하나에 해당하는 값으로 결정하는 것인, 음성 발화 양식에 따른 발화자의 감정인식 방법.

청구항 4

제 3 항에 있어서,

상기 (b) 단계는

상기 제 1 특징 요소 중 웃음 빈도수 및 웃음 길이에 기초하여 상기 배일런스 값을 결정하고, 상기 제 1 특징 요소 중 휴지기에 기초하여 상기 어라우절 값을 결정하는 것인, 음성 발화 양식에 따른 발화자의 감정인식 방법.

청구항 5

제 3 항에 있어서,

상기 (b) 단계는

상기 웃음 빈도수가 기 설정된 기준 웃음 빈도수를 초과하는 경우, 초과하는 정도에 기초하여 상기 배일런스 값을 포지티브 배일런스 값으로 결정하는 것인, 음성 발화 양식에 따른 발화자의 감정인식 방법.

청구항 6

제 3 항에 있어서,

상기 (b) 단계는

상기 휴지기가 기 설정된 기준 휴지 길이보다 짧다면, 상기 휴지기의 길이에 기초하여 하이 어라우절 값을 결정하고, 상기 기준 휴지 길이보다 길다면 상기 휴지기의 길이에 기초하여 로우 어라우절 값을 결정하는 것인, 음성 발화 양식에 따른 발화자의 감정인식 방법.

청구항 7

제 3 항에 있어서,

상기 (c) 단계는

상기 제 2 특징 요소 중 말접침 빈도수 및 길이에 기초하여 배일런스 및 어라우절 값을 결정하고, 상기 제 2 특징 요소 중 대화 반응의 빈도수 및 반응 시간에 기초하여 상기 배일런스 값을 결정하는 것인, 음성 발화 양식에 따른 발화자의 감정인식 방법.

청구항 8

제 7 항에 있어서,

상기 (c) 단계는

상기 말접침 빈도수 및 길이에 기초하여, 네거티브 배일런스 및 하이 어라우절을 결정하는 것인, 음성 발화 양식에 따른 발화자의 감정인식 방법.

청구항 9

제 7 항에 있어서,

상기 (c) 단계는

상기 대화 반응의 횟수가 기 설정된 기준 이상이며, 상기 대화 반응의 길이가 기 설정된 기준 이하이면, 상기 대화 반응 횟수 및 길이에 기초하여, 하이 어라우절 값을 결정하는 것인, 음성 발화 양식에 따른 발화자의 감정인식 방법.

청구항 10

음성 발화 양식에 따른 발화자의 감정인식을 제공하는 서버에 있어서,

음성 발화 양식에 따른 발화자의 감정인식 방법을 수행하기 위한 프로그램이 기록된 메모리; 및

상기 프로그램을 실행하기 위한 프로세서;를 포함하며,

상기 프로세서는, 상기 프로그램의 실행에 의해,

(a) 제 1 사용자의 음성 신호, 상기 제 1 사용자 및 상기 제 1 사용자와 서로 다른 제 2 사용자 간 대화 음성 신호를 수신하는 단계;

(b) 상기 음성 신호에서 제 1 특징 요소를 추출하고, 상기 대화 음성 신호에서 제 2 특징 요소를 추출하는 단계; 및

(c) 상기 제 1 특징 요소 및 제 2 특징 요소에 기초하여 상기 제 1 사용자의 감정을 판단하는 단계;

를 실행하고,

상기 제 1 특징 요소는 웃음 빈도수, 웃음 길이 및 휴지기를 포함하고,

상기 제 2 특징 요소는 상기 제 1 사용자 및 제 2 사용자의 말접침 빈도수 및 길이, 대화 반응 횟수 및 대화 반응 길이를 포함하되,

상기 (c) 단계는

상기 특징 요소들에 기초하여 배일런스(valence)-어라우절(arousal) 감정 모델 상의 배일런스 값 및 어라우절 값을 결정하여 상기 제 1 사용자의 감정을 판단하는, 서버.

청구항 11

제 1 항, 제 3 항 내지 제 9 항 중 어느 한 항에 따른 음성 발화 양식에 따른 발화자의 감정인식 방법을 수행하는 프로그램이 기록된 컴퓨터 판독가능 기록 매체.

발명의 설명

기술 분야

[0001] 본 발명은 음성의 발화에 따른 행동 양상을 이용하여 발화자의 감정을 인식하는 장치 및 방법에 관한 것이다.

배경 기술

[0002] 기존의 음성을 활용한 감정인식 기술들은 음성 신호 중 가장 기본적인 피치나 각종 주파수 정보에 기반하여 음향 모델을 만들고, 이를 이용하여 사용자의 감정을 인식하고 분류하고자 하였다.

[0003] 하지만 음성 정보만을 활용하여 감정을 판단하는 것에는 인식률의 한계가 존재하기에 이를 상호 보완할 수 있는 음성 신호 이외에 새로운 특징요소를 필요로 하고 있는 추세이다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) 공개특허공보 제 10-2014-0050130 호(명칭: 사용자 음성으로부터 감정정보를 확인하는 방법 및 장치)

발명의 내용

해결하려는 과제

[0005] 본 발명은 전술한 문제점을 해결하기 위하여, 본 발명의 일 실시예에 따라, 발화자의 음성으로부터 발화자의 감정을 추출하고, 발화의 행동 양식적 특징요소에 따른 발화자의 감정을 분류하는 것에 목적을 두고 있다.

과제의 해결 수단

[0006] 상기한 기술적 과제를 달성하기 위한 기술적 수단으로서 본 발명의 일 실시예에 따른, 음성 발화 양식에 따른 발화자의 감정인식 방법에 있어서, (a) 제 1 사용자의 음성 신호, 제 1 사용자 및 제 1 사용자와 서로 다른 제 2 사용자 간 대화 음성 신호를 수신하는 단계; (b) 음성 신호에서 제 1 특징 요소를 추출하고, 대화 음성 신호에서 제 2 특징 요소를 추출하는 단계; 및 (c) 제 1 특징 요소 및 제 2 특징 요소에 기초하여 제 1 사용자의 감정을 판단하는 단계;를 포함하는 것인, 음성 발화 양식에 따른 발화자의 감정인식 방법일 수 있다.

[0007] 또한, 다른 일 실시예로 음성 발화 양식에 따른 발화자의 감정인식을 제공하는 서버에 있어서, 음성 발화 양식에 따른 발화자의 감정인식 방법을 수행하기 위한 프로그램이 기록된 메모리; 및 프로그램을 실행하기 위한 프로세서;를 포함하며, 프로세서는, 프로그램의 실행에 의해, (a) 제 1 사용자의 음성 신호, 제 1 사용자 및 제 1 사용자와 서로 다른 제 2 사용자 간 대화 음성 신호를 수신하는 단계; (b) 음성 신호에서 제 1 특징 요소를 추출하고, 대화 음성 신호에서 제 2 특징 요소를 추출하는 단계; 및 (c) 제 1 특징 요소 및 제 2 특징 요소에 기초하여 제 1 사용자의 감정을 판단하는 단계;를 실행하는 명령어를 을 포함하는, 서버를 포함할 수 있다.

발명의 효과

[0008] 전술한 본 발명의 과제 해결 수단에 의하면, 음성 발화의 행동 양식에 따른 발화자의 감정인식을 제공하는 장치 및 방법을 사용하면, 발화자의 음성에서 나타나는 행동 양식의 특징 요소를 통하여 감정을 파악하고, 이를 이용하여 음성 정보와 융합한 실시간 개인 맞춤형 서비스를 제공할 수 있게 되는 것이다.

도면의 간단한 설명

- [0009] 도 1은 본 발명의 일 실시예에 따른, 음성 발화 양식에 따른 발화자의 감정인식 시스템을 도식화한 블록도이다.
- 도 2는 본 발명의 일 실시예에 따른, 베일런스-어라우절 감상 모델을 나타낸 예시도이다.
- 도 3은 본 발명의 일 실시예에 따른, 웃음 발생 빈도/길이 모듈(103a)을 이용한, 특징 요소를 추출하는 과정을 나타낸 동작 흐름도이다.
- 도 4는 본 발명의 일 실시예에 따른, 휴지 발생 길이 모듈(103b)을 이용한, 특징 요소를 추출하는 과정을 나타낸 동작 흐름도이다.
- 도 5는 본 발명의 일 실시예에 따른, 말걸침 발생 빈도 모듈(103c)을 이용한, 특징 요소를 추출하는 과정을 나타낸 동작 흐름도이다.
- 도 6은 본 발명의 일 실시예에 따른, 대화 반응 발생 빈도/길이 모듈(103d)을 이용한, 특징 요소를 추출하는 과정을 나타낸 동작 흐름도이다.
- 도 7은 본 발명의 일 실시예에 따른, 음성 발화 양식에 따른 발화자의 감정인식 방법을 나타낸 동작 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0010] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0011] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미하며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0012] 이하의 실시예는 본 발명의 이해를 돕기 위한 상세한 설명이며, 본 발명의 권리 범위를 제한하는 것이 아니다. 따라서 본 발명과 동일한 기능을 수행하는 동일 범위의 발명 역시 본 발명의 권리 범위에 속할 것이다.
- [0013] 도 1은 본 발명의 일 실시예에 따른, 음성 발화 양식에 따른 발화자의 감정인식 시스템을 도식화한 블록도이다.
- [0014] 도 1을 살펴보면 본 발명의 일 실시예에 따른, 음성 발화 양식에 따른 발화자의 감정 인식 시스템은 전처리 모듈(101), 구간 분할 모듈(102), 웃음 발생 빈도/길이 모듈(103a), 휴지 발생 길이 모듈(103b), 말걸침 발생빈도 모듈(103c), 대화 반응 발생 빈도/길이 모듈(103d), 감정 판별 모듈(104), 감정분류모델(105)을 포함할 수 있다.
- [0015] 전처리 모듈(101)은 사용자로부터 수신한 음성 신호에서 특징 요소를 추출하기 쉽도록 발화 음성을 가공하는 작업을 수행한다.
- [0016] 구간 분할 모듈(102)은 음성 신호에서 일정한 특징 요소를 추출할 수 있도록 음성 신호의 구간을 나누어 추출하는 역할을 수행한다.
- [0017] 다음은 특징 요소를 추출하는 4가지 모듈인 웃음 발생 빈도/길이 모듈(103a), 휴지 발생 길이 모듈(103b), 말걸침 발생 빈도 모듈(103c), 대화 반응 발생 빈도/길이 모듈(103d)이 있을 수 있다.
- [0018] 먼저 1인의 발화 음성 신호는 웃음 발생 빈도/길이 모듈(103a)과 휴지 발생 길이 모듈(103b)를 사용하여 특징 요소를 추출하고, 2인 이상의 대화는 앞의 두 모듈을 포함한, 말걸침 발생 빈도 모듈(103c)과 대화 반응 발생 빈도/길이 모듈(103d)을 추가적으로 사용된다.
- [0019] 자세히 설명하면, 웃음 발생 빈도/길이 모듈(103a)은 발화 음성 신호 중 발생하는 웃음을 인식하여 처리하는 모듈로, 웃음은 사람의 마음을 표정변화나 소리로 나타내는 방식의 하나로써 즐거움이나 여유 등을 표현하는 상태

를 지칭한다. 상세한 동작방법은 도 3을 참조하여 후술하도록 한다.

- [0020] 휴지기란 일시적인 조음활동이 중지하는 것을 일컫는 말로 단어와 단어, 어절, 문장 사이에 나타나는 쉼을 말하며, 휴지 발생 길이 모듈(103b)은 이러한 조음 활동이 끊어지는 구간의 길이를 분석하게 된다. 상세한 동작방법은 도 4를 참조하여 후술하도록 한다.
- [0021] 말걸침이란 첫 발화자의 발화가 종료되기 전 다음 발화자가 말을 끼어드는 현상을 말하며, 말걸침 발생 빈도 모듈(103c)을 통해 두 발화자의 대화에서 몇 번의 말걸침이 있었는지를 확인하게 된다. 해당 모듈의 상세한 동작방법은 도 5를 참조하여 후술하도록 한다.
- [0022] 대화 반응이란 첫 발화자의 발화가 종료된 직후 다음 발화자가 발화를 시작하기 전까지의 중간단계를 말한다. 대화 반응 발생 빈도/길이 모듈은 이러한 대화 반응이 발생하는 횟수와 그 구간의 길이를 확인하는 모듈로, 도 6을 참조하여 후술하도록 한다.
- [0023] 해당 모듈을 바탕으로 추출하는 특징 요소에 기초하여 배일런스(Valence)-어라우절(Arousal) 감상 모델상의 수치를 정하고, 사용자의 감정을 판단하는 척도로 사용될 수 있다. 배일런스-어라우절 감상 모델의 상세한 설명은 도 2를 참조하여 후술하도록 한다.
- [0024] 감정 판별 모듈(104)은 앞에서 추출된 발화 음성의 특징 요소를 통해 사용자의 감정을 판단하는 역할을 수행하고, 감정에 대한 판단을 내리기 위해 감정 분류 모델(105)이 사용될 수 있다.
- [0025] 감정 분류 모델(105)은 감정 판별 모듈(104)이 특징 요소를 통해 사용자의 감정을 판단할 수 있는 기준을 제시하는 일종의 데이터베이스이며, 예컨대 IEMOCAP(Interactive Emotional Dyadic Motion Capture) 데이터베이스가 사용될 수 있다.
- [0026] 도 2는 본 발명의 일 실시예에 따른, 배일런스-어라우절 감상 모델을 나타낸 예시도이다.
- [0027] 도 2를 살펴보면 본 발명의 일 실시예에 따른, 배일런스-어라우절로 이루어진 그래프를 확인할 수 있다.
- [0028] 배일런스는 사용자의 긍정 및 부정적인 상태를 측정한 것을 의미하고, 어라우절은 사용자의 각성도 혹은 흥분도를 측정한 것을 의미할 수 있다.
- [0029] 그래프의 가로축은 배일런스 값을 나타낸 것으로, 특징 요소가 우측으로 향할수록 포지티브 배일런스(Positive valence) 값에 가까워지고, 반대로 좌측으로 향할수록 네거티브 배일런스(Negative valence) 값에 가까워진다.
- [0030] 그래프의 세로축은 어라우절 값을 나타낸 것으로, 특징 요소가 위로 향할수록 하이 어라우절(High arousal) 값을 가지며, 반대로 아래로 향할수록 로우 어라우절(Low arousal) 값을 가지게 된다.
- [0031] 그래프의 사분면 중 제 1사분면에 해당하는 포지티브 배일런스-하이 어라우절의 값에 포함되는 감정은 대표적으로 행복(Happy), 흥분(Excited), 놀람(Surprise) 등이 포함될 수 있다.
- [0032] 또한, 제 2 사분면에 해당하는 네거티브 배일런스-하이 어라우절의 값에 포함되는 감정은 대표적으로 공포(Fear), 분노(Anger), 괴로움(Distressed) 등이 포함될 수 있다.
- [0033] 또한, 제 3 사분면은 네거티브 배일런스-로우 어라우절의 값에 포함되고, 대표적인 감정으로 슬픔(Sad), 혐오(Disgust) 등이 포함될 수 있다.
- [0034] 또한, 제 4 사분면에는 포지티브 배일런스-로우 어라우절의 값에 포함되고, 대표적인 감정으로 안정(Relaxed), 침착함(Calm), 졸음(Sleepiness) 등이 포함될 수 있다.
- [0035] 도 3은 본 발명의 일 실시예에 따른, 웃음 발생 빈도/길이 모듈(103a)을 이용한, 특징 요소를 추출하는 과정을 나타낸 동작 흐름도이다.
- [0036] 도 3을 살펴보면, 음성 발화 양식에 따른 발화자의 감정 인식 시스템이 사용자로부터 발화 음성을 수신하고, 특징 요소를 추출하기 위한 변환 단계를 가진다(S310).
- [0037] 해당 단계에서 사용자로부터 수신된 발화 음성은 전처리 모듈(101)과 구간 분할 모듈(102)을 통해 발화 구간을 특징 요소를 추출하기에 알맞게 변형된다.
- [0038] 다음은 웃음 발생 빈도/길이 모듈(103a)을 통해 사용자의 음성신호에서 웃음 발생 빈도 및 길이 값을 추출하는 단계를 가진다(S320).

- [0039] 단계(S320)에서 구간에 따른 웃음의 빈도수와 웃음의 길이를 수치화 하고, 이를 통해 배일런스 값을 정할 수 있다. 이는, 웃음 빈도수가 기 설정된 기준 웃음 빈도수를 초과하는 경우, 초과하는 정도에 기초하여 배일런스 값을 포지티브 배일런스 값으로 결정하게 된다.
- [0040] 마지막 단계로, 웃음 특징 요소를 판단하는 단계를 가진다(S330).
- [0041] 해당 단계에서 추출된 특징 요소를 토대로 사용자의 감정을 추측하게 된다. 선택적 실시예로, 웃음 발생 빈도가 0일 경우 네거티브 배일런스 값을 가지는 공포(Fear)라는 감정을 사용자에게 감지할 수 있고 다른 선택적 실시예로, 웃음의 발생 빈도가 기 설정된 값, 예컨대 1.9회 이상이고, 웃음의 길이가 870ms 이상이라면, 포지티브 배일런스 값을 가진 것으로 판단하고 흥분이나 행복이라는 감정을 사용자에게서 감지할 수 있다.
- [0042] 도 4는 본 발명의 일 실시예에 따른, 휴지 발생 길이 모듈(103b)을 이용한, 특징 요소를 추출하는 과정을 나타낸 동작 흐름도이다.
- [0043] 도 4를 살펴보면, 음성 발화 양식에 따른 발화자의 감정 인식 시스템이 사용자로부터 발화 음성을 수신하고, 특징 요소를 추출하기 위한 변환 단계를 가진다(S410).
- [0044] 도 3의 단계(S310)과 마찬가지로 사용자로부터 수신된 발화 음성은 전처리 모듈(101)과 구간 분할 모듈(102)을 통해 발화 구간을 특징 요소를 추출하기에 알맞게 변형된다.
- [0045] 다음은 휴지 발생 길이 모듈(103b)을 통해 사용자의 음성신호에서 휴지가 발생 시 그 길이 값을 추출하는 단계를 가진다(S420).
- [0046] 발화 음성에서 휴지가 발생하는 빈도를 측정하지 않는 것은, 전체적인 대화에서 그 발생빈도가 너무 높기 때문에 감정을 분류하는 요소로는 적합하지 않기 때문이다.
- [0047] 따라서 단계(S420)에서 구간에 따른 휴지의 길이를 수치화 하고, 기 설정된 길이와 비교하고, 이를 기초로 어라우절 값을 정할 수 있다. 이는, 휴지기가 기 설정된 기준 휴지 길이보다 길다면 휴지기의 길이에 기초하여 로우 어라우절 값을 결정하게 된다.
- [0048] 마지막 단계로, 휴지기 특징 요소를 판단하는 단계를 가진다(S430).
- [0049] 해당 단계에서 추출된 특징 요소를 토대로 사용자의 감정을 추측하게 된다. 선택적 실시예로 휴지의 기준점을 921ms로 하이-로우 어라우절을 구분할 수 있고, 기준점 이상의 수치를 가진다면 로우 어라우절, 기준점보다 높은 값을 가지면, 하이 어라우절로 분류하게 된다.
- [0050] 도 5는 본 발명의 일 실시예에 따른, 말걸침 발생 빈도 모듈(103c)을 이용한, 특징 요소를 추출하는 과정을 나타낸 동작 흐름도이다.
- [0051] 도 5를 살펴보면, 제 1 사용자의 발화 음성 수신 후 변환 단계(S510a)와 제 2 사용자의 발화 음성 수신 후 변환 단계(S510b)가 수행된다.
- [0052] 앞에서 진행한 도 3의 웃음 발생 빈도/길이 모듈(103a)과 도 4의 휴지 발생 길이 모듈(103b)와 다르게 본 단계(S510a 및 S510b)에서는 두 사람의 대화를 수신하고, 구간을 추출하게 된다.
- [0053] 다음으로 발화 음성 간의 중첩되는 구간을 추출하는 단계를 가진다(S520).
- [0054] 제 1 사용자와 제 2 사용자 간의 대화 중 말이 겹치는 상황의 몇 회인지를 카운팅하는 단계이다.
- [0055] 마지막으로 단계(S520)에서 카운팅한 말걸침 빈도수와 길이를 통해, 말걸침 특징 요소를 판단하는 단계를 가진다(S530).
- [0056] 말걸침 빈도수 및 길이에 기초하여, 네거티브 배일런스 및 하이 어라우절을 결정하게 된다. 예컨대, 말걸침 횟수는 기 설정된 빈도와 길이를 통해 제 1 사용자 및 제 2 사용자가 보유한 분노를 감지하는 요소로 사용될 수 있고, 선택적 실시예로 기준치보다 말걸침 횟수가 높다면, 사용자는 많은 분노를 느끼는 것으로 판단할 수 있다.
- [0057] 도 6은 본 발명의 일 실시예에 따른, 대화 반응 발생 빈도/길이 모듈(103d)을 이용한, 특징 요소를 추출하는 과정을 나타낸 동작 흐름도이다.
- [0058] 도 6을 살펴보면, 앞의 도 5의 단계(S510a 및 S510b)과 마찬가지로 제 1 사용자의 발화 음성 수신 후 변환 단계

(S610a)와 제 2 사용자의 발화 음성 수신 후 변환 단계(S610b)가 수행된다.

- [0059] 다음으로 제 1 사용자의 발화 음성 종료 후 제 2 사용자의 발화음성 시작 전까지 구간의 추출하는 단계를 가진다(S620).
- [0060] 이때, 대화 반응을 통해 하이-로우 어라우절의 특징을 분류할 수 있다. 이는, 대화 반응의 횟수가 기 설정된 기준 이상이며 대화 반응의 길이가 기 설정된 기준 이하이면, 상기 대화 반응 횟수 및 길이에 기초하여, 하이 어라우절 값을 결정하게 된다.
- [0061] 마지막으로 대화 반응 특징 요소를 판단하는 단계를 가진다(S630).
- [0062] 대화 반응 요소는 기 설정된 값보다 대화 발생빈도가 높거나, 반응 시간이 짧으면 하이 어라우절 값을 가질 수 있다. 선택적 실시예로 사용자간의 대화에서 대화 반응의 발생 빈도가 5.7회 이상이며, 그 반응 시간이 929ms 이하인 경우 하이 어라우절의 특징을 가진 감정집단으로 분류할 수 있다.
- [0063] 도 7은 본 발명의 일 실시예에 따른, 음성 발화 양식에 따른 발화자의 감정인식 방법을 나타낸 동작 흐름도이다.
- [0064] 도 7을 살펴보면, 음성 발화 양식에 따른 발화자의 감정인식 시스템이 사용자로부터 발화 음성 신호를 수신하는 단계를 가진다(S710).
- [0065] 사용자로부터 수신한 발화 음성 신호는 전처리 모듈(101)과 구분 분할 모듈(102)을 통해 특징 요소를 추출하기에 알맞도록 가공하는 과정을 거치게 된다.
- [0066] 또한 발화 음성 신호는 단일 사용자가 제공하는 음성 신호가 있고, 두 명 이상의 사용자가 대화를 나누는 복수의 발화 음성신호가 있을 수 있다.
- [0067] 다음으로 음성 신호에서 특징 요소를 추출하는 단계를 가진다(S720).
- [0068] 특징 요소는 앞에서 설명한 단일 사용자가 제공하는 음성 신호와 두 명 이상의 사용자가 나누는 대화에 따라 적용하는 모듈이 달라진다.
- [0069] 웃음 발생 빈도/길이 모듈(103a)과 휴지 발생 길이 모듈(103b)을 이용하여 단일 사용자가 제공하는 음성 신호에서 제 1 특징 요소를 추출할 수 있고, 2명 이상의 사용자의 대화는 웃음 발생 빈도/길이 모듈(103a), 휴지 발생 길이 모듈(103b)을 사용하여 추출한 제 1 특징 요소를 포함하여, 말걸침 발생 빈도 모듈(103c)과 대화 반응 발생 빈도/길이 모듈(103d)을 통해 추출한 제 2 특징 요소를 더 포함할 수 있다. 추후 발명이 고도화됨에 따라 음성에서 감정을 추출할 수 있는 추가적인 분석 모듈이 발명된다면, 추가적인 분석 모듈을 더 포함될 수 있다.
- [0070] 단계(S720)에서 추출한 특징 요소에 기초하여 사용자들의 감정을 판단하는 단계를 가진다(S730).
- [0071] 이때 특징 요소를 감정 분류 모델(105)을 통해 분류하고, 감정 판별 모듈(104)이 최종적으로 사용자의 감정 상태를 판단하는 단계를 가지게 된다.
- [0072] 이상에서 설명한 본 발명의 실시예에 따른 음성 발화 양식에 따른 발화자의 감정인식 장치 및 방법은, 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행 가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 이러한 기록 매체는 컴퓨터 판독 가능 매체를 포함하며, 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체를 포함하며, 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다.
- [0073] 이를 자세히 설명하면, 컴퓨터의 프로세서는 일종의 중앙처리장치로서 음성 발화 양식에 따른 발화자의 감정인식 방법을 제공하는 전체 과정을 제어한다. 여기서, 프로세서는 프로세서(processor)와 같이 데이터를 처리할 수 있는 모든 종류의 장치를 포함할 수 있다. 여기서, '프로세서(processor)'는, 예를 들어 프로그램 내에 포함된 코드 또는 명령으로 표현된 기능을 수행하기 위해 물리적으로 구조화된 회로를 갖는, 하드웨어에 내장된 데이터 처리 장치를 의미할 수 있다. 이와 같이 하드웨어에 내장된 데이터 처리 장치의 일 예로써, 마이크로프로세서(microprocessor), 중앙처리장치(central processing unit: CPU), 프로세서 코어(processor core), 멀티프로세서(multiprocessor), ASIC(application-specific integrated circuit), FPGA(field programmable gate array) 등의 처리 장치를 망라할 수 있으나, 본 발명의 범위가 이에 한정되는 것은 아니다.

[0074] 또한, 메모리는 음성 발화 양식에 따른 발화자의 감정인식 방법을 위한 프로그램이 기록된다. 또한, 프로세서가 처리하는 데이터를 일시적 또는 영구적으로 저장하는 기능을 수행한다. 여기서, 메모리는 자기 저장 매체(magnetic storage media) 또는 플래시 저장 매체(flash storage media)를 포함할 수 있으나, 본 발명의 범위가 이에 한정되는 것은 아니다.

[0075] 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, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

[0076] 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.

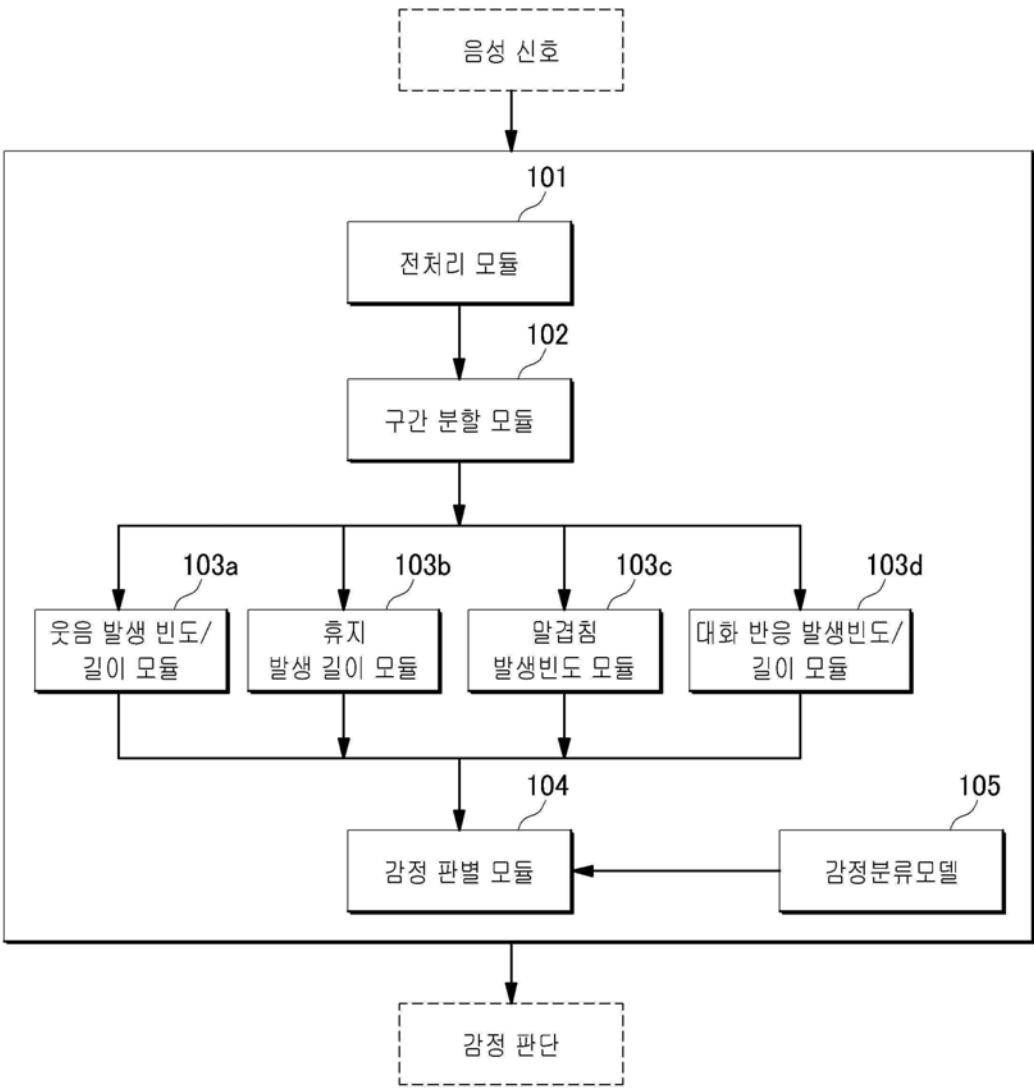
부호의 설명

[0077]

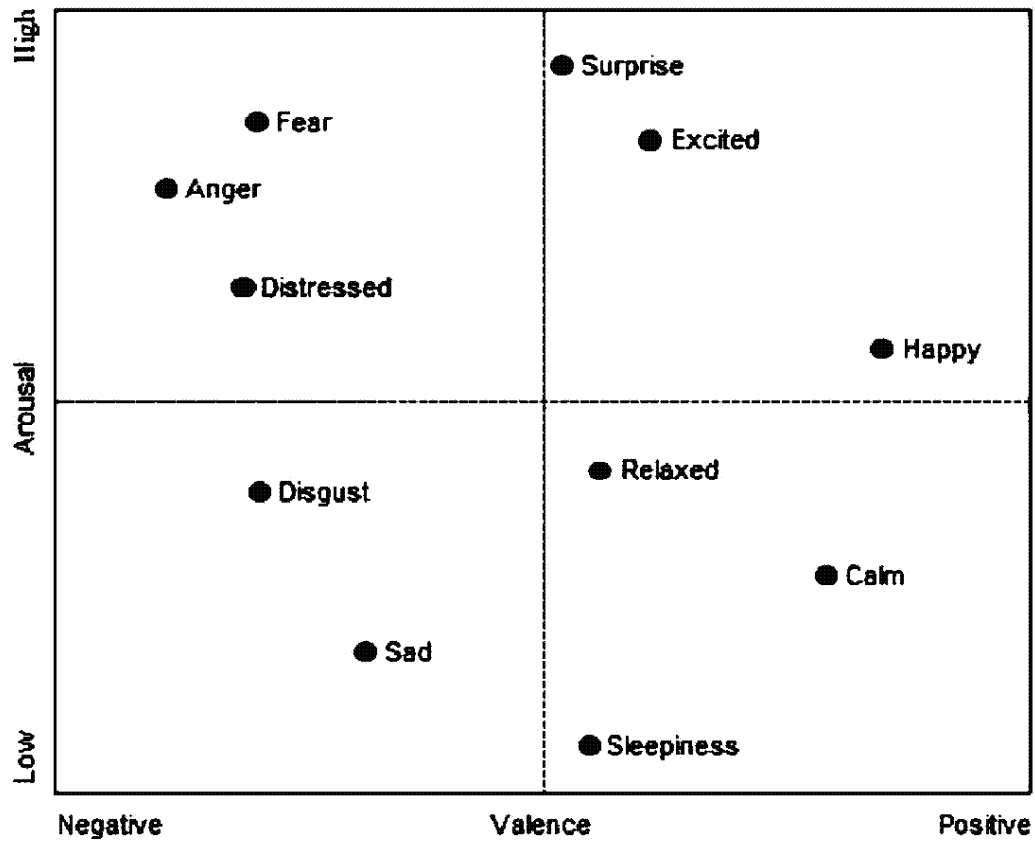
101: 전처리 모듈	102: 구간 분할 모듈
103a: 웃음 발생빈도/길이 모듈	103b: 휴지 발생길이 모듈
103c: 말걸침 발생빈도 모듈	
103d: 대화 반응 발생빈도/길이 모듈	
104: 감정 판단 모듈	
105: 감정 분류 모델	

도면

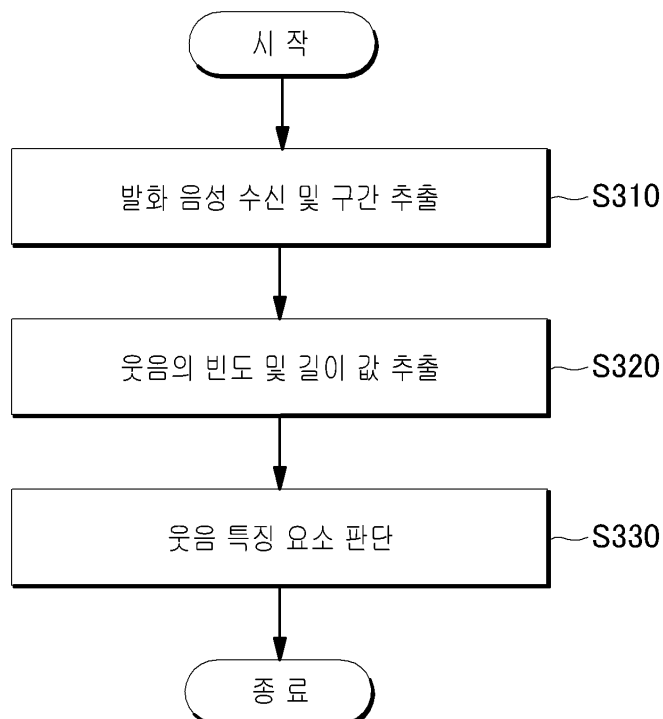
도면1



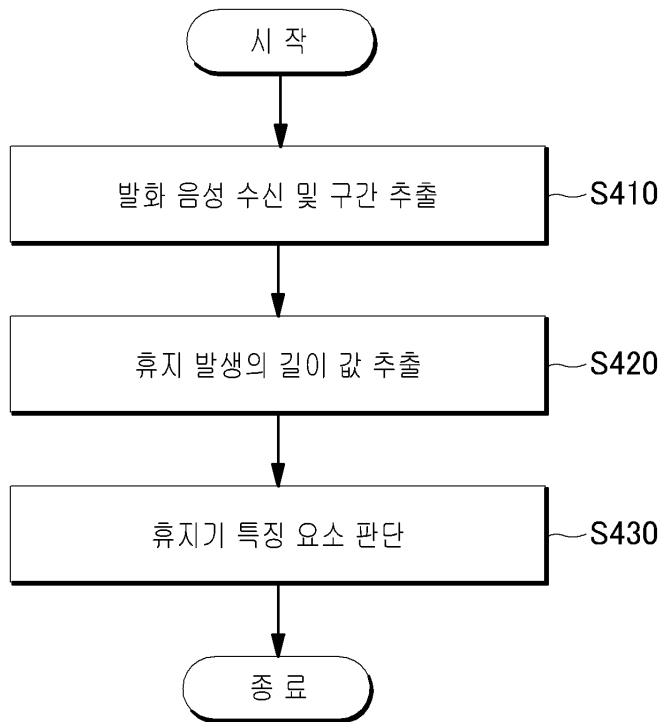
도면2



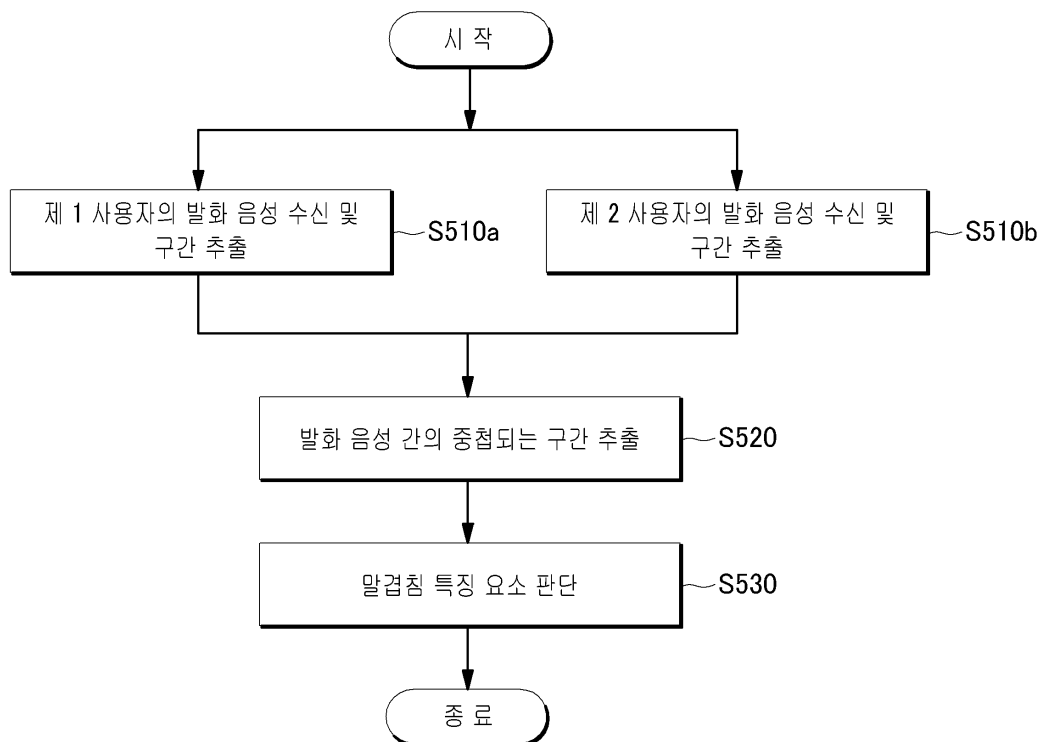
도면3



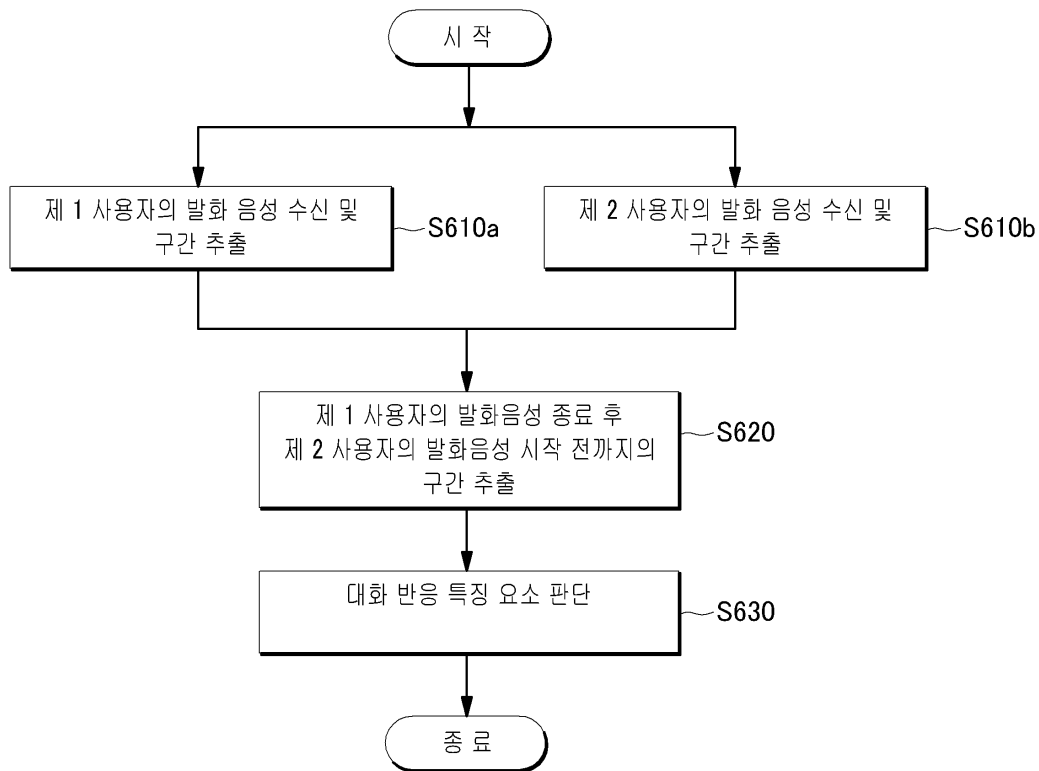
도면4



도면5



도면6



도면7

