



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0098104
(43) 공개일자 2010년09월06일

(51) Int. Cl.

G10L 21/00 (2006.01) G10L 17/00 (2006.01)

(21) 출원번호 10-2009-0017114

(22) 출원일자 2009년02월27일

심사청구일자 2009년02월27일

(71) 출원인

고려대학교 산학협력단

서울 성북구 안암동5가 1

(72) 발명자

육동석

서울 용산구 서빙고동 신동아아파트 1동 102호

이협우

경기도 안양시 동안구 갈산동 샘마을대우아파트 101-1002

(74) 대리인

특허법인추현

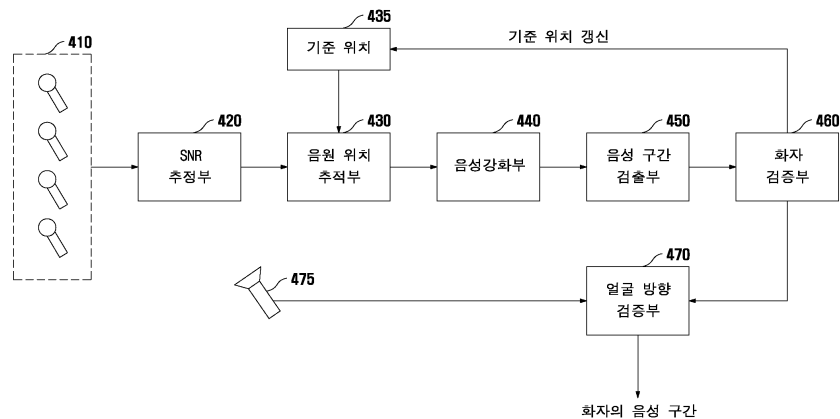
전체 청구항 수 : 총 12 항

(54) 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법 및 장치

(57) 요약

본 발명은 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법 및 장치에 관한 것으로, 본 발명의 일 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법은 마이크로폰 어레이에 입력된 오디오 신호에서 음성 구간을 검출하는 단계; 상기 검출된 음성 구간에서 화자 검증을 수행하는 단계; 상기 화자 검증에 성공한 경우, 카메라에 입력된 비디오 신호를 이용하여 화자의 얼굴을 검출하고 상기 화자의 얼굴 방향을 추정하는 단계; 및 상기 추정된 얼굴 방향이 미리 저장된 기준 방향과 일치하면, 상기 검출된 음성 구간을 상기 화자의 음성 구간으로 판단하는 단계를 포함한다. 본 발명에 의하면, 인식하고자 하는 화자의 위치와 방향을 고려하여 음성 구간을 검출하여 음성 인식 시스템의 성능을 향상시킬 수 있고, 화자의 위치가 시스템이 예상한 위치를 벗어나는 경우에는 적응적인 동작이 가능하다.

대표도 - 도4



특허청구의 범위

청구항 1

마이크로폰 어레이에 입력된 오디오 신호에서 음성 구간을 검출하는 단계;

상기 검출된 음성 구간에서 화자 검증을 수행하는 단계;

상기 화자 검증에 성공한 경우, 카메라에 입력된 비디오 신호를 이용하여 화자의 얼굴을 검출하고 상기 화자의 얼굴 방향을 추정하는 단계; 및

상기 추정된 얼굴 방향이 미리 저장된 기준 방향과 일치하면, 상기 검출된 음성 구간을 상기 화자의 음성 구간으로 판단하는 단계

를 포함하는, 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법.

청구항 2

제1항에 있어서,

상기 음성 구간을 검출하는 단계는,

마이크로폰 어레이에 입력된 오디오 신호를 이용하여 음원의 위치를 추정하는 단계; 및

상기 추정된 음원의 위치와 미리 저장된 기준 위치를 비교하여 잡음을 구별하는 단계를 포함하는 것을 특징으로 하는, 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법.

청구항 3

제2항에 있어서,

상기 화자 검증을 수행하는 단계는,

상기 화자 검증에 성공한 경우, 상기 기준 위치의 값을 상기 추정된 음원의 위치로 변경하는 단계를 포함하는 것을 특징으로 하는, 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법.

청구항 4

제2항에 있어서,

상기 음원의 위치를 추정하는 단계는,

상기 마이크로폰 어레이에 입력된 오디오 신호 중 일정 SNR 이상의 신호를 이용하는 단계인 것을 특징으로 하는, 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법.

청구항 5

제2항에 있어서,

상기 음성 구간을 검출하는 단계는,

상기 구별된 잡음을 제거하는 단계; 및

상기 잡음이 제거된 신호에서 단일 마이크로폰 기반으로 음성 구간을 검출하는 단계를 더 포함하는 것을 특징으로 하는, 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법.

청구항 6

제5항에 있어서,

상기 구별된 잡음을 제거하는 단계는,

상기 미리 저장된 위치와 다른 위치로 추정되는 음원의 신호를 제거하는 단계를 포함하는 것을 특징으로 하는, 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법.

청구항 7

마이크로폰 어레이에 입력된 오디오 신호를 이용하여 음원의 위치를 추정하는 단계; 및

상기 추정된 음원의 위치와 미리 저장된 기준 위치를 비교하여 일치하지 않는 횟수가 임계값 이상인 경우, 상기 오디오 신호에서 음성 구간을 검출하는 단계;

상기 검출된 음성 구간에서 화자 검증을 수행하는 단계;

상기 화자 검증에 성공한 경우, 카메라에 입력된 비디오 신호를 이용하여 화자의 얼굴을 검출하고 상기 화자의 얼굴 방향을 추정하는 단계; 및

상기 추정된 얼굴 방향이 미리 저장된 기준 방향과 일치하면, 상기 검출된 음성 구간을 상기 화자의 음성 구간으로 판단하는 단계

를 포함하는, 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법.

청구항 8

제7항에 있어서,

상기 화자 검증을 수행하는 단계는,

상기 화자 검증에 성공한 경우, 상기 기준 위치의 값을 상기 추정된 음원의 위치로 변경하는 단계를 포함하는 것을 특징으로 하는, 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법.

청구항 9

제 1 항 내지 제 8 항 중 어느 한 항에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법을 컴퓨터 시스템에서 실행하기 위한 프로그램이 기록된 컴퓨터 시스템이 판독할 수 있는 기록매체.

청구항 10

마이크로폰 어레이에 입력된 오디오 신호에서 음성 구간을 검출하는 음성 구간 검출부;

상기 검출된 음성 구간에서 화자 검증을 수행하는 화자 검증부; 및

상기 화자 검증에 성공한 경우, 카메라에 입력된 비디오 신호를 이용하여 화자의 얼굴을 검출하고 상기 화자의 얼굴 방향을 추정하고, 상기 추정된 얼굴 방향이 미리 저장된 기준 방향과 일치하면 상기 검출된 음성 구간을 상기 화자의 음성 구간으로 판단하는 얼굴 방향 검증부

를 포함하는, 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 장치.

청구항 11

마이크로폰 어레이에 입력된 오디오 신호를 이용하여 음원의 위치를 추정하는 음원 위치 추적부; 및

상기 추정된 음원의 위치와 미리 저장된 기준 위치를 비교하여 일치하지 않는 횟수가 임계값 이상인 경우, 상기 오디오 신호에서 음성 구간을 검출하는 음성 구간 검출부;

상기 검출된 음성 구간에서 화자 검증을 수행하는 화자 검증부;

상기 화자 검증에 성공한 경우, 카메라에 입력된 비디오 신호를 이용하여 화자의 얼굴을 검출하고 상기 화자의 얼굴 방향을 추정하고, 상기 추정된 얼굴 방향이 미리 저장된 기준 방향과 일치하면 상기 검출된 음성 구간을 상기 화자의 음성 구간으로 판단하는 얼굴 방향 검증부

를 포함하는, 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 장치.

청구항 12

제11항에 있어서,

상기 화자 검증부는,

상기 화자 검증에 성공한 경우, 상기 기준 위치의 값을 상기 추정된 음원의 위치로 변경하는 것을 특징으로 하

는, 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 장치.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 신호처리에 관한 것으로, 특히, 음성 인식 분야에서 시공간의 데이터를 이용하여 음성 구간을 검출하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 로봇이나 홈오토메이션과 같은 시스템은 항상 동작하는 도중에 입력되는 신호를 분석하여 사람이 지시하는 행동을 하여야 한다. 이를 위해서는 마이크로폰에 연속적으로 입력되는 신호 중에 사람의 목소리가 있는지 알아내야 화자인식이나 음성인식을 할 수 있다.

[0003] 통상, 음성인식은 기본적으로 인식하고자 하는 음성 패턴과 기준 패턴과의 유사도 판별에 의해 이루어진다.

[0004] 도 1은 종래의 음성 인식 시스템을 도시한 것이다.

[0005] 도 1의 음성 인식 시스템은 입력되는 신호로부터 음성 구간을 검출하는 음성구간 검출부(1)와, MFCC 계수를 사용하여 음성구간 검출부(1)에서 검출된 음성구간으로부터 그 특징을 추출하는 특징계수 추출부(2)와, HMM(Hidden Markov Model)과 VMS VQ(Variable Multi-Section Vector Quantization) 알고리즘을 이용하여 음성 신호를 인식하는 음성 인식부(3)와, 음성신호에 의해 학습된 단어모델 파라미터가 저장되어 있는 데이터 베이스(4)와, 음성 인식부(3)에서 인식된 음성신호에 대해 실효성을 판단하여 인식되는 단어를 출력하는 후처리부(5)로 구성된다.

[0006] 이와 같이 구성된 음성 인식 시스템에 있어서, 입력되는 신호로부터 정확한 음성구간을 검출하는 것은 음성 인식 시스템의 전처리 부분으로, 시스템의 성능을 좌우하는 전제조건으로서 매우 중요한 작업이다.

[0007] 음성인식시에 필수적인 음성구간 검출방법으로 다양한 기법이 사용되고 있다. 일반적으로 가장 많이 사용되는 방법은 음성신호의 시간 축상에서의 특징을 이용한 음성구간 검출방법이다. 다시 말해서, 음성신호는 음성구간에서 높은 에너지를 갖게 되고 음성 샘플 사이에 유사도가 매우 높을 뿐만 아니라, 최소한의 음성 지속 시간이 존재한다. 이와 같은 음성신호의 시간축상에서의 특징을 이용하여 배경잡음과 음성구간을 구분함으로써 음성 구간을 검출해내는 것이다.

[0008] 그런데 음성신호에 주변잡음이 심한 경우 음성신호의 상기 특징들이 잡음에 의해 손상되기 때문에 음성 구간의 검출이 어려워진다. 예를 들어, 신호 대 잡음비(SNR)가 0 dB인 경우에는 신호와 잡음의 에너지가 같기 때문에 에너지에 의한 잡음과 음성구간 구분은 불가능해진다.

[0009] 시스템에 입력된 모든 신호에 대해 화자인식이나 음성인식을 수행하면 올바른 결과를 내보낼 수 없을 뿐만 아니라 시스템의 불필요한 파워 소모가 발생할 수 있다. 시스템은 원하는 위치가 아닌 곳에서 발생하는 음성과 원하는 위치에서 발생하였지만 음성이 아닌 잡음을 무시하고 원하는 곳에서 발생한 음성만을 추출해야 한다. 다른 지점에서 발생하는 원하는 음성과 원하지 않는 잡음(원하지 않는 음성을 포함)이 동시에 입력되는 경우, 기존 음성 구간 검출 알고리즘들은 매우 낮은 성능을 보이게 된다. 또한, 화자가 원하는 위치에서 발화하는 경우에 다른 곳을 보고 있다면 시스템이 원하는 음성이 아닐 확률이 클에도 불구하고, 이를 구별하지 못하고 음성 인식을 의도하게 되는 문제가 있다.

발명의 내용

해결 하고자하는 과제

[0010] 따라서, 본 발명이 이루고자 하는 첫 번째 기술적 과제는 인식하고자 하는 음원의 위치와 화자의 얼굴 방향을 고려하여 음성 구간을 검출하여 음성 인식 시스템의 성능을 향상시킬 수 있는 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법을 제공하는 데 있다.

[0011] 본 발명이 이루고자 하는 두 번째 기술적 과제는 인식하고자 하는 음원의 위치와 화자의 얼굴 방향을 고려하여

음성 구간을 검출하여 음성 인식 시스템의 성능을 향상시킬 수 있는 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 장치를 제공하는 데 있다.

과제 해결수단

- [0012] 상기의 첫 번째 기술적 과제를 이루기 위하여, 본 발명의 일 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법은 마이크로폰 어레이에 입력된 오디오 신호에서 음성 구간을 검출하는 단계; 상기 검출된 음성 구간에서 화자 검증을 수행하는 단계; 상기 화자 검증에 성공한 경우, 카메라에 입력된 비디오 신호를 이용하여 화자의 얼굴을 검출하고 상기 화자의 얼굴 방향을 추정하는 단계; 및 상기 추정된 얼굴 방향이 미리 저장된 기준 방향과 일치하면, 상기 검출된 음성 구간을 상기 화자의 음성 구간으로 판단하는 단계를 포함한다.
- [0013] 바람직하게는, 상기 음성 구간을 검출하는 단계는, 마이크로폰 어레이에 입력된 오디오 신호를 이용하여 음원의 위치를 추정하고, 상기 추정된 음원의 위치와 미리 저장된 기준 위치를 비교하여 잡음을 구별할 수 있다. 이 경우, 상기 화자 검증을 수행하는 단계에서 상기 화자 검증에 성공한 경우, 상기 기준 위치의 값을 상기 추정된 음원의 위치로 변경할 수 있다. 한편, 상기 음원의 위치를 추정하는 단계는, 상기 마이크로폰 어레이에 입력된 오디오 신호 중 일정 SNR 이상의 신호를 이용하는 단계일 수 있다. 한편, 상기 음성 구간을 검출하는 단계에서, 상기 구별된 잡음을 제거하고, 상기 잡음이 제거된 신호에서 단일 마이크로폰 기반으로 음성 구간을 검출할 수도 있다. 이와 같이 구별된 잡음을 제거하는 단계에서, 상기 미리 저장된 위치와 다른 위치로 추정되는 음원의 신호를 제거할 수도 있다.
- [0014] 상기의 첫 번째 기술적 과제를 이루기 위하여, 본 발명의 다른 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법은 마이크로폰 어레이에 입력된 오디오 신호를 이용하여 음원의 위치를 추정하는 단계; 및 상기 추정된 음원의 위치와 미리 저장된 기준 위치를 비교하여 일치하지 않는 횟수가 임계값 이상인 경우, 상기 오디오 신호에서 음성 구간을 검출하는 단계; 상기 검출된 음성 구간에서 화자 검증을 수행하는 단계; 상기 화자 검증에 성공한 경우, 카메라에 입력된 비디오 신호를 이용하여 화자의 얼굴을 검출하고 상기 화자의 얼굴 방향을 추정하는 단계; 및 상기 추정된 얼굴 방향이 미리 저장된 기준 방향과 일치하면, 상기 검출된 음성 구간을 상기 화자의 음성 구간으로 판단하는 단계를 포함한다.
- [0015] 바람직하게는, 상기 화자 검증을 수행하는 단계에서 상기 화자 검증에 성공한 경우, 상기 기준 위치의 값을 상기 추정된 음원의 위치로 변경할 수 있다.
- [0016] 상기의 두 번째 기술적 과제를 이루기 위하여, 본 발명의 일 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 장치는 마이크로폰 어레이에 입력된 오디오 신호에서 음성 구간을 검출하는 음성 구간 검출부; 상기 검출된 음성 구간에서 화자 검증을 수행하는 화자 검증부; 및 상기 화자 검증에 성공한 경우, 카메라에 입력된 비디오 신호를 이용하여 화자의 얼굴을 검출하고 상기 화자의 얼굴 방향을 추정하고, 상기 추정된 얼굴 방향이 미리 저장된 기준 방향과 일치하면 상기 검출된 음성 구간을 상기 화자의 음성 구간으로 판단하는 얼굴 방향 검증부를 포함한다.
- [0017] 상기의 두 번째 기술적 과제를 이루기 위하여, 본 발명의 다른 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 장치는 마이크로폰 어레이에 입력된 오디오 신호를 이용하여 음원의 위치를 추정하는 음원 위치 추적부; 및 상기 추정된 음원의 위치와 미리 저장된 기준 위치를 비교하여 일치하지 않는 횟수가 임계값 이상인 경우, 상기 오디오 신호에서 음성 구간을 검출하는 음성 구간 검출부; 상기 검출된 음성 구간에서 화자 검증을 수행하는 화자 검증부; 상기 화자 검증에 성공한 경우, 카메라에 입력된 비디오 신호를 이용하여 화자의 얼굴을 검출하고 상기 화자의 얼굴 방향을 추정하고, 상기 추정된 얼굴 방향이 미리 저장된 기준 방향과 일치하면 상기 검출된 음성 구간을 상기 화자의 음성 구간으로 판단하는 얼굴 방향 검증부를 포함한다.
- [0018] 바람직하게는, 상기 화자 검증부는 상기 화자 검증에 성공한 경우, 상기 기준 위치의 값을 상기 추정된 음원의 위치로 변경할 수 있다.

효 과

- [0019] 본 발명에 의하면, 인식하고자 하는 화자의 위치와 방향을 고려하여 음성 구간을 검출하여 음성 인식 시스템의 성능을 향상시킬 수 있고, 화자의 위치가 시스템이 예상한 위치를 벗어나는 경우에는 적응적인 동작이 가능하다.

발명의 실시를 위한 구체적인 내용

- [0020] 이하에서는 도면을 참조하여 본 발명의 바람직한 실시 예를 설명하기로 한다. 그러나, 다음에 예시하는 본 발명의 실시 예는 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 다음에 상술하는 실시 예에 한정되는 것은 아니다.
- [0021] 본 발명의 일 실시 예에 따른 음성 구간 추출 방법에서는 원하는 음성이 발생하는 지점을 음성 인식 시스템이 미리 알고 있거나, 음원 위치 추적 기법을 통해 신호의 발생 위치를 음성 인식 시스템이 알고 있다고 가정한다. 시스템은 원하는 위치에서 발생한 잡음을 무시하고 원하는 위치 및 방향에서 발생한 음성만을 찾는다. 마이크로폰에 동시에 입력되는 음성과 잡음에 대하여 음성의 구간을 검출하기 위해서는 마이크로폰 어레이를 이용하는 것이 바람직하다. 마이크로폰 어레이는 복수의 마이크로폰으로 음원의 위치를 탐색할 수 있기 때문에 활용할 수 있는 정보가 많은 장점이 있다.
- [0022] 도 2는 본 발명의 일 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법의 흐름도이다.
- [0023] 먼저, 마이크로폰 어레이에 입력된 오디오 신호에서 음성 구간을 검출한다(S210). 이 과정에서, 신호의 세기를 이용하여 일정 SNR (Signal-to-Noise Ratio) 이상의 신호에 대해 음성 구간을 탐색할 수 있다.
- [0024] 다음, 검출된 음성 구간에서 화자 검증을 수행한다(S220). 화자 검증은 시스템에 원하는 화자의 음성 모델을 미리 구축하고, 시스템에 입력되는 음성 신호를 미리 구축된 음성 모델과 비교하는 방식으로 수행된다.
- [0025] 다음, 화자 검증에 성공한 경우(S230), 카메라에 입력된 비디오 신호를 이용하여 화자의 얼굴을 검출하고 화자의 얼굴 방향을 추정한다(S240). 화자 검증에 실패한 경우에는, 음성 구간 검출 과정(S210)부터 다시 시작한다.
- [0026] 다음, 비디오 신호를 이용하여 추정된 화자의 얼굴 방향이 시스템에 미리 저장된 기준 방향과 일치하면(S250), 음성 구간 검출 과정(S210)에서 검출된 음성 구간을 원하는 화자의 음성 구간으로 판단한다(S260). 만약 비디오 신호를 이용하여 추정된 화자의 얼굴 방향이 시스템에 미리 저장된 기준 방향과 일치하지 않으면, 음성 구간 검출 과정(S210)부터 다시 시작한다.
- [0027] 도 3은 본 발명의 다른 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법의 흐름도이다.
- [0028] 먼저, 마이크로폰 어레이에 입력된 오디오 신호를 이용하여 음원의 위치를 추정한다(S311).
- [0029] 다음, 위에서 추정된 음원의 위치와 시스템에 미리 저장된 기준 위치를 비교하여 일정 횟수 이상 불일치하는지 여부를 판단한다(S312). 이때, 불일치하는 횟수가 일정 횟수 미만인 경우, 해당 신호를 잡음으로 구분하거나 이를 제거한다(S313). 불일치하는 횟수가 일정 횟수 미만인 경우에는 기준 위치와 일치하는 음원을 찾는 과정을 거칠 수 있다.
- [0030] 이후에, 음성 구간을 검출할 수 있다. 바람직하게는, 신호의 세기를 이용하여 일정 SNR 이상의 신호에 대해 음성 구간을 탐색할 수 있다.
- [0031] 다음, 검출된 음성 구간에서 화자 검증을 수행한다(S320). 화자 검증은 시스템에 원하는 화자의 음성 모델을 미리 구축하고, 시스템에 입력되는 음성 신호를 미리 구축된 음성 모델과 비교하는 방식으로 수행된다.
- [0032] 화자 검증에 성공한 경우(S330), 시스템에 미리 저장된 기준 위치를 위에서 추정된 음원의 위치로 변경한다(S335).
- [0033] 다음, 카메라에 입력된 비디오 신호를 이용하여 화자의 얼굴을 검출하고 화자의 얼굴 방향을 추정한다(S340). 화자 검증에 실패한 경우에는, 음성 위치 추정 과정(S311)부터 다시 시작한다.
- [0034] 비디오 신호를 이용하여 추정된 화자의 얼굴 방향이 시스템에 미리 저장된 기준 방향과 일치하면(S350), 음원 위치를 추정(S311)하면서 검출한 음성 구간을 원하는 화자의 음성 구간으로 판단한다(S360). 만약 비디오 신호를 이용하여 추정된 화자의 얼굴 방향이 시스템에 미리 저장된 기준 방향과 일치하지 않으면, 음성 위치 추정 과정(S311)부터 다시 시작한다.
- [0035] 도 4는 본 발명의 또 다른 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 장치의 블록도이다.

- [0036] 마이크로폰 어레이(410)는 복수의 마이크로폰을 구비하여 오디오 신호를 입력받는다.
- [0037] SNR 추정부(420)는 마이크로폰 어레이(410)에 입력된 오디오 신호의 세기를 이용하여 오디오 신호의 SNR을 구한다. 일정 SNR 이상의 신호에 대해서만 음성 구간을 탐색하게 하는 것이 바람직하다.
- [0038] 음원 위치 추적부(430)는 마이크로폰 어레이에 입력된 오디오 신호를 이용하여 음원의 위치를 추적한다. 이 경우에도 일정 SNR 이상의 신호에 대해서만 음원의 위치를 추적하는 것이 바람직하다. 음원 위치 추적부(430)는 추정된 음원의 위치와 시스템에 미리 저장된 기준 위치를 비교하여 불일치시에 해당 음원을 잡음으로 판단할 수 있다. 음원 위치 추적부(430)는 기준 위치와 추정된 음원의 위치를 비교하여 원하지 않은 방향의 신호(음성을 포함한 잡음)를 제거할 수 있다.
- [0039] 음원 위치 추적부(430)에서 일정 방향으로의 오류가 연속적으로 발생하는 경우, 음성 강화부(440)는 해당 방향의 신호에 대해 음성을 강화할 수 있다. 즉, 시스템이 예상한 방향은 아니더라도, 다른 방향에서 화자의 음성 신호가 지속적으로 발생하는 경우에는 해당 방향에서의 음성 신호를 잡음으로 간주하는 것이 아니라, 정당한 음성 신호로서 용인하는 것이다. 이 경우에, 화자 검증부(460)는 정당한 음성 신호로 용인된 신호에 화자 검증기법을 적용하고, 해당 신호가 시스템이 인식하고자 하는 화자인 경우에는 시스템에 저장된 기준 방향을 변경한다. 이에 따라, 시스템은 인식하고자 하는 화자의 위치를 추적하면서 음성 인식을 수행할 수 있다.
- [0040] 한편, 음원 위치 추적부(430)에서 기준값으로 사용한 기준 위치는 메모리 기타 저장 수단(435)에 저장될 수 있다.
- [0041] 음성 강화부(440)는 전달된 오디오 신호의 음성을 강화하고 잡음을 제거한다. 음성을 강화하는 방법의 예로 특정 주파수 대역을 증폭시키는 방법이 있다.
- [0042] 음성 구간 검출부(450)는 음성 강화부(440)에서 강화된 음성을 단일 마이크로폰 기반의 음성 구간 검출 기법에 적용한다. 음성 강화부(440)에서 음성을 강화하는 이유는 음성 구간 검출 기법이 높은 SNR에서 더 좋은 성능을 보이기 때문이다.
- [0043] 화자 검증부(460)는 음성 구간 검출부(450)가 음성의 구간을 정하면 이 신호가 인식하고자 하는 사용자의 음성 인지 화자 검증을 수행한다.
- [0044] 화자 검증부(460)에서 화자가 인식하고자 하는 사용자로 판단되면, 얼굴 방향 검증부(470)는 카메라(475)에 입력된 비디오 신호에서 얼굴을 검출하여 그 얼굴의 방향을 추정한다. 얼굴 방향 검증부(470)는 추정된 얼굴의 방향이 시스템이 위치한 쪽의 방향이면 음성 구간 검출부(450)에서 검출된 음성 구간을 인식하고자 하는 화자의 음성 구간으로 인식한다.
- [0045] 본 발명은 소프트웨어를 통해 실행될 수 있다. 바람직하게는, 본 발명의 일 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법을 컴퓨터에서 실행시키기 위한 프로그램을 컴퓨터로 읽을 수 있는 기록매체에 기록하여 제공할 수 있다. 소프트웨어로 실행될 때, 본 발명의 구성 수단들은 필요한 작업을 실행하는 코드 세그먼트들이다. 프로그램 또는 코드 세그먼트들은 프로세서 판독 가능 매체에 저장되거나 전송 매체 또는 통신망에서 반송파와 결합된 컴퓨터 데이터 신호에 의하여 전송될 수 있다.
- [0046] 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 장치의 예로는 ROM, RAM, CD-ROM, DVD±ROM, DVD-RAM, 자기 테이프, 플로피 디스크, 하드 디스크(hard disk), 광데이터 저장장치 등이 있다. 또한, 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 장치에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.
- [0047] 본 발명은 도면에 도시된 일 실시 예를 참고로 하여 설명하였으나 이는 예시적인 것에 불과하며 당해 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 실시 예의 변형이 가능하다는 점을 이해할 것이다. 그리고, 이와 같은 변형은 본 발명의 기술적 보호범위 내에 있다고 보아야 한다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위의 기술적 사상에 의해서 정해져야 할 것이다.

산업이용 가능성

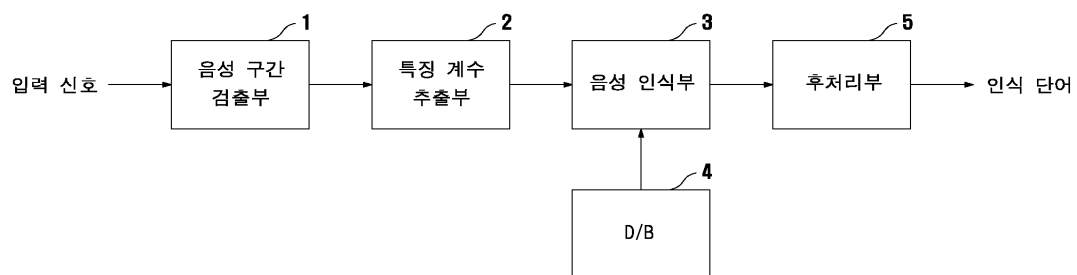
- [0048] 본 발명은 음성 인식 분야에서 시공간의 데이터를 이용하여 음성 구간을 검출하는 방법 및 장치에 관한 것으로, 인식하고자 하는 화자의 위치와 방향을 고려하여 음성 구간을 검출하여 음성 인식 시스템의 성능을 향상시킬 수 있는 음성 인식 장치 및 이를 포함하는 시스템에 적용될 수 있다.

도면의 간단한 설명

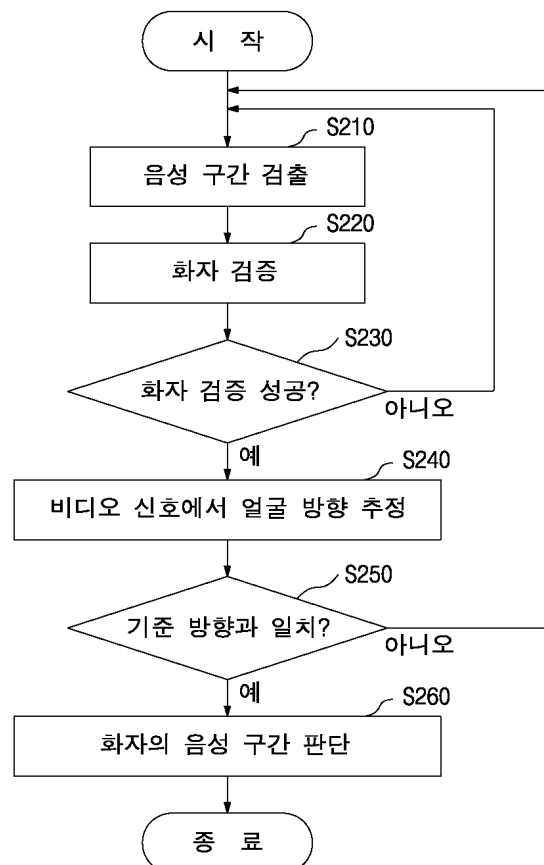
- [0049] 도 1은 종래의 음성 인식 시스템을 도시한 것이다.
- [0050] 도 2는 본 발명의 일 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법의 흐름도이다.
- [0051] 도 3은 본 발명의 다른 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 방법의 흐름도이다.
- [0052] 도 4는 본 발명의 또 다른 실시 예에 따른 오디오 및 비디오 정보를 이용한 시공간 음성 구간 검출 장치의 블록도이다.

도면

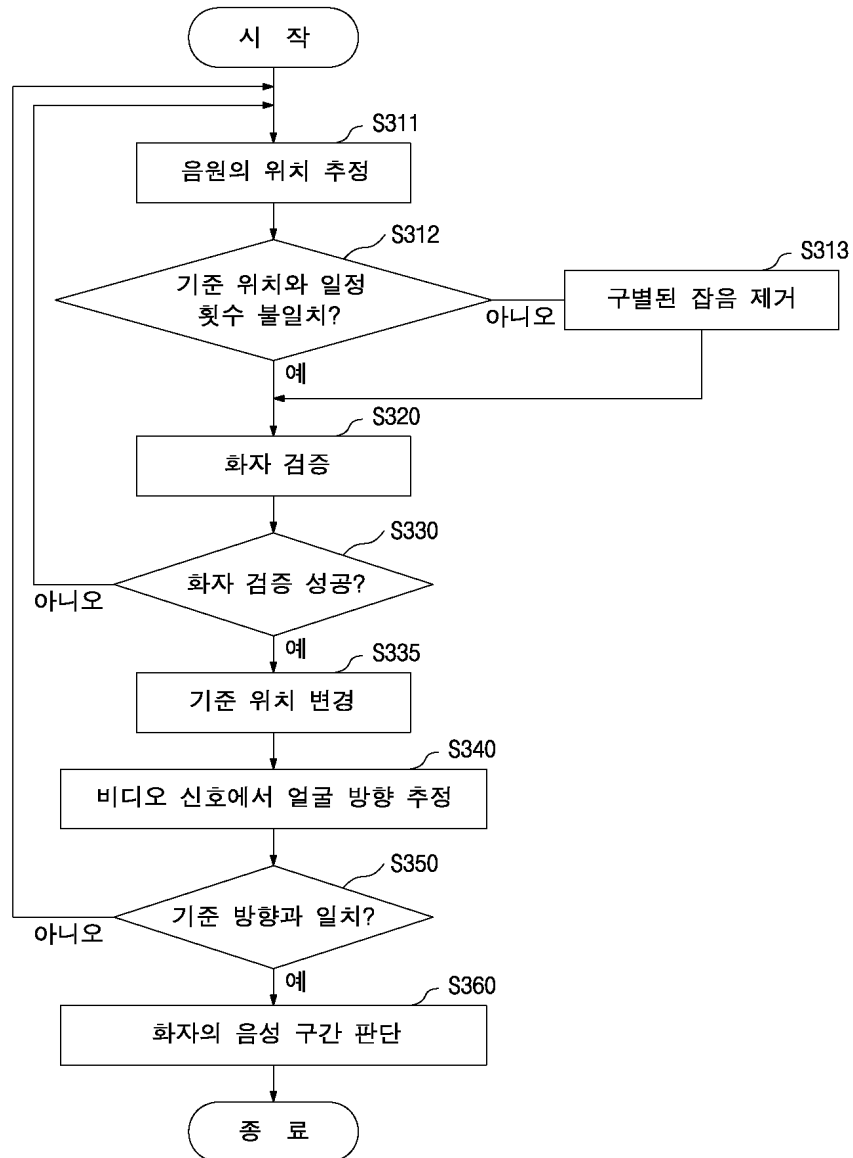
도면1



도면2



도면3



도면4

