



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0125322
(43) 공개일자 2017년11월14일

- | | |
|--|---|
| <p>(51) 국제특허분류(Int. Cl.)
G10L 17/02 (2013.01) G10L 17/12 (2013.01)
G10L 19/038 (2013.01)</p> <p>(52) CPC특허분류
G10L 17/02 (2013.01)
G10L 17/12 (2013.01)</p> <p>(21) 출원번호 10-2017-7020433</p> <p>(22) 출원일자(국제) 2015년02월27일
심사청구일자 없음</p> <p>(85) 번역문제출일자 2017년07월20일</p> <p>(86) 국제출원번호 PCT/KR2015/001953</p> <p>(87) 국제공개번호 WO 2016/137042
국제공개일자 2016년09월01일</p> | <p>(71) 출원인
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)</p> <p>(72) 발명자
고한석
서울시 용산구 이촌로88길 30 103동 1602호
김성수
경기도 부천시 소사구 은성로 4
(뒷면에 계속)</p> <p>(74) 대리인
리앤목특허법인</p> |
|--|---|

전체 청구항 수 : 총 11 항

(54) 발명의 명칭 사용자 인식을 위한 특징 벡터를 변환하는 방법 및 디바이스

(57) 요약

사용자의 발성을 포함하는 음성 신호로부터 특징 시퀀스를 추출하고, 상기 추출된 특징 시퀀스로부터 특징 벡터를 추출하고, 서로 다른 공분산 값에 기초한 확률값에 기초하여, 상기 검출된 특징 벡터의 차원 수를 줄이기 위한 변환 행렬을 획득하고, 상기 변환 행렬을 이용하여 상기 특징 벡터를 변환하는, 특징 벡터를 변환하는 방법이 개시된다.

(52) CPC특허분류

G10L 19/038 (2013.01)

(72) 발명자

노진상

경기도 안양시 만안구 안양천서로 289 108동 1702호

손수원

서울시 강남구 봉은사로 302 102동 1301호

이재원

서울시 강남구 선릉로126길 22 102동1001호

명세서

청구범위

청구항 1

특징 벡터를 변환하는 방법에 있어서,

사용자의 발성을 포함하는 음성 신호로부터 특징 시퀀스를 추출하는 단계;

상기 추출된 특징 시퀀스로부터 특징 벡터를 추출하는 단계;

서로 다른 공분산 값에 기초한 확률값에 기초하여, 상기 검출된 특징 벡터의 차원 수를 줄이기 위한 변환 행렬을 획득하는 단계;

상기 변환 행렬을 이용하여 상기 특징 벡터를 변환하는 단계를 포함하는, 방법.

청구항 2

제1항에 있어서, 상기 변환 행렬은

HLDA 변환 행렬인, 방법.

청구항 3

제1항에 있어서, 상기 변환 행렬을 획득하는 단계는

상기 변환 행렬의 유효한 차원 p 를 각 화자의 클래스 내 공분산 행렬에 대한 분산 행렬의 차원 별 누적 에너지에 기초하여 획득하는 단계를 포함하는, 방법.

청구항 4

제1항에 있어서, 상기 특징 벡터는

결합요인분석 방법에 의해 획득될 수 있는 i -벡터인, 방법.

청구항 5

제1항에 있어서,

상기 변환된 특징 벡터 및 각 스테이트의 특징 벡터에 대한 스코어링을 적어도 한번 수행하는 단계;

상기 스코어링 결과에 기초하여, 사용자를 인식하는 단계를 더 포함하는, 방법.

청구항 6

특징 벡터를 변환하는 디바이스에 있어서,

사용자의 발성을 포함하는 음성 신호를 수신하는 수신부;

상기 음성 신호로부터 특징 시퀀스를 추출하고, 상기 추출된 특징 시퀀스로부터 특징 벡터를 추출하고, 서로 다른 공분산 값에 기초한 확률값에 기초하여, 상기 추출된 특징 벡터의 차원 수를 줄이기 위한 변환 행렬을 획득하고, 상기 변환 행렬을 이용하여 상기 특징 벡터를 변환하는 제어부를 포함하는, 디바이스.

청구항 7

제6항에 있어서, 상기 변환 행렬은

HLDA 변환 행렬인, 디바이스.

청구항 8

제6항에 있어서, 상기 제어부는

상기 변환 행렬의 유효한 차원 p 를 각 화자의 클래스 내 공분산 행렬에 대한 분산 행렬의 차원 별 누적 에너지

에 기초하여 획득하는, 디바이스.

청구항 9

제6항에 있어서, 상기 특징 벡터는

결합요인분석 방법에 의해 획득될 수 있는 i-벡터인, 디바이스.

청구항 10

제6항에 있어서, 상기 제어부는

상기 변환된 특징 벡터 및 각 스테이트의 특징 벡터에 대한 스코어링을 적어도 한번 수행하고, 상기 스코어링 결과에 기초하여, 사용자를 인식하는, 디바이스.

청구항 11

제1항 내지 제5항 중 어느 한 항에 있어서, 상기 방법을 구현하기 위한 프로그램이 기록된 컴퓨터로 판독 가능한 기록 매체.

발명의 설명

기술 분야

[0001] 본 발명은 사용자 인식을 위한 특징 벡터를 변환하는 방법 및 장치에 대한 것이다. 보다 상세하게는 사용자 인식을 위해 사용자의 음성 신호로부터 사용자의 고유한 특성인 특징 벡터를 추출하고, 변환하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 화자 인식(speaker recognition) 기술은 사용자의 음성 신호로부터 사용자의 고유한 정보를 추출하여, 특정한 화자가 제시(claim)되었을 경우 발생된 음성이 제시 화자에 해당되는지 여부를 검증하는 기술이다.

[0003] 화자 인식을 위해서는 사용자가 입력한 음성 신호로부터 입력된 음성 신호의 고유한 특성인 특징 벡터를 추출하여야 한다. 특징 벡터는 높은 차수를 가지고 있어 화자 검증 시 계산량이 크므로, 특징 벡터를 추출하는 장치는 LDA(Linear discriminant analysis) 방법을 이용하여 특징 벡터를 변환함으로써 특징 벡터의 차수를 감소시킬 수 있다.

[0004] 그러나, LDA 방법에 의하면, 특징 벡터의 각 클래스가 등분산적(homoscedastic)이라고 가정하나, 실제 특징 벡터의 각 클래스는 이분산적(heteroscedastic)인 특징을 가질 수도 있다. 따라서, LDA 방법에 따라 특징 벡터의 차수가 감소되는 경우, 실제 데이터와는 다른 가정으로 인하여 화자 인식 시스템의 성능이 저하될 수 있다.

발명의 내용

과제의 해결 수단

[0005] 본 발명은 사용자 인식을 위해 사용자의 음성 신호로부터 사용자의 고유한 특성인 특징 벡터를 변환하는 방법 및 장치에 관한 것이다. 보다 상세하게는, 특징 벡터의 차원을 감소시키기 위해 특징 벡터를 추출하고 변환하는 방법 및 장치에 관한 것이다.

발명의 효과

[0006] 일 실시 예에 의하면, 각 클래스의 서로 다른 공분산 행렬을 고려한 HLDA 변환 매트릭스를 이용하여 i-벡터의 차원을 감소시킴으로써 화자 인식 성능이 향상될 수 있다.

도면의 간단한 설명

[0007] 도 1은 일 실시 예에 의한 사용자의 음성 신호로부터 특징 벡터를 추출하고 변환하는 디바이스를 나타낸 블록도이다.

도 2는 일 실시 예에 의한 분산 행렬의 고유값과 대응되는 에너지 값을 나타낸 그래프의 일 예이다.

도 3은 일 실시 예에 의한 특징 벡터를 추출하는 방법을 나타낸 순서도이다.

도 4는 일 실시 예에 의한 특징 벡터에 기초하여 사용자를 인식하는 방법을 나타낸 블록도이다.

도 5는 일 실시 예에 의한 특징 벡터에 기초하여 스코어링을 수행하는 방법을 나타낸 순서도이다.

도 6은 일 실시 예에 의한 디바이스의 내부 구조를 나타낸 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0008] 발명의 실시를 위한 최선의 형태

[0009] 일 실시 예에 의한 특징 벡터를 변환하는 방법에 있어서, 사용자의 발성을 포함하는 음성 신호로부터 특징 시퀀스를 추출하는 단계; 상기 추출된 특징 시퀀스로부터 특징 벡터를 추출하는 단계; 서로 다른 공분산 값에 기초한 확률값에 기초하여, 상기 검출된 특징 벡터의 차원 수를 줄이기 위한 변환 행렬을 획득하는 단계; 상기 변환 행렬을 이용하여 상기 특징 벡터를 변환하는 단계를 포함한다.

[0010] 더하여, 상기 변환 행렬은 HLDA 변환 행렬이다.

[0011] 더하여, 상기 변환 행렬을 획득하는 단계는 상기 변환 행렬의 유효한 차원 p 를 각 화자의 클래스 내 공분산 행렬에 대한 분산 행렬의 차원 별 누적 에너지에 기초하여 획득하는 단계를 포함한다.

[0012] 더하여, 상기 특징 벡터는 결합요인분석 방법에 의해 획득될 수 있는 i -벡터이다.

[0013] 더하여, 상기 변환된 특징 벡터 및 각 스테이트의 특징 벡터에 대한 스코어링을 적어도 한번 수행하는 단계; 상기 스코어링 결과에 기초하여, 사용자를 인식하는 단계를 더 포함한다.

[0014] 일 실시 예에 의한, 특징 벡터를 변환하는 디바이스에 있어서, 사용자의 발성을 포함하는 음성 신호를 수신하는 수신부; 상기 음성 신호로부터 특징 시퀀스를 추출하고, 상기 추출된 특징 시퀀스로부터 특징 벡터를 추출하고, 서로 다른 공분산 값에 기초한 확률값에 기초하여, 상기 추출된 특징 벡터의 차원 수를 줄이기 위한 변환 행렬을 획득하고, 상기 변환 행렬을 이용하여 상기 특징 벡터를 변환하는 제어부를 포함한다.

[0015] 발명의 실시를 위한 형태

[0016] 이하 본 발명의 바람직한 실시 예를 첨부한 도면을 참조하여 상세히 설명한다. 다만, 하기의 설명 및 첨부된 도면에서 본 발명의 요지를 흐릴 수 있는 공지 기능 또는 구성에 대한 상세한 설명은 생략한다. 또한, 도면 전체에 걸쳐 동일한 구성 요소들은 가능한 한 동일한 도면 부호로 나타내고 있음에 유의하여야 한다.

[0017] 이하에서 설명되는 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 한정해서 해석되어서는 아니 되며, 발명자는 그 자신의 발명을 가장 최선의 방법으로 설명하기 위한 용어로 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야만 한다. 따라서 본 명세서에 기재된 실시 예와 도면에 도시된 구성은 본 발명의 가장 바람직한 일 실시 예에 불과할 뿐이고, 본 발명의 기술적 사상을 모두 대변하는 것은 아니므로, 본 출원시점에 있어서 이들을 대체할 수 있는 다양한 균등물과 변형 예들이 있을 수 있음을 이해하여야 한다.

[0018] 첨부 도면에 있어서 일부 구성요소는 과장되거나 생략되거나 또는 개략적으로 도시되었으며, 각 구성요소의 크기는 실제 크기를 전적으로 반영하는 것이 아니다. 본 발명은 첨부한 도면에 그려진 상대적인 크기나 간격에 의해 제한되어지지 않는다.

[0019] 명세서 전체에서 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다. 또한, 명세서에서 사용되는 "부"라는 용어는 소프트웨어, FPGA 또는 ASIC과 같은 하드웨어 구성요소를 의미하며, "부"는 어떤 역할들을 수행한다. 그렇지만 "부"는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. "부"는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 "부"는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다. 구성요소들과 "부"들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 "부"들

로 결합되거나 추가적인 구성요소들과 "부"들로 더 분리될 수 있다.

- [0020] 아래에서는 첨부된 도면을 참고하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 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.
- [0021] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예를 설명한다.
- [0022] 도 1은 일 실시 예에 의한 사용자의 음성 신호로부터 특징 벡터를 추출하고 변환하는 디바이스를 나타낸 블록도이다.
- [0023] 일 실시 예에 의한 디바이스(100)는 사용자에게 의해 이용될 수 있는 단말 장치일 수 있다. 예를 들면, 디바이스(100)는 스마트 TV(television), UHD(ultra high definition) TV, 모니터, PC(Personal Computer), 노트북 컴퓨터, 휴대폰(mobile phone), 태블릿 PC, 내비게이션(navigation) 단말기, 스마트폰(smart phone), PDA(Personal Digital Assistants), PMP(Portable Multimedia Player) 및 디지털방송 수신기를 포함할 수 있다.
- [0024] 일 실시 예에 의한 디바이스(100)는 특징 벡터의 각 클래스 별로 서로 다른 분산값을 가질 수 있는 점을 고려하여 특징 벡터의 차원을 감소시킬 수 있다. 일 실시 예에 있어서, 클래스란 LDA 또는 HLDA 등의 기술에서, 다수의 데이터값이 분류될 수 있는 그룹을 의미한다. 디바이스(100)는 추출된 특징 벡터의 차원을 감소시키기 위해, 변환 행렬을 특징 벡터에 적용시킴으로써, 차원이 감소된 특징 벡터를 획득할 수 있다. 변환 행렬은 HLDA(heteroscedastic LDA) 방법을 통해 획득된 행렬에 기초하여 결정될 수 있다. HLDA 방법을 통해 변환 행렬을 결정하고, HLDA에 의한 변환 행렬을 이용하여 특징 벡터를 변환하는 경우, LDA 변환 행렬과는 달리 각 클래스의 분산값이 동일하다는 제한 사항이 해소(reject)될 수 있다.
- [0025] 디바이스(100)에 의해 추출된 특징 벡터는 사용자를 인식하기 위한 타겟 특징 벡터로 저장되거나, 테스트 특징 벡터로써 타겟 특징 벡터와 비교되어 사용자를 인식하기 위해 이용될 수 있다. 디바이스(100)는 추출된 특징 벡터 및 미리 저장된 특징 벡터를 이용한 스코어링(scoring)을 수행함으로써 사용자를 인식할 수 있다.
- [0026] 도 1을 참조하면, 음성 신호로부터 특징 벡터를 추출하는 디바이스는 특징 검출부(110), 특징 벡터 추출부(120) 및 특징 벡터 변환부(130)를 포함할 수 있다.
- [0027] 특징 검출부(110)는 MFCC(mel frequency cepstral coefficient) 방법을 이용하여 사용자의 음성 신호에 대한 특징값을 검출할 수 있다. MFCC는 인간의 귀가 가지는 비선형적인 주파수 특성을 이용하여 스펙트럼 기반의 유효한 특징 값을 검출하는 방법이다. 특징 검출부(110)에 의해 검출될 수 있는 음성 신호의 특징값은 상술된 MFCC 방법에 한하지 않고 다양한 방법에 의해 검출될 수 있다. 예를 들면, 음성 신호의 특징값은 음성 신호로부터 추출된 주파수 시퀀스에 대한 특징 파라미터 값으로 검출될 수 있다.
- [0028] 특징 벡터 추출부(120)는 특징 검출부(110)에 의해 검출된 특징값에 기초하여 음성 신호에 대한 특징 벡터를 추출할 수 있다.
- [0029] 특징 벡터 추출부(120)는 에르고딕 HMM(ergodic hidden markov model)을 이용하여 음성 신호를 분류할 수 있다. 특징 벡터 추출부(120)는 에르고딕 HMM을 이용하여 음성 신호를 음성 카테고리 및 대응되는 상태(state)로 분류할 수 있다. 음성 카테고리는 음성의 특징, 예를 들면, 주파수, 크기 등의 음성적인 특징에 따라 분류될 수 있다. 특징 벡터 추출부(120)는 에르고딕 HMM에 한하지 않고 다양한 방법을 이용하여 음성 신호를 분류할 수 있다.
- [0030] 한편, 결합요인분석(joint factor analysis) 방법에 의하면, 화자의 발화(speaker utterance)는 화자의 서브 스페이스와 채널의 서브 스페이스로 구성된 슈퍼 벡터로 표현될 수 있다. 그러나, 총 가변성 공간(total variability space)에서 화자의 발화는 수학식 1과 같이 하나의 스페이스로 구성된 슈퍼 벡터로 구성될 수 있다.

수학식 1

[0031] $M = m + T\omega$

[0032] 슈퍼 벡터 M은 화자의 발화를 나타내고, m은 화자와 채널에 독립적인 슈퍼 벡터, T는 총 가변성 행렬, ω 는 총 가변성 팩터(factor), 즉, i-벡터(i-vector)를 나타낸다. M, m, T 값은 결합요인분석 방법에 따라 획득될 수 있다.

[0033] i-벡터 ω 는, 이하 수학식 2 내지 4와 같이 바움-웰치 통계(baum-welch statistics)을 통하여 결정될 수 있다.

수학식 2

[0034]
$$N_c = \sum_{t=1}^L P(c|y_t, \lambda)$$

[0035] 수학식 2에서, N_c 는 CFxCF 차원의 대각 블록(diagonal block)로 구성된 행렬이며, y_t 는 특징 검출부(110)에 의해 검출될 수 있는 길이가 L 프레임인 특징 시퀀스 $y = \{y_1, y_2, \dots, y_L\}$ 중 하나를 나타낸다.

[0036] λ 는 GMM-UBM(Gaussian Mixture Models-Universal Background Model)에서의 파라미터인 $\lambda = \{w_c, m_c, \Sigma_c\}$, $c = 1, \dots, C$ 이다. λ 에 포함된 Σ 은 CFxCF 차원의 대각 공분산(diagonal covariance) 행렬이고, C는 가우시안 혼합 모델의 컴포넌트 차수이다. GMM-UBM은 패턴을 분류함에 있어 데이터들의 분포특성을 분류하는 방법에 관한 것으로, 파라미터 λ 에 따라 데이터 분포에 대한 모델이 결정될 수 있다.

수학식 3

[0037]
$$\hat{F}_c = \sum_{t=1}^L P(c|y_t, \lambda)(y_t - m_c)$$

[0038] 수학식 3에서, F는 특징 공간의 차수(feature space of dimension)를 나타낸다. 특징 공간은 특징 벡터가 정의될 수 있는 n차원의 공간을 의미한다.

[0039] 수학식 2 및 3의 N, F값은 바움-웰치 통계의 각 파라미터에 대한 확률 값에 기초하여 획득될 수 있다.

수학식 4

[0040]
$$\omega = (I + T' \Sigma^{-1} N T)^{-1} T' \Sigma^{-1} \hat{F}$$

[0041] 수학식 4와 같이, T, N, Σ 값에 기초하여, i-벡터 ω 값이 결정될 수 있다.

[0042] 각 상태(state)에는 GMM(Gaussian Mixture Models) 파라미터가 존재한다. 각 파라미터는 화자를 나타내는 개별 모델을 의미한다. GMM 파라미터는 수학식 5와 같이 나타낼 수 있다.

수학식 5

[0043]
$$\lambda_s = \{w_c^s, m_c^s, \Sigma_c^s\}$$

[0044] 그리고, 수학식 1은 각 상태에 대한 파라미터를 포함하여 수학식 6과 같이 나타낼 수 있다.

수학식 6

$$M = m_s + T_s \omega_s$$

[0045]

[0046] m_s 는 화자 및 채널에 독립적이고, 음성 카테고리(phonetic category)에 종속적인 슈퍼 벡터를 나타내며, T_s 는 음성 카테고리 s 에 대한 총 가변성 행렬, ω_s 는 음성 카테고리 s 에 대한 i -벡터를 나타낸다.

[0047] 특징 벡터 추출부(120)는 각 스테이트에 대한 ω_s 를 수학식 4에 따라 결정할 수 있다.

[0048] 특징 벡터 변환부(130)는 특징 벡터 추출부(120)에 의해 결정된 각 스테이트에 대한 ω_s 의 차원을 줄이기 위해 이하 수학식 7과 같이 HLDA(heteroscedastic LDA) 변환 매트릭스 A 를 q 차원의 ω_s 에 적용할 수 있다. 특징 벡터 변환부(130)에 의해 변환된 ω_s 의 차원이 감소됨에 따라 ω_s 를 포함하는 연산의 연산량이 감소될 수 있다.

수학식 7

$$\hat{\omega} = A\omega = \begin{bmatrix} A_{[p]} \omega \\ A_{[N-p]} \omega \end{bmatrix}$$

[0049]

[0050] A 는 $M \times N$ 매트릭스이며, A 는 첫번째 p 행까지 유효한 차원(useful dimension)을 포함하는 $A_{[p]}$ 와 나머지 $N-p$ 행을 포함하는 $A_{[N-p]}$ 를 포함한다. $N-p$ 차원은 불필요한 차원(nuisance dimension)으로 취급되어, $N-p$ 차원에서의 $A_{[N-p]} \omega$ 는 불필요한 정보로 취급될 수 있고, p 차원까지의 $A_{[p]} \omega$ 가 유효한 값으로 이용될 수 있다. 따라서, i -벡터

$\hat{\omega}$ 의 차원은 N 보다 낮은 차원인 p 차원으로 변환될 수 있다.

[0051] LDA에서는 모든 클래스의 공분산 행렬이 동일하다고 가정하나, 이러한 가정은 실제 데이터와 잘 부합되지 않는다. 따라서, 일 실시 예에 의한 특징 벡터 변환부(130)는 LDA 변환 매트릭스 대신 각 클래스의 서로 다른 공분산 행렬을 고려한 HLDA 변환 매트릭스를 i -벡터에 적용할 수 있다.

[0052] 특징 벡터 변환부(130)는 ω_s 를 HLDA 변환 매트릭스 A 를 이용하여 변환함으로써, ω_s 를 이용한 연산의 연산량을 줄일 수 있고, 실제 데이터와 다른 가정을 해소함에 따라 화자간 다양성을 늘리고, 동일 화자의 다양성을 줄일 수 있다.

[0053] 또한, 특징 벡터 변환부(130)는 이하 수학식 8과 같이 HLDA 변환 매트릭스 A 대신 LDA 및 HLDA의 변환 매트릭스가 결합된 $M \times q$ 차원의 UHLDA(unified HLDA) 변환 매트릭스 C 를 이용하여 ω_s 를 변환할 수 있다. 특징 벡터 변환부(130)는 수학식 7과 같이 A 대신 C 를 ω 에 적용함으로써 ω 값을 변환시킬 수 있다.

수학식 8

$$C = [A_{q/2} \ W_{q/2}]$$

[0054]

$$A_{q/2} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,q/2} \\ \vdots & \ddots & \vdots \\ a_{M,1} & \cdots & a_{M,q/2} \end{bmatrix} \quad W_{q/2} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,q/2} \\ \vdots & \ddots & \vdots \\ w_{M,1} & \cdots & w_{M,q/2} \end{bmatrix}$$

[0055]

[0056] W 는 $M \times N$ 의 LDA 변환 매트릭스이다. $A_{q/2}$ 및 $W_{q/2}$ 는 각각 A 와 W 의 $q/2$ 행의 서브 스페이스이다. 수학식 8의 $A_{q/2}$ 및 $W_{q/2}$ 에 한하지 않고, C 는 A 및 W 의 일부의 서브 스페이스로 구성될 수 있다.

[0057] LDA 및 HLDA의 변환 매트릭스 A 및 W 는 각각 LDA 및 HLDA 방법에 따라 획득될 수 있으며, 이에 한하지 않고, 다

양한 방법에 따라 획득될 수 있다.

[0058] 예를 들면, 상술된 HLDA 변환 매트릭스 A는 ML(maximum likelihood) 추정(estimation) 및 EM(Expectation Maximization) 알고리즘 또는, 다른 HLDA 방법인 SHLDA(smooth HLDA) 등에 기초하여 결정될 수 있다.

[0059] 특징 벡터 변환부(130)는 ML 추정 방법을 이용하여, 각 클래스의 평균값 및 공분산 값이 각 클래스 별로 동일한 경우와 서로 다른 경우에 대한 확률 값에 따라 HLDA 변환 행렬을 구할 수 있다. 특징 벡터 변환부(130)는 각 클래스의 평균값 및 공분산 값이 p차원까지는 서로 다른 값인 것으로 가정할 수 있고, n-p차원부터 n차원까지는 서로 동일한 값으로 가정할 수 있다.

[0060] 상술된 수학식 7의 유효한 차원 p는 이하 수학식 9 내지 11을 통해 결정될 수 있다.

수학식 9

$$\sum^{(j)} = \frac{1}{N_j} \sum_{i=1}^{N_j} (\omega_i^{(j)} - \bar{\omega}^{(j)}) (\omega_i^{(j)} - \bar{\omega}^{(j)})'$$

[0061] 수학식 9의 $\sum^{(j)}$ 는 j 화자의 클래스 내 i-벡터의 공분산 행렬을 나타낸다.

수학식 10

$$S_w = \sum_{j=1}^S \frac{1}{N_j} \sum^{(j)}$$

[0062] 수학식 10의 S_w 는 LDA 방법에서와 같이 각 화자들의 클래스 내 i-벡터의 공분산 행렬이 등분산적이라고 가정하고, 획득된 클래스 내 공분산 매트릭스를 나타낸다.

수학식 11

$$\sigma_{S_w} = \sum_{j=1}^S \frac{1}{N_j} (\sum^{(j)} - S_w) (\sum^{(j)} - S_w)'$$

[0063] 수학식 11의 σ_{S_w} 는 각 화자의 클래스 내 공분산 매트릭스에 대한 분산 행렬을 나타낸다. 분산 행렬 σ_{S_w} 로부터 고유 값 분해(Eigen value decomposition) 방법을 통해 고유값이 획득될 수 있다. 특징 벡터 변환부(130)는 내림차수로 정렬된 분산 행렬의 고유값으로부터 차원수 별로 누적 에너지를 구하여, 소정 에너지 이상이 분포된 차원 개수를 유효한 차원 p로 결정할 수 있다.

[0064] HLDA 방법에 의한 유효한 차원 p는 상술된 실시예에 한하지 않고 다양한 방법으로 획득될 수 있다.

[0065] 더하여, 특징 벡터 변환부(130)는 PLDA(probabilistic Linear Discriminant Analysis) 방법을 이용하여 변환된 특징 벡터의 노이즈 데이터를 제거할 수 있다.

[0066] 이하 도 2를 참조하여, 유효한 차원 p를 결정하는 방법을 더 자세히 설명하기로 한다.

[0067] 도 2는 일 실시 예에 의한 분산 행렬의 고유값과 대응되는 에너지 값을 나타낸 그래프의 일 예이다.

[0068] 도 2를 참조하면, 분산 행렬의 각 고유값과 대응되는 에너지 값을 나타낸 그래프가 도시되어 있다. 특징 벡터 변환부(130)는 상술한 바와 같이 분산 행렬의 고유값으로부터 차원수 별로 에너지를 구할 수 있다. 고유값의 누적 에너지는 0부터 해당 고유값까지의 그래프의 면적으로부터 결정될 수 있다.

[0069] 유효한 차원을 결정하기 위한 누적 에너지의 기준값을 전체 에너지의 90%라고 가정하면, 특징 벡터 변환부(130)는 누적 에너지가 전체 에너지의 90% 이상의 값을 가지는 고유값을 결정할 수 있다. 도 2의 그래프에서 고유

값 250인 지점에서의 누적 에너지가 전체 에너지의 90%가 되므로, 특징 벡터 변환부(130)는 250인 고유값과 대응되는 차수를 유효한 차원 p로 결정할 수 있다.

[0073] 도 3은 일 실시 예에 의한 특징 벡터를 추출하는 방법을 나타낸 순서도이다.

[0074] 도 3을 참조하면, 단계 S301에서, 디바이스(100)는 입력된 음성 신호로부터 특징값을 검출할 수 있다. 예를 들면, 디바이스(100)는 입력된 음성 신호로부터 MFCC 방법에 기초하여 특징 시퀀스를 검출할 수 있다.

[0075] 단계 S303에서, 디바이스(100)는 단계 S301에서 검출된 특징값에 대한 특징 벡터를 검출할 수 있다. 일 실시 예에 의한 특징 벡터는 결합 요인 분석 방법에 의한 i-벡터일 수 있으며, 바움-웰치 통계에 의하여 획득될 수 있다.

[0076] 단계 S305에서, 디바이스(100)는 각 클래스의 서로 다른 공분산 행렬을 고려한 변환 행렬을 획득할 수 있다. 예를 들면, 디바이스(100)는 ML 추정 방법에 따라, 각 클래스 간 서로 다른 공분산 값에 기초한 확률값에 기초하여, 변환 행렬을 획득할 수 있다.

[0077] 단계 S305에서 획득될 수 있는 변환 행렬은 HLDA 변환 행렬일 수 있다. HLDA 변환 행렬은 LDA 변환 행렬과는 달리, 각 클래스의 서로 다른 공분산 행렬에 기초하여 획득될 수 있다. 따라서, 디바이스(100)는 동일한 공분산 행렬이 적용되는 점을 가정하는 LDA 변환 행렬을 이용하는 것보다 실제 데이터의 공분산 행렬을 더 잘 반영하여 i-벡터를 변환할 수 있다.

[0078] 단계 S307에서, 디바이스(100)는 단계 S205에서 획득한 변환 행렬을 이용하여 특징 벡터를 변환할 수 있다.

[0079] 이하 도 4 내지 도 5를 참조하여, 특징 벡터에 기초하여 사용자를 인식하는 방법에 관하여 더 자세히 설명하기로 한다.

[0080] 도 4는 일 실시 예에 의한 특징 벡터에 기초하여 사용자를 인식하는 방법을 나타낸 블록도이다.

[0081] 도 4를 참조하면, 디바이스(100)는 410에서, 입력된 화자 발성을 포함하는 음성 신호로부터 MFCC 방법에 따라 특징값을 추출할 수 있다. 예를 들면, 특징값은 특징 시퀀스 y로 추출될 수 있다.

[0082] 디바이스(100)는 420에서, 스테이트 1의 슈퍼 벡터 m을 획득하고, 430에서, 특징 시퀀스 y에 기초하여 바움-웰치 통계에 따라 필요한 파라미터들을 획득할 수 있다. 440에서, i 벡터는 330에서 획득된 파라미터들과 총 가변성 행렬 T에 기초하여 획득될 수 있다.

[0083] 470에서 획득될 수 있는 i 벡터 역시 340의 i 벡터와 동일한 방법으로 450 내지 470을 통해 획득될 수 있다. 440에서 획득된 i 벡터는 현재 입력된 화자 발성을 포함한 음성 신호로부터 획득된 것인데 반하여, 470에서 획득될 수 있는 i 벡터는 사용자 인식을 위해 미리 획득된 특징 벡터일 수 있다.

[0084] 480에서, 디바이스(100)는 현재 입력된 음성 신호에 대한 i 벡터와 사용자 인식을 위해 비교하고자 하는 i 벡터를 이용하여 스코어링을 수행할 수 있다. 스코어링은 이하 수학적 식 12와 같이 수행될 수 있다.

수학적 식 12

$$scored(\omega_{target}, \omega_{test}) = \frac{\langle \omega_{target}, \omega_{test} \rangle}{\|\omega_{target}\| \|\omega_{test}\|}$$

[0085]

[0086] 수학적 식 12에서, ω_{target} 은 미리 획득된 i-벡터를 나타내고, ω_{test} 은 현재 입력된 음성 신호로부터 획득된 i-벡터를 나타낸다.

[0087] 수학적 식 12의 수식은 CDS(cosine distance scoring) 방법에 따른 것이며, 이에 한하지 않고, 다양한 방법에 따라 스코어링이 수행될 수 있다. 디바이스(100)는 수학적 식 12에 따라 획득된 스코어링 값에 따라 입력된 음성 신호의 화자를 식별할 수 있다. 디바이스(100)는 현재 입력된 음성 신호에 기초하여 획득된 i 벡터 값 및 각 스테이트에 대한 i 벡터 값에 대해 각각 스코어링을 수행함으로써, 입력된 음성 신호의 화자를 식별할 수 있다.

[0088] 도 5는 일 실시 예에 의한 특징 벡터에 기초하여 스코어링을 수행하는 방법을 나타낸 순서도이다.

[0089] 도 5를 참조하면, 단계 S501에서, 디바이스(100)는 입력된 음성 신호에 대한 특징 벡터를 획득할 수 있다. 특징 벡터는 상술된 특징 벡터를 추출하는 방법에 따라 획득될 수 있다. 일 실시 예에 의한 특징 벡터는 결합 요인

분석 방법에 따라 획득될 수 있는 i-벡터일 수 있다.

- [0090] 단계 S503에서, 디바이스(100)는 적어도 하나의 스테이트에 대한 특징 벡터를 획득할 수 있다. 디바이스(100)는 사용자 인식을 위하여 미리 저장된 특징 벡터 값을 획득할 수 있다. 디바이스(100)는 각 스테이트에 대한 특징 벡터를 하나 이상 획득할 수 있다.
- [0091] 단계 S505에서, 디바이스(100)는 단계 S501에서 획득된 특징 벡터와 단계 S503에서 획득된 적어도 하나의 특징 벡터에 대하여 스코어링을 수행함으로써 사용자 인식을 수행할 수 있다. 디바이스(100)는 단계 S501에서 획득된 특징 벡터 및 각 스테이트의 특징 벡터들에 기초하여 획득된 스코어링 값을 서로 비교함으로써 현재 입력된 음성 신호의 특징 벡터와 대응되는 스테이트를 결정할 수 있다. 디바이스(100)는 결정된 스테이트에 기초하여 입력된 음성 신호의 사용자를 인식할 수 있다.
- [0092] 이하 도 6을 참조하여, 디바이스의 내부 구성요소에 대하여 자세히 설명하기로 한다.
- [0093] 도 6은 일 실시 예에 의한 디바이스의 내부 구조를 나타낸 블록도이다.
- [0094] 도 6을 참조하면, 디바이스(600)는 수신부(610) 및 제어부(620)를 포함할 수 있다.
- [0095] 수신부(610)는 사용자의 발성을 포함하는 음성 신호를 수신할 수 있다. 예를 들어, 수신부(610)는 마이크를 통해 사용자의 발성을 포함하는 음성 신호를 수신할 수 있다.
- [0096] 제어부(620)는 수신부(610)에 의해 수신된 음성 신호에 기초하여 특징 벡터를 추출할 수 있다. 제어부(620)는 결합 요인 분석 방법에 따라 i-벡터를 추출할 수 있고, HLDA 변환 매트릭스를 이용하여 i-벡터의 차원을 감소시킬 수 있다. 또한, 제어부(620)는 현재 입력된 음성 신호와 대응되는 특징 벡터 및 각 스테이트에 대한 특징 벡터에 대해 스코어링을 수행함으로써 현재 입력된 음성 신호와 대응되는 화자를 식별할 수 있다.
- [0097] 일 실시 예에 의하면, 각 클래스의 서로 다른 공분산 행렬을 고려한 HLDA 변환 매트릭스를 이용하여 i-벡터의 차원을 감소시킴으로써 화자 인식 성능이 향상될 수 있다.
- [0098] 일부 실시 예에 의한 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 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 본 발명을 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.
- [0099] 비록 상기 설명이 다양한 실시예들에 적용되는 본 발명의 신규한 특징들에 초점을 맞추어 설명되었지만, 본 기술 분야에 숙달된 기술을 가진 사람은 본 발명의 범위를 벗어나지 않으면서도 상기 설명된 장치 및 방법의 형태 및 세부 사항에서 다양한 삭제, 대체, 및 변경이 가능함을 이해할 것이다. 따라서, 본 발명의 범위는 상기 설명에서보다는 첨부된 특허청구범위에 의해 정의된다. 특허청구범위의 균등 범위 안의 모든 변형은 본 발명의 범위에 포함된다.

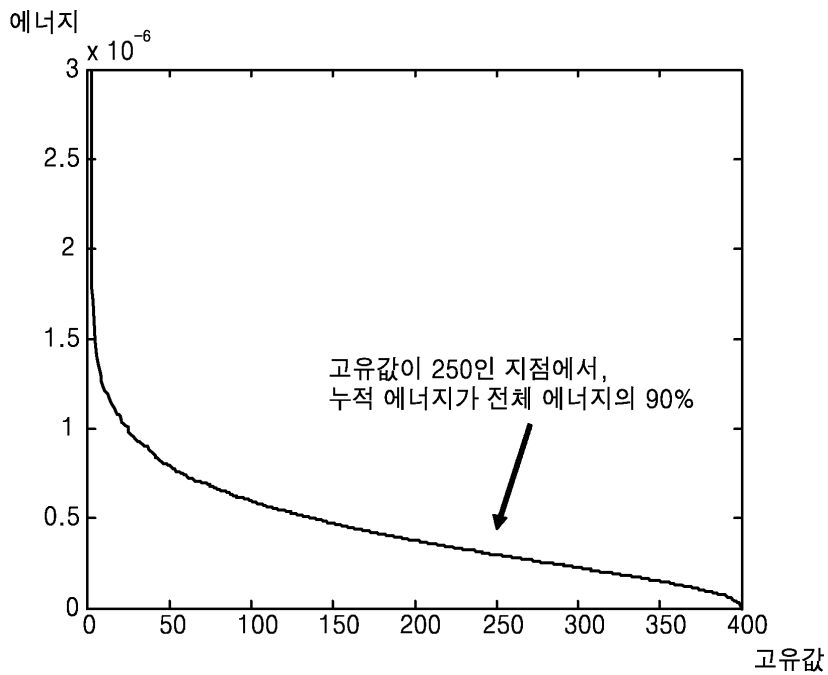
도면

도면1

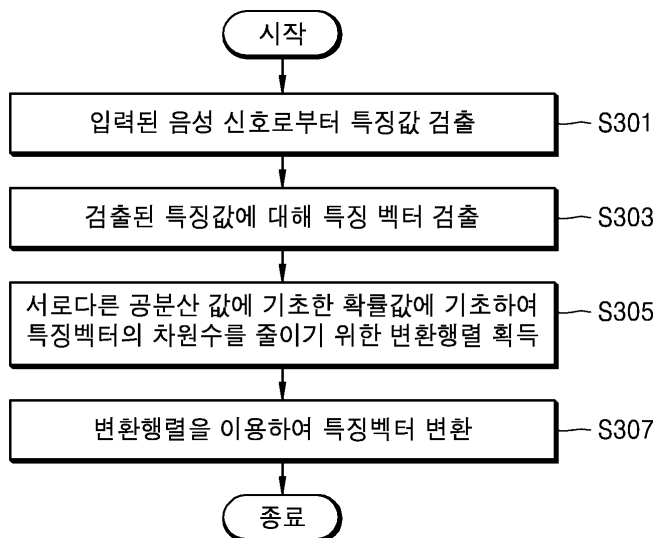
100



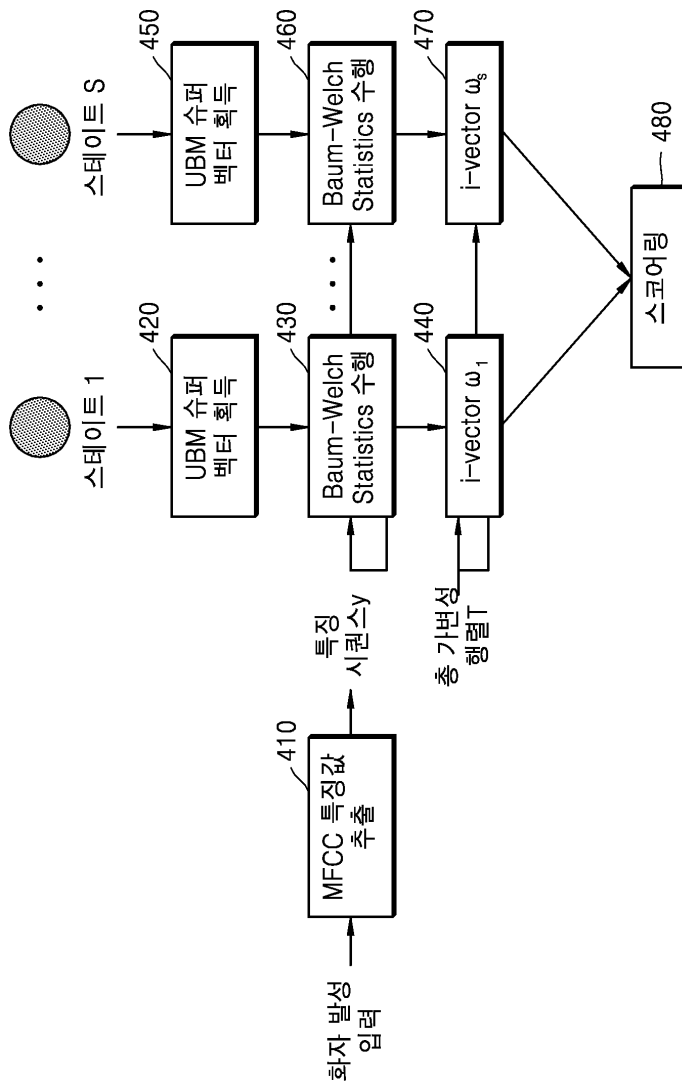
도면2



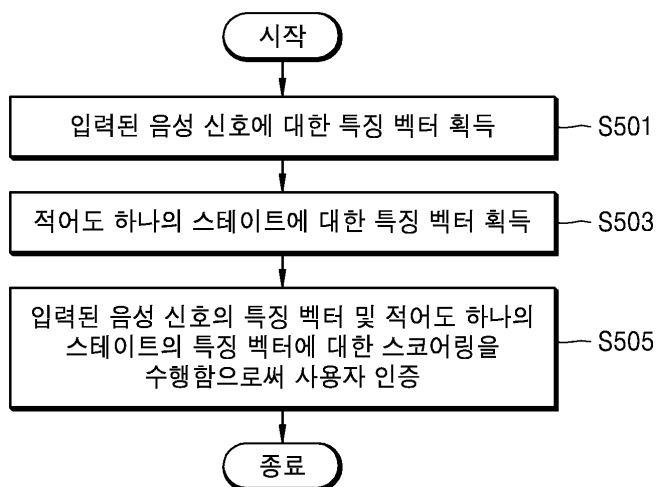
도면3



도면4



도면5



도면6

600

