

	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2012-0021428 (43) 공개일자 2012년03월09일
(51) 국제특허분류(Int. Cl.) G10L 15/14 (2006.01) G10L 15/00 (2006.01) G10L 21/02 (2006.01)		(71) 출원인 인하대학교 산학협력단 인천광역시 남구 인하로 100, 인하대학교 (용현동)
(21) 출원번호 10-2010-0074108		(72) 발명자 장준혁 서울특별시 서초구 서초대로52길 41, B동 301호 (서초동)
(22) 출원일자 2010년07월30일 심사청구일자 2010년07월30일		강상익 울산광역시 남구 둔질로7번길 16-2 (신정동)
		(74) 대리인 김건우

전체 청구항 수 : 총 4 항

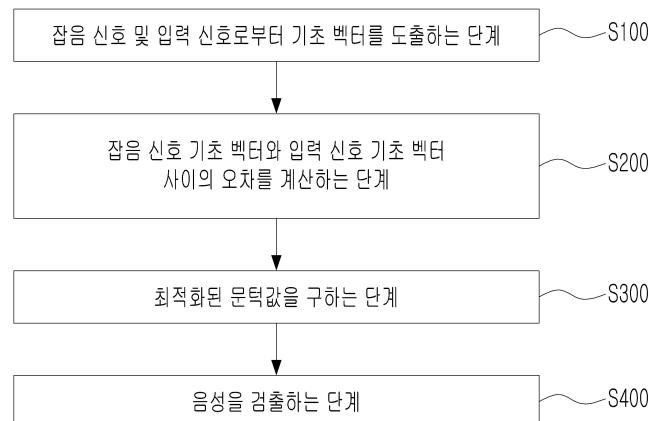
(54) 발명의 명칭 비음수 행렬 인수분해 기반의 음성 검출 방법

(57) 요약

본 발명은 비음수 행렬 인수분해 기반의 음성 검출 방법에 관한 것으로서, 보다 구체적으로는 음성 검출 방법으로서, (1) 비음수 행렬 인수분해 기법을 이용하여 잡음 신호 및 입력 신호로부터 기초 벡터를 도출하는 단계, (2) 상기 단계 (1)에서 도출한 잡음 신호 기초 벡터와 입력 신호 기초 벡터 사이의 오차를 계산하는 단계, 및 (3) 상기 단계 (2)에서 계산된 오차에 문턱 값을 적용하여 음성을 검출하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.

본 발명에서 제안하고 있는 비음수 행렬 인수분해 기반의 음성 검출 방법에 따르면, 잡음 신호로부터 추출한 기초 벡터들과 입력 신호와의 오차를 계산하고 이를 기반으로 음성의 활성 구간을 구분하되, 잡음 추정 구간에서의 오차 값 분포에 따라 잡음 환경을 추정하여 최적의 문턱 값을 선정하여 오차에 적용하여 음성 신호를 검출함으로써, 신호 대 잡음비가 낮아 상대적으로 열악한 잡음 환경에서도 우수한 성능의 음성 검출이 가능하다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호 2008-F-045-01

부처명 정보통신부

연구사업명 IT산업경쟁력강화산업

연구과제명 장애인 및 고령자를 위한 Digital Guardian 기술개발

주관기관 한국과학기술연구원

연구기간 2010.03.01 ~ 2012.02.28

특허청구의 범위

청구항 1

음성 검출 방법으로서,

- (1) 비음수 행렬 인수분해 기법을 이용하여 잡음 신호 및 입력 신호로부터 기초 벡터를 도출하는 단계;
- (2) 상기 단계 (1)에서 도출한 잡음 신호 기초 벡터와 입력 신호 기초 벡터 사이의 오차를 계산하는 단계; 및
- (3) 상기 단계 (2)에서 계산된 오차에 문턱 값을 적용하여 음성을 검출하는 단계를 포함하는 것을 특징으로 하는 비음수 행렬 인수분해 기반의 음성 검출 방법.

청구항 2

제1항에 있어서,

통계 모델 기반의 음성 검출 방법에 의해 추정된 음성부재구간에서 비음수 행렬 인수분해 기법을 이용하여 도출된 잡음 신호 및 입력 신호의 오차 값 분포에 따른 최적화된 문턱 값을 구하고,

상기 단계 (3)의 문턱 값을 상기 최적화된 문턱 값으로 하여 음성을 검출하는 것을 특징으로 하는 비음수 행렬 인수분해 기반의 음성 검출 방법.

청구항 3

제2항에 있어서, 상기 최적화된 문턱 값은,

각 잡음 환경별로 서로 다른 값인 것을 특징으로 하는 비음수 행렬 인수분해 기반의 음성 검출 방법.

청구항 4

제1항에 있어서,

상기 기초 벡터는 통계적 모델을 이용하여 추정된 a posteriori signal-to-noise ratio(SNR)와 a priori SNR인 것을 특징으로 하는 비음수 행렬 인수분해 기반의 음성 검출 방법.

명세서

기술분야

[0001] 본 발명은 음성 검출 방법에 관한 것으로서, 보다 구체적으로는 비음수 행렬 인수분해 기반의 음성 검출 방법에 관한 것이다.

배경기술

[0002] 음성과 비음성 구간을 검출하는 음성 검출기(voice activity detector, VAD)는 음성 부호화, 음성인식 그리고 음향학적 반향제거기 등 음성 통신 시스템에서 많이 적용된다. 특히, 음성 통신 시스템의 대역폭을 효율적으로 사용하기 위해서 필수적으로 요구되며 최근 Ephraim과 Malah의 연구에서 시작된 minimum mean square error(MMSE) 기반의 음성 향상 기법에 사용된 음성의 존재와 부재에 대한 통계적 모델을 음성 검출기에 적용한 것이 우수한 성능을 가진 것으로 알려져 있다.

[0003] 기존의 음성 검출 알고리즘에서는 음성과 잡음에 대한 통계적 모델을 기반으로 음성의 활성 여부를 판단하기 때문에, 신호대 잡음 비가 높은 신호에서는 비교적 정확한 음성 검출이 가능하지만 상대적으로 열악한 잡음 환경

에서는 음성 검출의 성능이 급격히 저하되는 단점이 있다.

[0004] 한편, 비음수 행렬 인수분해는 이미지 등의 패턴 학습과 패턴 인식에 우수한 성능을 보이는 알고리즘이다. 특히 다수의 입력 데이터에서 최적의 기초 패턴을 분리하여 이들의 선형 조합으로 전체 데이터를 근사할 수 있기 때문에 데이터 특징 추출에도 유용하다. 그러나 아직까지 비음수 행렬 인수분해를 이용한 효율적인 음성 검출 방법은 제안되지 못한 상황이다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 기존에 제안된 방법들의 상기와 같은 문제점들을 해결하기 위해 제안된 것으로서, 잡음 신호로부터 추출한 기초 벡터들과 입력 신호와의 오차를 계산하고 이를 기반으로 음성의 활성 구간을 구분하되, 잡음 추정 구간에서의 오차 값 분포에 따라 잡음 환경을 추정하여 최적의 문턱 값을 선정하여 오차에 적용하여 음성 신호를 검출함으로써, 신호대 잡음 비가 낮아 상대적으로 열악한 잡음 환경에서도 음성 검출의 성능이 우수한, 비음수 행렬 인수분해 기반의 음성 검출 방법을 제공하는 것을 그 목적으로 한다.

과제의 해결 수단

[0006] 상기한 목적을 달성하기 위한 본 발명의 특징에 따른 비음수 행렬 인수분해 기반의 음성 검출 방법은,
[0007] (1) 비음수 행렬 인수분해 기법을 이용하여 잡음 신호 및 입력 신호로부터 기초 벡터를 도출하는 단계;
[0008] (2) 상기 단계 (1)에서 도출한 잡음 신호 기초 벡터와 입력 신호 기초 벡터 사이의 오차를 계산하는 단계; 및
[0009] (3) 상기 단계 (2)에서 계산된 오차에 문턱 값을 적용하여 음성을 검출하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.

[0010] 바람직하게는,
[0011] 통계 모델 기반의 음성 검출 방법에 의해 추정된 음성 부재구간에서 비음수 행렬 인수분해 기법을 이용하여 도출된 잡음 신호 및 입력 신호의 오차 값 분포에 따른 최적화된 문턱 값을 구하고,
[0012] 상기 단계 (3)의 문턱 값을 상기 최적화된 문턱 값으로 하여 음성을 검출할 수 있다.

[0013] 더욱 바람직하게는, 상기 최적화된 문턱 값은,

[0014] 각 잡음 환경별로 서로 다른 값일 수 있다.

[0015] 바람직하게는,

[0016] 상기 기초 벡터는 통계적 모델을 이용하여 추정된 a posteriori signal-to-noise ratio(SNR)와 a priori SNR일 수 있다.

발명의 효과

[0017] 본 발명에서 제안하고 있는 비음수 행렬 인수분해 기반의 음성 검출 방법에 따르면, 잡음 신호로부터 추출한 기초 벡터들과 입력 신호와의 오차를 계산하고 이를 기반으로 음성의 활성 구간을 구분하되, 잡음 추정 구간에서의 오차 값 분포에 따라 잡음 환경을 추정하여 최적의 문턱 값을 선정하여 오차에 적용하여 음성 신호를 검출함으로써, 신호대 잡음 비가 낮아 상대적으로 열악한 잡음 환경에서도 우수한 성능의 음성 검출이 가능하다.

도면의 간단한 설명

[0018] 도 1은 본 발명의 일 실시예에 따른 비음수 행렬 인수분해 기반의 음성 검출 방법의 흐름을 도시한 도면.

도 2는 통계적 모델 기반의 음성 검출 방법을 통해 잡음 환경이라고 추정된 구간에서의 비음수 행렬 인수분해 결과의 분포를 도시한 도면.

도 3은 백색 잡음 환경에서 잡음 대 음성 신호 비에 대한 비음수 행렬 인수분해 기반의 음성 검출 결과를 도시한 도면.

도 4는 잡음 환경에 따른 음성 검출 확률의 변화를 도시한 도면.

발명을 실시하기 위한 구체적인 내용

[0019] 이하, 첨부된 도면을 참조하여 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있도록 바람직한 실시예를 상세히 설명한다. 다만, 본 발명의 바람직한 실시예를 상세하게 설명함에 있어, 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략한다. 또한, 유사한 기능 및 작용을 하는 부분에 대해서는 도면 전체에 걸쳐 동일한 부호를 사용한다.

[0020] 덧붙여, 명세서 전체에서, 어떤 부분이 다른 부분과 ‘연결’ 되어 있다고 할 때, 이는 ‘직접적으로 연결’ 되어 있는 경우뿐만 아니라, 그 중간에 다른 소자를 사이에 두고 ‘간접적으로 연결’ 되어 있는 경우도 포함한다. 또한, 어떤 구성요소를 ‘포함’ 한다는 것은, 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있다는 것을 의미한다.

[0021] 도 1은 본 발명의 일실시예에 따른 비음수 행렬 인수분해 기반의 음성 검출 방법의 흐름을 도시한 도면이다. 도 1에 도시된 바와 같이, 본 발명의 일실시예에 따른 비음수 행렬 인수분해 기반의 음성 검출 방법은, 잡음 신호 및 입력 신호로부터 기초 벡터를 도출하는 단계(S100), 잡음 신호 기초 벡터와 입력 신호 기초 벡터 사이의 오차를 계산하는 단계(S200), 및 음성을 검출하는 단계(S400)를 포함하여 구현될 수 있으며, 최적화된 문턱 값을 구하는 단계(S300)를 더 포함하여 구현될 수 있다.

[0022] 즉, 본 발명의 일실시예에 따른 비음수 행렬 인수분해 기반의 음성 검출 방법은, 먼저 비음수 행렬 인수분해 기법을 이용하여 잡음 신호 및 입력 신호로부터 기초 벡터를 도출한다(S100). 도출한 잡음 신호 기초 벡터와 입력 신호 기초 벡터 사이의 오차를 계산하여(S200), 계산된 오차에 문턱 값을 적용하여 음성을 검출하게 된다(S400). 이때, 음성 검출 성능을 향상시키기 위해 최적화된 문턱 값을 사용할 수 있는데, 이를 위해 통계 모델 기반의 음성 검출 방법에 의해 추정된 음성 부재구간에서 비음수 행렬 인수분해 기법을 이용하여 도출된 잡음 신호 및 입력 신호의 오차 값 분포에 따른 최적화된 문턱 값을 구할 수 있다(S300).

[0023] 이하에서는, 본 발명의 일실시예에 따른 비음수 행렬 인수분해 기반의 음성 검출 방법의 세부적인 내용에 대해 상세히 설명하도록 한다.

[0024] 본 발명의 일실시예에 따른 비음수 행렬 인수분해 기반의 음성 검출 방법을 위해서는 통계적 모델을 기반으로 하는 음성 검출 방법에 대한 분석이 선행되어야 한다. 먼저, 통계적 모델을 기반으로 하는 음성 검출 방법은, 시간 축 상에서 원래의 음성신호 $x(t)$ 에 잡음신호 $n(t)$ 이 인가된 입력신호 $y(t)$ 를 discrete Fourier transform(DFT)을 통해 주파수 축으로 변환하여 아래 수학적 식 1과 같이 표현한다.

수학적 식 1

$$Y(t) = X(t) + N(t)$$

[0026] 여기서, $Y(t)=[Y_1, Y_2, \dots, Y_M]$, $X(t)=[X_1, X_2, \dots, X_M]$ 및 $N(t)=[N_1, N_2, \dots, N_M]$ 는 각각 잡음에 오염된 음성신호, 원래의

음성신호 및 잡음신호의 DFT 계수 벡터를 나타낸다. 주어진 가설 H_0 , H_1 이 각각 음성의 부재와 존재를 표현한다고 하면 각 주파수 채널별로 다음 수학적 식 2와 같이 기술된다.

수학적 식 2

$$\begin{aligned} H_0: \text{speech absent} &: Y_k(t) = N_k(t) \\ H_1: \text{speech present} &: Y_k(t) = X_k(t) + N_k(t). \end{aligned}$$

음성과 잡음신호의 스펙트럼이 복소 가우시안 분포를 따른다는 가정으로부터 가설 H_0 와 H_1 을 조건으로 한 확률밀도함수는 다음 수학적 식 3과 같이 주어진다.

수학적 식 3

$$\begin{aligned} p(Y_k|H_0) &= \frac{1}{\pi\lambda_{d,k}} \exp\left\{-\frac{|Y_k|^2}{\lambda_{d,k}}\right\} \\ p(Y_k|H_1) &= \frac{1}{\pi[\lambda_{d,k} + \lambda_{x,k}]} \exp\left\{-\frac{|Y_k|^2}{\lambda_{d,k} + \lambda_{x,k}}\right\} \end{aligned}$$

여기서, $\lambda_{x,k}$ 와 $\lambda_{d,k}$ 는 각각 채널별 음성과 잡음의 분산이며, 이때 k 번째 주파수 밴드에 대한 우도 비는 다음 수학적 식 4와 같이 구한다.

수학적 식 4

$$A_k \equiv \frac{p(Y_k|H_1)}{p(Y_k|H_0)} = \frac{1}{1 + \xi_k} \exp\left\{\frac{\gamma_k \xi_k}{1 + \xi_k}\right\}$$

여기서, $\xi_k = \lambda_{x,k}/\lambda_{d,k}$ 와 $\gamma_k = Y_k/\lambda_{d,k}$ 는 각각 a priori signal-to-noise ratio(SNR)와 a posteriori SNR이다. 음성 부재 구간에서 갱신되는 잡음 신호로부터 구한 잡음 분산 $\lambda_{d,k}$ 를 이용하여 a posteriori SNR γ_k 를 추정하며, 또한 a priori SNR ξ_k 는 decision-directed(DD) 방식을 이용하여 다음 수학적 식 5와 같이 추정한다.

수학적 식 5

$$\hat{\xi}_k(t) = \alpha \frac{|\hat{X}_k(t-1)|^2}{\lambda_{d,k}(t-1)} + (1-\alpha)P[\gamma_k(t) - 1]$$

여기서, $|\hat{X}_k(t-1)|^2$ 은 이전 프레임에서 추정된 음성 신호의 k 번째 스펙트럼 성분의 크기에 대한 추정치이며, MMSE에 기반을 두어 구한다. 또한 α 는 가중치 값이며, 연산자 $P[?]$ 은 다음 수학적 식 6과 같이 정의된다.

수학적 식 6

$$P[x] = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

기존의 일반적인 통계적 모델 기반의 음성 검출기에 대한 결정식은 각각의 주파수 채널에서 구해진 우도 비를

기하 평균하여 다음 수식식 7과 같이 음성 검출 여부를 판단한다.

수식식 7

$$\log \Lambda(t) = \frac{1}{M} \sum_{k=1}^M \log \Lambda_k \begin{matrix} H_1 \\ > \beta \\ < \\ H_0 \end{matrix}$$

[0037]

여기서, M은 전체 주파수 대역의 개수이며, η 는 음성 검출 문턱 값이다.

[0038]

본 발명에서는 위와 같이 잡음 분산 $\lambda_{d,k}$ 를 이용하여 추정된 a posteriori SNR와 DD 방식을 이용하여 추정된 a priori SNR을 기초 벡터로 할 수 있다.

[0039]

단계 S100에서는, 비음수 행렬 인수분해 기법을 이용하여 잡음 신호 및 입력 신호로부터 기초 벡터를 도출한다.

[0040]

비음수 행렬 인수분해(Non-negative Matrix Factorization, NMF)는 PCA, VQ와 마찬가지로 기초 벡터들의 선형 결합으로 근사하여 행렬을 분해하는 기법으로서, NMF는 특별히 모든 성분이 비음수인 제약을 가진다. 비음수 성분으로 구성된 $n \times m$ 행렬 V는 $V \approx WH$ 형태, 즉 다음 수식식 8과 같이 인수분해된다.

[0041]

수식식 8

$$V_{i\mu} \approx (WH)_{i\mu} = \sum_{a=1}^r W_{ia} H_{a\mu}$$

[0042]

여기서, 행렬 V는 $n \times 1$ 벡터가 m개 결합된 것으로 볼 수 있으며, 행렬 W는 $n \times r$ 크기로 역시 r개의 $n \times 1$ 벡터 집합, H는 $r \times m$ 크기의 계수 행렬이 된다. 즉, m개의 n차 데이터 벡터들을 r개의 n차 기초 벡터들의 선형 결합으로 표현된다.

[0043]

일반적으로 행렬의 비음수 행렬 인수분해 과정은 행렬 W와 H를 반복적으로 갱신하여 행렬 V와 WH 간의 거리를 최소화하는 방향으로 근사한다. 따라서 어떠한 거리 함수 혹은 목적 함수를 적용할 것인가에 따라 성분 값 갱신 식이 달라진다. 일반적으로 널리 사용되는 목적 함수로는 Euclidean distance

[0044]

$\|V - WH\|^2 = \sum_{ij} (V_{ij} - \sum_{a=1}^r W_{ia} H_{aj})^2$ 이며 이것을 최소화하는 성분 값 갱신 식은 다음 수식식 9와 같다.

수식식 9

$$H_{a\mu} \leftarrow H_{a\mu} \frac{(W^T V)_{a\mu}}{(W^T WH)_{a\mu}}$$

$$W_{ia} \leftarrow W_{ia} \frac{(VH^T)_{ia}}{(WHH^T)_{ia}}$$

[0045]

본 발명에서는 위와 같이 잡음 분산 $\lambda_{d,k}$ 를 이용하여 추정된 a posteriori SNR와 DD 방식을 이용하여 추정된 a priori SNR을 기초 벡터로 두고, 시작 부분의 m+9개 프레임을 음성 부재 구간으로 가정한 후, m개 프레임의 슈

[0046]

퍼프레임들을 비음수 행렬 인수분해 하여 배경 잡음 신호의 기초 벡터 $\mathbf{W}$ 를 추출하고 1 프레임씩 이동하여 기초 벡터를 추출할 수 있다.

[0047] 단계 S200에서는, 단계 S100에서 도출한 잡음 신호 기초 벡터와 입력 신호 기초 벡터 사이의 오차를 계산한다. 비음수 행렬 인수분해 기법은 데이터 벡터를 기초 벡터들의 선형 결합으로 근사한다. 따라서 사전에 잡음 신호로부터 기초 벡터를 추출하여 입력 신호의 기초 벡터와의 거리를 계산하여 얻은 값으로 입력 신호와 잡음 신호의 유사도를 판단할 수 있다. 따라서 잡음 신호 기초 벡터와 입력 신호 기초 벡터 사이의 오차를 계산함으로써 오차 크기에 따른 유사도를 판단하여 입력 신호에서 음성을 검출하게 된다.

[0048] 단계 S300에서는, 통계 모델 기반의 음성 검출 방법에 의해 추정된 음성 부재구간에서 비음수 행렬 인수분해 기법을 이용하여 도출된 잡음 신호 및 입력 신호의 오차 값 분포에 따른 최적화된 문턱 값을 구한다.

[0049] 본 발명에서는 다양한 잡음 환경에 최적화된 문턱 값을 적용하기 위해 통계적 모델 기반의 음성검출기를 도입한다. NMF 결과를 사용하여 잡음 환경을 인식하여 잡음 환경에 해당하는 최적 문턱 값을 적용하기 위해서는 음성 구간을 최대한 배제하는 것이 바람직하다. 잡음 구간을 효과적으로 추정하기 위해 시작 N프레임에 대하여 통계적 모델 기반의 음성검출기를 적용하여 비음성일 경우 NMF 결과 값의 평균으로 잡음 환경을 구분한다.

[0050] 도 2는 통계적 모델 기반의 음성 검출 방법을 통해 잡음 환경이라고 추정된 구간에서의 비음수 행렬 인수분해 결과의 분포를 도시한 도면이다. 도 2에 도시된 바와 같이, NMF 결과 분포는 각각의 잡음 환경에 따라 명확하게 구분할 수 있다. 따라서 각각의 잡음 환경에 대해 최적화된 문턱 값을 다음 수학식 10과 같이 적용한다.

수학식 10

$$\Xi = \frac{1}{b} \sum_{t=1}^N \left\| \overline{\mathbf{W}}(t) - \mathbf{W}(t) \right\|^2 \Big|_{\log \Lambda(t) < \beta}$$

[0051]

수학식 11

$$\eta = \begin{cases} \eta_1 & \Xi < \Xi_0 \\ \eta_2 & \Xi_0 \leq \Xi < \Xi_1 \\ \eta_3 & \Xi_1 \leq \Xi < \Xi_2 \\ \eta_4 & \Xi_2 \leq \Xi \end{cases}$$

[0052]

[0053] 여기서, N은 잡음 환경을 추정하기 위한 프레임 사이즈이고, b는 N 프레임 안에서 $\log \Lambda(t) < \beta$ 인 프레임의 개수를 나타낸다. 주목해야 할 점은 통계적 모델기반의 음성검출 알고리즘을 적용하여 잡음환경에 최적화된 문턱 값을 적용함으로써 보다 향상된 음성 검출 결과를 얻을 수 있으며, 단순히 음성만을 검출하는 것이 아니라 특정 잡음 환경을 검출하여 후처리하는 방법으로 응용할 수 있다.

[0054] 단계 S400에서는, 단계 S200에서 계산된 오차에 문턱 값을 적용하여 음성을 검출한다. 이때, 단계 S300에서 도출된 최적화된 문턱 값 η 를 이용하면 본 발명의 일실시예에 따른 음성 검출 방법의 성능이 향상될 수 있다.

[0055] 총 10 개의 기초 벡터의 평균 $\overline{\mathbf{W}}$ 와 입력 신호의 기초 벡터들 사이의 거리를 계산하여 음성 활성 여부를 다음과 같이 판단한다. 특히, 단계 S300에서 도출된 최적화된 문턱 값 η 를 다음 수학식 12에 적용하여 최종적으로 음

성을 검출할 수 있다.

수학식 12

$$\begin{array}{c} H_1 \\ \|\bar{W} - W\|^2 > \eta \\ H_0 \end{array}$$

[0056]

[0057]

추출된 기초 벡터 $\bar{W}$ 는 배경 잡음 신호로부터 추출된 것이므로 이 기초 벡터들과 유사한 신호가 입력될 경우 거리는 상대적으로 작으며, 잡음 신호와 특성이 다른 음성 신호와의 거리는 상대적으로 매우 크다.

[0058]

도 3은 백색 잡음 환경에서 잡음 대 음성 신호 비에 대한 비음수 행렬 인수분해 기반의 음성 검출 결과를 도시한 도면이다. 도 3에는 백색 잡음 5dB인 SNR일 경우의 결과를 도시하였다. 도 3은 백색 잡음 환경에서 잡음 신호의 기초벡터 $\bar{W}$ 와 입력 신호의 기초벡터 W 사이의 거리를 보여주고 있다.

[0059]

도 4는 잡음 환경에 따른 음성 검출 확률의 변화를 도시한 도면이다. 도 4에는 NMF 기반의 음성 검출기에서 문턱 값 η '의 변화에 따른 음성 검출 확률 P_e (speech detection error probability)의 변화 곡선이다. 도 2에 도시된 바와 같이, NMF 결과가 잡음환경에 따라 각각의 최적화된 문턱 값을 가짐을 알 수 있다. 따라서 최적화된 성능을 위하여 잡음환경에 따른 문턱 값이 존재하는 것이 바람직하다. 음성 부재구간으로 가정한 시작 프레임에서 추출한 잡음 환경에서의 기초 벡터 $\bar{W}$ 는 입력신호가 동일한 잡음 환경이라면 0에 가까운 값을 가져야 하지만 그렇지 않다. 이것은 잡음이 stationary 혹은 non-stationary의 정도에 따라 잡음 신호의 기초 벡터와 입력 신호의 기초 벡터 간의 차이가 다른 값을 가지기 때문이다.

[0060]

실험결과

[0061]

본 발명에서 제안된 음성 검출 알고리즘의 성능을 평가하기 위하여 오경보(false alarm)와 누락(missing)을 포함한 음성 검출 오류 확률 P_e 를 측정하였다. 실험에 사용된 데이터는 총 230초의 깨끗한 음성 데이터에 음성과 비음성 부분을 10ms마다 수동으로 표시하였다. 분류된 음성 데이터의 음성 구간은 총 57.1%로 유성음 44.0%, 무성음 13.1%로 구성되었으며 잡음 환경을 만들기 위해 NOISEX-92 데이터베이스로부터 white, factory1, destroyer ops 잡음을 ETRI 데이터베이스로부터 street 잡음을 사용하였으며 각각 0, 5, 10, 15dB SNR로 원래의 음성신호에 더하여 사용하였다. NMF에서 기초 벡터를 구하기 위해 사용한 특징벡터는 a priori SNR ξ_k , a posteriori SNR γ_k 총 32차를 사용하였으며 $m=5$, $r=3$ 으로 사용하였다. 잡음 기초 벡터는 1 프레임씩 이동하면서 총 10개의 기초 벡터의 평균으로 사용하였으며 각각의 잡음을 추정하기 위하여 50 프레임에서의 NMF 결과 분포를 사용하여 분포의 평균값으로 잡음 환경을 구분하여 문턱 값을 결정하였다.

[0062]

음성 검출 실험 결과는 다음 표 1에 나타나 있다. 실험 결과를 분석해 보면 낮은 SNR에서 NMF 기반의 음성검출기에 최적화된 문턱 값을 적용한 실험 결과 두드러진 성능향상을 보였다. 이것은 NMF 기반의 음성인식기가 통계적 기반의 음성인식기보다 신호대 잡음 비의 영향을 덜 받는 것으로 볼 수 있으며 평균적으로 6.75%의 음성 검출 성능 향상을 확인할 수 있다.

표 1

Noise	Methods	SNR			
		0 dB	5 dB	10 dB	15 dB
street	LRT	40.38	38.28	29.45	24.33
	NMF	33.06	31.42	27.04	23.68
white	LRT	80.76	35.38	30.85	27.48
	NMF	55.84	25.66	23.30	21.65
factory1	LRT	64.11	45.84	33.74	23.33
	NMF	48.81	37.96	29.87	22.47
Destroyer ops	LRT	52.88	39.37	28.25	19.28
	NMF	44.67	36.16	26.01	17.96

[0063]

[0064]

이상 설명한 본 발명은 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에 의하여 다양한 변형이나 응용이 가능하며, 본 발명에 따른 기술적 사상의 범위는 아래의 특허청구범위에 의하여 정해져야 할 것이다.

부호의 설명

[0065]

S100: 잡음 신호 및 입력 신호로부터 기초 벡터를 도출하는 단계

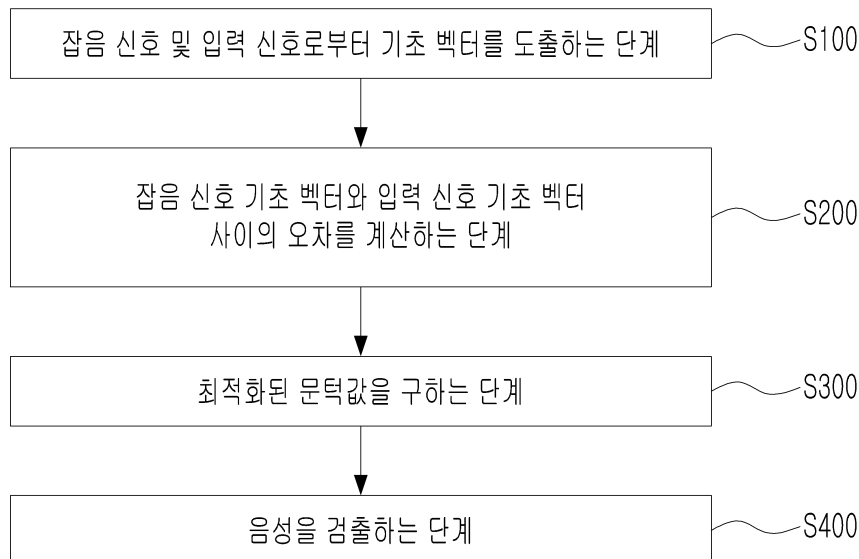
S200: 잡음 신호 기초 벡터와 입력 신호 기초 벡터 사이의 오차를 계산하는 단계

S300: 최적화된 문턱 값을 구하는 단계

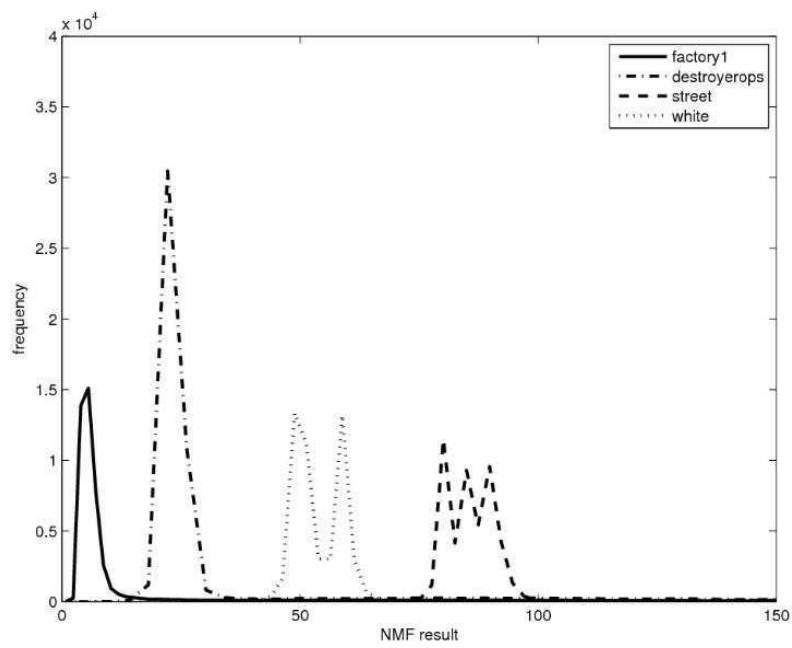
S400: 음성을 검출하는 단계

도면

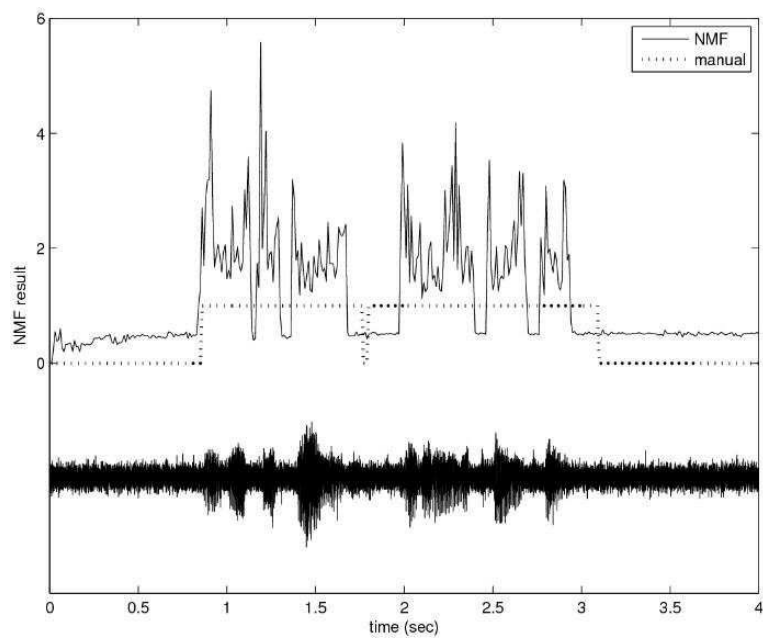
도면1



도면2



도면3



도면4

