



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2009년07월28일
(11) 등록번호 10-0909492
(24) 등록일자 2009년07월20일

(51) Int. Cl.

G10L 21/02 (2006.01)

(21) 출원번호 10-2008-0081622

(22) 출원일자 2008년08월20일

심사청구일자 2008년08월20일

(56) 선행기술조사문헌

Platt, "Probabilistic Outputs for SVMs and Comparisons to Regularized Likelihood Methods" Advances in Large Margin Classifiers, MIT Press, March 1999

조규행 외 2명, "Smoothed Global Soft Decision에 근거한 음성 향상 기법", 전자공학회 논문지, 제44권 SP편 제6호, pp.734-739, 2007년 11월

Hsuan-Tien Lin et al. "A note on Platt's probabilistic outputs for support vector machines", Machine Learning, Vol.68, No.3, pp.267-276, October 2007

US7403894 B2

전체 청구항 수 : 총 6 항

심사관 : 정성윤

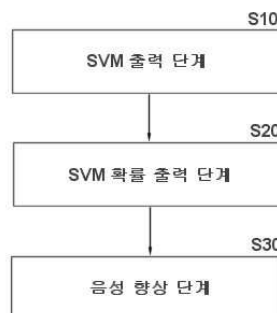
(54) 서포트 벡터 머신의 확률 출력을 이용한 음성 향상 방법

(57) 요약

본 발명은 서포트 벡터 머신(Support Vector Machine; SVM)의 확률 출력을 이용한 음성 향상 방법에 관한 것으로, 보다 구체적으로는 (1) 입력신호로부터 구해진 우도비(Likelihood Ratio; LR)를 이용하여 각 프레임별 음성 존재 유무에 대한 SVM 출력을 구하는 SVM 출력 단계; (2) 상기 SVM 출력 단계에서 구한 SVM 출력으로부터 시그모이드(sigmoid) 함수를 이용하여 각 프레임별 음성 존재 확률에 대한 SVM 확률 출력을 구하는 SVM 확률 출력 단계; 및 (3) 상기 SVM 확률 출력 단계에서 구한 각 프레임별 음성 존재 확률에 대한 SVM 확률 출력을 이용하여 음성을 향상시키는 음성 향상 단계를 포함하는 것을 그 구성상의 특징으로 한다.

본 발명에서 제안하고 있는 SVM의 확률 출력을 이용한 음성 향상 방법에 따르면, 시그모이드(sigmoid) 함수를 이용하여 얻어진 SVM의 확률 출력에 의해 추정된 새로운 전역음성 부재 확률(GSAP)을 음성 향상 방법에 적용함으로써, 다양한 잡음 환경에서 개선된 음성 향상 품질을 제공할 수 있다.

대표도 - 도1



특허청구의 범위

청구항 1

서포트 벡터 머신(Support Vector Machine; SVM)의 확률 출력을 이용한 음성 향상 방법으로서,

- (1) 입력신호로부터 구해진 우도비(Likelihood Ratio; LR)를 이용하여 각 프레임별 음성 존재 유무에 대한 SVM 출력을 구하는 SVM 출력 단계;
- (2) 상기 SVM 출력 단계에서 구한 SVM 출력으로부터 시그모이드(sigmoid) 함수를 이용하여 각 프레임별 음성 존재 확률에 대한 SVM 확률 출력을 구하는 SVM 확률 출력 단계; 및
- (3) 상기 SVM 확률 출력 단계에서 구한 각 프레임별 음성 존재 확률에 대한 SVM 확률 출력을 이용하여 음성을 향상시키는 음성 향상 단계

를 포함하는 음성 향상 방법.

청구항 2

제1항에 있어서,

상기 SVM 출력 단계는,

- a. 상기 입력신호에 대하여 각 주파수별로 우도비를 계산하는 우도비 계산 단계; 및
- b. 구해진 상기 우도비를 SVM의 특징 벡터로 이용하여 각 프레임별 음성 존재 유무에 대한 분류를 출력하는 SVM 분류 출력 단계

를 더 포함하는 것을 특징으로 하는 음성 향상 방법.

청구항 3

제1항에 있어서,

상기 SVM 출력 단계는, 입력신호를 커널 함수를 이용하여 고차원 공간으로 사상시켜 고차원 특징 공간에서의, 입력신호의 각 프레임별 음성 존재 유무에 대한 분류를 실시하는 것을 특징으로 하는 음성 향상 방법.

청구항 4

제1항에 있어서,

상기 SVM 확률 출력 단계는, 다음 수식과 같은 시그모이드 함수를 이용하는 것을 특징으로 하는 음성 향상 방법.

$$P(H_1|f(\Lambda)) = \frac{1}{1 + \exp(Af(\Lambda) + B)}$$

여기서, $P(H_1|f(\Lambda))$ 는 각 프레임별 음성 존재 확률을 나타내고, $f(\Lambda)$ 는 상기 SVM 출력 단계에서 구한 각 프레임별 음성 존재 유무에 대한 SVM 출력을 나타내며, A와 B는 학습 데이터로부터 우도 추정법(maximum likelihood estimation)에 의해 구해진 파라미터를 나타낸다.

청구항 5

제4항에 있어서,

상기 A와 상기 B를 각각 -1.97, -0.122로 설정하는 것을 특징으로 하는 음성 향상 방법.

청구항 6

제1항에 있어서,

상기 음성 향상 단계는, 상기 SVM 확률 출력 단계에서 구한 각 프레임별 음성 존재 확률에 대한 SVM 확률 출력으로부터 전역음성 부재 확률(Global Speech Absence Probability; GSAP)을 구한 후, 이를 음성 향상에 적용하

는 것을 특징으로 하는 음성 향상 방법.

명세서

발명의 상세한 설명

기술 분야

- <1> 본 발명은 서포트 벡터 머신(Support Vector Machine; SVM)의 확률 출력을 이용한 음성 향상 방법에 관한 것으로, 보다 구체적으로는 시그모이드(sigmoid) 함수를 이용하여 얻어진 SVM의 확률 출력에 의해 추정된 새로운 전역음성 부재 확률(Global Speech Absence Probability; GSAP)을 음성 향상 방법에 적용함으로써, 다양한 잡음 환경에서 개선된 음성 향상 품질을 제공할 수 있는 음성 향상 방법에 관한 것이다.

배경 기술

- <2> 음성 인식과 부호화 등의 시스템에서 배경 잡음이 존재하는 경우 두드러진 성능 저하와 관련하여 음성 향상과 관련된 많은 연구가 진행되었다. 그 중 대표적인 음성 향상 기법으로 음성의 존재와 부재에 대한 확률을 기반으로 한 소프트 디시전(Soft Decision; SD) 기법이 널리 알려져 있으며, 특히 각각의 스펙트럼 성분별로 소프트 디시전을 수행하는 기존의 방법 대신 주어진 프레임에서 전역적으로 수행하는 전역 소프트 디시전(Global Soft Decision; GSD)을 기반으로 한 음성 향상 기법이 우수한 성능을 보이는 것으로 알려져 있다.
- <3> 한편, 패턴 인식 분야에서 널리 사용되고 있는 서포트 벡터 머신(Support Vector Machine; SVM)의 출력을 확률 값으로 변환하는 효과적인 방법이 제시되었다. SVM의 확률 출력을 GSD 기반 음성 향상 기법의 주요 파라미터인 전역음성 부재 확률(Global Speech Absence Probability; GSAP)로 대체한 새로운 스펙트럼 향상 기법을 제안해 볼 필요성이 있다.

발명의 내용

해결 하고자하는 과제

- <4> 본 발명은 상기와 같은 필요성 인식에서 제안된 것으로서, 시그모이드(sigmoid) 함수를 이용하여 얻어진 SVM의 확률 출력에 의해 추정된 새로운 전역음성 부재 확률(GSAP)을 음성 향상 방법에 적용함으로써, 다양한 잡음 환경에서 개선된 음성 향상 품질을 제공할 수 있는 음성 향상 방법을 제공하는 것을 그 목적으로 한다.

과제 해결수단

- <5> 상기한 목적을 달성하기 위한 본 발명의 특징에 따른 서포트 벡터 머신(Support Vector Machine; SVM)의 확률 출력을 이용한 음성 향상 방법은,
- <6> (1) 입력신호로부터 구해진 우도비(Likelihood Ratio; LR)를 이용하여 각 프레임별 음성 존재 유무에 대한 SVM 출력을 구하는 SVM 출력 단계;
- <7> (2) 상기 SVM 출력 단계에서 구한 SVM 출력으로부터 시그모이드(sigmoid) 함수를 이용하여 각 프레임별 음성 존재 확률에 대한 SVM 확률 출력을 구하는 SVM 확률 출력 단계; 및
- <8> (3) 상기 SVM 확률 출력 단계에서 구한 각 프레임별 음성 존재 확률에 대한 SVM 확률 출력을 이용하여 음성을 향상시키는 음성 향상 단계를 포함하는 것을 그 구성상의 특징으로 한다.
- <9> 바람직하게는, 상기 SVM 출력 단계는,
- <10> a. 상기 입력신호에 대하여 각 주파수별로 우도비를 계산하는 우도비 계산 단계; 및
- <11> b. 구해진 상기 우도비를 SVM의 특징 벡터로 이용하여 각 프레임별 음성 존재 유무에 대한 분류를 출력하는 SVM 분류 출력 단계를 더 포함할 수 있다.
- <12> 바람직하게는, 상기 SVM 출력 단계는, 입력신호를 커널 함수를 이용하여 고차원 공간으로 사상시켜 고차원 특징 공간에서의 분류를 실시할 수 있다.
- <13> 바람직하게는, 상기 SVM 확률 출력 단계는, 다음 수확식과 같은 시그모이드 함수를 이용하는 것을 특징으로 한다.

$$P(H_1|f(\Lambda)) = \frac{1}{1 + \exp(Af(\Lambda) + B)}$$

<14>

<15> 여기서, $P(H_1|f(\Lambda))$ 는 각 프레임별 음성 존재 확률을 나타내고, $f(\Lambda)$ 는 상기 SVM 출력 단계에서 구한 각 프레임별 음성 존재 유무에 대한 SVM 출력을 나타내며, A와 B는 학습 데이터로부터 우도 추정법(maximum likelihood estimation)에 의해 구해진 파라미터를 나타낸다.

<16>

더욱 바람직하게는, 상기 A와 상기 B를 각각 -1.97, -0.122로 설정하는 것을 특징으로 한다.

<17>

바람직하게는, 상기 음성 향상 단계는, 상기 SVM 확률 출력 단계에서 구한 각 프레임별 음성 존재 확률에 대한 SVM 확률 출력으로부터 전역음성 부재 확률(Global Speech Absence Probability; GSAP)을 구한 후, 이를 음성 향상에 적용하는 것을 특징으로 한다.

효과

<18>

본 발명에서 제안하고 있는 SVM의 확률 출력을 이용한 음성 향상 방법에 따르면, 시그모이드(sigmoid) 함수를 이용하여 얻어진 SVM의 확률 출력에 의해 추정된 새로운 전역음성 부재 확률(GSAP)을 음성 향상 방법에 적용함으로써, 다양한 잡음 환경에서 개선된 음성 향상 품질을 제공할 수 있다.

발명의 실시를 위한 구체적인 내용

<19>

이하에서는 첨부된 도면들을 참조하여, 본 발명에 따른 실시예에 대하여 상세하게 설명하기로 한다.

<20>

도 1은 본 발명의 일 실시예에 따른 서포트 벡터 머신(SVM)의 확률 출력을 이용한 음성 향상 방법의 구성을 나타내는 도면이다. 도 1에 도시된 바와 같이, 본 발명의 일 실시예에 따른 서포트 벡터 머신(SVM)의 확률 출력을 이용한 음성 향상 방법은, SVM 출력 단계(S10), SVM 확률 출력 단계(S20), 및 음성 향상 단계(S30)를 포함한다.

<21>

SVM 출력 단계(S10)는, 입력신호로부터 구해진 우도비(LR)를 이용하여 각 프레임별 음성 존재 유무에 대한 SVM 출력을 구하는 역할을 한다. SVM 출력 단계(S10)는, 입력신호에 대하여 각 주파수별로 우도비를 계산하는 우도비 계산 단계, 및 구해진 우도비를 SVM의 특징 벡터로 이용하여 각 프레임별 음성 존재 유무에 대한 분류를 출력하는 SVM 분류 출력 단계로 구성될 수 있다. 또한, SVM 출력 단계(S10)는, 입력신호를 커널 함수를 이용하여 고차원 공간으로 사상시켜 고차원 특징 공간에서의 분류를 실시할 수 있다.

<22>

SVM 확률 출력 단계(S20)는, SVM 출력 단계(S10)에서 구한 SVM 출력으로부터 시그모이드(sigmoid) 함수를 이용하여 각 프레임별 음성 존재 확률에 대한 SVM 확률 출력을 구하는 역할을 한다. SVM 확률 출력 단계(S20)에서 이용되는 시그모이드 함수로서는, 다음 수학식 1과 같은 시그모이드 함수가 사용될 수 있다.

수학식 1

$$P(H_1|f(\Lambda)) = \frac{1}{1 + \exp(Af(\Lambda) + B)}$$

<23>

<24>

상기 수학식 1에서, $P(H_1|f(\Lambda))$ 는 각 프레임별 음성 존재 확률을 나타내고, $f(\Lambda)$ 는 상기 SVM 출력 단계에서 구한 각 프레임별 음성 존재 유무에 대한 SVM 출력을 나타내며, A와 B는 학습 데이터로부터 우도 추정법에 의해 구해진 파라미터를 나타낸다. 바람직하게는, A와 B를 각각 -1.97, -0.122로 설정할 수 있다.

<25>

음성 향상 단계(S30)는, SVM 확률 출력 단계(S20)에서 구한 각 프레임별 음성 존재 확률에 대한 SVM 확률 출력을 이용하여 음성을 향상시키는 역할을 한다. 음성 향상 단계(S30)는, SVM 확률 출력 단계(S20)에서 구한 각 프레임별 음성 존재 확률에 대한 SVM 확률 출력으로부터 전역음성 부재 확률(Global Speech Absence Probability; GSAP)을 구한 후, 이를 음성 향상에 적용하도록 구성될 수 있다.

<26>

이하에서는, 먼저 GSD 및 SVM에 대하여 간단히 살펴본 후, 본 발명에서 제안하고 있는 SVM 확률 출력을 이용한 음성 향상 방법에 대하여 살펴보기로 한다.

<27>

1. GSD 기반 음성 향상 방법

<28>

먼저 시간축 상에서 원래의 음성신호 x에 잡음신호 d가 인가되어 입력신호 y를 만든다고 가정한다. 이를 이산

푸리에 변화(Discrete Fourier Transform; DFT)을 통해 주파수 축으로 변환하면 다음 수학식 2와 같이 표현할 수 있다.

수학식 2

<29>
$$Y(t) = X(t) + D(t)$$

<30> 여기서, $\mathbf{Y}(t) = [Y_1(t), Y_2(t), \dots, Y_N(t)]^T$, $\mathbf{X}(t) = [X_1(t), X_2(t), \dots, X_N(t)]^T$, 및 $\mathbf{D}(t) = [D_1(t), D_2(t), \dots, D_N(t)]^T$ 는 각각 t번째 프레임에서의 DFT 계수 벡터이다. H_0 , H_1 이 각각 음성의 부재와 존재를 나타낸 가설이라고 하면, 주어진 가정으로부터 다음 수학식 3과 같이 기술될 수 있다.

수학식 3

<31> H_0 : 음성 부재: $Y(t) = D(t)$

<32> H_1 : 음성 존재: $Y(t) = X(t) + D(t)$

<33> 음성신호와 잡음신호의 스펙트럼이 복소 가우시안 분포를 따른다고 가정하면, 가설 H_0 과 H_1 을 조건으로 한 확률 밀도함수는 다음 수학식 4 및 수학식 5와 같이 주어질 수 있다.

수학식 4

<34>
$$p(Y_k(t)|H_0) = \frac{1}{\pi \lambda_{d,k}(t)} \exp \left\{ -\frac{|Y_k(t)|^2}{\lambda_{d,k}(t)} \right\}$$

수학식 5

<35>
$$p(Y_k(t)|H_1) = \frac{1}{\pi [\lambda_{d,k}(t) + \lambda_{x,k}(t)]} \exp \left\{ -\frac{|Y_k(t)|^2}{\lambda_{d,k}(t) + \lambda_{x,k}(t)} \right\}$$

<36> 여기서, $\lambda_{x,k}(t)$ 와 $\lambda_{d,k}(t)$ 는 각각 음성신호와 잡음신호의 분산을 나타낸다. 각각의 주파수 채널별 성분이 통계적으로 독립이라고 가정하면, 입력신호 y의 t번째 프레임에서의 GSAP는 다음 수학식 6과 같이 표현될 수 있다.

수학식 6

<37>
$$p(H_0|Y(t)) = \frac{p(Y(t)|H_0)p(H_0)}{p(Y(t)|H_0)p(H_0) + p(Y(t)|H_1)p(H_1)}$$

$$= \frac{1}{1 + \frac{p(H_1)}{p(H_0)} \prod_{k=1}^N \Lambda(Y_k(t))}$$

<38> 여기서, $p(H_0) (= 1 - p(H_1))$ 는 음성 부재에 대한 사전 확률이며, $\Lambda(Y_k(t))$ 는 각 주파수 채널별 우도비(LR)를 나타내며 다음 수학식 7과 같다.

수학식 7

<39>
$$\Lambda(Y_k(t)) = \frac{p(Y_k(t)|H_1)}{p(Y_k(t)|H_0)}$$

$$= \frac{1}{1 + \xi_k(t)} \exp \left[\frac{\gamma_k(t) \xi_k(t)}{1 + \xi_k(t)} \right]$$

<40> 여기서, $\gamma_k(t) = |Y_k(t)|^2 / \lambda_{d,k}(t)$, $\xi_k(t) = \lambda_{x,k}(t) / \lambda_{d,k}(t)$ 는 각각 사후 신호 대 잡음비(a posteriori signal-to-noise ratio(SNR))와 사전 신호 대 잡음비(a priori SNR)이다. 다음 수학식 8 및 9와 같이, 음성 부재 구간에서 갱신되는 잡음 신호로부터 구한 잡음 분산 $\hat{\lambda}_{d,k}$ 을 이용하여 사후 SNR을 $\hat{\gamma}_k$ 로써 추정하며, 또한 사전 SNR을

decision-directed(DD) 방식을 이용하여 $\hat{\xi}_k$ 로서 추정한다.

수학식 8

$$\hat{\gamma}_k(t) = \frac{|Y_k(t)|^2}{\hat{\lambda}_{d,k}(t)}$$

수학식 9

$$\hat{\xi}_k(n) = \alpha \frac{|\hat{X}_k(t-1)|^2}{\hat{\lambda}_{d,k}(t-1)} + (1-\alpha)K[\hat{\gamma}_k(t)-1]$$

여기서, $|\hat{X}_k(t-1)|$ 은 이전 프레임에서 추정된 음성 신호의 k번째 스펙트럼 성분의 크기에 대한 추정치이며, 최소자승 평균오차(minimum mean-square error, MMSE)에 기초하여 구한다. 또한 α 는 가중치 값이며, 연산자 $K[|]$ 은 다음 수학식 10과 같다.

수학식 10

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

$\hat{\mathbf{X}}(t) = [\hat{X}_1(t), \hat{X}_2(t), \dots, \hat{X}_N(t)]^T$ 을 n번째 프레임에서의 깨끗한 음성의 추정치라고 할 때, 기존의 스펙트럼 향상 기법은 오염된 음성 신호의 각각의 주파수 성분 $Y_k(t)$ 에 이득 함수 $G()$ 를 적용하여 $\hat{X}_k(t)$ 을 추정한다. 이득 함수를 구하는 다양한 방법 중 본 논문에서는 음악 잡음(musical noise)에 우수한 성능을 보이는 Epharim-Malah의 잡음 압축 규칙(noise suppression rule)을 적용하며, 추정치 $\hat{\mathbf{X}}(t)$ 는 다음 수학식 11과 같이 표현된다.

수학식 11

$$\hat{X}_k(t) = G(\eta_k(t), \gamma_k(t)) Y_k(t)$$

여기서, 이득 함수 $G()$ 는 다음 수학식 12와 같이 주어진다.

수학식 12

$$G(\eta, \gamma) = \frac{\sqrt{\pi}}{2} \sqrt{\frac{\eta}{\gamma(1+\eta)}} \exp\left(-\frac{\gamma\eta}{2(1+\eta)}\right) \times \left[\left(1 + \frac{\gamma\eta}{1+\eta}\right) I_0\left(\frac{\gamma\eta}{2(1+\eta)}\right) + \left(\frac{\gamma\eta}{1+\eta}\right) I_1\left(\frac{\gamma\eta}{2(1+\eta)}\right) \right]$$

여기서, I_0 와 I_1 은 각각 수정된 베셀 함수의 0차 1차 버전이다. 또한, 소프트 디시전 기법을 바탕으로 음성의 존재와 부재에 대한 확률 정보를 이용하여 다음과 같이 수정된 이득 함수 $\tilde{G}$ 을 적용하여 $\hat{X}_k(t)$ 을 다음 수학식 13과 같이 추정한다.

수학식 13

$$\hat{X}_k(t) = \tilde{G}(\eta_k(t), \gamma_k(t)) Y_k(t)$$

여기서, $\tilde{G}$ 는 다음 수학식 14와 같다.

수학식 14

$$\tilde{G}(\cdot) = (1 - p(H_0|\mathbf{Y}(t))) G(\cdot)$$

2. 서포트 벡터 머신(Support Vector Machine; SVM)

Vapnik이 제안한 SVM은 통계적 학습 이론에 기반을 둔 패턴 분류기이다. 기존의 학습 알고리즘은 학습 집단을 이용하여 학습 데이터의 수행도를 최적화하기 위한 경험적 위험 최소화(Empirical Risk Minimization; ERM) 방

법을 기초로 하고 있는 반면에, SVM은 분류 오류 확률을 최소화하는 구조적인 위험 최소화(Structural Risk Minimization) 방법에 기초하고 있다.

<55> 2개의 클래스에 대한 학습 데이터가 $(x_1, z_1), \dots, (x_l, z_l), x \in R^n, z_i \in \{+1, -1\}$ 과 같이 주어졌을 때, 초평면에 대한 방정식은 $(wx)+b=0$ 과 같이 초평면에 직교하는 가중치 벡터 w 와 바이어스 b 로 표현된다. 이 때 SVM은 학습 데이터로부터 최적 초평면을 위한 서포트 벡터(support vector) x_i^* 를 구하며, 이를 바탕으로 임의의 입력 데이터 x 가 주어졌을 때 SVM의 출력 값은 다음 수학적 식 15와 같이 표현된다.

수학적 식 15

$$f(x) = (w^* \cdot x) + b^* \\ = \sum_{i=1}^M \alpha_i^* z_i^* (x_i^* \cdot x) + b^*$$

<56>

<57> 여기서, w^*, b^*, α_i^* 는 각각 학습 데이터로부터 구한 최적 가중치 벡터, 최적 바이어스, 라그랑지 승수이며, M 은 서포트 벡터의 총 개수이다.

<58> 일반적으로 입력 데이터의 경우 명확하게 선형 분류가 되지 않는 경우가 대부분이며, 이를 위해 커널 함수 K 를 이용하여 고차원 공간으로 사상시켜 고차원 특징 공간에서의 분류를 다음 수학적 식 16과 같이 실시한다.

수학적 식 16

$$f(x) = (w^* \cdot x) + b^* \\ = \sum_{i=1}^M \alpha_i^* z_i^* K(x_i^* \cdot x) + b^*$$

<59>

<60> 3. SVM 확률 출력을 이용한 음성 향상 기법

<61> SVM의 출력 값은 상기 수학적 식 15 및 16에서 보듯이 입력 데이터와 서포트 벡터들과의 내적의 합으로 일종의 거리를 나타내며, 출력 값의 크기에 따라 입력 데이터와 클래스간의 상관성을 보여준다. SVM으로부터 확률 값을 직접 구할 수는 없으므로, SVM 출력으로부터 확률 값을 구하는 여러 방법이 제시되었다. 이를 바탕으로 음성 신호의 각 프레임에서 음성의 존재 유무에 대한 SVM 확률 출력을 도출하면, 각 프레임에서 음성의 존재에 대한 확률을 얻는다.

<62> 본 발명에서는 오염된 임의의 입력신호로부터 깨끗한 음성신호를 추정하기 위하여 GSD 기법을 적용하며, 이 때 상기 수학적 식 14의 GSAP를 SVM 출력 값의 확률로써 대체한다. 구체적으로, 상기 수학적 식 7의 각 주파수별 우도비(LR)을 SVM의 특징 벡터로 이용하여 음성의 존재 유무에 대한 분류를 실시하며, 이를 Platt이 제안한 다음 수학적 식 17과 같은 시그모이드(sigmoid) 함수에 적용하여 SVM 확률 출력을 구한다.

수학적 식 17

$$P(H_1|f(\Lambda)) = \frac{1}{1 + \exp(Af(\Lambda) + B)}$$

<63>

<64> 여기서, A 와 B 는 각각 시그모이드(sigmoid) 함수의 파라미터이며, $f(\Lambda)$ 는 입력된 입력신호로부터 구해진 우도비(LR)에 대한 SVM 출력이다. 이때 A 와 B 는 학습 데이터로부터 우도 추정법(maximum likelihood estimation)에 의해 구해진다. 먼저 학습 데이터에 대한 i 번째 출력 $f_i(\Lambda)$ ($i=1, \dots, l$)의 목표 확률 값 t_i 를 학습 데이터의 클래스 라벨 z_i 를 이용하여 다음 수학적 식 18과 같이 정의한다.

수학적 식 18

$$t_i = \frac{z_i + 1}{2}$$

<65>

<66> 이때 두 파라미터에 의해 최적의 확률 값을 갖도록 최소화 문제를 최적화하는 다양한 알고리즘들 중 본 발명에서는 model-trust minimization 알고리즘을 이용하여 구한다. 구체적으로 상기 수학적 식 17의 파라미터 A, B 는

학습 데이터를 이용하여 로그 우도(log likelihood) 최소화

$\min \left[- \sum_i \{ t_i \log P(H_1 | f_i(\Lambda)) + (1 - t_i) \log P(H_0 | f_i(\Lambda)) \} \right]$ 에 의해 최적화된 값을 찾는다.

<67> 최종적으로 상기 수학적 식 17로부터 도출된 각 프레임별 음성 존재 확률 $P(H_1 | f(\Lambda))$ 를 이용하여 구한 깨끗한 음성의 추정치는 다음 수학적 식 19와 같다.

수학적 식 19

<68> $\hat{X}_k(t) = \tilde{G}^*(\eta_k(t), \gamma_k(t)) Y_k(t)$

<69> 여기서, $\tilde{G}^*(\cdot) = P(H_1 | f(\Lambda)) G(\cdot)$

<70> 도 2는 F16 잡음 SNR 5 dB에서의 기존의 GSAP(SEGSD)와 본 발명에서 제안된 SVM 확률 출력으로부터 구한 GSAP를 비교하여 나타내는 도면이다. 도 2로부터, 본 발명에서 제안한 방법에 의한 GSAP 값이 음성의 존재 여부에 대해 더욱 견실한 추정을 나타냄을 확인할 수 있다.

<71> 4. 실험 결과 비교 및 분석

<72> 먼저, SVM 확률 출력을 구하기 위해 음성 데이터로부터 학습을 실시한다. 사용된 음성 데이터는 각각 4명의 남성, 여성 화자가 발음한 깨끗한 음성을 8 kHz로 샘플링하였으며, 총 226초의 음성 데이터를 음성과 비음성 부분으로 10 ms마다 수동으로 표시하여 클래스를 구성하였다. 잡음 환경을 고려하여 vehicular, street, white 잡음을 각각 5, 10, 20 dB의 SNR로서 추가하였다. SVM의 특징 벡터 차수는 16개의 주파수 채널에서 구한 우도비(LR)로 총 16차로 구성하였다. 또한 음성 데이터의 비선형 특성을 고려하여 상기 수학적 식 16의 커널 함수 K는 radial basis function(RBF) 커널을 적용하여 학습을 실시하였다. 최종적으로 본 실험에서는 학습을 통하여 총 126,965개의 서포트 벡터가 생성되었다.

<73> 상기 수학적 식 17의 SVM 확률 출력을 위한 시그모이드 함수의 두 파라미터 A, B는 학습에 사용된 데이터로부터 구한 SVM 출력 값과 클래스를 이용하여 구하였으며, 각각의 파라미터 값은 A=-1.97, B=-0.122이다.

<74> 본 발명에서 제안된 음성 향상 방법의 음질 평가를 위해 PESQ와 MOS 테스트를 수행하였다. 이때, 실험에 사용된 음성 데이터가 학습 데이터와 같을 경우 SVM 분류에 대해 종속적인 성능을 보일 수 있으므로, 이를 피하기 위해 다른 화자와 잡음 환경으로 구성하였다.

<75> 첫 번째 실험으로서, PESQ 테스트를 위해 각각 3명의 남성, 여성 화자가 총 60개의 문장을 발음한 깨끗한 음성을 8 kHz로 샘플링하였으며, 세 가지 형태의 잡음을 추가하였다. 추가된 잡음은 NOISEX-92 데이터베이스의 vehicular, white, F-16 잡음이며, SNR은 각각 5, 10, 15 dB로 달리하였다. 표 1은 PESQ 테스트 결과를 본 발명에서 제안한 방법과 기존의 방법(SEGSD)을 비교하여 나타낸 것이다. 표 1로부터, PESQ 테스트 결과 다양한 잡음 환경에서 전체적으로 본 발명에서 제안된 음성 향상 방법이 우수한 성능을 보여주었으며, 특히 낮은 SNR에서 향상 폭이 큰 것을 확인할 수 있다.

표 1

<76>

잡음 유형	방법	SNR (dB)		
		5	10	15
White	SEGSD	1.918	2.276	2.605
	제안된 방법	1.945	2.293	2.621
Vehicle	SEGSD	3.323	3.594	3.845
	제안된 방법	3.376	3.640	3.884
F-16	SEGSD	2.289	2.640	2.951
	제안된 방법	2.352	2.679	2.970

<77> 두 번째 실험으로서, MOS 테스트를 수행하였다. MOS 테스트에 사용된 음성 데이터는 각각 3명의 남성, 여성 화자가 발음한 총 18개의 문장의 깨끗한 음성을 8 kHz로 샘플링하였으며, PESQ 테스트와 동일하게 잡음 환경을 구성하였다. 잡음 제거 알고리즘에 적용하여 잡음이 제거된 음성 데이터를 대상으로 10명의 청취자에 의해 MOS를 결정하였다. 표 2는 MOS 테스트 결과를 본 발명에서 제안한 방법과 기존의 방법(SEGSD)을 비교하여 나타내고

있으며, 또한 가설 검증 결과도 함께 나타내고 있다. 표 2로부터, 대부분의 잡음 환경에 대하여 본 발명에서 제안된 SVM 확률 출력을 이용한 GSD 음성 향상 방법이 기존의 GSD 음성 향상 방법보다 우수한 성능을 나타냄을 확인할 수 있는데, 이는 SVM의 우수한 분류 성능에 의해 도출한 확률 값이 GSAP로써 더욱 적합하기 때문인 것으로 판단된다.

표 2

잡음 유형	방법	SNR (dB)		
		5	10	15
White	SEGSD	1.74 (± 0.32)	2.38 (± 0.36)	3.02 (± 0.36)
	제안된 방법	1.77 (± 0.31)	2.47 (± 0.34)	3.03 (± 0.35)
	가설 테스트	NW	NW	NW
Vehicle	SEGSD	3.47 (± 0.21)	3.69 (± 0.20)	4.14 (± 0.19)
	제안된 방법	3.63 (± 0.23)	3.74 (± 0.26)	4.23 (± 0.18)
	가설 테스트	B	NW	B
F-16	SEGSD	2.54 (± 0.33)	2.96 (± 0.31)	3.68 (± 0.37)
	제안된 방법	2.55 (± 0.27)	2.98 (± 0.30)	3.71 (± 0.42)
	가설 테스트	NW	NW	NW

5. 결론

본 발명에서는 GSAP의 추정을 위해 SVM 출력을 확률 값으로 변환하는 새로운 GSAP 추정 알고리즘을 제안하였다. 효과적인 GSAP의 추정을 위해 입력 신호의 각 프레임에서 구한 우도비(LR)를 SVM의 특징 벡터로 사용하여, 시그모이드 함수로부터 확률 값을 도출한다. 최종적으로 SVM의 확률 출력을 기존의 GSAP와 대체하여 구성된 제안된 음성 향상 방법의 성능을 평가하기 위해 PESQ 및 MOS 평가를 하였으며, 다양한 잡음 환경에서 본 발명에서 제안된 GSAP 추정 방법이 기존의 방법보다 향상된 결과를 나타내었다.

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.

도면의 간단한 설명

도 1은 본 발명의 일 실시예에 따른 서포트 벡터 머신(SVM)의 확률 출력을 이용한 음성 향상 방법의 구성을 나타내는 도면.

도 2는 F16 잡음 SNR 5 dB에서의 기존의 GSAP(SEGSD)와 본 발명에서 제안된 SVM 확률 출력으로부터 구한 GSAP를 비교하여 나타내는 도면.

<도면 중 주요 부분에 대한 부호의 설명>

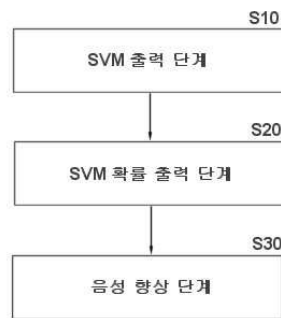
S10: SVM 출력 단계

S20: SVM 확률 출력 단계

S30: 음성 향상 단계

도면

도면1



도면2

