



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2008-0077720
(43) 공개일자 2008년08월26일

(51) Int. Cl.

G10L 11/02 (2006.01)

(21) 출원번호 10-2007-0017243

(22) 출원일자 2007년02월21일

심사청구일자 2007년02월21일

(71) 출원인

인하대학교 산학협력단

인천 남구 용현동 253 인하대학교

(72) 발명자

장준혁

서울 서초구 서초동 1660-31 B동 301호

조규행

서울 금천구 독산4동 1021-3

(뒷면에 계속)

(74) 대리인

김진우

전체 청구항 수 : 총 2 항

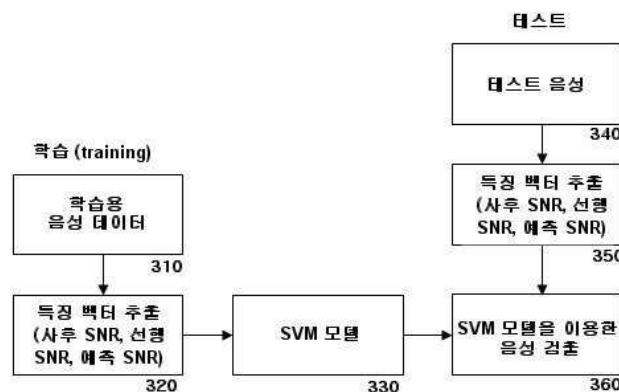
(54) 사후 신호대 잡음비, 선행 신호대 잡음비 및 예측 신호대잡음비를 특징 벡터로 이용하는 서포트 벡터머신에 기초한 음성 검출 방법

(57) 요약

본 발명은 사후 SNR(a posteriori SNR), 선행 SNR(a priori SNR) 및 예측 SNR(predicted SNR)을 결합하여 이를 특징 벡터로서 이용하는 서포트 벡터머신(Support Vector Machine; SVM)에 기초한 음성 검출 방법에 관한 것으로서, 보다 구체적으로는 (1) 학습용 음성 데이터로부터 사후 SNR, 선행 SNR 및 예측 SNR을 추출한 후, 이들을 결합하여 훈련용 특징 벡터를 생성하는 단계와, (2) 생성된 상기 훈련용 특징 벡터를 이용하여 SVM 모델을 생성하는 단계와, (3) 테스트하고자 하는 음성 데이터로부터 사후 SNR, 선행 SNR 및 예측 SNR을 추출한 후, 이들을 결합하여 테스트 특징 벡터를 생성하는 단계와, (4) 추출한 상기 테스트 특징 벡터를 상기 SVM 모델 생성 단계에서 생성한 상기 SVM 모델에 대입하여 음성을 검출하는 단계를 포함하는 SVM에 기초한 음성 검출 방법에 관한 것이다.

본 발명의 SVM에 기초한 음성 검출 방법에 따르면, 사후 SNR, 선행 SNR 및 예측 SNR을 결합하여 이들을 SVM에 대한 특징 벡터로서 사용함으로써, 실시간으로 음성 신호를 처리해야 하는 이동 통신 환경에 적합하게 하면서도 음성 검출 성능을 대폭 향상시킬 수 있다.

대표도 - 도3



(72) 발명자

박윤식

서울 은평구 불광1동 268-5 401호

이계환

경기 광명시 광명4동 한진아파트 110동902호

특허청구의 범위

청구항 1

사후 신호대 잡음비(a posteriori SNR), 선행 신호대 잡음비(a priori SNR) 및 예측 신호대 잡음비(predicted SNR)를 특징 벡터로서 이용하는 서포트 벡터머신(Support Vector Machine; SVM)에 기초한 음성 검출 방법으로서,

학습용 음성 데이터로부터 사후 SNR, 선행 SNR 및 예측 SNR을 추출한 후, 이들을 결합하여 훈련용 특징 벡터 $X = \{x_1, x_2, \dots, x_N\}$ 을 생성하는 단계;

생성된 상기 훈련용 특징 벡터를 다음 수학식 1에 적용하여 최적의 가중벡터 w_* 와 최적의 바이어스 b_* 을 구하여 SVM 모델을 생성하는 단계;

<수학식 1>

$$f(x) = \text{sign}(w_*^T K(x_i, x) + b_*)$$

(여기서, T는 전치행렬을 의미하며, 함수 $K(x, x_i)$ 는 다음 수학식 2와 같이 정의되는 커널 함수임.

<수학식 2>

$$K(x, x_i) = \Phi(x) \cdot \Phi(x_i)$$

여기서, Φ 는 고차원 공간으로 사상시켜주는 사상함수임.)

테스트하고자 하는 음성 데이터로부터 사후 SNR, 선행 SNR 및 예측 SNR을 추출한 후, 이들을 결합하여 테스트 특징 벡터를 생성하는 단계;

추출한 상기 테스트 특징 벡터를 상기 SVM 모델 생성 단계에서 생성한 상기 SVM 모델에 대입하여 음성을 검출하는 단계

를 포함하는 방법.

청구항 2

제1항에 있어서,

상기 커널 함수로서 다음 수학식 3과 같이 정의되는 RBF(Radius Basis Function)를 사용하는 방법.

<수학식 3>

$$K(x, x_i) = \exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right)$$

(여기서, σ 는 상기 특징 벡터 집합의 표준편차임.)

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

<20> 본 발명은 서포트 벡터머신(Support Vector Machine; SVM)에 기초한 음성 검출 방법에 관한 것으로서, 특히 사후 SNR(a posteriori SNR), 선행 SNR(a priori SNR) 및 예측 SNR(predicted SNR)을 결합하여 이를 특징 벡터로서 이용하는 SVM에 기초한 음성 검출 방법에 관한 것이다.

<21> 음성 검출기(voice activity detection; VAD)는 가변 전송률 음성 부호화가 필요한 다중 접속 기술의 한정된 주

과수 대역을 효율적으로 사용하기 위한 핵심적인 부분이다. 구체적으로 우수한 가변 전송률 음성 부호화기의 실현을 위해서는 음성/비음성 구간을 검출하는 음성 검출 성능이 주요한 요소로 작용한다. 이와 같은 이유로 다양한 형태의 음성 검출 알고리즘이 제안되었으며, 대표적인 것으로는 선형예측 부호화 파라미터, 에너지 레벨, 포먼트(formant) 모양, 영교차율, cepstral 특징(cepstral features), 음성신호의 적응 모델링, 주기적인 척도 등이 있다. 최근에는 패턴 인식에 근거한 음성 검출 방법과 선형 예측 부호화 잔류 신호의 고차 큐물런트(cumulants)를 이용하는 음성 검출 방법이 새로운 방법으로 제안되었다. 특히, 에너지 차이, 영교차율, 스펙트럼 차이들은 ITU-T G.729 Annex. B에 채택되었으며, 비슷한 방식들이 3GPP2의 Selectable Mode Vocoder(SMV)와 ETSI AMR VAD option2에 적용되었다.

<22> 한편, Enqing은 ITU-T G.729 Annex B 음성 검출기의 파라미터를 Support Vector Machine(SVM)에 적용하여 SVM 기반의 음성 검출 방법의 우수성을 최초로 제시하였다. 여기서, SVM이란 통계적 학습 이론으로서 분류 문제에 있어서 일반화 능력이 우수하여 패턴 인식의 여러 분야에서 응용되고 있는 기법으로서, 최근에는 음성 검출 방법으로도 여러 연구에서 이용되고 있다. 구체적으로 Ramirez는 자동차 잡음 환경에서의 음성을 대상으로 잡음을 제거한 후, subband SNR(Signal-to-Noise Ratio; 신호 대 잡음비) 추정치와 long-term SNR 추정치를 특징 벡터로서 사용하여 SVM 기법을 음성 검출 방법에 적용함으로써, SVM 기반의 음성 검출 방법의 우수한 성능을 보여주었다. 그러나 Ramirez의 방법에서 사용한 파라미터 중 long-term SNR의 경우 subband SNR보다 우수한 성능을 보여주지만, 이동 통신 환경에서는 실시간으로 음성 신호를 처리해야 하기 때문에 현실적으로 적용하기가 어렵다는 문제점이 있다.

<23> 한편, 음성 통계 모델에 기초한 기존의 음성 검출 방법에 있어서 사용되는 3가지 주요 파라미터는, 사후 SNR(*a posteriori* SNR), 선행 SNR(*a priori* SNR) 및 예측 SNR(predicted SNR)이다. 사후 SNR은 음성 부재 구간에서 측정된 잡음의 전력(power)과 입력 신호의 전력으로부터 직접 도출할 수 있으며, 선행 SNR은 이전 프레임에서 추정된 SNR 추정치와 현재 프레임의 SNR 추정치를 이용하여 스무딩된 SNR을 도출하는 DD(decision-directed) 추정법을 적용하여 계산할 수 있다. 또한, 예측 SNR은 음성의 통계 모델에 기초하여 도출한 음성 부재 확률을 이용하여 이전 프레임에서 현재 프레임을 추정하도록 한 SNR 파라미터로서, 음성 검출에 있어서 매우 효과적인 것으로 밝혀지고 있다.

<24> 따라서 실시간으로 음성 신호를 처리해야 하는 이동 통신 환경에 적합하면서도 음성 검출 성능이 우수한 SVM에 기초한 음성 검출 방법을 도출하기 위해서, SVM을 이용한 음성 검출 방법에 상기 3가지 파라미터를 접목해 볼 필요성이 있다.

발명이 이루고자 하는 기술적 과제

<25> 본 발명은, 상기와 같은 문제점 및 필요성에 대한 인식에서 비롯된 것으로서, 사후 SNR, 선행 SNR 및 예측 SNR을 결합하여 이들을 SVM에 대한 특징 벡터로 사용함으로써, 실시간으로 음성 신호를 처리해야 하는 이동 통신 환경에 적합하면서도 음성 검출 성능이 우수한 SVM에 기초한 음성 검출 방법을 제안하는 것을 그 목적으로 한다.

발명의 구성 및 작용

<26> 상기한 목적을 달성하기 위한 본 발명의 특징에 따른, SVM에 기초한 음성 검출 방법은,

<27> (1) 학습용 음성 데이터로부터 사후 SNR(*a posteriori* SNR), 선행 SNR(*a priori* SNR) 및 예측 SNR(predicted SNR)을 추출한 후, 이들을 결합하여 훈련용 특징 벡터 $X = \{x_1, x_2, \dots, x_N\}$ 을 생성하는 단계;

<28> (2) 생성된 상기 훈련용 특징 벡터를 다음 수학적 식 1에 적용하여 최적의 가중벡터 w_* 와 최적의 바이어스 b_* 을 구하여 SVM 모델을 생성하는 단계;

수학적 식 1

<29>
$$f(x) = \text{sign}(w_*^T K(x, x) + b_*)$$

<30> 여기서, T는 전치행렬을 의미하며, 함수 $K(x, x_i)$ 는 다음 수학적 식 2와 같이 정의되는 커널 함수이다.

수학식 2

<31>

$$K(x, x_i) = \Phi(x) \cdot \Phi(x_i)$$

<32>

여기서, Φ 는 고차원 공간으로 사상시켜주는 사상함수이다.

<33>

(3) 테스트하고자 하는 음성 데이터로부터 사후 SNR, 선행 SNR 및 예측 SNR을 추출한 후, 이들을 결합하여 테스트 특징 벡터를 생성하는 단계;

<34>

(4) 추출한 상기 테스트 특징 벡터를 상기 SVM 모델 생성 단계에서 생성한 상기 SVM 모델에 대입하여 음성을 검출하는 단계를 포함하는 것을 그 특징으로 한다.

<35>

본 발명의 다른 특징에 따른 SVM에 기초한 음성 검출 방법은,

<36>

커널 함수로서 다음 수학식 3과 같이 정의되는 RBF(Radius Basis Function)를 사용하는 것을 그 특징으로 한다.

수학식 3

<37>

$$K(x, x_i) = \exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right)$$

<38>

여기서, σ 는 상기 특징 벡터 집합의 표준편차이다.

<39>

이하에서는 본 발명의 실시예를 도면을 참조하여 상세히 설명한다.

<40>

본 발명의 상세한 설명에서는, 먼저 SVM 기법에 대하여 간단히 설명한 후, 다음으로 본 발명에서 SVM에 대한 특징 벡터로서 사용하고자 하는 사후 SNR, 선행 SNR 및 예측 SNR을 추출하는 과정을 살펴본 뒤, 본 발명에 따른 SVM에 기초한 음성 검출 방법을 설명한다. 마지막으로, 본 발명의 음성 검출 방법에 따른 음성 검출 결과를 기존의 음성 검출 방법에 따른 음성 검출 결과와 비교하도록 한다.

<41>

1. SVM(Support Vector Machines)의 이해

<42>

Vapnik이 제안한 SVM은 통계적 학습 이론에 기반을 둔 패턴 분류 기법이다. 기존의 학습 알고리즘은 학습 집단을 이용하여 학습 데이터의 수행도를 최적화하기 위한 경험적 위험 최소화(Empirical Risk Minimization; ERM) 방법을 기초로 하고 있는 반면, SVM은 분류 오류 확률을 최소화하는 구조적인 위험 최소화(Structural Risk Minimization; SRM) 방법에 기초하고 있다.

<43>

도 1은 선형 분리(linear separation)를 이용한 SVM을 나타내는 도면이다. 도 1에 도시된 바와 같이, 선형 SVM에 있어서 2개의 클래스를 구분할 수 있는 초평면(Hyperplane)은 무수히 많으나, 2개의 클래스(110, 120) 사이에서 가장 가까운 점들의 거리 ρ (마진; margin)(150)을 최대화하도록 하면 유일한 해로 초평면을 구할 수 있다.

<44>

일반적으로 ρ 을 최대화하는 초평면의 방정식은 최적의 가중벡터 w_* 와 바이어스 b_* 로 다음 수학식 4 및 5와 같이 표현된다.

수학식 4

<45>

$$w_*^T \cdot x_i + b_* = 0$$

수학식 5

<46>

$$\rho = 2 / \|w\|$$

<47>

이때 거리 ρ 을 최대화하기 위해서는, 다음 수학식 7을 만족하면서 다음 수학식 6을 최소화하여야 한다.

수학식 6

<48>

$$\Phi(w) = 1/2 w^T w$$

수학식 7

<49>
$$d_i(w^T x_i + b) \geq 1 \text{ for } i = 1, \dots, N$$

<50> 상기 수학식 6을 최소화하기 위해 Lagrangian의 안장점(saddle point)을 찾는 문제와 KKT(Karush-Kuhn-Tucker) 조건을 이용하여 Lagrange Multiplier를 찾는 Wolfe dual problem으로 변화시킬 수 있다. 변화된 식을 최대화 하는 값을 가지고 최적 가중벡터 w_* 와 바이어스 b_* 을 구할 수 있다. 구해진 최적 가중벡터와 바이어스에 따라 임의의 입력패턴 x 는 다음 수학식 8에 의해 분류된다.

수학식 8

<51>
$$f(x) = \text{sign}(w_*^T x + b_*)$$

<52> 그러나 일반적인 입력 패턴의 경우 명확하게 선형 분리가 되지 않는 경우가 대부분이므로, 상기 수학식 8을 그대로 적용할 수는 없다. 즉, 선형 SVM을 비선형 SVM으로 확장할 필요가 있는바, 이에 대해서는 도 2를 참조하여 이하에서 상세하게 설명한다.

<53> 도 2는 선형 분리가 되지 않는 경우에 있어서, 커널 함수를 이용하여 고차원 공간(high dimensional space)으로 SVM을 확장하는 것을 나타내는 도면이다. 도 2에 도시된 바와 같이, 입력 패턴의 선형 분리가 불가능한 경우 비선형 특성을 가진 SVM을 사용하게 되는데, 비선형 SVM은 커널(kernel) 함수를 사용하여 선형 분류가 가능한 고차원 공간으로 확장된 특징 공간(feature space)을 가지고 ρ 를 최대화 하는 값을 찾는다.

<54> 고차원의 공간으로 확장시킬 경우 어느 정도는 원 공간에서의 거리 관계를 보존 시킬 필요가 있기 때문에, 커널 함수는 고차원 공간으로의 사상 함수 $\phi(x)$ 을 사용해 다음 수학식 9와 같이 정의한다.

수학식 9

<55>
$$K(x, x_i) = \phi(x)^T \phi(x_i)$$

<56> 여기서, 중요한 점은 커널 트릭(kernel trick)을 사용함으로써 사상 함수에 대한 구체적인 설정 없이도 분류 함수를 구현할 수 있다는 것이다. 본 발명에서는 다음 수학식 10과 같은 RBF(Radial-Basis Function) 커널 함수를 사용하였다.

수학식 10

<57>
$$K(x, x_i) = \exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right)$$

<58> 커널 함수를 사용할 경우, 선형 SVM과 동일한 방법으로 최적 가중벡터 w_* 와 바이어스 b_* 을 구할 수 있으며, 결론적으로 비선형 SVM은 다음 수학식 11로 정리될 수 있다.

수학식 11

<59>
$$f(x) = \text{sign}(w_*^T K(x, x) + b_*)$$

<60> 2. 특징 벡터의 추출

<61> 본 발명에서는, 기존의 통계 모델에 기초한 음성 검출 방법에서 사용되는 주요 파라미터들인 사후 SNR, 선행 SNR 및 예측 SNR을 결합하여 SVM에 대한 특징 벡터로서 새롭게 제안한다.

<62> 원래의 음성신호 $x(n)$, 잡음신호 $d(n)$, $x(n)$ 에 $d(n)$ 가 인가된 입력신호(잡음음성신호) $y(n)$ 에 대한 DFT(Discrete Fourier Transform) 계수를 각각 $X_k(n)$, $D_k(n)$, $Y_k(n)$ (여기서, $Y_k(n)$ 은 n 번째 프레임에서의 k 번째 주파수 성분을 나타냄)이라고 하면, 음성 검출 기법에서 일반적으로 사용되고 있는 기본 가정은 다음 수학식 12와 같다.

수학식 12

$$H_0: \text{speech absent} : Y(n) = D(n) \\ H_1: \text{speech present} : Y(n) = X(n) + D(n)$$

<63>

<64>

상기 수학식 12의 가정에서, Y, X, D는 각각 잡음음성신호, 음성신호, 잡음신호의 DFT 계수 벡터들이다. 음성신호와 잡음신호의 스펙트럼이 복소수 가우시안 분포를 따른다고 가정하면, 상기 가정에 기초한 확률 밀도 함수는 다음 수학식 13 및 수학식 14와 같이 주어진다.

수학식 13

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

<65>

수학식 14

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

<66>

<67>

여기서, $\lambda_{x,k}$ 와 $\lambda_{d,k}$ 는 각 채널별 음성과 잡음의 분산이며, 파라미터로 쓰인 사후 SNR와 선행 SNR은 각각 다음 수학식 15 및 수학식 16과 같이 정의될 수 있다.

수학식 15

$$\gamma_k(n) \equiv \frac{|Y_k(n)|^2}{\lambda_{d,k}(n)}$$

<68>

수학식 16

$$\eta_k \equiv \frac{\lambda_{x,k}(n)}{\lambda_{d,k}(n)}$$

<69>

<70>

상기 수학식 15에서 알 수 있듯이, 특징 벡터의 첫 번째 요소인 사후 SNR은 잡음음성신호(전력)와 음성 부재 구간에서의 잡음 신호로부터 추정되는 잡음신호의 분산 $\lambda_{d,k}$ 을 이용하여 구할 수 있다.

<71>

특징 벡터의 두 번째 요소인 선행 SNR은 DD 추정법을 이용하여 다음 수학식 17과 같이 추정할 수 있다.

수학식 17

$$\hat{\eta}_k(n) \equiv \alpha \frac{|\hat{X}_k(n-1)|^2}{\lambda_{d,k}(n-1)} + (1-\alpha)P[\gamma_k(n)-1], \quad 0 \leq \alpha \leq 1$$

<72>

<73>

여기서, $|\hat{X}_k(n-1)|^2$ 은 이전 프레임에서 추정된 음성 신호의 k번째 스펙트럼 성분의 크기에 대한 추정치이며, 연산자 P[]는 다음 수학식 18과 같이 정의된다.

수학식 18

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

<74>

<75>

특징 벡터의 세 번째 요소인 예측 SNR의 개념은 최근에 소개되었는데, 음성구간에서의 잡음 전력 갱신을 고려하여 배경 잡음과 음성 각각에 long-term smoothed 전력 스펙트럼을 사용하여 추정하는데, 이때 사용되는 잡음과 음성의 분산에 대한 추정치는 다음 수학식 19와 같이 표현된다.

수학식 19

$$\begin{aligned} \hat{\lambda}_{d,k}(n+1) &= \zeta_d \hat{\lambda}_{d,k} + (1-\zeta_d) E[|D_k(n)|^2 | Y_k(n)] \\ \hat{\lambda}_{x,k}(n+1) &= \zeta_x \hat{\lambda}_{x,k} + (1-\zeta_x) E[|X_k(n)|^2 | Y_k(n)] \end{aligned}$$

여기서, $\hat{\lambda}_{d,k}(n)$ 와 $\hat{\lambda}_{x,k}(n)$ 는 $\lambda_{d,k}(n)$ 와 $\lambda_{x,k}(n)$ 의 추정치이며, ζ_d 와 ζ_x 는 잡음과 음성의 정상(stationary) 가정을 고려한 스무딩 파라미터이다.

상기 수학식 19에서 음성의 존재와 부존재를 고려하면, 현재 프레임에서의 잡음과 음성 분산 전력의 추정치는 다음 수학식 20 및 수학식 21과 같다.

수학식 20

$$\begin{aligned} E[|D_k(n)|^2 | Y_k(n)] &= E[|D_k(n)|^2 | Y_k(n), H_0] p(H_0 | Y_k(n)) \\ &\quad + E[|D_k(n)|^2 | Y_k(n), H_1] p(H_1 | Y_k(n)) \end{aligned}$$

수학식 21

$$\begin{aligned} E[|X_k(n)|^2 | Y_k(n)] &= E[|X_k(n)|^2 | Y_k(n), H_0] p(H_0 | Y_k(n)) \\ &\quad + E[|X_k(n)|^2 | Y_k(n), H_1] p(H_1 | Y_k(n)) \end{aligned}$$

여기서, 다음 수학식 22 내지 수학식 25를 이용하여 상기 수학식 20 및 수학식 21을 전개할 수 있다.

수학식 22

$$E[|D_k(n)|^2 | Y_k(n), H_0] = |Y_k(n)|^2$$

수학식 23

$$E[|X_k(n)|^2 | Y_k(n), H_0] = 0$$

수학식 24

$$E[|D_k(n)|^2 | Y_k(n), H_1] = \left(\frac{\hat{\xi}_k(n)}{1 + \hat{\xi}_k(n)} \right) \hat{\lambda}_{d,k}(n) + \left(\frac{1}{1 + \hat{\xi}_k(n)} \right)^2 |\hat{Y}_k(n)|^2$$

수학식 25

$$E[|X_k(n)|^2 | Y_k(n), H_1] = \left(\frac{1}{1 + \hat{\xi}_k(n)} \right) \hat{\lambda}_{x,k}(n) + \left(\frac{\hat{\xi}_k(n)}{1 + \hat{\xi}_k(n)} \right)^2 |\hat{Y}_k(n)|^2$$

$$\hat{\xi}_k(n) = \frac{\hat{\lambda}_{x,k}(n)}{\hat{\lambda}_{d,k}(n)}$$

여기서,

여기서, $p(H_0 | Y_k(n)) (= 1 - p(H_1 | Y_k(n)))$ 는 각 프레임에서의 음성 부재 확률이다. 최종적으로 상기 수학식 19에서 추정된 잡음과 음성의 전력을 이용하여 예측 SNR $\hat{\xi}_k(n)$ 값의 추정치를 구한다. 이 예측 SNR은 직전 프레임에서 현재 프레임의 값을 추정한다는 면에서 “예측(predicted)”이라는 용어가 사용되었고, 기존의 선행 SNR보다 음성 검출 방법의 성능 향상에 우수하다고 알려져 있다.

3. 본 발명에 따른 SVM을 이용한 음성 검출 방법

다음으로, 도 3을 참조하여 본 발명에 따른 SVM에 기초한 음성 검출 방법의 구체적인 구현 방법을 설명하기로

한다.

<90> 도 3은 본 발명에 따른, 사후 SNR, 선행 SNR 및 예측 SNR을 결합하여 특징 벡터로써 사용하는 SNR에 기초한 음성 검출 방법의 구체적인 구현 단계들을 나타내는 블록도이다. 도 3에 도시된 바와 같이, 본 발명에 따른 음성 검출 방법은 크게 학습(training) 단계와 테스트(test) 단계로 나누어진다. 학습 단계는, 학습용 음성 데이터(310)로부터 사후 SNR, 선행 SNR 및 예측 SNR을 추출한 후 이들을 결합하여 학습용 특징 벡터를 생성하는 단계(320)와, 생성된 학습용 특징 벡터를 앞서 언급한 수학적 식 1에 적용하여 최적의 가중벡터와 최적의 바이어스를 구하여 SVM 모델을 생성하는 단계(330)를 포함한다. 이렇게 학습 단계에 의해 SVM 모델이 결정되면, 테스트 단계로 들어간다. 테스트 단계는, 테스트하고자 하는 음성 데이터(340)로부터 사후 SNR, 선행 SNR 및 예측 SNR을 추출한 후 이들을 결합하여 테스트 특징 벡터를 생성하는 단계(350)와, 생성된 테스트 특징 벡터를 SVM 모델 생성 단계(330)에서 생성한 SVM 모델에 대입하여 음성을 검출하는 단계(360)를 포함한다.

<91> 4. 실험 결과

<92> 본 발명에서 제안된 SVM에 기초한 음성 검출 방법의 성능을 평가하기 위하여, G.729B 음성 검출 방법 및 SVM에 기초한 음성 검출 방법 중 기존에 제안되었던 방법들과 ROC(Receiver Operating Characteristics) 곡선을 이용하여 음성 검출 성능을 비교하였다. 실험에 사용된 데이터는 성능 평가를 위해 총 230초의 깨끗한 음성 데이터에 음성과 비음성 부분을 10 ms마다 표시하였다. 분류된 음성 데이터의 음성 구간은 총 57.1%로 음성 44.0%, 비음성 13.1%로 구성되었으며, 잡음 환경은 음성 데이터에 자동차, 도로 잡음을 각각 5dB SNR과 15dB SNR로 부가하였다.

<93> 도 4 및 도 5는 본 발명에 따른 SVM에 기초한 음성 검출 방법의 성능을 G.729B 음성 검출 방법 및 SVM에 기초한 기존의 음성 검출 방법들의 성능과 비교하는 도면으로서, 도 4 및 도 5는 각각 자동차 잡음을 5dB SNR과 15dB SNR로 부가한 경우에, 각각의 방법에 의한 음성 검출 결과를 ROC(Receiver Operating Characteristic) 곡선을 통하여 비교하는 도면이다. 도 4 및 도 5에서, 가로축인 FAR0은 음성이 아닌데 음성이라고 잘못 정보하는 확률(음성 오경보 확률)이고, 세로축인 HRO는 음성이 아닌 것을 정확하게 검출하는 확률(비음성 검출 확률)이다. 도 4 및 도 5를 통해, 자동차 잡음이 부가된 환경에서, 본 발명에서 제안된 음성 검출 방법이 기존의 음성 검출 방법보다 우수한 음성 검출 성능을 나타내는 것을 분명하게 확인할 수 있다.

<94> 도 6 및 도 7은 본 발명에 따른 SVM에 기초한 음성 검출 방법의 성능을 G.729B 음성 검출 방법 및 SVM에 기초한 기존의 음성 검출 방법들의 성능과 비교하는 도면으로서, 도 6 및 도 7은 각각 도로 잡음을 5dB SNR과 15dB SNR로 부가한 경우에, 각각의 방법에 의한 음성 검출 결과를 ROC 곡선을 통하여 비교하는 도면이다. 도 6 및 도 7을 통해, 도로 잡음이 부가된 환경에서도, 본 발명에서 제안된 음성 검출 방법이 기존의 음성 검출 방법보다 우수한 음성 검출 성능을 나타내는 것을 분명하게 확인할 수 있다.

<95> 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.

발명의 효과

<96> 본 발명의 SVM에 기초한 음성 검출 방법에 따르면, 사후 SNR, 선행 SNR 및 예측 SNR을 결합하여 이들을 SVM에 대한 특징 벡터로써 사용함으로써, 실시간으로 음성 신호를 처리해야 하는 이동 통신 환경에 적합하게 하면서도 음성 검출 성능을 대폭 향상시킬 수 있다.

도면의 간단한 설명

- <1> 도 1은 선형 분리(linear separation)를 이용한 SVM을 나타내는 도면.
- <2> 도 2는 선형 분리가 되지 않는 경우에 있어서, 커널 함수를 이용하여 고차원 공간으로 SVM을 확장하는 것을 나타내는 도면.
- <3> 도 3은 본 발명에 따른, 사후 SNR, 선행 SNR 및 예측 SNR을 결합하여 특징 벡터로써 사용하는 SNR에 기초한 음성 검출 방법의 구체적인 구현 단계들을 나타내는 블록도.
- <4> 도 4 및 도 5는 본 발명에 따른 SVM에 기초한 음성 검출 방법의 성능을 G.729B 음성 검출 방법 및 SVM에 기초한 기존의 음성 검출 방법들의 성능과 비교하는 도면으로서, 도 4 및 도 5는 각각 자동차 잡음을 5dB SNR과 15dB SNR로 부가한 경우에, 각각의 방법에 의한 음성 검출 결과를 ROC(Receiver Operating Characteristic) 곡선을

통하여 비교하는 도면.

<5> 도 6 및 도 7은 본 발명에 따른 SVM에 기초한 음성 검출 방법의 성능을 G.729B 음성 검출 방법 및 SVM에 기초한 기존의 음성 검출 방법들의 성능과 비교하는 도면으로서, 도 6 및 도 7은 각각 도로 잡음을 5dB SNR과 15dB SNR로 부가한 경우에, 각각의 방법에 의한 음성 검출 결과를 ROC 곡선을 통하여 비교하는 도면.

<6> <도면 중 주요 부분에 대한 부호의 설명>

<7> 110, 210 : 클래스 1

<8> 120, 220 : 클래스 2

<9> 130, 230 : Support Vectors

<10> 140 : 선형 초평면

<11> 150 : 마진

<12> 240 : 비선형 초평면

<13> 250 : 커널 함수

<14> 310 : 학습용 음성 데이터

<15> 320 : 특징 벡터 추출(사후 SNR, 선행 SNR, 예측 SNR)

<16> 330 : SVM 모델

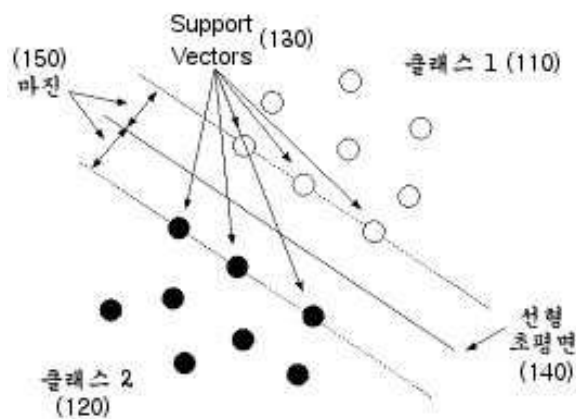
<17> 340 : 테스트 음성

<18> 350 : 특징 벡터 추출(사후 SNR, 선행 SNR, 예측 SNR)

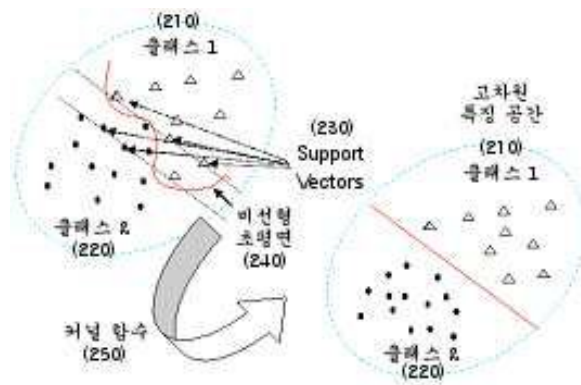
<19> 360 : SVM 모델을 이용한 음성 검출

도면

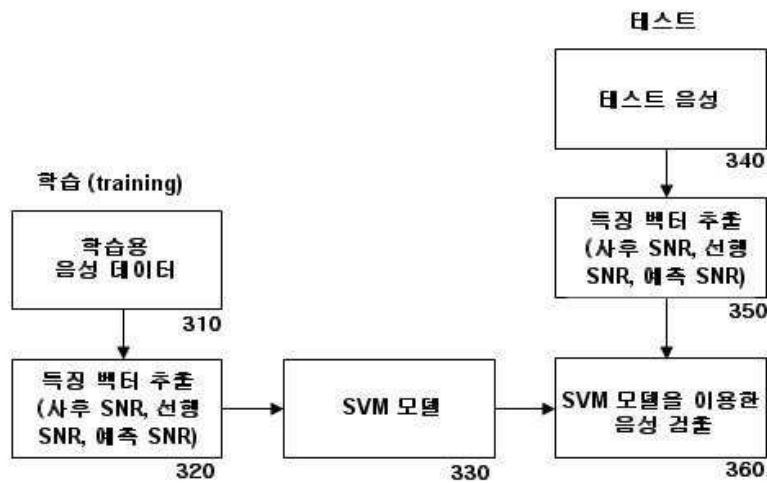
도면1



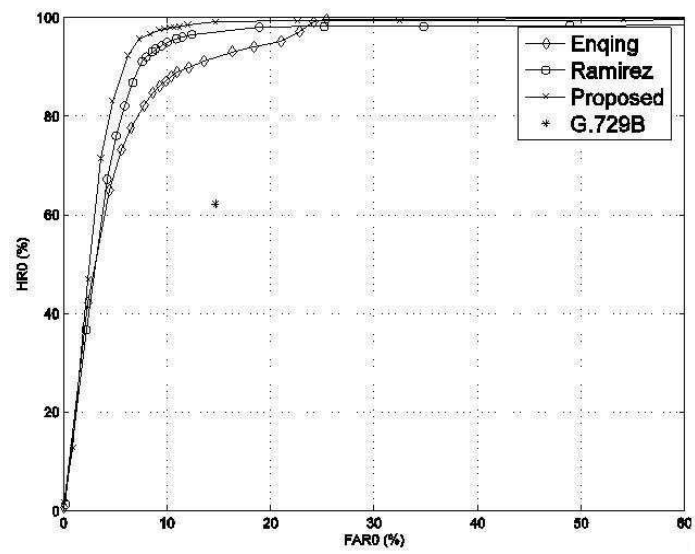
도면2



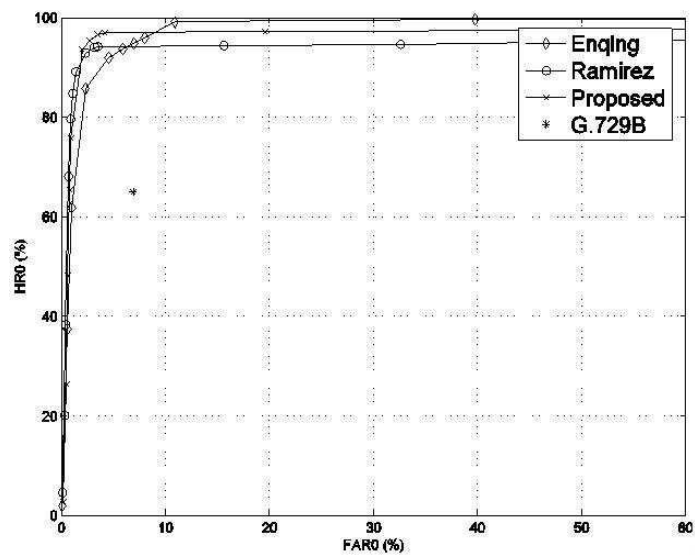
도면3



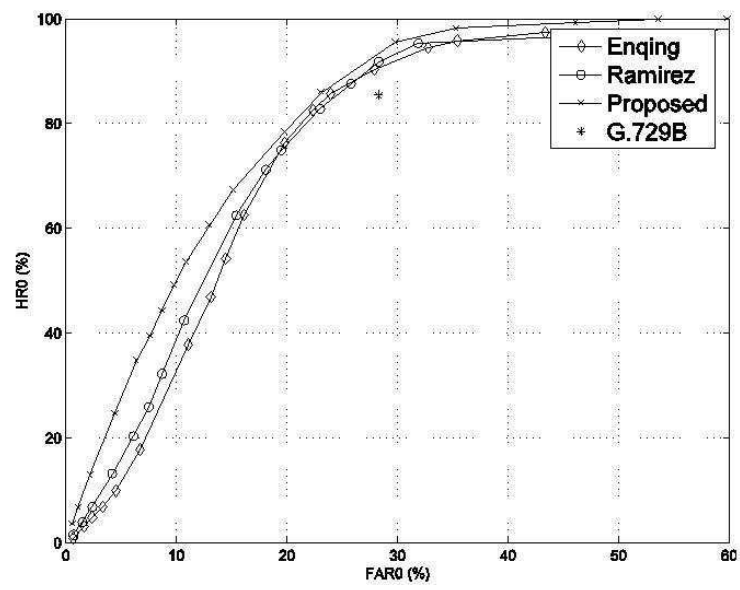
도면4



도면5



도면6



도면7

