



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0021780
(43) 공개일자 2010년02월26일

(51) Int. Cl.

G10L 19/02 (2006.01) G10L 19/06 (2006.01)

(21) 출원번호 10-2008-0080388

(22) 출원일자 2008년08월18일

심사청구일자 2008년08월18일

(71) 출원인

인하대학교 산학협력단

인천 남구 용현동 253 인하대학교

(72) 발명자

장준혁

서울 서초구 서초동 1660-31 B동 301호

이계환

경기 광명시 광명4동 한진아파트 110동902호

(74) 대리인

특허법인태동

전체 청구항 수 : 총 7 항

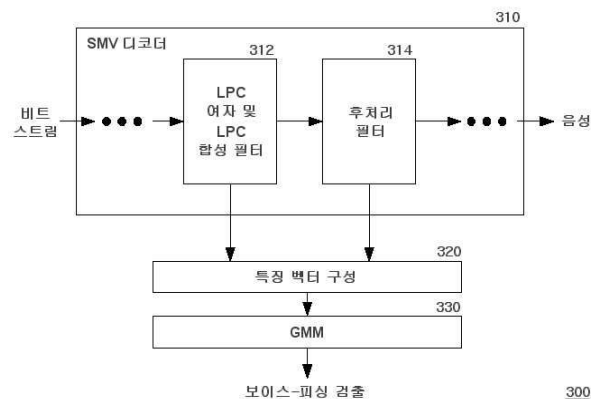
(54) 선택 모드 보코더에 기초한 보이스-피싱 검출 방법

(57) 요약

본 발명은 보이스-피싱 검출 방법에 관한 것으로서, 특히 선택 모드 보코더(SMV)에 기초한 보이스-피싱 검출 방법에 관한 것이다. 본 발명에 따른 SMV에 기초한 보이스-피싱 검출 방법은, (1) 훈련용 음성 신호에 대한 SMV 복호화 과정에서 자동으로 추출되는 파라미터들을 특징 벡터로서 이용하여 가우시안 혼합 모델(Gaussian Mixture Model; GMM)을 구성하는 단계; 및 (2) 입력 음성 신호에 대한 SMV 복호화 과정에서 자동적으로 추출되는 파라미터들을 특징 벡터로서, 상기 단계 (1)에서 구성된 상기 GMM에 적용하여, 상기 입력 음성 신호에 대한 보이스-피싱 여부를 결정하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.

본 발명의 보이스-피싱 검출 방법에 따르면, 3GPP2의 표준 코덱인 SMV 복호화 과정에서 자동으로 추출되는 파라미터들을 특징 벡터로서 이용하여 보이스-피싱을 검출함으로써, 보이스-피싱을 효과적으로 예방할 수 있다. 특히, 별도의 특징 벡터 추출 과정 없이, SMV 복호화 과정에서 자동으로 추출되는 파라미터들을 특징 벡터로서 이용한다는 면에서, 매우 경제적인 보이스-피싱 예방방법이라고 할 수 있다.

대표도 - 도3



특허청구의 범위

청구항 1

선택 모드 보코더(Selectable Mode Vocoder; SMV)에 기초한 보이스-피싱(voice-phishing) 검출 방법으로서,

(1) 훈련용 음성 신호에 대한 SMV 복호화 과정에서 자동으로 추출되는 파라미터들을 특징 벡터로서 이용하여 가우시안 혼합 모델(Gaussian Mixture Model; GMM)을 구성하는 단계; 및

(2) 입력 음성 신호에 대한 SMV 복호화 과정에서 자동으로 추출되는 파라미터들을 특징 벡터로서, 상기 단계 (1)에서 구성된 상기 GMM에 적용하여, 상기 입력 음성 신호에 대한 보이스-피싱 여부를 결정하는 단계

를 포함하는 보이스-피싱 검출 방법.

청구항 2

제1항에 있어서,

상기 단계 (1)의 특징 벡터와 상기 단계 (2)의 특징 벡터로서 동일한 종류의 파라미터들을 이용하는 것을 특징으로 하는 보이스-피싱 검출 방법.

청구항 3

제2항에 있어서,

상기 파라미터들에는, SMV 복호화 과정 중 상기 입력 음성 신호에 대하여 추출된 LSF(Line Spectral Frequency)를 LPC(Linear Prediction Coding)로 변환하는 과정에서 자동으로 추출되는 파라미터들이 포함되는 것을 특징으로 하는 보이스-피싱 검출 방법.

청구항 4

제2항에 있어서,

상기 파라미터들에는, SMV 복호화 과정 중 상기 입력 음성 신호에 대하여 구해진 LPC를 음성 신호로 바꾸는 후 처리 과정에서 자동으로 추출되는 파라미터들이 포함되는 것을 특징으로 하는 보이스-피싱 검출 방법.

청구항 5

제1항에 있어서,

SMV 복호화 과정에서 자동으로 추출되는 상기 파라미터들을 조합하여 특징 벡터로서 이용하는 것을 특징으로 하는 보이스-피싱 검출 방법.

청구항 6

제5항에 있어서,

반사 계수(Reflection Coefficients; RC), LSF의 제1 계수, 피치 지연(pitch lag), 수정된 피치 지연(corrected pitch lag) 파라미터들을 조합하여 특징 벡터로서 이용하는 것을 특징으로 하는 보이스-피싱 검출 방법.

청구항 7

제1항에 있어서,

상기 단계 (2)에서, 상기 입력 음성 신호에 대하여 구성된 상기 특징 벡터를 상기 GMM에 적용한 결과를 우도(likelihood)로서 구한 후, 이를 이용하여 상기 입력 음성 신호에 대한 보이스-피싱 여부를 결정하는 것을 특징으로 하는 보이스-피싱 검출 방법.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 보이스-피싱 검출 방법에 관한 것으로서, 특히 선택 모드 보코더(SMV)에 기초한 보이스-피싱 검출 방법에 관한 것이다. 보다 구체적으로는, 본 발명은, SMV 복호화 과정에서 자동으로 추출되는 파라미터들을 특징 벡터로서 사용하여 보이스-피싱 검출을 위한 가우시안 혼합 모델(Gaussian Mixture Model; GMM)을 구성한 후, 구성된 GMM 모델을 이용하여 입력 음성 신호의 보이스-피싱 여부를 결정하는 보이스-피싱 검출 방법에 관한 것이다.

배경 기술

[0002] 최근 정보 기술 기반의 산업이 발전함에 따라 인터넷이나 핸드폰 등이 생활화되고 있다. 이러한 현상에 발맞추어, 쇼핑물 사거나 고객 정보 해킹 등과 같은 사이버 범죄가 꾸준히 늘어나는 추세이다. 특히, 보이스-피싱(Voice Phishing; 전화 사기)과 같은 신종 경제 범죄의 경우 특별한 조치를 취하기 어렵기 때문에 해가 갈수록 큰 폭으로 증가하고 있으며, 그 수법 또한 갈수록 지능화 되고 있다. 보이스-피싱 범죄는 전 세계적으로 전년도 대비 약 34%라는 큰 폭의 증가세를 보이고 있는데, 특히 다른 나라에 비해 정보 기술 인프라가 생활화되어 있는 우리나라의 경우 더욱 심각한 문제가 되고 있다.

[0003] 일반적인 보이스-피싱의 경우, 대부분의 피해자가 가해자들의 거짓말에 속아 생기는 경우가 대부분이다. 거짓말과 관련하여 많은 연구가 진행되었고, 그 결과 사람은 거짓말을 하는 경우 심리적인 반응을 나타내는 것으로 밝혀졌다. 즉, 사람이 거짓말을 하게 되면, 과장된 웃음을 짓는다든지, 얼굴의 표정이 부자연스러워진다든지, 손을 가만히 있지 못한다든지, 눈을 자주 깜박인다든지, 목소리가 높아진다든지 하는 현상을 자주 관찰할 수 있다. 이러한 현상을 Ekman은 비언어적 누출(Nonverbal Leakage)이라는 표현을 사용하여 정의하였다. 특히, Ekman은, 사람의 목소리의 기본 주파수(Fundamental Frequency)가 진실을 말할 때보다 거짓말을 말할 때 더 높아진다는 사실과 모든 거짓말에는 감정이 개입된다는 사실을 밝혀냈다.

[0004] 그 피해가 점점 심각해지고 있는 보이스-피싱 범죄에 대처하기 위하여, 상기와 같은 거짓말에 대한 연구 결과를 효과적으로 적용해 볼 필요성이 있다.

발명의 내용

해결 하고자하는 과제

[0005] 본 발명은 기존에 제안된 방법들의 상기와 같은 문제점들을 해결하기 위해 제안된 것으로서, 거짓말에 대한 연구 결과를 이용하여 보이스-피싱을 효과적으로 검출할 수 있는 방법을 제안하되, 특히 3GPP2(3rd Generation Partnership Project 2)의 표준 코덱인 SMV 복호화 과정에서 자동으로 추출되는 파라미터들을 이용하여 별도의 특징 벡터 추출 과정 없이 보이스-피싱을 검출함으로써, 보이스-피싱을 효과적이고 경제적으로 검출할 수 있는 방법을 제공하는 것을 그 목적으로 한다.

과제 해결수단

[0006] 상기한 목적을 달성하기 위한 본 발명의 특징에 따른 선택 모드 보코더(Selectable Mode Vocoder; SMV)에 기초한 보이스-피싱(voice-phishing) 검출 방법은,

[0007] (1) 훈련용 음성 신호에 대한 SMV 복호화 과정에서 자동으로 추출되는 파라미터들을 특징 벡터로서 이용하여 가우시안 혼합 모델(Gaussian Mixture Model; GMM)을 구성하는 단계; 및

[0008] (2) 입력 음성 신호에 대한 SMV 복호화 과정에서 자동으로 추출되는 파라미터들을 특징 벡터로서, 상기 단계 (1)에서 구성된 상기 GMM에 적용하여, 상기 입력 음성 신호에 대한 보이스-피싱 여부를 결정하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.

[0009] 바람직하게는, 상기 단계 (1)의 특징 벡터와 상기 단계 (2)의 특징 벡터로서 동일한 종류의 파라미터들을 이용할 수 있다.

[0010] 더욱 바람직하게는, 상기 파라미터들에는, SMV 복호화 과정 중 상기 입력 음성 신호에 대하여 추출된 LSF(Line Spectral Frequency)를 LPC(Linear Prediction Coding)로 변환하는 과정에서 자동으로 추출되는 파라미터들이 포함될 수 있다.

[0011] 더욱 바람직하게는, 상기 파라미터들에는, SMV 복호화 과정 중 상기 입력 음성 신호에 대하여 구해진 LPC를 음

성 신호로 바꾸는 후처리 과정에서 자동으로 추출되는 파라미터들이 포함될 수 있다.

[0012] 바람직하게는, SMV 복호화 과정에서 자동으로 추출되는 상기 파라미터들을 조합하여 특징 벡터로서 이용할 수 있다.

[0013] 더욱 바람직하게는, 반사 계수(Reflection Coefficients; RC), LSF의 제1 계수, 피치 지연(pitch lag), 수정된 피치 지연(corrected pitch lag) 파라미터들을 조합하여 특징 벡터로서 이용할 수 있다.

[0014] 바람직하게는, 상기 단계 (2)에서, 상기 입력 음성 신호에 대하여 구성된 상기 특징 벡터를 상기 GMM에 적용한 결과를 우도(likelihood)로서 구한 후, 이를 이용하여 상기 입력 음성 신호에 대한 보이스-피싱 여부를 결정할 수 있다.

효 과

[0015] 본 발명의 보이스-피싱 검출 방법에 따르면, 3GPP2의 표준 코덱인 SMV 복호화 과정에서 자동으로 추출되는 파라미터들을 특징 벡터로서 이용하여 보이스-피싱을 검출함으로써, 보이스-피싱을 효과적으로 예방할 수 있다. 특히, 별도의 특징 벡터 추출 과정 없이, SMV 복호화 과정에서 자동으로 추출되는 파라미터들을 특징 벡터로서 이용할 수 있다는 면에서, 매우 경제적인 보이스-피싱 예방법이라 할 수 있다.

발명의 실시를 위한 구체적인 내용

[0016] 이하에서는 첨부된 도면들을 참조하여, 본 발명에 따른 실시예에 대하여 상세하게 설명하기로 한다.

[0017] 1. 3GPP2 SMV

[0018] 본 발명에서 사용되는 음성 부호화기인 SMV(Selectable Mode Vocoder)는 ETSI(European Telecommunications Standards Institute)에 의하여 3GPP2의 가변전송률로 표준화된 최신 코덱이다. SMV는 전송 환경과 상태에 따라서 다음 표 1과 같이 다양한 데이터 전송률(Data Rate)을 가지는 알고리즘에 기초하여 효율적인 음성 통신을 제공한다.

표 1

패킷 유형	패킷 당 비트
Rate 1	171 (8.55 kbps)
Rate 1/2	80 (4.0 kbps)
Rate 1/4	40 (2.0 kbps)
Rate 1/8	16 (1.0 kbps)
Blank	0

[0020] 또한, 데이터 전송률과 음질 사이의 절충 관계를 고려하여 모드 0(Premium), 모드 1(Standard), 모드 2(Economy) 및 모드 3(Capacity Saving)과 같은 4가지 모드에서 동작한다. SMV의 데이터 전송률은 음성 신호의 각 프레임의 특성과 SMV의 모드에 따라 선택되는 전송률 결정 알고리즘(Rate Determination Algorithm; RAD)에 의하여 선택되며, 8 kHz의 샘플링을 기준으로 할 경우 약 20ms마다 전송률을 선택한다.

[0021] 도 1은 SMV의 부호화 과정을 개략적으로 보여주는 도면이다. 도 1에 도시된 바와 같이, SMV의 부호화는 먼저 묵음 개선(Silence Enhancement), 고대역 통과 필터(High Pass Filter), 잡음 제거(Noise Suppression) 및 적응 틸트 필터(Adaptive Tilt Filter)와 같은 전처리 과정(Pre-Processing)을 거치게 된다. 전처리 과정을 거친 입력 신호는 프레임 단위마다 지각 가중치가 적용된 음성(Perceptually Weighted Speech), 개회로 피치 검출(Open Loop Pitch Detection), 신호 수정(Signal Modification) 및 LPC 분석(Linear Prediction coding Analysis)을 이용하여 프레임별 음성 검출, 클래스 분류 및 전송률을 결정한 뒤 신호를 비트스트림으로 출력하게 된다. 여기서, SMV는 입력 신호에 대하여 묵음(Silence), 잡음 유사한 무성음(Noise-Like Unvoiced), 무성음(Unvoiced), 시작음(Onset), 비정상 유성음(Non-Stationary Voiced), 및 정상 유성음(Stationary Voiced)과 같은 총 여섯 가지의 클래스로 분류를 하며, 전송률에 따라 정상 유성음을 가지는 타입 0 프레임과 그 외의 모든 타입 1 프레임의 두 가지로 프레임을 분류한다.

[0022] SMV의 복호화의 경우 전송된 비트스트림에서 추출한 데이터 전송률에 대한 정보에 따라 복호화를 하는 방식이 다르다. 도 2a 및 2b는 일반적인 SMV의 복호화 과정을 나타내는 도면이다. SMV의 일반적인 복호화

과정에서는, 전송된 비트스트림에서 LSF(Line Spectral Frequency)를 추출한 뒤, 데이터 전송률에 따라 조금씩 다른 방법을 사용하여 LPC를 추출한다. 그 후, 적응 필터(Adaptive Filter), 고대역 통과 필터(High Pass Filter), 및 신호 확대(Signal Up Scaling)로 구성되는 후처리 과정을 거쳐 출력 음성 신호로 바뀌게 된다.

2. 본 발명에서 제안하는 보이스-피싱 검출 방법

지금까지 연구된 보이스-피싱 검출 방법들은 사람이 거짓말을 할 경우에 발생하는 감정의 변화에 기초한 검출 방법이 대부분이었다. 또한 이산여현변환과 같이 별도의 복잡한 계산량을 필요로 하는 특징 벡터를 사용하여 보이스-피싱을 검출하려는 시도도 있었다. 하지만 본 발명에서는 휴대폰으로 전송되는 신호를 SMV 복호화기에서 음성 신호로 변경하는 복호화 과정에서 자동으로 추출되는 파라미터만을 사용하여 별도의 특징 벡터 추출 과정 없이 보이스-피싱을 검출할 수 있는 효과적인 방법을 제안한다.

본 발명에서 제안하고 있는 보이스-피싱 검출 방법은, 도 2a 및 2b에 도시된 SMV 복호화 처리 과정 중에서 비트스트림에서 추출된 LSF를 LPC로 변환하는 과정에서 자동으로 추출되는 파라미터들과, 구해진 LPC를 음성 신호로 바꾸는 후처리 과정에서 자동으로 추출되는 파라미터들을 이용하여, 보이스-피싱을 검출할 수 있는 가장 효과적인 특징 벡터를 구성한 뒤, 동일 특징 벡터로 이미 훈련된 GMM을 사용하여 실시간으로 전송되는 비트스트림에 대하여 우도를 구하게 된다. 구해진 우도를 비교하여 가장 큰 우도를 가진 값에 대한 플래그(Flag)를 선택하게 된다. 도 3은 본 발명의 일 실시예에 따른 보이스-피싱 검출 시스템의 구성을 나타내는 도면이다. 도 3에 도시된 바와 같이, 본 발명의 일 실시예에 따른 보이스-피싱 검출 시스템은, LPC 여자(excitation) 및 합성 필터(312) 및 후처리 필터(314)를 포함하여 구성되는 SMV 디코더(310), SMV 디코더(310)로부터 자동으로 추출되는 파라미터들을 이용하여 특징 벡터들을 구성하는 특징 벡터 구성부(320), 상기 특징 벡터 구성부에서 구성된 특징 벡터를 적용하여 보이스-피싱을 검출하는 GMM(330)을 포함한다. 이하에서, 본 발명에서 실제로 보이스-피싱 검출에 사용된 특징 벡터들과 GMM 알고리즘을 상세히 설명하기로 한다.

(1) 복호화 과정에서의 중요 특징 벡터들

a. Line Spectral Frequency(LSF)

음성 프레임의 마지막 1/4부분에 중심을 둔 윈도우를 이용하여 계산된 LPC 계수를 단위 원에 맵핑시켜 변환한 값을 말한다. SMV 복호화 과정에서 전송된 비트스트림에서 추출한 값을 이용하여 다음 수학적 식 1과 같은 방법으로 얻어진다.

수학적 식 1

$$lsf(i) = \epsilon(i) + \sum_{p=1}^{O_{Rate}^{MA}} b_p^{Rate}(i) \cdot \epsilon_p(i), \quad i = 1, \dots, 10$$

여기서, ϵ 은 양자화 예측 에러를, ϵ_p 는 p번째 프레임의 양자화 예측 에러를, i는 차수를 각각 나타낸다.

O_{Rate}^{MA} 는 전송률에 따른 이동평균 예측 계수의 차수를, b^{Rate} 는 전송률에 따른 LSF 예측 계수 값을 각각 나타낸다.

b. Linear Prediction Coding(LPC)

SMV의 부호화 과정에서, 자기상관함수와 반사 계수를 사용한 Levinson-Durbin 알고리즘을 사용하여 구해진다. 복호화 과정에서는 얻어지는 LPC 계수 값은 부호화 과정에서 LPC에서 LSF로의 변환 과정의 역변환을 이용하여 얻어진다.

c. LSF의 제1 계수(First-LSF)

전송된 비트스트림으로부터 구한 LSF의 첫 번째 계수를 나타낸다.

d. 장구간 예측(Long-Term Prediction) 이득

프레임의 특성에 따른 적응 코드북(Adaptive CodeBook; ACB)의 이득 값으로 나타낸다.

e. LPC 이득

다음 수학적 식 2와 같이, 반사 계수를 사용하여 구해진 LPC의 예측 이득 값이다.

수학식 2

$$G_{LPC} = -10 \cdot \log_{10} \left(\prod_{i=1}^{10} (1 - k^2(i)) \right), \quad i = 1, \dots, 10$$

여기서, k 는 반사 계수를 나타낸다.

f. 피치 지연(pitch lag), 피치 이득

SMV의 부호화 과정의 개회로 피치 검출 과정에서 추출된 피치 값을 전송된 비트스트림과 적응 코드북을 통해 나타내어진다.

g. 수정된 피치 지연(corrected pitch lag, CPL)

복호화된 피치 지연(l_p)과 반사계수(k), 프레임 타입 그리고 불량 프레임 지시 변수(bad frame indicator, bfi)를 통해 다음 수학식 3과 같이 나타내어진다.

수학식 3

$$\begin{aligned} & \text{Type0:} \\ & CPL(m) = l_p(m-1), \quad \text{if } k(1) < 0 \text{ and } bfi \text{ is } 1 \\ & CPL(m) = l_p(m-1) + 1, \quad \text{otherwise} \\ & \text{Type1:} \\ & CPL(m) = \min(147, \max(17, 3 \cdot A + B)), \quad \text{if } bfi \text{ is } 1 \\ & CPL(m) = l_p(m-1) + 1, \quad \text{otherwise} \\ & A = 0.5 \cdot (l_p(m-1) - l_p(m-2)), \\ & B = \frac{l_p(m-3) + l_p(m-2) + l_p(m-1) - 3 \cdot A}{3} \end{aligned}$$

여기서, m 은 현재 프레임을 나타낸다.

h. 반사계수(Reflection Coefficients; RC)

자기상관함수(R)와 LPC 계수(α)를 이용하여 다음 수학식 4와 같이 계산된다.

수학식 4

$$k_m(i) = - \frac{R_m(i) + \sum_{j=1}^{i-1} \alpha_j^{i-1} \cdot R_m(i-j)}{E^{i-1}}$$

(2) 가우시안 혼합 모델(Gaussian Mixture Model; GMM)

본 발명에서 제안하고 있는 보이스-피싱 검출 방법에 사용되는 GMM은, EM(Expectation Maximization) 알고리즘을 기초로 주어진 데이터 집합에 대한 분포 밀도를 복수의 가우시안 확률밀도함수로 모델링 한 뒤 실제 입력 데이터에 대하여 각 모델별 우도를 계산하고 최대 우도를 가지는 모델을 선택하는 패턴 인식 방법 중 하나이다. 앞서 설명한 LPC 합성 과정과 후처리 과정에서 자동으로 추출되는 N 개의 D 차원 특징 벡터를 $X = \{x_1, x_2, \dots, x_N\}, x_t \in R^D$ 라고 하면, M 개의 혼합 성분(Mixture Component)으로 구성되는 가우시안 확률밀도함수의 모델에 대한 우도(likelihood)는 다음 수학식 5와 같이 구할 수 있다.

수학식 5

$$\begin{aligned} p(x_t | \lambda) &= \sum_{i=1}^M p_i b_i(x_t) \\ b_i(x_t) &= \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x_t - \vec{\mu}_i)^T (\Sigma_i)^{-1} (x_t - \vec{\mu}_i) \right\} \end{aligned}$$

$$0 \leq p_i \leq 1, \sum_{i=1}^M p_i = 1$$

여기서, GMM을 위한 정상 모델 λ_G 와 Phishing 모델 λ_P 는 다음 수학적 6과 같이 가우시안 혼합 성분 밀도의 가중치(Mixture Weight: p_i), 평균 벡터(Mean Vector: $\vec{\mu}_i$), 및 공분산 행렬(Covariance Matrix: Σ_i)로 구성된다.

수학적 6

$$\lambda = \{p_i, \vec{\mu}_i, \Sigma_i\}, i = 1, \dots, M$$

$$p_i = \frac{1}{T} \sum_{t=1}^N p(i|x_t, \lambda)$$

$$\vec{\mu}_i = \frac{\sum_{t=1}^N p(i|x_t, \lambda) \vec{x}_t}{\sum_{t=1}^N p(i|x_t, \lambda)}$$

$$\Sigma_i = \frac{\sum_{t=1}^N p(i|x_t, \lambda) \vec{x}_t^2}{\sum_{t=1}^N p(i|x_t, \lambda)} - (\vec{\mu}_i)^2$$

여기서, i 번째 성분의 사후 확률(A Posteriori Probability)은 다음 수학적 7과 같이 주어진다.

수학적 7

$$p(i|x_t, \lambda) = \frac{p_i b_i(\vec{x}_t)}{\sum_{k=1}^M p_k b_k(\vec{x}_t)}$$

구성된 모델 파라미터 λ 는 EM 알고리즘을 사용하여 $p(x|\lambda') \geq p(x|\lambda)$ 가 되는 새로운 모델 λ' 를 정해진 문턱값에 도달할 때까지 반복하여 사후 확률을 구한다. 구해진 사후 확률 중 가장 큰 우도를 가진 모델을 구한 뒤 입력 값에 대한 모델별 우도 값을 비교하여 보이스-피싱 검출을 하게 된다(아래 수학적 8 참조).

수학적 8

$$\hat{P} = \underset{1 \leq V \leq S}{\operatorname{argmax}} \sum_{t=1}^N \log p(\vec{x}_t | \lambda_V)$$

$S = 2, (1: \text{TRUE}, 2: \text{FALSE})$

3. 실험결과

본 발명에서 제안하고 있는, 음성 코딩 파라미터에 기초한 보이스-피싱 검출 방법의 성능을 평가하기 위해서 보다 실제에 가까운 데이터의 수집이 필요하였다. 따라서 재연이 아닌 실제 보이스-피싱 가해자의 목소리를 수집하는 방법을 통하여 데이터를 모았다. 실제 가해자의 목소리는 국내외 각종 인터넷 사이트에서 모았으며, 음성 변조된 경우 등은 제외시켰다. 다음 표 2는 음성 데이터들에 대한 세부 정보를 나타낸다.

표 2

	남자	여자
일반적인 전화통화 목소리	5명 (약 3분)	3명 (약 1분 30초)
보이스-피싱 가해자 목소리	5명 (약 3분)	3명 (약 1분 30초)

[0066] 수집된 데이터는 모두 8 kHz로 샘플링하였으며, 16 bit로 양자화하였다. 보이스-피싱 검출을 위한 특징 벡터로서, 앞서 설명한 SMV의 복호화 과정에서 자동으로 추출되는 LPC 10차, LSF 10차, RC 10차를 이용하였으며, 16개의 혼합 성분을 사용한 GMM을 구성하여 실험을 수행하였다. 도 4a 내지 4c는 각각의 특징 벡터에 대한 수신기 동작 특성(Receiver Operating Characteristic; ROC) 곡선을, 남자, 여자, 남자 + 여자인 경우로 나누어 나타낸다. 또한, 다음 표 3은 각각의 특징 벡터에 대한 평균 오차율(Equal Error Rate; EER)을 나타낸다.

표 3

[0067]

특징 벡터	EER		
	남성	여성	남성 + 여성
LPC	24.64 %	37.74 %	41.08 %
LSF	29.05 %	42.60 %	46.77 %
RC	23.38 %	36.96 %	39.59 %

[0068] 인식 결과 기존의 감정 인식에서 많이 사용되었던 LPC에 비해 남성의 경우 RC가 더욱 좋은 인식률을 보였다. 하지만 여성의 경우 LPC가 좋은 인식률을 보였으며, 남성과 여성의 데이터를 혼합하여 인식 성능을 확인한 결과 RC가 가장 좋은 성능을 나타내었다.

[0069] 비교 결과를 보면 전체적으로 남성보다 여성의 경우 인식 성능이 떨어지는 것을 확인할 수 있는데, 이것은 수집된 데이터가 부족했기 때문이라고 생각된다. 더 많은 데이터를 포함한다면, 여성의 경우에도 남성의 경우와 비슷한 인식률을 나타낼 것이라 생각된다. 또한 LPC와 수식적으로 동등한 LSF의 경우 RC와 LPC에 비해서 낮은 인식률을 보이는데, 이것은 효과적인 양자화를 위해 단위 원상의 값들로 변환하였기 때문이라고 예상된다.

[0070] 가장 인식 성능이 우수한 RC 특징 벡터의 인식 성능을 보완하기 위해서, 후처리 과정에서 추출되는 특징 벡터들과 결합하여 새로운 특징 벡터를 만들어 실험을 하였다. 실험결과는 다음 표 4와 같다.

표 4

[0071]

특징 벡터	EER		
	남성	여성	남성 + 여성
RC + First LSF	22.95 %	35.91 %	38.89 %
RC + LTP 이득	24.38 %	35.84 %	40.93 %
RC + LPC 이득	23.48 %	36.13 %	39.87 %
RC + pitch lag	24.27 %	32.20 %	36.48 %
RC + CPL	22.95 %	30.30 %	36.15 %
RC + pitch 이득	25.20 %	36.71 %	39.60 %

[0072] 실험 결과, 피치와 관련된 특징 벡터에서 기존의 RC 성능보다 우수한 성능을 나타내었으며, 나머지 특징들은 기존의 RC 성능에 비해 낮은 성능을 나타내었다. 또한, RC + LPC, RC + LSF, 및 RC + LPC + LSF의 경우도 RC에 비해 낮은 성능을 나타내었다.

[0073] 이와 같은 결과는 사람의 감정이 사람의 목소리를 특징지을 수 있는 성도의 특성을 고려하기 때문이라 생각된다. 특히, 피치의 경우 보이스-피싱에 큰 영향을 미치는 요소인데, 이는 Ekman의 연구 결과에도 잘 나와 있듯이, 거짓말을 할 경우 일반적인 경우에 비해서 높은 기본 주파수를 가지기 때문이라 생각된다. 도 5는 실험에 사용된 남성 및 여성 데이터에 대한 피치 분포와 일반적인 전화 통화 시의 남성 및 여성의 피치 분포를 나타내고 있다.

[0074] 본 실험을 위하여 수집한 가해자의 목소리(False) 및 일반적인 전화 목소리(True)의 피치에 대한 히스토그램과, OGI 데이터베이스를 사용하여 추출한 피치의 히스토그램을 비교하여 본 결과, 일반적인 전화 목소리 데이터베이스인 OGI 데이터베이스의 결과와 실제 수집된 일반 목소리(True)는 거의 비슷한 피치 분포를 나타내었으나, 가해자의 목소리(False)의 경우 일반적인 목소리에 비해 약 50~100 Hz 정도 높은 피치 분포를 보였다. 따라서 가장 좋은 성능을 보여주었던 피치, 피치 지연, 및 LSF의 첫 번째 계수를 RC와 결합하여 최종적인 보이스-피싱 검출 성능을 테스트하였다. 도 6a 내지 6c와 표 5는 최종적으로 행해진 실험에 대한 ROC 곡선과 EER을 각각 나타

내고 있다.

표 5

[0075]

특징 벡터	EER		
	남성	여성	남성 + 여성
RC + First LSF + CPL + Pitch Lag	22.89 %	29.95 %	33.60 %

[0076]

최종적으로 제안된 특징 벡터를 이용한 실험 결과, 남성의 경우 많은 성능향상을 보이지는 않았지만 여성의 경우 많은 성능 향상을 보였으며, 남성과 여성의 데이터를 혼합한 경우에도 우수한 성능을 보였다. SMV의 복호화 과정에서 자동으로 추출되는 특징 벡터 중 가장 우수한 성능을 보인 것을 알 수 있었으며, 피치 정보가 감정과 보이스-피싱 검출에 큰 영향을 주는 것을 분명하게 확인할 수 있었다.

[0077]

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

[0078]

도 1은 SMV의 부호화 과정을 개략적으로 보여주는 도면.

[0079]

도 2a 및 2b는 일반적인 SMV의 복호화 과정을 나타내는 도면.

[0080]

도 3은 본 발명의 일 실시예에 따른 보이스-피싱 검출 시스템의 구성을 나타내는 도면.

[0081]

도 4a 내지 4c는 각각의 특징 벡터에 대한 수신기 동작 특성(Receiver Operating Characteristic; ROC) 곡선을 나타내는 도면.

[0082]

도 5는 실험에 사용되어진 남성 및 여성 데이터에 대한 피치 분포와 일반적인 전화 통화 시의 남성 및 여성의 피치 분포를 나타내는 도면.

[0083]

도 6a 내지 6c는 최종 특징 벡터(RC + First LSF + Pitch + Pitch Lag)에 대한 ROC 곡선을 나타내는 도면.

[0084]

<도면 중 주요 부분에 대한 부호의 설명>

[0085]

210: LPC 합성 필터

[0086]

220: 후처리 필터

[0087]

300: (본 발명의 일 실시예에 따른) 보이스-피싱 검출 시스템

[0088]

310: SMV 디코더

[0089]

312: LPC 합성 필터

[0090]

314: 후처리 필터

[0091]

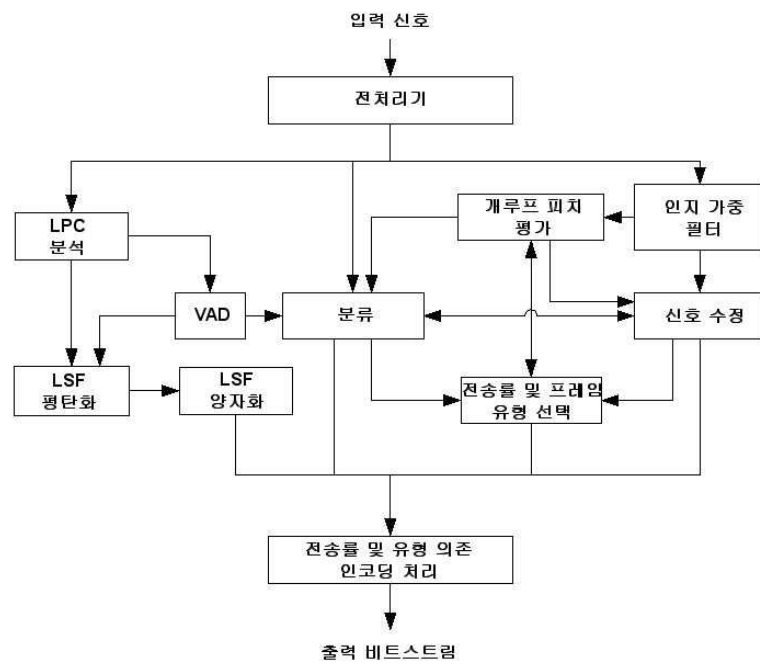
320: 특징 벡터 구성부

[0092]

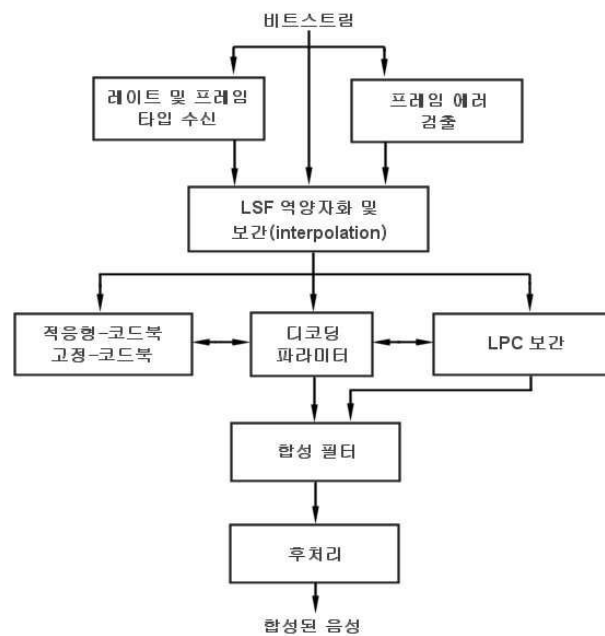
330: GMM

도면

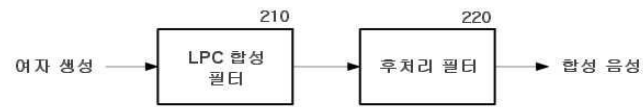
도면1



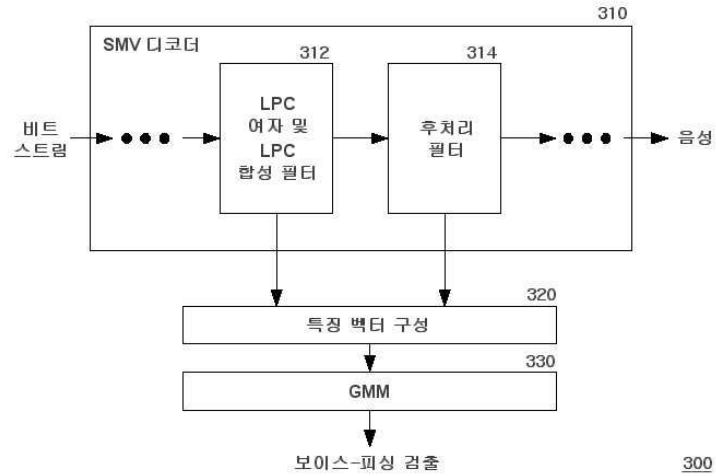
도면2a



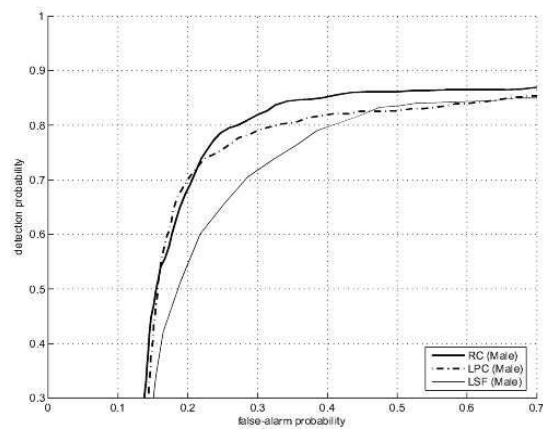
도면2b



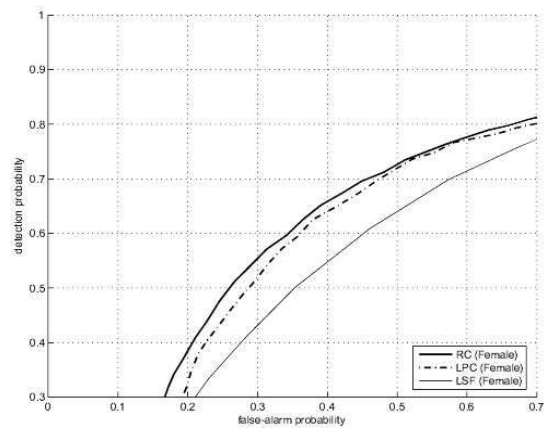
도면3



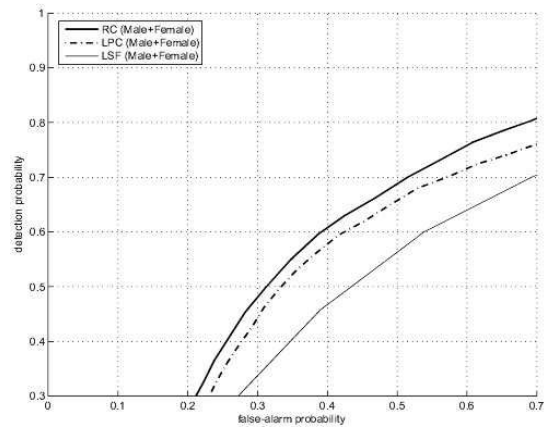
도면4a



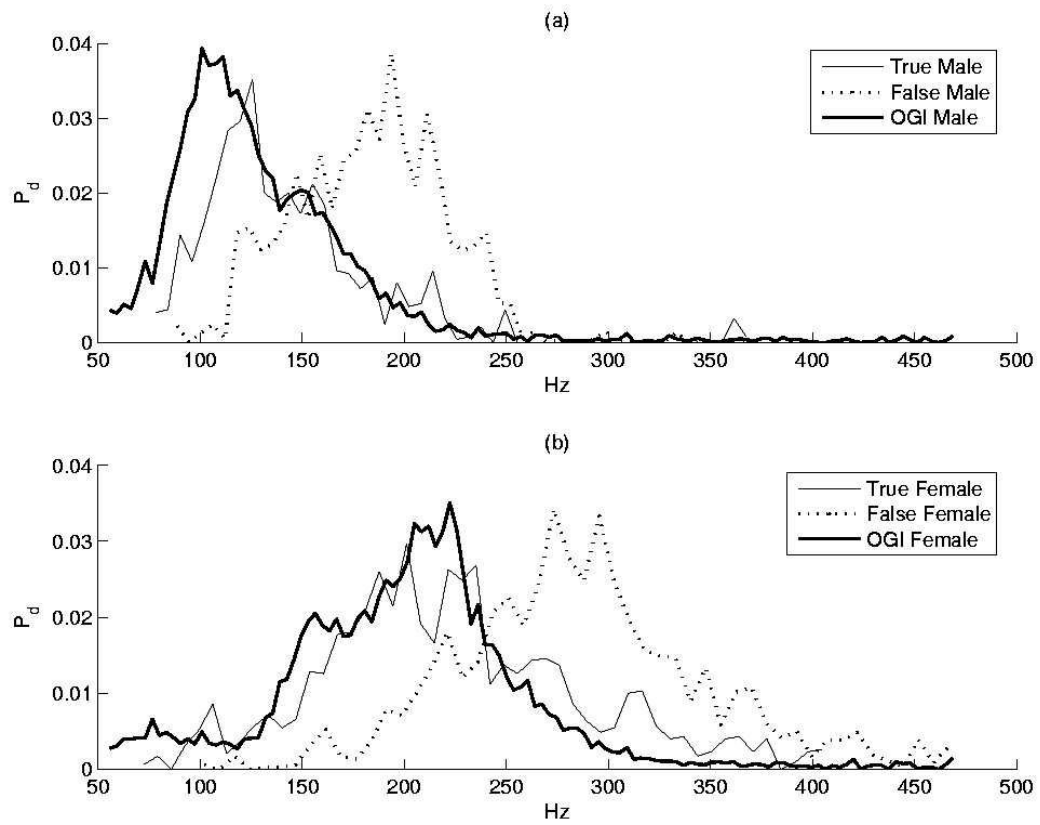
도면4b



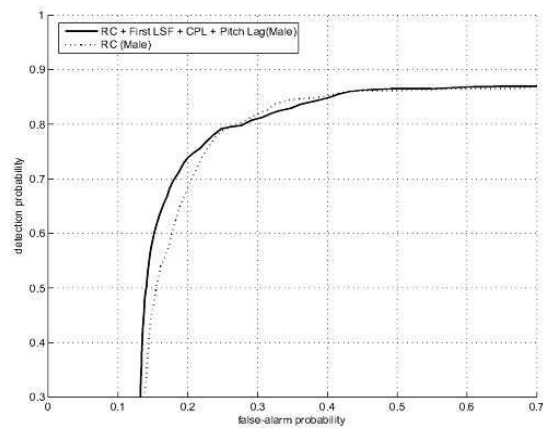
도면4c



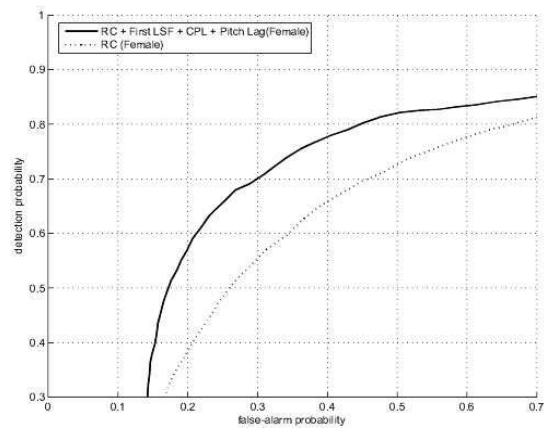
도면5



도면6a



도면6b



도면6c

