

논문 2008-45SP-4-11

3GPP2 SMV 기반의 보이스 피싱 검출 알고리즘

(Voice-Pishing Detection Algorithm Based on 3GPP2 SMV)

이 계 환*, 장 준 혁**

(Kye-Hwan Lee and Joon-Hyuk Chang)

요 약

본 논문에서는 보이스 피싱 (Voice Pishing) 예방을 위한 알고리즘을 3GPP2 Selectable Mode Vocoder (SMV) 코딩 파라미터를 기반으로 제안한다. 상대방 휴대폰에서 전송된 신호를 기반으로 SMV의 복호화 과정에서 자동적으로 추출되는 중요 특징벡터만을 사용하여 Gaussian Mixture Model (GMM)을 구성하고 이를 기반으로 보이스 피싱 예방을 위한 검출 알고리즘을 제안하였다. 실험 결과 제안된 코딩 파라미터 기반의 보이스 피싱 알고리즘이 전화사기 예방에 우수한 성능을 보인 것을 알 수 있었다.

Abstract

We propose an effective voice-pishing detection algorithm based on the 3GPP2 selectable mode vocoder (SMV). The detection of voice pishing is performed based on a Gaussian mixture model (GMM) using decoding parameters of the SMV directly extracted from the decoding process of the transmitted speech information in the mobile phone. The experimental results indicate that SMV decoding parameters are effective in discriminating between general voice and phisher's voice and the performance is significantly acceptable when the proposed technique is applied.

Keywords : 보이스 피싱, Selctable Mode Vocoder (SMV), Gaussian Mixture Model (GMM)

I. 서 론

최근 정보기술 기반의 산업이 점차 발전함에 따라 인터넷이나 핸드폰 등이 점차 생활화 되었다. 이러한 현상에 발마추어 쇼핑몰사기나 고객정보 해킹 등과 같은 사이버범죄가 꾸준히 늘어나는 추세이다. 특히 전화사기 (Voice Pishing)와 같은 신종경제 범죄의 경우 특별한 조취를 취하기 어렵기 때문에 해가 갈수록 큰 폭으로 증가하고 있다. 그 결과 갈수록 지능화 되고 있는 전화사기는 전 세계적으로 전년도 대비 약 34%라는 증가를 보이고 있으며, 다른 나라에 비해 정보기술 인프라가 생활화처럼 되어 있는 우리나라의 경우 더욱 심각한 현실이다.

일반적인 전화사기의 경우 많은 피해자들이 가해자들의 진실과 같은 거짓말에 속아서 생기는 경우가 대부분이다. 이러한 거짓말에 대해 많은 연구가 진행되었고, 그 결과 사람은 거짓말에 대한 심리적인 반응을 나타내는 것으로 알려져 있다^[1]. 사람이 거짓말과 같은 표현을 통해서 긴장을 하게 되면, 말로써는 긴장되지 않는다고 하지만, 과장된 웃음이나 얼굴의 표정이 부자연스러워지며, 손을 가만히 있지 못하고, 눈을 자주 깜박인다든지, 목소리가 높아지는 것과 같은 현상을 자주 관찰할 수 있다. 이러한 현상을 Ekman은 비언어적 누출 (Nonverbal Leakage)이라는 표현을 사용하여 정의하였다. 특히 목소리의 기본 주파수 (Fundamental Frequency)의 경우 진실을 말할 때 보다 거짓말을 할 때 더 높다는 사실과 모든 거짓말에는 감정이 개입되었다는 사실을 밝혀냈다^[2].

본 논문에서는 이러한 전화사기를 예방하기 위해서 보이스 피싱 검출 알고리즘을 제안한다. 가해자의 휴대

* 학생회원, ** 정회원, 인하대학교 전자공학부
(Department of Electronics Engineering, Inha University)
접수일자: 2007년11월6일, 수정완료일: 2008년4월21일

폰에서 전송된 음성 신호를 기반으로 3GPP2 표준 코덱인 Selectable Mode Vocoder (SMV)의 부호화 과정에서 자동적으로 추출되는 중요 특징벡터의 비교, 분석을 통해 참말과 거짓말에 대한 차이점을 밝히고, 이를 사용하여 가우시안 혼합모델 (Gaussian Mixture Model, GMM)을 구성한다. 그리고 추출한 특징벡터가 효과적인 보이스 피싱 검출을 위한 특징이 될 수 있는지 검증한다. 다양한 실험 결과 새로이 제안된 음성 코딩 파라미터를 이용한 보이스 피싱 알고리즘이 우수한 성능을 보인 것을 알 수 있었다.

본 논문의 구성으로는, II장에서는 실험에서 사용되어진 SMV에 대해서 기술하고, III장에서는 제안된 보이스 피싱 알고리즘에 대해 기술한다. IV장에서는 실험 결과 비교 및 분석에 대해 기술하였으며, 마지막으로 V장에서는 결론을 맺는다.

II. 3GPP2 Selectable Mode Vocoder

본 논문에서 사용되어진 음성 부호화기인 SMV는 ETSI에 의해 3GPP2의 가변전송률로 표준화 되어진 최신 코덱이다. SMV는 전송 환경과 상태에 따라서 표 1과 같이 다양한 데이터 전송률 (Data Rate)을 가지는 알고리즘에 기반하여 효율적인 음성 통신을 제공한다³⁻⁵.

또한, 데이터 전송률과 음질 사이의 절충관계를 고려하여 모드 0 (Premium), 모드 1 (Standard), 모드 2 (Economy) 그리고 모드 3 (Capacity Saving)과 같은 4가지 모드에서 동작한다. 그리고 SMV의 데이터 전송률은 음성신호의 각 프레임의 특성과 SMV의 모드에 따라 선택되는 전송률 결정 알고리즘 (Rate Determination Algorithm, RAD)에 의해 선택되며, 8 kHz의 샘플링을 기준으로 할 때 약 20ms마다 전송률을 선택한다.

그림 1은 입력신호에 대한 개략적인 SMV의 부호화

표 1. SMV의 데이터 전송률
Table 1. Data rates of SMV.

Packet Type	Bits per Packet
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

과정을 보여주고 있다. SMV의 부호화는 묵음 개선 (Silence Enhancement), 고대역 통과 필터 (High Pass Filter), 잡음 제거 (Noise Suppression) 그리고 적응 틸트 필터 (Adaptive Tilt Filter)와 같은 전처리 (Pre-Processing)를 거치게 된다. 전처리 과정을 거친 입력신호는 프레임 단위마다 지각가중치가 적용된 음성 (Perceptually Weighted Speech), 개회로 피치 검출 (Open Loop Pitch Detection), 신호 수정 (Signal Modification) 그리고 LPC 분석 (Linear Prediction coding Analysis)을 이용하여 프레임 별 음성검출 및 클래스분류 그리고 전송률을 결정한 뒤 신호를 비트스트림으로 출력하게 된다. 여기서 SMV는 입력신호에 대해 묵음 (Silence), 잡음 유사한 무성음 (Noise-Like Unvoiced), 무성음 (Unvoiced), 시작음 (Onset), 비정상 유성음 (Non-Stationary Voiced) 그리고 정상 유성음 (Stationary Voiced)과 같이 총 여섯 가지의 클래스로 분류를 하며, 전송률에 따라 정상 유성음을 가지는 타입 0 프레임과 그 외의 모든 타입 1 프레임의 두 가지로 프레임을 분류한다.

SMV의 부호화의 경우 전송되어진 비트스트림에서 추출한 데이터 전송률에 대한 정보에 따라 부호화를 하는 방식이 다르다. 그림 2는 전송률에 관계없이 일반적인 SMV의 부호화 과정을 보여준다. SMV의 부호화는 일반적으로 전송되어진 비트스트림에서 LSF (Line Spectral Frequency)를 추출한 뒤, 데이터 전송률에 따라 조금씩 다른 방법을 사용하여 LPC를 추출한다. 그리고 적응 필터 (Adaptive Filter), 고대역 통과 필터

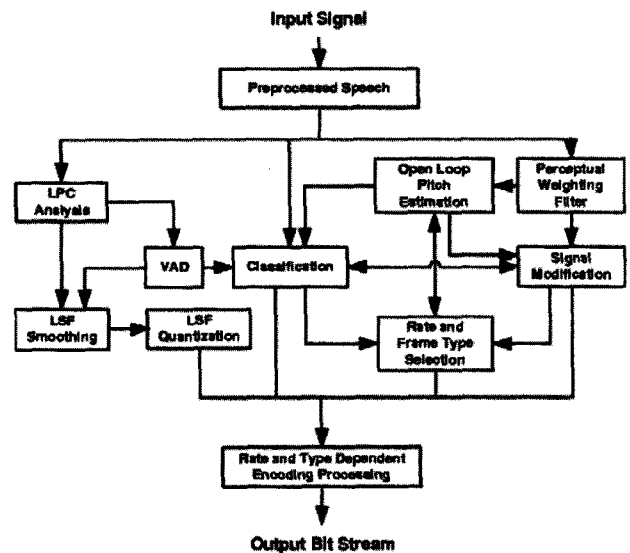


그림 1. SMV 부호화 처리과정의 다이어그램
Fig. 1. Flow diagram of SMV encoding process.

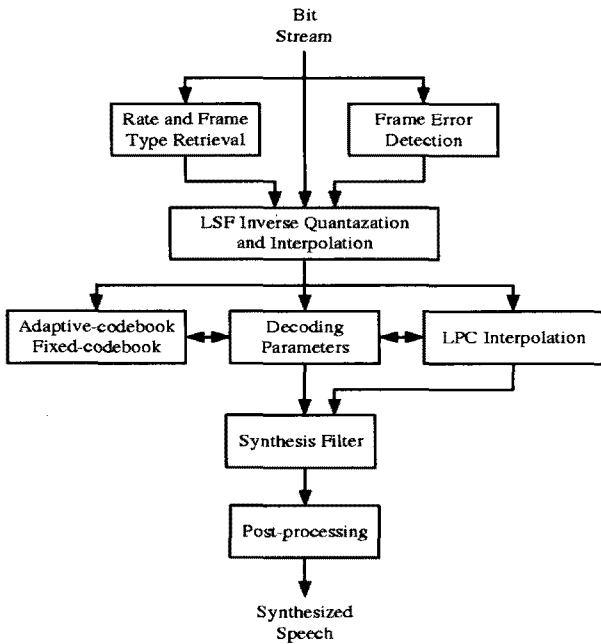


그림 2. SMV 복호화 처리과정의 다이어그램
Fig. 2. Flow diagram of SMV decoding process.

(High Pass Filter) 그리고 신호 확장 (Signal Up Scaling)으로 구성되는 후처리과정을 거쳐 출력 음성 신호로 바뀌게 된다.

III. 보이스 피싱 검출을 위해 제안된 알고리즘

지금까지 연구되어진 보이스 피싱 검출 알고리즘은 사람이 거짓말 할 경우 감정의 변화에 기반한 검출 방법이 대부분이었다^[6~7]. 또한 이산여현변환과 같이 별도의 복잡한 계산량을 필요로 하는 특징벡터를 사용하여 이루어진 연구가 이루어 졌다^[8~9]. 하지만 본 논문에서는 휴대폰으로 전송되어지는 신호를 SMV 복호화기에서 음성신호로 변경하는 복호화 과정에서 자동적으로 추출 되어지는 파라미터만을 사용해서 별도의 특징벡터 추출과정 없이 보이스 피싱 검출을 위한 효과적인 알고리즘을 제안한다.

제안된 알고리즘은 그림 2의 SMV의 복호화 처리과정 중에서 비트스트림에서 추출된 LSF를 LPC로 변환하는 과정에서 자동적으로 추출되는 특징벡터와 구해진 LPC를 음성 신호로 바꾸는 후처리 과정에서 자동적으로 추출되어지는 특징벡터를 이용하여 보이스 피싱 검출을 위한 가장 효과적인 특징벡터를 구성한 뒤 동일 특징벡터로 이미 훈련된 GMM을 사용하여 실시간으로 전송되는 비트스트림에 대해서 우도 (Likelihood)를 구한다. 구해진 우도를 비교해서 가장 큰 우도를 가진 값

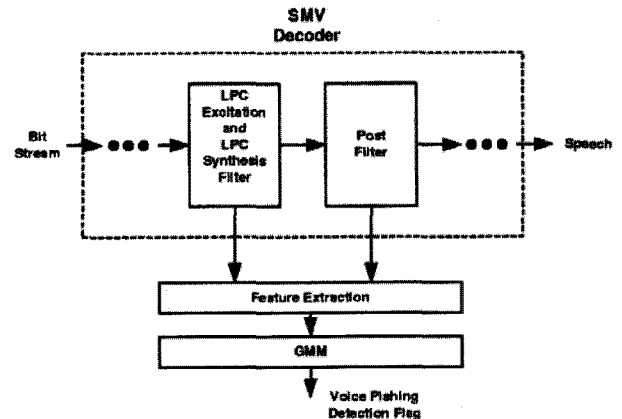


그림 3. 보이스 피싱 검출을 위해 제안된 SMV 알고리즘 블록도
Fig. 3. Diagram of the proposed SMV-based algorithm for oice pishing detection.

에 대한 플래그 (Flag)를 선택하게 된다. 관련하여 그림 3은 제안된 알고리즘의 블록도를 보여주고 있으며, 실제로 보이스 피싱 검출에 사용된 특징벡터들과 GMM 알고리즘을 다음과 같이 기술한다.

1. 복호화 과정에서의 중요 특징벡터

1.1. Line Spectral Frequency (LSF)

음성 프레임의 마지막 1/4부분에 중심을 둔 윈도우를 이용하여 계산된 LPC 계수를 단위원에 맵핑시켜 변환한 값이며, SMV 복호화 과정에서 전송된 비트스트림에서 추출한 값을 이용하여 다음과 같은 방법으로 얻어진다.

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

$$i = 1, \dots, 10$$

여기서 ϵ 은 양자화 예측 에러이며, ϵ_p 는 p 번째 프레임의 양자화 예측 에러를, i 는 차수를 나타낸다. O_{Rate}^{MA} 는 전송률에 따른 이동평균 예측 계수의 차수이며^[10], b^{Rate} 는 전송률에 따른 LSF 예측 계수값을 나타낸다.

1.2. Linear Prediction Coding (LPC)

SMV의 부호화과정에서 자기상관함수와 반사계수를 사용한 Levinson-Durbin 알고리즘^[11]을 사용하여 구해진다. 복호화 과정에서는 얻어지는 LPC 계수 값은 부호화 과정에서 LPC에서 LSF로의 변환과정의 역변환을 이용하여 얻어진다^[12].

1.3. LSF의 첫 번째 계수 (First-LSF)

전송된 비트스트림으로부터 구한 LSF의 첫 번째 계수를 나타낸다.

$$k_m(i) = - \frac{R_m(i) + \sum_{j=1}^{i-1} \alpha_j^{i-1} \cdot R_m(i-j)}{E^{i-1}} \quad (4)$$

1.4. 장구간 예측 (Long-Term Prediction) 이득

프레임 특성에 따른 적응 코드북 (Adaptive Codebook, ACB)의 이득 값으로 나타낸다 [5].

1.5. LPC 이득

다음과 같이 반사계수를 사용하여 구해진 LPC의 예측 이득 값이다.

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

$i = 1, \dots, 10$

여기서 k 는 반사계수이다.

1.6. 피치 지연 (Pitch Lag), 피치 이득

SMV의 부호화 과정의 개회로 피치 검출 과정에서 추출되어진 피치값을 전송된 비트스트림과 적응 코드북을 통해 나타내어진다.

1.7. 수정된 피치 지연 (Corrected Pitch Lag, CPL)

다음과 같이 전송된 비트스트림으로부터 추출해낸 피치 지연과 프레임 타입, 불량 프레임 지시변수 (Bad Frame Indicator, BFI), 반사계수를 이용하여 수정된 피치 지연이다.

Type 0:
 If $k(1) < 0$ and BFI is 1
 $CPL(m) = l_p(m-1),$
 Otherwis
 $CPL(m) = l_p(m-1) + 1,$ (3)

Type 1:
 If BFI is 1
 $CPL(m) = \min(147, \max(17, 3 \cdot A + B)),$
 Otherwis
 $CPL(m) = l_p(m-1) + 1,$
 $A = 0.5 \cdot (l_p(m-1) - l_p(m-2)),$
 $B = \frac{l_p(m-3) + l_p(m-2) + l_p(m-1) - 3 \cdot A}{3}$

여기서 m 은 현재 프레임을 나타낸다.

1.8. 반사계수 (Reflection Coefficients, RC)

자기상관함수 (R)와 LPC 계수 (α)를 이용하여 계산된다.

2. Gaussian Mixture Model (GMM)

제안된 보이스 피싱 검출에 사용되는 GMM은 EM (Expectation Maximization) 알고리즘을 기반으로 주어진 데이터 집합에 대한 분포밀도를 복수개의 가우시안 확률밀도함수로 모델링 한 뒤 실제 입력 데이터에 대해 각 모델별 우도를 계산하고 최대 우도를 가지는 모델을 선택하는 패턴인식 방법 중 하나이다 [13~14].

이미 기술한 LPC 합성과 후처리 과정에서 자동적으로 추출한 되어지는 N 개의 D 차원 특징벡터를 $X = \{x_1, x_2, \dots, x_N\}, x_t \in R^D$ 라고 하면, M 개의 혼합 성분 (Mixture Component)으로 구성되는 가우시안 확률밀도함수의 모델에 대한 우도는 다음과 같이 구할 수 있다.

$$p(x_t | \lambda) = \sum_{i=1}^M p_i b_i(x_t)$$

$$b_i(x_t) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x_t - \vec{\mu}_i)^T (\Sigma_i)^{-1} (x_t - \vec{\mu}_i) \right\}$$

$$0 \leq p_i \leq 1, \sum_{i=1}^M p_i = 1 \quad (5)$$

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

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

$$p_i = \frac{1}{T} \sum_{t=1}^T p(i | \vec{x}_t, \lambda)$$

$$\vec{\mu}_i = \frac{\sum_{t=1}^T p(i | \vec{x}_t, \lambda) \vec{x}_t}{\sum_{t=1}^T p(i | \vec{x}_t, \lambda)} \quad (6)$$

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

여기서 i 번째 성분의 사후확률 (A Posteriori Probability)은 다음과 같이 주어진다.

$$p(i|\vec{x}_t, \lambda) = \frac{p_i b_i(\vec{x}_t)}{\sum_{k=1}^M p_k b_k(\vec{x}_t)} \quad (7)$$

구성되어진 모델 파라미터 λ 는 EM 알고리즘을 사용하여 $p(x|\lambda') \geq p(x|\lambda)$ 가 되는 새로운 모델 λ' 를 정해진 문턱 값에 도달할 때까지 반복하여 사후확률을 구한다. 구해진 사후확률 중 가장 큰 우도를 가진 모델을 구한 뒤 입력 값에 대한 모델별 우도 값을 비교하여 보이스 피싱 검출을 하게 된다^[15].

$$\hat{P} = \underset{1 \leq V \leq S}{\operatorname{argmax}} \sum_{t=1}^N \log p(\vec{x}_t | \lambda_V) \quad (8)$$

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

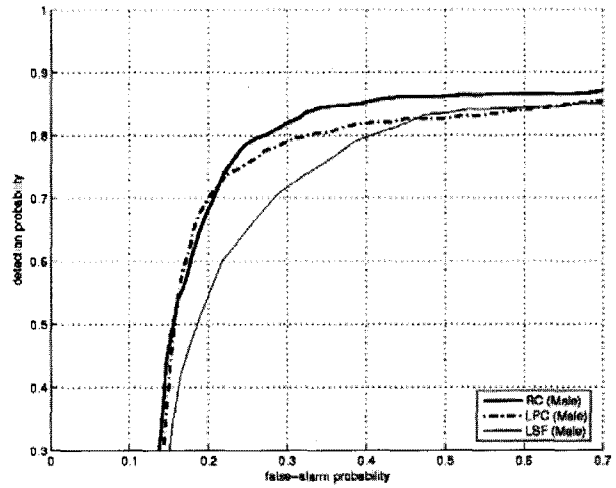
IV. 실험결과 분석 및 비교

본 논문에서 제안한 음성코딩 파라미터 기반의 보이스 피싱 검출 알고리즘의 성능 평가를 위해서는 보다 실제에 가까운 데이터의 수집이 필요하다. 따라서 직접 상황을 연출하여 데이터를 수집하는 방법보다는 재연이 아닌 실제 보이스 피싱 가해자의 목소리를 수집하는 방법을 통하여 데이터를 모았다. 실제 가해자의 목소리는 한국어를 대상으로 국내외 각종 인터넷 사이트에서 모았으며, 음성변조와 같은 경우는 제외 시켰다. 표 2는 음성 데이터들에 대한 세부정보를 보여준다.

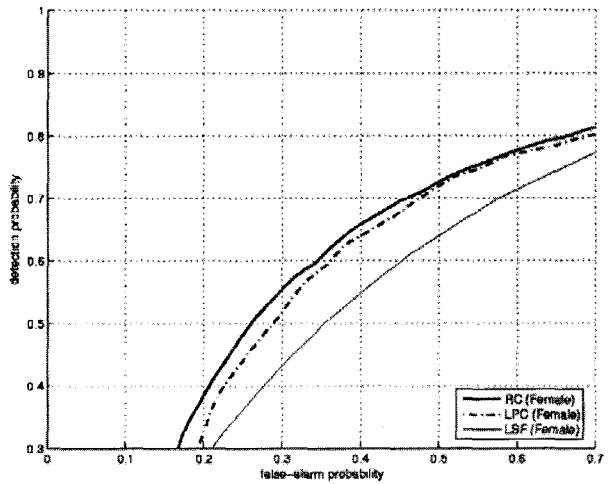
수집되어진 데이터는 모두 8 kHz로 샘플링 하였으며, 16 bit로 양자화 하였다. 보이스 피싱 검출을 위한 특징벡터로 III장에서 기술한 SMV의 복호화 과정에서 자동적으로 추출되어지는 LPC 10차, LSF 10차, RC 10차를 사용하여 특징벡터 구성하였으며, 16개의 혼합성분을 사용한 GMM을 구성하여 실험을 하였다. 그림 4

표 2. 수집된 데이터에 대한 세부 정보
Table 2. Information of collected data.

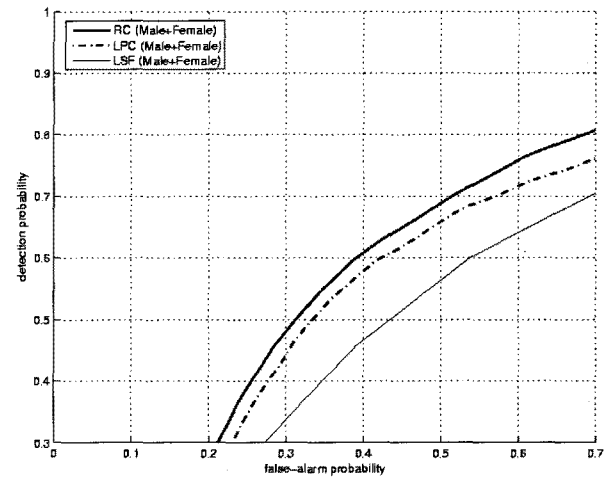
	Training		Test	
	남	녀	남	녀
일반적인 전화통화 목소리	5 명 (약 3분)	3 명 (약 1분 30초)	5 명 (약 3분)	3 명 (약 1분 30초)
보이스 피싱 가해자 목소리	5 명 (약 3분)	3 명 (약 1분 30초)	5 명 (약 3분)	3 명 (약 1분 30초)



(a)



(b)



(c)

그림 4. ROC에 기반한 인식 성능 비교 (a) 남자 (b) 여자 (c) 남자+여자
Fig. 4. The voice pishing detection performance based on ROC. (a) Male (b) Female (c) Male+Female

표 3. 특징벡터에 따른 EER
Table 3. EER according to the feature vectors.

특징벡터	EER		
	Male	Female	M+F
LPC	24.64%	37.74%	41.08%
LSF	29.05%	42.60%	46.77%
RC	23.38%	36.96%	39.59%

와 표 3은 각각 특징벡터의 Receiver Operating Characteristic (ROC) 커브와 Equal Error Rate (EER) 을 보여준다.

인식 결과 기존의 감정인식에서 많이 사용되어졌던 LPC에 비해 남자의 경우 RC가 보다 좋은 인식률을 보였다 [9]. 하지만 여자의 경우 LPC가 좋은 인식률을 보였으며, 남, 녀의 데이터를 혼합하여 인식성능을 확인한 결과 RC가 가장 좋은 성능을 나타내었다.

비교 결과를 보면 전체적으로 남자보다 여자의 경우 인식성능이 떨어지는 것을 확인할 수 있는데 이것은 수집된 데이터가 부족했기 때문이라고 생각된다. 따라서 더 많은 데이터를 포함 한다면 남자의 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 때문에 낮은 인식률을 보일 것이라 예상된다.

가장 인식 성능이 우수한 RC 특징벡터에 인식 성능을 보완하기위해서 후처리 과정에서 추출되어지는 특징 벡터들과 결합하여 새로운 특징벡터를 만들어 실험을 하였다. 실험결과 표 4와 같은 인식 성능을 보였다.

실험 결과 피치와 관련된 특징벡터에서 기존의 RC 성능보다 우수한 성능의 결과를 보였고, 나머지 특징들은 기존의 RC성능에 비해 낮은 성능 인식을 보였다. 또

표 4. 특징벡터에 따른 EER
Table 4. EER according to the feature vectors.

특징벡터	EER		
	Male	Female	M+F
RC + First LSF	22.95%	35.91%	38.89%
RC + LTP Gain	24.38%	35.84%	40.93%
RC + LPC Gain	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 Gain	25.20%	36.71%	39.60%

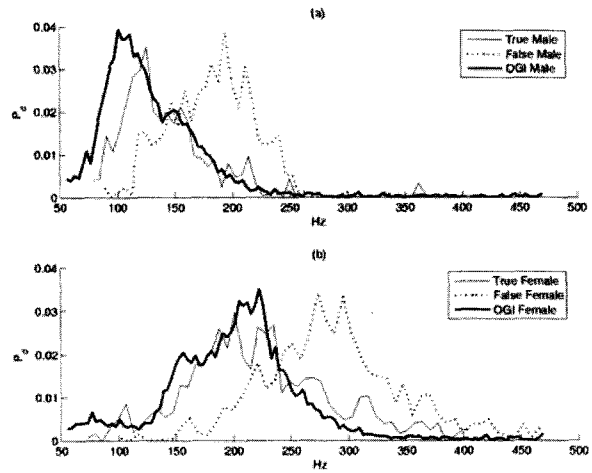


그림 5. 거짓말과 일반적인 전화통화시 피치 지연 분포
Fig. 5. Pitch lag distribution according to general speech and lie speech.

한 RC+LPC, RC+LSF 그리고 RC+LPC+LSF의 경우 역시 RC에 비해 낮은 성능 인식을 보였다.

이 같은 결과는 사람의 감정이 사람의 목소리를 특징 지을 수 있는 성도의 특성을 고려하기 때문이라 생각된다. 특히 피치의 경우 보이스 피싱에 큰 영향을 미치는 요소인데, 이는 Ekman [2]의 연구 결과에도 잘 나와 있듯이 거짓말 할 경우 일반적인 경우에 비해서 높은 기본주파수를 가지기 때문이라 생각된다. 그림 5는 실험에 사용되어진 남, 녀 데이터에 따른 피치 분포와 일반적인 전화통화시 남, 녀의 피치 분포를 보여주고 있다.

제한된 보이스 피싱 검출을 위해 수집한 가해자의 목소리 데이터 (False)와 일반적인 전화목소리 (True) 그리고 OGI 데이터베이스 [16]를 사용하여 추출한 피치의 히스토그램의 결과 일반적인 전화목소리 데이터베이스인 OGI 데이터베이스의 결과와 실제 수집된 일반 목소리와의 거의 비슷한 피치 분포를 볼 수 있으며 가해자의 목소리의 경우 일반적인 목소리에 비해 약 50~100 Hz 정도 높은 피치분포를 보이는 것을 볼 수 있다. 따라서 가장 좋은 성능을 보여주었던 피치와 피치 지연 그리고 LSF의 첫 번째 계수를 RC와 결합하여 최종적인 보이스 피싱 검출 성능을 테스트하였다. 관련하여 그림 6과 표 5는 최종적으로 행해진 실험의 ROC 커브와 EER을 보여주고 있다.

최종적으로 제안된 특징벡터를 이용한 실험 결과 남자의 경우 많은 성능향상을 보이지는 않았지만 여자의 경우 많은 성능향상을 보였고, 남, 녀 데이터를 혼합한 경우에도 우수한 성능을 보였다. SMV의 복호화 과정에서 자동적으로 추출되는 특징벡터 중 가장 우수한 성

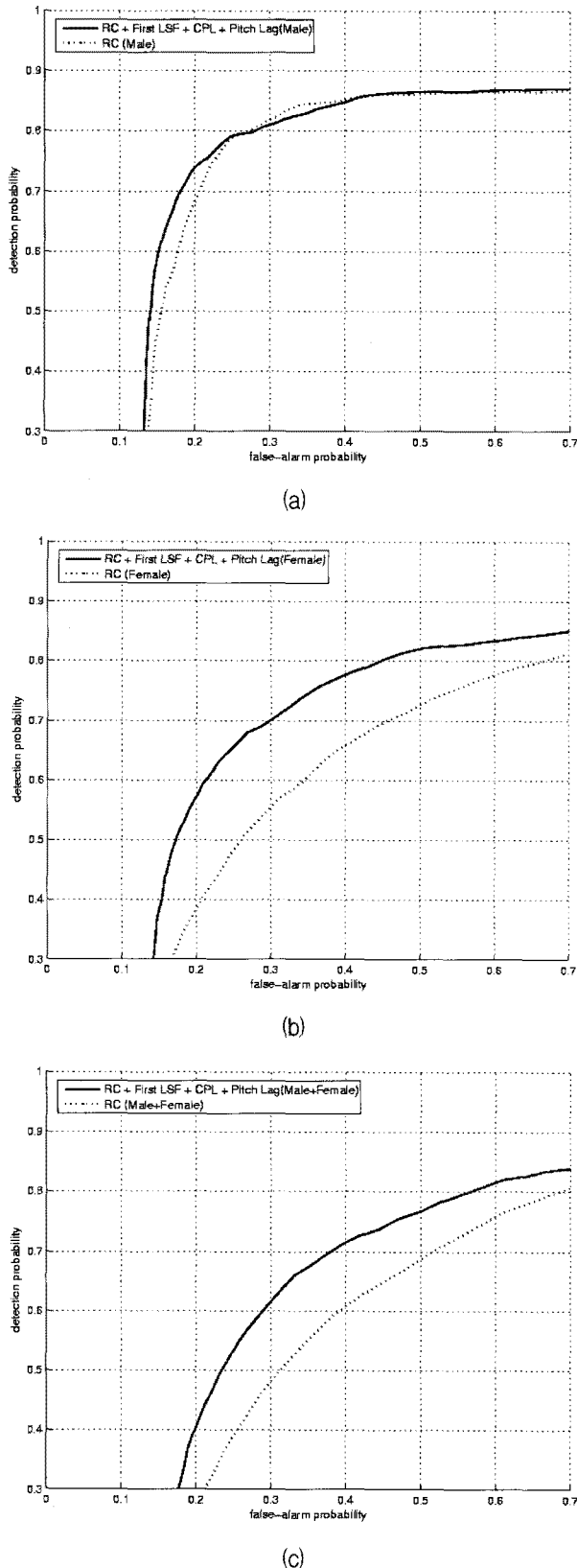


그림 6. 제안된 알고리즘의 ROC 성능비교
 (a) 남자 (b) 여자 (c) 남자+여자
 Fig. 6. The performance of proposed algorithm based on ROC. (a) Male (b) Female (c) Male+Female

표 5. 제안된 알고리즘의 EER
 Table 5. EER according to the proposed algorithm.

특징벡터	EER		
	Male	Female	M+F
RC + First LSF + CPL + Pitch Lag	22.89%	29.95%	33.60%

능을 보인 것을 알 수 있었으며, 피치 정보가 감정과 보이스 피싱 검출에 큰 영향을 주는지 알 수 있었다.

V. 결론

본 논문에서는 전화사기 예방 위한 음성 코딩 파라미터 기반의 보이스 피싱 검출 알고리즘을 제안하였다. 제안된 알고리즘은 사용자의 휴대폰으로 전송되는 데이터를 음성으로 복호화하는 과정에서 자동으로 생성되는 중요 파라미터를 특징벡터 선택한 뒤 GMM을 이용하여 검출을 하였다. 실험 결과 가해자의 목소리에 따른 피치의 특징분포를 알 수 있었으며, SMV의 복호화 과정의 핵심 특징벡터를 사용한 검출이 우수한 성능을 보인 것을 알 수 있었다. 또한 간접적인 방법이 아닌 직접적인 방법을 통해서 데이터를 수집할 수 있다면 그 성능은 더욱 향상 될 것이다. 특히, 휴대폰의 발신자 등록이 되어 있지 않은 상대방과 통화 시에 이용하거나, 차후 공인인증이 되어있지 않은 전화로부터 걸려온 전화 통화에 한정해 이용시 더 강한 성능을 보일 것으로 기대된다. 마지막으로, 보이스 피싱 검출 및 감정 인식 분야에서 보다 적합한 특징벡터와 패턴 인식기에 대한 다양한 시도와 연구가 진행 되어야 할 것이다.

감사의 글

본 연구는 정보통신연구진흥원의 IT 핵심기술개발사업의 일환으로 수행하였음.

참고 문헌

[1] Furedly J. J., Davis C., and Gurevich M., "Differentiation of deception as a psychological process:A psychophysiological approach," *Psychophysiology*, vol. 25, no. 6, pp.683-688, 1988.

[2] Ekman P., Friesen W. V., and Scherer K., "Body movement and voice pitch in deceptive interaction," *Semiotica*, vol. 16, no. 1, pp. 23-27,

- 1976.
- [3] Greer S. C., and Dejacó A., "Standardization of the selectable mode vocoder," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 953-956, 2001.
- [4] Yang G., Shlomot E. B., Thyssen J., Huan-yu S., and Murgia C., "The SMV algorithm selected by TIA and 3GPP2 for CDMA applications," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 709-712, 2001.
- [5] 3GPP2 Spec., "Software distribution for selectable mode vocoder (SMV), service option 56, specification," *3GPP2-C. Roa30-0, v3.0*, 2005.
- [6] Gadallah M. E., Matar M. A., and Alqezawi A. F., "Speech based automatic lie detection," *16th National Radio Science Conference, C22/1-C33/8*, 1999.
- [7] Harry H., and James D. H., "Voice Stress Analyzer Instrumentation Evaluation," *Final Report for Department of Defense Counterintelligence Field Activity Contract-FA 4814-04-0011*, 2006.
- [8] Daniel N., Kjell E., and Kornel L., "Emotion Recognition in spontaneous speech using GMM," *INTERSPEECH*, pp. 809-812, 2006.
- [9] Tsang-Long P., Yu-Te C., and Jun-Heng Y., "Emotion recognition from Mandarin speech signals," *International Symposium on Chinese Spoken Language Processing*, pp. 301-304, 2004.
- [10] Ohmuro H., Moriya T., Mano K., and Miki S., "Coding of LSF parameters using interframe moving average prediction and multi-stage vector quantization," *IEEE Workshop on Speech Coding for Telecommunications*, pp.63-64, 1993.
- [11] Gaunard P., Mubibanqiey C. G., Couvreur C., and Fontaine V., "Automatic classification of environmental noise events by hidden Markov models," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3609-3612, 1998.
- [12] Kabal P., Prakash R., and Ramachandran, "The computation of line spectral frequencies using Chebyshev polynomials," *IEEE trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 6, pp. 1419-1426, 1986.
- [13] Bishop C. M., *Neural networks for pattern recognition*, Oxford University Press, UK, 1995.
- [14] Duda R. O., Hart P. E., and Stork D. G., *Pattern classification*, John Wiley & Sons, INC., 2001.
- [15] Dempster A. P., Land N. M., and Rubin D. B., "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp.1-38, 1977.
- [16] Muthusamy Y. K., Cole R. A., and Oshika B. T., "The OGI multi-language telephone speech corpus," *International Conference on Spoken language Processing*, vol. 2, pp. 895-989, 1992.

 저 자 소 개



이 계 환(학생회원)
 2007년 인하대학교 전자전기
 공학부 학사.
 2007년~현재 인하대학교
 전자공학과 석사과정
 <주관심분야 : 디지털신호처리>



장 준 혁(정회원)
 1998년 경북대학교 전자공학과
 학사.
 2000년 서울대학교 전기공학부
 석사.
 2004년 서울대학교 전기컴퓨터
 공학부 박사.

2000년~2005년 (주)넷더스 연구소장
 2004년~2005년 캘리포니아 주립대학,
 산타바바라(UCSB) 박사후연구원
 2005년 한국과학기술연구원(KIST) 연구원
 2005년~현재 인하대학교 전자공학부 조교수
 <주관심분야 : 음성 신호처리, 오디오 신호처리,
 통신 신호처리, 휴먼/컴퓨터 인터페이스>