

잡음환경에 강인한 음성분류기반의 패킷손실 은닉 알고리즘

Packet Loss Concealment Algorithm Based on Robust Voice Classification in Noise Environment

김형국[†], 류상현

(Hyoung-Gook Kim[†] and Sang-Hyeon Ryu)

광운대학교 전자공학과

(접수일자: 2013년 7월 25일; 채택일자: 2013년 9월 13일)

초 록: 실시간 VoIP 네트워크는 지연, 지터 그리고 패킷손실과 같은 네트워크 장애요소로 인해 품질저하가 발생한다. 본 논문은 VoIP 음질 향상을 위해 잡음환경에 강인한 음성분류기반의 패킷손실 은닉 알고리즘을 제안한다. 제안된 방식에서는 음성신호로부터 추출된 다양한 특징들을 분석하고 이를 기반으로 획득된 적응적인 문턱값을 사용하여 수신단에 도착한 패킷을 분류한다. 정확한 신호분류 결과는 패킷손실 은닉에 사용된다. 그리고 선형 예측기반의 손실패킷 은닉은 연속적으로 패킷을 은닉하거나 손실된 패킷복원 시 발생하는 메탈릭 아티팩트를 제거함으로써 고품질의 음성을 제공한다.

핵심용어: VoIP, 패킷손실 은닉, 음성분류, 적응적인 문턱값 갱신 방법

ABSTRACT: The quality of real-time Voice over Internet Protocol (VoIP) network is affected by network impairments such as delays, jitters, and packet loss. This paper proposes a packet loss concealment algorithm based on voice classification for enhancing VoIP speech quality. In the proposed method, arriving packets are classified by an adaptive thresholding approach based on the analysis of multiple features of short signal segments. The excellent classification results are used in the packet loss concealment. Additionally, linear prediction-based packet loss concealment delivers high voice quality by alleviating the metallic artifacts due to concealing consecutive packet loss or recovering lost packet.

Keywords: VoIP, Packet loss concealment, Voice classification, Adaptive thresholding approach

PACS numbers: 43.72.Ja, 43.72.Kb, 43.60.Dh

1. 서 론

VoIP는 최근 음성통신 분야에서 급부상하고 있는 기술이다. 하지만 VoIP는 패킷손실, 패킷 지연 그리고 지터와 같은 장애 요인 등으로 인해 양질의 서비스를 제공하지 못하고 있다. 따라서 현재 지연을 줄이고, 지터를 일정하게 유지하며, 손실을 복원시키기 위한 연구들이 이루어지고 있다.^[1-2]

패킷손실을 은닉하거나 손실된 패킷을 복원하기 위해 다양한 방식들이 연구되고 있으며, 크게 두 가

지로 분류할 수 있다. 첫 번째로는 손실된 프레임들을 이전에 수신된 프레임의 피치 복제를 이용하거나 패턴 매칭을 이용하여 교환하는 “파형교환”방식이고, 두 번째는 이전에 수신된 신호들의 모델 파라미터를 추정하거나 보간하여 손실된 신호 부분을 복원하는 “모델 기반의 반복”방식이다.^[3]

모델기반의 패킷손실 은닉 알고리즘은 파형교환 방식에 비해 네트워크를 통해 패킷단위로 전달되어 재구성된 음성의 질을 효과적으로 향상시킨다. 하지만 연속적으로 발생한 패킷손실에는 피치 추정 오류가 발생하여 사용자에게 자연스러운 음성을 제공하는데 한계가 있다. 따라서 사용자에게 자연스러운 음성을 제공하기 위해 정확한 피치 추정을 이용하여

[†]Corresponding author: Hyoung-Gook Kim (hkim@kw.ac.kr)
Department of Wireless Communication Engineering, Kwangwoon University Room 405, Chambit Building, 447-1, Wolgye-Dong, Nowon-Gu, Seoul 139-701, Republic of Korea
(Tel: 82-2-940-5574, Fax: 82-2-913-5006)

모델 기반으로 자연스럽게 합성된 음성을 생성하는 과정과 손실구간동안 서서히 음성을 줄이는 과정이 필요하다.

본 논문에서는 상기 위에서 언급한 문제들을 해결하고 수신단 기반의 VoIP 음성통화품질을 향상시키기 위해서 잡음환경에 강인한 음성분류기반의 패킷손실 은닉 알고리즘을 제안한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 제안된 방법에 대해서 설명하고, 3장에서는 제안된 방법의 성능을 확인하기 위한 실험의 결과를 설명한다. 그리고 마지막 4장에서 결론을 서술한다.

II. VoIP 수신단 기반의 패킷손실 은닉 알고리즘

VoIP 수신단의 전체 신호처리 과정은 우선적으로 i 번째 패킷이 IP 네트워크로부터 VoIP 수신단에 도착하면, 수신단에서는 패킷정보를 추출하고 지터버퍼에 패킷을 저장한다. 시스템의 지연을 줄이기 위해서 지터 버퍼는 일정한 시간 간격(10ms단위)으로 패킷들을 디코딩하여 출력한다. 지터 버퍼로부터 디코딩된 각 음성프레임은 음성프레임 버퍼에 저장되고, 저장된 음성프레임은 음성과 비음성 그리고 묵음 또는 배경음으로 분류된다. 음성프레임을 분류한 결과는 패킷손실 은닉에 사용된다. 출력을 위한 연속적인 음성프레임 가운데, i 번째 음성프레임이 음성프레임 버퍼에 존재하지 않는다면, 패킷이 손실된 것으로 인식하고 손실 은닉이 수행된다. 패킷손실 은닉부는 손실된 패킷을 복원하기 위해서 디코더에 손실 이후의 음성프레임을 요청한다. 패킷손실 은닉을 수행한 후에는 복원된 음성프레임과 원래 음성프레임간의 불연속점을 제거하기 위해서 병합 및 스무딩을 수행한다. 디지털-아날로그 컨버터는 일정한 시간 간격(20ms)으로 프로세싱된 음성프레임을 아날로그 신호로 변환하여 사용자의 스피커로 전달한다.

2.1 적응적 음성분류기

음성분류 시 발생하는 문제를 해결하기 위해 많은 알고리즘들이 제안되었다. 일반적인 방식은 음성신

호로부터 특징값을 추출하고 미리 결정된 문턱값보다 크거나 작은 특징값들을 이용하여 유성음과 무성음을 분류하는 방식이다. 그리고 이전 연구에서는 Hurst 상수를 이용한 통계적인 분류 방식의 선형적인 모델을 제안되었다. 하지만 이런 통계적인 모델 기반의 유성음/무성음 분류 방식은 다량의 학습 데이터와 연산량을 요구하기 때문에 실질적인 VoIP 어플리케이션에 적용하기에는 적합하지 않다.

본 논문의 제안된 방식에서는 단구간 에너지(STE, Short Time Energy), 자기상관도(AC, Auto-Correlation), 그리고 ZCR(Zero Crossing Rate)과 같은 특징값들을 기반으로 변동되는 잡음환경에서도 음성 및 비음성 구간을 구분할 수 있는 임계값을 자동으로 갱신시킴으로써 유성음, 무성음, 묵음 분류의 정확도를 높인다.

Fig. 1은 제안된 음성 분류기의 전체 흐름도이다.

(첫 번째 단계) 20ms 음성 프레임에 해밍 윈도우를 적용하여 STE를 계산한다. 이와 동시에 AC를 계산한다.

(두 번째 단계) STE 문턱값 θ_{start} 는 p 개 프레임 이내의 값으로 계산되고($0 < p < 7$), 이 값은 아래 식에 적용하여 음성/비음성구간을 구분하는데 사용된다.

$$S_{i,k} = \begin{cases} \text{비음성구간} & \text{if } STE_{i,k} \leq \theta_{start} \\ \text{음성구간} & \text{otherwise} \end{cases}, \quad (1)$$

$S_{i,k}$ 는 음성프레임 버퍼에 저장된 i 번째 음성프레임이다.(i 번째 음성구간에서 k 번째 도착 패킷) 이와 동시에 AC의 문턱값 ψ_{start} 가 계산되어 아래 식 2에

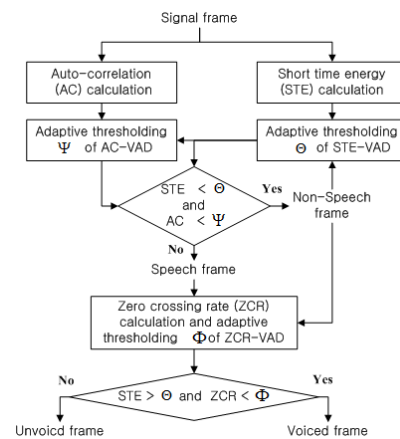


Fig. 1. Overall flow chart of the voice classification.

적용하여 음성/비음성구간을 구분하는데 사용된다.

$$S_{i,k} = \begin{cases} \text{비음성구간} & , \text{ if } AC_{i,k} \leq \Psi_{start} \\ \text{음성구간} & , \text{ otherwise} \end{cases} \quad (2)$$

(세 번째 단계) 유성음/무성음/배경음 신호를 분류하기 위해 Fig. 2와 같은 방법으로 비음성구간 동안 STE의 문턱값 $\theta_{i,k}$, AC의 문턱값 $\Psi_{i,k}$, 그리고 ZCR의 문턱값 $\Phi_{i,k}$ 를 갱신한다.

Fig. 2에서 $E_{i,k}$ 는 비음성구간 음성프레임의 에너지값, cnt_{up} 는 에너지 상승구간의 수, 그리고 cnt_{down} 는 에너지 하강구간의 수이다. 그리고 각 파라미터가 실제로 적용되는 수치의 범위는 $\alpha_b(0 < \alpha_b < 1)$, $\gamma(0 < \gamma < 1)$, $\alpha_{up}(> 3)$, $\nu(0 < \nu < 1)$, $N_{up}(> 2)$, $N_{down}(> 2)$, $\alpha_{down}(> 1)$ 이다.

(네 번째 단계) $STE_{i,k}$ 가 $\theta_{i,k}$ 보다 작고 $AC_{i,k}$ 가 $\Psi_{i,k}$ 보다 작다면, 그때 $S_{i,k}$ 는 비음성구간(목음 또는 배경잡음)으로 결정된다. 반대 경우, $S_{i,k}$ 는 음성구간으로 결정된다.

(다섯 번째 단계) $S_{i,k}$ 가 음성구간으로 결정되어지고 $STE_{i,k}$ 가 $\theta_{i,k}$ 보다 크고 $ZCR_{i,k}$ 가 $\Phi_{i,k}$ 보다 작다면, $S_{i,k}$ 는 유성음으로 구분된다. 이와 반대의 경우에는 $S_{i,k}$ 는 무성음으로 구분된다.

음성분류 결과는 음성품질의 향상을 위해 제안된 패킷손실 은닉알고리즘에서 효과적으로 사용된다.

$$\begin{aligned} & \text{If}(S_{i,k} = \text{비음성구간}) \\ & E_{i,k} = \alpha_b \cdot E_{i,k-1} + (1 - \alpha_b) \cdot STE_{i,k} \\ & \text{If}(E_{i,k} > E_{i,k-1}) \\ & \theta_{i,k} = \begin{cases} \gamma \cdot \theta_{i,k-1} + (1 - \gamma) \cdot \alpha_{up} \cdot E_{i,k} & , \text{ if } cnt_{up} > N_{up} \\ \theta_{i,k} & , \text{ otherwise} \end{cases} \\ & \Psi_{i,k} = \nu \cdot \Psi_{i,k-1} + (1 - \nu) \cdot \theta_{i,k} \\ & \Phi_{i,k} = \nu \cdot \Phi_{i,k-1} + (1 - \nu) \cdot \theta_{i,k} \\ & \text{else} \\ & \theta_{i,k} = \begin{cases} \gamma \cdot \theta_{i,k-1} + (1 - \gamma) \cdot \alpha_{down} \cdot E_{i,k} & , \text{ if } cnt_{down} > N_{down} \\ \theta_{i,k-1} & , \text{ otherwise} \end{cases} \\ & \Psi_{i,k} = \nu \cdot \Psi_{i,k-1} + (1 - \nu) \cdot \theta_{i,k} \\ & \Phi_{i,k} = \nu \cdot \Phi_{i,k-1} + (1 - \nu) \cdot \theta_{i,k} \end{aligned}$$

Fig. 2. Method of adaptive thresholding.

2.2 패킷손실 은닉 알고리즘

손실된 패킷들을 복원하는 기본 이론은 이전과 이후 프레임의 피치주기 정보를 사용하여 원 신호와 가장 밀접한 피치주기를 갖는 신호를 생성하는 것이다. 이 알고리즘은 선형 예측 분석에 의해 수신된 이전 음성의 잔여 신호를 추출하고, 주기적인 반복을 이용하여 손실된 음성의 여기신호 근사치를 생성하고, 이 여기 신호를 사용하여 합성된 음성을 생성하는 것이다.

제안된 패킷손실 은닉 알고리즘(PLC, Packet Loss Concealment)은 G.722의 패킷손실 은닉 알고리즘의 성능을 향상시키기 위해서 회귀적인 LP(Linear Prediction) 분석과 합성(LPAS, LP Analysis and Synthesis)을 기반으로 하여 추정된 피치 주기를 사용한다. 제안된 회귀적인 LPAS 알고리즘은 단구간 및 장구간 PLC로 구분되어 수행된다. 단구간 PLC에서는 AC로부터 추정된 피치 주기를 통해 손실 패킷들의 복원을 위한 스무딩한 여기신호를 생성하고, 짧은 구간 패킷손실 구간 복원을 위해서 작은 음성 구간을 반복하였을 때 생성될 수 있는 부자연스러운 인공음을 효과적으로 줄인다. 만일 연속적인 패킷들이 손실된다면, 장구간 PLC를 통해 이전 합성된 신호를 LP필터에 회귀적으로 입력하여 새로운 여기 신호를 생성한다. 이렇게 생성된 여기 신호를 필터링하여 재구

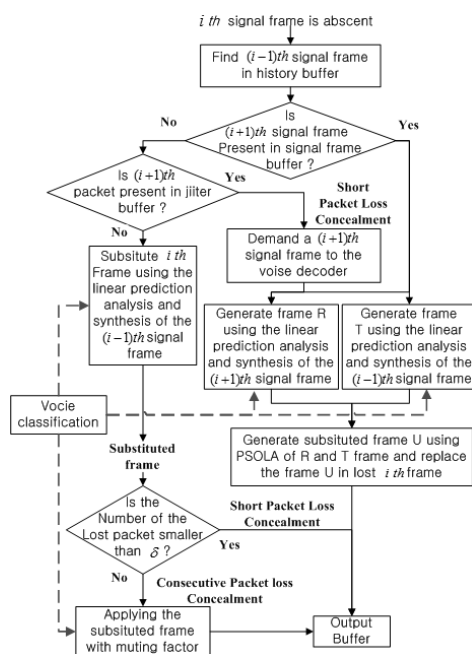


Fig. 3. Overall flow chart of packet loss concealment.

성된 신호로 합성하고 복원된 구간동안 점차적으로 소리를 줄임으로써 사용자 청각에 불편함을 주는 금속음을 자연스럽게 제거한다.

Fig 3은 패킷손실 은닉 알고리즘의 전체적인 블록도이며, 단 구간과 장구간 패킷 손실 은닉의 두 부분으로 구성된다.

손실된 패킷의 수가 $\delta(\geq 1)$ 보다 작다면, 추정된 피치 주기와 같은 주기성을 가진 신호를 이전에 수신된 신호에서 찾아 그 부분을 반복하여 합성된 신호를 생성하는 단 구간 패킷손실 은닉을 수행한다. 예를 들어, 손실패킷의 재구성에 이용할 수 있는 $i+1$ 번째 음성프레임이 없다면($i+1$ 번째 입력될 패킷 또는 $i+1$ 번째 음성프레임이 음성프레임 버퍼와 지터 버퍼에서 발견되지 않을 경우), 제안된 알고리즘은 히스토리 버퍼에 저장되어있는 이전 $i-1$ 번째 음성프레임을 이용하여 LPAS 방식을 통해 손실된 패킷을 은닉한다. 반면에 $i+1$ 번째 음성 프레임이 이용가능하다면, 제안된 알고리즘에서는 LPAS를 통해 이전 $i-1$ 번째 음성 프레임으로부터 음성프레임 T 를 생성하고, $i+1$ 번째 음성프레임을 이용하여 음성프레임 R 을 생성한다. 음성프레임 R 과 음성프레임 T 사이에서 PAOLA(Peak Alignment OverLap-Add) 방식을 적용함으로써 부드럽고 자연스러운 음성 프레임 U 를 생성하여 손실된 패킷 구간에 대체한다.

장구간 PLC에서는 회귀적 LPAS 알고리즘을 적용하여 생성된 여기 신호를 반복함으로써 연속적으로 손실된 패킷을 대체한다. 복원된 대체신호는 식(3)을 이용하여 뮤팅팩터에 따라 샘플단위로 신호의 크기를 줄여 금속성 인공음을 완화시키고 음성 품질향상을 위해 손실구간을 단위로 신호의 크기를 선형적으로 줄여준다.

$$\begin{aligned} &\text{if } S_{i,k} = \text{유성음} \\ & y_{i,k}(n) = \widehat{s}_{i,k}(n) \cdot \max(\lambda_j(n-1) - \Delta_V) \\ &\text{else} \\ & y_{i,k}(n) = \widehat{s}_{i,k}(n) \cdot \max(\lambda_j(n-1) - \Delta_{LV}). \end{aligned} \quad (3)$$

위의 식에서 음성구간에서 신호를 줄이는 뮤팅팩터 $\Delta_V = \zeta + \zeta \cdot \sin(2\pi a/b)$, ($0 < a < 1$, $0 < b < 1$, $0.001 < \zeta < 0.01$)와 비음성구간에 신호를 줄이는 뮤팅팩터 Δ_{LV} ($0.01 < \Delta_{LV} < 0.1$)는 음성 분류에 의존

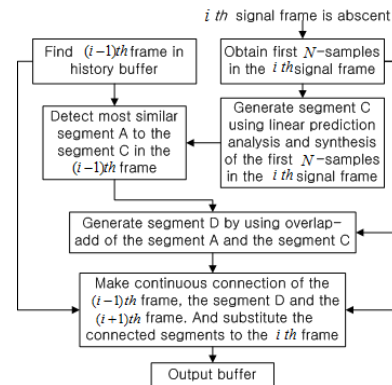


Fig. 4. Overall flow chart of merge and smoothing method.

적이며 샘플단위로 갱신한다.

2.3 병합 및 스무딩 방법

병합부는 음성프레임에서부터 대체프레임까지 그리고 대체프레임에서부터 이어지는 음성프레임까지의 천이영역에서 두 신호들 사이의 스무딩한 보간을 실시한다. 제안된 병합 및 스무딩 방법의 전체 흐름도는 Fig. 4에 설명되어 있다.

병합 및 스무딩 방법에 의한 스무딩한 보간은 아래와 같이 수행된다.

(첫 번째 단계) 우선 i 번째 음성프레임에서 N 개 샘플들을 획득하고 참조 세그먼트 C 를 생성하기 위해 LPAS로 입력한다.

(두 번째 단계) 참조 세그먼트 C 와 가장 유사한 음성구간 A 를 히스토리 버퍼의 이전 $i-1$ 번째 음성프레임의 샘플들에서 찾는다.

(세 번째 단계) 추정된 음성 프레임 D 는 음성 프레임 A 와 음성프레임 C 사이에서 PAOLA사용하여 생성한다.

(네 번째 단계) $i-1$ 번째 음성 프레임 D 와 $i+1$ 번째 음성 프레임을 병합하여 새로운 구간 O 를 생성한다. 음성 프레임 D 는 $i-1$ 번째 음성 프레임과 $i+1$ 번째 음성 프레임의 신호 왜곡을 방지하는데 사용된다. 세그먼트 O 는 i 번째 음성 프레임으로 대체된다.

III. 실험결과

제안된 패킷손실 은닉 알고리즘의 성능을 측정하기 이전에 제안된 음성 분류기를 클린과 노이즈 환

경에서 기존의 일반적인 음성분류방식들과 비교하였다. 실험군은 각각 다섯 명의 남성과 여성의 화자들이 말한 문장을 16 kHz로 샘플링 하였고, 10분 길이의 한국어 문장을 사용하였으며, 세 가지 다른 유형의 잡음(배틀잡음, 자동차잡음, 백색잡음)을 clean 과 SNR 20 dB에서 5 dB까지 5 dB단위로 감소하는 잡음강도로 각각의 실험군을 구성하였다. 그리고 제안된 방식의 성능을 측정하기 위해 아래와 같은 세 가지 측정방식을 사용하였다.^[4] (1) AVDE(The Averaged Voicing Decision Error)는 결정된 음성구간이 실질적으로 음성구간이 아닌 프레임의 비율을 나타낸다. (2) AGPE(The Averaged Gross Pitch Error)는 피치 추정부와 F0의 상대적인 에러가 문턱값의 20%보다 높은 기본 참 값들의 결정이 유성음인 프레임의 비율이다. (3) AFPE(The Averaged Fine Pitch Error)는 상대적인 에러, F0의 분포의 표준 편차로 정의된다. 상대적인 에러 F0는 문턱 값의 20%보다 아래인 값을 의미한다.

Table 1은 클린 음성과 세 가지 노이즈 상태에 대한 피치 트래킹 결과의 평균값을 나타낸다.

제안된 음성 분류기를 기존의 음성분류방식인 SRH(Summation of Residual Harmonics), 칼만 방식과 비교한 결과, AVDE, AGPE 그리고 AFPE가 모두 기존의 방식보다 좋은 성능을 보였다,

다음으로 제안된 패킷손실 은닉 알고리즘을 평가하기 위해 세 개의 모듈[SIP(Session Initiation Protocol) 신호, 음성 데이터 전달 그리고 네트워크 트래픽 에뮬레이터]로 구성된 테스트베드를 구축하였다.^[7] 구축된 테스트베드에서 SIP 신호는 SIP프록시 서버와 두 개의 클라이언트를 포함하고 있으며, C++를 이용하여 개발된 VoIP 어플리케이션을 클라이언트의 모바일 디바이스에 구현하였다. SIP 신호 메시지들은 처음에 클라이언트로부터 SIP 프록시 서버로 전송되며, 그 후 클라이언트로 다시 전송된다. 음성 데이

Table 1. Pitch tracking results.

Method	Perposed method	SRH ^[5]	Kalman ^[6]
AVDE	3.57	11.21	6.56
AGPE	2.53	2.64	3.27
AFPE	2.73	2.98	3.36

터 전송 모듈에서, 클라이언트 A와 B들은 각각 접속 지점(접속지점 1, 2)의 허브 1과 2로 연결되고, 네트워크 트래픽 에뮬레이터 모듈에서 트래픽 생성부는 다른 네트워크 망들과 WLAN의 연결 지연, 지터, 패킷손실을 시뮬레이트 하기 위해 사용되었다. 트래픽 클라이언트는 접속 지점1을 통해 RTP 패킷들을 수신하고, 접근 지점 2로 전송하는 구조를 갖는다. 사용된 음성 샘플들은 각각 5명의 남성과 여성 화자들에 의해 발생되어 5분 동안 통화하는 16 kHz로 샘플링된 음성신호로서, 15,000개의 패킷들로 구성되었다.

위의 테스트베드를 이용하여 총 3가지의 실험군을 구성하였으며, 네트워크 환경에 대한 정보는 다음 Table 2와 같다.

음질저하 정도를 측정하기 위해 PESQ를 이용한 객관적인 음성 품질 실험, 총 버퍼 지연(TBE), 지터 추정 에러(JER), 그리고 지연 패킷손실(LPL)을 측정하였다. PESQ는 ITU-T에서 지정한 객관적인 음질평가 지표이며, 1에서부터 5까지의 점수를 제공한다.^[8] 여기서 1은 들을 수 없을 정도로 낮은 품질을 의미하고 5는 매우 좋은 품질을 의미한다.

제안된 방식(PM)의 성능을 비교하기 위해서 논문을 기반으로 구현한 두 가지 방식과 비교하였다. M1는 지연 스파이크 검출^[1]과 손실패킷 은닉^[3]을 이용

Table 2. Statistics of network packet loss.

Trace	Average of network delay(ms)	Variance of network delay(ms)	Maximum jitter(ms)	Network packet loss (%)
1	25.38	7.46	48	1.93
2	34.80	13.37	152	3.97
3	79.98	31.22	371	3.97

Table 3.The performance of the proposed algorithm.

Trace	Method	TBE(ms)	JER(ms)	LPL(%)	PESQ
1	M1	42.54	16.66	0.85	2.912
	M2	35.30	22.31	0.10	3.256
	PM	27.76	13.25	0.04	3.523
2	M1	29.35	29.35	3.18	2.156
	M2	28.42	28.42	1.99	2.626
	PM	20.45	20.45	0.78	3.156
3	M1	44.09	77.61	8.61	1.837
	M2	77.85	77.45	7.58	2.351
	PM	61.43	68.18	2.67	2.835

한 적응적인 최소 평균 제곱 재생 알고리즘 방식이고, M2는 K-Erlang 분표를 따르는 지터 추정방법^[9]을 사용한 음성출력제어와 손실은닉알고리즘을 사용한 방식이다. 시뮬레이션은 광대역에서 G.722.2 음성 코덱을 사용하였다.

Table 3은 PM이 중간 지터, 높은 지터, 2% 패킷손실, 그리고 4% 패킷 손실에서 비교 방식 M1, M2보다 더 좋은 성능을 보임을 나타낸다. Table 3에 나타난 바와 같이 PM은 실험군 1에서 사용자가 이용할 수 있는 PESQ 스코어(>3.5)를 획득하였으며, 지터 정도와 패킷손실 비율이 증가할 때, PEQS 스코어는 감소하는 양상을 보였지만 참조 모델들에 비해 스코어 차이가 더 안정적인 모습을 나타내고 있다. PM은 입력된 패킷들 음성분류결과에 따라 손실복원 방법을 다르게 적용하여 기존보다 더 좋은 음질의 출력신호를 생성한다.

이에 따라 PM은 네트워크 상태의 급격한 변화와 다양한 손실 패턴 환경에서 버퍼링 지연을 최소화하는데 매우 적합한 방식이라고 판단된다.

IV. 결 론

본 논문에서는 잡음환경에 강인한 음성분류기반의 패킷손실 은닉 알고리즘을 제안하고, 성능을 측정하였다. 잡음환경에 강인한 음성분류의 결과로 획득되어지는 피치 추정은 IP 네트워크로 전달되는 음성의 품질을 향상하기 위해 제안된 패킷손실 은닉에 효과적으로 사용되었다. 그리고 기존의 방식들과 비교하였을 때, 제안된 방식은 더 적은 연산량으로 사용자에게 더 높은 음성품질을 제공하였다. 또한 실험 결과는 수신단 기반의 향상된 알고리즘이 실질적인 모바일 VoIP 어플리케이션들에서 사용될 수 있다는 것을 입증하였다.

감사의 글

이 논문은 2013년도 정보(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업지원을 받아 수행된 것임(NRF-2013R1A1A2007601).

References

1. A. Shallwani and P. Kabal, "An adaptive playout algorithm with delay spike detection for real-time VoIP," in Proc. of IEEE CCECE 2003, **2**, 997-1000 (2003).
2. M. J. Kim, and C. H. Kwon, "Dynamic Redundant Audio transmission for Packet Loss Recovery in VoIP Systems," (in korean) J. Acoust. Soc. Kr. **21**, 349-360 (2002).
3. H. Sanneck, A. Stenger, K. B. Younes, and B. Girod, "A new technique for audio packet loss concealment," in Proc. of GLOBECOM'96, 48-52, (1996).
4. W. Chu and A. Alwan, "Reducing f0 frame error of f0 tracking algorithms under noisy conditions with an unvoiced/voiced classification frontend," in Proc. of IEEE ICASSP, 4769-4772, (2009).
5. C. Shahnaz, W.-P. Zhu, and M. O. Ahmad, "Pitch estimation based on a harmonic sinusoidal autocorrelation model and a time-domain matching scheme," IEEE Trans. on ASLP. **20**, 1, 322-335 (2012).
6. M. G. Christensen, "A method for low-delay pitch tracking and smoothing," in Proc. of ICASSP, 345-348 (2012).
7. S. L. Ng, S. Hoh, and D. Singh, "Effectiveness of adaptive codec switching VoIP application over heterogeneous networks," 2nd Int'l Conf. on Mobile Technology, Applications and Systems, 7-13 (2005).
8. ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," ITU (2001).
9. H. Li, G. Zhang, and W. Kleijn, "Adaptive playout scheduling for VoIP using the K-Erlang distribution," in Proc. of EUSIPCO, 1494-1498 (2010).

저자 약력

▶ 김 형 국(Hyung-Gook Kim)



2002년 7월: 독일 SIEMENS/Cortologic AG
책임연구원
2005년 3월: 독일 베를린 공과대학교
Assistant Professor
2007년 2월: 삼성종합기술원 수석연구원
2007년 3월 ~ 현재: 광운대학교 전파공
학과 부교수

▶ 류 상 현(Sang-Hyeon Ryu)



2012년 2월: 광운대학교 전파공학과 학사
2012년 3월 ~ 현재: 광운대학교 전파공학
과 석사과정