한국음향학회지 제32권 제2호 pp. 157~166 (2013)
The Journal of the Acoustical Society of Korea Vol.32, No.2 (2013)
http://dx.doi.org/10.7776/ASK.2013.32.2.157

pISSN : 1225-4428
eISSN : 2287-3775

G.729 코덱의 패킷 손실 영향 모델을 이용한 비 침입적 음질 예측 기법

Non-Intrusive Speech Quality Estimation of G.729 Codec using a Packet Loss Effect Model

이민기†, 강홍구

(Min-Ki Lee† and Hong-Goo Kang)

연세대학교 전기전자공학과 디지털 신호처리 연구실

(접수일자: 2012년 11월 21일; 수정일자: 2013년 1월 10일; 채택일자: 2013년 1월 21일)

초 록: 본 논문은 패킷 손실의 영향을 이용한 비 침입적 음질 평가 방법을 제안한다. 패킷 손실은 패킷 기반의 통신 시스템에서 음질을 저하시키는 주된 요소이며 그 영향은 코덱에 내장된 패킷 손실 은닉 알고리듬에 의해 결정된다. 패킷 손실 영향을 반영한 음질평가 시스템을 위해 VoIP 에서 협대역 코덱으로 사용되는 코덱 중 하나인 G.729를 선택하였으며, 음성 특징에 따른 패킷 손실 영향을 구분하기 위해서 G.729 코덱의 음성 파라미터를 이용한 한 음성 특성 분류기를 설계하였다. 이후, 각각의 패킷 특성에 따른 음질 저하의 정도를 수치화하기 위해 원 PESQ-LQ 점수와 상관계수를 최대화하는 음질 저하 가중치를 반복적으로 구하였으며, 최종 음질 저하는 가중합으로 구하였다. 그 결과 제안한 모델과 PESQ-LQ의 상관계수는 칩입 모델 에서는 0.8950를, 비 침입 모델 에서는 0.8911의 결과를 나타내었다.
핵심용어: 비 침입적, 음질평가, 음질예측, 패킷손실은닉, G.729, PESQ, VoIP

ABSTRACT: This paper proposes a non-intrusive speech quality estimation method considering the effects of packet loss to perceptual quality. Packet loss is a major reason of quality degradation in a packet based speech communications network, whose effects are different according to the input speech characteristics or the performance of the embedded packet loss concealment (PLC) algorithm. For the quality estimation system that involves packet loss effects, we first observe the packet loss of G.729 codec which is one of narrowband codec in VoIP system. In order to quantify the lost packet affects, we design a classification algorithm only using speech parameters of G.729 decoder. Then, the degradation values of each class are iteratively selected that maximizes the correlation with the degradation PESQ-LQ scores, and total quality degradation is modeled by the weighted sum. From analyzing the correlation measures, we obtained correlation values of 0.8950 for the intrusive model and 0.8911 for the non-intrusive method.
Key words: Speech quality estimation, Packet loss concealment, G.729, PESQ, VoIP
PACS numbers: 43.72.-p, 43.72.Kb

I. 서 론

VoIP(Voice over Internet Protocol)을 이용한 음성통신 시스템은 기존의 PSTN(Public Switched Telephony Network) 대비 상대적으로 저렴한 IP 네트워크를 이용하기 때문에 기업들의 통화 비용 절감 요구와 맞물려 기존의 음성 통신 환경을 대체할 차세대 기술로 각광받아 왔다. 하지만 비순차 패킷을 가지는 IP 네트워크의 구조로 인하여 채널 열화가 발생하게 되면 실시간 순차 전송 시스템을 가지는 VoIP 에서 패킷의 지연(delay), 지터(jitter)가 발생하게 되고, 이는 패킷 손실을 야기하여 QoS(Quality of Service)를 보장하지 못하는 원인이 된다. 이로부터 VoIP 서비스가 성공적으로 이루어지기 위해서는 시스템을 모니터

†**Corresponding author:** Min-Ki Lee (minikey@dsp.yonsei.ac.kr)
2nd E.E. Bldg. B601, Sinchon-dong 134, Yonsei University, Seodaemun-gu, Seoul 120-749, Republic of Korea
(Tel: 82-2-2123-4534, Fax: 82-2-364-4870)

링 하여 음질 저하를 최소화 할 수 있도록 시스템을 적극적으로 보완해야 하며, 이 때 적절한 시스템 평가 방법이 필요하게 된다.

통신 시스템의 음질 평가 방법으로는 크게 전송단과 수신단의 양 단 신호를 직접 비교하는 침입적 방식(intrusive method), 수신단의 신호만으로 음질을 평가하는 비 침입적 방식(non-intrusive method), 그리고 각 시스템의 통신 파라미터를 취합해서 음질을 평가하는 파라메트릭 방식(parametric method) 의 세 가지로 구분할 수 있으며,[1] ITU(International Telecommuniation Union)는 침입적 방식의 P.862[2]와 P.863,[3] 비침입적 방식의 P.563[4] 그리고 파라메트릭 방식의 E-모델(E-model)[5]을 표준화하여 제안하고 있다.

파라메트릭 방식인 E 모델은 시스템 설계에 반영하기 위한 목적으로 만들어졌으며, 음질 저하에 영향을 미치는 모든 요소를 고려하여 실제 시스템의 각종 파라미터에 적용해서 0점부터 100 사이의 등급 요소(rating factor), R을 산출한다. 이는 음질평가에 사용되는 1점부터 5점 사이의 MOS 스케일로 변환할 수 있지만 정해진 각종 장비 및 통신 환경에 따른 고정된 파라미터를 사용하기 때문에 모니터링 용도로는 사용하지 않는다.

E 모델을 구성하는 파라미터 중에서 장비 손상 요소(equipment impairments) 에 해당하는 패킷 손실은 손실된 패킷의 특성에 따라 음질 저하에 미치는 영향이 달라지는 특징이 있다. 예컨대 손실된 패킷이 묵음이나 무성음의 경우에는 음질에 미치는 영향이 거의 없거나 적은 반면 유성음의 경우에는 체감 음질이 크게 저하되며,[6] 같은 유성음 구간이더라도 음성의 시작점인 온셋(onset)에 발생한 패킷 손실이 음성의 중간이나 끝 부분에 발생한 경우보다 음질의 저하가 매우 크다.[7]

Ding(2007)[8]은 이와 같이 패킷 손실률이 음질 저하에 미치는 요소를 세분화하여 음질 평가에 반영한 비 침입적 음질평가 시스템을 제안하였다. 정적인 E-모델로부터 시변 특성을 가진 패킷 손실, 노이즈 추정 및 템포럴 클리핑(temporal clipping) 등을 고려하였으며, 패킷 손실이 음질에 미치는 영향을 계산하기 위해 손실된 패킷의 정보를 허미트 다항식(hermite polynomial)을 이용하여 보간(interpolation)하고 이를 토대로 한 S/U/V 분류기를 통해 패킷 손실에 의한 영향을 세분화 하였다. Lee(2007)[9]는 패킷의 S/U/V 분류에서 보다 확장하여 SMV(Selectable Mode Vocoder)[10]의 6개 클래스 정보를 사용하였으며 이전 프레임과 이후 프레임까지 고려하게 되면 단순히 패킷 손실률을 쓴 것 보다 정확한 음질 추정이 가능하지만 손실된 패킷의 클래스 정보를 정확하게 알아야 하는 제약이 있었다.

본 논문은 VoIP에서 주로 사용되는 협대역 코덱 중 하나인 G.729[11] 코덱을 대상으로 한 패킷 손실 환경에서의 파라메트릭 기반의 비침입적 음질 평가 방식을 제안하였다. 패킷 손실의 영향은 패킷의 특성에 의존한다는 가정 하에 G.729 코딩 파라메터를 이용한 패킷 특성 분류기를 제안하였으며, 이를 기반으로 이전 프레임과 이후 프레임의 패킷까지 동시에 고려한 패킷의 전이 특성 분류 작업을 수행하였다. 이후 패킷 손실의 영향을 모델링하는 과정을 설명하고, 비침입적 방식을 위한 수신단 기반의 패킷 특성 분류를 위해 보간법을 사용하여 손실된 패킷의 클래스를 추정하였다. 이를 토대로 PESQ-LQ[12]의 음성품질 측정 결과를 추정하는 비 침입 모델을 제안하고, 원 패킷 정보를 이용하는 침입 모델과 ITU의 비 침입적 음질평가 방식인 P.563[4]의 결과와 비교하였다.

II. G.729 코딩 파라미터 기반의 음성 특성 분류기

본 장에서는 음성 신호의 특성에 따라 패킷 손실의 영향을 관찰하기 위한 G.729 코덱의 복호화기 기반 음성 특성 분류기를 제안한다.

패킷 손실 환경에서는 에러의 영향에 의한 음질 저하를 줄이기 위하여 복호화기에 내장된 패킷 손실 은닉 알고리듬(PLC: Packet Loss Concealment algorithm)을 수행하는데, 일반적으로 LPC 합성 필터의 대역폭을 넓혀주고(bandwidth extension), 피치 간격을 서서히 늘림과 동시에 게인(gain) 을 서서히 줄여서 체감 음질 저하를 최소화한다. 하지만 이러한 PLC는 특성 분류를 위한 특징벡터를 왜곡시키기 때문에 합성된 신호 대신 PLC를 거치기 전에 수신된 원 파라메터를 기반으로 한 음성 특성 분류 작업을 수행할 필요가 있다.

따라서 이번 장에서는 G.729 코덱의 코딩 파라메터 중 스펙트럼에 해당하는 LPC와 파워를 이용한 묵음(S: Silence), 무성음(U: Unvoiced), 그리고 유성음(V: Voiced) 모델링을 수행한다. 이를 위해 SMV 코덱의 음성 특성 분류기를 이용해 분류된 G.729 패킷의 묵음, 무성음, 유성음에 해당하는 LPC 스펙트럼과 파워를 모델링 하였고, 이를 대상으로 파라메트릭 기반의 7개의 클래스로 특성 분류 작업을 수행하였다. 이후 비 침입적 방식을 위한 수신단 기반의 음성 특성 분류기를 위하여 패킷 에러에 의한 특성 분류 에러를 최소화하기 위한 파라메터 보간 작업을 한 후 패킷 손실 환경에서의 특성 분류 정확도를 비교하였다.

2.1 G.729 코덱 패킷의 특징벡터 추출

음성 부호화기에서 널리 쓰이는 CELP(Codebook Excited Linear Prediction) 기반의 음성 복호화기는 Fig. 1과 같이 고정 코드북(fixed codebook)과 적응 코드북(adaptive codebook)으로 여기신호를 합성하고 LPC 합성필터를 통과시켜 음성을 합성한다.[13]

Fig. 1에서 나타낸 CELP 기반의 G.729 음성 부호화기는 합성된 음성 신호, $\hat{S}(z)$를 다음의 식(1)로 표현할 수 있다.

$$\hat{S}(z) = \frac{1}{(1-A(z))}\frac{1}{(1-g_p z^{-D})}\cdot g_c \cdot E(z) \qquad (1)$$

여기에서 $E(z)$는 고정 코드북, g_c는 고정 코드북 이득값, $A(z)$는 LPC 합성 필터, g_p는 적응 코드북 이득값, 그리고 D는 적응 코드북 지연값을 나타낸다.

G.729 코덱의 특성 분류를 위해 $A(z)$의 필터 계수 및 $\hat{S}(z)$의 파워, $\hat{P}_s$를 특징 벡터로 사용하였다. 이

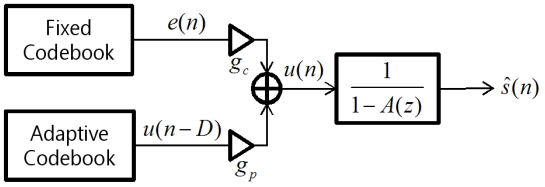

Fig. 1. Block diagram of CELP based G.729 speech decoder.

Table 1. Averaged first and second reflection coefficients and power.

Class	RC_1		RC_2		PWR (dB)	
	Avg.	Std.	Avg.	Std.	Avg.	Std.
S	-0.494	0.357	-0.10	0.226	15.53	5.62
U	-0.327	0.596	0.394	0.372	42.00	8.24
V	-0.979	0.036	0.758	0.239	61.69	4.37

때, 스펙트럼 정보는 LPC를 물리적인 의미를 가지는 반사 계수로 변환하였으며 SMV 코덱의 특성분류기에 의해 구분된 각각의 S/U/V 상태에서 1차 반사계수(RC_1), 2차 반사계수(RC_2) 및 파워(PWR)의 평균값과 분산을 구하였다. 이를 위해 ITU-T supplement 23의 database[14] 중에서 두 명의 남성과 두 명의 여성화자가 발화한 영어음성 중 40개의 샘플을 선택하여 대표적인 음성 특징인 묵음(S), 무성음(U), 유성음(V)에 해당하는 평균과 분산을 구하였으며 그 결과를 Table 1에 나타내었다.

2.2 G.729 코덱 패킷의 특성 분류

음성 특성에 따른 PLC의 음질 저하 영향을 보다 세부적으로 모델링하기 위해서 S/U/V의 3개 클래스 중에서 무성음을 화이트 노이즈 특성인 것(U1)과 아닌 것(U2)으로 나누고, 유성음은 불안정한 유성음(V5)과 안정적인 유성음(V6)으로 구분하였다. 또한 음성 특성의 전이 환경을 고려하기 위하여 무성음이 끝나는 패킷을 무성음 오프셋(U3), 그리고 유성음이 시작하는 패킷을 유성음 온셋(V4) 으로 정의하였으며 제안된 클래스는 Table 2에 나타내었다. 이와 같이 정의된 클래스로 분류하기 위해 사용된 특징 벡터인 1차 반사계수(RC1) 와 파워(Ps) 의 임계치를 Table 1과 분포에 근거하여 다음과 같이 정하였다.

파워의 경우 임계치를 22 dB, 34 dB로 정하였으며 U/V 사이의 경계선은 유성음 임에도 불구하고 상대적으로 파워가 적은 온셋 구간을 고려하여 46 dB로 설정하였다. 여기에 스펙트럼의 기울기를 반영하는 1차 반사계수를 사용하여 Table 2의 S/U1/U2/V6 분류를 수행하였다. 1차 반사계수가 0에 가까울수록 화이트 노이즈 특성의 스펙트럼 기울기를 가지며 -1에

Table 2. Category of the proposed speech classification.

Num.	Index	Description
0	S0	Silence
1	U1	Noise-like Unvoiced
2	U2	Unvoiced
3	U3	Unvoiced Offset
4	V4	Onset
5	V5	Non-stationary Voiced
6	V6	Stationary Voiced

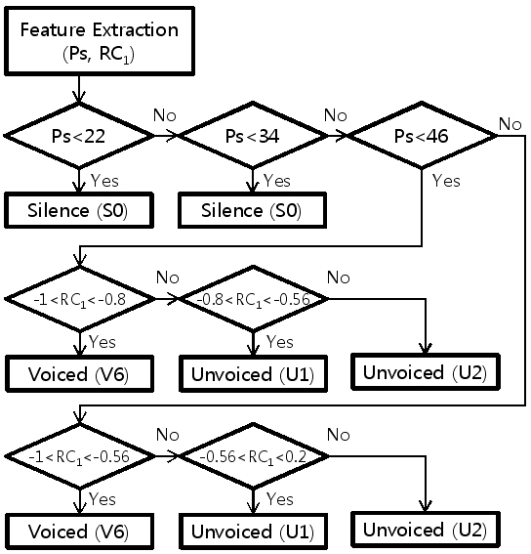

Fig. 2. Flow chart of the proposed G.729 packet classification.

가까울수록 저주파 성분이, 1에 가까울수록 고주파 성분이 많음을 의미한다. 따라서 스펙트럼 기울기에 따른 무성음 구간의 영향을 관찰하기 위하여 무성음 구간을 노이즈 특성의 무성음, U1과 파찰음 특성을 가지는 무성음, U2로 구분하였으며 그 순서도를 Fig. 2에 나타내었다.

S/U1/U2/V6 분류 작업 이후 각 클래스 간의 천이 구간을 고려하여 후보정 작업을 수행한다. 무성음으로 분류된 패킷 중 이후에 묵음이나 유성음이 올 경우 무성음 오프셋(U3), 유성음으로 분류된 패킷인 경우 이전에 무성음이나 묵음이 있으면 온셋(V4), 그리고 4개의 부 프레임에 해당하는 1차 반사계수의 분산이 0.013보다 큰 경우 비정적 유성음(V5)으로 정의

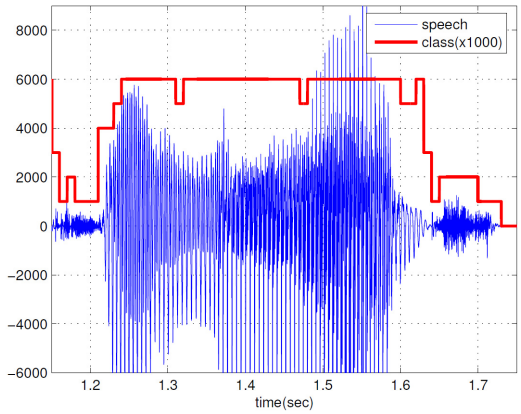

Fig. 3. An example of the proposed classification (Female).

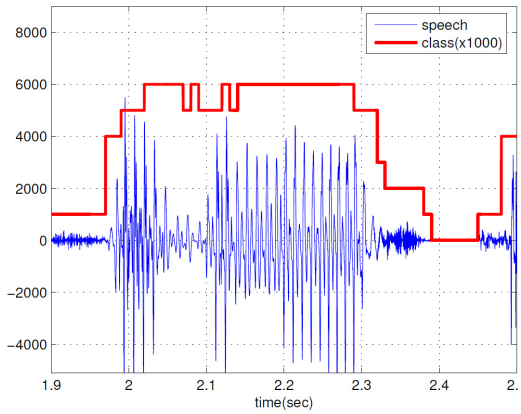

Fig. 4. An example of the proposed classification (Male).

하였다. Fig. 3과 Fig.4는 여성화자 및 남성화자에 대한 최종 음성 분류 결과의 예를 나타내었다.

2.3 패킷 손실 환경을 위한 보간 작업

패킷 에러가 발생하게 되면 일반적으로 수신단에서 과거 파라미터를 토대로 외삽법(extrapolation)을 사용하여 음성을 합성해낸다. 이러한 이유로 예측 기반의 CELP 음성 부호화기는 정적인 구간이 아닌 비정적 유성음, 혹은 천이구간에서 에러가 진행되는 결과를 야기하게 되며, 코딩 파라미터를 기반으로 한 음성 분류 작업에도 영향을 미치게 된다. 본 분류 작업에서 사용되는 G.729 코딩 파라미터 중에서 LPC 계수만 패킷 손실 이후 한 프레임까지 에러의 영향을 받기 때문에 패킷 손실 이후 2프레임의 값과 패킷 손실 이전의 값 간에 선형 보간 작업을 하고 패

Table 3. Correctness of the packet classification for the lost packet (%).

Orig. \ Est.	S0	U1	U2	U3	V4	V5	V6
S0	99.65	0.25	0.01	0.05	0.05	0	0
U1	39.88	33.9	24.82	0.28	0.58	0.43	0.12
U2	34.61	1.92	14.19	22.2	26.96	0.02	0.1
U3	33.02	8.24	4.82	12.11	4.14	19.16	18.5
V4	0.37	0.4	24.71	6.13	16.3	2.99	49.1
V5	0.26	0.19	0.31	0.32	23.76	24.84	50.33
V6	0	0.09	0.02	0.17	0.06	0.38	99.29

Table 4. Correctness of the packet classification for the lost packet with linear interpolation (%).

Orig. \ Est.	S0	U1	U2	U3	V4	V5	V6
S0	99.62	0.29	0.0023	0.015	0.075	0	0.012
U1	5.72	90.28	0.54	0.67	2.14	0.097	0.52
U2	1.61	6.06	92.30	0.0061	0.033	0	0.0045
U3	4.36	5.75	0.0039	84.98	0.23	0.68	3.59
V4	0.90	2.98	0.019	0.20	90.93	1.28	3.70
V5	0.11	0.54	0	0.97	1.68	74.54	22.05
V6	0.05	0.27	0.0005	0.53	0.44	0.73	97.94

킷 분류 작업을 수행하였다.

보간 작업의 영향을 관찰하기 위하여 ITU-T Supplement 23의 database 중 영어 샘플 8개를 선택하여 0.125% 부터 12.5% 사이의 패킷 에러를 발생시켰으며 G.729의 복호화기에서 보간 하기 전과 후의 패킷 클래스 예측률을 Table 3과 Table 4에 정리하였다. 특히 음질에 큰 영향을 미치는 온셋 구간의 예측 성공률은 16.3%에서 90.9%로 크게 향상됨을 확인할 수 있다.

III. 패킷 손실 환경의 음질 예측 모델

VoIP에서 음질을 저하시키는 요소로는 잡음, 지터, 지연값 초과 등으로 인한 패킷 손실 등이 있으며 이들 중에서 패킷 손실은 음질을 저하시키는 주된 요소 중 하나임과 동시에 시변 특성을 가지고 있기 때문에 실시간 음질을 예측하기 위한 하나의 척도로 사용된다. 하지만 같은 패킷 손실이 일어나더라도 랜덤(random) 패킷 손실보다 지속적인 버스트(burst) 패킷 손실이 음질에 더 큰 영향을 주기 때문에 기존의 파라메트릭 방식에서는 채널 에러의 패턴을 버스트 한 정도에 따라 모델링을 달리 한다.[6-8]

하지만 근본적으로는 패킷 손실률 보다 손실된 패킷이 어떤 음성 정보를 담고 있었는지에 따라 음질 저하에 미치는 정도가 달라지기 때문에 단순한 통계적 모델보다 음성 정보 특징에 따라 나누는 것이 보다 더 합리적이다. 따라서 패킷의 클래스 별로 음질 저하의 정도를 달리하여 음질 예측 방법을 제안하였으며 전체적인 개념도를 Fig. 5에 나타내었다. 패킷 특성 분류는 패킷 손실이 발생한 현재 뿐만 아니라 이전, 이후 프레임까지 고려하여 음질의 영향을 모델링하여 음질 예측에 반영하였다.

3.1 패킷 손실에 의한 음질 저하 모델

파라메트릭 방식의 ITU-T 표준인 E-모델은 다음의 수식(2)를 이용하여 전송 평가 인수 R을 계산하게 되며, 음질에 반영되는 요소로 크게 다음의 다섯 가지로 구분할 수 있다.[5]

$$R = R_0 - I_s - I_d - I_e + A \,. \tag{2}$$

여기에서 R0는 기본적인 신호 대 잡음 비율(basic signal to noise ratio), Is는 동시 손상 인수(simultaneous impairment factor), Id는 지연 손상 요소(delay impairment factor), Ie는 장비 손상 요소(equipment impairment factor), 그리고 A는 편리성(advantage factor)을 나타낸다. 각각의 요소는 정해진 장비나 실측된 데이터를 통해 변환 가능한 수식으로 정의되어 있으며 이를

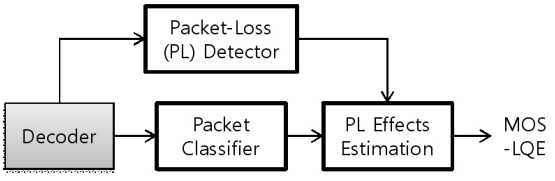

Fig. 5. Block diagram of the proposed speech quality estimation with the packet loss model.

토대로 전송 평가 지수 R을 계산할 수 있다. 이들 중에서 패킷 손실은 장비 손상 요소, Ie에 해당되며 패킷 손실률에서 버스트 정도를 반영하여 최종 결과값을 산출한다.

이처럼 E모델은 각 요소가 음질에 미치는 영향이 선형적이라고 가정한다. 마찬가지로 Ding(2007)과 Lee(2007)는 수식(3)과 같이 패킷 손실에 의한 손상 요소를 고려한 음질 예측 방식을 제안하였다.

$$MOS_e = MOS_c - DMOS_{PL}. \quad (3)$$

여기에서 MOS_c는 사용되는 음성 코덱에 의해 결정되는 고정된 MOS(Mean Opinion Score) 이며 ITU-T supplement 23의 주관적 음질 평가 결과의 평균값을 사용하였다. 본 논문에서 사용된 G.729 코덱의 경우 Sup23의 condition 1에 해당하며 총 96개의 주관적 음질 평가 결과 평균 3.8333, 표준편차 0.8786의 값을 가진다.

$DMOS_{PL}$은 음질 저하를 일으키는 요소 중 패킷 손실에 의한 영향이다. Ding(2007)은 이를 유성음과 무성음의 가중합으로 계산하였으며, Lee(2007)은 SMV의 6개 클래스 정보를 이용하여 각 클래스에 대한 가중합으로 모델링 하였고, 다음 수식(4)의 선형 모델로 나타낼 수 있다.[9]

$$D_{PL} = \mathbf{w}^T \mathbf{c}. \quad (4)$$

여기에서 는 N개의 클래스에 대한 음질저하 가중치이며, **c** 는 각 클래스에서 손실된 패킷의 수이다. 최적의 **w** 를 구하기 위해서 k번째 클래스에 해당하는 w(k)을 -1.0부터 5.0까지 변화시켰을 때 수식(5)과 같이 PESQ-LQO 값과 상관관계가 가장 높은 w(k)를 선택하도록 하였다. 이때 가중치를 유성음 클래스의 값으로 정규화하기 위하여 이에 대해서만 1.0으로 주고 나머지 구간은 0으로 초기화한다.

$$\arg\max_{w(k)} \left[\left| Corr\left\{ D_{PL}, DMOS_{PESQ-LQO} \right\} \right| \right]. \quad (5)$$

이후 실제 MOS 값으로 변환하기 위해 3차 회귀 분석을 수행하였으며 다음의 수식(6)을 통해 값을 예측한다.

$$\hat{D}_{MOS-PL} = a_0 + a_1 \cdot D_{PL} + a_2 \cdot D_{PL}^2 + a_3 \cdot D_{PL}^3. \quad (6)$$

3.2 패킷 손실 가중 모델링

수식(4)의 **w**를 모델링하기 위하여 패킷 손실이 반영된 훈련 데이터(training database)를 만들었다. 음성 샘플은 ITU-T coded speech database의 영어 샘플 8개를 선택하였다. 8초 길이를 가지는 각각의 샘플은 G.729 코덱에 의해 10 ms에 해당하는 800 프레임으로 부호화 된다. 선택된 8개의 샘플을 1개부터 100개 까지 무작위로 각각 100번의 패킷 에러를 발생시켜, 0.125~12.5%의 패킷 손실률을 가지는 총 80,000개의 음성 데이터를 만들었으며 앞서 보간 작업의 클래스 예측률을 계산할 때 사용된 환경과 동일하다. 이렇게 복호화된 음성 신호와 원 음성을 이용하여 PESQ-LQ 값을 측정하였고, 이를 추정해야 하는 목표값으로 간주하였다.

w의 초기값으로 정적 유성음(V6) 패킷의 음질 저하 가중치를 1.0으로 고정시키고, 나머지는 0으로 설정하였다. 이후 각 클래스 별로 -1.0부터 5.0까지 변화시켰을 때 PESQ-LQ 값들과 가장 상관관계가 높은 값을 선택하였으며 가중치의 변화가 수렴하는 횟수인 5번 반복하여 최종적으로 구해진 가중치를 Fig. 6과 Fig. 7에 나타내었다. 그림에서 x축의 숫자는 연속된 세 개의 패킷의 특성 번호를 나타낸 것이고 가운데 숫자가 손실된 패킷의 특성 번호를 나타낸다. y축은 "666"으로 특성 분류된 패킷의 가중치를 1로 제한하였을 때 다른 특성을 가진 패킷에서 수식(5)를 만족하는 가중치를 나타낸다. Fig. 6은 원 클래스 정보를 알 수 있는 침입 모델의 가중치이고 Fig. 7은 복호화단에서 원 클래스를 추정한 정보를 가지고 만들어진 비 침입 모델의 가중치이다. 여기에서 선형 보간 작업 결과 불안정 특성 구간의 검출이 평활(smoothing)하게 되어 비 안정적 유성음(V5)에 대한 정확도는 배제하였다.

Fig. 6의 가중치를 수식(4)를 이용해서 선형 합을

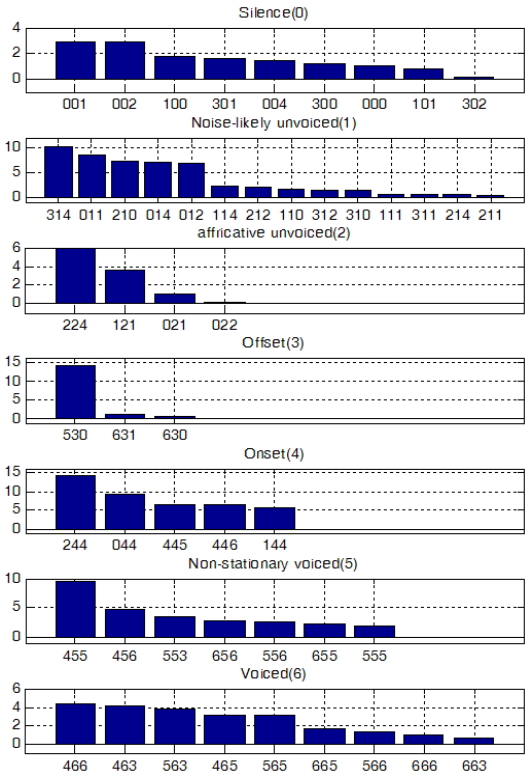

Fig. 6. Weighting values of the intrusive three-packet model.

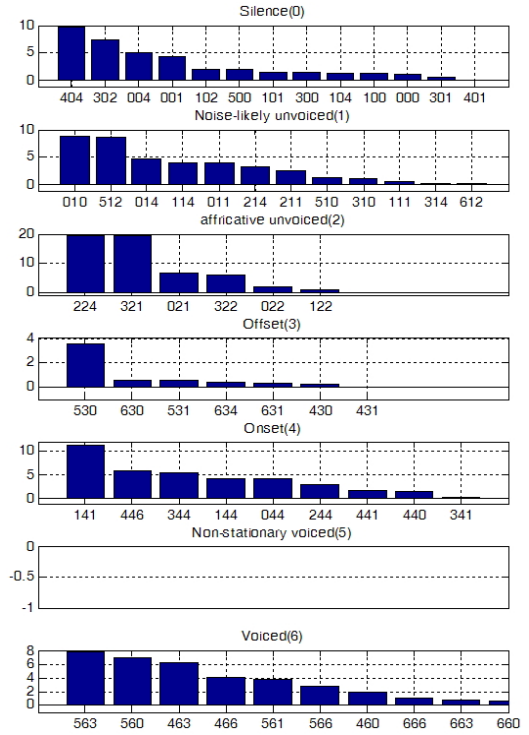

Fig. 7. Weighting values of the non-intrusive three-packet model.

구하고, 이를 MOS 스케일로 변환하기 위하여 음질 평가에서 널리 사용되는 수식(6)의 3차 회귀 분석을 수행하였다. 그 결과 원 클래스 정보를 알고 있는 침입적 모델을 가정하였을 경우 수식(6)의 각 계수 는 1.348e-2, 1.746e-2, -7.112e-5, 1.264e-7 이며, 회귀 방정식을 적용한 후의 상관계수는 0.9322를 나타내었다. 또한 복호화 단에서 클래스를 추정하는 Fig. 7의 비 침입 모델에 해당하는 수식(6)의 계수는 1.290e-2, 1.788e-2, -6.957e-5, 1,069e-7 이며, 3차 회귀 방정식 적용한 후의 상관계수는 0.9182 이다.

IV. 실험 결과

제안된 패킷 손실 가중치를 반영한 음질 평가 모델의 성능 평가를 위해 ITU-T coded speech database 에서 훈련(Training) 과정에서 쓰이지 않은 다른 영어 샘플 8개를 선택하여 실험군 1(Test 1)로 정의하고, 훈련 과정에서 쓰인 데이터베이스와 전혀 다른 NTT의 multi-lingual speech database[15] 의 영어 샘플 중 8개를 이용한 실험군 2(Test 2)를 정의하였다. 이들 실험군 1과 실험군 2에 0.125%부터 12.5% 의 패킷 손실률에 해당하는 G.729 코덱의 패킷 손실 환경을 적용하였으며 각각 4,000개의 테스트용 샘플을 만들었다. 이를 수식(4)의 패킷 손실 가중 모델을 적용하였으며, 원 패킷의 클래스 정보를 알 수 있는 침입 모델과 손실된 패킷의 클래스 정보를 추정하는 비 침입 모델의 성능 평가를 위해 수식(3)의 에 해당하는 PESQ-LQ의 저하 값과 수식(6)의 간의 상관계수를 측정하여 Table 5에 정리하였다. 그 결과 침입 모델의 경우 실험군 1인 ITU-T DB의 상관계수는 0.9217, 그리고 실험군 2인 NTT DB의 상관계수는 0.8932이며 비 침입 모델의 경우 실험군 1은 0.9086, 그리고 실험군 2

Table 5. Correlation coefficient between estimated D-MOSPL and D-PESQ.LQ.

	Intrusive 3-packet model		Non-intrusive 3-packet model	
Database	weighted sum	3rd regression	weighted sum	3rd regression
Training	0.9138	0.9322	0.9019	0.9182
Test 1	0.9089	0.9217	0.8971	0.9086
Test 2	0.8820	0.8932	0.8767	0.8898

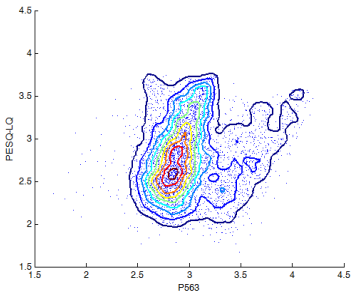

Fig. 8. P.563 vs PESQ-LQ (corr:0.1671).

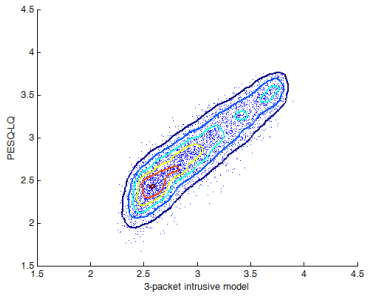

Fig. 10. Intrusive 3-packet model vs PESQ-LQ (corr:0.8950).

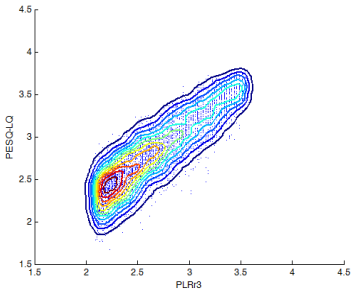

Fig. 9. PLR-reg3 vs PESQ-LQ (corr:0.8761).

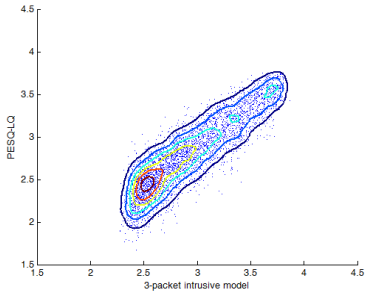

Fig. 11. Non-intrusive 3-packet model vs PESQ-LQ (corr:0.8911).

Table 6. Correlation coefficient, Mean Square Error (MSE) and percentage of absolute error between estimated MOS and PESQ-LQ.

	Corr.	MSE	Portion corresponding to the absolute error (%)							
			<0.1	<0.2	<0.3	<0.4	<0.5	<0.6	<0.7	<1.0
P.563	0.1671	0.3291	13.55	28.30	39.90	52.25	64.50	72.03	77.78	92.23
PLR-reg3	0.8761	0.1450	28.08	55.09	75.20	89.05	96.05	98.93	99.83	100
Int.3-reg3	0.8950	0.1495	29.45	55.90	76.33	89.30	95.58	98.53	99.48	100
N.Int.3-reg3	0.8911	0.1448	33.35	60.63	79.23	90.80	96.38	98.83	99.6	100

는 0.8898의 상관계수 값을 가지는 것을 확인하였다. 실험군 2인 NTT DB의 경우 언어가 실험군 1과 같은 영어 임에도 불구하고 일본에서 녹음한 환경 및 발음 차이로 인해 예측 에러가 상대적으로 큰 것을 확인할 수 있었다.

이를 토대로 얻어진 $DMOS_{PL}$을 수식(3)에 적용하여 최종적인 MOS값을 추정하였다. 마찬가지로 침입 모델과, 수신단 기반의 비 침입 모델, 두 가지로 실험하였으며, 비 침입적 음질평가 방식의 비교를 위해 패킷 손실률을 MOS 스케일로 3차 회귀 분석한 결과와 P.563의 결과를 PESQ-LQ와 비교하였다. 이때 비교군의 패킷 손실률에 대한 3차 회귀 분석은 PESQ-LO 값을 대상으로 하여 침입/비 침입 모델링에 사용한 80,000개의 훈련 데이터를 똑같이 사용하였으며 MOS값으로 변환하기 위한 3차 회귀식의 계수는 3.5679, -2.1581e-1, 1.1663e-2, 3.0703e-4이다. Fig. 8부터 Fig. 11은 각각 P.563, 패킷 손실률 - 3차 회귀분석 모델(PLR-reg3), 그리고 제안한 3패킷 침입 모델(Int.3-reg3)과 비 침입 모델(N.Int3-reg3)의 결과를 PESQ-LQ와 비교한 산점도를 나타낸다. 그리고 PESQ-LQ와 예측된 MOS 간의 상관관계(ρ)와 절대오차의 평균 제곱 에러(MSE), 그리고 각각의 절대오차 크기 포함 비율을 Table 6에 정리하였다. 실험군 1의 4,000개의 테스트 샘플에 대해 실험한 결과 P.563은 0.1671의 상관계수를 보여준 반면 패킷 손실률 - 3차 회귀분석 모델(PLR-reg3)은 0.8761, 침입 모델(Int.3-

reg3)은 0.8950, 비 침입 모델(N.Int3-reg3)은 0.8911의 상관계수를 나타내었다. 또한 제안된 침입/비 침입 모델이 다른 모델에 비해 상대적으로 적은 절대오차를 가지는 것을 확인할 수 있다.

연산량을 간접적으로 비교하기 위해 PC에서 제안된 비 침입적 3패킷 모델과 PESQ, 그리고 P.563의 수행 시간을 측정하였다. 쿼드코어 CPU인 AMD Phenom II X4 960T CPU와 8기가 메모리를 가진 PC에서 4000개 테스트 샘플을 가진 실험군1과 실험군 2에 대해 MOS값을 추정했을 때, 제안된 비 침입 방식의 패킷 가중 모델의 경우 G.729의 비트 스트림 으로부터 디코딩 과정을 포함해서 총 4분 39초가 걸렸으며 신호 기반의 침입적 음질평가 방식인 PESQ의 경우 21분 47초, 그리고 신호 기반의 비 침입적 음질평가 방식인 P.563의 경우 2시간 10분 11초가 걸렸다. G.729 코덱의 디코딩 과정을 포함한 제안된 음질평가 방식이 PESQ에 비해 4.68배, 그리고 P.563에 비해 28배 빠른 수행속도를 보여주었다.

V. 결 론

본 논문은 G.729 코덱을 이용한 패킷 손실 환경에서 비 침입적 방식의 파라메트릭 음질 예측 기법을 제안하였다. 기존의 파라메트릭 음질평가 방식은 패킷 손실의 영향을 버스트와 랜덤 특성의 두 가지로 구분하여 음질의 저하 정도를 반영하였으나 원 입력 신호의 특징을 반영하지 못하기 때문에 매 시간 변하는 음질을 평가하지 못하는 근본적인 한계가 있었다.

본 논문에서 제안한 패킷 손실 가중 모델은 패킷의 음성 특성에 따라 손실이 발생할 경우 음질에 미치는 영향을 구체적으로 수치화하기 위해 G.729의 수신단에서 코딩 파라미터를 기반으로 한 음성 특성 분류 작업을 선행하였고, 패킷 손실 환경에서 코딩 파라미터의 선형 보간 작업을 통해 원 패킷의 클래스를 추정하였다. 이를 토대로 각 패킷의 클래스 별 음질 저하 가중치의 선형합으로 음질의 저하도를 모델링 하였으며 3차 회귀 방정식을 통해 예측된 MOS와 PESQ-LQ를 비교하였다. 그 결과 제안된 비 침입 모델에서 PESQ-LQ와 0.8911의 상관계수를 가지는 것을 확인하였으며 패킷 손실률의 3차 회귀 분석 모델인 0.8761보다 높고 침입 모델의 0.8955와 비교하였을 때 크게 떨어지지 않는 안정적인 성능을 보여주었다.

제안된 모델을 패킷 손실률의 3차 회귀분석 모델과 비교하였을 때 패킷 손실률이 높아질수록 다양한 패킷 에러의 패턴을 반영하지 못하고 절대오차가 커지는 반면 제안된 비 침입 모델은 높은 패킷 손실률에도 오차가 커지지 않는 것을 확인할 수 있었다. 또한 신호 기반의 비 침입적 음질평가 표준 모델인 P.563보다 제안된 클래스 정보 기반의 비 침입적 모델이 0.1671에 비해 0.8911이라는 높은 PESQ-LQ와 상관계수를 보여주었으며 각 클래스 별로 가중치를 갖고 있기 때문에 G.729의 복호화 과정을 포함하더라도 침입적 방식인 PESQ보다 4.68배 빠른, 굉장히 낮은 연산량으로 실시간 음질을 평가할 수 있는 장점이 있다.

References

1. A.W. Rix, J. G. Beerends, D. S. Kim, P. Kroon, and O. Ghitza, "Objective Assessment of Speech and Audio Quality Technology and Applications," IEEE Trans. Audio Speech Lang. Process. **14**, 1890-1901 (2006).
2. ITU-T Recommendation P.862, *Perceptual Evaluation of Speech Quality (PESQ) : An Objective Method for End-To-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs*, 2001.
3. ITU-T Recommendation P.863, *Perceptual Objective Listening Quality Assessment*, 2011.
4. ITU-T Recommendation P. 563, *Single-Ended Method for Objective Speech Quality in Narrowband Telephony Applications*, 2004.
5. ITU-T Recommendation G.107, *The E-model : A Computational Model for Use in Transmission Planning*, 2009.
6. S. R. Broom, "VoIP quality assessment : taking account of the edge-device," IEEE Trans. Audio Speech Lang. Process. **14**, 1977-1983 (2006).
7. L. Sun, and E. C. Ifeachor, "Perceived speech quality prediction for voice over IP-based networks," IEEE ICC. 2573-2577 (2002).
8. L. Ding, Z. Lin, A. Radwan, M. S. El-hennawey, and R. A. Goubran, "Non-intrusive single-ended speech quality assessment in VoIP," Speech Commun., **49**, 477-489 (2007).
9. M. K. Lee, K. T. Kim, H. G. Kang, and D. H. Youn, "Speech quality estimation using packet loss effects in CELP-type speech coders," Proceedings Interspeech 2007, 1697-1700 (2007).

10. 3GPP2 C.S0030-0 Ver 3.0, *Selectable Mode Vocoder (SMV) Service Option for Wideband Spread Spectrum Communication Systems*, 2004.
11. ITU-T Recommendation G.729, *Coding of Speech at 8kbit/s using Conjugate-Structure Algebraic Code-Excited Linear Prediction (CS-ACELP)*, 2007.
12. ITU-T Recommendation P. 862.1, *Mapping Function for Transforming P. 862 Raw Result Score to MOS-LQO*, 2003.
13. A. M. Kondoz, *Digital Speech; Coding for Low Bit Rate Communication Systems* (John Wiley & Sons, 1994).
14. ITU-T Recommendation P.Sup23, *ITU-T Coded Speech Database*, 1998.
15. "Multi-Lingual Speech Database for Telephonometry," NTT Advanced Technology Corporation (NTT-AT), 1994.

저자 약력

▸ 이 민 기(Min-Ki Lee)

2004년 2월: 연세대학교 기계전자공학부 학사
2006년 8월: 연세대학교 전기전자공학과 석사
2006년 9월~현재: 연세대학교 전기전자공학과 박사

▸ 강 홍 구(Hong-Goo Kang)

1989년 2월: 연세대학교 전기전자공학과 학사
1991년 2월: 연세대학교 전기전자공학과 석사
1995년 8월: 연세대학교 전기전자공학과 박사
1996년~2002년: Senior Technical Staff Member, AT&T Labs-Research
2002년~2005년: 연세대학교 전기전자공학과 조교수
2005년~2011년: 연세대학교 전기전자공학과 부교수
2011년 9월~현재: 연세대학교 전기전자공학과 정교수