

VoIP 코더들의 프레임손실은닉 알고리즘 성능평가

한승호* 문광* 한민수*

* 한국정보통신대 음성/음향 정보 연구실

Performance Evaluation of Frame Erasure Concealment Algorithms in VoIP Coders

Seung Ho Han*, Kwang Moon* and Minsoo Hahn*

* Speech and Audio Information Lab, Information and Communications Univ.

space0128@icu.ac.kr

Abstract

Frame erasures cause speech quality degradation in wireless communication networks or packet networks. The degradation becomes worse when consecutive frame erasures occur. Speech coders have a frame erasure concealment(FEC) mechanism to compensate for frame erasures. It is meaningful to evaluate the performance of FEC mechanisms for frame erasures that occur in communications networks. In this paper, various frame erasures are designed. And the FEC algorithms of speech coders are evaluated and analyzed with the Perceptual Evaluation of Speech Quality(PESQ). It is found that the performances vary in accordance with frame erasure types, frame erasure rates, and utterance lengths.

I. 머리말

최근 인터넷과 같은 IP(Internet Protocol) 네트워크를 통한 통신의 사용이 증가하고 있다. 따라서 IP 네트워크를 통한 음성의 전송 역시 증가하고 있다. 그러나 IP 네트워크는 지연(delay), 지터(jitter), 반향(echo) 등과 같은 네트워크의 고유 특성 때문에 손실 없는 데이터의 전송을 보장하지 못한다. 이러한 수신단에서의 데이터 손실을 프레임 손실 또는 패킷 손실이라고 부른다. 데이터가 손실되면 품질의 저하가 발생하나 데이터의 전송은 실시간으로 이루어지기 때문에 재전송하는 것이 불가능하다. 그러므로 전송된 프레임이 손실된 경우 손실을 은닉하고 손실된 프레임과 유사한 새로운 신호를 만들어 내야 한다. 이러한 과정을 프레임손실은닉(frame erasure concealment) 또는 패킷손

실은닉(packet loss concealment)이라 한다. VoIP에 사용되는 음성 코더들은 이러한 손실에 대비하여 고유의 프레임손실은닉 알고리즘을 내장하고 있다. 프레임손실은닉 알고리즘은 그 방법에 따라 크게 2 가지로 분류된다. 새로운 복원신호를 만들기 위해 복호된 이전 음성신호를 사용하는 PCM 기반 알고리즘과, 손실없이 수신된 이전 프레임의 피치, LP 계수, 코드북 이득 등을 사용하는 파라미터 기반 알고리즘이 있다.

각 알고리즘에 따라 그 특성과 성능이 다르므로 그들의 특성을 비교하고 성능을 평가하는 것이 필요하다. 다양한 프레임 손실을 발생시키기 위해서 ITU-T에서 제공하는 EID(Error Insertion Device)를 이용하였다 [1]. 또한, 프레임손실은닉 알고리즘 성능 평가를 위해 ITU-T에서 제공하는 PESQ 값을 이용하였다 [2]. 본 실험에 사용된 코더는 PCM기반의 G.711과 CELP기반인 G.723, G.729, AMR-NB, AMR-WB이다. AMR-WB은 16 kHz의 대역폭을 가지고 있고 나머지 코더들은 8 kHz의 대역폭을 가진다.

II. 코더들의 프레임손실은닉 알고리즘

코더들에 내장된 프레임손실은닉 알고리즘을 간단히 살펴보면 다음과 같다. G.711은 PCM 기반 알고리즘을 가지고 있다. G.711은 버퍼링(buffering)을 통해 이전 신호로부터 피치 구간(pitch period)을 계산한 후 피치 구간을 반복해서 삽입한다 [3]. 나머지 코더들은 파라미터 기반 알고리즘을 가지고 있다. G.723.1과 G.729는 유성음/무성음 분류에 기초로 하여 유성음과 무성음 분류에 따라 서로 다른 여기(excitation) 신호를 발생시킨다. G.723.1은 무성음인 경우에는 랜덤 넘버 발생기(random number generator)를 이용해서, 유성음인 경우에는 피치 예측을 통해 여기신호가 만들어진다. 반

면에 G.729는 무성음일 경우에는 고정형 코드북(fixed codebook) 만을 이용해서, 유성음일 경우에는 적응형 코드북(adaptive codebook) 만을 이용해서 여기신호를 만든다 [4, 5]. AMR-NB, AMR-WB의 경우에는 현재 프레임과 이전 프레임의 상태에 기반한 7 개의 state를 가진 state machine을 가진다. State는 연속적으로 손실된 프레임 수를 나타내며 이 state machine에 기초하여 코드북의 이득(gain)을 조절하여 여기신호를 만든다 [6, 7].

III. 실험 환경

1. 프레임 손실의 발생

네트워크 채널은 Gilbert-Eliot 모델이나 Bellcore 모델 등으로 모델링 된다. Gilbert-Eliot 모델은 state 간의 이동 확률과 각 state 에서의 에러발생 확률을 가지는 binary-state Markov 모델이다. Bellcore 모델은 N-state Markov 모델로서 각 state에서 각각의 steady state 확률을 가진다. [8, 9]. 손실을 발생시키기 위해 ITU-T Recommendation G.191에 포함된 STL(Software Tool Library)를 이용한다. STL은 두 모델에 기초하여 랜덤(random) 프레임 손실과 버스트(burst) 프레임 손실을 발생시키는 EID를 제공한다. EID 안에 파라미터들을 조절해 줌으로써 다양한 프레임 손실을 발생 시킬 수 있다.

2. 실험 DB

DB는 문장길이, 화자의 성별을 고려한 한국어 음성 100문장을 사용하였다 DB는 16비트 모노 PCM 데이터로 16 kHz의 표본화 주파수(sampling rate)를 가진다. 협대역(narrow-band) 코더인 G.711, G.723.1, G.729, AMR-NB에서 실험하기 위해 위의 16 kHz 데이터를 8 kHz로 다운 샘플링(down-sampling)하여 사용하였다. DB는 음성 길이가 프레임손실은닉 알고리즘에 미치는 영향을 고려하기 위해 4초 미만의 음성, 4초 이상 8초 미만의 음성, 8초 이상 12초 미만의 음성으로 구성되었다. 또한, DB는 각각 50명의 남자음성과 여자음성으로 구성된 총 100개의 음성 데이터가 사용되었으며 그 구성은 표1과 같다.

3. 음질평가

코더의 프레임손실은닉 알고리즘의 성능평가를 위하여 음질 평가가 필요하다. 그 방법으로는 MOS 테스트 방법이 널리 사용되나 본 논문에서 테스트하는 총 데이터의 수가 44,100 개이므로 MOS 테스트를 하기는 사실상 불가능하다. 따라서 본 논문에서는 MOS 테스

트를 모델링하여 상대적인 음성 비교 평가 점수를 제공해 주는 ITU-T Recommendation P.861에 포함된 PESQ 측정 도구를 음질 평가를 위해 사용하였으며, 그 결과는 기준음성에 대해 열화된 음성의 상대적인 PESQ 값 (-0.5 ~ 4.5)을 생성한다 [2].

표 1. 실험 DB

Utterance Length	Gender	Male Speech data	Female Speech data	Total
0~4 sec (DB1)		20	20	40
4~8 sec (DB2)		15	15	30
8~12 sec (DB3)		15	15	30
Total		50	50	100

IV. 실험

실험은 2가지 목적을 가지고 수행된다. 첫째는 발생 길이에 따른 알고리즘의 평가이다. 둘째는 다양한 손실률을 가지는 랜덤 또는 버스트 프레임 손실에 대한 알고리즘의 평가이다.

1. 실험 데이터

랜덤과 버스트 프레임 손실에 대한 프레임손실은닉 알고리즘의 평가를 위해, 랜덤 프레임 손실은 1%, 3%, 5%, 10%, 15%, 20%의 손실률을 가지고, 버스트 프레임 손실은 1%, 3%, 5%, 10%, 15%의 손실률을 가지고 발생시켰다. 그리고 인위적인 조절을 통해 항상 연속적인 2, 3개의 프레임 손실을 가지는 랜덤 형태의 3%, 5%, 10%, 15%, 20%의 손실률을 가지는 프레임 손실을 발생시켰다.

이런 손실을 발생시킨 이유는 다음과 같다 비록 EID를 이용해 버스트 프레임 손실을 발생시켰지만 이 경우에는 일정 부분에서 여러 개의 프레임 손실이 동시에 날아가게 되는 경우가 대부분이다. 즉, 프레임 손실이 발생하는 구간이 집중되어 있다는 것을 의미한다. 하지만 랜덤 프레임 손실을 발생시켜 프레임 손실이 발생한 프레임 앞에 인위적으로 손실을 1, 2개 추가로 발생시키면 항상 연속적으로 2, 3개의 프레임 손실되게 된다. 따라서 임의로 발생하는 연속적인 프레임 손실에 대한 알고리즘의 평가가 가능하게 된다.

위와 같이 총 발생된 손실률의 종류는 21개이다. 한편, 비트율(bit rate)에 따라 G.711, G.729는 1개, G.723.1은 2개, AMR-NB은 8개, AMR-WB는 9개의 모드가 각각 존재한다. 따라서 코더들의 총 비트율 모드 수는 21개이다 그러므로 실험에 사용한 총 데이터 수는 44,100개(21(손실형태)*21(비트율)*100(DB))이다.

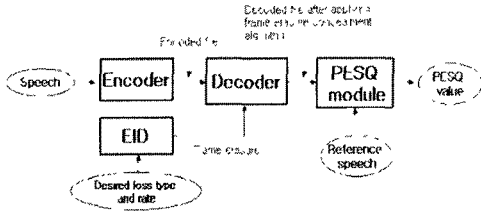


그림 1. 실험 절차

2. 실험 절차

실험절차는 그림 1과 같다. 먼저, 3.2절에서 설명한 음성 데이터를 각 코더에서 부호화한다. 그리고 4.1절에서 설명한 손실형태나 손실률에 따라 EID를 이용해 프레임 손실을 발생시킨다. 그 다음 손실이 포함된 정보를 복호화한다. 복호화시 코더에서는 내장된 손실은 너과정을 거치게 된다. 마지막으로, 복호화된 음성파일 음성 데이터를 PESQ 모듈에서 평가하여 성능을 비교한다.

V. 결과

G 711은 PCM 기반 코더이고, AMR-WB는 광대역(wide-band) 코더이므로 다른 코더들과 직접적으로 비교할 수 없다 그리고 각 코더별로 비트율의 차이가 존재하므로 역시 같은 기준으로 비교하는 것은 무리가 있다. 따라서 은닉알고리즘의 성향을 비교해야한다.

1. 발생 길이

발성 길이에 따른 성능을 비교하기 위해 각 DB별 PESQ 값의 표준편차를 구하면 표2, 3과 같다 이 값이 큰 것은 발성 길이에 대한 프레임손실은닉 알고리즘 성능의 일관성이 떨어지는 것을 의미한다 발성 길이가 짧을수록 성능의 일관성이 떨어지며 버스트 프레임 손실에서 그 영향이 더 큰 것을 확인하였다. 랜덤 프레임 손실에 대해선 AMR 코더들이 발성 길이에 영향을 많이 받는 것을 알 수 있으며 G 729는 발성 길이에 대한 영향을 가장 적게 받는 것을 확인하였다. 버스트 프레임 손실에서는 전체적으로 발성 길이에 영향을 많이 받으며, G.723.1이 가장 큰 영향을 받았다.

표 2, 랜덤 프레임 손실에 대한 PESQ값들의 표준편차

	G.711	G.723.1	G.729	AMR-NB	AMR-WB
DB1	0.2417	0.2952	0.2674	0.3520	0.3384
DB2	0.2199	0.2407	0.1936	0.2567	0.2445
DB3	0.1926	0.1602	0.1781	0.2169	0.1846
Average	0.2181	0.2320	0.2130	0.2752	0.2558

표 3 버스트 프레임 손실에 대한 PESQ값의 표준편차

	G.711	G.723.1	G.729	AMR-NB	AMR-WB
DB1	0.3607	0.4466	0.4135	0.4485	0.4526
DB2	0.2648	0.2810	0.3104	0.2852	0.2864
DB3	0.2054	0.2714	0.2410	0.2210	0.2488
Average	0.2770	0.3330	0.3216	0.3182	0.3292

2. 랜덤과 버스트 프레임 손실

일반적으로 랜덤 프레임 손실에서의 프레임손실은닉 알고리즘 성능이 버스트 프레임 손실에서의 성능 보다 우수했다. 손실률 증가에 따른 성능의 일관성을 비교하기 위하여, 표4, 5에서 보는 것과 같이 1% 손실에 대한 PESQ 값으로 각 손실률에서의 PESQ 값을 정규화 시킨다.

표 4. 랜덤 프레임 손실에 대한 성능평가 결과

	G.711	G.723.1	G.729	AMR-NB	AMR-WB
1%	1.0000	1.0000	1.0000	1.0000	1.0000
3%	0.9196	0.8789	0.9073	0.8905	0.8867
5%	0.8683	0.8122	0.8514	0.8096	0.8154
10%	0.7881	0.7020	0.7496	0.6993	0.7106
15%	0.7288	0.6228	0.6806	0.6020	0.6183
20%	0.6846	0.5608	0.6025	0.5537	0.5652

표 5. 버스트 프레임 손실에 대한 성능평가 결과.

	G.711	G.723.1	G.729	AMR-NB	AMR-WB
1%	1.0000	1.0000	1.0000	1.0000	1.0000
3%	0.8752	0.8323	0.8588	0.8892	0.9042
5%	0.8239	0.7610	0.7982	0.8324	0.7845
10%	0.7231	0.6493	0.6847	0.6913	0.6503
15%	0.6120	0.5409	0.5802	0.5945	0.5597

표 6. 랜덤 프레임 손실과 버스트 프레임 손실에 대한 프레임손실알고리즘의 성능비교

	Random	Burst	Difference
G711	3.4667	3.2893	0.1774
G723.1	3.3308	3.3314	-0.0006
G729	3.3948	3.1138	0.2810
AMR-NB	3.1956	3.2544	-0.0588
AMR-WB	3.2480	3.1698	0.0782

랜덤 손실에 대해서는 표 4에서 보는 것과 같이 손실률이 증가함에 따라 G 711, G 729의 경우 알고리즘 성능의 일관성이 잘 유지되는 것을 볼 수 있고, G 723.1, AMR-NB, WB의 경우는 성능이 빠르게 나빠지는 것을 볼 수 있다 즉, 높은 손실률에서는 낮은 손실률에서 보여주던 성능을 보여주지 못한다는 것을 의미한다

다. 버스트 손실에 대해서는 표 5에서 보는 것과 같이 손실률이 증가함에 따라 G.711, AMR-NB의 경우는 프레임손실은닉 알고리즘의 성능이 떨어지는 비율이 다른 코더들 보다 좋음을 알 수 있고, AMR-WB, G.729, G.723.1의 경우는 높은 프레임 손실률에서 더욱 성능 악화가 심해지는 것을 알 수 있다.

랜덤, 버스트 프레임 손실에 대한 프레임손실은닉 알고리즘의 성능은 1~15% 랜덤, 버스트 프레임 손실에 대한 프레임손실은닉 알고리즘 결과들의 평균값들의 차를 이용하여 비교한다. 결과는 표 6에서 확인할 수 있다. G.723.1과 AMR-NB에서는 이 차가 음의 값을 나타냈다. 이는 버스트 프레임에 대한 프레임손실은닉 알고리즘의 성능이 랜덤 프레임에 대한 성능보다 더 좋음을 의미한다. 그리고 AMR-WB에서는 그 차이가 적었다. 이는 랜덤과 버스트 프레임 손실에 대한 성능의 차이가 거의 없다는 것을 의미한다. 이와 같은 결과는 G.723.1, AMR-NB, AMR-WB에 내장된 프레임손실은닉 알고리즘이 랜덤 프레임 손실에 강인하지 않기 때문이다. G.723.1의 경우에는 설계단계부터 버스트 프레임 손실에 대해서만 강인하도록 설계되었고, AMR 코더들 역시 연속적인 에러에 기초한 state machine을 가지고 있기 때문에 이러한 결과를 보이는 것이다 [5, 6, 7]. 반대로 G.729는 이 차이가 가장 컸는데, 이는 G.729가 적응형 이득 조절기를 가지고 있지 못해, 에러 전파(error propagation) 현상이 발생하기 때문이다.

표 7. 연속된 2개의 프레임 손실 발생에 대한 성능평가 결과

	G.711	G.723.1	G729	AMR-NB	AMR-WB
3%	-0.0061	-0.0599	0.0622	-0.0266	0.0658
5%	0.0281	0.0296	0.1310	-0.0130	0.0567
10%	0.0743	0.2418	0.1722	-0.0007	0.1466
15%	0.0749	0.0945	0.1363	-0.0801	0.1456
20%	0.1095	0.1043	0.1443	-0.0199	0.2297

표 8. 연속된 3개의 프레임 손실 발생에 대한 성능평가 결과

	G.711	G.723.1	G729	AMR-NB	AMR-WB
3%	0.0390	-0.1414	0.1876	-0.0764	-0.0783
5%	0.0241	0.0075	0.1879	-0.1600	-0.0330
10%	0.0233	0.1191	0.2945	-0.0721	0.1745
15%	0.1477	0.1690	0.2749	-0.1704	0.2086
20%	0.1745	0.2119	0.3016	-0.0206	0.2324

인위적으로 항상 2, 3개의 프레임 손실이 발생하도록 조절한 프레임 손실에 대한 결과는 표7, 8과 같다 결과 분석을 위해 랜덤 프레임 손실이 발생했을 때의 결과와 연속된 2, 3개의 프레임 손실이 발생했을 때의

결과의 차를 이용했으며 이 값이 작을수록 연속적인 프레임 손실에 강인하다는 것을 의미한다. AMR-NB는 다른 코더들과는 달리 이 차이가 항상 음의 값을 보였다. 이것은 AMR-NB의 경우는 랜덤 프레임 손실이 발생되었을 때보다 오히려 연속적인 2, 3개의 프레임 손실이 발생했을 때의 성능이 더 좋을 수 있음을 시사한다.

VI. 맺음말

본 연구는 코더들의 프레임손실은닉 알고리즘의 성능을 실험적인 방법에 의해 분석하였다. 여러 가지 손실들에 대해 실험적으로 프레임손실은닉 알고리즘을 수행함으로써 각 코더들의 프레임손실은닉 알고리즘 성능에 대한 객관적인 비교가 가능했고 각 알고리즘의 특성도 알 수 있었다. 이는 각 코더에 내장되어 있는 프레임손실은닉 알고리즘의 성능개선을 위한 자료로 사용될 수 있을 것이다.

참고문헌

- [1] ITU-T, Software tools for speech and audio coding standardization, ITU-T Recommendation G.191, 2000.
- [2] ITU-T, Perceptual evaluation of speech quality (PESQ), ITU-T Recommendation P.862, 2001
- [3] ITU-T, A high quality low-complexity algorithm for packet loss concealment with G.711, ITU-T Recommendation G.711 Appendix I, 1999.
- [4] ITU-T, Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s, ITU-T Recommendation G.723.1, 1996.
- [5] ITU-T, Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited-linear-prediction (CS-ACELP), ITU-T Recommendation G.729, 1996.
- [6] ETSI, Universal mobile telecommunications system AMR speech codec; Error concealment of lost frames, ETSI TS 126 091, 2002.
- [7] ETSI, Universal mobile telecommunications system AMR speech codec, wideband Error concealment of lost frames, ETSI TS 126 191, 2002.
- [8] *Mushkin, M, Bar-David, I*, "Capacity and coding for the Gilbert-Elliott channels," IEEE Trans. on inform. Theory, Vol. 35, Nov. 1989, pp 1277-1290.
- [9] Vijay K. Varma, "Testing Speech Coders for Usage in Wireless Communications Systems," Proc. of IEEE workshop on Speech Coding for Telecommunication, Oct. 1993, pp 94-94.
- [10] 문광, "Improvement of VoIP Packet Loss Concealment for CELP-based coder", 석사학위논문, ICU, 2004.