

Time Scale Modification 을 이용한 인터넷 패킷 손실의 보상

오윤학*, 허호영*, 김희동**, 전준현***, 김형순*

*부산대학교 전자공학과, **한국외국어대학교, ***한국통신 멀티미디어 연구소

Concealment of Internet Packet Loss Using Time Scale Modification

Yoon Haak Oh*, Ho Young Hur*, Hee Dong Kim**, Jun Hyun Jeon***, Hyung Soon Kim*

*Dept. of Electronics Eng., Pusan National Univ., ** Dept. of Information and Comm.Eng.,

Hankuk University of Foreign Studies, ***Korea Telecom Multimedia Tech. Research Lab.

E-mail: {ntoh1, hyher, kimhs}@hyowon.cc.pusan.ac.kr, kimhd@ice.hufs.ac.kr

요 약

인터넷에서 전송 지연에 의해 발생한 패킷 손실은 인터넷이라는 공유망을 사용함으로써 불가피하게 발생하게 된다. 이러한 패킷 손실은 음성과 같이 실시간 처리가 필요한 미디어를 전송할 경우 매우 심각한 문제를 야기시킨다. 본 논문에서는 인터넷을 통한 음성통신에서 전송 지연으로 인해 발생한 패킷 손실을 time scale modification(TSM) 기법을 이용하여 복원하는 방법들을 검토하였다. 청취평가 결과, 송수신단 양쪽에서 TSM 을 적용할 경우 기존의 가장 우수한 접근방식인 재생성(regeneration) 방식과 거의 대등한 음질을 얻을 수 있었다.

1. 서 론

오늘날의 패킷교환망은 대용량의 데이터 전송에 사용될 뿐만 아니라 음성과 화상 전송을 위해서도 사용되고 있다. 하지만 패킷교환망은 최소 비트율이나 최대 지연에 대한 제약이 없기 때문에 네트워크 혼잡해 질 경우 네트워크 지연에 의한 패킷 손실이 불가피하며, 이는 실시간 처리가 필요한 응용 분야에서 심각한 문제를 야기시킨다. G.723.1 음성부호화 방식에서는 손실된 패킷에 대해 재생성 방법을 이용하여 복원하는데, 3 프레임 이상의 손실이 발생하면 처리하지 못하는 문제가 있다. 본 논문에서는 손실된 패킷을 보상하기 위해 TSM 을 이용한 2 가지 방법을 검토한다. 첫번째 방법은 손실된 패킷의 앞과 뒤에 있는 음성 데이터를 TSM 으로 늘려줌으로써 손실을 보상하는 것이다. 두번째 방법은 인터넷 지연에 의한 손실이 적게 발생하도록 송수신단에서 TSM 을 이용하여 음성 테이

터를 압축시켜서 전송하고, 손실 패킷을 G.723.1 음성부호화 방식의 복원방법으로 복원한 후 음성 데이터를 TSM 으로 다시 늘려주는 방법이다. 본 논문에서는 TSM 방법 중 비교적 적은 계산량으로도 우수한 성능을 나타내는 Synchronized OverLap and Add (SOLA)방법[1]을 사용하였다. 청취실험 결과 두 번째 방법의 경우 기존의 가장 우수한 방식인 재생성 방식과 거의 유사한 수준의 음질을 얻을 수 있음을 확인하였다.

2. G.723.1 음성 부호화기

G.723.1 은 DSP Group 의 TrueSpeech 를 기반으로 ITU-T 에 의해 표준화된 음성 부호화기이다. 이 표준안은 1992 년 비디오폰의 상용화에 따른, 음성부호화기의 표준화 요구에 의해 ITU-T 에서 표준화를 진행하여 1995 년 말에 완성되었다. 표준화 과정에서의 요구조건은 전화 대역폭 모뎀을 비디오 신호와 더불어 전송되기 위해 가능한 낮은 전송율(상한 8Kbps)을 갖도록 하고, ITU 의 G.726(32Kbps ADPCM)과 유사한 정도의 음질을 가져야 한다는 것이었다. 이러한 요구조건에 따라 표준화된 G.723.1 음성 부호화기는 5.3 및 6.3Kbps 의 두 가지 전송율로 동작 가능하며, 각각의 전송율에서 음성을 가장 잘 모델링하도록 최적화되었다. 음질은 MOS(Mean Opinion Score)로 3.98 이며, 이는 아날로그 전화보다 2% 낮은 값으로 상당히 우수한 편이다. 또한 G.723.1 부호화 방식은 1997 년 말에 VoIP 포럼에서도 표준안으로 채택되었다.

3. 네트워크 전송지연에 의한 문제[2]

일반적으로 음성통신에서의 지연시간은 프레임처리를 위한 지연, 음성 처리를 위한 지연 그리고 데이터 전송 지연을 들 수 있다. 저 전송율 음성 부호화기는 입력음성을 프레임 단위로 처리하기 때문에 한 프레임의 지연이 기본적으로 생기게 된다. 이러한 지연은 실제 시스템의 구현과는 상관없이 부호화기가 프레임 단위로 동작하기 때문에 생기는 지연이므로 Algorithmic 지연이라고도 한다. 그리고, 실제 통신시스템의 송신단에서 음성을 부호화하고 수신단에서 복호화하는데 걸리는 지연을 음성 처리를 위한 지연이라고 한다. 마지막으로 데이터 전송 지연은 부호화된 음성 데이터를 수신단으로 보내는데 걸리는 시간으로 채널 지연이라고도 한다. Algorithmic 지연과 음성 처리를 위한 지연 그리고 데이터 전송 지연을 합쳐서 단방향 지연이라고 하며 단방향 지연의 최대 허용시간은 일반적으로 반향이 없을 때 30 ms 정도, 반향이 있을 경우 20~25 ms 정도이다.

이러한 전송 지연 중 인터넷을 통한 음성통신의 경우 Algorithmic 지연과 음성 처리를 위한 지연은 일반적으로 40-50ms 정도이지만 데이터 전송 지연이 심하게 변하기 때문에 프레임들이 비순차적으로 들어오거나 프레임이 손실되는 문제가 생기게 된다. 비순차적인 프레임은 시퀀스 번호를 이용하여 재배열함으로써 복구할 수 있지만 전송 지연에 의한 프레임 손실은 트래픽이 심한 경우 40%까지 손실되는 것으로 나타나고 있어서 매우 심각한 문제를 야기한다. 그러므로 인터넷을 통한 실시간 멀티미디어 통신을 위해서는 이러한 전송 지연에 의한 문제, 특히 패킷 손실의 문제를 해결하는 것이 중요하다.

4. 기존의 패킷 손실 보상방법

패킷 손실이 생길 경우 복원하는 방법은 세가지로 나눌 수 있다[3]. 첫번째는 삽입(insertion) 방법으로서, 수신단에서 손실된 프레임을 묵음이나 잠음으로 대체하거나 이전 프레임을 반복해서 사용하는 것이다. 이 방법은 계산량은 작지만 음질이 상당히 떨어진다. 두번째로 보간(interpolation) 방법이 있는데, 손실된 프레임과 유사하다고 예상되는 대체 프레임을 얻기 위해서 패턴 비교나 보간을 수행한다. 본 논문에서 다루는 TSM 접근 방법도 이 부류에 속한다. 이 방법은 삽입 방법에 비해 계산량은 많지만 더 나은 음질을 가진다. 세번째 방법인 재생성(regeneration) 방법은 손실된 프레임의 앞과 뒤 프레임들

의 정보를 이용하여 그것으로부터 손실 프레임의 대체 프레임을 생성한다. 이 방법은 앞의 두 방법에 비해 복잡하지만 더 나은 음질을 얻을 수 있다.

본 논문에서 비교 수단으로 선정한 G.723.1 음성부호화기의 손실 패킷 보상 방법은 두 프레임이하의 손실에서 스펙트럼 포락선 정보는 이전 프레임들은 보간해서 사용하고, 여기 신호는 유·무성음을 판단하여 이전 프레임의 주기적인 여기 신호를 이용하거나 크기가 보상된 임의의 여기 신호를 사용한다. 세 프레임이상의 손실에 대해서는 앞의 두 프레임까지는 위 방법으로 복원하고 세번째 프레임부터는 복원하지 못하고 신호가 없어지게 된다[4]. 본 논문에서 검토한 방법은 전송 지연이 심하게 될 경우 음성 신호를 시간적으로 줄이거나 늘림으로써 프레임 손실이 생긴 부분을 복원하는 방법이다.

5. TSM 을 이용한 패킷 손실 보상방법

음성신호의 시간축 변환(time-scale modification(TSM))이란 기본 주파수 및 스펙트럼 형태 등 원래의 신호 특성을 그대로 유지하면서 발음속도만 빠르게 또는 느리게 변환시키는 것이다. 이러한 시간축 변환 방법에는 단구간 Fourier 해석에 의한 방법[5], 정현파 모델에 의한 방법[6], synchronized overlap and add(SOLA)방법[1], pitch synchronized overlap and add(PSOLA)방법[7] 등이 있다. 본 논문에서는 이들 시간축 변환 방법 중 비교적 적은 계산량으로도 우수한 성능을 나타내는 SOLA 방법을 사용했다. 이 방법은 음성신호의 시간축 변환을 위해 단구간 신호들을 중첩 가산(overlap and add)해서 더하기 전에 상호상관함수를 이용하여 단구간 신호들의 동기를 맞추는 방법을 사용한다.

TSM 을 이용하여 패킷 손실을 보상하는 방법으로서 본 논문에서는 다음 두 가지 접근 방법을 검토하였다.

5.1 수신단만 변경하는 방식

프레임 손실에 대해 수신단에서만 처리할 경우 그림 1 과 같은 방법을 사용하게 된다. 송신단에서는 음성의 시간축 변경 없이 저 전송율 부호화기를 통해 전송하고 수신단에서 복호화한 음성데이터에서 손실된 프레임의 앞뒤 프레임을 시간축으로 늘려서 보상해주게 된다. 프레임 손실이 심하면 시간축 변경을 많이 하게 될 때도 생기는데 이럴 경우 음질이 심하게 떨어진다. 이 방식은 Sanneck 등에 의해 제안되었다[8].

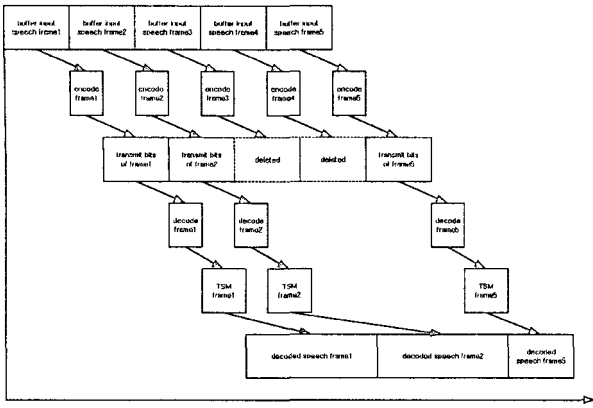


그림 1. 수신단에서만 시간축 변경을 적용하는 경우

5.2 송신단, 수신단을 모두 변경하는 방식

전송지연에 의한 프레임 손실이 있을 경우, 손실이 적게 발생하도록 송신단에서 입력음성 신호를 미리 시간축으로 0.5 배에서 1 배 사이의 값(실험에서는 0.7 배를 사용)으로 줄여서 네트워크로 전송하고 수신단에서는 다시 시간축으로 늘리게(실험에서는 1.43 배를 사용) 된다. 이때 음성신호의 시간축 변경 방법은 SOLA 를 사용한다. 이 방법을 3 가지 방식으로 적용할 수 있다. 첫번째 방식은 송신단에서 입력음성신호를 줄여서 보내고 수신단에서 손실 프레임에 대해서 TSM 으로 복원하고 다시 늘리는 것이다. 두번째 방식은 송신단에서 입력음성신호를 줄여서 보내고 수신단에서 손실 프레임에 대해서, G.723.1 음성부호화 방식에서 사용하는 복원방법을 이용하여 복원하고 다시 늘리는 것이다. 세번째 방식은 송신단에서 입력 음성신호를 줄여서 보내고 수신단에서 손실 프레임에 대해서 복원하지 않고 다시 늘린 다음에 손실 프레임에 대해서 TSM 으로 복원하는 방식이다. 위 세가지 방법 중 복원했을 때 가장 음질이 우수한 두번째 방식을 실험에 사용했다. 그림 2 는 전체 시스템에서 음성이 어떤 식으로 처리되는지 예를 든 것으로 부호화 전단계와 복호화 후에 시간축 변경을 한다. 이 방법은 송신단과 수신단 양쪽에서 TSM 을 적용함으로써 추가적인 지연을 발생시키지만 TSM 을 두번하는 과정에서의 음질 저하는 거의 없었다.

6. 실험 및 고찰

네트워크에서의 패킷 손실은 전형적으로 패킷 손실 확률로서 특징지어진다. 실험 결과에 따르면 패킷 사이의 시간 간격이 작을수록 손실 확률이 높아진다. 그리고, 이전 패킷이 손실되었을 때 다음 패킷이 손실될 확률은 이전 패킷의 손실 여부를 고려하지 않았을 때보다 큰데, 이는

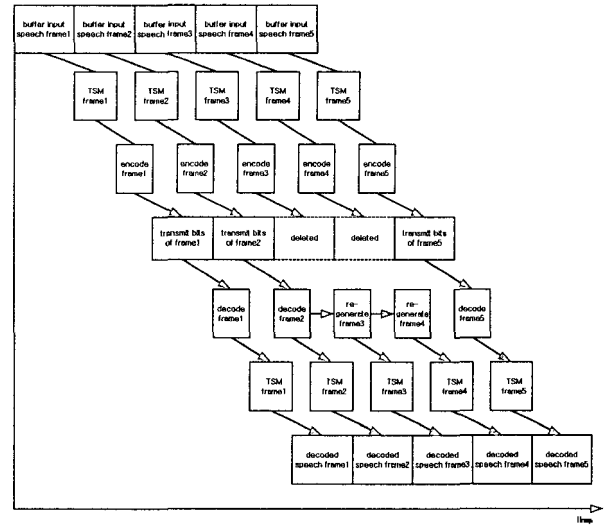


그림 2. 송신단과 수신단에 시간축 변경을 하는 경우 (손실에 대해 재생성 방법을 사용)

네트워크에서의 패킷 손실에 연관성이 있고 집중적으로 발생한다는 것을 의미한다. 본 논문에서는 참고문헌 [2]의 실험적인 측정치에 근거하여 패킷 손실을 모델링 하였다.

본 논문에서는 전송지연으로 인한 프레임 손실을 복원하는 방법으로서 기존의 G.723.1 내부에서의 regeneration 방법과 TSM 을 이용한 두 가지 방법을 비교하였다. 실험에 사용된 음성은 16 비트 및 8kHz 로 샘플링 된 5 개의 문장으로 그 내용은 다음과 같다

- (1) 힘이 더 세다고 하기로 결정했습니다.
- (2) 자산관리과 좀 바꿔주시겠어요?
- (3) 저 안전관리과 번호가 몇 번 이지요?
- (4) 수고하십니다. 회계과 좀 부탁할까요?
- (5) 음 저 총무부의 운영관리실 번호가 몇 번 이지요?

청취평가는 비교대상이 되는 두 음성 중에서 어떤 것이 사람이 듣기에 더 명료한 소리를 합성하는가를 판단하는 형태로 진행하였다. 20 명의 청취자가 실험에 참여하였으며, 청취자 자신이 어떤 음성 데이터가 어떠한 방법으로 복원된 신호인지를 알 수 없도록 음성 데이터의 순서를 랜덤하게 제시하였다. 그리고, 청취자가 비교음성 중에서 어느 쪽이 명료한지 판단할 수 없을 경우에는 구별 안됨란을 두어 판단을 유보하도록 허용하였다.

표 1 에서 기존 방식은 현재 가장 우수한 방식으로 알려진 재생성 방식에 의거하여 G.723.1 에서 채택된 패킷 손실 복원방법을 사용한 것이고, TSM 방법 1 은 프레임 손실에 대해 수신단에서 TSM 으로 복원하는 것이다. 그리고, TSM 방법 2 는 송신단에서 입력음성신호를 0.7 배로

줄여서 보내고 수신단에서 손실 프레임에 대해서, G.723.1에서 사용하는 패킷 손실 복원방법을 이용하여 복원하고 다시 1.43 배 늘리는 방법이다. 표 1(a)의 결과로부터 TSM 방법 1의 경우 기존 방식에 비해 명료도가 확실히 떨어짐을 확인할 수 있다. 이에 반하여 표 1(b)에 나타난 것처럼 TSM 방법 2의 경우 평균적으로 볼 때 기존 방식에 비해 약간 뒤떨어지는, 거의 대등한 결과를 얻을 수 있었다. 그러나, 문장에 따른 선호도의 편차가 매우 커서 이에 대한 분석검토 및 보완 실험이 요망된다.

표 1. 청취자 선호도 조사 결과 (%)

(a) 기존 방식과 TSM 방법 1의 비교

	기존방식	TSM 방법 1	구별 안됨
문장(1)	40	15	45
문장(2)	30	30	40
문장(3)	15	30	55
문장(4)	55	10	35
문장(5)	55	10	35
평균	39	19	42

(b) 기존 방식과 TSM 방법 2의 비교

	기존방식	TSM 방법 2	구별 안됨
문장(1)	0	91	9
문장(2)	64	9	27
문장(3)	9	82	9
문장(4)	82	0	18
문장(5)	82	18	0
평균	47	40	13

7. 결 론

인터넷에서 전송 지연에 의한 패킷 손실은 음성 전송과 같이 실시간 처리가 필요한 경우 매우 심각한 문제를 야기한다. 본 논문에서는 TSM을 이용하여 수신단에서 패킷 손실을 보상하는 방법과 송신단에서 음성 데이터를 압축시켜 전송하고 손실 패킷은 수신단에서 G.723.1 음성부호화기의 패킷 손실 복원방법으로 복원하여, TSM으로 다시 늘려주는 방법을 검토하였다. 청취평가 결과, 이들 중 두번째 방법의 경우 기존의 가장 우수한 방법인 재생성 방법에 비해서 문장에 따른 편차가 크기는 하지만 평균적으로 거의 대등한 명료도를 가짐을 확인하였다.

앞으로 단순히 패킷 손실의 보완이 아니라 traffic 상황

에 따라 적응적인 변환비율을 가지는 TSM을 통해 전송 지연 편차를 보상하는 방식에 대해 계속 연구할 계획이다.

참 고 문 헌

- [1] S. Roucos and A. M. Wilgud, "High Quality Time-Scale Modification for Speech," Proc. ICASSP, pp. 493-496, Apr. 1986.
- [2] J. C. Bolot, "Characterizing End-to-End Packet Delay and Loss in the Internet," Journal of High-Speed Networks, vol. 2, no. 3, pp. 305-323, Dec. 1993.
- [3] C. S. Perkins, O. Hodson and V. Hardman, "A Survey of Packet-Loss Recovery Techniques for Streaming Audio," IEEE Network Magazine, Sep./Oct. 1998.
- [4] ITU Recommendation, "G.723.1 - Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbps/s," Mar. 1996.
- [5] M. R. Portnoff, "Time-scale Modification of Speech Based on Short-time Fourier Analysis," IEEE Trans. Acoustic., Speech, Signal Processing, vol. ASSP-29, no. 3, pp. 374-390, Jun. 1981.
- [6] T. F. Quateri and R. J. McAulay, "Shape Invariant Time-scale and Pitch Modification of Speech," IEEE Trans. Signal Processing, Vol. 40, no. 3, pp. 497-510, Mar. 1992.
- [7] E. Moullines and F. Charpentier, "Pitch Synchronous Waveform Processing for Text-to-Speech Synthesis Using Diphones," Speech Communication, vol. 9 (5/6), pp. 453-467, 1990.
- [8] H. Sanneck, A. Stenger, K. Ben Younes, B. Girod, "A New Technique for Audio Packet Loss Concealment," Global Internet, London, Nov. 1996.