

텔레메틱스 기반의 VoIP 음성 통화품질 향상을 위한 수신단 구조

VoIP Receiver Structure for Enhancing Speech Quality Based on Telematics

김형국* 서광덕**
(Hyoung-Gook Kim) (Kwang-Duk Seo)

요약

텔레메틱스 기반의 VoIP 음성통화는 지연, 지터 그리고 패킷손실과 같은 네트워크 장애요소로 인해 품질저하가 발생된다. 본 논문에서는 이런 장애요소로 인한 통화품질 문제를 해결하기 위하여 수신단 기반의 통화품질 향상 알고리즘을 제안한다. 제안된 방식은 손실은닉, 역동적인 지터추정을 이용한 적응적인 플레이아웃 버퍼 스케줄링, 그리고 전이영역에 존재하는 두 신호간의 스무딩 방법으로 구성된 플레이아웃 콘트롤 및 신호합성부를 통해 고품질의 음성통화를 가능하게 한다. 성능 측정 결과, 제안된 알고리즘은 기존 알고리즘에 비해 높은 PESQ와 낮은 버퍼링 지연을 보여주었다.

Abstract

The quality of real-time voice communication over Internet Protocol networks based on telematics is affected by network impairments such as delays, jitters, and packet loss. To resolve this issue, this paper proposes a receiver-based enhancing method of VoIP speech quality. The proposed method enables users to deliver high-quality voice using playout control and signal reconstruction, which consists of concealment of lost packets, adaptive playout-buffer scheduling using active jitter estimation, and smooth interpolation between two signals in a transition region. The proposed algorithm achieves higher Perceptual Evaluation of Speech Quality (PESQ) values and low buffering delay than the reference algorithm.

Key words : Voice over Internet Protocol, Playout Control, Loss Concealment

I. 서론

통합네트워크 환경으로 발전되고 있는 인터넷을 중심으로 VoIP는 다양한 영역과 융합하여 기존에

제공되던 유선 VoIP를 벗어나 모바일 VoIP 형태로 급격히 진화하고 있으며, 최근에는 이를 기반으로 텔레메틱스 분야와 결합하여 자동차 내에서 고품질의 음성 및 영상통화 서비스를 이용할 수 있는 기

† 이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2012-0001941).

* 주저자 : 광운대학교 전파공학과 부교수

** 공저자 및 교신저자 : 광운대학교 전파공학과 석사과정

† 논문접수일 : 2011년 11월 8일

† 논문심사일 : 2012년 2월 13일

† 게재확정일 : 2012년 3월 27일

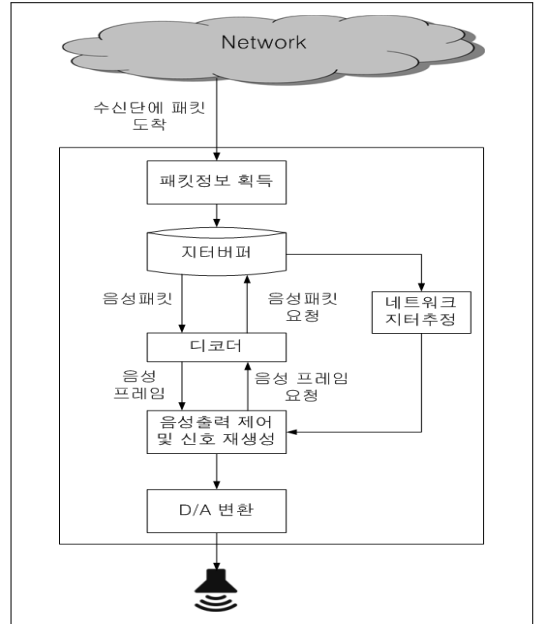
술들이 개발되고 있다. 그러나 운행 중의 자동차환경 내에서는 지속적으로 네트워크 환경이 변화되며, IP 네트워크 망을 이용한 음성통화시 네트워크 저해요소가 심하게 발생하기 때문에 정확한 네트워크 환경 예측과 버퍼링지연이 적은 효율적인 패킷 스케줄링에 의한 고품질 음성출력이 필요하다. 이러한 문제점을 해결하기 위해 VoIP기반의 손실은닉 및 플레이아웃 스케줄링 알고리즘들이 연구 및 개발되어 오고 있다[1].

본 논문에서는 텔레메틱스 환경에서 안정적인 음성 통신을 제공할 수 있는 VoIP 수신단 구조를 제안한다. 수신단 구조에서는 수신단에 도착하는 패킷들로부터 추정된 지터를 이용하여 플레이아웃을 적응적으로 스케줄링하고, 손실된 음성신호를 복원 및 병합함으로써 버퍼링지연이 적은 고품질의 통화품질을 제공한다.

II. VoIP 수신단 구조

본 논문에서 제안하는 VoIP 수신단 구조는 <그림 1>과 같다.

수신단은 크게 지터버퍼, 디코더 그리고 음성출력 제어 및 신호 재생성부로 구성된다. 송신단으로부터 전송된 20ms 길이 단위의 음성패킷은 IP네트워크망을 통해 수신단에 도착한다. 도착한 패킷의 헤더로부터 패킷의 생성시간, 도착시간 그리고 네트워크 지연 등의 패킷정보를 획득한 후 지터버퍼에 저장된다. 획득된 패킷정보를 이용하여 네트워크 지터추정부에서는 네트워크 지터를 추정한다. 디코더는 수신된 패킷을 출력하기 위해 일정한 시간 간격으로 지터버퍼에 패킷을 요청하게 되고 지터버퍼는 저장된 패킷을 디코더로 전달한다. 디코더는 패킷을 받아 20ms 길이의 음성 프레임 형태로 디코딩하여 음성출력 제어 및 신호 재생성부로 전달한다. 음성출력 제어 및 신호 재생성부는 추정된 네트워크 지터를 참조하여 네트워크 지연과 음성의 끊김 현상 등을 줄이기 위한 음성 신호처리 수행을 하며, 이를 위해 디코더로 추가적인 음성 프레임을 요청한다. 음성출력 제어 및 신호재생성부는 음성

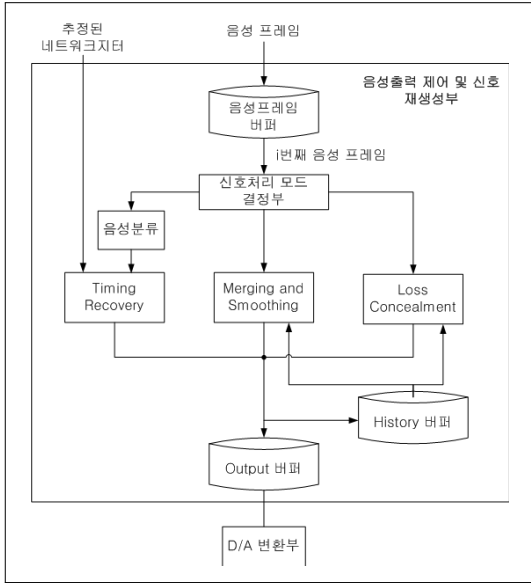


<그림 1> VoIP 수신단 구조
<Fig. 1> Structure of Receiving Part in VoIP

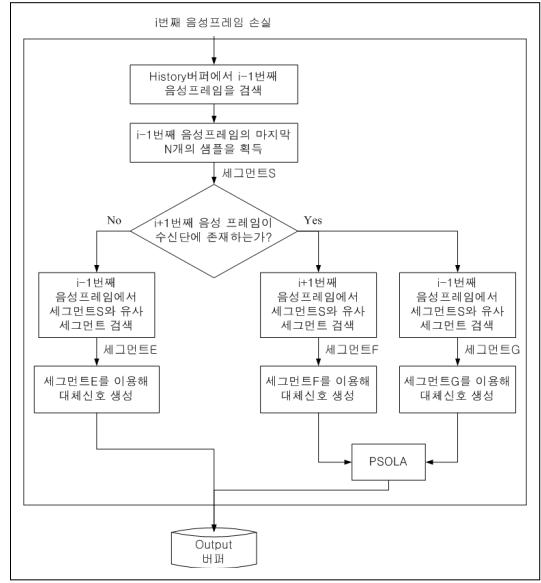
프레임의 압축 및 신장, 손실복원 그리고 병합 및 스무딩의 신호처리 과정을 거쳐 D/A변환부로 전달하고 최종적으로 스피커를 통해 사용자에게 향상된 음성 통화품질을 제공한다.

1. 음성출력 제어 & 신호 재생성부

<그림 2>는 본 논문에서 제안된 음성출력제어 및 신호 재생성부의 구조를 나타낸다. 음성출력제어 및 신호 재생성부에서는 디코딩된 음성 프레임을 저장하는 음성프레임 버퍼, 출력된 음성 프레임을 저장하는 History 버퍼 그리고 D/A변환부로 전달되기 이전 신호처리가 수행된 음성프레임을 저장하는 Output버퍼를 이용하여 기존에 사용되었던 구조와 달리 입력된 음성 프레임에 대해 패킷 스케줄링을 수행하는 타이밍복원, 손실 은닉, 음성패킷 병합 및 스무딩 등의 3가지 모드중의 한 모드를 결정하여 신호처리를 수행한다[2]. 3가지 각각의 모드에서 사용된 손실은닉, 병합 및 스무딩, 타이밍복원은 본 논문에서 제안된 구조에 맞추어 고품질 음성획득을 위해 개선된 방식이 적용되었다.



〈그림 2〉 음성출력 제어 및 신호 재생성부 블록도
 〈Fig. 2〉 Block Diagram of Playback Control and Signal Reconstruction



〈그림 3〉 손실 은닉 블록도
 〈Fig. 3〉 Block Diagram of Loss Concealment

- 손실 은닉(Loss Concealment): *i*번째 음성 프레임이 출력되기 위해서는 IP네트워크 망을 통해 수신되는 과정에서 손실 및 지연으로 인해 수신단의 음성 프레임 버퍼에 *i*번째 프레임이 저장되어 있지 않을 경우 손실은닉이 수행된다.
- 병합 및 스무딩(Merging and Smoothing): *i-1*번째 음성 프레임의 손실은닉이 수행되어 졌다면, 복원된 프레임과 정상 수신된 *i*번째 음성프레임과의 불연속점을 제거하기 위해 병합 및 스무딩을 수행한다.
- 타이밍복원(Timing Recovery): *i*번째 프레임이 존재하고 *i-1*번째 프레임이 손실되지 않은 경우에는 지터추정부를 통해 추정된 지터를 참조하여 버퍼링지연을 줄이기 위해 신호의 신장 및 압축을 통해 출력되는 음성프레임의 길이를 조절한다.

1) 손실 은닉

손실된 음성프레임의 신호는 손실 전/후 신호와 유사하기 때문에 전/후 신호를 이용하여 복원하며,

복원 방법은 다음 <그림 3>과 같다. *i-1*번째 프레임을 출력된 음성프레임이 저장되어 있는 History버퍼에서 검색한다. 검색된 프레임의 마지막 N개의 샘플(세그먼트S)을 획득한다. 음성프레임버퍼와 지터버퍼의 *i+1*번째 프레임 존재 유무에 따라 손실 은닉은 두 가지 방법으로 나누어진다.

첫 번째, *i+1*번째 프레임이 수신단에 존재하지 않는다면 History버퍼의 *i-1*번째 프레임을 이용하여 *i*번째 프레임을 복원한다. *i-1*번째 프레임에서 세그먼트 S와 유사한 세그먼트를 찾는다. 찾아진 세그먼트를 E라 정의하고, 세그먼트 E를 반복적으로 배열함으로써 손실된 음성프레임을 은닉한다.

유사 세그먼트는 다음 식 (1)을 이용해 찾는다.

$$R(k) = \frac{\sum_{t=0}^{M-1} |x_{i-1}(t+k) - s(t)|}{\sqrt{\sum_{t=0}^{M-1} x_{i-1}(t+k)} \cdot \sqrt{\sum_{t=0}^{M-1} s(t)}} \quad (1)$$

여기서 N은 프레임 S에서 획득한 샘플개수, M은 History버퍼의 샘플개수이다. *s*(*t*)와 *x*(*t*)는 각각 획득한 N개의 신호와 History버퍼의 신호를 나타낸다.

이때, $R(k)$ 중 최소의 값을 가지는 k 부터 N 개까지의 세그먼트를 유사세그먼트로 정의한다.

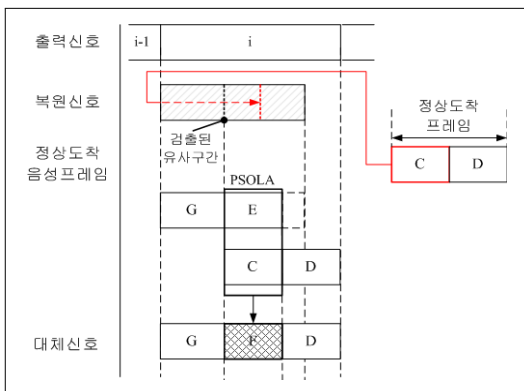
두 번째, $i+1$ 번째 음성프레임이 수신단에 존재한다면 $i-1$ 번째 프레임과 $i+1$ 번째 프레임을 사용하여 i 번째 프레임을 복원한다. 이는 첫 번째 방법과 달리 $i-1$ 번째 프레임과 $i+1$ 번째 프레임을 모두 사용하여 음성프레임을 복원함으로써 본래의 i 번째 음성 프레임과 유사한 신호를 생성할 수 있다.

$i-1$ 번째 프레임과 $i+1$ 번째 프레임에서 세그먼트S와 유사한 세그먼트를 찾는다. 찾아진 세그먼트를 세그먼트 F, 세그먼트 G라 정의한다. 세그먼트 F와 세그먼트 G를 반복적으로 배열하여 각각 음성프레임을 생성 후, PSOLA(Pitch Synchronous Overlap an Add) 방식을 이용해 최종적인 대체신호를 생성한다[3].

2). 병합 및 스무딩

<그림 4>는 병합 및 스무딩 방법을 나타내며, 다음 순서로 수행된다.

- ① 손실 은닉을 통해 출력된 $i-1$ 번째 음성 프레임 이후 신호를 생성한다.
- ② 정상 도착한 i 번째 음성 프레임을 세그먼트 C, D로(각각 10ms) 나누고 요청한 복원신호에서 세그먼트 C와 유사구간을 식(1)을 통해 검색한다. 검색된 유사구간 이전 신호는 세그먼트 G라 정의하고, 유사구간으로부터 10ms 길이를 세그먼트 E라 정의한다.
- ③ 세그먼트 E와 C에 대해 PSOLA를 수행하여 세그먼트 F를 생성한다.
- ④ 세



<그림 4> 음성프레임 병합 및 스무딩
<Fig. 4> Signal Frame Merging

그먼트 G, 세그먼트 F, 세그먼트 D를 병합하여 정상 도착한 i 번째 프레임을 대체한다.

3) 타이밍복원

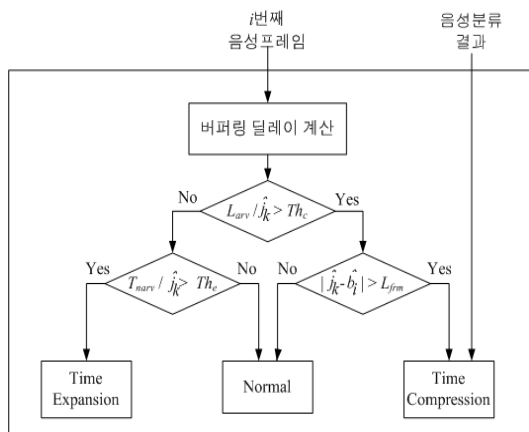
다음 <그림 5>는 타이밍복원 알고리즘을 나타낸다.

<그림 5>에서 j_k 는 지터추정부에서 k 번째로 도착한 패킷에 추정된 네트워크 지터를 나타내며, \hat{b}_i 는 i 번째 프레임이 가지는 버퍼링 지연을 나타낸다. L_{frm} 은 음성프레임의 길이를 의미하고 L_{arv} 는 현재 수신단에 남아있는 음성프레임의 길이, T_{narv} 는 현재 수신단에 패킷이 도착하지 않은 시간, Th_c 는 압축 문턱값, Th_e 는 신장 문턱값이다.

시간 복원에서는 음성패킷의 압축과 신장의 두 가지 모드 중의 하나의 모드를 결정하기 전에, 현재 수신단에 가지는 버퍼링 지연 \hat{b}_i 를 다음 식 (2)와 같이 계산한다.

$$\hat{b}_i = \hat{b}_{i-1} + \alpha_b \cdot (p_i - \alpha_i) \tag{2}$$

여기서 p_i 는 i 번째 프레임이 스피커로 출력된 시간, α_i 는 i 번째 프레임이 수신단에 도착한 시간, α_b 는 스무딩 파라미터이다. 버퍼링 지연을 계산한 후 i 번째 프레임의 시간 복원 모드를 결정한다.

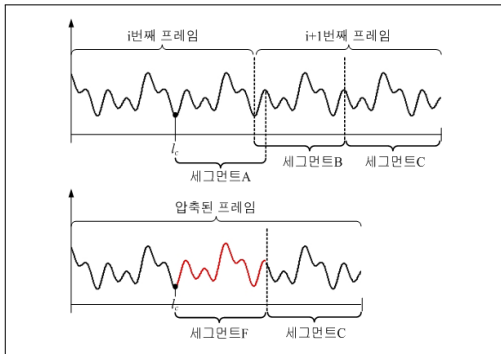


<그림 5> 시간 복원 방식 결정 블럭도.
<Fig. 5> Diagram for sub-process decision logic of timing recovery

L_{arv} 와 j_k 의 비율이 Th_c 보다 작고, T_{narv} 와 j_k 의 비율이 Th_c 보다 크다면 신장을 수행한다. 추정된 네트워크 지터는 현재 네트워크에서 발생하는 지연의 변화를 나타내며, 이러한 네트워크 지연변화만큼 수신단에 패킷이 도착하지 않는다면 현재 네트워크 지연이 증가하고 있다는 것을 의미하며, 이로 인한 수신단측 패킷 손실을 방지하기 위해 신장을 수행한다. L_{arv} 와 j_k 의 비율이 Th_c 보다 크고, j_k 와 \hat{b}_i 의 편차가 L_{frm} 보다 작다면 정상출력을 한다. 이는 버퍼링 지연이 증가하지 않는 상황에서 패킷이 일시적으로 많이 들어오는 경우 압축을 방지하고 출력을 안정된 네트워크 상황에 맞추기 위해서다.

L_{arv} 와 j_k 의 비율이 Th_c 보다 크고, j_k 와 \hat{b}_i 의 편차가 L_{frm} 보다 크다면 압축을 수행한다. 추정된 j_k 보다 많은 프레임이 수신단에 남아있다면 현재 네트워크 상황이 좋아져 패킷이 빨리 도착함을 의미하고, 이로 인해 버퍼링 지연이 증가하는 것을 막기 위해 압축을 수행한다. 음성분류결과가 무성음, 무음으로 판별이 났다면 최종적으로 압축을 수행한다. L_{arv} 와 j_k 의 비율이 Th_c 보다 작고 T_{narv} 와 j_k 의 비율이 Th_c 보다 작다면, 정상출력을 한다. 추정된 네트워크 지터보다 수신단에 패킷이 일찍 도착하게 된다면 네트워크 상황에 큰 변화 없이 패킷이 일정하게 도착하고 있음을 의미하므로 정상출력을 하게 된다.

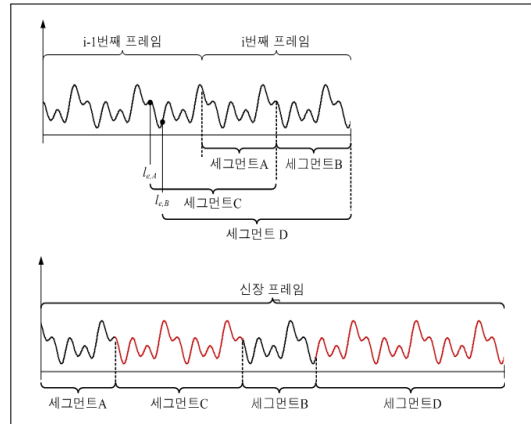
<그림 6>은 음성프레임 압축방법을 나타낸다.



<그림 6> 음성프레임 압축방법
<Fig. 6> Signal Frame Compress Algorithm

연속적인 프레임을 압축하기 위해 $i+1$ 번째 음성 프레임을 세그먼트 B, C로 구분한다. i 번째 프레임에서 세그먼트 B와 유사한 세그먼트 A를 찾고 시작점을 l_{cA} 라 한다. 세그먼트 A와 세그먼트 B를 **Overlap-add**를 수행하여 세그먼트 F를 생성하여 세그먼트A를 대체한다. 대체된 세그먼트 F와 세그먼트 C를 이어주어 압축된 음성프레임을 생성한다.

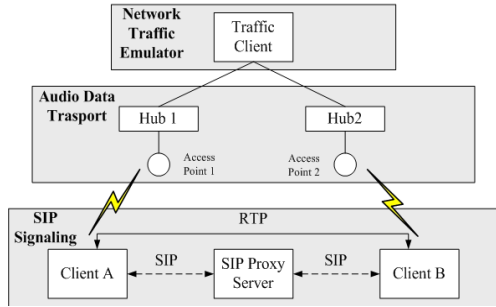
<그림 7>은 음성프레임 신장방법을 나타낸다. i 번째 음성프레임을 신장하기위해 세그먼트 A, B로 구분한다. **History**버퍼의 $i-1$ 번째 음성프레임을 획득하여 세그먼트A와 유사구간인 l_{cA} 를 식(1)을 이용해 찾고 세그먼트A의 마지막 샘플까지를 세그먼트 C라 정의한다. 세그먼트 B와 유사구간인 l_{cB} 를 식(1)을 이용해 찾고, 세그먼트 B의 마지막 샘플까지를 세그먼트 D라 정의한다. <그림 7>과 같이 세그먼트 A, C, B 그리고 D의 순서로 병합하여 i 번째 프레임의 대체신호인 신장신호를 생성한다.



<그림 7> 음성프레임 신장방법
<Fig. 7> Signal Frame Expand Algorithm

III. 실험 결과 및 고찰

본 논문에서 제안하는 수신단 구조의 성능을 측정하기 위해 **Session Initiation protocol(SIP) signal, Audio Data Transport, Network Traffic Emulator**로 구성된 테스트 베드는 <그림 8>과 같다.



〈그림 8〉 테스트베드 구조
(Fig. 8) Testbed Structure

〈표 1〉 VoIP 수신단 성능측정 실험군
(Table 1) Test Database for Performance of Mobile VoIP Receiver

구분	네트워크 지연 평균	네트워크 지연 분산	네트워크 패킷손실(%)
실험군A	30.74	7.25	0
실험군B	25.38	7.46	1.93
실험군C	24.82	8.30	3.99
실험군D	36.27	11.48	0
실험군E	34.80	13.37	3.97
실험군F	47.17	17.88	1.99

Client는 오디오 신호를 패킷화하여 Audio Data Transport 모듈로 보내며, Access Point에서는 패킷을 받아 Traffic Client로 전송한다. Traffic Client에서는 지연, 지터 그리고 패킷손실을 주어 반대편 Access Point로 전달하여 음성 신호를 전달한다.

위 테스트 베드를 이용하여 총 6개의 실험군을 구성하였으며, 네트워크 환경에 대한 정보는 다음 <표 1>과 같다.

<표 2>는 실험군에 대한 기존 알고리즘과 제안된 알고리즘의 버퍼링 지연과 PESQ 측정 결과를 나타낸다.

비교를 위해 사용한 기존 알고리즘은 시간복원 및 손실 은닉의 신호처리와 K-Erlang 분포를 따르는 지터추정방법을 사용한 음성출력제어를 결합한 수신단 구조이다[4, 5]. 기존 알고리즘과 비교하여 제안된 알고리즘에서는 수신단 패킷정보로부터 정확도가 높은 지터를 추정하여 신호처리결정부에 적용

〈표 2〉 VoIP 수신단 성능측정 결과
(Table 2) The Performance of the proposed Mobile VoIP Receiver

구분	기존 알고리즘		제안된 알고리즘	
	버퍼링 지연 측정결과	PESQ 측정결과	버퍼링 지연 측정결과	PESQ 측정결과
실험군A	40.56 ms	4.365	35.68 ms	4.542
실험군B	34.11 ms	3.703	27.81 ms	3.859
실험군C	30.60 ms	3.334	26.78 ms	3.562
실험군D	57.94 ms	3.927	53.41 ms	4.052
실험군E	45.42 ms	3.004	41.73 ms	3.378
실험군F	49.18 ms	3.021	44.73 ms	3.711

함으로써 <표 2>에 나타난 바와 같이 모든 실험군에서 제안된 알고리즘이 기존 알고리즘에 비해 버퍼링지연이 감소하였고, PESQ 또한 상승하였다. 특히, 실험군A와 실험군D에서는 네트워크 패킷손실이 없기 때문에 PESQ가 4이상의 고품질 음성신호를 획득할수 있는 반면에, 실험군C와 실험군E에서는 네트워크 패킷손실이 다른 실험군들에 비해 높기 때문에 PESQ가 상대적으로 낮음을 알 수 있다.

IV. 결 론

본 논문에서는 텔레메틱스 기반의 VoIP에서 통화품질 향상을 위한 수신단 구조를 제안하였다. 제안된 알고리즘은 총 6개의 실험군에 대해 평균 지연 33.02ms, 평균 PESQ 3.850로 기존 수신단 구조에 비해 지연 9.94ms감소하였고, PESQ는 0.291상승하였다. 제안된 알고리즘은 음성 통화시 발생하는 네트워크 저해요소를 효과적으로 해결함으로써 사용자에게 고품질의 음성 통화를 제공할 수 있다.

참 고 문 헌

[1] R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne, "Adaptive playout mechanisms for packetized audio applications in wide area network," *IEEE Infocom Conference on Computer Communication*, vol. 2, pp.680-688, June, 1994.

- [2] J.-H. Chen, "Packet loss concealment for predictive speech coding based on extrapolation of speech waveform," *ACSSC2007 Conference Record of the Forty-First Asilomar Conference on Signal, Systems and Computers*, pp.2088-2092, 2007.
- [3] D. Dorran, "Audio time-scale modification," *Dublin Institute of Technology Doctoral Thesis*, 2005.
- [4] N. Aoki, "VoIP packet loss concealment technique taking account of pitch variation in pitch waveform replicatoin," *TENCON 2004 IEEE Region 10 Conference*, vol. 3, pp.52-55, 2004.
- [5] H. Li, G. Zhang, and W. Kleijin, "Adaptive playout scheduling for VoIP using the K-Erlang distribution," *The 2010 European Signal Processing Conference*, pp.1494-1498, 2010.

저자소개



김 형 국 (Kim, Hyoung-Gook)

2007년 3월 ~ 현 재 : 광운대학교 전파공학과 부교수
2005년 4월 ~ 2007년 2월 : 삼성종합기술원 수석연구원
2002년 8월 ~ 2005년 3월 : 독일 베를린 공과대학교 Assistant Professor
1999년 1월 ~ 2002년 7월 : 독일 SIEMENS/Cortologic AG 책임연구원



서 광 덕 (Seo, Kwang-Duk)

2011년 3월 ~ 현 재 : 광운대학교 전파공학과 석사과정
2011년 : 광운대학교 전파공학과 공학사