

VoIP의 음성품질/대역효율 개선을 위한 음성패킷 처리

김재원[†], 손동철^{**}

요 약

본 논문에서는 10msec 프레임의 가변전송률 G.729 음성부호화기를 설계하여 VoIP에서의 대역 효율을 개선하고, 부전송률 부호화기 도입을 통한 인터넷 프로토콜에서 발생하는 패킷 손실구간의 음성 품질을 개선할 수 있는 방안을 도출하였다. 가변 전송률 음성부호화기 설계는 음성 통화중 발생하는 약 60% 정도의 묵음 구간을 활용하는 기법으로서 활성 음성구간은 8kbps로 전송하고, 비 활성 음성구간은 1kbps로 전송함에 의하여 고정 전송률 방식의 음성 부호화기에 비하여 평균 패킷 전송량을 약 50% 감소시켜 대역 효율을 개선할 수 있다. 제안 방법의 성능은 동일 프레임 크기를 갖는 ITU-T G.729B 방식과 감지 활성도의 변화분과 음성품질 손상 구간의 비율을 기준으로 비교 평가하였다. 그리고 인터넷 환경에서의 패킷 손실에 의한 음성 품질 저하 방지는 4kbps 부전송률 음성부호화기 도입과 오류 발생 전후의 패킷에 의한 오류은닉 방법을 활용하였으며, 성능은 재생 음성품질로서 평가하였다. 본 논문에서 설계한 가변 전송률 부호화 방식은 고정 전송률 방식에 대비하여 평균 음성 패킷의 전송량을 1/2로 감소시켜 대역효율 개선이 가능하며, 감소된 전송률을 손실 패킷구간에 활용하는 경우 8kbps 고정 전송률 방식과 동일 대역이용에서 3dB의 음성품질 개선이 가능하여 VoIP 성능 개선이 가능하리라 사료된다.

Voice Packet Processing Scheme for Voice Quality and Bandwidth Efficiency in VoIP

Jae-Won Kim[†], Dong-Chul Sohn^{**}

ABSTRACT

In this paper, We present an efficient variable rate speech coder for spectral efficiency and packet processing technique for packet loss compensation of a voice codec with 10msec frame in VoIP service. Through disconnecting the users from the spectral resource during silence interval of about 60% period, a variable rate voice coder based on a voice activity detection(VAD) can increase spectral gain by two times. The performance of the method was analyzed by variation of detected voice activity factor and degraded speech frame ratio under various background noise level, and compared those of G.729B of ITU-T 8kbps standard speech codec. A method to compensate lost packets utilized addition of recovery data to a main stream and error concealment scheme for speech quality enhancement, the performance is verified by reconstructed speech quality. The proposed scheme can achieve spectral gain by two times or enhance speech quality by 3dB through reserved bandwidth of VAD. Therefore, the proposed method can enhance a spectral efficiency or speech quality of VoIP.

Key words: VoIP(인터넷 전화), Variable Rate Coder(가변전송률 부호화기), Speech Quality Enhancement(음성 품질 향상), Error Concealment(오류 효과 최소화)

※ 교신저자(Corresponding Author): 김재원, 주소: 경상북도 경주시 효현동 산42-1번지(780-712), 전화: 054)770-5276, FAX: 054)748-2895, E-mail: jwkim@kyongju.ac.kr
접수일: 2003년 5월 19일, 완료일: 2003년 12월 22일

[†] 경주대학교 컴퓨터전자공학부 조교수

^{**} 천안대학교 정보통신공학부 조교수
(E-mail: dcon@cheonan.ac.kr)

※ 이 논문은 2001년도 한국학술진흥재단의 지원에 의하여 연구 되었음.(KRF-2001-002-E00063)

1. 서 론

음성 서비스는 실시간 전달 요구 특성으로 인하여 bursty한 트래픽[1]에 적합한 패킷교환 망보다는 일정한 채널을 보장하는 회선 교환 망이 적합한 특성을 지니지만, VoIP는 망 통합이 가져다주는 경제적인 측면에서의 요구이고, 나아가 최근에는 개별적 망운용 비용 및 서비스 제공의 제한성 극복을 위하여 다양한 유무선 통신망의 인터넷 프로토콜에 의한 통합 망으로의 진화를 목표로 많은 연구가 진행되고 있다. 하지만 VoIP는 패킷 전송 지연, 패킷 손실 등에 의한 음성 서비스 품질저하 문제를 내포할 수 밖에 없으며, 현재 기술 단계에서는 효율적인 패킷전달 기술을 통한 QoS(Quality of Service)해결[2] 방법이 대두되고 있다. VoIP 음성 품질을 결정하는 요소로는 음성 부호화 방식, 패킷 지연시간(delay), 패킷 도착시간의 상이성에 의한 jitter, 인터넷 서비스의 best efforts 특징으로 인한 패킷 손실(Lost Packet) 등이며[3], 지금까지 진행된 VoIP의 음성 품질 개선 방안들은 송신 패킷을 목적지까지 안전하게 전송하기 위한 우선 순위 부여, 음성 패킷에 비하여 큰 크기를 가지는 데이터 패킷의 분할 전송에 의한 음성 패킷 지연시간 감소, 패킷 전송 과정에서 발생할 수 있는 패킷 손상(corruption)을 대처할 수 있는 FEC (Forward Error Correction) 방법들이 제안되고 있지만 음성 패킷 손실의 영향을 완벽하게 해결될 수 없다.

본 논문에서는 현재 VoIP와 유무선 통합망의 액세스 게이트웨이에서 사용이 제안되고 있는 G.711, G.723.1, G.729 음성부호화 기법중 알고리즘 지연시간의 최소화, 효율적 패킷전송이 가능하면서 가변 전송률 기법을 수용할 수 있는 G.729 알고리즘을 근거로 가변 전송률 부호화기를 설계하고 그 성능을 분석하였으며, 손실 패킷 구간에서의 음성 품질을 개선하기 위한 부전송률 부호화기를 통한 음성 패킷 처리 기술을 설계하였다. 음성 활성화도 측정에 근거한 가변 전송률 부호화기 설계는 활성화 음성구간과 묵음 구간을 음성의 특징 파라미터에 의하여 구분하게 되는데, 묵음 구간에서는 신호의 정적인 특성에 의하여 적은 비트로서도 표현함으로써 스펙트럼 효율을 향상시키고자 한다[4-6]. 그리고 손실 패킷 구간의 음성 품질 보상을 위해서 주 전송률(8kbps) 음성부호화기로 사용한 G.729 알고리즘과 동일한 프레임 크기를 갖는 부 전송률(4kbps) 음성부호화기를 설계하여 2가

지 음성 패킷을 수신측으로 전송하여 주 전송률 부호화기의 패킷이 손실된 구간의 프레임에 부 전송률 음성부호화기의 패킷을 활용함에 의하여 음성 품질을 보상하고자 한다. 본 논문의 구성은 2장에서는 가변 전송률 부호화기의 설계로서 음성 활성화도 측정 방안과 묵음 구간의 표현을 위한 1kbps 부호화기의 설계에 대하여 설명하였으며, 3장에서는 패킷 손실 구간에서의 음성을 보상하는 방안과 부 전송률 부호화기 설계에 대하여 설명하였다. 4장에서는 설계 알고리즘의 성능 분석과정으로 음성활성도 측정 방안은 동일 프레임 크기를 갖는 ITU-T G.729B DTX 방식과 대비하였으며, 손실 패킷 처리 방안은 추가로 전송되는 부 프레임에 의하여 개선되는 음성품질을 객관적인 평가 기준을 통하여 분석하였다. 그리고 결론에서는 설계된 알고리즘의 성능을 토대로 제안 방법이 VoIP에 적용 가능함을 확인하였다.

2. 가변 전송률 부호화기 설계

가변 전송률 부호화기의 동작은 10msec 구간의 입력 음성 샘플을 음성 활성화도 측정 방안으로 활성화 음성과 묵음 구간으로 판별하여, 활성화 음성 구간은 8kbps G.729 부호화기로 부호화를 수행하고, 묵음 구간에서는 설계한 1kbps로 부호화를 수행하게 된다.

2.1 음성활성도 측정 알고리즘

음성 발생체계는 음원생성, 성도의 영향, 입술을 통한 방사 과정으로 요약되는데, 이러한 특성은 시간에 따른 음성파형의 변화, 주파수 영역에 따른 에너지 분포, 공진에 의한 포먼트 주파수 분포, 여기신호 형태등이 있으며 이를 토대로 유성음(Voiced Speech), 무성음(Unvoiced Speech), 그리고 묵음(Silence)으로 구분 가능하다[7]. 본 논문에서 설계된 음성 활성화도 측정 알고리즘의 전체적인 흐름은 그림 1에 나타난 바와 같이 입력음성 80샘플 (10msec)를 수신하여 실제 음성의 존재 여부를 결정하는 프레임 단위의 음성 활성화도 측정을 수행한다. 그 과정은 오프셋이 제거된 음성 대역 에너지, 1kHz 이하의 저주파 대역에서의 에너지 분포, 음성 신호의 영교차율을 계산하게 된다. 그리고 원래 음성 신호에서 LPC 상관 계수가 제거된 신호를 이용한 Peakiness 측정값을 구하는 과정을 수행한다.

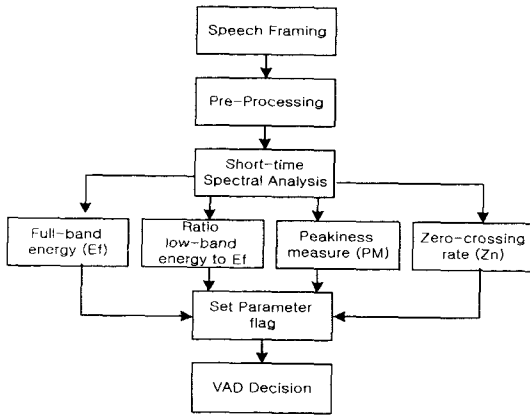


그림 1. 설계된 음성활성도 측정 알고리즘의 동작 흐름도

2.2 음성 특징 파라미터별 세부 동작

음성을 구성하는 유성음과 일부 무성음 구간은 큰 에너지 분포를 가지기 때문에 음성과 묵음 구간의 판별을 위하여 가장 기본적이고 효율적인 방법으로 프레임 에너지를 이용하게 된다. 음성 대역의 프레임 에너지($E_f(n)$)은 음성 분석 과정에서 생성되는 음성 신호의 자기 상관계수 $R(0)$ 의 정규화된 값을 이용하여 다음과 같이 정의된다.

$$E_f(n) = 10 * \log_{10}[R(0)] \tag{1}$$

추출된 음성프레임 에너지와 다양한 배경잡음 레벨에서의 시뮬레이션을 통하여 추출된 실험적 에너지 임계치 값 ($T_e(B)$)에 근거하여 묵음 구간이라고 판단된 프레임을 나타내기 위한 플래그 값 (f_{e_sil})은 다음과 같이 정의된다.

$$f_{e_sil} = \begin{cases} 1, \dots E_f(n) < T_e(B) \\ 0, \dots otherwise \end{cases} \tag{2}$$

$$T_e(B) = -(7.84 * 10^{-7}) * B^2 + 2.58 * B - 22000m \tag{2-1}$$

그러나 배경잡음이 다양하게 변화하는 환경에서는 입력 음성신호의 에너지 측정만으로는 정확한 음성 활성도 측정에 문제가 발생하게 되는데[8-11], 이의 보완을 위하여 음성의 여기신호가 성문과 입술 사이의 성도를 통과하면서 발생하는 공진특성에 의하여 결정되는 한 세트의 포만트를 분석함에 의하여 성능의 보완이 가능하다. 첫 번째 포만트 주파수에 많은 에너지 분포가 이루어지며, 이 주파수의 분포가 유성음의 경우 1kHz 이하에, 무성음의 경우 높은 주

파수 대역에 분포하게 되어 배경잡음에 의하여 손상된 음성신호의 유/무성음 판별을 위하여 주파수에 따른 에너지 분포가 음성 활성도 측정의 파라미터로 활용 가능하다. 1kHz 이하의 주파수 대역에 분포된 에너지 계산은 (식 3)과 같이 정의된다.

$$E_l(n) = 10 * \log_{10} \left[\frac{1}{N} h^T R \right] \tag{3}$$

여기서 h 는 1kHz에서 Cutoff 주파수를 갖는 FIR 필터의 임펄스 응답을 나타내며, R 은 Toeplitz 자기 상관 계수 행렬을 나타낸다. 그리고 음성 전체 대역에 대한 1kHz 이하의 대역에서의 에너지 분포는 $R_{lb} = E_l(n)/E_f(n)$ 와 같이 정의된다. 프레임 에너지에 대한 저주파 대역 에너지 분포를 통하여 유성음과 무성음을 배경잡음으로부터 구분하기 위한 플래그 값 (f_{lb})은 (식 4)와 같이 정의되어 진다.

$$f_{lb} = \begin{cases} 0, \dots (T_{lb1} \leq R_{lb} \leq T_{lb2}) \\ 1, \dots otherwise \end{cases} \tag{4}$$

여기서 T_{lb1} , T_{lb2} 는 무성음, 유성음 저주파 대역의 에너지 분포를 위한 실험적 임계치로서 0.95, 0.98을 사용하였다. 그리고 오프셋이 제거된 음성신호의 인접 샘플간의 부호 변화 즉 영교차율은 음성 신호의 스펙트럴 분포를 반영하는데, 단순한 연산 과정 만으로도 유용하게 활용될 수 있는 파라미터가 된다. 시간에 따른 진폭의 변화가 유성음 구간에서는 주기적이고 높은 진폭 레벨을 갖기 때문에 낮은 영교차율을 가지며, 낮은 신호레벨과 랜덤 잡음과 같은 특성을 갖는 무성음에서는 높은 값을 갖게 된다. 음성 신호의 영교차율 ($Z_x(n)$)은 식(5)과 같이 정의되어 진다.

$$Z_x(n) = \frac{1}{2} \sum_{k=1}^N |\text{sgn}[X(n-k)] - \text{sgn}[X(n-k-1)]| \tag{5}$$

여기서 $\text{sgn}(X)$ 는 X 가 0보다 클 때 1의 값, 그렇지 않은 경우에는 0의 값을 가지게 되며, 유/무성음 판단을 위한 두 플래그 값 (f_{z_vcc} , f_{z_unv})이 식(6)과 같이 각각 정의된다.

$$f_{z_vcc} = \begin{cases} 1, \dots Z_x(n) < T_{z1} \\ 0, \dots otherwise \end{cases}$$

$$f_{z_unv} = \begin{cases} 1, \dots Z_x(n) < T_{z2} \\ 0, \dots otherwise \end{cases} \tag{6}$$

여기서 T_{z1} , T_{z2} 는 유성음, 무성음 구간의 영교차를 임계치로서 20, 10을 각각 사용하였다. 위의 3가지 파라미터들은 음성 생성 모델에 근거한 신호의 특징을 나타내는 파라미터이기 때문에 높은 레벨의 배경잡음이 존재하는 환경에서 동일 프레임내에 나머지 샘플들에 비하여 훨씬 큰 절대값을 갖는 몇 개만의 음성샘플들이 존재하는 경우에 활성음성과 배경잡음의 판별에 어려운 점이 발생하였다. 이러한 현상은 유성음 구간의 시작과 끝부분에 해당하는 프레임, 파열음 구간에서 빈번히 발생하게 된다. 이러한 구간에서의 효율적인 동작 방안으로는 큰 진폭을 갖는 소수의 샘플들의 효과를 크게 반영함에 의하여 가능하게 되는데 이러한 개념은 원 음성 신호에서 단구간 상관관계가 제거된 LPC 잔여신호의 RMS (Root Mean Square) 값을 이용함에 의하여 가능하게 된다. LPC 잔여신호를 이용한 peakiness 측정값 (PM)은 식 (7)과 같이 정의된다.

$$PM(n) = \frac{\sqrt{\frac{1}{N} \sum_{k=1}^N r^2(n+k)}}{\frac{1}{N} \sum_{k=1}^N |r(n+k)|} \quad (7)$$

여기서 $r(n)$ 은 음성 신호에서 단구간 상관 관계인 LPC가 제거된 LPC 잔여신호를 나타내며, N 은 10msec 구간의 음성샘플 갯수에 해당하는 프레임 크기를 나타내며 그 값은 80이다. PM 값을 이용한 두 플래그 값 (f_{pm_u} , f_{pm_l})이 식 (8)와 같이 정의된다.

$$f_{pm_u} = \begin{cases} 1, \dots PM(n) > T_{pm2} \\ 0, \dots otherwise \end{cases}$$

$$f_{pm_l} = \begin{cases} 1, \dots T_{pm1} < PM(n) < T_{pm2} \\ 0, \dots otherwise \end{cases} \quad (8)$$

여기서 T_{pm1} , T_{pm2} 는 유성음, 파열음을 포함한 무성음의 PM값의 상한, 하한 임계값을 나타낸다.

음성의 특징적인 파라미터들을 근거로 식(2), 식(4), 식(6), 식(8)을 통하여 추출되어진 파라미터 각각에 대한 플래그 값들을 활성음성 판별에 대하여 보수적인 방법으로 최종적인 음성 활성화 측정이 식(9)와 같이 이루어진다. 음성 프레임 에너지, 영교차율을 통한 무성음 판별, PM의 상한치를 통한 유성음을 추출하여 활성음성으로 판단하고 나머지 구간의 보정을 위하여 각 파라미터의 조합을 통하여 부가적인

활성 음성 프레임을 판단하게 된다.

$$VAD(n) = (f_{c_sil}) | (f_{z_uv}) | (f_{pm_u}) | (f_{c_sil} \& f_{lb} \& f_{z_vcc} \& f_{pm_l}) \quad (9)$$

여기서 ‘|’, ‘&’는 논리적인 ‘Not’, ‘Or’, ‘And’을 나타내고 있으며, VAD 결과 값이 1인 경우 활성음성, 0인 경우 배경잡음 구간으로 판단하게 된다.

2.3 목음 구간에서의 음성 부호화 기법

목음 구간에서는 부호화 과정은 그림 2과 표 1에 나타난 바와 같이 주기적인 피치 성분을 갖지 않는 특성 때문에 여기 신호는 랜덤한 신호 특성을 갖으면서 균일 분포를 가지는 pseudorandom 코드 벡터에 의하여 여기신호 벡터를 형성하고, 신호의 단구간 상관관계를 추출을 위하여 LSF(Line Spectrum Frequency) 7-bit, 여기신호 이득 표현을 위하여 2-bit, 랜덤 여기신호의 생성을 위한 seed 1-bit 등으로 구성하여 10msec 구간에 10-bit를 할당함에 의하여 1kbps의 부호화기를 구성하였다.

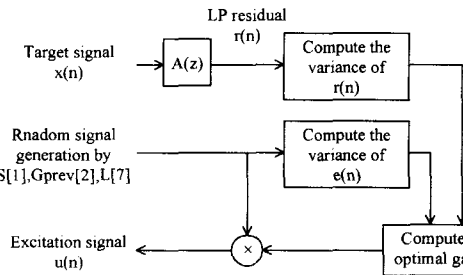


그림 2. 목음 구간에서의 음성 생성 모델

표 1. 1kbps 부호화기의 비트 할당

파라미터	할당 부호어	프레임당 비트
LSF	L[7]	7
코드북 이득	G[2]	2
여기신호 Seed	S[1]	1
전체 비트		10

3. 손실 패킷구간의 음성품질 보상

효율적인 손실패킷 복원은 손실 패킷의 특성에 부합되게 복구함이 타당함대[12,13] VoIP와 같이 많은 노드에 대규모의 패킷을 전송할 때 수신단에 도착하

는 음성패킷은 손실 패킷에 대한 복원의 도입이 필요하다. 손실 패킷 처리 방법에는 재전송과 같은 송신 측 처리방법과 수신단에서 수신패킷의 오류 효과를 최소화 시키는 오류 은닉 방법이 있다[14,15]. 본 논문에서는 실시간 전달 특성을 갖는 음성 패킷 서비스를 고려하여 오류 은닉 방법을 도입하였다. 이 방법은 음성 신호의 높은 단구간 자기 상관 관계를 이용하여, 손실 패킷 구간을 부 전송률로서 전송된 패킷에 의한 보간법에 의하여 음성을 재생하는 구조를 갖고 있으며, 손실된 패킷 주위에서 디코더의 상태를 알아내고 그것을 통해 손실된 패킷 구간의 음성 품질을 향상시키는 방법이다[16].

3.1 부가 패킷 전송을 통한 손실 패킷 복원

부가 음성 패킷을 이용한 손실 패킷 구간에서의 음성 품질 개선 방안은 입력 음성을 서로 다른 두가지 전송률의 패킷으로 구성하여 서로 다른 시간대에 전송하는 방법을 이용하였다. 주 전송률을 갖는 G.729 8kbps/1kbps 정보와 4kbps로 설계된 부 전송률 부호화기를 동시에 부호화를 수행하여, 서로 다른 시간에 코딩된 음성 패킷 정보를 하나의 패킷으로 형성한 후 전송하고 수신단에서는 비트 오류가 발생하지 않은 구간에 대해서는 주 전송률로 수신한 패킷으로 음성을 재생하고, 패킷 손실 구간은 부 전송률로 코딩된 정보를 통하여 음성을 재생하게 된다. 본 논문에서는 G.729 알고리즘을 이용하여 주 전송률 패킷을 구성하며, G.729 알고리즘의 양자화 및 코드북 검색과정의 재설계를 통하여 4kbps 부호화기를 설계하여 부 전송률 패킷을 구성하였다. G.729 알고리즘은 10msec 프레임 크기를 갖고 있어 프레임 지연에 의한 음성품질 열화는 거의 나타나지 않으며, 2개의 서로 다른 음성 부호화기가 동일 알고리즘을 이용하며 파라미터의 검색 및 양자화 과정에서만 서로 다른 과정을 수행함으로써 음성 부호화에 요구되는 계산량의 증가는 거의 나타나지 않는 장점을 지닌다. 그림 3은 수신단에서 이루어지는 음성 재생 과정을 나타내고 있다. 정상적 패킷을 수신하면 원래의 주 전송률 음성 패킷을 이용하며, 손실 패킷을 만나면 다음 패킷에 포함되어 전송되어진 손실 패킷에 대한 부 음성정보 패킷을 이용하여 손실 패킷 구간의 음성을 재생하게 된다. 부가 패킷 전송에 의한 4kbps의 전송률 증가는 음성 활성화 측정 방법의 도입을

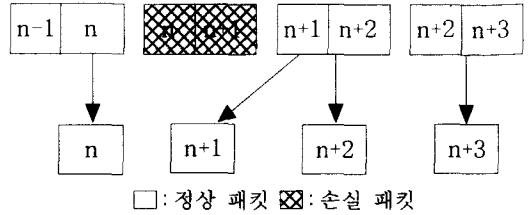


그림 3. 손실 패킷 구간의 패킷 복원 과정

통한 이득으로 상쇄 가능하여[17,18], 설계 방식을 활용하게 되면 G.729와 동일 전송률에서 패킷 손실 보상 결과를 도출할 수 있다.

3.2 4Kbps 부전송률 음성부호화기 설계

패킷 손실 구간에서의 음성 품질 개선을 위하여 이용된 4Kbps 부전송률 음성부호화기는 8kbps G.729의 5msec 코드북 검색 서브 프레임의 크기를 10msec로 확장하고, LSF 양자화의 2단계 벡터 양자화중 첫 번째 양자화 방법만을 이용하고 프레임간의 보간법을 도입하여 전송 비트를 절감하였다. 4kbps에서 이용한 고정 코드북 검색을 위한 구조는 표 2에 나타나 있는데, grid index 방법을 도입하고 각 펄스의 위치는 3 또는 4비트로 표현되어 펄스의 위치로 13비트, 부호로서 4비트, Grid로서 1비트를 사용하여 총 18비트를 이용하였다. 이를 통하여 4kbps 부 전송률 부호화기에서 할당된 비트의 구성은 표 3에 나타나 있다.

표 2. 4kbps 부전송률 부호화기의 고정 코드북 구조

Pulse	Sign	Positions
i0	s0	m0 : 0, 10, 20, 30, 40, 50, 50, 70
i1	s1	m1 : 2, 12, 22, 32, 42, 52, 62, 72
i2	s2	m2 : 4, 14, 24, 34, 44, 54, 64, 74
i3	s3	m3 : 6, 16, 26, 36, 46, 56, 66, 76 8, 18, 28, 38, 48, 58, 68, 78

4. 실험 및 결과

4.1 가변 전송률 부호화기 설계

설계된 가변 전송률 부호화기의 성능 확인을 위하여 이용된 음성 샘플은 잡음이 존재하지 않는 환경에서 남녀 각 2명의 화자에 의하여 녹음된 1,145 프레임의 음성 샘플을 이용하였으며 이 음성샘플들의 음성

활성도는 약 0.47이었다. 그리고 이동 환경에서 발생하는 잡음을 고려하여 차량, 거리잡음 등을 채취하여 신호대 잡음비를 50dB~10dB구간에서 10dB 간격을 갖게 디지털 합성방법으로 잡음이 첨가된 음성 샘플을 구성하였다. 알고리즘 성능 항목은 감지 활성화도 (Detected Activity), 음질 저하 프레임 비율 (Clipped Frame Rate) 등 2가지 객관적인 평가 기준을 사용하였다. 전자는 동일 음성 샘플의 무잡음 음성과 잡음이 포함된 음성 샘플에서의 감지 활성화도의 변화를 반영하여 스펙트럼의 효율 특성을 판단하였으며, 후자는 무잡음 음성 샘플에서는 실제 음성이 존재하는 구간이었으나 배경잡음에 의하여 음성의 특징 파라

미터가 손상되어 묵음 구간으로 판단되는 경우로서 이러한 프레임이 증가되면 음성부호화에 의하여 재생된 음성 품질이 손상되게 된다. 이와 같이 스펙트럼 효율과 음성 품질의 변화를 객관적인 평가 기준으로 평가하였다.

그림 4는 설계 알고리즘에서 이용된 파라미터들의 적합성 여부를 위하여 높은 배경잡음의 환경에서 각 파라미터 값의 변화를 나타내고 있다. (a)는 신호대 잡음비가 10dB인 잡음이 포함된 음성 샘플을 나타내고 있으며, (b)는 0~4kHz 대역에서의 프레임 에너지를 나타내고 있는데, 배경잡음에 의하여 활성 음성 구간의 판단이 어려운 구간이 존재함을 볼 수

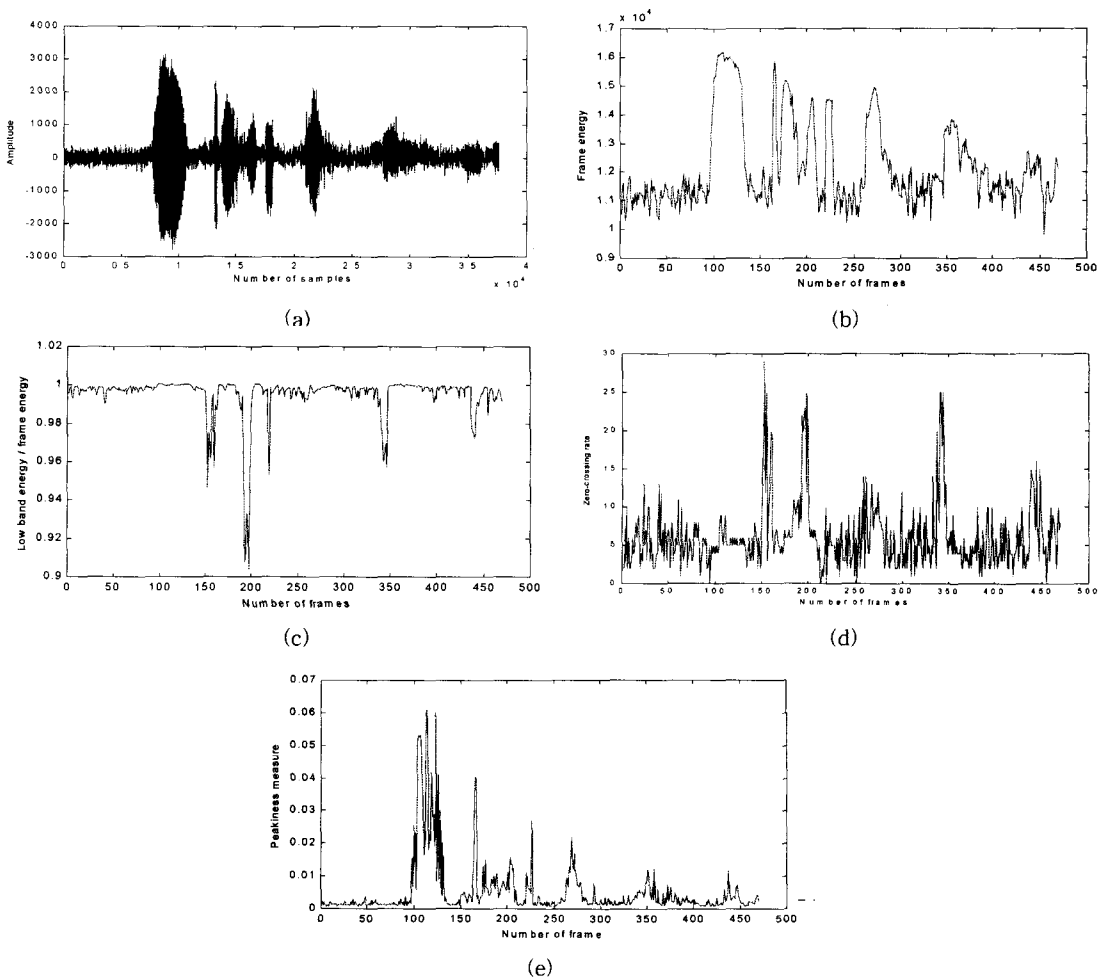
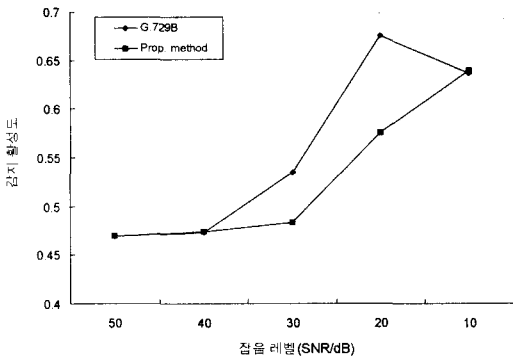


그림 4. 10dB SNR 잡음 환경에서의 파라미터들의 특성

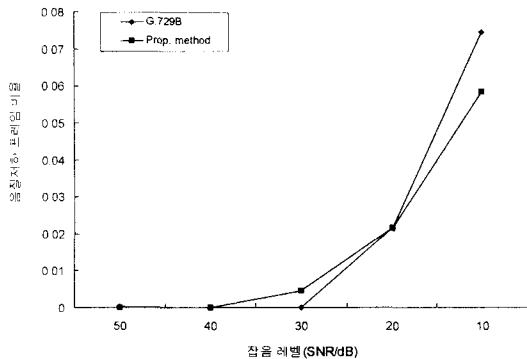
(a) 10dB SNR 음성샘플, (b) 프레임 에너지, (c) 저주파 대역의 에너지 분포, (d) 음성 샘플의 영교차율, (e) LPC 잔여신호의 Peakiness 측정값

있다. (c)는 전체 음성대역에너지에 대한 1kHz 이하의 저주파 대역의 에너지 분포의 비율 나타내고 있는데, 목음 구간과 유성음 구간은 비교적 높은 값을 갖고 있지만 무성음 구간에서는 낮은 값을 갖고 있음을 확인 가능하다. (d)는 영교차율을 나타내고 있다. 저주파 대역의 에너지 분포와는 반대로 무성음 구간에서 높은 값을 유지함을 알 수 있으며, (e)는 LPC 잔여 신호의 Peakiness 측정값을 표현하고 있는데, 유성음 구간과 목음으로 판별될 수 있는 일부 무성음 구간에서 큰 값을 나타냄을 알 수 있다.

그림 5는 배경잡음 레벨에 따른 평균 감지 활성도와 음질저하 프레임 비율의 변화를 나타내고 있다. 30, 20dB SNR에서 거의 동일한 음질저하 프레임 비율에서 G.729B VAD는 높은 감지 활성도를 나타내고 있으며, 10dB SNR에서는 동일한 값의 감지 활성도에서 설계된 알고리즘이 G.729B VAD에 비하여 낮은 음질저하 프레임 비율을 나타내어 제안 방법이



(a)



(b)

그림 5. 배경잡음 레벨 변화에 따른 감지 활성도, 음질저하 프레임 비율 변화

높은 배경잡음이 존재하는 환경에서 양호한 음성 서비스 품질을 얻을 수 있음을 확인하였다.

표 4와 표 5는 20dB, 10dB SNR의 잡음 레벨을 갖는 차량, 거리, Babble잡음 음성샘플에서의 설계된 알고리즘과 G.729B의 감지 활성도와 음질저하 프레임 비율 변화를 나타내고 있다. 20dB SNR에서는 차량 잡음과 Babble 잡음 환경에서 제안된 알고리즘이 G.729B VAD에 비하여 두가지 성능 요소에서 양호한 특성을 나타내고 있으며, 3가지 잡음환경에 대한 평균 성능 또한 동일한 음질저하 프레임 비율에서 G.729B VAD에 비하여 약 10% 낮은 감지 활성도를 나타냄을 알 수 있다. 그리고 10dB SNR에서는 동일한 감지 활성도에서 G.729B VAD에 비하여 약 1.6% 낮은 음질저하 프레임 비율을 나타냄을 확인하였다.

설계 알고리즘과 G.729B VAD 알고리즘의 성능을 비교 검토한 결과, 배경잡음이 적은 음성 샘플에서는 모두 양호한 성능을 나타내었으나 높은 배경잡음 레벨을 갖는 음성샘플에서는 G.729B VAD는 급격히 성능이 열화되는 특성을 나타내었으며, 설계된 알고리즘은 완만한 성능 저하를 나타내어 유무선 통합망에서 이동통신망에서 발생하게 되는 다양하게

표 4. 20dB SNR 음성 샘플에서의 두 알고리즘의 비교 성능

음성 샘플	무잡음 음성샘플에 대비 감지 활성도 증가(%)		음질저하 프레임비율(%)	
	제안 방안	G.729B	제안 방안	G.729B
차량 잡음	11.57	8.47	2.04	2.42
Babble 잡음	17.89	20.79	1.39	2.77
거리 소음	3.35	22.36	3.04	1.25
평균	10.60	20.54	2.16	2.15

표 5. 10dB SNR 음성 샘플에서의 두 알고리즘의 비교 성능

음성 샘플	무잡음 음성샘플에 대비 감지 활성도 증가(%)		음질저하 프레임비율(%)	
	제안 방안	G.729B	제안 방안	G.729B
차량 잡음	17.96	12.53	6.21	7.87
Babble 잡음	24.85	20.38	4.26	8.04
거리 소음	8.03	16.96	7.05	6.38
평균	16.95	16.62	5.84	7.43

변화하는 배경잡음을 갖는 환경에서 강건하게 동작할 수 있는 부호화 방법임을 확인하였다.

4.2 패킷 손실 구간의 음성 품질 개선

실제 인터넷 환경에서 수신 패킷들은 2~5% 정도의 패킷 손실을 겪고 있는데, 본 논문에서는 다양한 패킷 손실 영역에 대하여 시뮬레이션을 수행하였으며, 패킷 손실에 의한 음성 품질 저하는 객관적 음성 품질 평가 방법인 SNR을 이용하였다. 본 논문에서는 설계 방법에 대하여 평균 패킷 손실율이 0~10% 구간에 대하여 성능을 측정할 때, 5% 패킷 손실 환경 하에서도 11dB 이상의 SNR을 유지하였으며, 가변 전송률 방식에 의한 전송률 감소 이득을 음성 품질 개선에 활용한 경우 이 방법을 도입하지 않은 G.729 알고리즘에 대비하여 표 6에 나타낸 바와 같은 성능 개선 효과를 도출할 수 있었다.

제안 방법은 G.729 알고리즘에 비하여 완만한 음성 품질 열화를 나타내고 있음을 알 수 있으며, 현재 일반적으로 받아들여지고 있는 인터넷 상에서의 최대 패킷 손실률인 5% 정도에서도 약 11dB의 SNR 특성이 유지되어 VoIP 서비스에서 적절한 음성품질이 가능함을 확인하였으며, 패킷 손실률이 증가함에 의해서도 음성 품질의 열화가 급격하지 않음을 확인하였다.

표 6. 평균 패킷 손실률과 SNR 결과

평균 패킷 손실률	G.729 [dB]	제안 방법[dB]
0%	15.30	15.30
1%	12.80	14.20
5%	8.10	11.42
10%	6.00	9.05

5. 결 론

본 논문에서는 인터넷 프로토콜의 VoIP 음성 서비스 성능 개선 방안의 일환으로 송신 음성패킷의 전송률을 감소시키기 위한 가변 전송률 음성부호화기를 설계하였다. 음성 부호화기 알고리즘의 지연에 의한 성능 저하를 최소화하기 위하여 VoIP 및 차세대 통신망의 액세스 게이트웨이에서 사용 예정인 음성부호화 방식중 전송지연과 효율적 패킷 전송이 가능한 10msec 프레임 크기의 G.729 방식을 채택하여

활성음성 구간에서는 8kbps, 묵음 구간에서는 1kbps로 전송 가능한 부호화 방안을 설계하였으며, 성능은 배경잡음 레벨에 따른 감지 활성도, 음성품질 저하 프레임의 비율 변화로서 평가한 결과 동일한 프레임 크기를 갖는 G.729 알고리즘과 유사 음성품질에서 대역 효율을 2배 증대 가능함을 확인하였으며, G.729B의 음성활성도 측정 알고리즘에 비하여 배경잡음에 많은 환경에서 강건하게 동작함을 확인하였다. 그리고 가변 전송률 방식의 도입에 의하여 감소된 전송률 이득을 부 프레임의 패킷 전송으로 활용함에 의하여 8kbps G.729와 동일한 전송률, 5% 프레임 손실에서 제안방법은 약 11dB의 SNR을 유지하여 제안 방법을 도입하지 않는 경우에 비하여 3.3dB의 성능 개선 효과를 도출 가능하여 유무선 통합 통신망에서의 VoIP에 적용 가능한 음성패킷 처리 방법으로 활용 가능함을 확인하였다.

참 고 문 헌

- [1] A. Das, E. Paksoy, and A. Gersho, Multimode and Variable-Rate Coding of Speech, Speech Coding and Synthesis, (W. B. Klein and K.K Paliwal editors), pp. 257-288, Elsevier 1995.
- [2] Y. Snir, Y. Ramberg, and J. Strassner, R.Cohen, QoS Policy Information Model, Internet Draft, June 2002.
- [3] Tsung-Li Wu, S. Flix Wu, and Fengmin Gong, "Securing QoS: Threat to RSVP Message and their Countermeasure," Technical Report of NCSU, Dec. 18, 1999.
- [4] K. Srirani, P. K. Varshney, and J. G. Shanthikumar, "Discrete-time analysis of integrated voice/data multiplexers with and without speech activity detectors," IEEE J. Selected Areas in Communications, vol. SAC-1, pp. 1124-1132, December 1983.
- [5] M. Rodriguez and H. Tawfik, "Outage performance in cellular systems with discontinuous transmission," IEE Electronics Letters, vol. 29, no. 15, pp. 1329-1330, July 1993.
- [6] Jae Won Kim, Min Sik Seo, Byung Sik Yoon, Song In Choi, and Young Gap You, "A Voice

Activity Detection Algorithm for Wireless Communication Systems with Dynamically Varying Background Noise," IEICE Trans. on Communications, Vol. E83-B, No. 2, pp. 414-418, Feb. 2000.

[7] E. Paksoy, K. Srinivasan and A. Gersho, "Variable bit-rate CELP coding of speech with phonetic classification," European Transactions on Telecommunications, vol. 5, no. 9, pp. 591-601, September 1994.

[8] K. Srinivasan and A. Gersho, "Voice activity detection for cellular networks," Proc. IEEE Workshop Speech Coding for Telecommunications, pp. 85-86, October 1993.

[9] R. A. Goubran and H. M. Hafez, "Background acoustic noise reduction in mobile telephony," Proc. 36th IEEE Soc. Conf. On Vehicular technology, pp. 72-76, May 1986.

[10] J. Stegmann and G. Schroder, "Robust voice activity detection based on the wavelet transform," Proc. IEEE VTC97, pp. 99-100, 1997.

[11] F. Beritelli, S. Casale, and A. Cavallaro, "A voice activity detector for mobile communications based on fuzzy logic," Proc. European Symposium Intelligent Techniques (ESIT97), pp. 91-95, Italy, March 1997.

[12] J.C. Bolot and A. Vega-Garcia, The case for FEC based error control for packet audio in the internet, Internet Draft, June 2002.

[13] C. Perkins, O. Hodson, V. Hardman, "A Survey of Packet Loss Recovery Techniques for Streaming Audio," IEEE Network, Sep. 1998.

[14] M. Podolsky, C. Romer, S. McConne, "Simulation of FEC based Error Control for Packet Audio on the Internet," Proc. IEEE INFOCOM

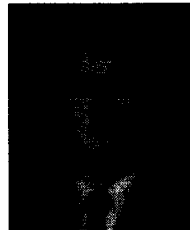
'98, April 1998.

[15] C. Montminy, T. Aboulnaser, "Improving the performance of ITU-T G.729A for VoIP," Proc. ICME 2000, August 2000.

[16] 김은중, 이인성, "인터넷 환경에서 FEC 기능이 추가된 AMR 음성부호화기를 이용한 오디오 스트림 전송," 한국통신학회논문지, 제26권 제12A호, pp. 2027-2035, 2001 12월.

[17] M. Fine, K. McCloghrie, K. Chan, S. Hahn, C. Bell, A. Smith, and F. Reichmeyer, Differentiated Services Quality of Service Policy Information Base, Internet Draft, June 2002.

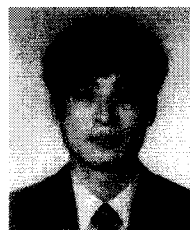
[18] T. Painter and A. Spanias, "Perceptual Coding of Digital Audio," Proc. of IEEE, vol. 88, No. 4, pp 451-515, April 2000.



김재원

1986년 2월 경북대학교 전자공학과(공학사)
 1988년 2월 경북대학교 전자공학과(공학석사)
 2000년 2월 충북대학교 정보통신공학과(공학박사)
 1991년 1월~2000년 2월 한국전자통신연구원 선임연구원

2000년 3월~현재 경주대학교 컴퓨터전자공학부 조교수
 관심분야 : 이동통신, 음성신호처리, VoIP, 멀티미디어통신



손동철

1983년 2월 경북대학교 전자공학과(공학사)
 1985년 8월 경북대학교 컴퓨터공학(공학석사)
 2001년 8월 충북대학교 정보통신공학과(공학박사)
 1983년 12월~1999년 5월 한국전자통신연구원 선임연구원

2002년 3월~현재 천안대학교 정보통신공학부 조교수
 관심분야 : 무선통신, 인터넷기술, VoIP, 멀티미디어통신