인터넷 전화에서 통화품질 향상을 위한 묵음 처리 기법

황인환[0] 최대석 이정태
부 산 대 학 교 컴 퓨 터 공 학 과
{hih93, dschoi1, jtlee}@hyowon.pusan.ac.kr

In-Hwan Hwang[0] Dae-Seok Choi Jung-Tae Lee
Dept. of Computer Engineering, Pusan National University

요 약

본 연구에서는 인터넷전화에서 전체 음성 트래픽의 56% 이상을 차지하는 묵음을 제거해줌으로써 네트워크 트래픽을 줄여 통화품질을 향상시키기 위한 방안을 제안하고 그 성능을 분석하였다. 묵음을 검출하기 위해서 평균 에너지값을 이용하는 방법을 사용하였으며, 묵음을 제거함으로써 발생하는 음성과 묵음간의 부자연스러움에 대한 해결 방안 및 음성이면서 묵음구간에 포함되어 제거되는 프레임에 대한 보상기법을 제안하였다.

1. 서 론

인터넷 전화는 기존 전화시스템에 비해 저렴한 가격과 타 서비스와 통합 및 다양한 부가 서비스의 제공 측면에서 많은 장점을 가진다. 그러나 상대적으로 낮은 음질로 인하여 사용자의 요구를 충족시키지 못하고 있다. 이는 현재 인터넷 망이 Best-effort 서비스를 지원하고 있으며 실시간성을 보장할 수 없기 때문이다. 인터넷 전화 개발을 위한 주요 기술로는 음성 코딩 및 압축, 인터넷상의 실시간 데이터 전송, 손실된 음성 패킷에 대한 복구 기술 등이 있다. 본 연구에서는 이러한 기술 중 음성코딩 및 압축기술에 있어서 좀 더 나은 통화품질을 얻을 수 있는 방안을 제시하고자 한다.

음성 통화 시 묵음이 차지하는 비율은 평균 56%에 이른다 [4]. 인터넷을 통한 음성통화에서 실제로는 묵음이지만 음성과 똑같은 크기로 데이터가 처리됨으로써 발생하는 네트워크 상의 트래픽을 제거해 줌으로써 음질이 좋아질 수 있다. 이러한 묵음을 검출해 내기 위한 여러 알고리즘이 제시되어 있으며 그 중 가장 간단하면서도 효율적인 방법으로 에너지값을 이용하는 방법이 있다[3]. 본 연구에서는 평균 에너지값을 이용한 묵음처리의 방법을 제시하고 이를 이용한 묵음 처리 시 손실되는 음성구간에 대한 복구방법, 그리고 각 코덱에 따른 통화품질의 영향을 분석하였다.

2. 묵음의 처리

음성 통화는 대화자가 번갈아가며 이야기하기 때문에 50% 이상이 묵음 시간이다. 또 쌍방이 모두 묵음 시간인 경우도 발생될 수 있으므로, 이 점을 고려하면 음성 통화에서 묵음시간의 비율은 56%에 이른다. 음성을 인터넷을 통해 전송하는 경우 일정 크기의 샘플을 주기적으로 전송하므로 이 때 묵음도 전부 디지털 신호로 변환되어 주기적으로 전송된다. 즉, 묵음일 때 전송되는 데이터들은 대역폭을 낭비하게 되는 것이다. 이러한 대역폭의 낭비는 통화품질의 저하를 초래하므로 묵음구간을 검출하여 제거해주는 것이 통화품질의 향상에 도움이 될 것이다.

2.1 최적 음성 프레임의 크기

묵음을 검출하기 위한 방법 중 에너지값을 이용한 방법으로는 한 프레임내의 평균치를 구하여 이 값과 임의의 경계값을 비교하는 방법, 한 프레임내의 묵음으로 간주되는 샘플의 비율에 따라 묵음을 검출하는 방법, 그리고 묵음 샘플이 임의의 갯수 이상 연속적으로 나타나는 경우 그 프레임을 묵음으로 간주하는 방법 등이 있다. 그 중 평균 에너지를 이용하는 방법은 가장 간단하면서도 묵음을 대체로 정확하게 검출해 낼 수 있다.

평균 에너지를 이용하는 방법은 한 프레임내의 평균에너지를 구한 후, 구해진 평균값과 임의로 설정한 경계값을 비교하여 평균값이 경계값의 범위 내에 들어오면 묵음으로 간주하는 방법이다. 이 때 프레임의 크기와 묵음을 검출하기 위한 경계값을 설정이 중요한 문제가 된다.

보통 PSTN망을 이용한 전화에서는 음성을 8000Hz의 속도로 샘플링하므로 125us마다 한개의 샘플이 상대방에게 전송된다. 패킷 교환망에서는 음성을 전송하기 위해 UDP 나 TCP 같은 전송 계층 프로토콜과 IP와 같은 네트워크 계층의 프로토콜을 사용하는데 이들 프로토콜은 보낼 음성정보를 패킷으로 만들어야 한다. 따라서, 송신측에서 녹음되는 음성을 패킷단위로 모음으로써 패킷을 생성하는 시간이 상대방에게 음성이 들리게 되는 시점에 영향을 주게 되는 것이다. 이는 다음 식으로 표현 가능하다.

$$T_{playback} \geq P_{packetization} + N_{delay} \quad (1)$$

$$P_{packetization} = P_{size} * 0.125ms \quad (2)$$

여기서, P_{size}는 패킷크기, N_{delay}는 통신망에서의 패킷전송시간, $P_{packetization}$은 패킷을 생성하는데 걸리는 시간을 나타낸다.

인터넷전화의 toll-quality를 보장하기 위해서는 200ms 내에 음성이 수신자에게 들려야 한다[1]. 보통 국내망에서 패킷의 단방향 평균지연시간은 전체 패킷의 95%가 45ms 이내이다[5]. 따라서, 200ms 이내에 음성의 재생이 이뤄져야 한다는 조건하에서 식(1)에 의한 최대 패킷 생성 시간을 계산하면 155ms이다. 즉 1 초당 8000 개의 샘플을 녹음 시 1240 개의 샘플에 해당한다.

한편, 네트워크를 통해 전송할 수 있는 음성 패킷의 최소길이는 네트워크상의 지터(Jitter)를 고려해서 계산할 수 있다. 즉 지터를 보상할 수 있을 정도의 크기를 가진 음성 패킷을 생성한다면 끊어짐 없이 음성을 재생시킬 수 있다.

따라서

$$P_{size} \geq J_{jitter} / 0.125ms \qquad (3)$$

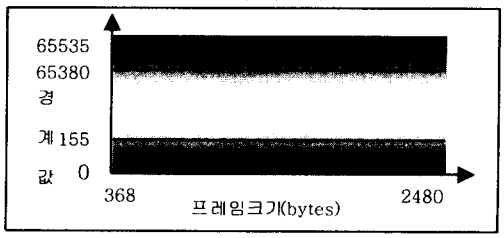
여기서 J_{jitter}는 네트워크상의 지터이다. 국내망에서의 단방향 지터의 평균은 23ms이므로[5] 이를 보상할 수 있도록 패킷의 최소크기를 결정할 경우 184 개의 샘플에 해당하게 된다. 즉 국내망에서 최적 음성 프레임의 크기는 184에서 1240개의 샘플사이에서 결정되어야 한다. 즉 한 샘플 당 16bits를 할당 할 경우 368~2480 bytes 이다.

2.2 묵음 검출을 위한 경계치

묵음 검출을 위한 적절한 경계값의 범위를 측정하기 위해 음의 끊어짐을 최소화하면서 묵음의 비율이 56%가 되는 구간을 기준으로 하였으며, 측정 조건은 PCM방식, 8kHz의 샘플링율, 모노, 샘플당 16bits 로 설정하였다. 이 경우 프레임의 크기는 368~2480bytes 범위 내로 한정하였다.

묵음 검출은 측정장소에 따른 여러 가지 변화요인을 생각할 수 있는데, 일반적으로 인터넷전화의 사용장소는 일반 가정집 정도의 소음크기를 가진 대체로 조용한 곳이라는 가정을 하였다. 일반 가정에서의 자연스런 배경잡음에 대한 평균에너지 레벨을 측정하여 평균을 산출한 결과 0~115, 65420~65535 이내의 범위에서 측정 되었으며, 묵음으로 처리될 수도 있는 화자의 아주 작은 목소리의 측정결과 255 이상과 65280 이하의 값으로 측정되었다. 따라서 경계값의 범위는 115 ~ 255, 65280 ~ 65420 이내에서 결정되어야 한다.

실제 묵음의 비율과 MOS를 통한 음질을 고려하여 가장 적절한 경계값의 측정결과 그림2.1에서와 같이 0 ~ 155, 65380 ~ 65535 범위를 묵음구간으로 설정하는 것이 가장 효율적임을 알 수 있었다. 한편 이 실험을 통해 묵음구간은 프레임의 크기와 거의 연관이 없음을 알 수 있었다.



[그림2.1] 프레임크기에 따른 묵음 검출 경계값 측정

2.3 압축 기법에 따른 묵음의 영향

인터넷 전화에서 많이 사용되는 압축 코덱으로는 GSM (Global System for mobile communication)과 G.723.1 등이 있다. 이러한 압축 방법들이 묵음의 검출과 경계값 설정 등에 어떻게 영향을 미칠 지에 대한 분석이 이뤄져야 한다. 이를 위해 각 압축 기법을 실제로 인터넷 전화에 적용하여 음질에 대한 MOS 및 묵음비율을 고려한 테스트를 통해 영향을 파악 하였다.

GSM은 음성 통화를 위한 전송 속도가 13kbps로 한 프레임 당 20ms의 크기를 가진다. 즉 8kHz 샘플링 시 160 개의 샘플로 한 프레임이 구성되며 본 연구에서는 한 샘플 당 16bits를 할당 하므로 320bytes 단위로 프레임의 크기가 결정된다. 이때 지터와 네트워크 지연을 고려한 프레임 크기에 따른 묵음검출 경계값의 범위 또한 그림2-1과 같은 결과를 보였다.

G.723.1은 5.3kbps와 6.3kbps의 전송속도를 가지는 압축방식으로 한 프레임 당 30ms의 크기를 가진다. 즉 8kHz, 16bit PCM방식에서 한 프레임 당 240 개의 샘플로 구성된다. 즉 한 샘플당 16bits를 할당할 경우 480bytes단위로 프레임 크기가 결정된다.

두 압축기법에 따른 묵음 검출 경계값의 측정에서 일반 PCM 방식과 유사한 경계값을 나타내었으며 두 압축기법 모두 경계값이 거의 같게 나왔다. 즉 압축을 하기 전에 묵음처리가 이루어지므로 경계값의 범위와 압축기법 사이에는 큰 영향이 없었으며 적절한 경계값의 범위로는 0~155, 65380~65535이다.

2.4 묵음 구간의 처리

묵음으로 간주되는 구간에 대한 적절한 처리를 해주기 위해 송신측에서의 묵음 또는 묵음정보의 전송여부와 수신측의 묵음이나 묵음정보의 이용여부에 따라 4 가지의 경우를 생각할 수 있다.

첫째, 송신측에서 묵음 또는 묵음정보를 전송하고 수신측에서 묵음구간을 그대로 출력하는 경우. 둘째, 송신측에서 묵음정보를 전송하고 수신측에서 배경잡음을 삽입하여 처리하는 경우. 셋째, 송신측에서 묵음구간에 대한 아무런 정보를 전송하지 않고, 수신측에서 아무런 처리를 하지 않는 경우. 넷째, 송신측에서 묵음구간에 대한 정보를 전송하지 않고 수신측에서 배경잡음을 삽입하여 처리하는 경우이다. 위 4가지 경우 중, 첫번째 경우는 수신측에서는 묵음에 대한 아무런 조치를 안하므로 묵음정보를 굳이 전송할 필요가 없다. 그리고 네번째 경우는 송신측에서 묵음에 관한 어떠한 정보도 제공하지 않으므로 수신측에서 묵음구간에 대한 배경잡음 삽입 등의 처리가 어렵다. 그래서 본 연구에서는 두번째와 세번째 방법에 대해서 검토하였다.

먼저, 송신측에서 묵음구간에 대한 정보를 보내고, 수신측에서는 이 정보를 이용하여 배경잡음을 삽입하는 경우를 보면 묵음에 대한 데이터량을 최소화 하면서 묵음구간을 수신측에 알려주는 방법이 필요하다. 통화시작 직후 75ms정도까지는 음성이 포함되지 않은 배경잡음 또는 묵음 구간으로 간주될 수 있으므로[2] 송신측에서는 통화 시작 직후의 75ms 이내에 포함되는 프레임의 에너지 값을 수신측으로 전송한다. 송신측에서는 음성 구간일 경우 음성데이터를 전송하고 묵음 구간일 경우 한 바이트의 묵음 정보를 수신측으로 보내 묵음임을 알린다. 수신측에서는 음성정보가 수신될 경우에는 그대로 재생하고 묵음정보가 수신될 경우에는 통화 시작 시 송신측에서 보낸 에너지 값을 음성이 들어 올 때까지 재생한다. 이와 같은 방법의 장점은 수신측에서 전화가 끊어졌는지 통화 중인지 구별이 가능하다는 것과 묵음과 음성과의 차이를 줄임으로써 음질이 자연스러워진다는 점이다. 단점으로는 송신측에서 묵음구간을 알려주기 위한 메커니즘이 필요하며 처리 시간이 증가하고 묵음정보를 나타내는 프레임만큼의 트래픽이 발생하게 된다는 점과 수신측에서의 묵음구간에 대한 배경잡음 삽입 메커니즘이 필요하다는 점이다.

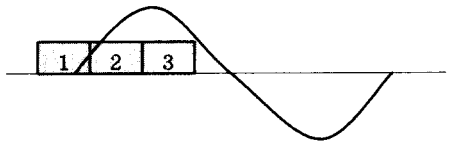
송신측에서 묵음구간에 대하여 아무런 정보를 보내지 않고, 수신측에서도 아무런 처리를 하지 않는 경우, 장점으로는 묵음구간임을 알리기 위한 메커니즘 부분이 빠지므로 처리 시간을 줄일 수 있다는 점과 묵음에 해당하는 프레임만큼의 데이터량을 줄이므로 네트워크 부하를 줄일 수 있다는 점이다. 단점으로는 수신측에 아무 소리도 들리지 않으면 끊어졌는지 통화 중

인지 구별이 불가능하다는 것과 묵음과 음성과의 차이로 인한 부자연스러움이 발생한다는 점이다. 따라서, 송신측에서 묵음정보를 전송하고 수신측에서 묵음정보를 이용해 배경잡음을 재생함으로써 자연스러운 음을 만들어 내는 것이 좀 더 음질을 높일 수 있는 방법이다.

2.5 부분 묵음의 처리

한 프레임에서 평균에너지를 이용한 묵음처리를 해줌으로써, 실제로는 음성이지만 평균치가 묵음구간에 속해 묵음으로 간주되는 경우, 이 프레임 내에 포함되는 음성의 시작부분과 끝부분의 손실에 대해 보상을 해주어야 한다. 그림2.2에서와 같이 2, 3번 프레임은 평균치가 묵음 범위를 넘어 음성으로 처리되지만 1번 프레임은 평균치가 묵음구간에 속해 음이 시작되는 부분이 포함되어 있지만 묵음으로 처리된다. 음의 손실부분은 특히 음성이 시작되는 부분에서 음의 끊어짐을 크게 느낄 수 있다. 그러므로 음의 시작부분이 손실 되는 것에 대한 처리에 중점을 두고 살펴 보기로 하자.

이와 같은 부분 묵음을 처리하기 위한 방법으로, 현재 프레임이 음성이고 이전 프레임이 묵음이면 이전 프레임과 현재 프레임을 모두 전송하고 그 이외에는 현재 프레임만을 전송하는 방법을 사용한다. 이 방법의 장점은 실제로 묵음이 아니지만 묵음으로 간주될 수 있는 음성의 시작부분에 대한 음의 끊어짐 현상을 막을 수 있다는 점이다. 단점으로는 현재 프레임과 이전 프레임을 검사하는데 소요되는 처리 시간이 증가하고 한 프레임에 해당하는 크기만큼 음성의 지연이 발생한다는 것이다. 그러나 음질이 좋아지는 효과가 크므로 부분 묵음에 대한 보상을 해주는 것이 좋다.



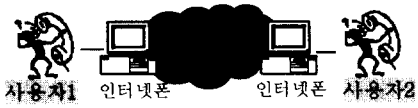
1 : 평균값이묵음의범위 내에 들어가서 묵음으로 처리되는프레임

2, 3 : 평균값이묵음의 범위를 넘어 음성으로 처리되는프레임

[그림 2.2] 부분 묵음의 발생

3. 제안한 기법의 성능 시험

제안한 묵음처리 기법의 성능을 평가하기 위하여 그림 3.1과 같이 시험망을 구성하고 인터넷폰을 이용하여 LAN 및 서울과 부산사이의 인터넷망을 통해 음질을 측정하였다.



[그림 3.1] 성능평가를 위한 시험망의 구성도

제안한 묵음처리 기법을 적용하였을 경우와 그렇지 않은 경우의 성능을 비교, 분석하기 위하여 50명의 사용자를 대상으로 MOS(Mean Opinion Score) 를 측정하였으며 음성의 명료도, 끊김 현상의 발생 빈도, 지연 등을 MOS 측정의 기준으로 하였다. MOS 값의 범위는 가장 나쁜 음질을 나타내는 1 에서 가장 좋은 음질을 나타내는 5 까지이다. 측정 결과 LAN에서는 표3.1과 같이 묵음처리를 한 경우와 하지 않은 경우에 크게 차이가 없었으며 서울과 부산사이의 인터넷망을 통한 측정에서는 표 3.2와 같은 결과를 보여 묵음처리를 한 경우의 통화품질이 더욱 우수함을 알 수 있었다.

[표 3.1] LAN환경에서의 묵음처리 기법의 성능

	명료성	끊김	지연	평균
묵음처리를 하지 않은경우	4.4	4.6	4.5	4.5
묵음처리를 한경우	4.4	4.7	4.5	4.5

[표 3.2] 인터넷 환경에서의 묵음처리 기법의 성능

	명료성	끊김	지연	평균
묵음처리를 하지 않은경우	4.1	4.0	3.9	4.0
묵음처리를 한경우	4.2	4.3	4.3	4.3

4. 결 론

본 연구에서는 인터넷전화에서 전체 음성 트래픽의 56% 이상을 차지하는 묵음을 제거해줌으로써 네트워크 트래픽을 줄여 통화품질을 향상시키기 위한 방안을 제안하고 그 성능을 분석하였다. 묵음 검출을 위해서는 평균에너지를 이용하는 방법을 사용하였으며 묵음범위를 결정하는 경계값의 에너지 레벨은 0~155, 65380~65535 임을 알 수 있었다.

그리고 묵음의 제거로 인한 음의 부자연스러움을 해결하기 위해 송신측에서 보내온 1byte의 묵음정보를 이용하여 수신측에서 배경잡음을 재생해줌으로써 자연스런 통화가 가능하였으며, 음성의 시작부분에서 묵음으로 간주되어 음이 끊어지는 현상은 현재 프레임이 음성일 경우 이전프레임을 검사하여 묵음이면 이전 프레임을 먼저 전송하고 현재 프레임을 전송하는 방법을 사용함으로써 음의 끊어짐을 보상할 수 있었다.

제안한 기법의 성능을 평가하기 위하여 시험망을 구축하여 묵음처리 전과 묵음처리 후의 통화품질을 MOS로 비교 분석한 결과, LAN에서는 유사한 통화품질을 보였으며 장거리 인터넷망에서는 묵음처리를 한 경우가 좀 더 나은 통화품질을 보였다.

참고 문헌

[1] ITU-T Recommendation G.114, "One-way transmission time" ,1996

[2] Lucy Liao and Mark A Gregory, "Algorithms For Speech Classification", Proceeding of Fifth International Symposium on Signal Processing and its Application, ISSPA '99, Brisbane, Australia, pp 22-25 August, 1999

[3] Mr. Chris Rose, Dr. Rovert W. Donaldson, " Real -time Implementation and Evaluation of an adaptive silence deletion algorithm for speech compression ", IEEE Pacific Rim Conference on communications, Computers and Signal Processing, May 9-10,1991

[4] 이정태, "음성의 묵음시간을 활용한 데이터 통신 프로토콜의 설계" 공학박사 학위논문, 서울대학교, 1989

[5] 황원주, "인터넷 전화 서비스를 지원하기 위한 Differentated Service의 성능평가", 공학석사 학위논문, 부산대학교, 2000