

VoIP를 위한 음질 평가 기준 연구

조 아 서, 박 상 옥, 박 영 철*, 윤 대 회
연세대학교 전기·전자공학과, 연세대학교 정보기술학부*

The Criterion of Speech Quality Measurement for VoIP

A Seo Cho, Sang Wook Park, Young Chul Park*, Dae Hee Youn
Electrical and Electronics Eng., Yonsei University,
Division of Information Technology, Yonsei University*

요 약

VoIP 음성 통신 시스템에서 통화를 할 때, 네트워크 상황이 나빠짐에 따라 시간 지연, 패킷 손실, 지터 등의 QoS 파라미터에 의한 영향이 발생하므로 통화 품질이 떨어지게 된다. 통화 품질을 개선하기 위해서는 통화 품질과 QoS 파라미터와의 관계를 명확히 파악하고 그에 대한 개선 방법을 연구해야 한다. 따라서 본 논문에서는 통화 품질과 QoS 파라미터와의 상관관계를 회귀 분석을 통해 도출해 내었다. 제시된 음질 평가 기준은 QoS 파라미터만을 가지고 음질을 예측하기 때문에 계산량이 매우 적으며, 음질 평가 수행 중에 음성 통신 시스템에 거의 영향을 미치지 않는다는 장점을 가지고 있다.

수요가 증가하고 있는 이 시점에서 VoIP의 단점을 보완하여 더 나은 통신 환경을 구현하기 위해서는 통화품질과 QoS 파라미터와의 상관관계를 정확하게 파악할 필요가 있다.

기존의 객관적 음질 평가 방법들은 실제 원 음성과 왜곡 음성을 사용하여야 하므로 실제 통신 시스템에서 사용하기가 어렵다. 특히 VoIP와 같이 대역폭이 제한된 음성 통신 시스템은 테스트 신호의 사용으로 그 성능이 저하될 수 있다. 따라서 본 논문에서는 QoS 파라미터와 음질과의 상관관계를 분석하였다. 제안된 방법은 QoS 파라미터만을 고려하므로 계산량이 매우 작으며 음질 평가 중 음성 통신 시스템에 영향을 거의 주지 않는다는 장점을 가지고 있다.

I. 서 론

최근 기존 전화망이 아닌 데이터 망을 통해 음성 통화를 이용할 수 있는 새로운 통신 기술인 VoIP(Voice over Internet Protocol)의 수요가 급증하고 있다. VoIP 기술은 데이터 망을 이용하여 음성통화를 가능하게 함으로써 기존 전화망에 소요되던 비용을 혁신적으로 줄일 수 있는 기술이다. 따라서 VoIP는 통신비용을 줄일 수 있다는 장점 때문에 그 수요가 급격히 늘고 있을 뿐만 아니라, 그동안의 대부분의 수익을 국제 전화나 시외 전화 등에 의존했던 통신 사업자나 음성 교환기 위주의 통신 장비를 개발해 판매해 왔던 업체들도 변화를 적극 수용하려는 움직임을 보이고 있다. 하지만 VoIP는 완벽한 QoS(Quality of Service)가 보장되지 않는다는 단점을 가지고 있다. 즉, 네트워크 상황에 따라 통화품질이 저하될 수 있다는 것이다. 통화 품질을 저하시키는 요인으로 지연, 지터, 패킷 손실, 반향 등의 QoS 파라미터를 지적할 수 있다. VoIP의

II. 음성 신호 데이터베이스 구축

본 논문에서의 실험을 수행하기 위해 다양한 네트워크 상황을 QoS 파라미터, 즉, 지터와 패킷 손실, 패킷 사이즈, 그리고 지터의 분포도를 변화시켜가면서 왜곡된 음성 신호 데이터를 만들었다.

1. VoIP 환경에서의 QoS 파라미터

일반 전화와 달리 인터넷을 이용한 통신에서는 저 전송률 부호화기를 사용하고, 네트워크의 상황에 따라 패킷 전송 경로가 매번 바뀔 수 있기 때문에 음질 저하가 일어날 수 있다. 그 뿐만 아니라 네트워크 상황 악화, 장비상의 문제 등에 대해서도 음질 저하는 발생하게 된다. 음질에 영향을 주는 각 상황을 파라미터로 나타낸 것을 QoS 파라미터라고 한다. 데이터 망을 통한 음성 통신에서 음질에 영향을 미치는 QoS 파라미터는 대단히 많지만 본 논문에서는 네트워크 상황에 따른 음질 저하에 대해 논의할 것

* 본 연구는 삼성 전자의 지원에 의하여 수행되었음.

이므로 지연(delay), 지터(jitter), 패킷 손실(packet loss) 등의 네트워크 파라미터에 대해서만 고려하였다.

지연은 입력 신호의 변화에 대하여 출력 신호의 변화가 즉각 그것에 따르지 않고 시간적으로 늦어지는 현상을 의미하는 것이다. 이는 음질에 직접적인 영향을 미치지 않지만 실제 음성 통화에서 양방향 지연이 400ms 이상 발생할 경우 원만한 의사소통이 어렵게 된다. 또한 지연의 변화율을 나타내는 지터는 송신부에서 일정한 간격으로 패킷을 전송하였을 때 수신부에 도달하는 패킷은 송신부에서 전송한대로 일정한 간격이 아닌 각기 다른 간격으로 도달하게 되는 현상을 의미한다. 이는 라우터에서의 큐잉 지연과 각 음성 패킷들이 다른 경로로 전송되기 때문에 일어나는 현상이다. 따라서 제 시간에 도달하지 못한 패킷은 복호화되지 못하므로 음질에 역시 영향을 미치게 된다. 마지막으로 패킷 손실은 라우터가 혼잡할 때, 즉, 라우터가 처리할 수 있는 패킷의 수보다 더 많은 양의 패킷이 도달했을 때 라우터가 패킷을 버림으로써 발생하는 현상이다. VoIP에서는 UDP를 사용하기 때문에 손실된 데이터를 복원하지 않는다. 따라서 음성 통신에서의 패킷 손실은 음질에 크게 영향을 미친다.

2. 실험 장비

실제 네트워크 상황에 따라 왜곡된 음성을 얻기 위해 실험에 사용한 장비는 녹음한 음성 데이터를 저장하기 위한 PC, 네트워크 상황을 제어해줄 수 있는 리눅스 기반 네트워크 에뮬레이터인 NISTNET (National Institute of Standards and Technology), VoIP 게이트웨이인 SMG-400, 그리고 실제 통화와 유사한 환경을 만들기 위한 일반 전화기로 이루어져 있다.

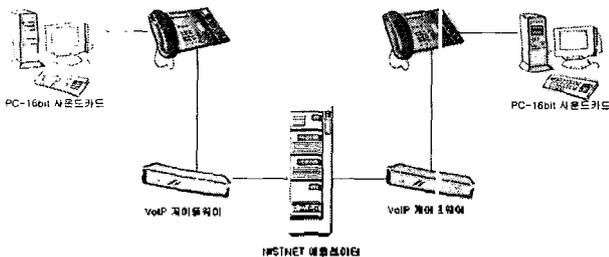


그림1. 실험 수행을 위한 음성 통신 시스템 블록도

3. 녹음 조건

본 논문에서 이루어지는 음질 평가에 이용되는 음성 신호는 ITU-T Rec. P.830과 ITU-T Rec. P.800의 형식과 녹음 기준에 따른 NTT-AT의 "Multi-Lingual Speech Database for Telephony 1994"에서 한국인 여자 화자 2명 그리고 한국인 남성 화자 2명의 음성 신호를 선택하였다. 녹음을 하기 위한 음성 부호화기로는 G.723.1 6.3 kbps와 G.729A 8.0 kbps를 사용했으며 녹음 작업을 수행

하기 전에 먼저 실험에 사용할 네트워크 상황을 지정해야 하는데 본 논문에서는 실측 데이터를 바탕으로 하여 네트워크 상황을 정리하였다.

실측 데이터는 국내 사이트 간, 혹은 국외 사이트간의 ping을 측정한 데이터로 국내 데이터와 국외 데이터로 나뉜다. 각 데이터들은 측정된 시간이나 날짜가 서로 다르므로 네트워크의 트래픽 또한 데이터별로 다르게 나타난다. 전체적으로 국내의 데이터를 종합하여 분석한 결과 평균 지연이 15ms에서 350ms정도, 지연의 표준편차가 0ms에서 70ms정도, 그리고 패킷 손실이 0에서 50% 정도 사이의 값을 보인다. 일반적으로 국내 데이터의 시간 지연이나 지터, 패킷 손실율의 값이 크기가 작은 편이고 국외 데이터의 경우는 국내에 비해 큰 값을 나타냄으로써 네트워크 상황이 좋지 않음을 보인다. 본 실험에서 네트워크 상황을 추출하기 위해 국내 데이터 50개와 국외 데이터 20개를 분석하여 지터 값이 주로 나타나는 0에서 20ms에 대해서는 조밀하게 나누고 나머지 20ms에서 70ms사이는 간격을 넓게 나누었다. 또한 패킷 손실 값이 주로 나타나는 0에서 10%에 대해서는 역시 조밀하게 나누어 네트워크 상황을 추출하고 나머지 20ms에서 50ms까지는 넓은 간격을 두고 네트워크 상황을 추출하였다. 또한 패킷 사이즈의 영향을 고려하기 위해 각 음성 부호화기의 경우에 대해 1배수, 2배수 그리고 3배수까지 녹음하였고, 평균 지연은 한 방향으로 150ms로 지정하였다. 따라서 여러 가지 네트워크 상황에 대해 총 3120개의 음성 샘플을 녹음하였다

각 네트워크 상황을 부여하여 왜곡된 음성 신호들을 컴퓨터 사운드 카드를 통해 원 신호와 동일한 조건인 8 kHz, 16 bit 표본화율로 변환시켰다. 또한 수신된 음성 신호는 전송 과정에서 에너지가 다소 달라질 수도 있기 때문에 원음과 동일하게 26 dBov(dB overload)로 정량화 하여 객관적 음질 평가 시, 에너지 차이가 왜곡으로 계산되지 않도록 하였다.

III. 음질 평가

각 QoS 파라미터와 음질과의 상관관계를 도출해 내기 위해 네트워크 상황에 맞추어 왜곡된 음성 신호에 대해 음질 평가를 수행해야 한다. 주관적 음질 평가 방법은 제 감 음질과 직접적인 관계가 있으나 많은 시간과 노력, 비용이 소모되므로 사용자의 음질 평가와 상관관계가 높은 객관적 음질 평가 방법으로 주관적 음질을 예측하는 것이 바람직하다.

본 논문에서는 다양한 QoS 파라미터와 음질과의 상관관계를 도출하기 위한 기준 음질 평가 방법으로 주관적 음질평가 대신 객관적 음질 평가법인 PESQ(Perceptual Evaluation of Speech Quality)를 사용하였다. PESQ는 아날로그 연결, 음성 부호화기, 패킷 손실, 가변 지연 등 다양한 실제 네트워크 상황을 고려하여 음질을 평가하기 때문에 기존의 심리 음향 모델에 비해 네트워크의 영향을 받아 왜곡된 음성 신호에 대해 주관적 음질과 상관관계가

높은 음질 평가 법이므로 본 연구에서 구축한 음성 데이터베이스의 기준 음질 평가법으로 적합하다.

IV. QoS 파라미터와 음질과의 상관관계 도출

왜곡된 음성 데이터베이스에 대한 객관적 음질 평가 PESQ를 수행하여 각 데이터에 대한 PESQ MOS 점수를 계산하였다. 본 절에서는 주어진 MOS 값을 가지고 음질과 QoS 파라미터 간의 상관관계를 조사하였다.

1. 회귀 분석

회귀분석은 변수들 간의 함수관계를 분석하고 모형화하는 통계적 기법이다. 음질이 QoS 파라미터의 값에 따라 변동한다면, 이들 두 변수사이의 함수관계를 추정하여 음질을 추정할 수 있고, QoS 파라미터가 음질에 미치는 영향을 분석할 수 있을 것이다.

2. QoS 파라미터를 이용한 음질 예측

여러 가지 QoS 파라미터가 음질에 미치는 영향을 분석하기 위해서는 각 QoS 파라미터를 조합하여 얻은 네트워크 상황을 지정하고 그에 따라 왜곡된 음성 신호에 대한 음질 평가를 수행하여 음질과 각각의 QoS 파라미터와의 관계를 분석해야 한다. 본 논문에서는 음성 부호화기 G.723.1과 G.729A의 각각의 경우에 대해 패킷 손실의 10가지 경우, 지터의 13가지 경우, 패킷 사이즈의 3가지 경우를 각각 조합하여 네트워크 상황을 부여한 후 그에 알맞게 녹음된 왜곡된 음성을 가지고 음질 평가를 수행하였으며 나타난 결과를 토대로 음질과 QoS 파라미터인 패킷 손실, 지터, 패킷 사이즈와의 관계를 찾아내었다.

(1) 6.3 kbps G.723.1 음성 부호화기

본 절에서는 G.723.1음성 부호화기를 이용한 음성 통신 시스템에서 390가지의 네트워크 상황을 부여하여 왜곡 음성 데이터베이스를 구축한 후 각 음성을 음질 평가 하여 얻은 결과로 다중 회귀 분석을 통하여 음질과 QoS 파라미터 사이의 관계를 도출하였다. PESQ로 측정된 MOS (Mean Opinion Score)를 기준 음질로 지정한 후 QoS 파라미터로 기준 음질을 예측할 수 있는 함수 관계를 회귀 분석을 통해 찾아내었다.

예측 MOS를 구하기 위해 3가지의 QoS 파라미터가 사용되었다. 따라서 네트워크 상황을 측정하여 각 QoS 파라미터를 얻을 수 있으면 음질 평가를 실시하지 않고도 음질을 예측할 수 있게 되는 것이다. 다음 수식은 회귀 분석 결과로 예측 음질을 구할 수 있는 수식이다.

$$Y = -0.030 \times PL - 0.016 \times J + 0.084 \times PS + 2.691$$

여기에서 PL 는 패킷 손실율이고 J 는 지터, PS 는 패킷 당 프레임 수를 나타낸다. 1차 회귀 분석으로 찾아낸 수식을 이용하여 음질을 예측해 본 결과 중간 정도의 MOS 값은 비교적 정확히 예측되는데 비해 MOS가 낮거나 혹은 높은 경우에 대해서는 정확한 예측이 이루어지지 않는 것을 확인할 수 있었다. 따라서 QoS 파라미터를 이용하여 보다 정확한 음질을 예측하기 위해서 1차적으로 예측된 MOS를 이용하여 또 한번의 회귀 분석을 통해 최종 MOS를 예측하게 된다. 다음 수식은 최종 음질을 예측할 수 있는 식이다.

$$MOS = 0.616 \times Y^2 - 1.028 \times Y + 1.497$$

여기에서 Y 는 1차 예측 MOS, MOS 는 최종 예측 MOS이다.

Correlation Coefficient	0.9505
SEE	0.2118

표1. 회귀 분석 결과 : G.723.1

표1은 G.723.1에서의 다중 회귀 분석 결과를 나타낸다. 다음 그림2와 그림3은 도출된 수식으로 구한 예측 MOS와 기준 PESQ MOS와의 상관관계를 보여준다.

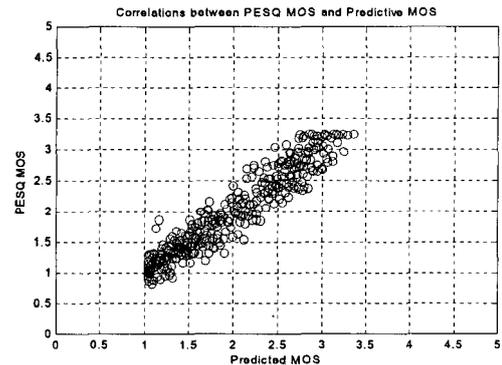


그림3. 기준 PESQ MOS와 예측 MOS의 상관 관계 : G.723.1

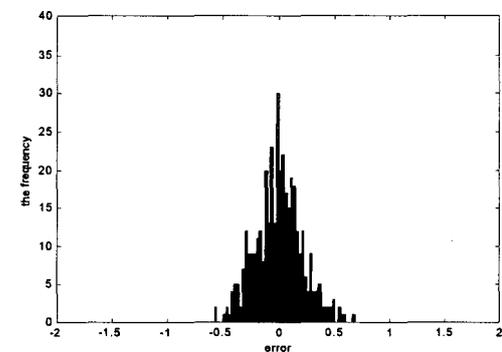


그림4. PESQ MOS와 예측 MOS간의 오차에 대한 분포도 : G.723.1

위의 그림에서 기준 MOS와 예측 MOS는 0.5 이내의 오차를 갖는 것을 알 수 있다.

(2) 8.0 kbps G.729A 음성 부호화기

본 절에서는 G.729A 음성 부호화기를 이용하여 얻은 왜곡 음성 데이터베이스를 이용하여 음질과 QoS 파라미터 사이의 관계를 도출하였다. 회귀 분석 결과 예측 MOS를 얻을 수 있는 수식은 다음과 같다.

$$Y = -0.021 \times PL - 0.017 \times J + 0.036 \times PS + 2.462$$

$$MOS = 0.444 \times Y^2 - 0.581 \times Y + 1.236$$

여기에서 PL 는 패킷 손실, J 는 지터, 그리고 PS 는 패킷 사이즈를 의미한다. 또한 Y 는 1차 예측 MOS, MOS 는 최종 예측 MOS를 나타낸다.

Correlation Coefficient	0.9495
SEE	0.1780

표2. 회귀 분석 결과 : G.729A

표2는 G.729A에서의 다중 회귀 분석 결과를 나타낸다. 다음 그림5와 그림6은 도출된 수식으로 구한 예측 MOS와 기준 PESQ MOS와의 상관관계를 보여준다.

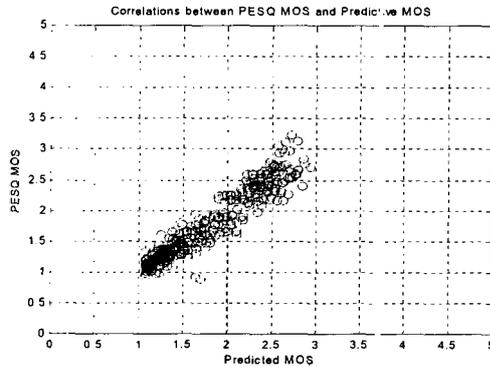


그림5. 기준 PESQ MOS와 예측 MOS의 상관 관계 : G.729A

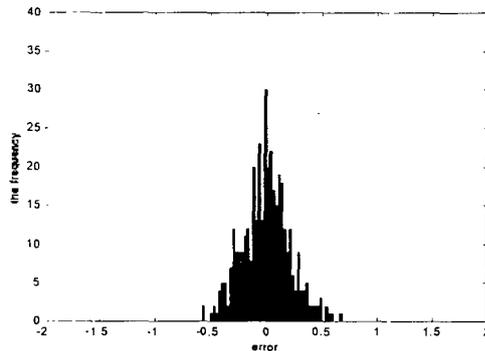


그림6. PESQ MOS와 예측 MOS간의 오차에 대한 분포도 : G.729A

역시 기준 MOS와 예측 MOS는 0.5 이내의 오차를 갖는 것을 알 수 있다.

V. 결론

본 논문에서는 실제 VoIP 음성 통신 시스템에서 음질에

영향을 주는 QoS 파라미터와 음질과의 관계를 도출하였다. 지터, 패킷 손실 등의 QoS 파라미터는 음성 통신 시스템에서 음질에 가장 큰 영향을 미치는 것으로 이들을 통해 음질을 예측하고자 하였다. 이들 음질과 QoS 파라미터 간의 관계를 도출해내기 위해 실제 음성 통신 시스템에 각 네트워크 상황을 부여하여 데이터베이스를 구축한 후, 대표적 객관적 음질 평가인 PESQ로 음질평가를 수행하였다. 음질 평가 결과를 토대로 음질과 QoS 파라미터와의 상관관계를 분석하였다. 1차적으로 분석된 결과를 확인해 본 결과 큰 MOS와 작은 MOS에 대해서 정확한 예측 음질을 얻을 수 없음을 확인하고, 보다 더 정확한 음질을 예측하기 위해 또 한번의 회귀 분석을 수행하였다. 그 결과 예측 음질과 기준 음질의 상관 계수는 G.723.1 음성 부호화기의 경우 0.9505, G.729A의 경우 0.9495 정도의 높은 상관관계를 갖는다는 것을 확인하였다.

VI. 참고 문헌

1. Antony W. Rix, John G. Beerends, Michael P. Hollier and Andries P. Hekstra, "PESQ - the New ITU standard for end-to-end speech quality assessment", *AES 109th Convension*, 2000.
2. Samir Mohamed, Francisco Cervantes-Perez, Hossam Afifi, "Integrating Networks Measurements and Speech Quality Subjective Scores for Control Purpose", *INFOCOM 2001. Proceedings. IEEE*, Volume: 2, pp. 641-649, 2001
3. Bart Duysburgh, Stefaan Vanhastel, "On the influence of best-effort network conditions on the perceived speech quality of VoIP connections", *Computer Communications and Networks, 2001. Proceedings*, pp. 334-339, Tenth International Conference on, 2001.
4. N.R. Draper, H.Smith, *Applied Regression Analysis*, Jon Wiley & Sons, New York, 1981.
5. A. Lakaniemi, J.Oosti, V.I.Raisanen, "Subjective VoIP Speech quality evaluation based on network measurements," *Communications, 2001. ICC 2001. IEEE International Conference on*, Volume: 3, 2001.
6. Thomas J. Kostas, Michael S. Borella, Ikhtlaq Sidhu, Guido M. Schuster, Jacek Grabiak, Jerry Mahler, "Real-Time Voice Over Packet-Switched Networks", *IEEE*, 1998.