

심리 음향 켈스트럼 평균 차감법을 이용한 이동 전화망에서의 음질 평가

Speech Quality Measure in a Mobile Communication System
Using PLP Cepstral Distance with CMS

윤종진 · 박상욱 · 박영철 · 윤대희 · 차일환*

(J.-J. Yun · S.-W. Park · Y.-C. Park · D.-H. Youn · I.-H. Cha)

ABSTRACT

For the set up, management and repair of a mobile communication system, continuous estimation of speech quality is required. Speech quality measurement can be conducted by listener's judgement in a subjective test such as MOS (Mean Opinion Score) test. However, this method is laborious, expensive and time-consuming, it is advisable to predict subjective speech quality via objective measures.

This paper presents a robust objective speech quality measure, PLP-CMS (Perceptual Linear Predictive-Cepstral Mean Subtraction), which can predict subjective speech quality in mobile communication systems. PLP-CMS has a high correlation with subjective quality owing to PLP (Perceptual Linear Predictive) analysis and shows a robust performance not being influenced by PSTN (Public Switched Telephone Network) channel effects due to CMS (Cepstral Mean Subtraction).

To prove the performance of our proposed algorithm, we carried out subjective and objective quality estimation on speech samples which are variously distorted in a real mobile communication system. As a result, we demonstrated that PLP-CMS has a higher correlation with subjective quality than PSQM (Perceptual Speech Quality Measure) and PLP-CD (Perceptual Linear Predictive-Cepstral Distance).

Keyword: PLP-CMS, MOS, speech quality measure, channel effect

I. 서 론

시간과 장소에 구애받지 않고 자유로운 통신을 하고자하는 인간의 욕구에 부합하여 최근 몇 년간 이동 전화가 널리 보편화되었다. 이러한 이동 전화의 품질을 결정하는 가장 중요한 요소는 통화 성공률과 통화 품질이며, 특히 통화 품질 즉, 음질은 사용자의 청각에

* 연세대학교 전기·컴퓨터공학과 음향 음성 및 신호처리 연구실

따른 주관에 의해 결정된다. 그러므로 이동 전화망의 설치 및 유지, 보수를 위해서는 지속적인 음질 평가가 반드시 필요하다.

현재 이동전화망을 통해 전송되는 음성 신호는 음성 부호화기로 압축된 디지털 신호들로 표현되며, 이 신호들이 재생되는 음성 신호의 음질에서 차지하는 중요도는 크게 차이가 있다. 그러므로 이동 전화망의 음질은 음성 신호의 특성을 고려한 방법으로 평가하여야 한다. 음질 평가는 기본적으로 여러 사용자의 반복 청취실험을 통한 주관적 음질 평가 방법에 의해 이루어질 수 있다. 하지만 이 방법은 이용자의 체감 음질과 직접적인 관계가 있다는 장점에도 불구하고 다양한 환경에서 반복해서 수행하기에는 많은 시간과 노력, 비용이 소모되므로 적합하지 않다. 그러므로 주관적 음질과 상관관계가 높은 객관적 척도로 주관적 음질을 예측하는 것이 바람직하다.

· 이 논문에서는 실제 CDMA (Code Division Multiple Access) PCS (Personal Communication System) 이동 전화망에서 얻은 음성 신호에 대하여 주/객관적 음질 평가를 수행하여 실제 전화망에서의 음질 평가를 위해서는 채널에 의한 영향을 고려할 필요가 있음을 확인하였으며, 제안된 PLP-CMS (Perceptual Linear Predictive-Cepstral Mean Subtraction)와 PLP-CD (Perceptual Linear Predictive-Cepstral Distance)[1], PSQM (Perceptual Speech Quality Measure)[2][3]에 대하여 성능 비교를 하였다.

II. 음질 평가 방법

음질 평가는 평가하고자 하는 음성 신호를 음질 평가 실험자가 직접 듣고 주관적 판단에 따라 음질을 평가하는 주관적 음질 평가 방법과 계산에 의한 원 음성 신호와 왜곡 음성 신호의 대수적 차이를 이용하는 객관적 음질 평가 방법이 있다.

1. 주관적 음질 평가

주관적 음질 평가는 양방향 청취에 의한 대화식 평가법과 단방향 청취에 의한 청취식 방법이 있다. 이중 대화식 평가법은 완벽한 통신 시스템이 완성된 단계에서만 이루어질 수 있으며, 음성 부호화기의 개발 단계나 이동 전화망의 설치 단계등에서는 적용이 불가능하다. 그러므로 수행이 용이한 청취식 평가법이 보편적으로 사용된다.

청취식 평가법에는 왜곡된 음성 신호만을 듣고 평가를 내리는 절대 음질 평가법(ACR: Absolute Category Rating), 원 음성 신호와 왜곡된 음성 신호를 듣고 왜곡 정도를 평가하는 왜곡 음질 평가법(DCR: Degradation Category Rating)과 두 개의 왜곡된 음성 신호를 듣고 비교하는 비교 음질 평가법(CCR: Comparison Category Rating)이 있다[4]. 본 논문에서는 실제 전화 이용 환경과 가장 유사하게 비교 음성 신호 없이 음질 평가를 실시하는 절대 음질 평가법의 MOS(Mean Opinion Score) 평가방법을 이용하였다. MOS 평가는 실험자가 왜곡된 음성 신호를 들은 후, 표 1에 제시된 바와 같이 음성 신호의 왜곡정도를 5단계로 평가하는 방법이다.

표 1. MOS 평가 등급

Rate	Speech Quality	Level of Distortion
5	Excellent	Imperceptible
4	Good	Just perceptible but not annoying
3	Fair	Perceptible and slightly annoying
2	Poor	Annoying but not objectionable
1	Bad	Very annoying and objectionable

2. 객관적 음질 평가

기존의 객관적 음질 평가 방법은 평가가 이루어지는 영역에 따라 시간 영역에서의 평가, 주파수 영역에서의 평가 및 심리 음향 영역에서의 평가로 분류할 수 있다. 시간 영역에서의 평가는 신호대 잡음비(SNR: Signal-to-Noise Ratio), 구간별 신호대 잡음비(SegSNR: Segmental SNR)가 있다. 이 방법들은 원 음성 신호와 왜곡된 음성 신호간의 자승 평균 오차를 이용하는 방법으로 32kbps이상의 전송률을 갖는 파형 부호화기의 성능 평가에 적합하다[5].

주파수 영역에서의 평가는 스펙트랄 거리 (SD: Spectral Distance), 대수적 스펙트랄 거리(LSD: Log Spectral Distance)[6], 선형 예측 캡스트랄 거리(LPC-CD: Linear Predictive Coefficient-Cepstral Distance) 등이 있다. 이러한 방법 중 LPC-CD는 음성 신호의 포락선 성분의 차이를 계산하며, 16~32 kbps의 전송률을 갖는 음성 부호화기의 성능 평가에 적합하다[7].

심리 음향 영역에서의 평가는 원 음성 신호와 왜곡된 음성 신호에 인간의 청각적 특성을 반영한 심리 음향 모델을 이용하여 두 음성 신호의 왜곡정도를 측정하는 방법으로 BSD (Bark Spectral Distortion), PLP-CD (Perceptual Linear Predictive-Cepstral Distance)[1], MNB (Measuring Normalizing Blocks)[8][9], PSQM (Perceptual Speech Quality Measure) 등이 있다. 특히 PSQM과 MNB는 ITU-T의 권고안으로 PSQM은 4 kbps이상의 음성 부호화기의 성능 평가에, MNB는 4 kbps미만의 음성 부호화기와 비트 에러, 프레임 삭제가 발생하는 음성 부호화기의 성능 평가에 적합하다.

3. 지연 시간 측정 및 동기화

객관적 음질 평가 방법은 원 음성 신호와 입력 음성 신호의 시간축 동기화가 이루어지지 않았다면 입력 음성 신호의 음질과 원 음성 신호의 음질이 동일하다 하더라도 매우 왜곡된 음질로 평가할 정도로 정상적인 음질 평가를 수행할 수 없다. 그러므로 객관적 음질 평가 방법에 있어 동기화는 매우 중요하며 객관적 음질평가를 수행하기 위해서는 두 음성 파형의 정확한 시간축 동기화가 반드시 이루어져야 한다.

III. 음성 신호 데이터베이스

이 논문에서 이루어지는 주관적 음질 평가와 객관적 음질 평가에 이용되는 음성 신호

는 ITU-T Rec. P.830[10]의 기준에 따른 NTT-AT의 "Multi-Lingual Speech Database for Telephonometry 1994"에서 한국인 남성화자 2명 그리고 한국인 여성화자 2명의 음성 신호를 선택하였다. 또한 전화기의 송/수화기의 대역 특성을 고려하기 위하여 ITU-T Rec. P.48[11]의 Modified-IRS (Modified-Intermediate Reference System) 필터링 하였으며, 객관적 음질 평가에서 음성 신호의 에너지 차가 왜곡으로 계산되지 않도록 하기 위하여 -26 dBov로 정량화 하였다.

이 논문에서의 실험은 이동 전화의 사용량이 많고, 다양한 전파 환경이 있는 서울에서 실시하였으며, 가능한 모든 실제 이동 전화 사용 환경을 고려하기 위하여 차량이동 환경과 도보 이동 환경 등 모든 조건에서 고르게 실시하였다. 핸드 프리킷을 이용하여 DAT (Digital Audio Tape)에 연결된 이동 전화 단말기를 통해 원 음성 신호를 재생하여 전송하면, PSTN의 반대편에서는 수신된 음성 신호를 DAT에 녹음하였다. 이렇게 얻은 data를 두 종류로 분류하였는데, 일반 교환기뿐 아니라 사설 교환기를 더 거치는 과정에서 선형 필터링으로 인해 채널의 영향을 많이 받은 음성 신호들을 A set, 일반 교환기만을 거친 음성 신호들을 B set으로 하였다.

이렇게 얻은 CDMA 전화망에서 왜곡된 음성 신호들을 PC의 사운드 카드를 통해 원음과 동일한 조건인 16 bit, 8 kHz 표본화율로 변환시킨 후 각 set별로 왜곡 정도가 다양한 음성 신호들을 고르게 분류하여 40개씩 골라 최종적으로 80개의 음성 신호에 대하여 MOS 평가를 수행하였다. MOS 평가에는 이동 전화사용자 30여명이 참여하였다. 또한 수신된 음성 신호는 전송과정에서 에너지가 다소 달라졌으므로 원음과 동일하게 -26 dBov로 정량화하였다.

IV. 채널 환경에 의한 영향

본 장에서는 III장에서 얻은 음성 신호 데이터 베이스에 대해 실시한 주/객관적 음질 평가를 고찰한 뒤, 이러한 음성 신호에 대해 기존의 음질 평가 방법인 PSQM과 PLP-CD가 보이는 성능 저하에 대해 기술한다.

1. 채널 영향에 따른 음성 신호의 왜곡

일반적인 이동전화 환경에서 음성 신호는 주로 그림 1과 같은 경로를 통해 전달된다. 또한 양쪽이 모두 이동 전화 환경인 경우 왜곡의 발생이 어느 이동 전화의 환경으로 인한 것인지 알 수 없기 때문에 음질 평가 실험 역시 그림 1과 같은 경로를 통해 음성 신호를 전송하게 된다.

이 과정에서 디지털 시스템 영역에서는 음성 패킷 손실에 의한 음질 저하가 발생하며, 아날로그 시스템 영역에서는 PSTN 경로 과정에서 선형 필터링으로 인해 스펙트럼 왜곡 등의 음질 저하가 생기게 된다[16][17]. 이러한 왜곡 중 PSTN 경로에 따른 채널 영향에 의한 음질 저하는 주관적 음질에 크게 영향을 미치지 않으며, 모든 시간 영역에 존재하는 특성으로 인해 전화 사용자들에게 있어서는 미미한 왜곡으로 인식되거나 심지어 전혀 인식되지 않기도 한다. 예를 들면, 서로 다른 장소에서 사용하는 전화에서 통화를 하더라도

사용자는 통화하는데 있어 큰 차이를 느끼지 못하며, 심한 경우에도 그다지 통화 자체에 어려움을 겪지는 않는다. 그러나 이러한 차이는 주파수 영역에서 음성 신호를 비교하여보면 그림 2와 같이 서로 다른 특성을 갖고 있음을 알 수 있다.

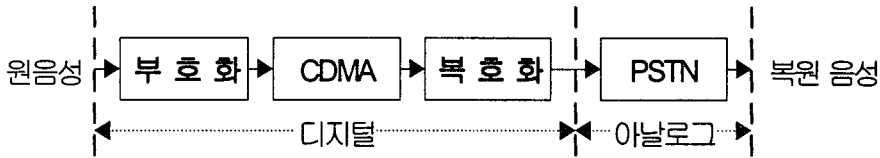


그림 1. 이동 전화 경로

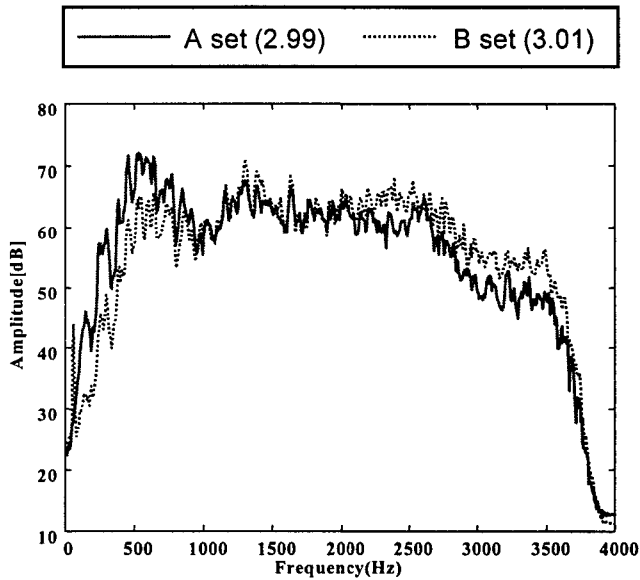


그림 2. 채널 환경에 따른 음성 신호의 스펙트럼 변화

이 논문에서의 실험에 의하면 그림 2와 같이 다른 채널 경로의 영향으로 인해 스펙트럼의 포락선이 크게 다른 두 음성 신호에 대하여 실험자들은 거의 동일한 MOS 점수를 부여함을 확인하였다. 그러나 그림 2와 같은 스펙트럼의 차이는 객관적 음질 평가에 큰 영향을 미치며 성능 저하를 야기할 수 있다. 그림 2는 A set의 MOS 점수 2.99인 음성 신호의 스펙트럼과 B set의 MOS 점수 3.01인 음성 신호의 스펙트럼을 보이고 있다.

이러한 현상으로 인해 PSTN 경로가 다른 것 이외에는 모든 녹음 환경이 동일한 A set과 B set의 음성 신호에 대해 MOS와 PSQM 및 PLP-CD 결과의 분포도 그림 3을 보면, 각 set의 음성 신호들은 유사한 MOS 점수의 분포를 보이고 있으나 PSQM과 PLP-CD의 결과 값들은 각 set에 따라 확연히 이원화된 분포를 보인다. 그림 3에서 가로

측은 MOS 점수, PSQM 값, PLP-CD 값을 나타내며, 세로축은 각 값의 발생 빈도를 나타낸다. 단, MOS 점수는 점수가 클수록 양질의 음질이나 PSQM 값과 PLP-CD 값은 왜곡된 음성 신호의 왜곡 정도를 나타내므로 값이 작을수록 좋은 음질임을 의미한다.

그림 3의 결과에 따르면 실험자들은 A set과 B set에 대하여 유사한 MOS 점수를 부여했으나 기존의 객관적 방법들은 채널 경로에 의한 A set과 B set의 스펙트럼 차이로 인해 A set과 B set에 대하여 매우 상이한 점수로 계산함을 알 수 있다. 즉, PSQM의 경우 B set에서 음질이 가장 좋지 않은 음성 신호는 PSQM 값이 5이나 이는 A set에서는 음질이 가장 좋은 값에 해당하여, A set의 MOS 4점의 음성 신호와 B set의 MOS 1점의 음성 신호와 유사한 음질의 음성 신호로 평가하고 있다. PLP-CD 역시 동일한 결과를 보이고 있다.

그러므로 PSQM과 PLP-CD의 경우 혼련 과정이 이루어진 채널 환경이 아닌 다른 환경에서 성능을 평가 할 경우 채널 영향의 변화로 큰 성능 저하가 발생하게 된다. 이러한 현상은 음성 부호화기의 성능 평가를 목적으로한 기존의 음질 평가 알고리즘에서는 고려하지 않은 점으로 음질 평가 알고리즘의 성능에 큰 영향을 끼치며, 전화망에서의 음성 인식에서는 이미 주지되고 있다[12].

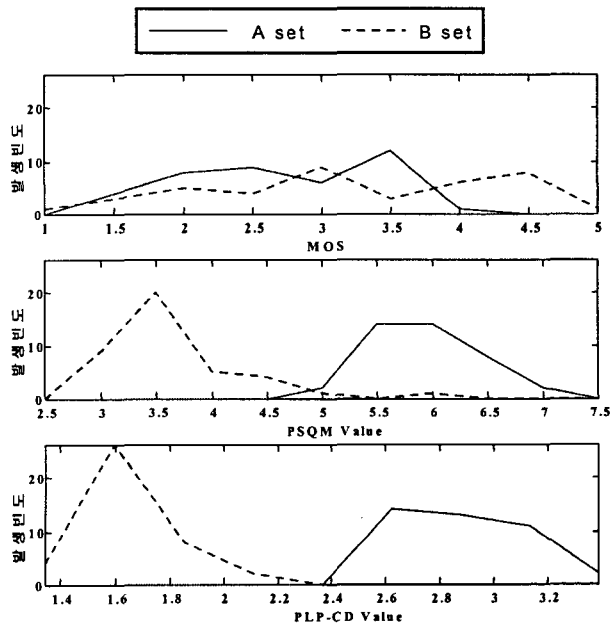


그림 3. CDMA 음성에 대한 MOS와 PSQM, PLP-CD의 분포

2. 채널 환경에 따른 기존 알고리즘의 성능 비교

서로 다른 채널 경로의 음성 신호, A set과 B set 각각에 대하여 PSQM와 PLP-CD의 객관적 음질 평가 결과를 비교하면 그림 4, 5와 같다. 그림 4, 5에서 r 은 상관 계수를 나타내며, 곡선은 회귀 분석에 이용된 함수의 곡선을 나타낸다.

PSQM 방법의 경우 A set에 대해서는 0.79, B set에 대해서는 0.92의 매우 높은 주관적 음질과의 상관 관계를 보이고 있다. 또한 PLP-CD의 경우 A set에 대해서는 0.51의 상관 관계를, B set에 대해서는 0.80의 상관 관계를 보인다

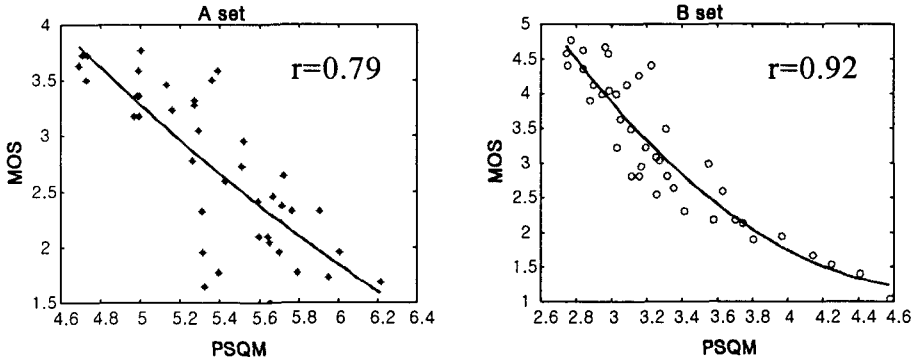


그림 4. CDMA 음성 신호 A set과 B set 각각에 대한 MOS 점수와 PSQM 결과의 상관도

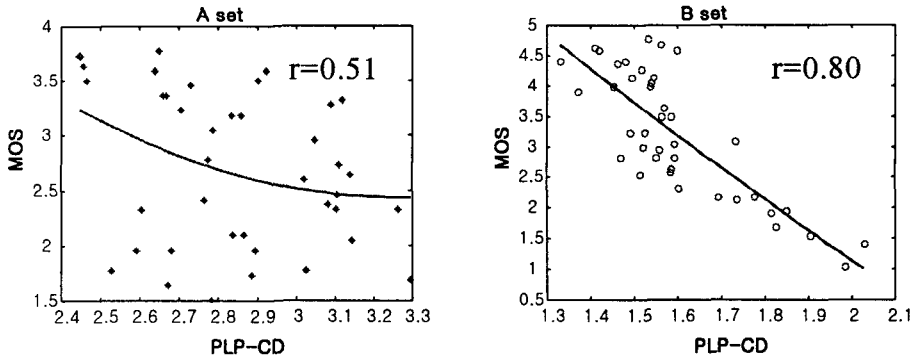


그림 5. CDMA 음성 신호 A set과 B set 각각에 대한 MOS 점수와 PLP-CD 결과의 상관도

PLP-CD의 경우 상대적으로 채널의 영향을 많이 받은 A set의 음성 신호에 대해 낮은 상관 관계를 갖는 것으로 보아 채널 환경이 좋지 않은 환경에 대해서는 더욱 나쁜 성능을 보임을 예상할 수 있다.

이러한 기존의 두 가지 객관적 음질 평가 방법에 대해 A set과 B set 모두에 대해 주관적 음질과의 상관관계를 비교하면 그림 6, 7과 같다. 이 경우 PSQM 방법은 그림 6에서와 같이 0.59, PLP-CD는 그림 7에서와 같이 0.49의 매우 낮은 상관 관계를 보이고 있으며, 이는 앞의 각 set에 대한 상관계수와 비교할 때 채널 환경의 변화에 따른 영향으로 극심한 성능저하가 발생했음을 알 수 있다. MOS 결과와 객관적 방법의 결과의 상관도 그림

6, 7을 보면 그림 4와 그림 5에서와는 달리 분포점들이 이원화되어 있다. 이는 A set에 대한 분포와 B set에 대한 분포점들이 이원화된 것으로 그림 2와 그림 3에서 관찰한 바와 같이 채널 경로에 따른 스펙트럼의 변화가 각각 다르기 때문에 비롯된 것이다. 따라서 기존의 음질 평가 방법인 PSQM과 PLP-CD는 채널 경로가 다양한 이동 전화 환경에서 사용할 수 없으며 이에 대한 대안이 필요하다.

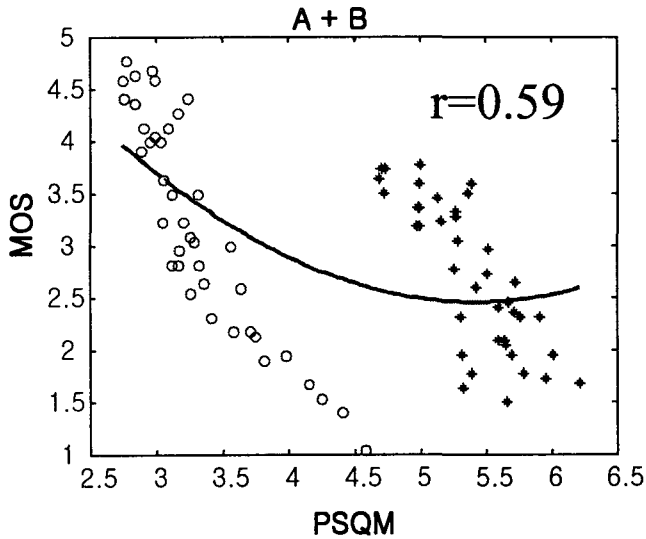


그림 6. CDMA 음성 신호 A set과 B set, 모두에 대한 MOS 점수와 PSQM 결과의 상관도

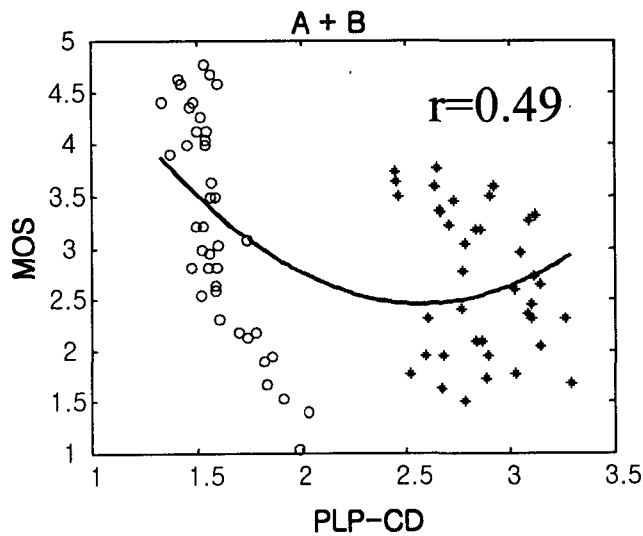


그림 7. CDMA 음성 신호 A set과 B set, 모두에 대한 MOS 점수와 PLP-CD 결과의 상관도

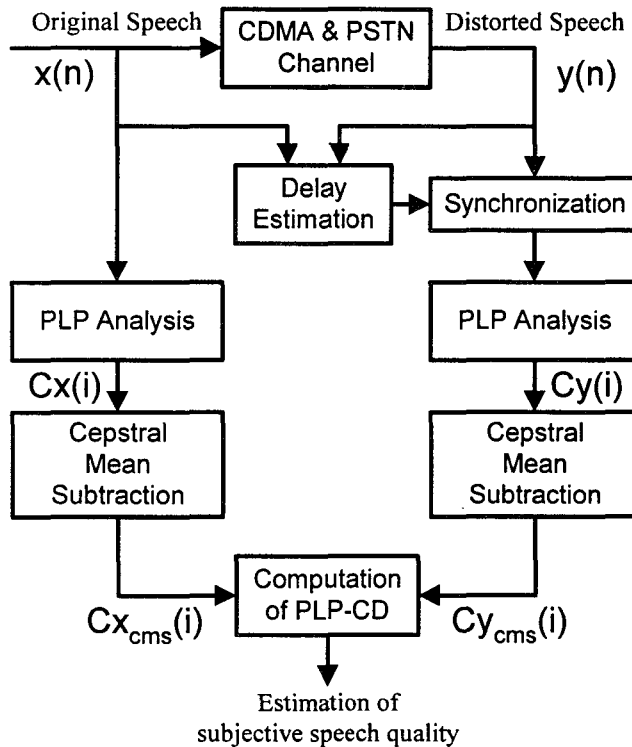


그림 8. 제안된 PLP-CMS의 블록도

V. 음질 평가 알고리즘

IV장의 실험에서 본 바와 같이 기존의 음질 평가 방법은 채널 경로의 영향으로 성능 저하가 발생하므로 본 논문에서는 이러한 다양한 채널 환경에도 일관된 성능을 갖는 PLP-CMS를 제안한다. PLP-CMS 알고리즘의 개괄적 블록도는 그림 8과 같으며, 계산 과정은 지연 추정, 심리음향 선형 예측 그리고 캡스트랄 평균 차감(Cepstral Mean Subtraction: CMS) 및 거리 계산의 네 단계로 이루어졌다.

PLP-CMS는 원 음성에 대한 왜곡 음성의 지연 시간을 측정하여 두 음성 신호를 동기화 시키며, 입력 음성 신호의 스펙트럼에 인간의 청각적 특성을 반영한다. 또한 캡스트랄 평균 차감단에서는 음성 신호 전구간에 걸쳐 영향을 주고 있는 채널에 의한 영향을 제거한다. 마지막으로 두 음성 신호의 PLP 캡스트랄 거리를 계산하여 입력된 음성의 왜곡 정도를 측정한다.

1. 지연 추정

일반적으로 두 음성 신호의 상호 상관으로 입력신호의 지연 시간을 측정한다. 그러나 CDMA PCS 음성 신호는 음성 부호화기, 채널 에러와 PSTN의 복합적인 영향으로 원 음

성 신호의 특성이 유지되지 않으며, 부분적으로는 음성 신호가 전혀 복원되지 않는 경우도 있다. 따라서 본 논문에서는 ANSI Standard T1.801.04-1997[13]의 알고리즘을 실험에 적합하게 개선하여 이용하였다.

원 음성 신호와 왜곡된 음성 신호의 포락선을 맞추어 넓은 영역에서 지연 시간을 측정 한 뒤, 주파수 영역에서 짧은 구간의 스펙트럼을 이용하여 미세한 지연 시간을 측정하는 ANSI의 알고리즘은 파형이 완벽하게 보존된 신호에 대해서는 정확한 지연 시간을 계산하나, 짧은 구간의 스펙트럼을 비교하기 때문에 CDMA PCS 음성 신호와 같이 다양한 형태의 왜곡을 지닌 음성 신호의 지연 시간 측정에 적합하지 않다. 실제로 실험 결과 ANSI의 알고리즘은 CDMA 망에서 왜곡된 음성 신호의 지연 시간 계산에 있어 많은 오차를 보였다.

따라서 이 논문에서는 짧은 구간의 스펙트럼을 이용하는 대신 CDMA 망에서도 대체적으로 유지되는 음성 신호의 전 구간 파형을 이용하였다. 즉, 음성 신호의 포락선을 이용하여 넓은 영역에 대해 낮은 분해능으로 대체적인 지연 시간을 계산하여 계산량을 줄였으며, 시간 영역에서 두 음성 신호간의 전구간 상호 상관 관계를 이용하여 세밀한 지연 시간을 계산하였다. 이러한 방법을 이용하여 모든 음성 신호에 대하여 동기화를 수행할 수 있었다.

2. 심리 음향 선형 예측 분석

심리 음향 선형 예측 분석단에서는 이미 많은 실험에 의해 입증된 사람의 청각적 특성의 모델링[1],[14],[18],[20]을 이용하여 제안된 알고리즘이 보다 높은 상관관계를 갖고 주관적 음질을 예측 할 수 있도록 하였다.

- 임계 대역 분석
- 등감곡선 프리엠퍼시스
- 주관적 소리 세기의 비선형성 보정

인간의 귀는 저주파 대역에서는 세밀한 청각 분해능을 갖으나 고주파 대역에서는 그렇지 못하며, 일정한 임계 주파수 영역에서는 각 주파수별 에너지가 상호 영향을 끼친다고 알려져 있다[20]. 이러한 특성을 고려한 것이 임계 대역 분석이다[18]. 또한 사람의 귀는 주파수 별로 자극에 대해 다르게 민감하다. 예를 들면 100 Hz의 톤이 1,000 Hz의 톤과 동일한 크기로 들리려면 35 dB 더 큰 세기(intensity)를 가져야 한다. 이러한 청각적 특징을 고려하여 각각의 주파수에 대하여서도 사람에게 의해 지각되는 소리의 크기가 동일하도록 보정할 필요가 있는데, 이를 표현한 것이 등감 곡선이며, 이에 의한 소리의 단위를 phon이라 한다. 뿐만 아니라 phon 단위의 물리적 소리 크기 변화는 실제로 사람이 인지하는 sone 단위의 소리 크기의 변화와는 주파수 대역별로 다른 특성을 보이는데 이를 비선형 관계로 모델링하여 보정하였다.

3. 캡스트랄 평균 차감 및 거리 계산

제안된 PLP-CMS는 두 음성 신호간의 거리 측정 인자로 PLP 캡스트랄 계수를 이용

한다. PLP 캡스트럴 계수는 인간의 청각적 특성을 고려한 스펙트럼 포락선을 표현하는 계수이다. 그러나 LPC와는 달리 각 계수간의 상관 관계가 없어 거리 측정 계산에 용이한 Euclidean Distance 방법을 적용할 수 있으며, 객관적 음질 평가에 있어서도 우수한 성능을 보인다[1][19].

또한 PLP-CMS는 캡스트럴 평균 차감법[12]을 이용함으로써 다양한 PSTN 경로에 의해 음성 신호 전구간에서 발생하는 스펙트럼 영향을 제거하여 어떠한 PSTN 경로를 거친 음성 신호에 대해서도 신뢰할 수 있는 객관적 음질 평가를 수행하도록 하였다. IV장에서 이루어진 실험에 의하면 각각 다른 채널을 거친 음성 신호는 동일한 MOS점수를 갖더라도 스펙트럼 특성이 매우 다르며 이로 인해 음질 평가 알고리즘의 성능 저하가 발생하였다.

따라서 PLP-CMS는 다음과 같이 평균 차감법을 이용한 PLP 캡스트럴 거리 계산으로 원 음성 신호에 대한 왜곡된 음성 신호의 왜곡 정도를 식 (1)과 같이 계산한다.

$$PLP-CMS = \frac{10}{\log 10} \sqrt{2 \sum_{i=1}^N (Cx_{cms}^k[i] - Cy_{cms}^k[i])^2} \quad (1)$$

여기서, $Cx_{cms}^k[i] = Cx[i] - E\{Cx[i]\}$, $Cy_{cms}^k[i] = Cy[i] - E\{Cy[i]\}$ 이며, $Cx[i]$, $Cy[i]$ 는 원 음성 신호와 왜곡된 음성 신호의 k 번째 프레임의 i 번째 PLP 캡스트럴 계수, $E\{Cx[i]\}$, $E\{Cy[i]\}$ 는 원 음성 신호와 왜곡된 음성 신호의 모든 프레임의 평균 PLP 캡스트럴 계수, N 은 PLP 캡스트럴 차수이다.

VI. 실험 결과 및 분석

본 장에서는 주관적 음질과 객관적 음질과의 상관성 분석을 통한 객관적 음질 평가 방법의 성능 비교 방법에 대해 설명한다. 또한 제안된 음질 평가 방법, PLP-CMS의 실험 결과를 분석한 뒤, PSQM과 PLP-CD와의 성능 비교를 통해 PLP-CMS가 기존의 방법들에 비해 우수한 성능을 갖음을 보인다.

1. 주관적 음질과 객관적 음질과의 상관성 분석

객관적 음질 평가 결과로써 주관적 음질을 예측하기 위하여 식 (2)의 2차 회귀 분석 함수[15]를 사용하였다. 이를 통하여 객관적 음질 평가 결과의 범위를 MOS 점수와 동일한 범위를 갖도록 하였다.

$$\hat{y} = \beta_0 + \beta_1x + \beta_2x^2 \quad (2)$$

여기서 x 는 객관적 음질 평가 결과, \hat{y} 는 객관적 음질 평가 결과로 예측된 MOS 값이

며, β_i 는 회귀 분석 계수이다.

또한 객관적 음질 평가 방법의 성능을 평가하기 위하여 실제 MOS 값과 식 (2)로부터 예측된 MOS 값과의 상관 계수를 식 (3)과 같이 이용하였다. 상관 계수의 값이 1이면 객관적 음질이 주관적 음질을 완벽하게 예측함을 의미한다.

$$r = \sqrt{\frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}} \quad \text{단, } (0 \leq r \leq 1) \quad (3)$$

여기서 \bar{y} 는 평균 MOS 값, \hat{y}_i 와 y_i 는 예측 및 실제 MOS값이다.

2. 제안된 PLP-CMS의 실험 결과 및 성능 분석

제안된 알고리즘 PLP-CMS의 경우 그림 9과 같이, A set과 B set에 대해 각각 0.89와 0.88의 매우 높은 상관관계를 보이고 있다. 특히 A set의 상관 관계가 PLP-CD의 경우 0.51이었음에 반해 PLP-CMS의 경우 0.89로 비교적 채널 환경이 양호한 B set의 경우보다 높다. 이는 PLP-CMS가 심리 음향 모델의 적용이 적절했을 뿐만 아니라 캡스트랄 평균 차감법의 이용으로 채널의 영향을 최소화했음을 알 수 있다.

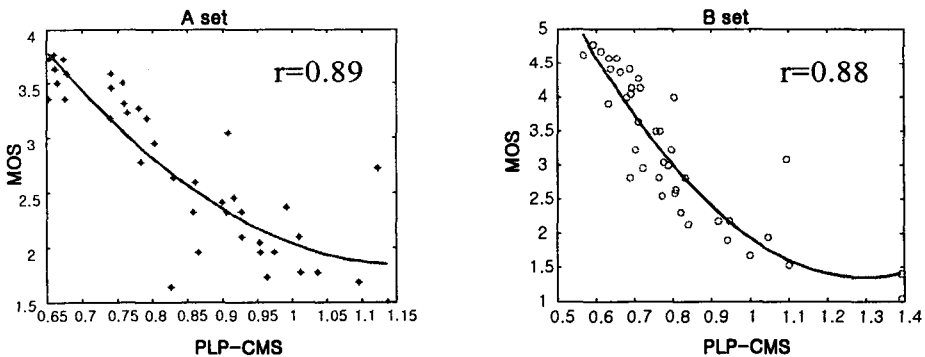


그림 9. CDMA 음성 신호 A set과 B set 각각에 대한 MOS 점수와 PLP-CMS 결과의 상관도

또한 MOS 점수와 PLP-CMS 결과의 분포도 그림 10을 보면 PSQM이나 PLP-CD와는 다르게 A set의 결과가 B set의 결과가 MOS 분포와 같이 거의 동일한 영역에 분포한다. 특히 PLP-CMS와 PLP-CD를 비교하면 그림 3에서 본 바와 같이 A set과 B set의 채널 영향이 서로 달랐음에도, PLP-CMS는 PLP-CD에 비해 캡스트랄 평균 차감법의 적용으로 채널 경로의 영향을 효과적으로 제거함을 알 수 있다.

이렇게 채널 경로가 다른 환경에 대해 고려한 PLP-CMS 방법의 A set과 B set 모두에 대한 상관도와 상관 계수를 보면 그림 11과 같다. 그림 10으로부터 예상한 바와 같이

PSQM와 PLP-CD와는 달리 A set과 B set의 분포가 이원화되지 않고 하나의 회귀 분석 함수의 곡선을 중심으로 일관되게 분포하며 이에 따라 PSQM과 PLP-CD는 이 경우 0.59와 0.49로 매우 작은 상관 계수를 보인 반면 PLP-CMS는 0.89의 높은 상관 계수를 보이고 있다. 이로 보아 PLP-CMS는 채널 환경이 다른 음성 신호에 대해서도 일관된 성능을 갖음을 확인할 수 있다.

PLP-CMS의 성능을 평가하기 위하여 녹음한 CDMA 시스템의 PCS 음성 신호에 대하여 PSQM과 평균 차감법을 이용하지 않은 PLP-CD와 함께 객관적 음질 평가를 수행하여 객관적 음질 평가 방법의 성능 지표인 주관적 음질과의 상관 계수를 비교하였다.

PSQM의 결과를 보면 서로 다른 채널 환경의 음성 신호 A set과 B set에 대한 주관적 음질과의 상관 관계는 0.79와 0.92로 우수한 성능을 보이고 있다. 그러나 A set과 B set 모두에 대한 상관 관계는 0.59로 크게 낮아 졌다. 또한 PLP-CD의 경우 채널에 의한 영향을 많이 받은 A set에 대해 0.51의 상관 관계를, 상대적으로 채널에 의한 영향이 적은 B set에 대해서는 0.80의 상관 관계를 보이고 있다. 그리고 모든 set에 대한 상관 관계는 0.49로 크게 낮은 상관 관계를 보였다. 이는 그림 2에서 본 바와 같이 두 방법 모두 채널의 영향을 고려하지 않았기 때문이며, 이로 인해 실제 이동 전화 환경에서 이용하기에 적합치 않다는 것을 알 수 있다.

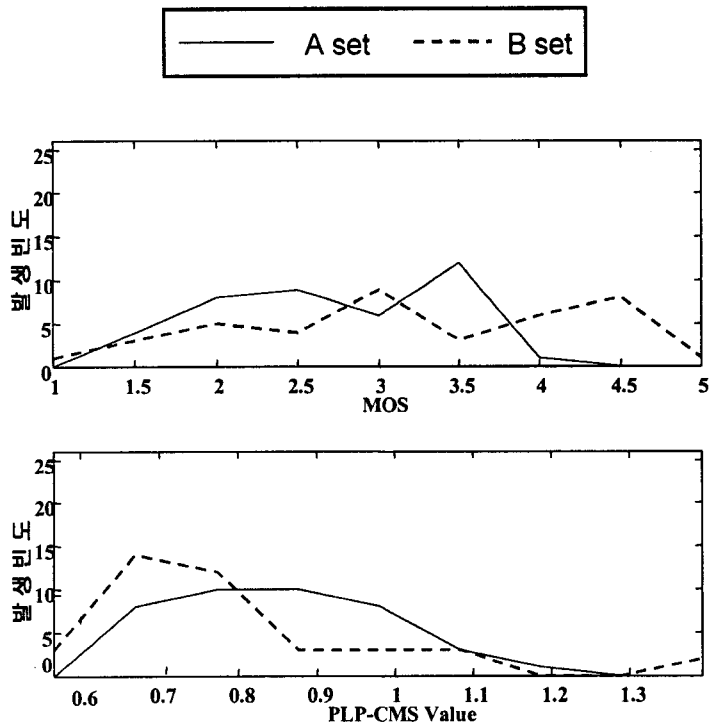


그림 10. CDMA 음성에 대한 MOS와 PLP-CMS 분포

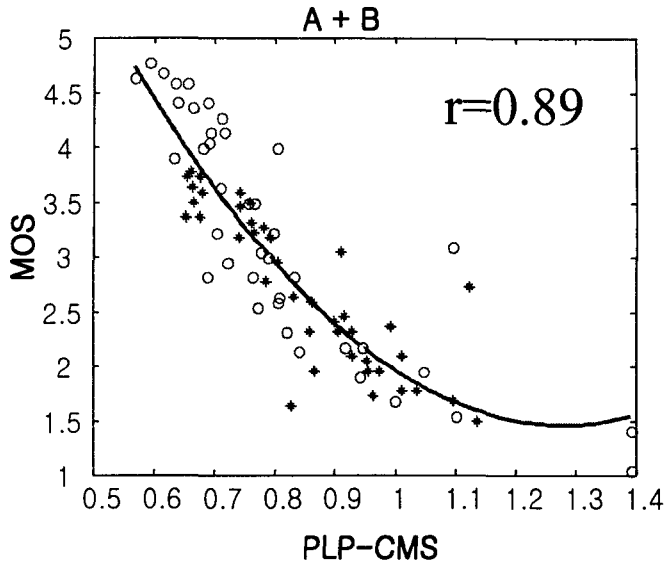


그림 11. CDMA 음성 신호 A set과 B set, 모두에 대한 MOS 점수와 PLP-CMS 결과의 상관도

표 2. CDMA 음성 신호와 객관적 음질 평가 방법별 상관 관계

	CDMA PCS		
	A	B	A+B
PSQM	0.79	0.92	0.59
PLP-CD	0.51	0.80	0.49
PLP-CMS	0.89	0.88	0.89

반면 PLP-CMS의 경우 채널의 영향의 정도에 무관하게 각 set에 대하여 PSQM과 PLP-CD보다 높거나 유사한 0.89와 0.88의 상관 관계를 보이고 있으며, 두 가지 set 모두에 대해서도 성능의 저하 없이 0.89의 상관 관계를 유지하고 있다. 실제로 PLP-CMS 결과의 분포를 보면 각 set에 대한 분포가 MOS 분포와 같이 동일한 영역에 함께 분포하고 있음을 알 수 있으며, 이는 채널 환경이 다른 음성 신호에 대하여서도 일관된 성능을 보임을 의미한다.

VII. 결 론

실제 이동 전화망에서 왜곡된 음성 신호 중 채널 경로에 의한 영향이 많은 음성 신호와 비교적 채널 경로에 의한 영향이 적은 음성 신호를 녹음하여 주관적 음질 평가를 수행하였다. 이러한 음성 신호에 대해 기존의 대표적 객관적 음질 평가 방법인 PSQM과 PLP-CD에 대해 성능을 확인한 결과 각각의 채널 경로의 음성 신호에 대하여서는 일정

이상의 상관 관계를 보이나 채널 경로에 대한 구분 없이 모든 음성 신호에 대해 실험한 결과 채널 경로에 따른 영향으로 인해 심한 성능 저하가 일어나며 이에 대한 대책이 필요함을 확인하였다.

따라서 이 논문에서는 객관적 음질 평가 방법인 PLP-CMS를 제안하였다. PLP-CMS는 심리 음향 선형 예측 분석을 이용하여 주관적 음질과의 상관 관계를 높였으며, 캡스트럼 평균 차감법을 통하여 채널 경로에 따른 영향과 무관하게 일관된 성능을 갖는다. 이는 실제 CDMA PCS 이동 전화 음성 신호에 대하여 PLP-CMS 결과와 주관적 음질과의 상관 관계로써 확인하였다. 즉, PSQM과 PLP-CD와 같은 기존의 방법들은 서로 다른 채널 경로를 거친 음성 신호에 대한 상관 계수가 0.59와 0.49로 크게 낮아졌으나 PLP-CMS는 0.89로 여전히 높은 상관 계수를 유지하였다.

본 알고리즘은 보다 다양한 이동 전화 환경에 대해 최적화가 진행 중이며, 이를 이용한 자동 음질 평가 시스템 개발도 추진 중에 있다. 또한 본 논문에서 고려하지 못한 이동 전화망의 음성 신호에서 나타나는 가변 지연(Variable Delay)등을 고려한 알고리즘의 성능 개선을 위해 연구중이다.

참 고 문 헌

- 윤종진 · 박상욱 · 박영철 · 차일환. 1999. "심리 음향 캡스트럼 평균 차감법을 이용한 CDMA망에서의 음질 평가." 제12회 신호처리합동학술대회, pp. 851-854.
- 박상욱. 1997. *이동 전화 자동 음질 평가 시스템 구현*. 연세대석사학위논문.
- 양병곤. 1997. "인간의 청각척도에 관한 고찰." *음성과학 2*, 한국음성과학회.
- 이성주 · 김희동 · 김형순. 1998. "LPC 캡스트럼 거리 기반의 천이구간 정보를 이용한 음성의 가변적인 시간축 변환." *음성과학 3*, 한국음성과학회.
- 이정학 · 김진숙. 1999. "소리의 특성 및 청지각 기능에 대한 고찰." *음성과학 5(1)*, 한국음성과학회.
- Nynek Hermansky. 1990. "Perceptual Linear Predictive(PLP) analysis of speech." *J. Acoust. Soc. Am.*, vol 87, pp. 1738-1752.
- ITU-T Rec. P.861. 1998. "Objective Quality Measure of Telephone-band (300-3,400 Hz) Speech Codecs."
- John G. Beerends., & Jan A Stermerdink. 1994. "A Perceptual Speech-Quality Measure Based on a Psychoacoustic Sound Representation." *J. Audio Eng. Soc.* vol 42(3), pp. 115-123.
- ITU-T Rec. P. 800. 1996. "Methods for Subjective Determination of Transmission Quality."
- N. Kitawaki. 1992. "Quality assessment of coded speech." *Advances in Speech Signal Processing*, S. Furui and M. Sondi, Ed. New York: Marcel Dekker, pp. 357-385.
- K. H. Lam., O. C. Au., C. C. Chan., K. F. Hui., & S. F. Lau. 1995 "Objective Speech Measure for Chinese in Wireless Environment." *Proc. of ICASSP*, vol 1, pp. 277-280.
- Nobuhiko Kitawaki., Hiromi Nagabuchi., & Kenzo Itoh. 1998. "Objective Quality Evaluation for Low-Bit-Rate Speech Coding Systems." *IEEE J. on Select Areas in Comm*, vol 6. no 2, pp. 242-248.

- Stephen Voran. 1999. "Objective Estimation of Perceived Speech Quality - Part I: Development of the Measuring Block Technique." *IEEE Trans. on Speech and Audio Processing*, vol 7, no 4, pp. 371-382.
- Stephen Voran. 1999. "Objective Estimation of Perceived Speech Quality - Part I: Development of the Measuring Block Technique." *IEEE Trans. on Speech and Audio Processing*, vol 7, no 4, pp. 383-390.
- ITU-T Rec. P.830. 1996. "Subjective Performance Assessment of Telephone-Band and Wideband Digital Codecs."
- ITU-T Rec. P.48. 1988. "Specification for an Intermediate Reference System."
- R. J. Mammone., X. Zhang., & R. P. Ramachandran. 1996. "Robust Speaker Recognition - A Feature-based Approach." *IEEE Signal Processing Mag*, pp. 58-71.
- ANSI Standard T1. 801. 04-1997. 1997. "Multimedia Communications Delay, Synchronization, and Frame Rate Measurement." New York.
- S. Wang., A. Sekey., & A. Gersho. 1992 "An Objective Quality Measure for Predicting Subjective Quality of Speech Coders." *IEEE J. on Select Areas in Comm*, vol 10(5), pp. 819-829.
- N. R. Draper., & H. Smith. 1981. *Applied Regression Analysis*. Jon Wiley & Sons, New York.

접수일자: '99. 9. 18.

게재결정: '99. 10. 15.

▲ 윤 종 진

서울특별시 서대문구 신촌동 134
연세대학교 전기·컴퓨터공학과
음향 음성 및 신호처리 연구실(우: 120-749)
e-mail: jongjin@radar.yonsei.ac.kr

▲ 박 상 욱

서울특별시 서대문구 신촌동 134
연세대학교 전기·컴퓨터공학과
음향 음성 및 신호처리 연구실 (우: 120-749)
e-mail: latest@lethe.yonsei.ac.kr

▲ 박 영 철

서울특별시 서대문구 신촌동 134
연세대학교 전기·컴퓨터공학과
음향 음성 및 신호처리 연구실(우: 120-749)
e-mail: young@lethe.yonsei.ac.kr

▲ 윤 대 회

서울특별시 서대문구 신촌동 134
연세대학교 전기·컴퓨터공학과
음향 음성 및 신호처리 연구실(우: 120-749)
e-mail: dhyoun@radar.yonsei.ac.kr

▲ 차 일 환

서울특별시 서대문구 신촌동 134

연세대학교 전기·컴퓨터공학과

음향 음성 및 신호처리 연구실(우: 120-749)

e-mail: iwcha@radar.yonsei.ac.kr