

아날로그 셀룰라 시스템을 위한 자동 음질 평가기 개발

Development of an Automatic Speech Quality Evaluator for Analog Cellular System

박 상 옥*, 최 용 수*, 정 성 교*, 윤 대 회*, 이 충 용*

(Sang Wook Park*, Yong Soo Choi*, Sung Kyo Jung*, Dae Hee Youn*, Chung Yong Lee*)

*본 연구는 SK Telecom(주) 연구비 지원에 의해 이루어진 것입니다.

요 약

본 논문에서는 아날로그 이동 전화 환경에서의, 객관적인 음질 평가 척도를 사용하여 주관적 음질을 추정하는 이동전화 자동 음질평가 시스템을 개발하였다. 이동전화의 통화 품질을 유지하기 위해서는 이동전화의 네트워크를 계속하여 체크하는 것이 매우 중요하다. 주관적 음질 평가는 사람의 체감을 직접 나타내는 것이므로 실제적인 음질을 평가하는데 중요한 척도가 되지만, 인력과 시간이 많이 소모되므로 다양한 지역에서 지속적으로 음질을 평가하는데 부적절하다. 이러한 문제를 해결하기 위하여 객관적 음질평가 척도를 이용하여 주관적 음질 평가 척도를 예측하는 자동 음질 평가 시스템이 필수적이다.

반복된 실험을 통하여 BSD(Bark Spectral Distance)가 주관적 음질 평가 척도와 높은 상관관계가 있음을 확인하였으며 원래의 음성과 이동 전화 채널을 통과한 음성과의 BSD를 측정된 후 이를 바탕으로 MOS(Mean Opinion Score)를 추정하는 자동 음질 평가 시스템(Automatic Speech Quality Evaluator)을 개발하였다.

ABSTRACT

This paper presents an automatic speech quality evaluator for Korean analog cellular phone. Traditionally, speech quality under various noise environments has been determined using subjective tests schemes such as Mean Opinion Score (MOS) and the tests are quite expensive and highly dependent on test conditions. As a result, it is required to develop an automatic quality evaluation system that can provide a good estimate of subjective quality using the easily obtained objective measure. From the results of repeated subjective and objective quality tests, it was observed that Bark Spectral Distance (BSD) based on the psychoacoustic model is most highly correlated to MOS as previously reported. In this paper, we propose a real-time automatic speech quality evaluator that predicts MOS from the measured BSD between the original speech and the transmitted speech signal.

I. 서 론

인간 생활 전반에 걸쳐 정보 및 통신 문화가 큰 비중을 차지하게 됨에 따라 시간과 장소에 구애받지 않고 자유로운 통신을 하기 위한 이동 통신 서비스에 대한 수요가 급증하고 있다. 이를 위하여 무선 매체인 전파를 이용하여 제 1세대 이동 통신 시스템으로 분류되는 아날로그 이동 전화 시스템이 상용 서비스 중에 있다. 무선 통신의 주요 수단인 이동 전화와 같은 이동 통신 시스템에서 아날로그 라디오 채널을 통해 전송되는 대부분의 음성은 채널

채널 잡음 및 배경 잡음에 의하여 왜곡된다.

이동 전화 사용자가 느끼게 되는 서비스 만족도는 주로 통화 성공률과 통화 품질로 이루어진다. 이중 통화 품질은 사용자가 귀로 직접 느끼는 부분으로 주관적 음질 평가에 의해서 측정된다. 주관적 음질 평가는 미리 녹음된 테이프를 사용하여 평가 대상자들에 의하여 직접 이루어지므로, 대도시등과 같이 다양한 전파 환경과 수많은 기지국 환경에서 주관적 평가를 수행하는 것은 많은 물적 비용과 시간을 필요로 한다.

이러한 문제를 해결하기 위하여 신호대 잡음비(Signal-to-Noise Ratio : SNR), 구간별 신호대 잡음비(Segmental SNR : SegSNR), 선형 예측 켈프스트랄 거리(Linear Predictive Coefficient-Cepstral Distance : LPC-CD)[1] 그리고 바크

* 연세대학교 전자공학과

접수일자 : 1998년 6월 12일

(Bark) 주파수 영역에서 스펙트럼의 차이를 나타내는 바크 스펙트럴 거리(Bark Spectral Distance:BSD)[2]등과 같이 쉽게 평가할 수 있는 객관적 음질평가 척도를 이용하여 DAM(Diagnostic Acceptability Measure)[3], MOS(Mean Opinion Score)[4]등과 같은 주관적 음질 평가 척도를 예측하는 연구가 활발히 진행되어 왔다[5][6][7]. S. Quackenbush[5]는 기존의 SNR, SegSNR 그리고 LPC-CD등과 같은 객관적 음질 평가 척도와 DAM과의 상관관계를 연구하였으며, Wang[6]은 인간의 심리 음향을 반영한 BSD로 MOS를 예측하는 방법을 제안하였고, 또 CELP 음성 부호화기에도 도입하였다. 그 밖에도 낮은 전송율을 갖는 음성 부호화기의 음질평가에 대하여 많은 연구가 있었다 [8][9][10].

한국의 아날로그 이동 전화 환경에서의 음질 평가 시스템에 관한 연구가 공식적으로 보고된 바는 없지만, 이동 전화 채널을 통과한 음성과 원음과의 단순한 LPC거리 측정을 이용한 음성 인식율을 바탕으로 통화 품질을 측정하는 방법이 사용되었고, 그 성능은 비교적 만족스럽지 못한 수준이라 알려져 있다[11].

본 논문의 목적은 이러한 기존의 음질 평가 시스템을 개선하여 한국의 아날로그 이동 전화 환경에 적합한 자동 음질 평가 시스템을 개발하는 것이다.

반복 음질 평가 실험을 수행한 결과, [7]에서 보고된 바와 같이, 인간의 심리 음향 모델을 적용한 BSD가 MOS와 높은 상관 관계를 가지고 있으며, SNR, SegSNR, LPC-CD등이 MOS와 낮은 상관 관계를 가지고 있음을 확인하였다. 본 논문에서는, 이러한 실험 결과를 바탕으로 원래의 음성과 이동 전화 채널을 통과한 음성과의 BSD를 측정 한 후 이를 바탕으로 MOS를 추정하는 자동 음질 평가 시스템(Automatic Speech Quality Evaluator)을 개발하였다.

지금까지의 보고된 자동 음질 평가 시스템[7][12]은 오프 라인(off-line)으로 녹음된 음성 데이터를 사용하여 주관적 음질을 예측하였으나, 본 논문에서는 제안된 자동 음질 평가 시스템은 TI(Texas Instrument)사의 TMS320C30 [13] 프로세서를 사용하여 실시간 구현하였다. 주 시스템과 디지털 신호처리부도 이루어진 자동 음질 평가 시스템은 아날로그 이동 전화 채널을 통과한 음성과 원 음성과의 BSD를 계산하고, 이를 이용하여 MOS를 예측한다.

본 논문의 구성은 다음과 같다. 2장에서 객관적 음질 평가 방법들과 주관적 음질 평가 방법에 대해서 설명하며 3장에서는 음질 평가 실험과 성능 평가에 대해서 자세히 기술한다. 실시간 음질 평가 시스템에 대하여 4장에서 설명을 한 후, 마지막으로, 5장에서 결론 및 추후 연구과제로 끝을 맺는다.

II. 음질 평가 방법

음질 평가 방법은 크게 객관적 방법과 주관적 방법으로 구분할 수 있다. 객관적 음질 평가 방법은 원 음성과 왜곡된 음성의 대수적 차이를 나타내는 것으로 SNR, SegSNR,

LPC-CD, BSD등이 주로 사용된다. 반면에 주관적 음질 평가는 원 음성과 왜곡된 음성과의 차이를 사람이 직접 듣고 평가하는 것이다. 주관적 평가는 사람의 체감을 직접 나타내는 것이므로 실제적인 음질을 평가하는데 중요한 척도가 되지만, 인력과 시간이 많이 소모되므로 다양한 지역에서 지속적으로 음질을 평가하는데 부적절하다. 그러므로 주관적 음질과 상관도가 높은 객관적 평가 방법을 사용하여 주관적 음질을 예측하는 것이 보다 바람직하다. 본 장에서는 이러한 객관적, 주관적 음질 평가 방법들에 대해서 살펴본다.

2.1 객관적 음질 평가

원 음성과 비교하여 왜곡된 음성의 음질을 평가하기 위해서는 두 음성 파형의 동기화가 필수적이다. 많은 객관적 척도가 동기화의 정확도에 따라 민감하기 때문에 시간축 정렬은 객관적 음질 평가 척도를 계산하기 전에 반드시 수행되어야 한다. 음성의 동기화 방법에는 여러 가지가 있다. 예를 들어, 원음과 왜곡된 음성의 자승 평균 오차를 최소화 되도록 동기 시킬 수 있다[7]. 이러한 방법은 제한된 메모리를 요구하는 실시간 시스템에 적용이 불가능하므로, 본 논문에서는 그림 1과 같이 의사 잡음(Pseudo Noise : PN)신호를 원 음성 앞부분에 실어 보낸 후, 원 의사 잡음 신호와의 상호상관값을 최대로 하는 점을 찾음으로써 동기화시킨다. 의사 잡음 신호는 채널 등을 통과하여도 시작점이 정확하게 일치하게 되었을 때 상호상관값의 절댓값이 최대치를 갖는 특성을 보인다.

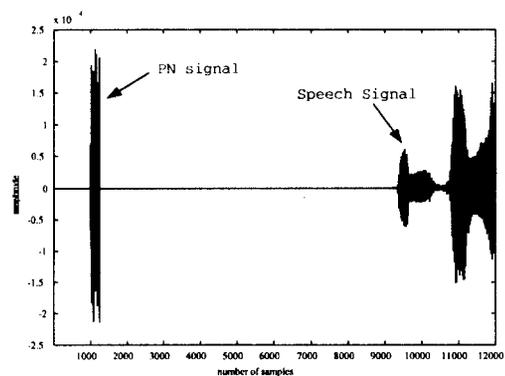


그림 1. 테스트를 위한 신호 구성

객관적 음질 평가에 주로 사용되는 SNR은 원 음성과 왜곡된 음성 파형 간의 자승 오차 평균이며 다음과 같이 정의된다[5].

$$SNR = 10 \log_{10} \frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1} [x(n) - y(n)]^2} \quad [dB], \quad (1)$$

여기서 $x(n)$, $y(n)$, N 은 각각 원음성, 왜곡된 음성, 샘플

플 수이다. SNR은 단지 원 음성과 왜곡된 음성 궤형 간의 대수적 차이를 나타내므로, 위상 왜곡이 심한 아날로그 이동 전화 시스템이나, 심리 음향을 고려한 낮은 전송률을 갖는 음성부호화기의 주관적 음질을 예측하는데는 적합하지 않다.

SNR은 음성 신호 전체에 대한 계산이므로, 음성 신호 중 에너지가 큰 부분에 크게 영향을 받는다. 그러나, 음성 신호는 에너지가 큰 부분과 작은 구간이 반복되는 비정적 신호(nonstationary)이므로 SNR을 사용하는 것 보다 구간별로 SNR을 구하여 이것의 통계적 특성을 이용하는 것이 바람직하다. 이러한 방법을 SegSNR이라 하며 다음과 같이 정의된다[5].

$$SegSNR = \frac{1}{L} \sum_{k=1}^L SNR_k \quad [dB], \quad (2)$$

여기서, SNR_k 은 k번째 구간에서의 SNR을 나타내며, L은 총 프레임 갯수이다. SegSNR은 위에서 설명하듯이 SNR에 비하여 음성 신호의 특성을 잘 반영하고 있으므로 SNR에 비하여 주관적인 음질과 상관관계가 높다.

선형 예측 계수(Linear Predictive Coefficient)는 음성의 주파수 스펙트럼 포락선 모양을 나타낸다. LPC-CD는 이 포락선 성분의 대수적 차이를 계산한 것이다[1].

$$LPC_CD \approx [2 \sum_{i=1}^K (C_x(i) - C_y(i))^2]^{1/2} \quad [dB], \quad (3)$$

여기서, $C_x(i)$, $C_y(i)$, K는 각각 원 음성과 왜곡된 음성의 LPC 캡스트랄 계수 및 차수이다.

2.1.1 심리 음향 모델을 이용한 객관적 음질 평가

바크 스펙트랄 거리 (BSD)[2][6]는 인간의 청각적 임계 대역(auditory critical band)을 기본으로 두고, 인간 청각을 모델링하는 필터 뱅크를 사용하는 주관적인 접근 방법이다. 인간의 심리 음향을 고려하므로 음질의 주관적인 평가와의 상관관계가 매우 높다고 알려져 있다[7]. 인간의 청각 특성은 800Hz 이상의 주파수에 대해서는 주파수가 증가함에 따라 청각의 분해능이 감소하고, 중간 주파수 영역에서 보다 민감하다[14][15][16]. 이러한 특성을 반영한 것이 바크 스펙트럼이다.

임계 대역 스펙트럼 분해능(critical-band spectral resolution)은 기저막(basilar membranc)의 주파수 분해능을 반영하기 위해, 음성의 파워 스펙트럼을 다음의 변환 관계식을 사용하여 헤르츠(Hertz) 주파수 영역에서 바크 주파수 영역으로 변환한다.

$$\Omega(\omega) = 6 \ln \{ |/20\pi + [(\omega/120\pi)^2 + 1]^{0.5} \} \quad (4)$$

여기서 ω 는 rad/s로 표현되는 각 주파수이다. 바크 주파수 영역으로 변환된 파워스펙트럼은 (5)와 같이 $\Psi(\Omega)$ 로 표현되는 임계 대역 마스킹 곡선과 이산 컨볼루션(discrete

convolution)를 한다.

$$\Psi(\Omega) = \begin{cases} 0 & \Omega \leq 1.3 \\ 10^{2.5(\Omega+0.5)} & -1.3 \leq \Omega \leq -0.5 \\ 10^{-\frac{1}{2}(\Omega-0.5)} & -0.5 \leq \Omega < 2.5 \\ 0 & 2.5 \leq \Omega \end{cases} \quad (5)$$

여기서 $\Psi(\Omega)$ 는 비대칭이며, 저주파수에서는 +10dB/Bark를, 고주파수에 대해서는 -25dB/Bark를 갖는다. (5)는 임계 대역 주파수 응답을 근사화한 것으로 기저막에 있는 청신경의 대역 통과 필터에 해당된다. (6)의 $F(\omega)$ 와 근사화된 임계 대역 마스킹 곡선 $\Psi(\Omega)$ 와의 이산 컨볼루션은 임계 대역 필터링을 의미한다.

$$\Theta(\Omega_i) = \sum_{\omega=-1.3}^{2.5} F(\Omega - \Omega_i) \Psi(\Omega) \quad (6)$$

주파수에 따라 다른 인간의 청각 민감도를 반영하고, 40dB 레벨의 청각 민감도를 근사화하기 위하여 다음과 같이 equal-loudness preemphasis를 수행한다.

$$\Xi[\Omega(\omega)] = E(\omega) \Theta[\Omega(\omega)] \quad (7)$$

여기서 $E(\omega)$ 는 등감곡선(equal-loudness curve)을 나타낸 것으로 아래와 같이 근사화 된다.

$$E(\omega) = \frac{(\omega^2 + 56.8 \times 10^6) \omega^4}{(\omega^2 + 6.3 \times 10^6)^2 (\omega^2 + 0.38 \times 10^9)} \quad (8)$$

Intensity-loudness power law는 소리 강도(intensity)와 지각하는 크기(perceived loudness) 사이의 비선형성을 근사화하여 phon에서 sonic으로의 변환을 의미한다. 또한 임계 대역 스펙트럼 크기의 분산을 줄여준다.

$$\Phi(\Omega) = \Xi(\Omega)^{0.33} \quad (9)$$

실제 시스템 설계시에는 계산량 감소를 위해 (6)의 컨볼루션과 (7)의 preemphasis는 파워 스펙트럼 영역에서 가중합(weight sum) 방법에 의해 하나의 과정으로 간단히 구현될 수 있다[2][6].

$$\Xi[\Omega(\omega_i)] = \sum_{\omega=\omega_{i-1}}^{\omega_i} w_i(\omega) F(\omega) \quad (10)$$

그림 2는 각 임계 대역 가중 함수 $w_i(\omega)$ 를 보여준다.

그림 3은 원 신호와 동신 채널을 통과한 왜곡된 신호와의 BSD 계산과정이다. 원 신호 $x(n)$ 과 왜곡된 신호 $y(n)$ 에 대하여 50% 중첩의 Hamming 윈도우를 씌운다. FFT를 수행하여 파워 스펙트럼 $P_x(\omega)$ 과 $P_y(\omega)$ 를 구한후, 위에서 설명한 심리 음향 모델을 이용하여 각각의 바크 스펙트럼을 구하고, 그 차이를 계산한 것이 BSD이다.

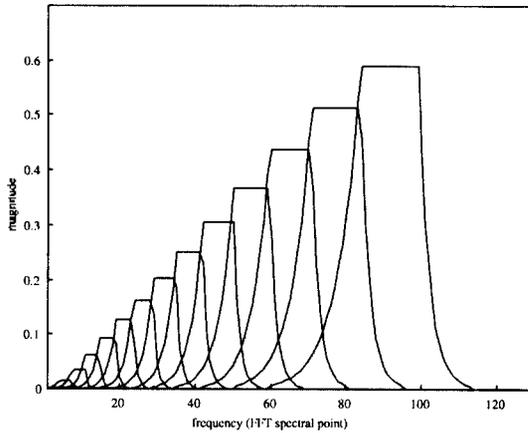


그림 2. 입계 대역 가중 함수

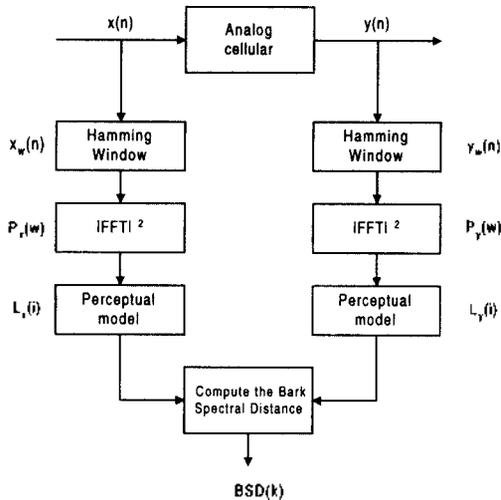


그림 3. 원 신호와 통신 채널을 통과한 왜곡된 신호와의 BSD 계산 과정

2.2 주관적 음질평가

채널을 통과한 음성을 주관적으로 평가하는 방법에는 여러 가지가 있다. 주로 사용되는 방법에는 DCR(Degradation Category Rating), CCR(Comparison Category Rating), ACR(Absolute-Category Rating)[4]등이 있다. DCR은 고품질 음질의 기준 음성과 왜곡된 음성의 차이를 5등급으로 나타내며, 왜곡의 정도가 작을 때 주로 사용된다. CCR은 DCR이 변형된 형태로 소음 제거 시스템의 평가 등 입력 음성의 음질 향상에 사용되는 방법이다. 본 논문에서는 음질의 전체적인 평가에 주로 사용되는 ACR을 사용하여 주관적인 음질을 평가한다.

ITU-T(International Telecommunication Union - Telecommunication)는 권고안 P.800[4]에서 주관적 음질 평가로 ACR을 제시하고 있다. ACR에서는 음질평가의 레벨로 주로 MOS를 사용한다. MOS는 음질손상의 척도로 5등급의 손상척도를 제시한다.

III. 음질 평가 실험

3.1 음질 평가 실험

객관적 음질 평가와 주관적 음질 평가를 위해서 남·녀 화자가 발음하여 얻은 5~8초 길이의 두 음성 신호를 사용하였다.

남·녀 아나운서가 두 문장을 릴 테이프에 녹음하였다. 원 음성을 얻기 위해 16 비트 A/D 컨버터를 사용하여 8kHz 표본화율로 다시 A/D 변환을 하였다.

다양한 왜곡 음성 데이터를 얻기 위해서, 그림 4와 같이 차량으로 장소를 옮겨가며, 원 음성 데이터를 차량에 설치된 아날로그 이동 전화 채널을 통과시켜 40회에 걸쳐 녹음하였다. 이동되는 차량에서는 노트북 컴퓨터에 저장된 음성 데이터를 반복하여 재생시켰으며, 일반 유선 전화를 사용하는 쪽에서는 멀티미디어 컴퓨터의 16비트 사운드 카드를 이용하여 8 kHz로 녹음 하였다.

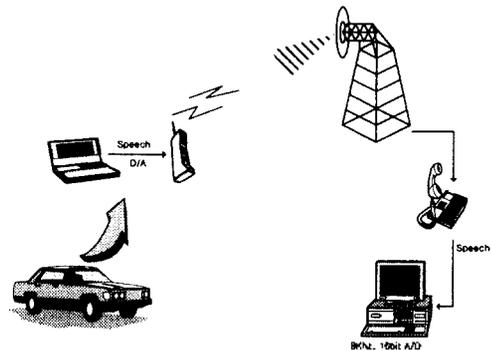


그림 4. 음질 평가 실험 환경

객관적 음질 평가 척도를 계산하기 위하여 개인용 컴퓨터를 사용하였다. MOS 테스트는 평가 환경의 변화에 영향을 받지 않도록, 20명의 청취자가 HIFI 오디오용 헤드폰을 사용하여 동시에 시행하였다. 또 청취자의 피로를 막고, 정확한 평가를 위해 1회에 30분을 초과하지 않도록 하였다.

객관적 음질평가 척도로부터 주관적 음질을 예측하기 위해서는 이차함수가 널리 사용되고 있으며 비교적 우수한 성능을 나타낸다[17].

$$\widehat{MOS} = ax^2 + bx + c \tag{11}$$

여기서 x 는 객관적 음질평가 척도이며, a, b, c 는 선형 회기(Least Square Regression) 방법[18]을 사용하여 구한다.

음질 평가 시스템의 성능 평가를 위해서 아래 식을 사용하여 상관 계수를 구한다[19].

$$r = \sqrt{\frac{\sum (\widehat{MOS}_i - m_x)^2}{\sum (MOS_i - m_y)^2}} \tag{12}$$

여기서 m_y 는 평균 MOS값, \widehat{MOS}_i 는 예측된 MOS값이

다. 위 식은 (11)을 사용하여 MOS를 예측했을 때 예측된 MOS 값과 실제 MOS값과의 오차 정도를 의미한다. $r=1$ 은 MOS 예측 오차가 없음을 의미한다.

그림 5는 남자 문장-a 경우, SNR, SegSNR, LPC-CD, BSD와 MOS와의 상관 관계를 그래프로 나타낸 것이다. 실선은 이미 실험된 객관적 음질 평가 척도와 MOS를 바탕으로 선형회귀방법으로 구한 (11)의 예측 함수이고, 각 점들은 실제 객관적 척도에 대한 MOS 값을 나타낸다. 예측 함수와 점의 분포가 일치 할수록 높은 상관관계를 나타낸다. 그림 5 (a)는 SNR과 MOS의 낮은 상관관계를 보여준다. 그림에서 SNR은 시간축 상에서 원 음성과 왜곡된 음성의 단순 곱셈의 비교이므로, 곱셈 변화가 심한 아날로그 이동 전화 채널에서 주관적 음질을 예측하는 척도로는 적합하지 않음을 알 수 있다. 그림 5 (b)는 SegSNR로서 SNR에 비하여 더 높은 상관관계를 보여준다. 그림 5 (c)의 LPC-CD는 스펙트럼 포락선의 차이를 나타내므로 위상 변화에 그다지 민감하지 않다. BSD의 경우인 그림 5 (d)는 MOS와 가장 높은 상관관계를 나타낸다.

전체 음성 데이터에 대해서 식(12)를 이용하여 SNR, SegSNR, LPC-CD, BSD등의 객관적 음질평가 척도와 주관적 음질평가 척도인 MOS와의 상관계수를 구하고 표 1에 정리하였다.

위의 표에서 표준편차 s는 실제 MOS_i에서 예측된 MOS_i를 변 오차의 표준편차이다. 표 1에서 보면, [7]에서 보고된 바와 같이 인간의 심리 모델을 적용한 BSD가

표 1. 객관적 음질 평가 척도와 MOS와의 상관 계수

		SNR	SegSNR	LPC-CD	BSD
남 문장-a	상관계수(r)	0.396	0.476	0.646	0.869
	표준편차(s)	1.195	1.144	0.993	0.643
이 문장-a	상관계수(r)	0.398	0.695	0.741	0.891
	표준편차(s)	1.307	1.024	0.819	0.646
남 문장-b	상관계수(r)	0.420	0.457	0.528	0.927
	표준편차(s)	1.426	1.398	1.334	0.588
이 문장-b	상관계수(r)	0.632	0.677	0.615	0.870
	표준편차(s)	1.187	1.141	1.207	0.754

MOS와 높은 상관관계를 가지고 있으며, SNR, SegSNR, LPC-CD등이 MOS와 낮은 상관관계를 보여준다. 결론적으로 한국어 아날로그 이동 전화 음질 평가 척도에 BSD를 사용하는 것이 적합함을 알 수 있다.

IV. 실시간 자동 음질 평가 시스템

본 장에서는 제한된 자동 음질 평가 시스템의 실시간 구현에 대하여 기술한다. 실시간 시스템은 Texas Instrument사의 TMS320C30 DSP 프로세서 1개를 사용하였으며, TMS320C30 assembly[20]를 사용하여 실시간 구현되었다. 자동 음질 평가 시스템은 디지털 신호처리부와 주 시스템의 1조로 이루어진다. 디지털 신호처리부는 음성의 A/D, D/A, BSD계산, MOS 예측 등의 신호처리 전반을 담당하

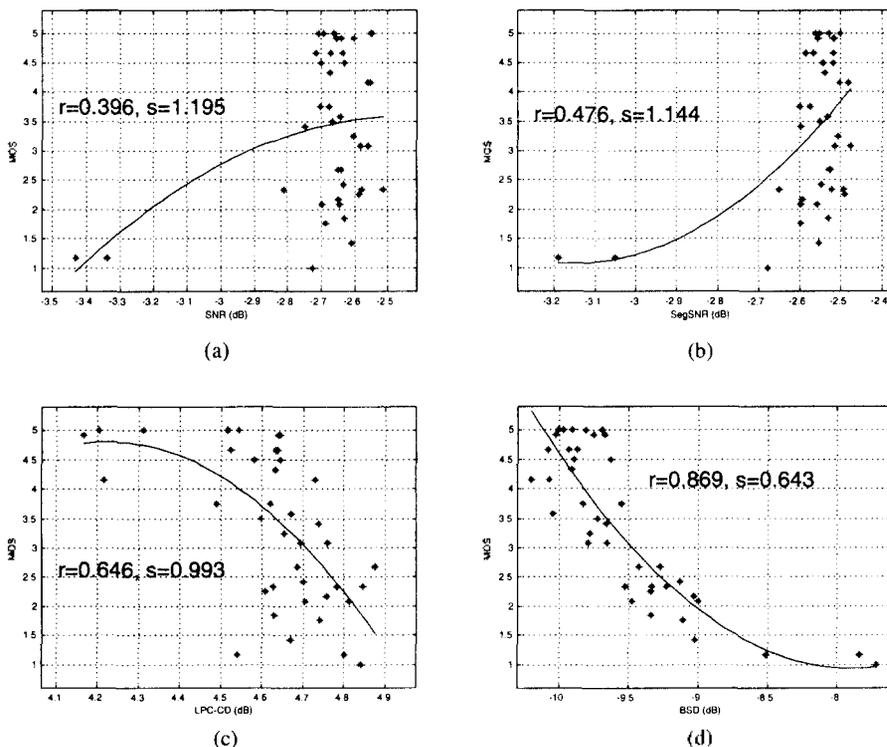


그림 5. 객관적 음질 평가 척도의 MOS와의 상관 관계; (a) SNR, (b) SegSNR, (c) LPC-CD, (d) BSD

고, 주 시스템은 디지털 신호처리부와 직렬 포트를 통하여 통신하여 디지털 신호처리부를 제어한다.

그림 6은 제안된 실시간 자동 음질 평가 시스템을 나타낸 것이다. 시스템 A에 의하여 아날로그 이동 전화 채널을 통과한 음성은 디지털 신호처리부 B에서 A/D 변환된 후 BSD를 구한다. 디지털 신호처리부 B는 계산한 BSD로 MOS를 예측한 후 주 시스템 B에 MOS값을 전달한다. 또, 이와 반대로 B에서 A로의 경로도 성립된다.



그림 6. 자동 음질 평가 시스템.

4.1 실시간 시스템 하드웨어

디지털 신호처리부는 Texas Instrument사의 TMS320C30 DSP Chip을 CPU로 사용한다. 메모리는 33 MHz TMS320C30이 제로 웨이트 상태로 동작하기 위하여 15 ns의 접근시간을 가지는 32K 워드의 SRAM과 디지털 신호처리부를 운영하는 어셈블리 프로그램과 음성 데이터를 저장하는 512K 워드의 EPROM으로 구성되어있다. 음성 데이터를 왜곡없이 깨끗하게 주고 받기 위하여 16비트 시그마델타 A/D 컨버터와 16-비트 Monolithic D/A 컨버터를 사용하였다.

BSD를 실시간으로 구하기 위한 어셈블리어는 평시에는 속도가 느린 EPROM에 저장되어있다가 프로그램 수행시에는 보다 빠른 SRAM에서 수행이 된다.

SRAM은 15ns의 접근속도를 가지고, EPROM은 120ns의 접근속도를 가지므로 4웨이트의 동작상태를 갖는다. 이 두가지 동작 상태를 동시에 만족시키기 위해서는 하드웨어적으로 RDY신호를 발생시키는 회로가 필요하다. 하드웨어 RDY신호의 발생과 어드레스 디코드를 위하여 GAL22V10을 사용하였는데, 이것은 최대 12개의 입력과 10개의 출력을 가지고 있으며, 순차회로와 조합회로 등, 두가지 모두 프로그램이 가능하다. GAL22V10 내부의 플립 플롭을 사용하여 선택적으로 SRAM을 접근할때 0웨이트, EPROM을 접근할때 4웨이트를 발생시키는 회로를 설계하였다.

디지털 신호처리부는 주 시스템과 통신을 하기 위해서 RS-232C 포트[21][22]를 사용한다. RS-232C포트는 대부분의 PC의 직렬 통신 포트로서 사용된다. RS-232C는 EIA (Electronic Industry Association)에서 제정된 규약으로 DTE (data terminal equipment)와 DCE(data carrier equipment) 사이의 기계적, 전기적, 논리적 인터페이스에 대한 표준안이다. 이것은 25pin이나 9pin의 커넥터가 사용되어지는데, 본 논문에서는 보다 간단한 구현을 위하여 TD, RD와 접지신호만을 연결하여 사용하였다. 본 논문에서는 별도의 하드웨어 없이 9600bps, 1 stop bit, no parity의 전

이중방식(Full Duplex)을 구현하였다.

4.2 실시간 시스템 소프트웨어

주 시스템에서 제어 신호를 받은 디지털 신호처리부는 자기상관값을 이용하여 시작점을 검출한다. 시작점을 찾아낸 후 BSD를 구하고 MOS 예측함수를 사용하여 MOS를 추정한 후 주 시스템에게 예측된 MOS를 보낸다. 주 시스템은 전달 받은 MOS값을 보조 기억장치에 저장한다. 그림 7은 제안된 실시간 시스템의 흐름도를 나타낸다.

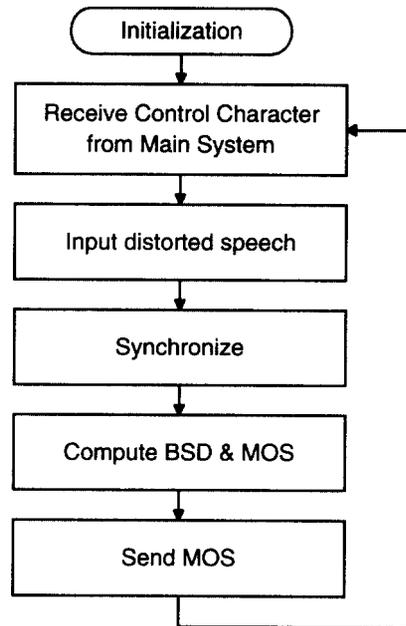


그림 7. 실시간 시스템 흐름도

본 논문에서는 빠른 주파수 스펙트럼 계산을 위해 radix-2 FFT를 사용하였고, BSD계산에 필요한 원음의 바크 스펙트럼 값들을 미리 테이블에 저장하여 불필요한 계산을 줄였다.

시간축 정렬이 완료되면, 신호처리부는 8KHz의 표본화율을 갖는 음성 데이터를 40샘플씩 버퍼링한다. 저장된 40개의 음성 데이터에 80샘플의 길이를 갖는 50% 중첩의 Hamming 윈도우를 씌우고, zero padding을 하여 256 포인트 FFT를 한다. FFT를 수행하여 얻어진 음성의 파워 스펙트럼을 15개의 임계대의 필터를 통과시킨 후 Intensity-loudness 보정을 거쳐 바크 스펙트럼을 구한다. 미리 계산된 원 음성의 바크 스펙트럼과 실시간으로 구한 왜곡된 음성의 바크 스펙트럼을 비교하여 BSD를 계산한다. 예측함수를 사용하여 예측된 MOS는 다시 직렬 포트를 통하여 주 시스템에 전달된다. 그림 8은 BSD를 계산하는 실시간 소프트웨어의 흐름도이다.

표 2에 각 루틴 당 소요시간을 정리하였다. 가장 많은 CPU 시간을 소모하는 부분은 FFT이고 프로그램 전체 CPU 점유율은 12.59%에 불과하여 남은 계산 능력에 음질 평가

와 관련된 다른 알고리즘을 사용할 수 있도록 최적화 된 것을 알 수 있다.

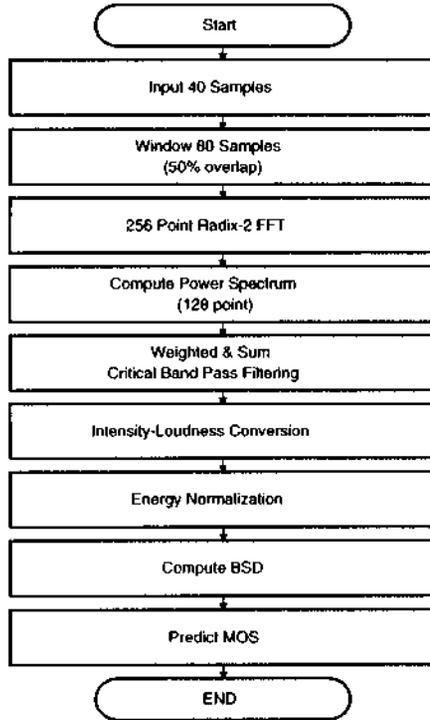


그림 8. BSD 계산 소프트웨어 흐름도

표 2. 각 루틴별 계산 시간 및 CPU 점유율

	계산시간(cycle)	CPU 점유율
256 point FFT	7870	9.54%
파워 스펙트럼	648	0.79%
임계 대역 가중 함수	546	0.66%
intensity 보정	1128	1.37%
Bark Spectral Distance	57	0.07%
에너지 정규화	62	0.08%
기타	79	0.09%
합계	10390	12.59%

V. 결 론

본 논문에서는 아날로그 이동 전화 환경에서, 객관적인 음질 평가 척도를 사용하여 주관적 음질을 추정하는 시스템을 개발하였다.

반복 실험을 통하여 인간의 심리음향 모델을 적용한 BSD가 MOS와 높은 상관관계를 가지고 있으며, SNR, SegSNR, LPC-CD등이 MOS와 낮은 상관관계를 가지고 있음을 확인하였다. 본 연구에서는 이러한 실험 결과를 바탕으로 BSD를 계산하여 MOS를 예측하는 실시간 음질 평가 시스템을 개발하였다.

자동 음질 평가 시스템 개발을 위하여 C언어를 사용

하여 컴퓨터 모의 실험을 하였으며, Texas Instrument사의 TMS320C30 프로세서 1개를 사용하여 실시간 시스템으로 구현하였다.

현재의 이동 전화 환경은 아날로그 이동 전화 시스템에서 디지털 이동 전화 시스템으로 활발히 옮겨가고 있으며, PCS(Personal Communication Service)등의 새로운 이동 통신 서비스를 제공 중이다. 이러한 디지털 이동 전화 시스템들은 아날로그 이동 전화 시스템과는 다른 특성을 가지고 있다. 그러므로, 추후 과제로, 새로운 이동 전화 시스템 환경의 효율적인 운영을 위한 자동 음질 평가 시스템의 연구가 활발히 진행 중이다.

참 고 문 헌

1. Sadaoki Furui, M. Mohan Sondhi, *Advance in Speech Signal Processing*, Dekker.
2. Nynck Hermansky, "Perceptual Linear Predictive (PLP) analysis of speech", *J. Acoust. Soc. Am.*, vol. 87, pp1738-1752, April 1990.
3. W.B. Kleijn, K.K. Paliwal, *Speech Coding and Synthesis*, Elsevier, 1995.
4. ITU-T, "Method for subjective determination of transmission quality", *Rec. P.800*, Aug 1996.
5. S. Quackenbush, T. Barnwell, and M. Clements, *Objective Measures of Speech Quality*, Englewood Cliffs, NJ:Prentice Hall, 1988.
6. Shihua Wang, et al, "An Objective Measures for Predicting Subjective Quality of Speech", *IEEE J. Select. Areas Commun.*, vol 10, No5, pp819-829, June 1992.
7. K.H. Lam, O.C. Au, et al, "Objective Speech Quality Measures for Cellular Phone", in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vol 1, pp487-490, May 1996.
8. R. Kubichek, D. Atkinson, A. Webster, "Advances in Objective Quality Assessment", *IEEE Globecom*, pp1765-1770, 1991.
9. Mohamed M. Mehy, Tarek N. Saadawi, "A Perceptually-Based Objective Measures for Speech Coders Using Abductive Network", in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, pp479-482, May 1996.
10. Simao F. Campos Neto, Franklin L. Corcoran, "Performance Assessment of Tandem Connection of Cellular and Satellite-Mobile Coders", in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, pp1379-1382, May 1997.
11. 김영렬, SK텔레콤, 개인정보.
12. K.H. Lam, O.C. Au, et al, "Objective Speech Measures for Chinese in Wireless Environment", in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol 1, pp277-280, May 1995.
13. Texas Instrument, *TMS320C3x User's Guide*, 1992.
14. 차일환, 옥성권, *소음진동공학*, 서울 문연당, 1996.
15. E. Zwicker, H. Fastl, *Psychoacoustics Facts and Models*, Heidelberg: Springer-Verlag, 1990.
16. Thomas D. Rossing, *The Science of Sound*, Addison Wesley, 1990.
17. ITU-T, "Models for Predicting Transmission Quality from Objective Measurements", *Supplement 3 (Series P)*, March

1993.

- 18. Strang, *Linear Algebra and its Applications*, second edition, Academic Press INC. 1980.
- 19. N.R. Draper, H. Smith, *Applied Regression Analysis*, John Wiley & Sons, New York, 1981.
- 20. Texas Instrument, *TMS320 Floating-Point DSP Assembly Language Tools*.
- 21. Texas Instrument, *Developing a Full-Duplex UART on the TMS320C3X*, TMS320DSP Designer Notebook, Number 58.
- 22. Hans-Peter Messmer, *THE INDISPENSIBLE PC HARDWARE BOOK*, Addison Wesley, 1994.

▲박 상 욱(Sang-Wook Park)

한국음향학회지 vol.15, no.6, 1996 참조

1998년 2월 : 연세대학교 대학원 전자공학과 졸업(공학석사)

1998년 3월 ~ 현재 : 연세대학교 대학원 전자공학과 박사과정

▲최 몽 수(Yong-Soo Choi)

한국음향학회지 vol.17, no.2, 1998 참조



▲정 성 교(Sung-Kyo Jung)

1997년 2월 : 연세대학교 공과대학 전자공학과 졸업(공학사)

1997년 3월 ~ 현재 : 연세대학교 대학원 전자공학과 석사과정

※주관심분야 : 음성부호화, 실시간 시스템

▲윤 대 희(Dae-Hee Youn)

한국음향학회지 vol.13, no.1, 1994 참조

▲이 충 욱(Chung-Yong Lee)

한국음향학회지 vol.17, no.5, 1998 참조