

마크 코히어런스 함수를 이용한 이동 전화 음질 평가

정희원 박상욱*, 류승균*, 박영철*, 윤대희*

A New Objective Speech Quality Measure Over Mobile Communication Using Bark Coherence Function

Sang-Wook Park*, Seung-Kyun Ryu*, Young-Cheol Park* and Dae-Hee Youn* *Regular Members*

요 약

음질 평가에는 주관적 음질 평가법과 객관적 음질 평가법이 있는데, 주관적 음질 평가법은 사람이 직접 듣고 평가하므로 실제 체감 음질을 나타낸다. 그러나 많은 사람들에게 의하여 직접 평가되므로 비용과 시간이 많이 소모되는 단점이 있다. 객관적 음질 평가법은 수학적인 계산에 의하여 원음과 왜곡음의 유사성을 비교하는 것으로 빠르고 비용이 적게 드나 실제 체감 음질과는 거리가 있다. 본 논문에서는 객관적 음질 평가 척도로 BCF (Bark Coherence Function)을 제안한다. BCF는 심리 음향 영역에서 코히어런스 함수를 정의한 것으로 기존의 객관적 음질 평가법에 비하여 주관적 음질과 상관관계가 높고 계산량이 적다. CDMA 이동 전화 시스템의 음성 데이터와 회기분석 결과, BCF가 ITU-T 표준안인 PSQM (Perceptual Speech Quality Measure)와 MNB (Measuring Normalizing Block)에 비하여 높은 상관관계를 갖음을 입증하였다

ABSTRACT

A new methodology for perceptual quality measure is presented. The new method defines bark coherence function (BCF) as a new cognition module. False prediction errors are occasionally observed in previously developed perceptual models when they are applied to the end-to-end speech quality measurement of communication systems. Those errors are mainly caused by the linear distortion of the analogue interface of the system being evaluated. The BCF itself normalizes those effects of linear filtering, so that it is ideal for the speech quality assessment of mobile communication systems. To evaluate the performance of the new perceptual model, the regression analysis was performed with CDMA digital cellular, CDMA personal communication service (PCS) and speech codec's. The correlation coefficients computed using the BCF showed noticeable improvements over the perceptual speech quality measure (PSQM) and the measuring normalizing block (MNB) that are recommended by ITU-T. Robustness of the BCF to various conditions was also tested

I. 서 론

이동 전화 시스템에 있어서 음질 평가는 매우 중요한 일 중의 하나이다. 음질 평가는 주로 청취자들에 의해서 직접 평가되는 주관적 음질 평가 방법과 수학적인 계산으로 얻어지는 객관적 음질 평가법이 있는데, 이중 주관적 음질평가방법은 사람에게 의해서 직접적으로 평가되므로 실제 사용자의 체감 음질을

가장 잘 표현할 수 있다. 그러나 주관적 음질 평가는 시간과 비용이 많이 소모되므로 객관적 음질 평가 척도를 사용하여 주관적 음질을 예측하는 것이 보다 바람직하다. 효과적인 객관적 음질 평가 척도는 다양한 왜곡환경과 음성 부호화 알고리즘에 대해서 주관적 음질과 상관 관계가 높아야 한다.

지난 20년동안 객관적 음질평가에 대한 여러 연구가 지속되어왔다. 음질을 평가하는 전통적인 방법으

* 연세대학교 전기전자공학과 음향 음성 및 신호처리 연구실
논문번호: , 접수일자:

로 신호대 잡음비(Signal-to-Noise Ratio : SNR)^[1], 구간별 신호대 잡음비(Segmental SNR : SegSNR)^[1] 등이 사용되었다. SNR은 시간 영역에서 원음과 왜곡음의 대수적 차이를 의미하는데 주로 파형 부호화기의 음질을 평가하는데 적합하다^[1]. 음성 주파수 스펙트럼의 포락선 차이를 나타내는 선형 예측 코스트랄 거리(Linear Predictive Coefficient-Cepstral Distance : LPC-CD)는 16k~32kbps 음성 부호화기의 음질을 평가하기 위하여 사용되었다^[2].

저전송률을 갖는 최신 음성부호화기는 나쁜 SNR에도 만족할 만한 음질을 보여주는 등 신호의 유사성에 기반한 객관적 음질평가법으로 주관적 음질을 예측하는데 부적당하다는 문제점이 있다. 최근에 와서 주관적 음질과 상관 관계를 높이기 위하여 인간의 청각적 특징을 이용하여 음질을 평가하는 방법들이 연구되었다^{[3][10]}. Hermansky^[4]는 입체 대역 스펙트럼 분해능(critical-band spectral resolution), 등감 곡선(equal-loudness curve), 주관적 소리 세기(intensity-loudness power law)를 이용하여 음성을 심리 음향 영역에서 분석하였으며, Wang^[5]은 이러한 인간의 심리 음향 영역에서 음성 스펙트럼의 차를 계산하는 BSD(Bark Spectral Distortion)로 MOS(Mean Opinion Score)를 예측하는 방법을 제안하였고, 또 CELP 음성 부호화기에도 도입하였다.

그 밖에도 낮은 전송률을 갖는 음성 부호화기의 음질평가에 대하여 많은 연구가 있었다. PSQM(Perceptual Speech Quality Measure)^[6]와 MNB(Measuring Normalizing Blocks)^[9]는 ITU-T에 의하여 권고안^[11]으로 채택되었는데 이 중 PSQM은 정교한 심리음향 모델을 적용하고 인식 효과(cognitive effect)를 사용하여, 4Kbps 이상의 전송률을 갖는 CELP 음성부호화기와 파형 음성부호화기 성능 평가에 매우 적합하다. 또, MNB는 PSQM의 적용 범위를 보완하여 비트 에러 그리고 프레임 삭제시 음성 부호화기의 성능을 평가한다.

심리 음향을 이용한 객관적 음질 평가법은 심리 음향 천이 모듈과 인식 모듈의 두 부분으로 이루어진다. 심리 음향 천이 모듈은 음성 신호를 인간이 귀에서 듣는 신호의 형태로 변환 시키는 부분을 의미하며, 인식 모듈은 천이 모듈로 변환된 원음과 왜곡음의 차이점을 정량적으로 계산하는 부분이다. 본 논문에서는 새로운 객관적 음질 평가법의 인식 모듈로 심리 음향 영역에서 코히어런스 함수^[12]를 정의한 바크 코히어런스 함수 (Bark Coherence

Function : BCF)를 제안한다. BCF는 일반적으로 널리 알려진 코히어런스 함수(Magnitude Squared Coherence : MSC)에 기반하여 인간의 청각적 특징을 표현하는 라우드니스 영역에서 자기 바크 스펙트럼과 상호 바크 스펙트럼을 구한다. BCF는 평가 대상이 되는 통신시스템의 아날로그 부분에서 발생하는 선형왜곡^{[13][14]}에 영향을 받지 않는다. 그 결과 기존 ITU-T 권고안인 PSQM과 MNB에 비하여 실제 이동전화 환경에서 보다 정확하게 주관적 음질을 예측한다.

본 논문의 구성은 다음과 같다. II장에서 기존의 주/객관적 음질 평가법에 대하여 알아보고, III장에서는 기존 객관적 음질평가법을 이동전화 환경에 적용시에 발생하는 문제점에 대하여 설명한다. IV장에서는 새로운 객관적 음질평가법인 BCF에 대하여 설명한 후, V장에서 BCF의 실험 및 성능을 평가하며 VI장에서 결론을 맺는다.

II. 음질평가

음질 평가는 주로 청취자들이 직접 듣고 판단하는 주관적 음질 평가법^[15]과 원음성과 왜곡음성의 대수적 차이를 이용하는 객관적 음질 평가법이 있다.

1. 주관적 음질평가

주관적 음질 평가법에는 청취방법에 따라 양방향 청취법과 단방향 청취법으로 나뉜다. 양방향 청취법인 대화식 청취법은 실시간 시스템이 완성된 단계에서 가능한 주관적 음질 평가법으로 음성 부호화기 개발 단계에서 쉽게 수행할 수 없다. 그러므로 평가가 단방향 청취식 방법이 일반적으로 사용된다.

단방향 청취식 방법에는 왜곡된 음성 신호만을 듣고 절대적 음질을 평가하는 절대 음질 평가법(ACR : Absolute Category Rating), 원음성과 왜곡음을 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)등이 있다. 본 논문에서는 음성부호화기 평가에 일반적으로 많이 사용되는 절대 음질 평가법을 사용하였다. 표1에 절대 음질 평가법에 사용되는 MOS (Mean Opinion Score) 등급을 나타내었다. 5점은 음질이 매우 우수함을 의미하며 1점은 음질이 매우 나쁨을 의미한다.

표 1. MOS 평가등급.

Rating	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

원음성과 왜곡된 음성을 대수적으로 비교하여 평가하는 객관적 음질 평가법은 두 신호의 시간축 동기화가 필수적이다. 본 논문에서는 비교적 널리 사용되는 방법인 상호 상관법을 이용하여 두 신호의 동기화를 수행하였다. 상호 상관법은 두 신호의 상호 상관을 구한 후 그 최대값을 시간축 동기기로 사용한다. 기존의 객관적 음질 평가법은 평가가 이루어지는 영역에 따라 시간 영역에서의 평가, 주파수 영역에서의 평가 및 심리 음향 영역에서의 평가로 분류된다.

SNR 척도는 원 음성과 왜곡된 음성 파형간의 자승평균오차를 계산한다^[1]. SNR은 단지, 원 음성과 왜곡된 음성 파형의 신호 파형만의 비교이므로, 아나로그 이동전화시스템이나 낮은 전송률을 갖는 음성부호화기의 주관적 음질을 예측하는데에는 적합하지 않다. 반면에 SegSNR은 SNR 보다 주관적 음질과 상관 관계가 높다. 또, 인간의 청각 특성을 이용한 심리 가중된(Perceptually Weighted) SNR인 Frequency-variant SNR(FVSNR)등도 있다.

LPC-CD^[2]는 원 음성과 왜곡된 음성의 선형예측성도 모델에 기본을 두고 있다. 파라미터는 LPC 자체가 직접 사용될 수도 있고, PARCOR, Log Area Ratio(LAR), Cepstral 계수 또는 LSP(Line Spectrum Pairs)등과 같이 변환될 수도 있다

주관적 음질 평가 결과와 상관관계가 높다고 알려진 심리 음향 영역에서의 평가는 원 음성과 왜곡 음성에 인간의 청각적 특성을 반영한 심리 음향 모델^[16]을 적용하여, 두 신호의 왜곡정도를 측정하는 방법으로 일반적으로 다음과 같은 세가지 청각적 특성을 이용한다.

- 임계 대역 분석
- 등감 곡선 보정
- 주관적 세기 보정

인간의 귀는 저수파 대역에서 세밀한 청각 분해

능을 갖으나 고주파 대역에서는 그렇지 못하며, 일정한 임계 주파수 영역에서는 각 주파수별 에너지가 상호 영향을 준다고 알려져있다. 이러한 특성을 고려한 것이 임계 대역 분석이다. 또한 인간의 귀는 주파수 별로 자극에 대하여 다르게 민감하다. 예를 들어 100Hz의 톤이 1000Hz의 톤과 동일한 크기로 들리려면 35dB 더 큰 세기를 가져야 한다. 이러한 청각적 특징을 고려하여 각각의 주파수에 대하여서도 사람에게 의해 지각되는 소리의 크기가 동일하도록 보정할 필요가 있는데, 이를 표현한 것이 등감 곡선이다. 주관적 세기 보정은 인간이 소리의 세기 변화에 따라 비선형적으로 반응하는 특징을 반영한 것이다.

이러한 심리 음향 영역에서 음성 스펙트럼의 차이를 계산하는 객관적 음질 평가법으로 BSD, PSQM, MNB 등이 있다. BSD는 심리 음향을 나타내는 바크 스펙트럼의 차이를 나타낸다. PSQM은 정교한 심리 음향 모델과 인지 모델을 이용하여 왜곡된 음성 신호의 음질을 평가한다. PSQM은 입력된 음성 신호의 스펙트럼에 대하여 위에서 설명한 심리 음향 특성과 정교한 인식 모듈을 통하여 왜곡된 음성 신호의 왜곡 정도를 평가하는 ND(Noise Disturbance)를 계산한다. 이러한 PSQM은 ITU-T에 의하여 1996년에 300-3400Hz 전화 대역 음성 부호화기의 성능 평가를 위한 음질 평가 권고안으로 채택되었다. 이 방법은 주로 전송률 4kbps 이상의 CELP 음성부호화기를 대상으로 하여 입력 음성 레벨, 화자 종속, 반복 부호화 (transcoding) 등의 항목을 평가하는데 적합하다^[11].

MNB는 전화대역 음성부호화기의 음질 평가 알고리즘인 PSQM보다 다양한 경우에 대해 적용시키기 위하여 고안되었다. PSQM은 전송률 4kbps 이상을 갖는 CELP 부호화기의 성능 평가에 있어서는 우수한 성능을 보인 반면, 무선 채널의 비트에러, ATM 네트워크의 셀 손실, 인터넷 폰의 패킷 손실 및 지연^[10]등의 최근에 개발된 음성 통신 시스템에서 발생하는 음질 저하에 대해서는 비교적 좋지 않은 결과를 보인다. MNB는 정교한 청각 모델보다는 인간이 음질에 따라 반응하는 모델을 기초로 하여 시간 영역과 주파수 영역에서 정규화된 블록값의 거리를 계산한 (Auditory Distance)으로써 음성 신호의 왜곡 정도를 평가한다. 또한 0과 1에서 접근선을 갖는 logistic 함수를 사용하여 주관적 음질과의 상관 관계를 높였다. 이 방법은 1998년에 ITU-T의 권고안으로 채택되었으며, 채널의 비트 에러나 프레

임 손실등의 왜곡 요인을 갖는 음성부호화기의 음질을 평가하는데 적합하다^[11]. MNB는 적은 계산량을 고려한 MNB I 과 주관적 음질과 보다 높은 상관관계를 갖는 MNB II가 있다. 두 알고리즘은 주파수 영역의 FMNB (Frequency MNB) 구조는 동일하고, 시간 영역의 TMNB (Time MNB)구조에서 차이가 있다.

III. 이동전화 음질평가

디지털 이동전화 시스템에서 일반적으로 음성은 이동전화기의 음성부호화기에 의하여 부호화되어 무선 채널을 통하여 기지국으로 전달 되어진 후, 다시 복호화 되어 일반 유선 전화 (Public Switched Telephone Network : PSTN) 사용자에게 청취된다. 그림 1에 디지털 이동전화에서 음성이 전달되는 과정을 간략히 나타내었다. 최근의 상용화된 표준 음성부호화기는 음성이 부호화/복호화되는 동안 음질의 손실이 거의 없는 뛰어난 성능을 나타낸다. 디지털 이동전화 시스템에서 음질 저하는 대부분 부호화된 음성이 무선 채널을 통하여 전달될 때 발생하는 채널에러로 효과적으로 복호화 되지 못할 경우에 발생한다. 선형왜곡은 통신시스템의 아날로그부분에 의하여 발생되는데, 음질에 거의 영향을 주지 않는다. 그러나 [14]에 보고된바와 같이 선형왜곡은 기존의 객관적 음질 평가 시스템의 성능을 저하시킨다.

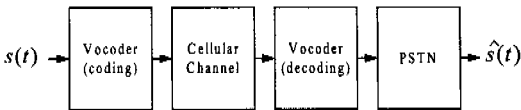


그림 1. 디지털 이동 전화 시스템에서 음성 전달 경로

1. 선형왜곡

본 절에서 아날로그 회로로 인하여 발생하는 선형왜곡이 객관적 음질 평가에 미치는 영향을 살펴 보기 위하여 CDMA PCS 시스템을 사용하여 실제 음성 데이터 베이스를 구축하였다. 구축된 음성데이터는 두가지 서로 다른 통과대역 특성을 가지는 경로를 통과하여 녹음 되어 두가지 군으로 분류된다. A군으로 분류된 음성은 앞에서 설명한 일반적인 디지털 이동전화 시스템의 음성 전달경로와 같이 그림 2와 같이 이동전화를 통하여 기지국에 도착한 후 국선 (Public Switched Telephone Network :

PSTN)과 사실 전화 교환기 (Private Branch Exchange : PBX)를 추가로 거친다. B군은 사실 교환기를 거치지 않는 점을 제외하고 A군과 동일하다. 일반적으로 음성은 국선(PSTN)에서 내선(PBX)를 거치면서 선형 왜곡 발생하며, 국선과 내선사이의 연결부분 뿐만 아니라 음성 통신시스템 내의 다양한 경로에 발생할 수 있다. 구축된 음성데이터를 이용하여 주/객관적 음질 평가를 수행하였다. 주관적 음질평가로 두군의 음성을 랜덤하게 섞어 30명의 청취자에 의하여 절대 음질평가법인 MOS평가법을 실시하였다. 청취자중 과반수 이상이 과거에 주관적 음질 평가 경험이 있었다. 객관적 음질평가를 위하여 ITU-T에 의하여 권고안으로 채택되었으며 기존의 객관적 음질평가법 중 주관적 음질과 상관관계가 높다고 알려져 있는 PSQM을 사용하였다.

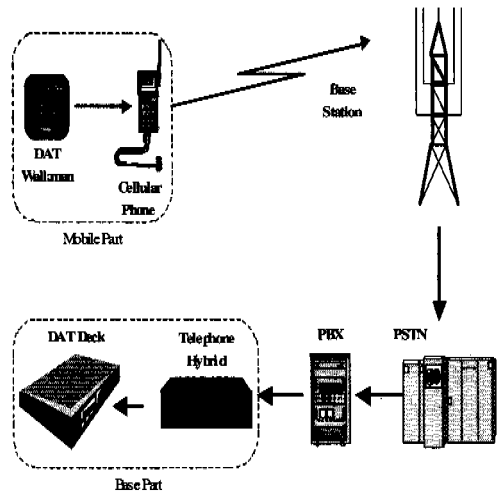


그림 2. A군 음성 데이터 녹음 경로

그림 3은 두 음성 데이터 군에 대한 주/객관적 음질 평가결과의 분포를 나타내었다. 각 분포도는 정규분포와 모양이 매우 비슷하며 그림 3 (a)의 MOS 분포도를 살펴볼 때 사실교환기를 추가로 통과한 A군과 통과하지 않은 B군의 분포도가 거의 비슷한 것으로 보아 사실교환기는 주관적 음질에 거의 영향을 미치지 않으며, 두 군의 음질이 비슷함을 알 수 있다. 그러나 객관적 음질평가인 PSQM의 분포도인 그림 3 (b)에서 A군이 B군에 비하여 전체적으로 PSQM값이 높음을 나타낸다. 이러한 경우 A군의 음질이 B군보다 나쁨을 예측하여 A군의 음성 전달경로가 열악하다는 잘못된 판단을 할 수 있다. 이러한 현상은 A군의 음성이 사실교환기를 통

과하면서 발생된 선형 왜곡현상에 기인한다. 또 다른 여러 실험을 통하여 선형 왜곡이 기존의 다른 객관적 음질 평가법들에 대해서도 성능을 저하시킬 것을 확인하였다.

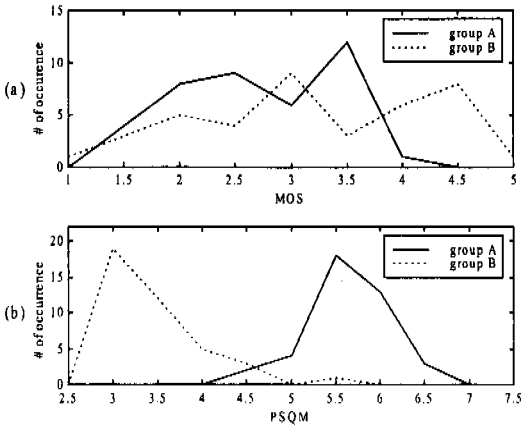


그림 3. CDMA 음성 데이터에 대한 주관적 음질과 객관적 음질의 분포도. (a) MOS분포도 (b) PSQM분포도.

2. 이득조정

정확한 객관적 음질 평가 척도를 구하기 위하여 수신단에서 전체 이득 조정이 필요하다. 또, 각 분석 프레임마다 천천히 변하는 이득을 보상하기 위하여 부분적으로 이득 조정이 요구된다. 이득 조정을 위하여, 원음성과 왜곡음 전체 에너지를 같은 수준으로 조정을 해야 하며, 이러한 이득 조정은 BSD나 PSQM 같은 원음과 왜곡음의 스펙트럼 차이를 구하는 객관적 음질 평가법에는 필수적이다.

예를 들어 음성의 특정 주파수 밴드에 첨가 잡음으로 인하여 왜곡이 발생하였다고 가정하면, 왜곡음의 전체 에너지는 첨가된 잡음만큼 증가하게 되고, 결과적으로 이득 조정은 첨가 잡음이 존재하지 않는 다른 주파수 영역의 에너지 수준이 낮아지게 한다. 그러므로 음성이 첨가 잡음으로 왜곡되었을 때 이득 조정은 또 다른 주파수 왜곡을 발생하게 한다. 그림 4는 이러한 이득 조정문제의 전형적인 예이다. 그림 4 (a)에서 실선은 원음의 바크 스펙트럼, 점선은 이득조정 전의 왜곡음 바크 스펙트럼, 그리고 쇄선은 이득 조정 후의 왜곡음 바크 스펙트럼을 의미한다. 그림에서 음성은 고주파 영역이 첨가 잡음으로 왜곡되었으며, 청취자는 실제로 이 잡음에 대하여 인지하고, 음질 저하를 느낀다. 그러나 그림 4 (b)에서와 같이 이득 조정후에는 실제 첨가 잡음으로

로 인한 왜곡이외에도 잡음이 존재하지 않은 주파수 영역에 새로운 스펙트럼 왜곡이 발생하였다. 이러한 경우 스펙트럼의 차이를 구하는 기존의 객관적 음질평가법은 음질을 잘못 평가할 확률이 높아진다.

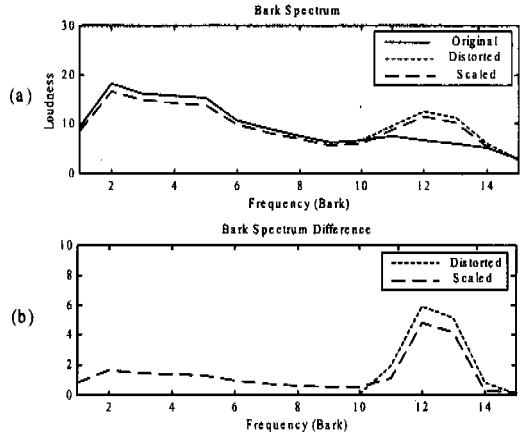


그림 4. 첨가잡음에 의한 이득조정 문제. (a) 원음의 바크 스펙트럼(실선), 이득조정 전의 왜곡음 바크 스펙트럼(점선), 이득조정 후의 왜곡음 바크 스펙트럼(쇄선) (b) 원음과 이득조정 전의 왜곡음 바크 스펙트럼 차이와 원음과 이득조정 후의 왜곡음 바크 스펙트럼 차이.

IV. 바크 코히어런스 함수

심리 음향을 이용한 객관적 음질 평가법의 최신 관심 분야는 심리 음향 모델 자체 보다는 인식 모듈이다. 심리 음향 모델은 인간의 청각 특성을 반영하며 비교적 잘 연구된 분야이다. 인간의 청각 주파수 분해능은 헤르츠 주파수를 기준으로 동일하지 않으며, 소리의 민감성은 각 주파수마다 다르다. 또, 소리의 세기와 관련된 라우드니스는 비선형적 특성을 지니고 있다. 인식 모듈은 소리 정보가 인간 두뇌에서 처리되는 과정을 모델링한다. 인식 모듈은 아직 연구가 미진하여 매우 단순화된 모델을 사용한다^[17]. 비록 PSQM이 첨가 소음과 제거 소음에 대한 비대칭 효과를 이용하며, MNB (Measuring Normalizing Block)가 스펙트럼 차이에 대한 청취자의 적응 알고리즘을 사용하나, 이러한 방법들은 다양한 전파 환경을 갖는 이동 전화 시스템의 주관적 음질을 예측하는데 문제점을 가지고 있다.

본 논문에서는 객관적 음질 평가법의 새로운 인식 모듈로 BCF(Bark Coherence Function)을 제안한다. BCF는 주파수 영역에서 두 신호의 상관관계

를 나타내는 코히어런스 함수(magnitude coherence function : MSC)^[12]을 인간의 심리 음향 반영한 바크(Bark) 영역에서 적용시킨 것이다.

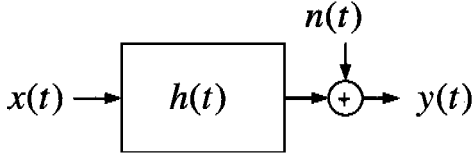


그림 5. 이동 전화 시스템의 간략화된 블록도

실제 음성 통신 시스템은 그림 5와 같이 간단하게 표현될 수 있다. 그림에서 $x(t)$ 는 입력 음성이며, $n(t)$ 는 첨가 잡음, 음성부호화기에 의하여 발생하는 왜곡 그리고 통신 채널 에러로 인한 비선형 왜곡이다. 또, $h(t)$ 는 선형 시스템으로 모델링되는 통신시스템의 아날로그부분이며 $y(t)$ 는 최종적으로 청취자에게 들리는 왜곡 음성이다. MSC는 선형 주파수 영역에서 다음과 같이 정의 된다.

$$\gamma_{xy}^2(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)} \quad (1)$$

여기서 $S_{xy}(f)$ 는 입력 신호 $x(t)$ 와 출력 신호 $y(t)$ 의 상호 파워 스펙트럼이고, $S_{xx}(f)$ 는 $x(t)$ 의 파워 스펙트럼, 그리고 $S_{yy}(f)$ 는 $y(t)$ 의 파워 스펙트럼이다. MSC는 원음과 왜곡음 사이의 주파수 영역에서 상관관계를 나타낸다. 다음으로 선형 시스템을 통과한 $x(t)$ 의 스펙트럼인 코히어런스 출력 파워 스펙트럼 (Coherent Power : CP)와 잡음 성분에 의한 출력 잡음 스펙트럼 (Non Coherent Power : NCP)를 식 (2)로 나타낼 수 있다^[18].

$$\begin{aligned} CP(f) &= \gamma_{xy}^2(f) |S_{yy}(f)|^2 \\ NCP(f) &= [1 - \gamma_{xy}^2(f)] |S_{yy}(f)|^2 \end{aligned} \quad (2)$$

BCP와 CP의 비인 잡음 대 신호 비 (Distortion-to-Signal Ratio : DSR)는 다음과 같이 표현된다.

$$DSR(f) = \frac{NCP(f)}{CP(f)} = \frac{1 - \gamma_{xy}^2(f)}{\gamma_{xy}^2(f)} \quad (3)$$

잡음이 없는 이상적인 경우 MSC는 1의 값을 갖고 이때, DSR는 0이 된다. 또 비선형 잡음만이 존재할 경우 MSC는 0이 되어 DSR은 무한대의 값을

갖게 된다. 일반적인 경우 통신 시스템은 비선형 잡음이 존재하므로 DSR은 0과 무한대 사이의 값을 갖는다.

선형 주파수 영역에서 MSC의 정의와 더불어, 심리 음향 영역에서 정의된 MSC인 BCF는 다음과 같이 정의된다.

$$BCF(b) = \frac{|L_{xy}(b)|^2}{L_{xx}(b)L_{yy}(b)} \quad (4)$$

여기서 b 는 바크 주파수, $L_{xx}(b)$ 와 $L_{yy}(b)$ 는 각각 $x(t)$ 와 $y(t)$ 의 자기 바크 스펙트럼이고, $L_{xy}(b)$ 는 $x(t)$ 와 $y(t)$ 의 상호 바크 스펙트럼이다. 그림 6은 상호 바크 스펙트럼과 자기 바크 스펙트럼을 구하는 과정이다. $x(t)$ 와 $y(t)$ 의 자기 스펙트럼과 상호 스펙트럼을 구한 후 [4] 그리고 [5]의 심리 가중 필터링을 하고 라우드니스 보정을 한다. 음성을 라우드니스 영역으로 천이 시키기 위해서 앞장에서 설명한 심리 음향 특성을 사용한다. 이러한 일련의 과정들은 이미 잘 알려져 있으며, 본 논문에서는 BSD [5]방법을 사용하였다.

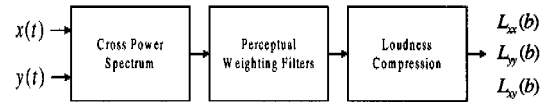


그림 6. 자기 바크 스펙트럼, 상호 바크 스펙트럼 계산.

객관적 음질 평가법의 인식 모듈로서 BCF의 성능을 평가하기 위하여 실제 음성을 사용하여 그림 7과 같이 원음과 왜곡음의 바크 스펙트럼과 BCF를 구하였다. 왜곡음은 저주파 영역이 감쇄되었으며 고주파 영역은 첨가 잡음으로 왜곡되었다. 이 경우, 청취자는 일반적으로 저주파 영역 감쇄 보다는 고주파 영역의 첨가 잡음에 대하여 불평을 하게 된다. 그런데, 스펙트럼의 차이를 구하는 BSD와 같은 기존의 객관적 음질 평가법은 고주파 영역 및 저주파 영역 모두에서 비슷한 정도의 주관적 음질 저하를 나타내므로, 청취자의 실제 주관적 음질과 일치하지 않는다. BCF는 선형 왜곡에는 영향을 받지 않고 고주파 영역의 첨가 잡음에만 영향을 받으므로 그림 7에서와 같이 비선형 왜곡인 고주파 영역의 첨가 잡음에만 1보다 작은 값을 나타내어 보다 정확하게 주관적 음질을 예측할 수 있다.

BCF를 계산한 후 최종적으로 왜곡 정도를 정량하기 위하여 다음과 같이 바크 신호대 잡음비(Bark

Distortion-to-Speech Ratio : BDSR)를 정의한다.

$$BDSR = \sum_{b=0}^N \frac{1 - BCF(b)}{BCF(b)} \quad (5)$$

BDSR는 심리 음향 영역에서 측정된 이동 전화 시스템의 음질을 의미한다.

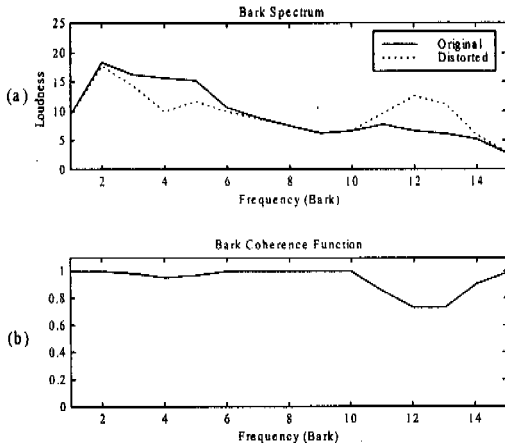


그림 7. 바크 스펙트럼과 BCF. (a) 원음과 왜곡음의 바크스펙트럼 (b) BCF.

V. 실험 및 결과

본 논문에서는 제안된 BCF의 성능을 평가하기 위하여 회귀분석^[18]을 수행하였다. 기존의 연구에서 일반적으로 회귀 함수로 2차 다항함수를 사용하였으나 이 2차 함수는 단순 증가 함수가 아니고, 또 쉽게 사용하는 주관적 음질 평가법인 MOS는 1점부터 5점사이의 값을 가지며, BCF와 같은 객관적 음질 평가법은 0점부터 무한대까지의 값을 갖게된다. 그러므로 기존의 2차 회귀 함수는 중심축의 오른쪽 부분의 새로운 심화 왜곡에 대하여 적절히 분석할 수 없다. 비록 logistic 함수를 사용하여 객관적 음질 평가값을 0부터 1사이 값으로 천이 시켜주는 방법^[9]이 제안되었으나, 이는 1점부터 5점사이의 값을 갖는 MOS값과 직관적이지 못하다. 본 논문에서는 1과 5사이를 점근선으로 갖는 다음과 같은 새로운 회귀 함수를 제안한다.

$$\hat{y} = 1 + \frac{4}{1 + \exp(ax + b)} \quad (6)$$

여기서 a 와 b 는 회귀 분석 계수이다. 제안된 회귀 함수는 무한한 범위의 객관적 음질 평가 척도 값을 유한한 범위로 천이함으로써, 상관 계수를 증가시키며, 단순 증가 함수가 아닌 다항 함수를 사용할 경우 발생할 수 있는, 부분적으로 객관적 음질 평가 척도값 증가시 예측된 음질이 향상되는 현상을 막는다. BCF의 성능을 나타내는 지수로 피어슨 상관 계수와 표준 추정 오차(SEE)를 사용하였다. 상관 계수는 객관적 음질 평가값과 MOS와의 선형성을 나타내며, 0에 가까울수록 낮은 상관 정도를, 1에 가까울수록 강한 상관 정도를 보여 준다. 표준 추정 오차는 실제 MOS와 예측된 주관적 음질인 예측된 MOS와의 차이를 통계적으로 나타낸다.

그림 8은 그림 3과 같은 음성 데이터를 가지고 BCF에 대하여 분석 한 결과이다. 그림에서 BDSR의 분포도가 MOS 분포도와 같이 비슷한 분포를 가짐으로 보아 BCF는 국선의 이날로그 시스템으로 인한 선형왜곡의 영향을 받지 않고, 실제 이동 전화 시스템의 음질을 평가하는데 적절함을 알 수 있다.

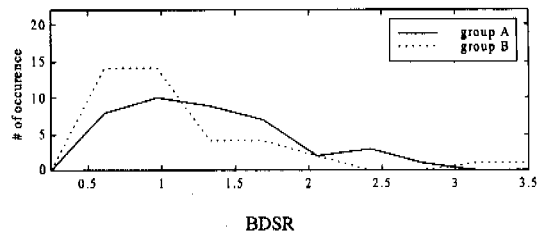


그림 8. BDSR 분포도

표 2는 한국의 CDMA PCS와 AMPS환경에 대한 음질평가 실험 결과이다. 테스트 I와 II는 각각 채널특성이 다른 PSTN을 통과한 데이터로 정리한 결과로 각 특성 별로 크게 A군과 B군으로 구분하였다. 테스트 I은 CDMA PCS의 경우로 그림 3 및 그림 8에서 설명한 음성 데이터이다. 사용된 문장은 각각 여성 화자 2명, 남성 화자 2명으로 NTT-AT의 다국어 음성 데이터 베이스에서 선택하였다. 이동 전화의 사용량이 많고 다양한 전파 환경이 있는 서울에서 실시하였으며, 가능한 모든 실제 이동 전화 사용 환경을 고려하기 위하여 차량이동 환경과 도보 이동 환경등 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. 또, 반대의

경우에도 녹음을 실시하였다. 이렇게 얻은 음성 신호 데이터를 두 종류로 분류하였는데, PSTN뿐만 아니라 사실 교환기를 더 거치는 과정에서 선형 왜곡으로 인해 채널의 영향을 많이 받은 음성 신호들을 A 군, PSTN만을 거친 음성 신호들을 B 군으로 하였다. 회귀 분석 결과, PSQM의 경우 A군과 B군에 대하여 각각의 상관계수는 0.798과 0.911으로 주관적 음질과 상관관계가 높으나 A군과 B군이 섞여 있을 때 상관계수는 0.558로 매우 떨어진다. 또, 표준 추정 오차는 각각 0.437과 0.435이나 A군과 B군이 섞인 경우는 0.772이다. 이는 A군과 B군에서 녹음된 데이터가 통과한 PSTN의 아날로그 특성이 틀리므로 선형 왜곡으로 인하여 PSQM값이 영향을 받기 때문이다. MNB는 A군과 B군이 섞여 있을 경우 각각의 경우에 비하여 성능의 저하가 심하지 않으나 "Per-Condition Correlation"에 최적화 되었으므로 전체적으로 성능이 떨어진다. BCF는 아날로그 시스템 특성에 영향을 받지 않으므로, A군의 경우 0.881, B군의 경우 0.926 그리고 A군과 B군이 섞여 있을 경우 0.903의 상관계수 값을 갖는 등, 모든 경우에 있어서 높은 상관관계를 보여준다. 테스트 II는 아날로그 이동 전화시스템의 결과로서 BCF가 가장 높은 상관계수 값을 갖는 등 테스트 I과 일치된 결과를 보인다.

표 3은 한국 CDMA 디지털 셀룰러시스템(Digital Cellular System : DCS)와 CDMA PCS 시스템에 대한 회귀분석 결과이다. 테스트 III은 CDMA DCS 그리고 테스트 IV은 CDMA PCS의 경우이다. 녹음 환경은 테스트 I, II와 비슷하며, 테스트 III, IV를 통하여 BCF의 성능이 뛰어난을 다시 확인할 수 있다. 녹음 데이터에서 가변 지연을 발견하였는데 이는 VoIP (Voice over Internet Protocol)와 같은 패킷 기반 음성 전송 시스템에서 자주 발견되며, 객관적 음질 평가 시스템의 성능을 저하시킨다. 테스트 IV보다 테스트 III서 가변 지연이 더 자주 발견되었

는데 이로 인하여 세가지 객관적 음질 평가법 모두 테스트 III이 테스트 IV에 비하여 낮은 상관계수를 나타낸다. 가변지연문제를 해결하기 위하여, 전처리 과정으로 상호상관을 이용한 음성 동기화를 행하였으며, 원음과 왜곡음의 객관적 음질 왜곡이 최저점으로 최종적인 동기화를 수행하였다. 테스트 IV에서 배경 잡음이 발견되었는데 이는 목음 구간에 낮은 가중치를 갖는 객관적음질 평가법의 성능을 저하시킨다. 이는 기존의 객관적 음질 평가법이 주로 음성 부호화를 평가하기 위하여 개발되었기 때문이다. PSQM은 이득보상으로 인하여 배경잡음의 에너지 수준이 감소되었기 때문에 이득보상이 필요 없는 BCF에 비하여 낮은 성능을 보인다.

표 3. DCS와 PCS에 대한 상관계수와 표준추정오차.

Test	Correlation Coefficient / SEE		
	PSQM	MNB II	BCF
III	0.692/0.671	0.667/0.755	0.840/0.505
IV	0.929/0.453	0.794/0.793	0.958/0.349

마지막으로 표 4에 전화대역 음성부호화에 대한 분석결과를 정리하였다. 테스트V는 2.4~16kbps 음성부호화기, DAT loop 그리고 MNRU (Modulated Noise Reference Unit)^[19](5~35dB)등을 포함하여 테스트 VI는 2.4~32kbps 음성부호화기 그리고 MNRU (5~25dB)등의 조건을 포함한다. 음성 데이터는 모두 미국식 영어로 녹음되었으며 주관적 평가는 각각 다른 미국 기업 연구실에서 주관으로 행하여졌다. 테스트 VI에서 64kbps PCM이 원음으로 사용되었다. 실험 결과 BCF가 비록 이동전화 음질 평가를 위하여 개발되었으나 음성부호화기 평가에도 기존의 음질 평가 방법인 PSQM과 MNB에 비하여 뛰어난 성능을 가짐을 확인하였다.

표 2. PCS와 AMPS 에 대한 상관계수와 표준추정오차

Test	Correlation Coefficient / SEE								
	PSQM			MNB II			BCF		
	A	B	A+B	A	B	A+B	A	B	A+B
I	0.798 / 0.437	0.911 / 0.435	0.558 / 0.772	0.763 / 0.468	0.878 / 0.506	0.764 / 0.600	0.881 / 0.342	0.926 / 0.399	0.903 / 0.400
II	0.890 / 0.616	0.904 / 0.614	0.852 / 0.727	0.548 / 1.130	0.516 / 1.230	0.530 / 1.179	0.923 / 0.520	0.880 / 0.680	0.891 / 0.632

표 4. 음성부호화에 대한 상관계수와 표준추정오차.

Test	Correlation Coefficient / SEE		
	PSQM	MNB II	BCF
V	0.916/0.322	0.875/0.389	0.913/0.327
VI	0.878/0.416	0.792/0.530	0.910/0.361

VI. 결론

본 연구는 심리 음향을 반영한 바크 주파수 영역에서 코히어런스 함수를 정의하고 이를 이용하여 CDMA 이동 전화 시스템에 대하여 객관적 음질평가를 수행하였다. 또, 회귀 분석을 위하여 MOS 평가 점수와 같은 범위를 갖는 새로운 회귀함수를 제안하였다. 실험 결과 BCF는 기존 ITU-T 권고안인 PSQM과 MNB에 비하여 주관적 음질과 높은 상관관계를 가짐이 입증되었다. 이동 전화 시스템에서 발견된 가변 지연 현상을 보상하기 위한 연구가 활발히 진행 중에 있다.

Acknowledgments

We wish to thank Peter Kroon of Lucent Technologies and Joshua Rosenbluth of AT&T for supplying English original and coded speech and associated MOS scores.

참고 문헌

[1] S. Quackenbush, T. Barnwell and M. Clements, *Objective measures of speech quality*, Prentice Hall, 1988.

[2] N. Kitawaki, H. Nagabuchi and K. Itoh, "Objective quality evaluation for low-bit-rate speech coding systems", *IEEE J. Select. Areas Commun.*, vol 6, No2, pp242-248, Feb 1988.

[3] R. Kubichek, D. Atkinson and A Webster, "Advances in objective voice quality assessment", in *Proc. of IEEE Global Telecom Conference*, vol 3, pp1765-1770, 1991.

[4] Nynek Hermansky, "Perceptual Linear Predictive (PLP) analysis of speech", *J. Acoust. Soc. Am.*, vol. 87, pp1738-1752, April 1990.

[5] 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.

[6] J. G. Beerends and J. A. Stemrindk, "A perceptual speech-quality measured based on a psychoacoustic sound representation", *J. Audio Eng. Soc.*, vol 42, No 3., pp115-123, March, 1994.

[7] W. Yang, M Dixon, and R. Yantorno, "A modified bark spectral distortion measure which uses noise masking threshold", in *Proc. IEEE Speech Coding Workshop*, pp. 55-56, 1997.

[8] D. S. Kim, O. Ghitza and P. Kroon, "A computational model for mcs prediction", in *Proc. IEEE Speech Coding Workshop*, pp141-143. 1999.

[9] S. Voran, "Objective estimation of perceived speech quality, Part I : Development of the measuring normalizing block technique", *IEEE Trans. on Speech and Audio Processing*, vol. 7, No. 4, pp371-382, July 1999.

[10] A. Rix, R. Reynolds, and M. Hollier, "Robust perceptual assessment of end-to-end audio quality", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp39-42, October 1999.

[11] ITU-T Rec. P.861, "Objective quality measurement of telephone-band speech codecs", 1998.

[12] Julius S. Bendat and Allan G. Piersol, *Engineering Applications of Correlations and Spectral Analysis*, John Wiley & Sons, 1980.

[13] M. G. Rahim and B. H. Juang, "Signal bias removal by maximum likelihood estimation for robust telephone speech recognition", *IEEE Trans. Speech and Audio Processing*, vol. 4, No. 1, pp19-30, 1996.

[14] A. Rix, R. Reynolds, and M. Hollier, "Perceptual measurement of end-to-end speech quality over audio and packet-based networks", *AES 106th Convention*, Munich, 8-11 May, pre-print no.4873, 1999.

[15] ITU-T Rec. P.800, "Method for subjective determination of transmission quality", 1996.

[16] E. Zwicker and H. Fastl, *Psychoacoustics Facts*

and Models, Springer-Verlag, 1990.

- [17] Markus Hauenstein, "Application of meddis' inner hair-cell model to the prediction of subjective speech-quality", in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process*, pp 545-548, 1998.
- [18] N.R. Draper, H. Smith, *Applied Regression Analysis*, John Wiley & Sons, New York, 1981.
- [19] ITU-T Rec. P.810, "Modulated noise reference unit (MNRU)", 1996.

박 상 욱(Sang-Wook Park)

한국통신학회논문지 제25권 제12B호, 2000년 12월
참조

류 승 균(Seung-Kyun Ryu)

1989년 2월: 연세대학교 전자공학과 졸업
1991년 2월: 연세대학교 전자공학과 석사
1991년 3월~현재: LG 전자
1997년 9월~현재: 연세대학교 전자공학과 박사과정

박 영 철(Young-Cheol Park)

한국통신학회 논문지 제25권 제6B호, 2000년 6월
참조

윤 대 희(Dae-Hee Youn)

한국통신학회 논문지 제25권 제6B호, 2000년 6월
참조