

# CDMA 이동통신 시스템에서의 주관적 음질을 추정하기 위한 객관적 척도

## Objective Measure for Estimating Subjective Voice Quality in Wireless Communication

백 금란(\*), 이 미숙(\*), 김 영렬(\*\*), 서성희(\*\*), 이 황수(\*,\*\*)

\*한국 과학 기술원 전기 및 전자공학부

\*\*SK 텔레콤 중앙연구원

### 요 약 문

본 논문에서는 CDMA(Code Division Multiple Access) 채널을 통과하면서 여러 가지 형태로 손상된 음성에 대한 주관적 음질 평가를 할 수 있는 객관적 척도에 대한 연구를 수행하였다. 즉, CDMA 채널을 통과한 음성 신호에 대하여 주관적 음질 평가 방법 중 가장 많이 사용되고 있는 MOS(Mean Opinion Score) 테스트를 수행하고, 이 MOS 테스트 결과를 추정할 수 있는 객관적도 알고리즘을 시뮬레이션 하였다. 이러한 연구 결과로 PSQM(Perceptual Speech Quality Measure)을 CDMA 채널 환경에 맞게 수정하여 우수한 성능의 객관적 음질 평가 방법을 얻었다.

### 제 1 장 서 론

무선 통신의 중요 수단인 이동 전화기와 같은 통신 시스템에서는 채널을 통해 전송되는 대부분의 음성이 채널 감쇠와 배경 음에 의해 왜곡된다. 따라서 서비스업자들이 고려해야 할 중요한 사항 중의 하나가 이동 통신 채널 상에서의

음질을 좋게 유지하는 것이다. 일반적으로 이동 전화기의 통화 품질을 측정하기 위해서는 먼저 서비스 가능 지역 내의 다양한 장소에서 반복 청취 실험에 의한 주관적인 평가를 실시해야 한다. 그러나 이 방법은 매우 수고스럽고 비용이 많이 들기 때문에 실제적으로 불가능하다. 따라서 이동 통신 시장의 활성화와 더불어 다양한 감음 환경과 채널 손상 하에서 녹음된 음성의 주관적 평가 결과를 정확히 추정할 수 있는 객관적도 알고리즘의 개발에 대한 연구가 활발히 진행되고 있다[5][6][7]. 객관적도 알고리즘 개발을 위해서는 먼저 이동 통신 채널을 통하여 음성 데이터를 녹음하고 이 음성 데이터에 대해 주관 평가를 실시한다. 그리고 이 데이터들을 이용하여 객관적도 결과를 구한다. 음 이미 평가한 주관적도와 객관적도 결과 사이의 매핑 함수식을 구한다. 즉 이 왜곡된 음성과 원 음성과의 차이를 구하여 객관적 음질을 평가한 후 이렇게 계산된 객관적 음질 척도로부터 주관적 음질 척도인 MOS를 예측한다.

## 제 2 장 음질 평가 방법

음질 평가 방법은 크게 수학적 표현식에 근거한 객관적 방법과 청취자들의 주관적 평가 결과에 근거한 주관적 방법으로 구분할 수 있다. 그 중에서 주관적인 평가 방법이 더 정확한 음질을 나타내지만, 이 방법은 시간과 비용이 많이 소요되고 일관성이 없다는 단점이 있다. 따라서 주관적 음질 평가 결과를 예측할 수 있는 객관적 평가 방법을 사용하므로써 이러한 주관적도의 단점을 보완할 수 있다. 본 장에서는 일반적으로 널리 사용되고 있는 객관적인 평가 방법 및 주관적인 평가 방법에 대하여 간단히 설명한다.

### 2-1 객관적 음질 평가 방법

#### 2-1-1 SegSNR(Segmental SNR)

객관적 음질 평가에 주로 사용되는 SNR은 원 음성과 왜곡된 음성 파형 간의 자승 오차 평균이다. 또한, 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 구간별 SNR을 구하여 이것의 통계적 특성을 이용하는 것이 바람직하며 이러한 방법을 SegSNR이라 한다[5].

#### 2-1-2 LPC\_CD(LPC-Cepstral Distance)

LPC(Linear Prediction Coefficient)는 음성의 주파수 스펙트럼 포락선 모양을 나타낸다. 그리고 LPC\_CD 방법은 이 포락선 성분의 대수적 차이를 계산한 것이다[1].

#### 2-1-3 BSD(Bark Spectral Distance)

마크 스펙트럴 거리(BSD)[2][6]는 인간의 청각적 임계 대역을 기본으로 두고, 인간 청력을 모-

델링하는 필터 뱅크를 사용하는 주관적인 접근 방법이다. 인간의 심리 음향을 고려하므로 음질의 주관적인 평가와의 상관 관계가 매우 높다고 알려져 있다[8]. 인간의 청각 특성은 800Hz 이상의 주파수에 대해서는 주파수가 증가함에 따라 청각의 분해능이 감소하고, 중간 주파수 영역에서 보다 민감하다. 이러한 특성을 반영한 것이 마크 스펙트럼이며, BSD는 원 신호와 왜곡된 신호의 마크 스펙트럼의 차를 구하는 방법이다.

#### 2-1-4 PSQM

PSQM(Perceptual Speech Quality Measure)은 인간이 어떤 소리를 들었을 때 그 소리를 어떻게 인식하는지를 모델링한 것이다. 즉, 신호의 물리적 표현을 인간의 심리 음향학적 표현방식으로 바꾸는 방법으로 3상에 보다 자세히 설명되어 있다.

### 2-2 주관적 음질 평가 방법

주관적도[5][6]는 크게 명료도 테스트와 자연도 테스트로 나뉜다. 명료도 테스트 방법에는 DRT (Diagnostic Rhyme Test), MRT(Modified Rhyme Test)등이 있고 자연도 테스트에는 DAM (Diagnostic Acceptability Measure), DCR (Degradation Category Rating), ACR(Absolute Category Rating), CCR(Comparison Category Rating)[3][4]등이 있다. 본 논문에서 사용한 ACR에서는 음질을 평가하는데 MOS[4]를 사용하여 다섯 단계로 음질을 평가한다.

### 2-3 객관적으로 부터 주관적 음질 예측

객관적 음질 평가 척도로 부터 주관적 음질을 예측하기 위해서는 이차함수가 널리 사용되고 있으며 비교적 우수한 성능을 나타낸다[7] 또한 음질평가 시스템의 성능을 평가하기 위해서 예측한 주관적도의 결과와 실제 주관적도 사이의 상

관 계수를 구한다[9]. 이 때 상관 계수가 1인 것은 MOS 예측 오차가 없음을 의미한다.

### 제 3 장 Modified PSQM 알고리즘

#### 3-1 ITU-T P.861 표준 PSQM 알고리즘

이 절에서는 보코더에 의해 왜곡된 전화선 대역 음성 신호에 대한 음질의 PSQM 알고리즘으로 측정하는 방법에 대해 설명하고 있다[10]. PSQM에서는 실제 환경에서 사람들이 어떤 소리를 듣고 그 소리를 어떻게 인지하는가를 모델링한다. PSQM에서는 왜곡된 신호와 원 신호에 대한 물리적인 표현을 귀의 내부에서 표현되는 방식에 필적하는 심리 음향학적 표현 방식으로 바꾼다. PSQM에서 왜곡된 신호의 음질은 음성 신호에 대한 내부적 표현의 차이에 따라 결정된다. 이 차이는 시간과 주파수의 함수로서 noise disturbance를 계산하는데 사용된다. PSQM의 평균인 noise disturbance는 왜곡된 음성 신호의 음질과 직접 연관된다. 음성 신호의 물리적 표현을 심리 음향학적 표현 방식으로 바꾸는 방법은 다음과 같은 세 단계에 의해 이루어진다.

1. time-frequency mapping
2. frequency warping
3. intensity warping

PSQM 알고리즘에서는 인지(perceptual) 모델링 외에 주관적 평가와 객관적 평가 결과 사이의 높은 상관도를 얻기 위해 인식(cognitive) 모델링을 사용한다.

#### 3-2 Modified PSQM(MPSQM) 알고리즘

앞 3-1절에서 설명된 PSQM은 대부분 코덱의 성능을 평가하는 데는 그 성능이 우수하지만 채널을 통과한 음성의 음질을 평가하는 데는 정확한 척도가 되지 못한다[10]. 특히 음성이 채널을 통과하는 과정에서 임의의 구간의 음성이 지

워지는 frame erasure 현상이 일어 나는 경우에는 더욱 그러하다. 실제 채널을 통과한 음성들을 살펴 보니 음성 신호 중에 연속적으로 지워진 frame이 발생됨을 발견할 수 있었다. 그러므로 본 논문에서는 이러한 frame erasure 효과를 고려하는 새로운 알고리즘을 제안하였다. 즉, MPSQM은 음성 중에 frame erasure가 난 비율(Frame Erasure Rate : FER) 변화에 따른 MOS 값을 구하고, FER과 MOS와의 관계를 exponential 함수로 모델링 하여 기존의 PSQM 알고리즘에 적용하였다.

우선, QCELP (Qualcomm Code Excited Linear Prediction) 코덱에서 frame erasure가 발생했다고 임의로 판정하였다. 그리고 FER를 1%에서 5% 까지 연속적으로 발생시켜 가며 MOS 테스트를 해 보았다. 이러한 실험 결과를 통해 아래 그림 3-1과 같이 exponential 함수로 FER과 MOS와의 관계를 모델링하게 되었다. 그리고 이 모델링에 의해 얻어진 exponential 함수 값의 역수를 PSQM의 마지막 단계인 noise disturbance에 적용 시켰다. 그림 3-2는 본 논문에서 제안한 MPSQM 알고리즘에 대한 설명을 블록화하여 보다 자세하게 설명한 것이다.

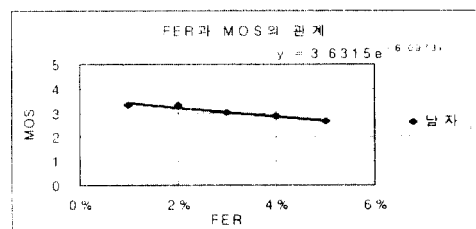
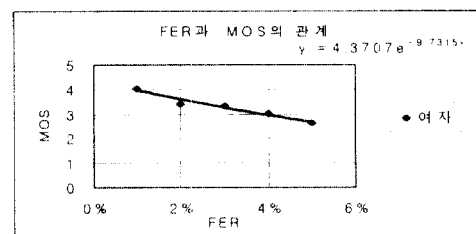


그림3-1 FER과 MOS의 관계

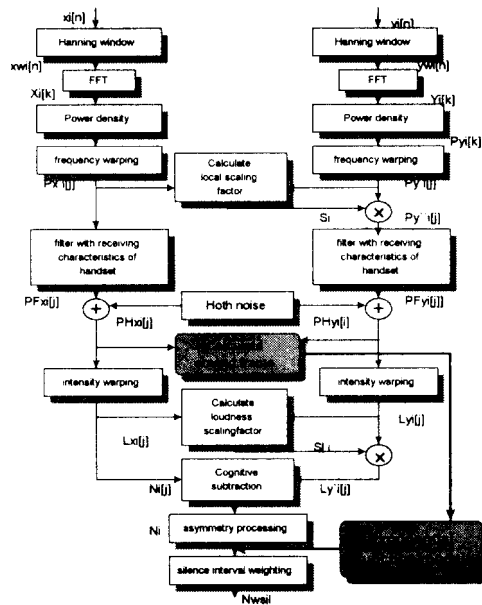


그림3-2 MPSQM 알고리즘 과정

## 제 4 장 실험 결과

### 4-1 음성 데이터 채취 및 MOS 테스트

먼저 reference 데이터를 채취하기 위하여 2-4 초 정도 길이의 문장 데이터를 구성하였다. 그리고 3명의 남자와 3명의 여성 화자가 발성한 음성 데이터를 스튜디오에서 DAT 테이프에 녹음하였다. 또한 왜곡된 음성 데이터를 채취할 때는 각 상비 사이의 임피던스 매칭에 주의하였다. 녹음 환경을 셋팅한 후에는 4가지 링크별로 데이터를 수집하였다. 데이터를 수집할 때는 MOS 점수가 고르게 분포하도록 하기 위하여 자동차로 이동하면서 데이터를 채취하여 채취된 데이터를 정렬한 후에 각 점수에 해당하는 데이터를 선정하였다. MOS 테스트에 사용된 데이터의 유형은 CDMA 시스템을 통하여 녹음한 데이터와 여러 인위적으로 만든 데이터 등이 있다. 보통 MOS 테스트 결과는 어떤 집단의 사람들이 테스트에 참여했느냐에 따라 그 결과가 다를 수 있다. 따라서 MOS 테스트를 할 때는 MNRU 신호에 대한 평가 결과

를 같이 참조한다. 위와 같이 준비된 데이터는 MOS 테스트를 위해서는 모두 랜덤하게 섞어야 한다. MOS 테스트를 할 때 먼저 청취자들 스스로 평가의 기준을 마련할 수 있도록 여러 유형의 음성 샘플들을 들려 주었다. 그리고 쉬는 시간 없이 너무 오랫동안 테스트를 하게 되면 집중력이 저하되기 때문에 한 번 테스트하는 데 걸리는 시간을 24분으로 하였다.

### 4 2 주관 및 객관적도의 상관 관계 분석

객관적도의 궁극적인 목적은 주관적도의 결과, 즉 MOS 테스트 결과를 정확하게 예측하는 데 있다. 객관적도 값을 구하는 데 고려되어야 할 중요한 문제 중에 하나는 원 신호와 전파선을 통한 왜곡된 신호 사이의 동기를 맞추는 것이다. 먼저 동기를 맞추기 위해 임의로 음성의 시작 부분이 존재한다고 생각하고 몇 샘플씩 이동시켜 가면서 객관적도 알고리즘을 수행시킨다. 이렇게 해서 얻은 객관적도 값 중에서 최소 값을 갖게 되는 부분에서 동기가 맞을 것이므로 그 때의 객관적도 값을 선택하였다. 이 때, 몇 샘플씩 이동 시키느냐가 객관적도 알고리즘에 미치는 영향을 알아 보기 위해 여러 경우의 지연에 대한 객관적도의 값을 비교해 보니, 곡선들은 서로 같은 모양을 갖고 있음을 알 수 있다. 특히, 64와 32 shift는 거의 비슷한 모양이지만 128 shift는 다소 차이가 있었다. 그러므로 본 실험에서는 64씩 샘플을 이동해 가면서 가장 적합한 곳을 동기로 맞추어 객관적도 값을 구하였다.

그림4-1은 지금까지 연구되어 온 여러 가지 객관적도 알고리즘 중 성능이 좋은 것을 선택해서 MOS와의 상관 관계를 조사한 결과이다. 표에서와 같이 인간의 심리 음향의 모델을 이용한 BSD와 PSQM이 MOS와 상관 관계가 높음을 알 수 있다. 그 중 PSQM은 BSD보다 더 높으므로 본 논문에서는 PSQM을 선택하여 실험하였다.

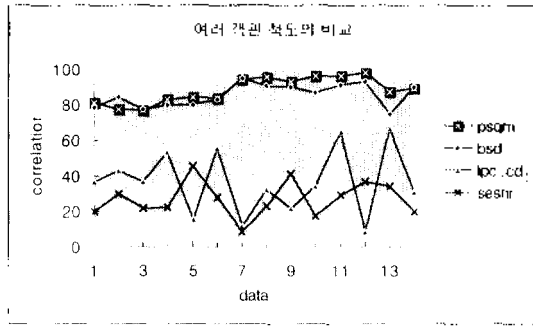


그림4-1 여러 가지 객관 척도의 결과 비교

#### 4-5 MPSQM의 성능 평가

이 장에서는 기존의 PSQM 알고리즘과 본 논문에서 개발한 MPSQM의 성능을 MOS와의 상관 관계 값을 이용하여 비교 평가하였다. 앞 절에서도 설명하였듯이 MPSQM은 그림 4-2에서 보여지듯이 지워진 frame 구간은 찾을 후 exponential 함수로 모델링한 값을 적용시켜 기존의 PSQM 알고리즘에 적용하였다. 그 결과 표 4-2에서 보여지는 것처럼 PSQM도 높은 성능을 보여 주지만 frame erasure가 많이 발생한 음성 신호(업 링크)에서는 성능이 떨어질 수 있다. 반면에 그림 4-3에서 보여지는 것처럼 MPSQM은 전체적인 성능이 PSQM보다 더 우수했다. 그러므로 채널을 통과하면서 생기는 여러 영향 중 frame의 손실이 객관 척도 알고리즘에 주는 영향을 줄여 그 성능을 개선 시킬 수 있었다.

음성에 있어 업선과 내선에 대한 차이보다는 링크가 어떻게 구성되었느냐가 결과를 좌우한다. 즉, 같은 음성을 업선으로 받은 것이나 내선으로 받은 것이나 결과가 비슷하게 나오지만 업링크나 다운링크에 따라서는 결과가 크게 다르다.

그림 4-4에서 알 수 있듯이 PSQM의 업링크 결과가 MPSQM의 업링크에 대한 결과 값보다 낮다. 그 이유는 음성 데이터의 파형들을 분석해 보니 음성 신호 중 업링크에서의 지워진 frame 수가 다운링크에 비해 많았다. 이러한 영향을 고려한 MPSQM 알고리즘에 의한 결과는 PSQM의

결과보다 높게 나왔음을 알 수 있다.

표4-2 PSQM과 MPSQM의 성능 비교

	데이터	PSQM	MPSQM
국선/Up link	남1	81.4	87.1
	남2	77.7	90.7
	남3	77.1	84.1
	여1	83.3	87.4
	여2	84.1	86.1
국선/Down link	남1	94.1	95.9
	남2	95.8	94.5
	남3	93.4	96.5
	여1	96.3	91.0
	여2	96.3	95.2
내선/Up link	남1	87.8	89.5
	여1	89.5	88.2

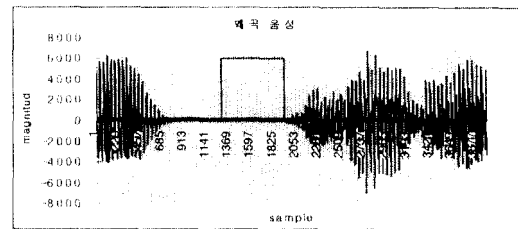
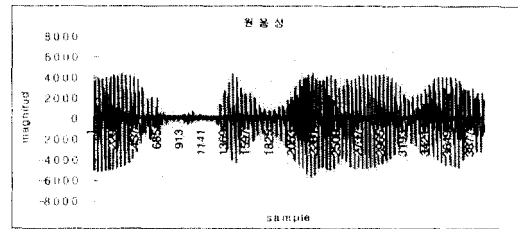


그림4-2 MPSQM에서 frame erasure 구간 검출

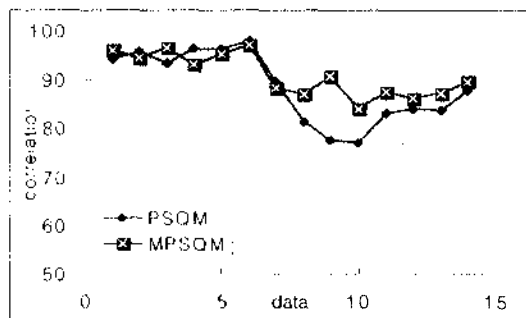


그림4-3 PSQM과 MPSQM의 성능 비교

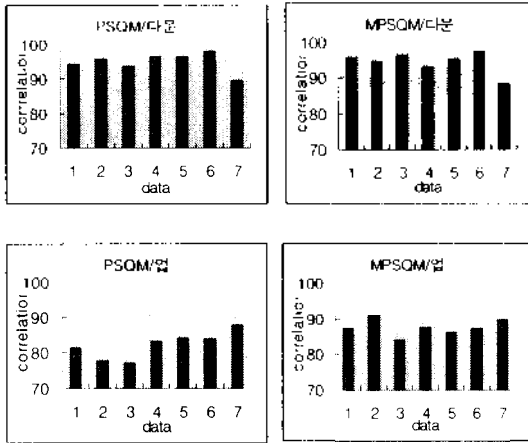


그림4-4 Up/Down link에 따른 비교

## 제 5 장 결 론

본 논문에서는 CDMA 채널을 통과하면서 여러 가지 형태로 손상된 음성에 대한 주관적 척도 결과를 예측할 수 있는 객관적도의 알고리즘에 대한 연구를 수행하였다. 반복 실험을 통하여 기존의 소개된 여러 객관적도 알고리즘 중에 인간의 심리음향의 모델을 적용한 BSD와 PSQM이 MOS와 가장 높은 상관 관계를 가지고 있으며 SegSNR와 LPC\_CD 등은 낮은 상관 관계를 가지고 있음을 확인하였다. 이 중 특히 PSQM은 좋은 성능을 보인다. 그러나 음성에 시위전 frame 수가 많으면 PSQM의 성능이 떨어짐으로 FER를 고려할 필요성이 있다. 따라서 기존의 PSQM을 수정하여 CDMA 채널 환경에 맞는 MPSQM 알고리즘을 제안하였다. 실험 결과 MPSQM 알고리즘은 기존의 알고리즘에 비하여 전체적인 성능과 신뢰성이 크게 향상되었다.

향후 연구 과제로는 실제의 통화 환경하에서 즉, 잡음이 섞인 음성 신호에 대해 고려하여 연구해야 하며, 같은 음성 신호에 대해 업 링크와 다운 링크의 결과가 왜 다른지에 관한 과제이다. 또한 코덱이나 화자의 변화에 독립적인 결과를 갖는 객관적도의 개발을 위한 연구가 진행 중이다.

## 참고 문헌

- [1] Sadsoki Furui, M. Mohan Sondhi. Advance in Speech Signal Processing, Dekker
- [2] Nynek Hermanky, Percepture Liner Prediction (PLP) analysis of speech, J. Acoust. Soc. Am., vol.87, pp1734-1752, 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] ITU-T "Model for predicity Transmission Quality from objective Measurements", Supplement3(Series p), March 1993.
- [8] K.H.Lam, O.C. Au, et al, Objective Speech Measures for Chinese in Wirelees Environment. in Proc. IEEE Int. Conf. Acoust., Speech Signsl Pricess., vol 1, pp277-280, May 1995
- [9] N.R.Draper, H.Smith, Applied Regression Analysis, John Willey & Sons, New York, 1981.
- [10] ITU-T RECOMMENDATION P. 861, Objective Quality Measurement of Telephone-Band (300-3400Hz) Speech Codecs.