디지털 이동통신망 환경 하에서 마스킹 효과를 이용한 객관적 음질 평가 척도

김광수[†] · 김민정[††] · 석수영[†††] · 정호열[††††] · 정현일[†††††]

요 약

본 논문에서는 이동전화망 환경 하에서의 음성의 통화품질 평가를 위해 마스킹 문턱치를 이용하는 객관적 음질평가법을 제안하고 실험을 통하여 그 유효성을 확인하였다. 현재까지 잘 알려진 BSD(Bark Spectral Distortion), PSQM(Perceptual Speech Quality Measure) 등의 성능을 먼저 분석하였다. 그 결과, MOS(Mean Opinion Score)와의 상관성이 이동통신 환경하에서 문헌상에 보고된 결과보다 성능이 저하됨을 확인하였다. 이동통신 환경하에서 보다 효율적인 객관적 음질평가척도 개발을 위하여 제안된 방법에서는 인간의 심리음향학적 마스킹 현상을 이용하였고, 성능 평가시 비교대상인 주관적 음질척도로는 DMOS(Degradation MOS)를 사용하였다. 디지털 이동통신망에서 수집된 음성 데이터에 대한 성능평가실험을 수행한 결과, BSD와 PSQM 같은 기존의 척도들에 비하여 평균 4%의 상관성능이 향상됨을 확인하였다.

An Objective Speech Quality Measure using Masking Effect under Digital Mobile Telephone Network Environment

Kwang-Soo Kim[†], Min-Joung Kim[††], Soo-Young Suk[†††], Ho-Youl Jung[††††] and Hyun-Yeol Chung[†††††]

ABSTRACT

In this paper, we propose a new objective speech quality measure using noise masking threshold for speech quality assessment of mobile telephone network environments, and verify the effectiveness of the proposed method through the experiments. For such a purpose, well known objective speech quality measures such as BSD and PSQM are first evaluated for digital mobile telephone network environments. However, these conventional methods does not have good performance under mobile networks environments compared to literary results. To be more effective objective speech quality measure under mobile telephone environments, the proposed method employs human psychoacoustic masking effect. The DMOS, instead of MOS, is used as a subjective speech quality measure for performance evaluation. The performance comparison are carried out with speech data collected from digital mobile telephone environments. As results, the proposed measure have and average 4% higher performance, in terms of correlation, than existing objective speech quality measures such as BSD and PSQM.

Key words: 객관적 음질평가척도, 음질평가, 디지털 이동통신

1. 서 론

최근 디지털 통신기술의 발전으로 이동전화의 보급이 급격히 증가되고 있다. 그러나 이동전화의 발성환경, 통신경로 등의 변화로 인하여 통화음의 품질이 크게 열화되어서 통화음질에 대한 불만을 호소하는 이용자가 늘고 있다. 이동전화망 환경하에서 이동 전화 사용자가 체감하는 음성의 통화품질을 알기 위해서는 반복 청취 실험에 의한 주관적인 평가를 실시해야 하지만 이 방법은 시간과 비용이 많이 소요되며,

[†] 정회원, 경운대학교 컴퓨터전자정보공학부 정보통신공학과
[††] 영남대학교 대학원 정보통신공학과(박사수료)
[†††] 영남대학교 대학원 정보통신공학과(박사과정)
[††††] 영남대학교 전자정보공학부 조교수
[†††††] 영남대학교 전자정보공학부 교수

평가자의 심리적 상황 변화에 따라 평가 결과가 달라지는 경우가 많아 평가의 일관성이 부족하다는 단점이 있다[1,2]. 따라서 객관적 음질평가 척도를 이용하여 주관적 음질을 예측하는 방법의 개발이 요구되어 오고 있다. 객관적 음질 평가 척도는 원래의 음성과 왜곡된 음성을 비교하여 자동적으로 주관적 점수를 추정하는 기법으로서 다양한 형태의 왜곡환경에 대해서 주관적 음질 평가 척도인 MOS(Mean Opinion Score) 또는 DMOS(Degradation MOS) 평가결과와 상관성이 높을수록 좋은 척도로 볼 수 있다[2-4].

지금까지 몇 몇 객관적 음질 평가 척도가 제안되어 통신 시스템, 코덱의 개발 및 평가에 이용되어 왔다[2,4,13]. 이들 음질평가척도들은 크게 시간영역, 스펙트럼 영역, 지각적 영역 기반 척도들로 구분된다. 일반적으로 사용되는 SNR(Signal-to-Noise Ratio), SegSNR(Segmental SNR) 등은 시간 영역 척도에 속하며, 주로 파형 부호화기의 음질을 평가하는데 사용되어왔다. 최근의 부호화기들은 인간의 발성 모델을 기반으로 설계되어 있기 때문에, 시간 영역이나 스펙트럼 영역 기반 음질평가척도들은 적합하지 못하다. 최근 들어 가장 널리 연구되어 오는 지각적 왜곡 척도는 원음성과 왜곡된 음성에 인간의 청각적 현상을 반영한 심리음향 모델을 적용하여 두 신호의 왜곡정도를 측정하는 방법이며 주관적 품질과 매우 높은 상관성을 보여주고 있다고 알려져 있다. 이 척도는 임계대역 분석(critical band analysis), 등감곡선 보정(equal-loudness preemphasis), 주관적 세기 보정(intensity-power law)의 세 가지 단계의 심리음향학적인 인간의 청각 특성을 고려하며 음성신호를 라우드니스 영역 또는 Bark 영역과 같은 지각적 영역으로 변환하고 청각 모델을 결합시킨다[2,4,13]. BSD(Bark Spectral Distortion)가 최초로 개발된 척도이다[12,13]. 한편 PSQM(Perceptual Speech Quality Measure)은 ITU-T(International Telecommunications Union-T)에 의하여 P.861 권고안으로 채택되었으며, 낮은 전송률을 가지는 음성 부호화기의 평가에 사용되는 심리음향학적 음질 평가 척도이다[4].

이와 같은 척도들은 대부분 부호화기의 평가를 그 목적으로 하고 있기 때문에, 유무선 통신환경에서 발생하는 여러 왜곡 요인들을 고려하지 않고 있다[3]. 그러나, 이동 전화와 같은 무선 전송채널 환경에는 채널의 대역 통과 특성과 부가되는 잡음, 다중 경로 페이딩, 프레임 삭제, 가변 지연 등이 복합적으로 존재한다[4,5]. 그러므로 이와 같은 다양한 상황을 반영할 수 있는 이동 전화의 주관적 음질을 간접적으로 예측할 수 있는 음질 자동평가 척도의 개발이 필요하다. 또한 다양한 환경왜곡 하에서 신뢰성이 높은 객관적 음질 평가 척도가 되기 위해서는 마스킹 효과와 같은 인간의 지각특성을 보다 효율적으로 고려하여야 한다.

또한 음질평가를 위한 주관적 음질척도로서는 왜곡된 음성만을 들려주는 MOS 보다는 원래의 음성과 왜곡된 음성을 동시에 들려주어 평가하게 하는 상대평가법으로서의 DMOS 평가가 더 효과적이라고 볼 수 있다[3].

따라서 본 논문에서는 이동전화 환경에 적합한 새로운 음질평가척도 개발을 위하여 기존의 평가척도들의 문제점을 검토한 후, 성능 개선을 위한 방법으로서 인간의 심리음향학적 현상인 마스킹 효과를 척도개발에 도입하고 그 유효성을 확인하고자 한다. 또한, 객관적 음질평가척도의 성능평가방법으로서 MOS 보다는 DMOS를 사용하여 보다 효율적인 성능평가가 이루어질 수 있도록 하였다.

본 논문의 구성은 다음과 같다. 2장에서 기존의 주관적 음질 척도와 객관적 음질 척도를 소개하고 3장에서는 본 논문에서 제안하는 마스킹 효과를 고려한 객관적 음질평가척도에 관하여 서술한다. 4장에서 성능평가실험 결과를 논의한 후, 5장에서 결론을 맺는다.

2. 음질의 평가방법

음질 평가방법에는 청취자들이 직접 듣고 판단하는 주관적 음질 평가방법과 원래의 음성과 왜곡된 음성과의 수치적 차이를 이용하는 객관적 음질 평가방법이 있는데 여기서는 기존의 방법에 대해 간단히 서술한다. 또한 객관적 음질 평가 척도의 성능을 평가하는 방법에 대해서도 살펴본다.

2.1 주관적 음질 척도

청취자의 주관적 등급에 기반하고 있는 음질 척도를 주관적 음질 척도라고 부른다. 이 척도들은 객관적 음질의 성능이 주관적 품질을 예측하는 능력에 의하여 평가되기 때문에 매우 중요한 역할을 한다. 청취자들은 정의된 등급에 따라 음성을 듣고 등급을

부여한다. 이 방법은 간단하기는 하나 시간과 비용이 매우 많이 소요된다. 주관적 품질 척도는 청취자 대부분의 청각적인 반응이 비슷하기 때문에 적절한 숫자의 청취자들에 의한 평가가 모든 청취자들을 대표할 수 있다고 가정한다. 가장 널리 사용되는 주관적 음질 척도는 MOS와 DMOS이다[1-3].

• MOS(Mean Opinion Score)

MOS는 음질을 추정하는 데에 가장 널리 쓰여지고 있는 방법으로 절대음질평가법(ACR : Absolute Category Rating)이다. 청취자들은 표1과 같은 5가지 등급을 사용하여 원래의 음성은 듣지 않고 평가대상이 되는 음성만을 청취한 후, 그 음성품질의 등급을 부여한다.

MOS의 장점은 음질에 대한 등급부여가 자유롭다는 반면 청취자들에 의한 등급의 변동이 매우 크다는 단점을 가지고 있다. 이 변동은 수많은 청취자들에 의하여 실시함으로써 보상할 수 있다. 따라서 적어도 40명 이상의 평가자에 의한 MOS 점수를 사용하도록 ITU-T P.800에서 권고하고 있다[3].

표 1. MOS와 음질 등급

등급	음질
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

• DMOS(Degradation MOS)

DMOS는 청취자들이 시험 대상인 원래의 음성과 왜곡된 음성을 비교청취하므로서 상대적인 음질열화(speech quality degradation)에 등급을 부여하는 상대적 음질평가법(DCR : Degradation Category Rating)이다[3]. 표2에 DMOS 평가의 5개 등급을 나타내었다. 그러나, DMOS 방법은 절대적인 왜곡량뿐 아니라 왜곡 형태에 대하여 의존할 가능성이 있기 때문에 여러 가지 다른 형태의 왜곡을 비교하기에는 어려움이 있다.

2.2 객관적 음질 척도

객관적 평가척도들은 여러 가지가 있으나 이들 척도의 성능 평가 시스템의 기본 구조는 그림 1과 같다.

표 2. DMOS와 Degradation 등급

등급	Degradation level
5	Inaudible
4	Audible but not annoying
3	Slightly annoying
2	Annoying
1	Very Annoying

그림 1에서처럼 원래의 음성과 왜곡된 음성신호로부터 주관적 음질평가척도와 객관적 음질평가척도를 각각 구하고, 객관적 음질평가척도로부터 추정된 주관적 음질평가척도는 MOS와 DMOS에 의하여 측정된 주관적 음질평가척도와의 상관성 분석에 의하여 그 성능을 평가한다. 상관성이 높을수록 주관적 음질을 잘 예측하는 객관적 음질평가척도이다.

시간영역이나 스펙트럼 영역기반 왜곡척도는 파형 부호화기의 평가에서는 많이 이용되어 왔다. 그러나, 최근의 부호화기는 단순한 음성파형의 재생보다는 음성발성 모델을 적용하여 원래의 음성과 동일한 음성을 합성하도록 설계되어 있기 때문에 거의 사용되지 않는다. 최근 들어 가장 널리 연구되어 오는 지각적 왜곡 척도는 원음성과 왜곡된 음성에 인간의 청각적 현상을 반영한 심리음향 모델을 적용하여 두 신호의 왜곡정도를 측정하는 방법이다. 심리음향 모델을 이용한 척도는 임계대역 분석(critical band analysis), 등감곡선 보정(equal-loudness preemphasis), 주관적 세기보정(intensity-power law)의 세 가지 단계의 심리음향학적인 인간의 청각 특성을 고려한다[2,4,13]. 이 척도에서는 음성신호를 라우드니스 영역 또는 Bark 영역과 같은 지각적 영역으로 변환하고 청각 모델을 결합시킨다. 여기서는 음질평가척도로서 현재 많이 이용되고 있는 BSD와 ITU에 의하여 권고되고 있는 PSQM을 평가기준으로 검토하기로 한다.

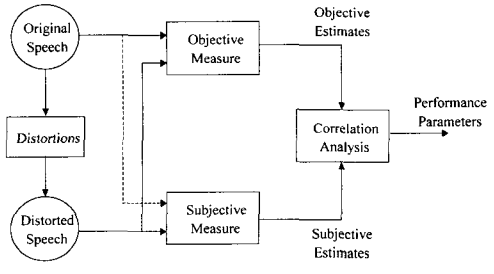
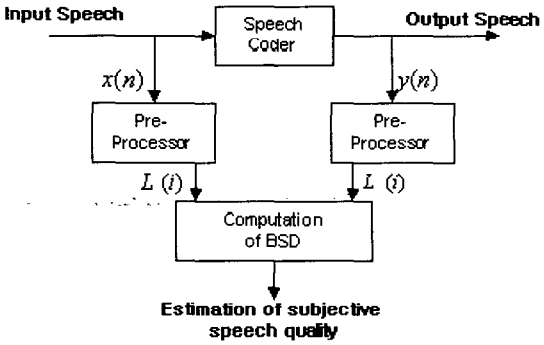


그림 1. 객관적 음질척도의 성능평가 시스템 구성

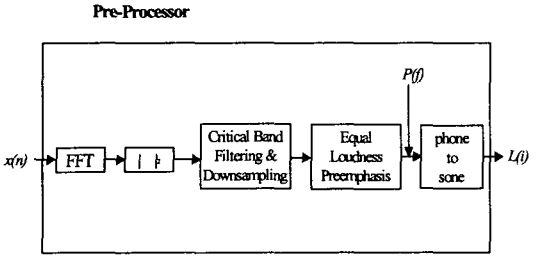
2.2.1 Bark Spectral Distortion(BSD)[13]

BSD 척도는 부호화에 의한 왜곡을 측정하기 위하여 개발되었다. 심리음향학적 모델을 사용한 최초의 객관적 음질평가척도로서 음질과 음성 라우드니스가 직접 관련되어 있다는 가정에 기초하고 있다. 소리에 대한 임계대역 특성을 Bark scale로 나타내는데 이 과정에서는 위에서 언급한 심리음향학적 청각모델의 세 가지 단계를 거쳐 음성신호의 power 스펙트럼 $P(f)$가 Bark 스펙트럼 $L(i)$으로 변환된다. 이를 그림 2에 나타내었으며 (a)는 BSD의 전체계산과정 블록도이고 (b)는 세 가지 단계의 pre-processor이다.

먼저 pre-processor 단계에서는 원음성 $x(n)$과 음성부호화기를 통과하여 왜곡된 음성 $y(n)$은 각각 동일한 과정을 거쳐 Back spectrum $L_x(i)$와 $L_y(i)$을 구하게 된다. 이 pre-processor 단계의 세부과정은 다음과 같다. 각 음성신호의 전력 스펙트럼 $|X(f)|^2$, $|Y(f)|^2$을 구하여 임계대역 필터링 과정을 거친다. 필터링은 크게 두단계로 나누어지는데, 식(1)의 관계식에 의하여 Herz-to-Bark 변환되고 이 때 사용되는 임계대역 필터 $F(b)$는 식(2)의 형태를 가진다.



(a) Calculatiion of BSD



(b) Pre-Processor in BSD

그림 2. BSD척도 계산과정

$$f = Y(b) = 600\sinh(b/6) \tag{1}$$

$$10\log_{10}F(b) = 7 - 7.5(b-0.215) - 17.5[0.196 + (b-0.215)^2]^{1/2} \tag{2}$$

여기서 f는 주파수(Hz)이고, b는 Bark-scale이다.

변환된 Bark 스펙트럼은 귀의 등감곡선(Equal-loudness curve)에 의하여 보정되어진다. 이때 보정을 위한 preemphasis filter는 아래의 형태를 가진다.

$$H(z) = \frac{(2.6 + z^{-1})}{(1.6 + z^{-1})} \tag{3}$$

마지막으로 P(phon)-to-L(sone) 변환을 거쳐 Bark 스펙트럼을 얻게 된다. Phon(P)이란 소리의 라우드니스 레벨의 단위이며, sone은 주관적인 라우드니스의 scale이다. i번째 라우드니스 $L(i)$는 다음 수식에 의하여 얻어진다.

$$L(i) = \begin{cases} 2^{(P-40)/10} & \text{if } P \geq 40 \\ (P/40)^{2.642} & \text{if } P < 40 \end{cases} \tag{4}$$

이상의 과정은 그림 2(b)의 preprocessor 단계에 해당된다[13].

BSD척도는 원 음성과 왜곡된 음성과의 Bark 스펙트럼 차이를 계산하며 식(4)로부터 구한다.

$$BSD^{(k)} = \sum_{i=1}^{N}[L_x^{(k)}(i) - L_y^{(k)}(i)]^2 \tag{5}$$

여기서, $L_x^{(k)}(i)$, $L_y^{(k)}(i)$는 k번째 입, 출력프레임의 Bark 스펙트럼을 나타내며, N은 임계대역의 수이다.

2.2.2 Perceptual Speech Quality Measure (PSQM)

PSQM은 ITU-T 권고안 P.861로 채택된 척도이다[2,4]. 이 척도는 부호화에 의한 왜곡만이 존재한다고 가정한 경우에 적용가능하도록 설계되어있다. 객관적 음질평가 척도 중에서 가장 정교한 심리음향학적 모델을 사용하여 기존의 여러 척도들과의 성능비교 실험결과 가장 우수한 성능을 나타낸다고 알려져 있다[2,4]. 최대한의 성능제고를 위하여 음성 신호를 심리음향학적 영역에서의 음성신호의 크기를 나타내는 라우드니스 영역으로 변환한다. 또한, 왜곡의 계산에서 묵음 부분에 작은 weight를 부가하거나 무시한다. PSQM에서는 음성을 perceptual 영역으로 변환하기 위하여 라우드니스 계산의 심리음향학적인 결과를 사용하며, 성능을 최적화하기 위하여 라우

드니스 계산의 절차를 변경한다. 그리고 전체적인 음성에서의 묵음 부분에서의 왜곡을 고려한다. 그림 3는 PSQM의 계산과정을 블록도로 나타내고 있다[2,4].

2.3 객관적 척도의 평가

객관적 음질척도로부터 주관적 음질(MOS)을 예측하기 위하여 통계적인 회귀분석을 수행하는데, 통상 식(6)과 같은 이차 회귀 함수가 널리 사용되고 있다[4,9,10,15,16].

$$\hat{x}(i) = ax(i)^2 + bx(i) + c \tag{6}$$

여기서, $x(i)$는 객관적 척도이고, $\hat{x}(i)$는 예측된 주관적음질척도인 MOS값이다.

객관적 음질 척도와 주관적 음질 척도와의 상관계수는 객관적 음질 척도를 평가하는 기준으로 사용되어 왔다. MOS값과 예측된 MOS값 사이의 상관계수(ρ)는 식(7)에 의하여 구해진다.

$$\rho = \frac{\sum_{i=1}^{N} \hat{x}(i)y(i)}{[\sum_{i=1}^{N} \hat{x}(i)^2 \sum_{i=1}^{N} y(i)^2]^{1/2}} \tag{7}$$

여기서, $\hat{x}(i)$는 측정된 주관적 MOS값이고, $y(i)$는 객관적 음질 척도로부터 예측된 MOS값이다.

상관계수 ρ=1인 경우, MOS 예측오차가 없음을 의미한다.

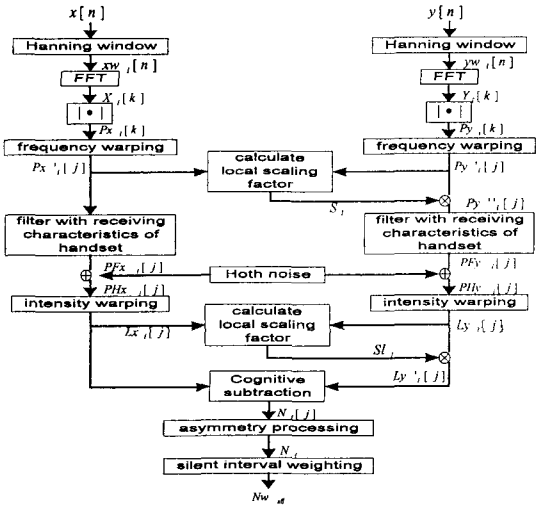


그림 3. PSQM 계산과정

이 절에서 기술한 기존의 객관적 음질평가척도들은 무잡음 환경에서의 음질평가를 위하여 개발되었기 때문에 부호화에 의한 왜곡만이 존재하는 경우에는 우수한 성능을 보이는 것으로 보고되어있다. 그러나 실제 이동통신 환경에서는 부호화에 의한 왜곡뿐만이 아니라, 페이딩, 부가잡음, 채널왜곡 등 많은 요인에 의하여 왜곡이 발생하게 된다. 이러한 요인들에 의한 왜곡이 부가될 경우 그 성능이 저하된다는 문제점을 가지고 있다.

3. 잡음 마스킹을 이용한 객관적 음질평가척도

본 절에서는 위에서 기술한 바와 같은 객관적 음질 평가척도의 성능을 저하시키는 여러 가지 요인들 중 무선 전송채널 환경하에서의 채널의 대역 통과 특성과 부가되는 잡음에 의한 영향을 줄이기 위하여 심리학적 마스킹 효과를 이용하는 객관적 음질평가척도를 제안한다. 제안하는 방법은 BSD에서 사용되는 perceptual 모델에 마스킹 효과를 추가적으로 고려하여 구성한다. 본 절에서는 먼저 잡음마스킹 문턱치에 대하여 살펴보고, 이 문턱치를 이용하는 객관적 음질평가척도를 기술한다.

3.1 잡음 마스킹 문턱치(Noise Masking Threshold : NMT)의 계산[7]

사람의 청각은 여러 대역에서의 음성신호가 동시에 들릴 때 특정 레벨 이하의 음성은 지각하지 못하는 특성이 있다. 이러한 특성을 마스킹 효과라고 부르며 음성이나 오디오신호의 압축에 적용되어왔다. 따라서 잡음 마스킹 문턱치 이하의 음성신호는 들리지 않으나, 왜곡량의 계산에는 포함되어 있기 때문에 잡음 마스킹의 문턱치를 적용하여 지각적인 왜곡량의 계산과정에서 제외하는 것이 더욱 바람직하다. 따라서, 객관적 음질 평가 척도의 계산시 실제로 귀에 들리는 가청왜곡(audible distortion)만을 계산하도록 하기 위해 잡음마스킹 문턱치를 도입하였다. 그림 4에 이 과정을 블록다이아그램으로 나타내었으며 잡음마스킹 문턱치 계산과정은 MPEG에서의 오디오 신호의 압축에 사용되는 것과 같은 과정을 거친다[7].

이 척도는 크게 라우드니스의 계산과 잡음마스킹 문턱치(NMT : Noise Masking Threshold)의 계산의 두 가지 단계로 나누어진다. 라우드니스 계산에서

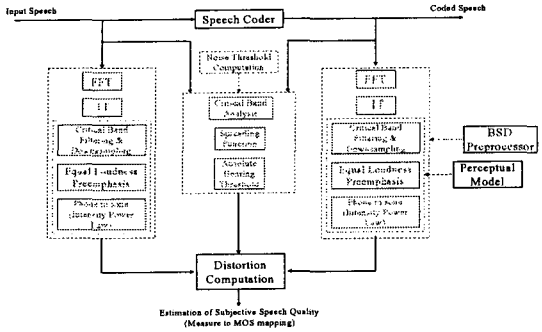


그림 4. 잡음마스킹 문턱치의 적용

는 음성신호를 라우드니스 영역으로 변환하는데 2.2.1절에서 설명한 방법과 동일한 절차를 거친다. 잡음마스킹 문턱치 계산에서는 임계대역 분석, spreading 함수의 적용, spreading 마스킹 문턱치의 계산, absolute threshold를 고려하여 추정한다. 먼저, FFT에 의한 전력스펙트럼 $P(w)$으로부터 계산되는 각 임계대역별 에너지의 합을 다음 식과 같이 얻는다.

$$B_i = \sum_{\omega=bl_i}^{bh_i} P(\omega) \tag{8}$$

여기서 bl_i는 임계대역 i의 하한, bh_i는 임계대역 i의 상한이며 B_i는 임계대역 i에서의 에너지이다.

Spreading 함수는 임계대역들 사이의 마스킹 효과를 추정하기 위하여 사용되며 행렬의 형태를 가진다. (8)식에서 구한 임계대역 에너지에 convolution 연산을 통하여 spread 임계대역 스펙트럼을 (9)식과 같이 구할 수 있다.

$$C_i = S_{ij} * B_i \tag{9}$$

여기서 C_i는 spread 임계대역 스펙트럼, S_{ij}는 행렬형태의 spreading 함수이고 i는 마스킹되는 신호의 bark 주파수, j는 마스킹 신호의 bark 주파수이다.

마스킹 문턱치 계산은 크게 두 가지이다. 첫번째는 tone 마스킹 잡음에 대한 것이며 C_i이하 1.45+idB로 추정되고 i는 bark 주파수이다. 두번째는 tone을 마스킹하는 잡음에 대한 것인데 C_i이하 5.5dB로 추정된다. 그 신호가 잡음에 가까운가 tone신호에 가까운가를 결정하기 위해서 SFM(Spectral Flatness Measure)을 사용하며 전력스펙트럼의 기하평균(Gm)대 산술평균(Am)의 비로서 정의된다. dB 단위로 변환된 SFM은 (10)식과 같다.

$$SFM_{dB} = 10\log_{10}\frac{Gm}{Am} \tag{10}$$

또한 tonality의 계수 α를 사용한다.

$$\alpha = \min\left(\frac{SFM_{dB}}{SFM_{dB_{\max}}}, 1\right) \tag{11}$$

즉, $SFM_{dB_{\max}}$ = -60dB 이면 신호는 tone에 가깝고, 0dB이면 잡음에 가깝다고 추정할 수 있다.

각 대역별 마스킹 에너지에 대한 offset(O_i)은 다음 식과 같이 놓는다.

$$O_i = \alpha(14.5+i) + (1-\alpha)5.5 \tag{12}$$

문턱치 offset은 spread 임계대역 스펙트럼에서 차감되어 아래 식과 같이 spread 문턱치 추정 T_i가 된다.

$$T_i = 10^{\log_{10}(C_i - (O_i/10))} \tag{13}$$

이렇게 구하여진 문턱치는 에너지에 대한 재정규화(renormalization)에 의하여 다시 bark 영역으로 바뀌어지게 되며 T_i'로 표시한다. spreading 함수의 비안정성 때문에 deconvolution 대신 재정규화을 사용한다. Bark 영역에서의 문턱치값은 절대 문턱치와의 비교에 의하여 결정되어진다.

3.2 잡음 마스킹 문턱치를 이용한 객관적 음질평가척도의 제안

앞 절에서 계산된 잡음 마스킹 문턱치를 NMT라 표기하고 아래과 같이 고려하여 최종적인 perceptual 왜곡에서 제외시킨다. $L_x(n)$과 $L_y(n)$은 원래의 음성과 왜곡된 음성의 라우드니스 벡터이다. $D_{xy}(n)$은 라우드니스 벡터 사이의 차이이고, $NMT(n)$는 위에서 서술한 과정을 거쳐서 추정되어진 잡음마스킹 문턱치이다.

n-번째 프레임의 perceptual 왜곡 $D(n)$을 계산하기 위하여, $M(n,i)$는 n-번째 프레임의 perceptible 왜곡을 나타내며, I는 i-번째 임계대역이다. 왜곡이 perceptible이라면, $M(n, i)$는 1아니면 0중의 하나가 된다.

Perceptible 왜곡은 아래 식과 같이 잡음 마스킹 threshold $NMT(n,i)$에 대한 n-번째 프레임의 i-번째 라우드니스 차이($D_{xy}(n, i)$)를 비교하여 얻어진다.

$$M(n,i)=0, \quad \text{if} \quad D_{xy}(n,i)\le NMT(n,i)$$
$$M(n,i)=1, \quad \text{if} \quad D_{xy}(n,i)> NMT(n,i) \tag{14}$$

n-번째 프레임의 perceptual 왜곡은 잡음 마스킹 threshold 보다 더욱 큰 라우드니스 차이의 합으로써 정의되며 다음과 같이 계산된다.

$$Dist(n)=\sum_{i=4}^{18} M(n,i)D_{xy}(n,i) \tag{15}$$

여기서, $M(n, I)$와 $(D_{xy}(n, i)$는 각각 perceptible 왜곡과 n-번째 프레임에서의 i-번째 임계대역에서의 라우드니스 차이이며 $Dist(n)$은 n-번째 프레임의 perceptual 왜곡, 18은 임계대역의 수를 나타내며 1~3번 대역은 전화망의 경우에 있어서는 filter-out 되었다고 가정하여 제외되었다[4,9].

이러한 잡음마스킹 문턱치의 계산은 음성신호가 아닌 tone 신호 또는 협대역 잡음과 같은 정상상태 신호를 사용한 심리음향학적 실험에 기반하고 있다.

그러나, tone과 같은 신호를 사용하여 계산되는 잡음마스킹 문턱치는 음성과 같은 정상상태가 아닌 신호에 대하여 적용하는 것은 적절하지 않을 것으로 판단된다. 따라서 본 논문에서는 잡음마스킹 문턱치에 scaling factor를 적용하여 이를 척도설계에 반영하는 방법을 제안하며 scaling factor를 적용한식(14)를 다시 쓰면 아래의 식과 같다.

$$M(n,i)=0, \quad \text{if} \quad D_{xy}(n,i)\le \beta NMT(n,i)$$
$$M(n,i)=1, \quad \text{if} \quad D_{xy}(n,i)> \beta NMT(n,i) \tag{16}$$

여기서, $M(n, I)$는 n번째 프레임의 I번째 perceptible 왜곡이고, $(D_{xy}(n, I))$는 n번째 프레임의 I번째 라우드니스 차이이며, β는 scaling factor이다.

0.01의 step size로 scaling factor를 0.0에서 1.0까지 변화시켜 본 결과 0.76인 경우에 가장 우수한 성능을 나타내었다. Scaling을 적용한 척도의 계산과정을 그림5에 나타내었다. 제안하는 scaled 잡음마스킹 문턱치를 적용한 부분은 그림에 modified 잡음마스킹 문턱치로 나타내었다.

3.3 DMOS의 고려

앞에서 언급한 바와 같이 MOS의 장점은 음질에 대한 등급부여가 자유로운 반면에 청취자들에 의한 등급의 변동이 매우 크며 시간과 비용이 많이 소요되는 단점을 가지고 있다. 이를 이용한 객관적 음질 평

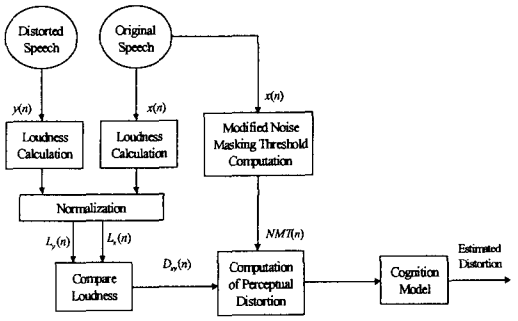


그림 5. Scaled 잡음마스킹 문턱치를 적용한 객관적 음질 평가척도

가 척도는 원래의 음성과 왜곡된 음성을 비교하여 주관적 음질을 예측하여 MOS 결과와 비교하기 때문에 부정확한 평가가 내려질 가능성이 있다. MOS와는 달리 상대적인 주관적 음질을 평가하는 DMOS는 청취자에게 원래의 음성과 왜곡된 음성을 들려주고 원 음성에 대한 왜곡에 대하여 상대적인 등급을 부여한다. 이는 그 절차상 객관적 음질평가척도에 가깝다고 볼 수 있다. 따라서 본 논문에서는 MOS 대신 DMOS에 대하여 객관적 음질 척도 성능을 평가하는 방법을 도입하기로 한다.

4. 평가 실험결과 및 분석

3.2와 3.3절에서 제안한 객관적 음질평가척도의 유효성을 평가하기 위해 다음과 같이 평가실험을 실시하였다. 평가를 위한 음성 데이터베이스 작성을 위한 원 음성은 연속음성 데이터베이스에서 각각 남성화자 2인과 여성화자 2인이 발성한 8초 정도의 길이를 가진 5문장을 선택하여 디지털 셀룰라 이동전화망을 통과시켜서 실제 환경에서의 음성데이터를 수집하였다. 음성데이터들은 주·야간 각3회에 걸쳐 차량을 정지 또는 정속 운행(60km 정도)하면서 이동전화의 사용량이 많고 전파환경이 다양한 대도시인 서울과 대구에서 수집하였다. DAT에 연결된 이동전화 단말기를 통하여 원 음성을 전송한 후, 일반 전화망에 연결된 유선 전화기를 통하여 DAT로 녹음하여 데이터베이스를 구축하였다. 이 과정을 나타내면 그림6과 같다.

주관적인 음질 평가는 DMOS를 사용하였다. 또한 MOS와의 비교를 위하여 MOS를 이용하는 주관적 음질평가법을 병행하였다. 청취시험에는 40명의 남

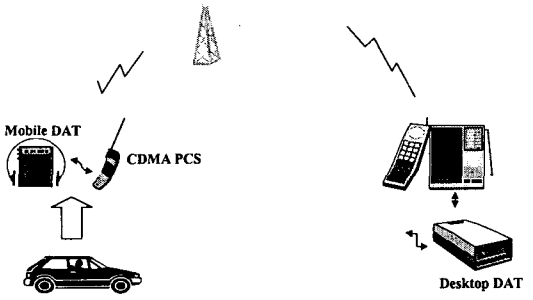


그림 6. 음성 데이터의 수집과정

녀가 참여하며, 방음장치가 되어있는 어학실습실에서 왜곡된 음성만을 들려주어 MOS를 측정하고 원음성과 왜곡된 음성을 차례로 들려주어 평가용지에 그 음성품질 등급을 부여하여 평균등급을 구한다. 이 전체적인 과정은 주관적 음질평가에 관한 ITU-T P.800 권고안에 따라 수행되었다[3].

디지털 셀룰라 이동전화망 환경을 통해 수집된 실제 음성에 대해 척도별로 얻어진 MOS와의 상관계수는 그림7과 같이 나타났다.

왜곡된 음성에 대해 MOS와의 상관성을 비교 조사한 결과 BSD와 PSQM척도는 0.813과 0.852의 평균 상관성을 보여주어 주관적 음질을 어느 정도 잘 예측할 수 있음을 알 수가 있으며 부호화 왜곡만이 존재하는 무잡음 환경뿐 아니라 실제 이동통신 전화망에 있어서도 이들 지각적 척도가 비교적 효과적인 객관적 척도임을 확인 할 수 있었으나, 문헌에 소개된 부호화에 의한 왜곡만이 존재하는 경우보다 이동통신망 환경에서는 성능이 약 평균 13%정도가 열화됨을 알 수 있었다.

또한, scaled 마스킹을 적용하는 제안된 방법의 경우 0.868의 가장 높은 상관성을 보였다. 이는 지각적 음질평가 척도인 BSD와 PSQM에 비하여 평균 5% 이상의 상관성능이 향상되어 이동통신망 환경에서 가장 유효한 객관척도임을 확인할 수 있었다.

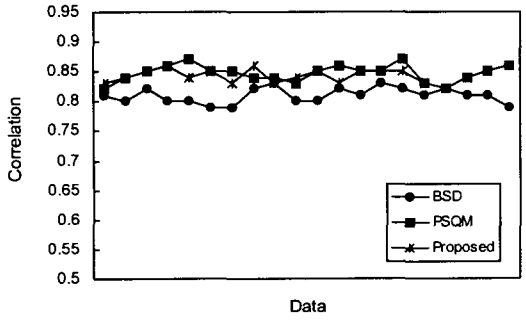


그림 7. MOS와 척도별 상관계수

DMOS와의 척도별 상관성 실험결과는 그림8과 같다.

DMOS와의 상관성능 실험 결과, BSD와 PSQM 그리고 제안한 객관적 음질 평가척도 모두에 대해서 평균 4%의 상관성능 향상을 보여, DMOS에 의한 평가방법이 더욱 유효함을 확인할 수 있었다.

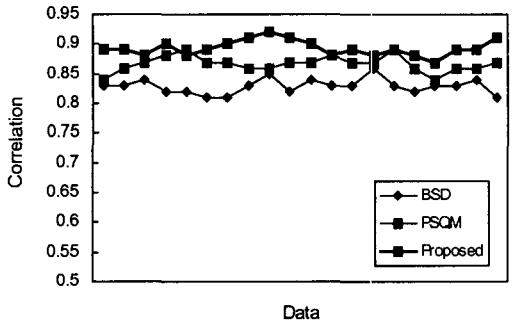


그림 8. DMOS와 척도별 상관계수

5. 결 론

본 논문에서는 이동전화망 환경 하에서의 음성의 통화품질 평가를 위한 객관적 평가척도로서 scaled 마스킹 문턱치를 이용하는 방법을 제안하였다. 또한 주관적 음질평가척도로는 DMOS(Differential Mean Opinion Score) 평가기법을 이용하였다. 실험 결과, BSD와 PSQM 척도가 각각 0.813과 0.852의 평균 상관성능을 보여, 이동통신 환경에서의 통화품질 평가 척도로서는 지각적 척도들이 비교적 유효한 것으로 나타났다. 그러나, 실제 이동전화망 환경에서 수집된 음성 데이터에 대해서는 문헌상에 나타났던 것보다 약 13%정도 상관성능이 열화됨을 알 수 있었다. 제안한 방법의 경우에는 MOS와의 0.868의 상관성능을 보여 기존의 두 가지 방법보다 약 5%의 성능향상이 있음을 확인하였다. 또한, MOS 대신 DMOS를 주관적 음질평가척도로 사용하였을 경우, 모든 객관적 음질평가척도의 상관성능이 약 평균 4% 정도 향상됨을 확인하였다.

참 고 문 헌

[1] J.G.Beerends and J.A.Stemerdink "A perceptual audio quality measure based on psychoacoustic sound representation", J. Audio Eng. Soc., vol.40, pp.963-978, Dec. 1992.

[2] J.G.Beerends and J.A.Stemerdink "A perceptual speech quality measure based on psychoacoustic sound representation", J. Audio Eng. Soc., vol. 42, pp.115-123, Mar. 1994.

[3] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality", Geneva, 1996.

[4] ITU-T Recommendation P.861, "Objective quality measurement of telephone-band (300 ~3400 Hz) speech codecs", Geneva, 1996.

[5] N.S.Jayant and P.Noll "Digital Coding of Waveforms : Principles and Applications to Speech and Video", Prentice Hall, 1984.

[6] C.Jin and R.Kubicheck, "Vector Quantization Techniques for Output-Based Objective Speech Quality" Proc. ICASSP, pp.491-494, 1996.

[7] J.Johnston, "Transform coding of audio signals using perceptual noise criteria", IEEE J. on Select. Areas in Commun., vol. SAC-6, pp.314-323, 1988.

[8] N.Kitawaki, H. Nagabuchi, and K.Itoh, "Objective quality evaluation for low-bit-rate speech coding systems", IEEE J. Select. Areas Commun., vol .6, pp.242-248, Feb. 1988.

[9] K.Lam, O.Au, C.Chan, K.Hui, and S.Lau, "Objective speech quality measure for cellular phone", ICASSP, vol. 1, pp.487-490, 1996.

[10] M.M.Meky and T.N.Saadawi, "A perceptually-based objective measure for speech coders using abductive network," ICASSP, vol. 1, pp.479-482, 1996.

[11] S.R.Quackenbush, T.P.Barnwell III and M.A. Clements, "Objective Measures of Speech Quality", Prentice-Hall, Englewood Cliffs, 1988.

[12] S.Voran and C.Sholl, "Perception-based objective estimators of speech quality", IEEE Speech Coding Workshop, pp.13-14, Annapolis, 1995.

[13] S.Wang, A.Sekey, and A.Gersho, "An objective measure for predicting subjective quality of speech coders", IEEE J. Select. Areas Commun., vol. SAC-10, pp.819-829, June 1992.

[14] E.Zwicker and H.Fastl, "Psychoacoustics : Facts and Models", Springer-Verlag, 1990

[15] 김광수, 정호열, 정현열, "시간/주파수 마스킹을 이용한 이동전화망에서의 객관적 음질평가척도에 관한 연구", 제 17회 음성통신 및 신호처리 학술대회 논문집, 2000.

[16] Kwang-Soo, Kim, Ho-Youl Jung and Hyun-Yeol Chung, "Evaluation of Objective Speech Quality Measures for the CDMA Wireless Telephone Network", Proceedings on International Conference on Speech Processing '99, 1999.

김 광 수

1994년 8월 경남대학교 전자공학과(공학사)
1998년 2월 영남대학교 대학원 전자공학과(공학석사)
2002년 8월 영남대학교 대학원 전자공학과(공학박사)
2001년 3월~현재 경운대학교 컴퓨터전자정보공학부 전임강사
관심분야 : 음성분석 및 인식, 음성 및 오디오 신호처리, 음질평가 등

김 민 정

1999년 영남대학교 대학원 멀티미디어통신공학과 (공학석사)
1999년~현재 영남대학교 대학원 정보통신공학과(박사수료)
관심분야 : 디지털신호처리, 음성처리, 음성인식, 화자인식 등

석 수 영

1998년 계명대학교 물리학과 (이학사)
2000년 영남대학교 대학원 멀티미디어통신공학과 (공학석사)
2000년 3월 ~ 현재 영남대학교 대학원 정보통신공학과(박사과정)
관심분야 : 디지털신호처리, 문자인식, 음성인식 등

정 흐 열

1988년 8월 아주대학교 전자공학과(학사)
1990년 8월 아주대학교 전자공학과(석사)
1998년 4월 INSA de Lyon, France (Ph.D)
1999년 3월 ~ 현재 영남대학교 전자정보공학부 전임강사
관심분야 : 디지털 신호처리, 디지털 워터마킹, MPEG, JPEG

정 현 열

1975년 2월 육군사관학교/영남대학교 전자공학과(학사)
1981년 8월 영남대학교 전자공학과(석사)
1989년 3월 동북대학교 정보공학과(공학박사)
1989년 3월 ~ 현재 영남대학교 전자정보공학부 교수
관심분야 : 디지털 신호처리, 문자인식, 음성인식