특집

오디오 품질 측정 기술

박호종 (광운대학교)

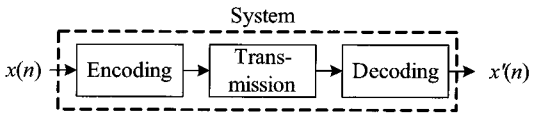
I. 서 론

멀티미디어 통신 및 방송에서 사용자가 청취하는 오디오 신호의 품질은 서비스의 종합 품질을 결정하는 매우 중요한 항목이다. 아날로그 통신 및 방송에서는 신호의 전송 과정에서 유입되는 잡음에 의하여 신호대 잡음비(SNR)가 감소하여 품질이 저하되므로, 채널의 성능에 따라 품질이 결정되는 비교적 단순한 구조를 가졌다. 그러나 최근 통신 및 방송이 디지털 방식으로 변경됨에 따라 전송 과정에서 다양한 디지털 처리가 수행되므로 출력 오디오 품질에 영향을 미치는 요인들이 크게 증가하였다. 예로, 채널의 성능에 따라 결정되는 전송 오류뿐만 아니라, 오디오 신호의 부호화 방법과 비트율(bit-rate)에 따라 오디오 품질에 큰 차이가 발생한다.

디지털 오디오 통신 및 방송에서 오디오 신호가 전송되는 과정은 <그림 1>과 같다. 최초의 아날로그 오디오 신호를 샘플링 하여 $x(n)$을 얻고, 이를 부호화(encoding) 하여 특정 비트율을 가지는 비트열을 출력한다. 수신단은 전송된 비트열에서 전송 오류를 보정하고 복호화(decoding)를 거쳐 최종 오디오 출력 신호 $x'(n)$를 얻는다. 이상의 전송 과정을 시스템이라 정의하고, 시스템에서 발생 하는 신호 왜곡에 의하여 $x'(n)$의 품질은 $x(n)$에 비하여 저하된다.

디지털 통신 및 방송에서 신호 왜곡의 가장 큰 이유는 부호화에 의한 정보 손실이며, 정보 손실량은 주로 부호화 방법과 비트율에 따라 결정된다. 따라서 서로 다른 시스템을 통하여 오디오를 전송할 경우, 각 시스템이 사용하는 부호화 방법과 비트율에 따라 품질 저하도 각각 다르게 나타난다.

오디오 통신 및 방송에서 다양한 목적에 따라 오디오 품질 측정이 필요하다. 오디오 서비스 사업자들은 낮은 비트율로 우수한 품질의 오디오 전송을 원하고, 다양한 부호화 방법과 비트율을 가지는 시스템의 품질을 비교 평가하여 최적의 방식을 선택하여 서비스를 제공한다. 또한 서비스 중에 실시간으로 오디오 품질을 계속 모니터링 하여 문제점을 점검하고 사용자의 불만을 사전에 방지하는 업무도 필요하다. 반면, 서비스 사용자들은 서비스를 받으면서 오디오 품질을 평



〈그림 1〉 디지털 오디오 신호의 전송 과정

가하고, 이를 바탕으로 서비스 만족도를 결정하고 서비스 개선 요구 등을 수행하게 된다.

시스템 개발자들은 자신의 시스템이 타사에 비하여 우수한 품질을 제공하여야 시장 경쟁력을 가지므로, 시스템이 제공하는 오디오 품질이 경쟁력 확보를 위한 핵심 항목이 된다. 오디오 부호화 설계자들은 새로운 방식의 부호화 알고리즘을 개발할 때 기존 방법과의 품질 비교를 통하여 향상된 오디오 품질을 제공하는 알고리즘을 개발한다.

이와 같이 오디오 품질은 서비스 제공자 및 사용자, 시스템 설계자, 개발자 등에게 매우 중요한 항목이고, 각자는 고유의 목적에 따라 오디오 품질을 측정하고 있다. 그러나 각자가 독자적인 방법으로 품질을 측정하면 통일된 기준이 없으므로 서로의 품질을 직접적으로 비교할 수 없어 활용 가치가 떨어진다. 또한 각자의 평가 방법에 의한 품질 측정 결과를 서로 신뢰하지 못하는 문제도 발생한다. 이 문제를 해결하기 위하여 표준화된 오디오 품질 측정 방법이 필요하며, 국제 표준화 기구에서 청각 이론과 통계 이론에 근거하여 체계적이고 정교하게 설계된 품질 측정 방법을 제정하였다. 현재 대부분이 표준화된 방법으로 품질을 측정하고 있으며, 표준 방법에 따라 정확하게 측정된 품질은 전 세계적으로 공인된 품질로 활용된다.

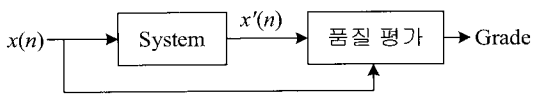
본 논문에서는 오디오 품질을 측정하는 기본 개념을 먼저 설명한다. 다음, 대표적인 세 가지 표준 오디오 품질 측정 방법을 소개하고, 각 방법의 세부 동작과 특징을 설명한다.

II. 오디오 품질 측정의 개요

오디오 신호는 수 많은 방법에 의하여 다양한 형태로 생성되므로 오디오 신호의 절대적 품질을 정의할 수 없다. 따라서 오디오의 품질은 <그림 2>와 같이, 평가할 대상 신호 $x'(n)$ (이를 평가음이라 함)의 품질을 원음 $x(n)$ (또는 기준 신호)의 품질과 비교하여 상대적 품질로 측정한다. 즉, 평가음의 품질이 원음 품질에 비하여 얼마나 저하되는지를 측정한다. 그리고 품질을 정량적으로 표시하여야 객관적인 자료로 활용하기 용이하므로, 품질은 점수(grade)로 표시한다.

오디오 품질은 원음과 평가음을 사람이 직접 듣고 평가하여야 가장 정확한 측정이 되며, 이를 주관적(subjective) 평가라 한다. 이 때, 개인별 선호도 또는 편견의 영향을 배제하기 위하여 가능한 많은 평가자들의 결과를 바탕으로 최종 품질을 결정해야 한다. 또한, 오디오 신호의 특성에 따라 오디오 품질이 변하므로 다양한 종류의 오디오 신호에 대한 품질을 측정하여 평균값을 구해야 한다. 따라서 주관적 평가는 정확한 품질 측정 결과를 제공하지만 많은 시간과 노력이 필요한 문제점을 가진다.

주관적 평가의 문제점을 근본적으로 해결하기 위하여 신호를 청취하지 않고 수학적으로 원음과 평가음의 차이를 분석하여 품질을 측정하는 방법을 사용할 수 있고, 이를 객관적(objective) 평가라 한다. 객관적 평가에는 평가자가 필요 없고, 컴퓨터 연산을 통하여 오디오 신호의 재생 시간보다 훨씬 빠르게 품질을 측정하므로 시간과 노력이 적게 필요한 장점이 있다. 그러나 청각 동작을 수학적으로 완벽하게 모델링 할 수 없으므



〈그림 2〉 오디오 품질 측정 과정

로 객관적 평가의 정확도는 주관적 방법에 비하여 많이 저하된다. 예로, 객관적 방법에서 높은 품질로 평가를 받은 오디오 신호를 실제 청취하면 심한 품질 저하를 느끼는 경우가 많다. 따라서 객관적 평가는 특별한 조건에서 제한적으로 사용하여야 하며, 주로 간단하게 대략적인 오디오 품질을 측정할 때 사용한다.

III. 주관적 오디오 품질 측정 기술

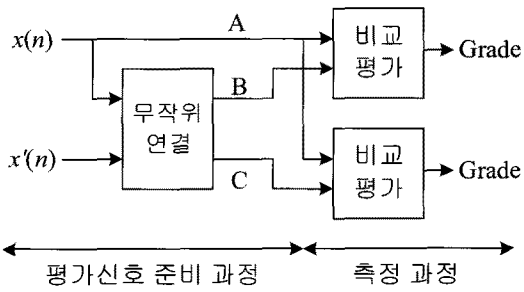
국제 표준화 기구인 ITU-R(International Telecommunications Union - Radiocommuni-cation Sector)에서 제정한 주관적 오디오 품질 측정 방법 중에 대표적인 두 가지 방법은 아래와 같다[1,2].

- ITU-R BS.1116 : Method for the subjective assessment of small impairments in audio systems including multichannel sound systems
- ITU-R BS.1534 : Method for the subjective assessment of intermediate quality level of coding systems

BS.1116은 품질 저하가 매우 적을 때 사용하고, BS.1534는 품질 저하가 중간 이상일 때 사용하도록 권고하고 있다.

1. BS.1116

BS.1116은 약간의 품질 저하가 발생하는 고품질의 오디오 시스템의 품질 측정을 위하여 제안되었다. 즉, <그림 1>에서 부호화 과정은 없으며, 시스템 동작으로 인하여 오디오 품질이 약간



〈그림 3〉 BS.1116 품질 측정 방법

저하될 때 미세한 품질 저하를 정교하게 측정할 때 사용된다.

BS.1116의 측정 방법은 "Double-Blind Triple Stimulus with Hidden Reference" 를 기반으로 한다. <그림 3>이 평가 신호 준비 및 측정 방법을 보여준다. 평가진행자는 원음 $x(n)$과 평가음 $x'(n)$으로부터 A, B, C로 표시된 세 신호를 만든다. 이 때, A는 항상 원음이고, B와 C는 원음과 평가음 중에서 하나씩이 무작위로 연결된다. 청취 평가자는 A, B, C로 표시된 세 신호를 청취한다. 물론, 평가자는 B와 C 중에서 어느 것이 원음인지 모르고, 이런 의미로 이를 Hidden Reference라 한다. 평가자는 A와 B를 청취하여 두 신호 사이의 품질 차이를 <표 1> 기준에 따라 1.0~5.0 사이에서 0.1 단위의 점수로 평가한다. 동일한 방법으로 A와 C 사이의 품질 차이를 평가한다. B와 C 중에서 하나는 원음이므로 두 점수 중에서 하나는 5.0이 되는 것이 이론적으로 정상이며, Hidden Reference의 목적

〈표 1〉 BS.1116 평가 기준

Impairment	Grade
Imperceptible	5.0
Perceptible, but not annoying	4.0
Slightly annoying	3.0
Annoying	2.0
Very Annoying	1.0

중에 하나가 평가 신뢰도를 평가하는 기준으로 사용하는 것이다.

품질 측정 진행에서 시간 제약은 없으며 평가자가 임의로 A, B, C 중에서 하나를 선택하여 청취 가능하고, 각 신호의 재생을 임의의 시점에서 시작할 수도 있다. 평가자가 A-B 및 A-C에 대한 평가 점수를 입력하면 $x'(n)$에 대한 품질 측정이 종료되고, 새로운 A, B, C 신호가 주어지고 청취자는 새로운 신호들을 청취하면서 두 번째 평가를 진행한다. 이 때, B와 C 신호의 연결 방법은 신호가 바뀔 때 마다 다시 무작위로 결정된다.

모든 평가자가 모든 오디오 신호에 대한 측정을 마치면 해당 시스템의 품질을 하나의 최종 점수로 표시한다. 평가자는 B와 C 중에 하나가 원음인 것을 알고 그를 기반으로 평가하므로, 평가음의 점수만으로 품질을 표시하면 왜곡된 결과를 얻는다. 따라서 오디오 품질은 식 (1)로 정의되는 SDG(subjective difference grade)로 표시한다.

$$\text{SDG} = \text{평가음 청취점수} - \text{원음 청취점수} \quad (1)$$

SDG는 이론적으로 음수값을 가지고, 0에 가까울수록 원음에 근접한 것을 의미한다. 모든 평가자와 오디오 신호에 대한 SDG를 통계처리 하여 평균값과 신뢰수준을 구하면 최종적인 품질이 결정된다.

BS.1116의 품질 측정에서 평가자는 미세한 품질 차이를 인지하여야 하므로 매우 높은 수준의 청각 능력을 가지는 전문가(expert)로 구성하여야 한다. 이를 위하여 평가 진행 전에 훈련 과정을 통하여 각 평가자의 청취 능력을 점검하고, 기준을 만족하지 못하는 평가자는 평가에서 배제한다. 또한, 평가 후에도 평가 결과를 바탕으로 신뢰도에 문제가 있는 평가자의 모든 점수는 최종 통계에서 배제 시킨다. BS.1116은 20명 정도의 평가자를 권고한다.

두 신호 사이의 품질 차이를 비교하려면 앞에 청취한 오디오 신호를 잠시 기억해야 하는데, 청각 기억력에 한계가 있으므로 평가에 사용하는 오디오 신호는 보통 10~25초 길이를 가진다. 오디오 청취는 스피커 청취 또는 헤드폰 청취가 모두 가능하다. 스피커로 청취하는 경우에는 청취 공간의 크기, 모양, 반사 특성, 청취자의 위치 등이 매우 중요하며, 표준안이 제시하는 조건을 만족해야 한다.

2. BS.1534

BS.1534는 중간 정도 이상의 오디오 품질 저하가 발생할 때 품질을 측정하는 방법이며, 부호화에 의하여 품질이 저하되는 디지털 통신 및 방송에서의 품질 측정에 가장 널리 사용된다. 이 방법은 "MUlti Stimulus test with Hidden Reference with Anchor(MUSHRA)" 방법에 기초하며, BS.1534를 통상적으로 MUSHRA 방법이라 한다.

특정 시스템의 품질을 단독으로 측정하는 경우도 있지만, 여러 시스템의 품질을 비교 평가하는 것이 필요한 경우도 있다. 예로, 오디오 방송 서비스의 최적 비트율을 결정 할 때, 각 비트율에 대한 품질을 측정하여 품질과 전송량을 기준으로 가장 우수한 비트율을 선택한다. 만일 각 시스템의 품질을 독립적으로 평가하고 그 결과들을 단순 비교하면, 각 평가에 포함되는 미세한 편차가 점수에 포함되어 동일한 기준으로 품질을 비교를 할 수 없다. 특히, 품질 저하가 큰 경우에는 편차가 더욱 커지게 되므로 BS.1116를 바탕으

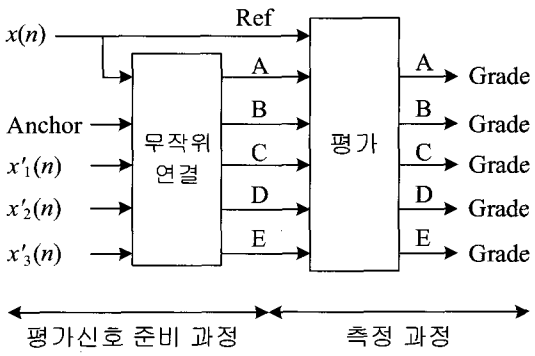
로 시스템 품질을 비교 평가 하는 것은 적절하지 않다.

여러 시스템의 품질을 정확하게 비교하기 위하여 MUSHRA는 각 시스템 평가음의 품질을 동시에 측정하는 전략을 사용한다. 또한, 중간 이상의 품질 저하에 대한 품질 기준을 잡기 위하여 원 신호를 3.5kHz 저역 통과시킨 Anchor 신호를 사용한다. 3.5kHz 저역통과 Anchor 신호 이외에 다른 Anchor 신호를 추가적으로 사용할 수 있다. 일반적으로, 7kHz 저역통과 신호, 잡음이 포함된 신호 등이 사용된다.

<그림 4>가 MUSHRA 평가 신호 준비 및 측정 방법을 보여준다. 평가진행자는 원음, Anchor, 여러 평가음으로부터 무작위로 A, B, C, ... 신호를 만들고, 여기에 포함된 원음을 Hidden Reference라 한다. 평가자는 공개된 원음과 A, B, C, ... 신호를 청취하면서 A, B, C, ... 신호의 품질을 0~100 사이의 점수로 각각 평가한다. 물론, 평가자는 A, B, C, ... 신호 중에서 어느 것이 원음 또는 Anchor인지는 모른다.

MUSHRA 평가를 편리하게 진행하기 위하여 <그림 5>의 화면으로 구성된 MUSHRA 평가 장치인 STEP 프로그램을 널리 사용한다[3]. 이를 기준으로 MUSHRA 과정을 설명하면, 청취자는 Ref(원음)과 A, B, C, D, E로 표시된 6개의 버튼을 클릭하여 해당 신호를 청취하고, 각 신호의 품질을 0~100 구간의 점수로 평가한다. 그림에서 보듯이 각 신호에 할당된 Scroll Bar를 이용하여 간단히 점수를 입력할 수 있다. 모든 점수 입력이 끝나고 NEXT를 클릭하면 현 신호에 대한 평가가 종료되고 다음 신호에 대한 평가가 진행된다.

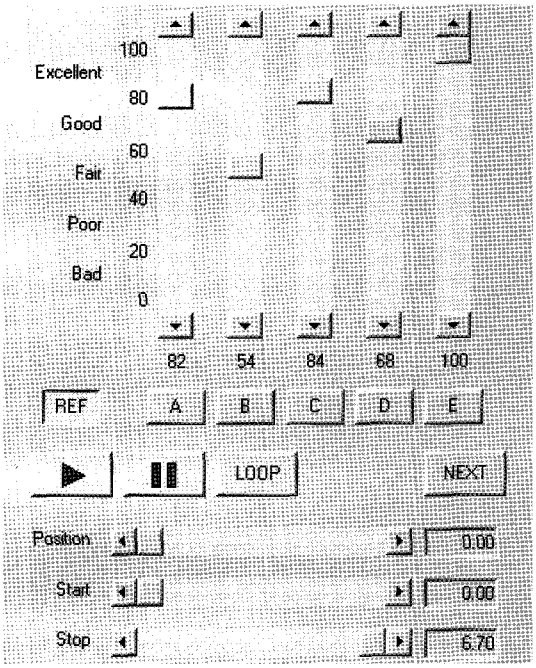
BS.1116과 동일하게 평가 진행에 시간 제약은 없으며 청취자가 임의로 Ref, A, B, C, D, E



〈그림 4〉 MUSHRA 품질 측정 방법

중에서 하나를 클릭하여 청취 가능하고, 화면 아래의 Time Bar를 사용하여 임의 시점에서의 재생 시작도 가능하다.

모든 평가자가 모든 오디오 신호에 대한 평가를 마치면, Hidden Reference와 Anchor 점수를 포함하여 각 시스템에 대한 모든 점수를 통계



〈그림 5〉 MUSHRA 평가를 위한 장치의 기본 화면

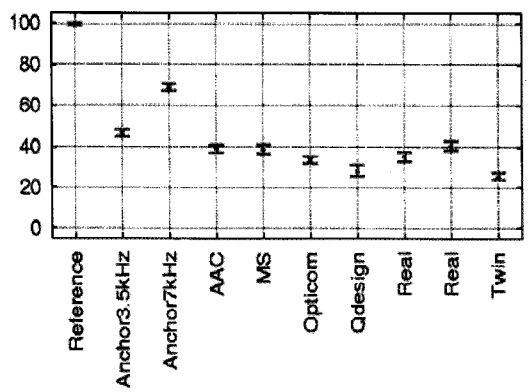
처리하여 최종 평가 결과를 제공한다. 먼저 식 (2)에 따라 시스템 j의 평균 점수 $\overline{u_j}$를 구한다.

$$\overline{u_j} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{k=1}^{M} u_{ijk} \quad (2)$$

여기서, i는 평가자 인덱스, j는 시스템 인덱스, k는 신호 인덱스이고, u_{ijk}는 해당 점수이다. 다음 $\overline{u_j}$의 표준편차 S_j를 구하고, 신뢰구간 δ_j를 식(3)에 따라 구한다.

$$\delta_j = t_{0.05}\frac{S_j}{\sqrt{NM}} \quad (3)$$

$t_{0.05}$는 95% 신뢰구간의 t 값이고, 최종적으로 각 시스템별로 $\overline{u_j}$와 $[\overline{u_j}-\delta_j, \overline{u_j}+\delta_j]$ 영역을 그래프로 나타낸다. 일반적으로 평가자 및 평가 신호의 수가 증가하면 신뢰구간 δ_j이 감소하여 통계적으로 보다 정확한 결과를 얻는다. 95% 신뢰구간 대신에 99% 신뢰구간을 사용하기도 하며, 99% 신뢰구간을 사용하면 δ_j 값이 증가한다. <그림 6>이 3.5kHz와 7kHz 저역통과 시킨 두 개의 Anchor를 사용하고 7개의 시스템에 대한 평가를 진행한 결과의 예를 보여준다.



〈그림 6〉 MUSHRA 평가의 결과 예

BS.1116에 비하여 평가자의 청각 능력에 대한 조건은 완화되지만, 평가를 진행하기 전에 훈련 단계에서 정확한 평가를 수행하지 못하는 평가자는 평가에서 제외시킨다. 또한 최종 통계처리에서 신뢰도가 떨어지는 평가자의 결과는 모두 배제하여 평가의 정확성을 확보하도록 한다.

평가자의 청각 기억에 한계가 있으므로 한 화면에 총 15개 이하의 신호를 청취하도록 제한한다. 이 때, 세 개는 원음, Hidden Reference, Anchor에 각각 해당하므로 실제 평가대상은 12개로 제한된다. 20명의 평가자가 적당하고, 각 신호의 길이는 20초로 제한된다.

Ⅳ. 객관적 오디오 품질 측정 기술

ITU-R에서 객관적 오디오 품질 측정을 위한 표준으로 BS.1387 "Method for objective measurement of perceived audio quality"를 제정하였다[4]. 보통 이 방법을 PEAQ (Perceptual Evaluation of Audio Quality)이라고 하고, BS.1387으로 측정한 품질 값을 PEQA 값이라도 한다.

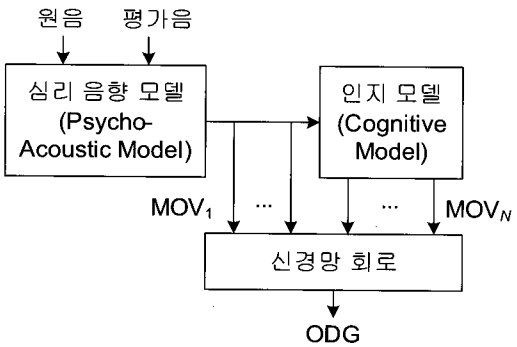
PEAQ의 기본 개념은, 원음과 평가음의 특성을 수학적으로 분석하여 두 신호를 청취할 때 청각이 인지하는 품질 차이를 수학적으로 계산하는 것이다. 물론, 단순히 수학적으로 두 신호 사이에 큰 차이가 있다고 그것이 곧바로 청각의 인지 차이로 나타나는 것은 아니므로 수학적 차이를 인지 차이로 변환시키는 것은 매우 복잡한 연산을 포함하는 어려운 과정이다.

PEAQ은 크게 두 가지 버전으로 구성된다. 기본 버전(Basic Version)은 수학 연산이 간단하지만 정확도가 떨어진 품질 측정 결과를 제공하

고, 고급 버전(Advanced Version)은 좀 더 복잡한 수학 연산을 사용하여 보다 정확한 품질 측정 결과를 제공한다. 기본 버전은 실시간 처리, 즉 오디오 재생 속도보다 빠른 품질 측정이 가능하도록 하여 오디오 서비스 중에 온라인으로 품질 평가를 할 수 있도록 한다.

PEAQ 과정을 간단히 정리하면 <그림 7>과 같다. 인간의 청각기관이 소리를 인지하는 과정은 기본적으로 주파수 영역에서 이루어지므로, 원음과 평가음을 주파수 영역으로 변환하고 심리음향모델에 따라 청각적 차이를 분석한다. 기본 버전은 푸리에 변환(Fourier transform)을 사용하고, 고급 버전은 푸리에 변환과 필터 뱅크를 동시에 사용하여 주파수 분석을 실시한다. 심리음향 모델의 결과로서 매스킹(masking)과 관련된 다수의 MOV (model output variable)을 출력한다.

인지 모델에서는 심리음향 모델 분석 결과를 바탕으로 실제 청각 기관의 동작을 상세히 모델링 하고, 수학적 차이를 청각 인지의 차이로 변환시켜 여러 개의 추가적인 MOV를 출력한다. 각 MOV는 청각 인지를 결정하는 특정 항목에서의 두 신호의 차이점을 나타낸다. 예로, MOV 중에



〈그림 7〉 PEAQ 흐름도

서 WinModDiff 변수는 두 신호의 포락선 차이의 평균을 보여주고, EHS는 두 신호 차이의 하모닉 구조를 보여준다. 각 MOV를 구하는 과정은 수학적으로 너무 복잡하여 본 논문의 수준에서 다루는 것이 불가능하므로 생략한다.

마지막으로, 인공 신경망(artificial neural networks)을 이용하여 MOV들을 결합하여 하나의 최종 점수 값을 구한다. 기본 버전은 11개의 MOV를 사용하고, 고급 버전은 5개의 MOV를 사용한다. 이렇게 구한 최종 점수는 ODG (objective difference grade)로 정의되고 아래의 의미를 가진다.

ODG=평가음 객관적 점수−원음 객관적 점수

즉, ODG는 식 (1)의 SDG를 수학적으로 구한 것에 해당하고, 보통 음수값을 가진다. 모든 오디오 신호에 대한 ODG의 평균을 구하여 해당 시스템의 품질에 해당하는 PEAQ 값으로 최종 출력한다.

청각을 수학적 모델로 정확하게 모델링 하는 것은 불가능하므로 PEAQ 값과 SDG 값 사이에는 오차가 발생한다. PEAQ의 성능, 즉 PEAQ를 통하여 해당 오디오 품질을 얼마나 정확하게 측정하였는지는 PEAQ와 SDG의 상관관계로 평가한다. PEAQ의 성능 평가 보고서에 의하면 PEAQ와 SDG는 약 0.84 정도의 상관관계 계수(correlation coefficient)를 가진다[4].

V. 결 론

디지털 오디오 통신 및 방송에서 오디오 품질은 서비스의 종합 품질을 결정하는 매우 중요한 항목이다. 오디오 품질 평가는 많은 기관에서 각

각의 목적에 따라 진행되는데, 평가 결과를 서로 비교하고 높은 수준의 신뢰도를 보장하기 위하여 표준화된 품질 측정 방법이 필요하다. 표준 방법에 따라 정확히 측정된 품질은 전 세계적으로 공인된 품질로 활용될 수 있는 장점을 가진다.

가장 널리 사용되는 표준 품질 측정 방법으로 ITU-R에서 제정한 주관적 평가를 위한 BS. 1116과 BS1534, 객관적 평가를 위한 BS.1387이 있다. BS.1116과 BS.1534는 평가자가 원음과 평가음을 직접 청취하여 두 신호 사이의 품질 차이를 점수로 출력한다. BS.1387은 두 신호를 심리음향 모델과 인지 모델을 기반으로 수학적으로 분석하여 청각 인지 차이를 다수의 MOV로 출력하고 이 값들을 인공 신경망에 입력하여 최종 품질 값을 구한다.

참고문헌

[1] Recommendation ITU-R BS.1116, "Method for the subjective assessment of small impairments in audio systems including multichannel sound systems," 1997.

[2] Recommendation ITU-R BS.1534, "Method for the subjective assessment of intermediate quality level of coding systems," 2003.

[3] Audio Research Labs, "Subjective Training and Evaluation Program (STEP)," White Paper, 2004.

[4] Recommendation ITU-R BS.1387 "Method for objective measurement of perceived audio quality," 2001.

저자소개

박 호 종

1986년 2월 서울대학교 전자공학과, 공학사
1987년 12월 Univ. of Wisconsin-Madison, 전기컴퓨터 공학과, M.S.
1993년 5월 Univ. of Wisconsin-Madison, 전기컴퓨터 공학과, Ph.D.
1993년 9월 ~ 1997년 08월 삼성전자 선임연구원
1997년 9월 ~ 현재 광운대학교 교수

주관심 분야 : 음성/오디오 신호처리, 멀티미디어 신호처리