

# 음질 평가법의 표준과 연구 동향 - 전송 처리음 분야

## Review of Standard Sound Quality Assessment Methods for the Transmitted and Processed Sounds

오원근

(Wongyun Oh)

순천대학교 멀티미디어공학과

(접수일자: 2012년 11월 27일; 수정일자: 2013년 2월 5일; 채택일자: 2013년 2월 14일)

**초 록:** 음질 평가는 좋은 소리를 만들기 위해 필수적인 요소이며, 음향의 특성과 대상 시스템에 따라 다양한 방법이 사용되고 있다. 본 논문에서는 음질 평가법의 전반적인 방법론 및 전송 처리된 음향 신호의 품질 평가법에 대해 ITU-T, ITU-R, IEC, 그리고 ANSI 등의 권고안에 기술된 국제 표준을 중심으로 요약하고 분석하였다. 분야별로는 음성 명료도, 음성 음질, 그리고 오디오 음질 평가법을 다루었으며, 현재 사용되는 권고안의 기술적인 내용과 최신 연구 동향 및 향후 발전 방향 등에 대해 기술하였다.

**핵심용어:** 국제표준, 음질 평가, 음성 명료도, 음성 음질, 오디오 음질

**ABSTRACT:** Assessing the quality of audio signals is an important consideration in making high quality sounds and various methods have been developed. This paper provides a general framework of sound quality and a technical overview of the international standard methods which are described in ITU-T, ITU-R, IEC and ANSI Recommendations in the speech intelligibility, speech quality, and audio quality areas. In addition, some recent findings and future works are included.

**Keywords:** International Standards, Sound quality assessment, Speech intelligibility, Speech quality, Audio quality

**PACS numbers:** 43.60. Dh, 43.71. Gv

### 1. 서 론

음향 기술의 궁극적인 목표는 좋은 음질을 구현하는 것이며, 이를 위해서 객관적이고 합리적인 음질 평가는 필수적이라 할 것이다. 이에 따라 다양한 음질 평가 방법이 실내 음향, 통신, 전기 음향 등의 분야에서 오랫동안 연구되어 왔으며, 일부는 국제 기구의 표준 권고안으로 권고되고 있다. 음질 평가법은 음원의 종류와 응용 분야에 따라 각각 다른 관점과 기법이 사용 된다. 본 논문에서는 음질 평가의 관점에서 음원을 현장음, 전송 처리음, 그리고 기기 작동음의 3 가지로 분류 하고 각각에 대한 표준 음질 평가

법에 대해 고찰하고자 한다.

현장음은 음원과 청취자가 가까운 거리에 있어 공간을 통해 전달되는 음을 듣는 것을 말하며, 콘서트, 연설회장, 회의, SR(Sound Reinforcement), PA(Public Address) 등의 음향 신호가 여기에 해당된다. 현장음의 음질에 대한 연구는 건축음향, 실내음향, 전기음향 등의 분야에서 주로 연구가 이루어지고 있다.<sup>[1-5]</sup>

전송 처리음은 통신 전송이나 매체에 저장된 소리를 말하며, 이에 대한 음질 평가는 20세기 들어서 통신 기술의 발달과 함께 시작되었다. 1920년대에 음성 명료도 주관 평가 방법, 1940년대에는 이용한 객관 평가법인 AI(Articulation Index)가 개발되었고, 이후에도 음질 측정 및 평가를 위한 다양한 방법이 연구되어 왔다.<sup>[4,9]</sup> 최근에는 디지털 오디오 신호처리

†Corresponding author: Wongyun Oh (owg@sncu.ac.kr)  
Department of Multimedia Engineering, Suncheon National University, 255 Jungang-ro, Suncheon City, Republic of Korea.  
(Tel: 82-61-750-3832, Fax: 82-61-750-3830)

기술의 발전으로 심리음향모델에 기반을 둔 지각부호화(perceptual coding) 방식의 압축을 이용한 오디오의 전송 및 저장 기술이 널리 사용되고 있다.<sup>[10,11]</sup> 이 분야의 음질 평가는 잘 통제된 주관 평가 또는 심리음향 기반의 객관평가 알고리즘이 주로 사용된다. 국제기구인 ITU(International Telecommunication Union)나 EBU(European Broadcasting Union) 등에서는 관련된 평가법을 권고안으로 만들어 음질 평가의 객관성과 통일성을 유지하기 위해 노력하고 있다.<sup>[12-28]</sup>

기기 작동음은 비교적 최근에 관심을 받기 시작한 분야로, 기기 동작 시에 발생하는 소음이나 복합음을 말한다. 이 분야의 음질은 소리 자체 보다는 기기의 사용감에 미치는 영향을 평가하게 되며, 가전, 자동차 등 다양한 분야에서 연구가 이루어지고 있다.<sup>[8, 29-31]</sup>

한편, 이와 같이 음질에 대한 많은 연구와 국제기구의 권고안에도 불구하고, 음질을 평가하는 경우에 음질의 정의와 평가 기준을 모호하게 사용하거나 주관적인 해석을 바탕으로 평가하는 경우가 종종 있는 것도 사실이다. 그 이유로는 평가자의 표준 음질 평가법에 대한 이해부족, 현재까지 개발된 음질 평가법과 인간 청감의 불일치로 인한 평가법에 대한 불신, 그리고 기존의 평가 방법이 특정 시스템에 부적합한 경우 등을 들 수 있을 것이다. 그러나 이러한 한계에도 불구하고 검증받은 표준 평가법을 이용해서 음질을 측정하는 것은 음질의 객관적인 상호 비교를 위해 매우 중요하다고 할 수 있다.

이러한 맥락에서 본 논문에서는 현재까지 연구된 다양한 음질 평가 기법에 대한 간략한 기술적 소개와 함께 향후 연구 동향에 대한 고찰을 하고자 한다. 본 편에서는 우선 전체적인 음질 평가의 프레임워크(framework)와 처리음에 대한 국제 표준 및 최신 연구 동향에 대해서 고찰하고, 후속 논문에서 현장음과 기기 작동음에 대해 다룰 예정이다.

처리음에 대한 표준 음질 평가 기법과 동향 등에 대한 요약 및 분석 리뷰는 다양한 연구자들에 의해 수행되었다. 주관 평가법에 대한 리뷰로는 Hoeg 등<sup>[32]</sup>의 EBU의 주관 평가법에 대한 요약 및 Bech와 Zacharov<sup>[33]</sup>의 주관 평가 이론, 방법론 그리고 응용에 재한 상세한 정리가 있다. 객관 평가법 리뷰에는 Rix 등,<sup>[34]</sup> Campbell 등<sup>[35]</sup>의 연구가 있으며, 종합적인 리

뷰에는 de Lima 등<sup>[36]</sup>의 주, 객관 평가법 및 음성, 오디오 음질에 대한 종합적인 리뷰와 Oh 등<sup>[37]</sup>의 멀티미디어 서비스 분야의 음질 평가 리뷰 연구가 있다. 본 논문에서는 이들의 연구보다 더 거시적인 측면에서 음질을 고찰하고, 보다 최신의 기술 동향에 대해 고찰하고자 한다.

논문의 구성은 2장에서는 전체적인 음질 평가 기법의 관점과 종류, 3장은 음성 명료도 평가 기법, 4장은 음성 음질 평가 기법, 5장은 오디오 음질 평가 기법, 그리고 6장은 결론의 순서로 이루어져 있다.

## II. 음질 평가 프레임워크

### 2.1 음질의 정의와 음질 평가 속성

음질(sound quality)이란 연구자에 따라 “청취자가 만족 또는 불만족으로 표현할 수 있는 음에 대한 청각적인 인상(auditory image)<sup>[37]</sup>”, “소리가 특정한 기술적인 목적이나 목표에 부합되는 정도”, 그리고 “원하는 소리의 특성(nature)과 비교했을 때 지각한 소리의 특성에 대한 평가” 등과 같이 정의하고 있다.<sup>[38]</sup> 이를 종합하면 음질이란 일반적으로 “소리에 대한 주관적, 기술적 만족 정도”라고 정의할 수 있다.

Berg 와 Rumsey<sup>[39]</sup>는 음질을 세분화 하여 음색 품질(timbral quality), 공간감 품질(spatial quality), 그리고 기술적 품질(technical quality)의 3가지로 분류하였다. 음색 품질은 음에 대한 청각적 인상에 대한 평가이며, 기술적 품질은 잡음, 왜곡 등과 같은 물리적 측정 지표에 대한 품질을 의미한다. 공간감 품질은 2채널 이상의 음원에 대해서 “음원과 음원이 있는 환경에 대한 3차원적인 속성” 또는 “음원의 위치에 대한 청각적 인식”등을 의미 한다.<sup>[39, 40]</sup> 공간감은 음질 평가 시 무시될 수 없는 요소이며, Rumsey와 Bech<sup>[41]</sup>는 음질 판단 시 공간감이 약 30%의 비중을 차지한다는 것을 밝혔다.

이러한 3분야 음질에 대한 구체적인 평가는 음질 속성(attribute)을 측정함으로써 이루어진다. 음질 속성은 기술적, 심리적인 구체적인 평가 또는 측정 항목을 말하며, 최종적인 음질은 각 속성을 선형 또는 비선형으로 조합하여 하나의 숫자로 표현하는 것이 일반적이다. Letowski<sup>[42]</sup>는 음질 평가 속성으로 MuRAL

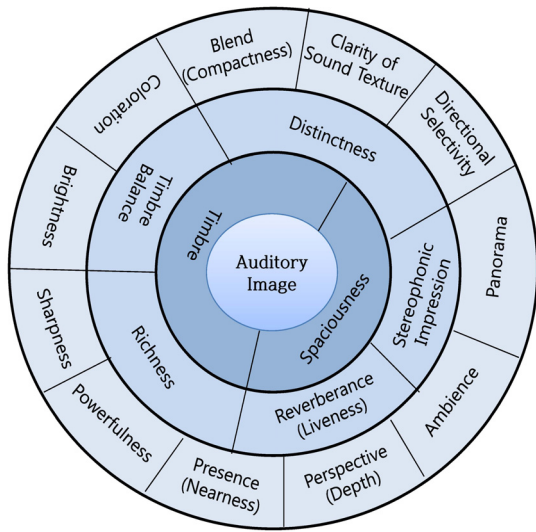


Fig. 1. MuRAL Hierarchical system.

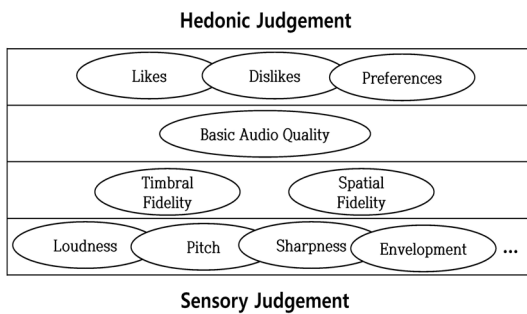


Fig. 2. Hypothetical hierarchy of audio quality attributes.

(Multidimensional auditory Assessment Language)이라는 계층적인 체계를 제안하였다(Fig. 1). 여기에서 동일한 동심원 상의 각 속성은 중요도가 동등한 서로 독립적인 요소로서, 중심에서 가까울수록 정서적인 항목, 멀어질수록 감각적인 항목을 의미한다.

또 다른 속성 구조로 Zielinski<sup>[42]</sup>는 음질 평가 시에 정량적으로 측정 가능한 속성과 측정 불가능한 감정적인 속성을 구분하여 Fig. 2와 같은 가상의 계층 구조의 속성 체계를 제안하였다.

이 계층 구조에서 하위 계층은 감각적 판단(sensory judgement)이 가능한 속성들이며, 이들은 가상의 지각 공간(perceptual space)에서 서로 독립적인 직교 변수로 구성되어야 한다.<sup>[43]</sup> 그림의 상위로 갈수록 감정적 또는 쾌락적 판단(hedonic judgement)과 관련된 요소이며 주관성이 점점 강하게 개입 된다. 최상위에 있는 감정 판단 속성(좋음, 싫음, 선호도 등)은 일

반적으로 인간이 음질에 대해 표현하는 최종적인 결과이며, 이들은 여러 하위 요소가 결합된 다차원적인 판단의 결과라고 볼 수 있다.

Zielinski<sup>[42]</sup>는 음질 평가 시 이러한 감정적 속성을 직접 평가하는 것은 주관성에 의한 편견(bias)이 생길 가능성이 높기 때문에 바람직하지 않음을 지적하였다. 그는 이러한 편견에 의한 영향을 줄이기 위해서는 감정적 판단을 직접 평가하는 것을 최대한 배제하고, 하위 계층에 있는 측정 가능한 인자를 조합하여 음질을 평가해야 한다고 제안하였다. 국제 표준인 ITU-R의 주관평가에서는 이렇게 하위 계층의 측정 가능한 속성을 조합하여 하나의 값으로 나타낸 것을 BAQ(basic audio quality)로 정의하고 있다.<sup>[20, 22]</sup>

## 2.2 음원의 종류와 평가 기준

인간이 청취하는 소리는 다양한 방법과 경로를 통해 생성되고 귀로 전달된다. 음질 평가는 이러한 음원의 종류와 전달 경로에 따라 서로 다른 관점과 평가 인자를 사용하게 된다. Blauert와 Jekosch<sup>[43]</sup>는 음질 평가를 제품 작동 음질(product-sound quality), 일반 음질(sound quality), 그리고 전송음 음질(sound-transmission quality)로 구분하였다. 그러나 이러한 분류는 음질 평가 시 평가의 기준이 혼용되는 단점이 있어, 본 논문에서는 음질 평가의 기준과 평가의 관점에 따라 다음과 같이 3 분야로 분류하였다.

- 전송 처리음(transmitted or processed sound)
- 현장 청취음(air-transmitted sound)
- 제품 작동음(product sound)

전송 처리음은 음질 평가 시 원음과의 동일성(authenticity)을 중요한 기준으로 삼는 경우를 말하며, 통신망 전송이나 부호화 된 소리가 여기에 해당된다. 음질 평가는 전송 경로에서 소리에 포함된 속성과 정보의 손상 정도를 평가하여, 평가 대상음이 원음과 동일한 경우 가장 음질이 좋다고 평가한다. 현장 청취음은 공간을 통해서 전달되는 소리를 듣는 경우이며, 실내 음향이나 전기 음향을 통한 재생이 여기에 해당된다. 이 분야 음질 평가의 특징은 실내의 적절한 잔향(reverberation) 등과 같이 소리 전달과

정에서 원음의 특성이 향상 될 수도 있다는 점이다. 따라서 현장 청취음은 원음의 변형을 심미적인 (aesthetic) 관점에서 평가하거나 정보 전달의 정도를 평가한다. 제품 작동음은 기계, 장비 등의 작동 시에 나오는 소리이다. 이 분야의 음질은 소리 자체의 물리적, 감성적 특성 보다는 해당 소리가 전체 제품에 미치는 감성 품질을 평가한다. 즉, 제품의 사용감과 부가가치를 높이는 소리를 좋은 음질을 가진 소리로 평가한다.

### 2.3 주관평가와 객관평가

음질 평가 방법은 크게 인간이 직접 음을 듣고 평가하는 주관 평가(subjective testing)와 측정장비를 이용하는 객관 평가(objective testing)로 구분된다. 주관 평가는 다수의 평가자가 직접 소리를 청취하고 해당 되는 음질 속성을 평가한 후 이를 통계적으로 처리하여 최종 음질 평가값인 MOS(Mean Opinion Score)를 구하는 방식이다. 이는 인간의 체감 음질을 직접적으로 반영할 수 있는 장점이 있으나, 청취 환경이나 평가자의 선호도 및 구성 등의 요인에 의해 결과가 왜곡되는 현상이 쉽게 발생할 수 있기 때문에 강인성(robustness)과 시불변성에 취약하다는 문제점이 있다. ITU, IEC 등의 주관평가와 관련된 권고안에서는 이러한 오차를 줄이기 위해 평가 절차, 청음 조건, 평가자 수준과 구성 등을 엄격하게 제시하고 있다.<sup>[12, 20, 22, 24, 33]</sup> 그러나 권고안에 규정한 대로 정확하게 주관 평가를 하기 위해서는 많은 시간과 비용이 소요되는 것이 단점이라 할 수 있다. 주관 평가의 종류에는 평가음만을 듣고 점수를 매기는 절대 평가 방식과 원음과 평가음을 교대로 듣고 비교해서 그 차이를 평가하는 상대 평가 방식이 있다.

객관 평가는 측정 기기를 이용하여 음질을 판단하는 방법이며, 그 종류에는 물리적 파라미터를 측정하는 방법과 심리음향모델을 이용한 방식이 있다. 물리적 파라미터 기반의 객관 평가는 임펄스 응답, 스펙트럼 특성 등과 같이 물리적인 데이터를 이용하는 방법으로써 건축음향이나 실내음향 분야에서 음질을 측정하는 경우에 주로 사용 된다.<sup>[4, 5, 44, 45]</sup>

심리음향모델 기반의 객관 평가는 보다 정교한 방식으로써, 심리음향에서 연구된 청각과 인지 모델

기반의 알고리즘을 사용하여 인간의 주관적인 평가 점수를 예측하도록 구성되어 있다. 종류에는 비교 방식(intrusive or double-ended measurement) 과 단독 방식(non-intrusive or single-ended measurement)이 있다. 비교 방식은 원음과 평가음을 비교하여 원음에 비해 평가음의 음질이 얼마나 저하되었는지를 수치로 나타낸다. 단독 방식은 평가음 하나만을 입력으로 사용하며, 내부적으로는 사전에 파라미터화시켜 저장해 놓은 이상적인 음질 기준과의 비교를 통해 평가값을 예측한다. 일반적으로 비교 방식이 주관평가와의 상관도(correlation)가 높지만, 실시간 음질 평가나 원거리 통신망에서의 음질 평가와 같이 원음을 입력으로 이용하기 곤란한 경우에는 단독 방식이 사용된다. 이러한 심리음향기반의 객관 평가법은 주관평가의 단점인 시간과 비용 문제를 줄일 수 있으며, 환경 요인에 의한 오차를 줄일 수 있다는 장점이 있어 최근 오디오 코덱, 디지털 통신, 멀티미디어 서비스 등에서 광범위 하게 사용되고 있다.<sup>[10, 11, 46]</sup>

### 2.4 음성과 오디오

음질 평가에서 음성(speech)과 일반 오디오는 다른 관점에서 평가된다. 음성에 대한 평가는 명료도 또는 이해도(intelligibility) 평가와 음질 평가로 나뉘며, 오디오는 음질만을 평가한다. 여기에서 명료도란 음성 정보의 전달 정도를 말하며, 음질의 좋고 나쁨과는 반드시 일치하지는 않는다. 즉, 음질은 좋지 않더라도 명료도가 높을 수 있으며, 반대의 경우도 가능하다.<sup>[4, 47]</sup> 예를 들어 기계 합성 음성의 경우 청감 음질은 좋지 않다고 평가되지만 명료도는 높으며, 반면 VoIP 망에서 일부 단어에서만 패킷 손실이 있는 경우에 평균적인 음질은 좋지만 명료도는 낮아질 수도 있다.<sup>[47]</sup> 음성 명료도와 관련된 국제 표준으로는 IEC 60268-16,<sup>[28]</sup> ANSI S3.2,<sup>[26]</sup> S3.5<sup>[27]</sup> 등이 있으며, 음성 음질과 관련된 표준은 ITU-T P.800,<sup>[12]</sup> P.862,<sup>[14]</sup> P.862.2,<sup>[16]</sup> 그리고 P.863<sup>[18]</sup>이 있다.

오디오의 평가에서는 음성과 달리 음질만을 평가하며, 정보 전달 정도를 별도로 평가하지는 않는다. 처리음의 경우 오디오 음질은 인간이 느끼는 절대 음질이 아니라, 통신망이나 코덱 등에 의한 원음의 음질 손상 정도를 측정하고 평가하는 것을 의미한다.

Table 1. The international standard assessment methods of sound quality.

Type	Assessment Methods		
Speech Intelligibility	Subjective		ANSI S3.2 (PB, MRT, DRT)
	Objective	Quantitative	IEC 60286 (STI) ANSI S3.5 (SII)
Speech Quality	Subjective	Absolute	ITU-T P.800 (ACR)
		Comparative	ITU-T P.800 (DCR, CCR)
	Objective	Intrusive	ITU-T P.862 (PESQ) ITU-T P.862.2 (W-PESQ) ITU-T P.863 (P.OLQA)
		Non-Intrusive	ITU-T P.563
Audio Quality	Subjective	Absolute	ITU-R BS.1284
		Comparative	ITU-R BS.1284 ITU-R BS.1116 ITU-R BS.1534 (MUSHRA)
	Objective	Intrusive	ITU-R BS.1387 (PEAQ)

다. 오디오 음질 평가와 관련된 표준은 ITU-R 권고안에 기술되어 있는 방법이 사용된다. 주관 평가법으로는 BS.1284,<sup>[22]</sup> BS.1116,<sup>[20]</sup> 그리고 BS.1534<sup>[24]</sup>가 있으며, 객관 평가법으로는 PEAQ로 알려진 BS.1387<sup>[25,48]</sup>이 있다. Table 1은 음성과 오디오의 주관, 객관 평가법 및 해당되는 국제 표준을 나타낸 것이다.

### III. 음성 명료도

음성 명료도란 화자(speaker)가 말한 음성에 담긴 정보가 듣는 사람에게 잘 전달되는가를 평가하는 척도이며, 일반적으로 듣는 사람이 정확하게 인지한 단어나 문장의 비율로 명료도를 나타내는 방법을 사용한다. 이러한 명료도 측정 방법은 1910년대 부터 시작되었으며, 이후 전화 및 세계대전 중의 통신 시스템의 확대로 점차 발전하였다. 현재는 유무선 통신, 네트워크, 그리고 오디오 코덱 등에서 음성 명료도의 평가를 위한 다양한 측정법이 개발되고 사용되고 있다.<sup>[49]</sup>

#### 3.1 명료도 주관 평가

명료도의 주관평가는 미리 정해진 어휘나 문장을 발음한 음성을 평가자가 직접 듣고 정확하게 인지한

비율을 통계적으로 처리하여 명료도를 나타내는 방법이다. 이 때 사용하는 평가 음성의 종류에 따라 음절 평가(nonsense syllable tests), 단어 평가(word tests), 문장 평가(sentence tests)로 구분 된다.

음절 평가는 자음-모음-자음(C-V-C)형태로 구성된 의미 없는 단음절을 사용하여 평가한다. 이는 각 단위 음절에 대한 명료도 분석에는 유용하나 평가 음절의 난이도를 조절하기 힘들고, 무의미한 단어를 사용하기 때문에 평가자에 대한 사전 훈련이 필요하다는 단점이 있다.<sup>[47]</sup>

단어 평가는 의미 있는 단음절 단어의 집합을 사용해서 명료도를 평가하는 방식이다. 평가에 사용되는 단어의 목록이나 구성 방법 등은 국가별로 달라 지는데, 미국 표준인 ANSI S3.2<sup>[26]</sup>는 영어에 적합한 명료도 평가를 위해 MRT(Modified Rhyme Test), DRT(Diagnostic Rhyme Test), 그리고 PB(Phonetically Balanced)와 같은 방법을 제시하고 있다.

MRT는 6개씩 묶어진 단음절 총 50세트를 사용하는 방법으로, 한 세트 내의 각 단어는 운율이 동일하며 첫째 또는 마지막 자음만 다르게 선정되어 있다. 평가자는 단어를 듣고 6개 중에서 청취했다고 여겨지는 단어를 선택하는 방식으로 평가를 진행한다. DRT는 초성이 다른 96개의 단음절어 쌍을 사용하는 데, 각 쌍은 발성 특성에 따라 6개 그룹으로 분류되어 있다. 평가자는 1쌍의 단어를 듣고 청취했다고 생각되는 단어를 둘 중에서 선택한다. PB는 2차 대전 중에 개발된 것으로 음성학적으로 균형을 이룬 1,000개의 단음절 어음을 사용하여 명료도를 평가한다. PB평가는 가장 널리 사용되는 방법이나, 다른 방법에 비해 청취자나 발화자에 대한 훈련이 더 많이 필요하고, 신호 대 잡음비(signal to noise ratio)의 변화에 민감하다는 단점이 있다.<sup>[50]</sup> Table 2는 MRT, DRT, PB 평가에 사용되는 단어 목록의 일부를 예시한 것이다.<sup>[50,51]</sup>

이러한 주관 평가에 있어 평가 환경과 절차는 결과에 영향을 미치기 때문에 그 조건을 엄격하게 통제하고 있다. 일반적으로 평가자와 발화자는 모두 해당 언어를 모국어로 구사하는 사람으로 구성되는 것이 원칙이며, 청력이나 발음에 이상이 없어야 한다. 또한 성별이나 연령이 편중되지 않도록 구성되

Table 2. Examples of MRT, DRT, and PB words list.

MRT Words			
went - sent - bent - dent - tent - rent hold - cold - told - fold - sold - gold pat - pad - pan - path - pack - pass ...			
DRT Words			
Voicing	Nasality	Sustentation	
veal - feel bean - peen gin - chin ...	meat - beat need - deed mitt - bit ...	vee - bee sheet - cheat vill - bill ...	
Sibilation	Graveness	Compactness	
zee - thee cheep - keep jilt - gilt ...	weed - reed peak - teak bid - did ...	yield - wield key - tea hit - fit ...	
PB Words			
List 1	List 2	List 3	List 4
are bad bar ...	awe bait bean ...	ache air bald ...	bath beast bee ...

고, 통계상의 오차를 줄이기 위해서는 적절한 인원을 대상으로 평가해야 한다. 또한 평가를 진행하는 사람은 사용하는 음향 장비의 사용에 익숙해야 하며, 전체적인 실험 과정의 계획과 최종 통계처리, 실험 중간의 피험자의 심리적인 요인에 대한 이해 등을 필수적으로 갖추어야 한다.<sup>[50]</sup>

또한 평가 단어는 해당 언어에서 사용하는 친숙한 단어로써, 일상 회화에서 사용하는 단어의 빈도와 같은 구성을 가져야 한다.<sup>[52,53]</sup> 따라서 나라별로 서로 다른 단어 목록을 평가에 사용하고 있는데, 한국어에 대해서는 1960년 신규식의 목록을 비롯하여 의료 및 건축음향 분야 등에서 많은 연구자에 의해 평가 단어 목록이 제안되어 왔다.<sup>[53-56]</sup> 그러나 오랫동안 표준화가 되지 않은 상태로 다양한 목록이 사용되었기 때문에 최근 회화의 단어 사용 빈도가 맞지 않는 문제<sup>[52]</sup>와 평가의 일관성과 신뢰성 문제<sup>[53]</sup>가 꾸준히 제기 되어 왔다.

최근 2009년에는 KSI ISO 8253-3 표준으로 한국어를 이용한 어음 청각검사(speech audiometry)의 표준 목록이 작성되었으며, 이는 국제 표준인 ISO 8253-3, Acoustics-Audiometric test methods-Part 3: Speech audiometry

Table 3. KS ISO 8253-3 Korean standard.

	Adult	Schooler	Preschooler
Bisyllabic word lists	36 Words (12×3 lists)	24 Words (12×2 lists)	12 Words (Pictures)
Monosyllabic word lists	200 Words (50×4 lists)	100 Words (25×4 lists)	100 Words (25×4 lists) (Pictures)
Sentence lists	80 Sentences (10×8 lists)	80 Sentences (10×8 lists)	40 Sentences (10×4 lists)

의 내용에서 한국어 검사 자료를 개발하여 한국산업 표준(KS)으로 개정한 것이다.<sup>[57]</sup> 이 표준은 이어폰을 통한 공기 전도(air conduction), 골 진동자(bone vibrator)를 통한 골 전도(bond conduction) 또는 스피커에서 제공하는 녹음한 어음을 사용하는 어음 청각 검사의 절차와 요구 조건을 기술한 것이나, 일정한 검토과정을 거쳐 통신, 오디오 코덱 등의 평가에도 사용될 수 있을 것으로 생각된다.

KS ISO 8253-3의 한국어음의 표준은 이음절어표(bisyllabic word list, BWL), 단음절어표(monosyllabic word list, MWL), 그리고 문장표(sentence list, SL)로 구성되어 있으며, 각 표는 연령별로 일반(만 13세 이상), 학령기 아동(만 6-12세), 그리고 학령전기 아동(만 3-5세) 용으로 구분되어 있다. Table 3은 표준어음표의 구성을 나타낸 것이다.

## IV. 음성 음질 평가

### 4.1 주관 평가

음성 음질에 대한 주관 평가는 ITU-TP.800<sup>[12]</sup> 권고안이 주로 사용된다. 여기에서 음성에 대한 청취 음질의 평가 방법에 대해서는 ACR(Absolute Category Rating), DCR(Degradation Category Rating), 그리고 CCR(Comparison Category Rating)이 기술되어 있다.

ACR은 절대 평가 방식이며, 일상적으로 사용되는 문구로 구성된 테스트 문장을 듣고 주관 평가 점수인 MOS-LQS를 1점(bad)~5점(excellent)으로 평가한다. DCR은 평가의 변별력을 높이기 위해 상대 비교를 하는 방법으로써, 평가자는 원 문장과 평가 문장을 순서대로 연속해서 듣고 평가 문장의 음질이 나빠지는 정도를 1점(very annoying)~5점(inaudible)으로

평가한다. CCR은 DCR과 유사하나 원 문장과 평가 문장의 순서를 무작위로 제시한다는 점이 다르다. 평가자는 두 번째 음성의 음질을 +3(much better)~-3(much worse)의 7점 척도로 평가한다. Table 4에 각 방법에서 사용되는 스케일을 나타내었다.<sup>[12,33]</sup>

한편, P.800 권고안에는 주관 평가시의 환경에 대해서도 엄격하게 규정하고 있다. 예를 들어 평가 문장의 녹음과 청취를 위한 공간의 환경은 방의 체적 30~120 m<sup>3</sup>, 잔향시간 500 ms 이하(최적 200~300ms), 실내 소음은 스펙트럼에 피크가 없는 30 dBA이하인 곳에서 시행되어야 하며, 평가자에 대해서는 다음 3가지 조건을 만족해야 한다.<sup>[12]</sup>

- 전화망 또는 관련된 일을 하지 않아야 한다.
- 최근 6개월간 어떠한 주관 평가도 참여하지 않았어야 하며, 특히 청취 주관 평가에는 최근 1년간 참여하지 않았어야 한다.
- 평가 문장을 사전에 들은 적이 없어야 한다.

이외에도 평가 조건이나 환경에서 생길 수 있는 바이어스를 제거하기 위한 다양한 조건들이 제시되어 있으며, 이를 지키지 않고 평가하는 경우 평가값의 정확성을 보장하기 힘들다고 할 수 있다.

Table 4. ACR, DCR and CCR scale.

(a) ACR scale	(b) DCR scale	(c) CCR scale
5: excellent	5: inaudible	3: much better
4: good	4: audible but not annoying	2: better
3: fair	3: slightly annoying	1: slightly better
2: poor	2: annoying	0: about the same
1: bad	1: very annoying	-1: slight worse
		-2: worse
		-3: much worse

Table 5. Applicability of ITU-T standards.

Recommendations	Bandwidth	Applications
P.862	Narrow Band(300-3400Hz)	Telephone
P.862.2	Wide Band(50~7000Hz)	Headsets
P.863	Narrow, Wide, Superwide Band (50~14000Hz)	Digital Telephony

### 4.2 비교 방식의 객관 평가

ITU-T의 음성 음질에 대한 비교 방식의 객관 평가는 권고안 P.862,<sup>[14]</sup> P.862.2,<sup>[16]</sup> 그리고 P.863<sup>[18]</sup>에 기술되어 있으며, 평가하고자 하는 대상 시스템의 주파수 대역폭에 따라 Table 5와 같이 선택적으로 사용된다.

P.862는 PESQ(Perceptual Evaluation of Speech Quality)라 부르며, 주로 협대역(narrow-band) 전화망에서 음성의 음질을 평가하기 위한 방법이다. 평가는 기준 음성과 평가 대상 음성을 입력 받아 전처리 과정을 거친 후, 인간의 청각 모델을 통과 시키고 두 신호의 차이를 인지 모델로 입력하여 -0.5~4.5 범위의 평가값을 출력한다. 개발 당시 표준화 과정에서 실시된 주관평가와의 비교 테스트 결과 0.935의 높은 상관도를 보였고, Hu 등<sup>[6]</sup>의 연구에서 음질 향상 알고리즘 평가 시에도 7개의 객관 평가법 중 가장 높은 상관도를 나타내었다.

P.862.2는 W-PESQ(Wideband PESQ)로 불리며, 헤드폰 등을 사용하는 경우를 위해 PESQ를 광대역(wide-band)으로 확장한 것이다. 전체적인 계산 과정은 PESQ와 동일하나, 전처리 과정에서 광대역 필터를 사용하여 7000 Hz까지 음성 입력을 받는 점이 다르다. W-PESQ는 언어마다 약간씩 다른 특성을 보이기도 하는데, G.722 코덱에 대한 평가에서 한국어와 일본어는 주관 평가값인 MOS-LQ에 비해 높은 MOS값을 출력하는 경향이 있다는 것이 알려져 있다.<sup>[14]</sup>

P.863은 최근 발전된 디지털 전화망에 적합한 음질 평가를 위해 가장 최근(2011년 1월)에 발표된 권고안이다.<sup>[18]</sup> 개발 단계에서는 P.OLQA(Perceptual Objective Listening Quality Assessment)로 알려져 있으며, 협대역에서 초광대역(super wide-band) 까지 통신선로에서 음성에 대한 음질을 평가할 수 있도록 고안된 방법이다. P.863은 두 가지 모드가 있는데 초광대역 모드는 헤드폰으로 양귀(diotic) 청취를 하는 경

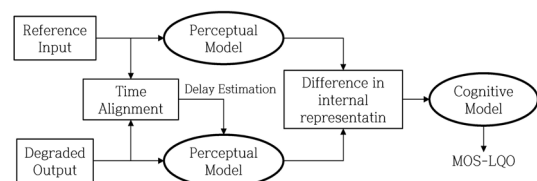


Fig. 3. The PESQ algorithm.

우에 사용하며, 협대역 모드는 전화기 등으로 모노 청취를 하는 경우에 사용한다. 발표 당시에 11개 언어에 대해 주관평가와의 비교 평가가 실시 되었으나,<sup>[18]</sup> 한국어는 여기에 포함되지 않았기 때문에 향후 P.863을 한국어의 객관 평가에 적용하기 위해서는 주관평가와의 상관관계 등에 대한 추가적인 연구가 필요할 것으로 보인다.

#### 4.3 단독(non-intrusive) 방식의 객관 평가

단독 방식은 원음과의 비교 없이 평가음만을 단독으로 사용해서 음질 값을 예측하는 평가 방식이다. 이는 비교 방식에 비해서, 주관 평가 MOS값과의 상관도가 떨어지는 단점이 있다. 그러나 원음과의 비교가 용이하지 않은 실시간 모니터링이나 장거리 통신과 같은 응용 분야에는 유용하게 사용할 수 있는 방법이다.

평가값을 계산하는 알고리즘은 미리 계산된 이상적인 신호음과 평가음과의 차이를 계산하거나,<sup>[58, 59]</sup> 또는 통신로에서 발생하는 음질 저하(degradation) 패턴이 인간의 목에서는 생성이 불가능하다는 점을 이용하여 통신로에서 추가된 신호의 영향을 예측하여 계산하는 방법<sup>[60]</sup> 등이 사용된다.

2004년에 ITU-T에서는 협대역 전화망의 음성 음질 평가를 위한 단독 방식의 평가법으로 당시에 제안되었던 ANIQUE(Auditory model for Nonintrusive Quality Estimation)<sup>[61]</sup>와 SEAM(Single-Ended Assessment Model)<sup>[60]</sup> 중에서 SEAM을 ITU-T P.563 권고안으로 채택하였다. 당시 P.563은 ITU의 벤치마크 결과 주관 평가와의 상관도가 평균 0.88로 나타났다.<sup>[19]</sup> (PESQ의 상관도 평균은 0.93) 또한 de Lima 등<sup>[36]</sup>의 연구에서는 PESQ는 0.91과 0.87 그리고 P.563은 0.79와 0.77의 상관도를 보였다. 최근 Yen 등<sup>[62]</sup>은 인간 청감의 관점에서 명료도, 명확함(clarity), 그리고 자연스러움(naturalness)의 3가지 파라미터를 추출하고 이를 이용하여 음질을 예측하는 방법을 제안하였다. 이를 ITU-T Supp.23<sup>[63]</sup> 데이터베이스를 이용하여 P.563과 비교한 결과 P.563은 평균 0.770, 제안된 방법은 0.788로 나타나 약간의 개선이 가능함을 보였다.<sup>[62]</sup>

이와 같이 단독 방식은 비교(intrusive) 방식에 비해 주관 평가와의 상관도가 낮은 편이다. 이는 인간이

청감적으로 파악하는 이상적인 음질에 대한 요소가 적절하게 파라미터로 만들어지지 못하고 있음을 의미하여, 추가적으로 음질 저하 요인이나 주관적인 음질을 파라미터화 하는 연구를 통해 주관 평가와의 상관도를 높이는 연구가 지속되어야 할 것으로 보인다.

## V. 오디오 음질

### 5.1 주관 평가

ITU-R 권고안에 기술되어 있는 오디오 음질의 주관 평가법은 BS.1283, BS.1116, BS.1284, BS.1285, 그리고 BS.1534<sup>[20-24]</sup> 등이 있는데, 각 방법은 음질의 손상 정도에 따라 선택적으로 사용된다.

음질의 손상 정도가 미세하여 섬세한 평가가 필요한 경우에는 BS.1116이 사용되며, 이는 오디오의 평가 시 음질의 차이를 민감하게 인지 가능하도록 평가단 구성, 평가 절차, 청음 조건 등의 평가 환경을 가장 엄격하게 규정하고 있는 권고안이다. 평가법은 기준음과 평가음, 그리고 기준음과 동일한 평가음 3개를 제시하는 double-blind triple-stimulus with hidden reference 방법을 사용한다. 이 방법에서 평가자는 (A, B, C)와 같이 세 개의 평가음을 청취한다. 이 때 A는 기준음이며 B, C는 평가 대상음인데, 둘 중 하나는 A와 동일하며 나머지 하나가 실제 평가할 음이다. 평가자는 A와 비교했을 때 B와 C의 음질 열화 정도를 “감지 불가(5점)~매우 거슬림(1점)”과 같이 5점 척도로 평가한다. 이러한 평가 절차는 미세한 손상이 있는 경우에 특히 변별력이 뛰어난 것으로 알려져 있다.

한편 음질 손상 정도가 어느 정도 있는 경우에는 BS.1116에서 규정한 만큼의 엄격함이 필요하지 않은 경우가 많다. 음질 손상 정도가 보통인 경우의 평가법은 BS.1284와 BS.1534에 기술되어 있다. BS.1534는 MUSHRA(Multi Stimulus test with Hidden Reference and Anchor)라고 불리는 방법으로 디지털 라디오, 모바일 멀티미디어 서비스 등과 같이 어느 정도 손상이 예측되는 중간 정도의 품질을 가진 오디오 시스템에 대한 주관 평가법이다. 평가 방법은 최대 15개를 하나의 그룹으로 묶어서 비교 평가하게 되는데, 하나의 그룹에는 1개의 기준음, 1개의 숨겨진 앵커



(anchor), 그리고 1개의 숨겨진 기준음이 포함된다. MUSHRA는 여러 평가음을 동시에 비교하기 때문에 다양한 상황에 대한 상대적인 평가를 동시에 할 수 있다는 장점이 있다. 점수는 0~100점 사이로 Excellent (80~100), Good(60~80), Fair(40~60), Poor(20~40), Bad (20이하)의 기준으로 평가한다. 구체적인 평가 속성은 BS.1284의 부록에 기술되어 있으며 BAQ(Basic Audio Quality)의 평가, 그리고 스테레오와 다채널의 경우는 음상 정보와 공간 정보에 대한 평가가 포함된다.

한편, 다채널 오디오의 음질 평가 시에는 공간적 품질도 중요한 평가 요소로 작용한다. 이러한 공간적 품질에는 음상 정보, 정위감과 같은 3차원 공간에서의 음원의 위치 정보 및 확산감 등과 같이 실내 환경에 대한 청각적 인상을 평가한다.<sup>[39, 40, 64-66]</sup>

## 5.2 객관 평가

현재 처리에 있어서 오디오 음질의 객관 평가법은 ITU-R BS.1387 권고안에 기술된 PEAQ(Perceptual Evaluation of Audio Quality)가 대표적으로 사용되고 있다.<sup>[25, 48, 67-70]</sup>

### 가. PEAQ

PEAQ는 인간의 심리음향모델(psychoacoustic model)과 두뇌의 인지 모델(cognitive model)을 이용하여, 인간이 실제 청감으로 느끼는 기준음과 평가음의 차이를 예측하는 알고리즘이다. 간략하게 나타낸 전체 계산 과정은 Fig. 4와 같다.

계산 과정은 다음과 같다. 먼저 기준음과 평가음은 청각을 모델링한 심리음향모델의 입력으로 들어가 심리음향적 처리과정을 거친다. 이때 청각 모델은 FFT기반과 필터뱅크(filterbank) 기반의 두 가지가 있다. FFT만을 사용한 것을 basic버전이라 하며, 이

는 실시간 응용과 같이 빠른 연산이 필요한 경우 사용한다. advanced 버전은 FFT와 필터뱅크를 이용한 모델로써 basic version에 비해 약 4배의 연산이 필요하지만 높은 시간 해상도를 얻을 수 있어 정밀한 계산이 필요한 경우에 사용한다. 심리음향 모델을 거친 후에는 기준음과 평가음의 음질의 차이를 나타내는 파라미터인 MOV(Model Output Variable)가 계산된다. MOV는 basic버전의 경우 11개, advanced버전의 경우 5개가 사용되며, 최종 음질 평가값을 계산하기 위한 인지 모델의 입력으로 사용된다. 인지 모델은 인공신경회로망을 사용하며, 최종적인 음질 저하 정도의 예측값인 -4에서 0 범위의 ODG(Objective Difference Grade)가 구해지게 된다. 여기에서 0은 음질 저하를 감지 불가(imperceptible)를 의미하며 -4는 음질 저하가 매우 거슬림(very annoying)을 의미한다.

이러한 PEAQ 알고리즘은 표준화 과정에서 CRC'97 음원 데이터베이스로 테스트한 결과 주관평가와의 상관도가 basic버전의 경우 0.837, advanced버전의 경우 0.851의 값을 보였으며,<sup>[25]</sup> 이는 주관 평가를 어느 정도 대신할 수 있을 정도의 비교적 높은 상관도라 할 수 있다.

PEAQ의 원래 목적은 오디오 코덱의 개발 및 평가, 미디어 서비스 평가, 전송 네트워크 계획 등의 평가에 사용하기 위해 개발되었다. 그러나 이외에도 PEAQ는 합성 음성 및 오디오, 잡음 감소 알고리즘, BSS(Blind Source Separation)의 평가에도 사용되었으며,<sup>[35]</sup> 음향기기의 음질 유사도 평가,<sup>[70, 71]</sup> MPEG-21 프레임워크에서 음에 대한 주관적 인상 측정,<sup>[72]</sup> 그리고 IP기반의 오디오 음질 평가<sup>[73]</sup> 등에도 성공적으로 이용되었다.

이와 같이 PEAQ는 현재 가장 광범위하게 이용되는 객관적 음질 평가법이지만 다음과 같은 몇 가지 한계도 가지고 있다.<sup>[35, 67]</sup> 첫째, MOV로 파라미터화되지 않은 요인에 대해서는 평가가 불가능하다. 둘째, 고품질 오디오에서 미세한 손상 평가를 위해 개발되었기 때문에 중간 정도 이상의 음질 손실에 대해서는 적합하지 않다. 셋째, 2채널 스테레오까지의 평가만 가능하다는 점 등이다. 따라서 2001년 PEAQ가 ITU BS.1387 권고안으로 만들어진 이후 이를 개선하거나 또는 새로운 접근 방식을 사용한 객관 평

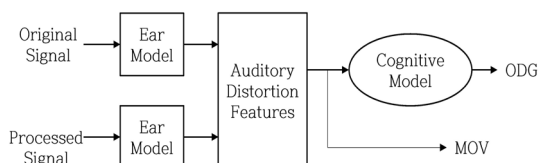


Fig. 4. The PEAQ algorithm.

가 알고리즘을 만들기 위한 연구가 꾸준히 지속되고 있다.

이러한 연구들로는 기존의 MOV에서 고려되지 않았던 비선형 왜곡이 있는 경우 음질 평가에 대한 연구는 Moore 등,<sup>[74,75]</sup> Tan 등,<sup>[76]</sup> 그리고 Huber 등<sup>[77]</sup>이 있고, 시변 왜곡(time-varying distortion)에 대해서는 Hardin과 Creusere의 연구가 있다.<sup>[78,79]</sup> 음질 손상이 중간 이상인 경우로 PEAQ를 확장하여 낮은 비트율 코덱 등의 평가에도 적용 가능하도록 한 것에는 Huber 등,<sup>[77]</sup> Vanam 등,<sup>[80,81]</sup> 그리고 Creusere 등<sup>[82]</sup>의 연구가 있다. 또한 Barbedo 등<sup>[83]</sup>은 새로운 인지 모델을 사용하여 PEAQ의 성능 향상을 꾀하였으며, Fernando 등<sup>[84]</sup>은 PEAQ에 퍼지(fuzzy) 논리를 사용한 방식을 제안하였다.

#### 나. 다채널 오디오의 객관 평가

PEAQ는 모노 및 2채널 스테레오까지의 객관 평가를 위해 고안되었다. 그러나 최근에는 5.1채널 이상의 다채널 오디오도 일반화 되었으며, 이에 대한 객관적 음질 평가 기법이 활발히 연구되고 있다.<sup>[85-88]</sup> Choi 등<sup>[86]</sup>은 다채널 오디오 코딩의 음질 평가에 사용할 객관 평가법을 제안하였다. 이 연구에서는 다채널 신호를 HRTF(head related transfer function)을 사용하여 2채널 신호로 변환한 다음, 음질 평가 요소로서 기존의 PEAQ의 basic version에서 사용된 MOV중 10개와 공간감을 평가하기 위해 양귀간 시간차이, 양귀간 레벨차이, 그리고 IACC(interaural cross-correlation coefficient)의 3가지 파라미터를 추가하여 평가하는 방법을 제안하였다. 또 다른 방법으로는 QESTRAL (Quality Evaluation of Spatial Transmission and Reproduction using an Artificial Listener)이 있다.<sup>[87, 88]</sup> 이는 다채널 오디오의 재생과 전송 시의 공간감 평가 목적으로 개발된 것으로, 14개의 공간감과 관련된 파라미터를 사용하여 공간감 품질을 예측하는 방법이다.

#### 다. PEAQ이외의 방식

PEAQ 이외에 다른 접근법을 사용한 객관 평가법에 대한 연구도 활발히 진행되고 있다. Manders 등<sup>[89]</sup>은 보청기에 사용되는 귀환 제거기(feedback canceller)의 음악 음질 평가를 위한 객관 평가법을 제안하였다.

이들은 기존의 평가법에서 사용된 음질 파라미터 11개를 사용하여, 귀환 제거기의 음질 평가에 가장 효과적인 metric의 조합을 찾는 방법을 연구하였다. 이와 같은 접근 방법은 기존의 평가법이 적합하지 않은 새로운 응용 분야에서 객관 평가 알고리즘을 개발하는 경우에 효과적으로 적용 될 수 있을 것이다. 또한 Kandadai 등<sup>[90]</sup>은 영상압축의 품질 평가를 위해 개발된 Mean Structural Similarity(MSSIM) measure를 오디오 품질 평가에 적용하는 연구를 수행하였으며, Yue 등<sup>[91]</sup>은 streaming audio에 대한 객관 평가법을 제시하였다. 그러나 이러한 연구에도 불구하고 PEAQ를 대체할 새로운 객관평가법은 아직 개발되지 않고 있으며, ITU에서는 PEAQ의 개선을 위한 새로운 작업계획을 작성하는 단계에 있다.<sup>[92]</sup>

#### 라. 절대평가 방식

오디오의 객관적 음질 평가에서 절대 평가 방식에 대한 연구는 많지 않으며, 이에 대한 표준안도 현재는 만들어지지 않은 상태이다. 그 이유로 일반 오디오의 음질 판단은 음성보다는 훨씬 다양한 요인을 가지고 있기 때문에 이를 적절하게 반영하는 이상적인 음질 기준을 파라미터 만드는 것이 쉽지 않기 때문이다. 오디오의 절대 객관 평가법은 통신이나 미디어 서비스 분야 외에도 PA시스템, SR시스템 분야나 실시간 방송 등에도 응용이 가능하기 때문에 이와 관련된 연구가 필요한 분야라 할 수 있다.

## VI. 결 론

본 논문에서는 현재 사용되는 음질 평가 기법의 현황과 연구 동향에 대해서 전반적으로 고찰하였다. 음원은 음질 평가의 기준에 따라 전송 처리음, 현장 청취음, 그리고 제품 작동음으로 구분된다. 본 논문에서는 그 중 전송 처리음에 대해서 음성 명료도, 음성 음질, 그리고 오디오의 음질 평가 방법에 대해 논하였다. ITU 등의 표준 권고안에 수록된 방법의 특징과 현재의 연구 동향 그리고 향후 연구 방향에 대해 기술하였으며, 후속 연구로는 본 논문에서 다루지 않은 현장 청취음과 기기 작동음 음질 평가의 표준 방법과 연구 동향에 대해 고찰할 예정이다.

## 감사의 글

본 논문은 2011년도 순천대학교 대학자체 연구비의 지원을 받아 수행한 결과이다.

## 참고 문헌

1. J. H. Ku, *Rayleigh's Acoustical Research* (Korea Studies Information, Paju, 2008).
2. S. H. Kang, *Spatial Acoustics* (Sound Media, Goyang, 2012).
3. L. Beranek, *Concert Halls and Opera Houses*, 2nd ed., (Springer-Verlag, New York, 2010).
4. G. Ballou, *Handbook for Sound Engineers*, 4th ed., (Elsevier, Oxford, 2008).
5. F. A. Everest and K. C. Pohlmann, *Master Handbook of Acoustics*, 5<sup>th</sup> ed. (McGraw-Hill, New York, 2009).
6. Y. Hu, and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. on Audio, Speech, and Lang. Proc.* **16**, 229-238 (2008).
7. S. Willsallen, and D. Cabrera, "Assessment of music audio quality in a sports stadium," AES 117<sup>th</sup> Convention, paper no. 6273 (2004).
8. H. Fastl and E. Zwicker, *Psychoacoustics-Facts and Models*, 3<sup>rd</sup> ed. (Springer, Berlin, 2007).
9. H. J. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.* **67**, 318-326 (1980).
10. T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE* **88**, 451-515 (2000).
11. A. Spanias, T. Painter, and V. Atti, *Audio Signal Processing and Coding* (Wiley, Hoboken, 2007).
12. ITU-T P.800, *Methods for Subjective Determination of Transmission Quality*, 1996.
13. ITU-T P.800.1, *Mean Opinion Score(MOS) Terminology*, 2006.
14. ITU-T P.862, *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*, 2001.
15. ITU-T P.862.1, *Mapping Function for Transforming P.862 raw Result Scores to MOS-LQO*, 2003.
16. ITU-T P.862.2, *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs*, 2007.
17. ITU-T P.862.3, *Application Guide for Objective Quality Measurement Based on Recommendations p. 862, p. 862.1 and p. 862.2*, 2007.
18. ITU-T P. 863, *Perceptual Objective Listening Quality Assessment*, 2011.
19. ITU-T p. 563, *Single ended Method for Objective Speech Quality Assessment in Narrow-band Telephony Applications*, 2004.
20. ITU-R BS.1116-1, *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*, 1994.
21. ITU-R BS.1283-1, *A Guide to ITU-R Recommendations for Subjective Assessment of Sound Quality*, 1997.
22. ITU-R BS.1284-1, *General Methods for the Subjective Assessment of Sound Quality*, 1997.
23. ITU-R BS.1285, *Pre-selection methods for the subjective assessment of small impairments in audio systems*, 1997.
24. ITU-R BS.1534-1, *Method for the Subjective Assessment of intermediate Quality Level of Coding Systems*, 2001.
25. ITU-R BS.1387-1, *Method for Objective Measurements of Perceived Audio Quality*, 1998.
26. ANSI S3.2-1989 (R1999), *Method for Measuring the Intelligibility of Speech over Communication Systems*, 1999.
27. ANSI S3.5-1997 (R2012), *American National Standard Methods for Calculation of the Speech Intelligibility Index*, 2012.
28. IEC 60268-16, *Sound System Equipment-Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index*, 2011.
29. M. Bodden, "Perceptual sound quality evaluation," in *Proc. InterNoise2000*, 1-6 (2000).
30. H. Fastl, *Psychoacoustics and sound quality, in communication acoustics*, edited by J. Blauert (Springer-Verlag, Berlin, 2005).
31. J. Kunio, "Using sound quality to improve your product," in *Intern. Appl. Tech. Conf. & Ex.*, 1-14 (2006).
32. W. Hoeg, L. Christensen, and R. Walker, "Subjective assessment of audio quality-the means and methods with in the EBU," *EBU Tech. Rev.*, 40-50 (1997).
33. S. Bech and N. Zacharov, *Perceptual Audio Evaluation-Theory, Method and Application* (Wiley, Atrium, 2006).
34. A. Rix, J. Beerends, D.-S. Kim, P. Kroon, and O. Ghitza, "Objective assessment of speech and audio quality technology and applications," *IEEE Trans. on Audio Speech and Lang. Proc.* **14**, 1890-1901 (2006).
35. D. Campbell, E. Jones, and M. Glavin, "Audio quality assessment techniques-a review, and recent developments," *Sig. Proc.* **89**, 1489-1500 (2009).
36. A. A. De Lima, F. P. Freeland, R. A. De Jesus, B. C. Bispo, L. W. P. Biscainho, S. L. Netto, a. a. de Lima, R. a. de Jesus, a. Said, a. Kalker, R. Schafer, B. Lee, and M. Jam, "On the quality assessment of sound signals," 2008 *IEEE Inter. Sym. Cir. and Sys.* **3**, 416-419 (2008).
37. W. Oh and S.-K. Lee, "Quality assessment of sound signals in multimedia and communication systems," *Comm. in Comp. and Inform. Science* **353**, 57-64 (2012).
38. S. Zielinski, F. Rumsey, and S. Bech, "On some biases encountered in modern audio quality listening tests-A review," *J. Audio Eng. Soc.* **56**, 427-451 (2008).

39. J. A. N. Berg and F. Rumsey, "Systematic evaluation of perceived spatial quality," in AES 24<sup>th</sup> Intern. Conf. on Multichannel Audio, paper no. 43 (2003).
40. F. Rumsey, "Spatial quality evaluation for reproduced sound: terminology, meaning, and a scene-based paradigm," J. Acoust. Soc. Am. **50**, 651-666 (2002).
41. F. Rumsey, and S. Bech, "On the relative importance of spatial and timbral fidelities in judgements of degraded multichannel audio quality," J. Acoust. Soc. Am. **118**, 968-976 (2005).
42. S. Zielinski, "On some biases encountered in modern listening tests," in Spatial Audio & Sensory Eval. Tech. (2006).
43. J. Blauert and U. Jekosch, "Concepts behind sound quality: some basic considerations," in InterNoise2003 (2003).
44. ISO 3382-1, *Acoustics-Measurement of Room Acoustic parameters-Part 1: Performance Spaces*, 2009.
45. ISO 3382-2, *Acoustics-Measurement of Room Acoustic parameters-Part 2: Reverberation Time in Ordinary Rooms*, 2008.
46. Y. Huang and J. Benesty(Eds.), *Audio Signal Processing for Next-Generation Multimedia Communication Systems* (Kluwer Academic Publishers, Norwell, 2004).
47. P. C. Loizou, *Speech Enhancement* (CRC Press, Boca Raton, 2007).
48. P. Kabal, *An examination and interpretation of ITU-R BS.1387: perceptual evaluation of audio quality*, (McGill Univ., Rep., 2003).
49. SFPE editor, "Speech intelligibility," Fire Protection Engineering, 16-18 (2002).
50. *Speech Intelligibility Papers*, <http://www.meyersound.com/support/papers/speech/index.htm>, 2013
51. N. A. Geoffroy, *Measuring Speech Intelligibility in Voice Alarm Communication Systems* (MS thesis, Worcester Polytechnic Institute, 2005).
52. S.-W. Byun, "Frequencies of Korean phonemes and reliability of Korean phonetically balanced word lists,"(in Korean), Kr. J. Otol. **44**, 485-489 (2001).
53. S.-W. Byun, S. M. Chung, H. S. Kim, and Y. M. Go, "A Survey of phonetically balanced words lists used in training hospitals in Korea," (in Korean), Kr. J. Otol. **48**, 1086-1090 (2005).
54. T. Y. Hahm, "Complementary study on construction of Korean word lists for speech audiometry," (in Korean), Inje Med. J. **7**, 1-19 (1986).
55. C. S. Yoon, S. W. Kim, and Y. K. Oh, "A study on the standardization of articulation testing method and its evaluation suitable for Korean language (I)," (in Korean), J.Arch. Instit. Kr. **4**, 117-125 (1988).
56. C. S. Yoon, S. W. Kim, and Y. K. Oh, "A study on the standardization of articulation testing method and its evaluation suitable for Korean language(I)," (in Korean), J.Arch. Instit. Kr. **5**, 95-108 (1989).
57. KS I ISO 8253-3:2009, *Acoustics-Audiometric test methods-Part 3: Speech Audiometry*, (in Korean), 2009.
58. H. Hermansky, "Perceptual linear prediction(PLP) analysis of speech," J. Acoust. Soc. Am. **87**, 1738-1752 (1990).
59. W. Li and R. F. Kubichek, "Output-based objective speech quality measurement using continuous hidden Markov models," in Proc. 7<sup>th</sup> Intern. Sym. Sig. Proc. and Its Appl., 389-392 (2003).
60. P. Gray, M. Hollier, and R. Massara, "Non-intrusive Speech Quality Assessment Using Vocal Track Models," Inst. Elect. Eng. Proc. Vis. Img. Sig. Proc. **147**, 493-501 (2000).
61. D. Kim, "ANIQUE: an auditory model for single-ended speech quality estimation," IEEE Trans. on Speech and Audio Proc. **13**, 821-831 (2005).
62. T.-Y. Yen, J.-H. Chen, and T.-S. Chi, "Perception-based objective speech quality assessment," in Proc. ICASSP, 4521-4524 (2009).
63. ITU-T, *P.Supp23 : ITU-T coded-speech database*, 2004.
64. F. Rumsey, "Subjective assessment of the spatial attributes of reproduced sound," in Proc. AES 15<sup>th</sup> Intern. Conf., 122-135 (1998).
65. J. Berg, "Evaluation of perceived spatial audio quality," in Proc. 9<sup>th</sup> World MultiConf. on Syst. Cyber. and Inform., 10-14 (2005).
66. F. Rumsey, "Spatial audio and sensory evaluation techniques context, history and aims," in Proc. Spatial Audio & Sensory Eval. Tech., 1-7 (2006).
67. S. H. Park, S. W. Ryu, J. Y. Park, and J. Shin, "Analysis and evaluation of PEAQ: Objective method for perceived audio quality measurement," (in Korean), in Proc. ITFE, 234-239 (2003).
68. M. Šalovarda, I. Bolkovac, and H. Domitrovic, "Estimating perceptual audio system quality using PEAQ algorithm," in 18<sup>th</sup> Intern. Conf. on Appl. Electromag. and Comm., 1-4 (2005).
69. G. Markovic, *Analysis of Methods for Objective Evaluation of Quality of Audio Signals and Application in Implementation of An Ecoder on A Class of Digital Signal Processors* (Ph.D Thesis, University of Novi Sad, 2006).
70. S. Lee, N. Choi, and K. Sung, "A study on the subjective quality assessment of sound," Inform. Comm. Mag. **22**, 1386-1396 (2005).
71. B. D. Jun, N. Choi, H.-W. Ko, and K. Sung, "Intelligent diagnostics for sound reproduction system by the use of PEAQ," Adv. in Neural Networks-ISNN, 382-389 (2006).
72. B. Feiten and I. Wolf, "Audio adaptation according to usage environment and perceptual quality metrics," IEEE Trans. on Multimedia **7**, 446-453 (2005).
73. E. S. Myakotnykh and S. U. Peter, "Towards a computational quality model for IP-based audio," in Proc. QoMEX, 110-115 (2009).

74. B. C. J. Moore and C.-T. Tan, "Perceived naturalness of spectrally distorted speech and music," *J. Acoust. Soc. Am.* **114**, 408-419 (2003).
75. B. C. J. Moore, C.-T. Tan, N. Zacharov, and V.-V. Mattila, "Measuring and predicting the perceived quality of music and speech subjected to combined linear and nonlinear distortion," *J. Audio Eng. Soc.* **52**, 1228-1244 (2004).
76. C.-T. Tan, B. C. J. Moore, N. Zacharov, and V.-V. Mattila, "Predicting the perceived quality of nonlinearly distorted music and speech signals," *J. Audio Eng. Soc.* **52**, 699-711(2004).
77. R. Huber and B. Kollmeier, "PEMO-Q - A new method for objective audio quality assessment using a model of auditory perception," *IEEE Trans. on Audio, Speech and Lang. Proc.* **14**, 1902-1911 (2006).
78. J. C. Hardin and C. D. Creusere, "Objective Analysis of Temporally Varing Audio Quality Metrics," in *Proc. 42<sup>nd</sup> Asilomar Conf.*, 1245-1249 (2008).
79. C. D. Creusere and J. C. Hardin, "Assessing the quality of audio containing temporally varying distortions," *IEEE Trans. on Audio, Speech and Lang. Proc.* **19**,711-720 (2011).
80. R. Vanam, "Scalable perceptual metric for evaluating audio quality," in *Proc. Rec. of the Thirty-Ninth Asilomar Conf.*, 319-323 (2005).
81. R. Vanam and C. D. Creusere, "Evaluating low bitrate scalable audio quality using advanced version of PEAQ and energy equalization approach," in *Proc. ICASSP*, 189-192 (2005).
82. C. D. Creusere, K. D. Kallakuri, and R. Vanam, "An objective metric of human subjective audio quality optimized for a wide range of audio fidelities," *IEEE Trans. on Audio Speech and Lang. Proc.* **16**, 129-136 (2008).
83. J. G. A. Barbedo, A. Lopes, "A New cognitive model for objective assessment of audio quality," *J. Audio Eng. Soc.* **53**, 22-31 (2005).
84. L. Abanto, G. Kemper, and J. Telles, "A novel fuzzy logic-based metric for audio quality assessment: objective audio quality assessment," in *Proc. Telecom. (CONATEL)*, 17-20 (2011).
85. S. George, S. Zielinski, F. Rumsey, "Initial developments of an objective method for the prediction of basic audio quality for surround audio recordings," *AES 120<sup>th</sup> Convention*, paper no. 6686 (2006).
86. I. Choi, B. G. Shinn-Cunningham, S. B. Chon, K.-M. Sung, "Objective measurement of perceived auditory quality in multi-channel audio compression coding systems," *J. Audio Eng. Soc.* **56**, 3-17 (2008).
87. F. Rumsey, S. Zielinski, P. Jackson, M. Dewhurst, R. Conetta, S. George, S. Bech, D. Meares, "QESTRAL(Part 1): quality evaluation of spatial transmission and reproduction using an artificial listener," *AES 125<sup>th</sup> Convention*, paper no. 7595 (2008).
88. R. Conetta, *Towards the automatic assessment of spatial quality in the reproduced sound environment* (Ph.D Thesis, University of Surrey, 2011).
89. A. J. Manders, D. M. Simpson, and S. L. Bell, "Objective prediction of the sound quality of music processed by an adaptive feedback canceller," *IEEE Trans. on Audio, Speech, and Lang. Proc.* **20**, 1734-1745 (2012).
90. S. Kandadai, J. Hardin, and C. D. Creusere, "Audio quality assessment using the mean structural similarity measure," in *Proc. ICASSP*, 221-224 (2008).
91. Y. Yue, X. Xiang, and W. Yaodu, "A novel objective method for evaluating the quality of streaming audio," in *Proc. IC-BNMT*, 555-559 (2009).
92. Y. Huh and K. Oh, "Report on ITU-R SG6 meeting," *TTA J.* **135**, 129-131 (2011).

## 저자 약력

### ▶ 오원근(Wongun Oh)



1989년 2월: 한양대학교 전자통신공학과 학사  
 1992년 2월: 한양대학교 전자통신공학과 석사  
 1997년 2월: 한양대학교 전자통신공학과 박사  
 1997년 3월~현재: 순천대학교 멀티미디어공학과 교수