

# IPTV에서 평가주기 변화가 주관적 영상 화질 평가에 미치는 영향 분석

정희원 하 상 용\*, 김 진 철\*, 정 운 영\*, 최 재 영\*\*, 종신회원 노 병 희\*\*\*

## Impact of the Varying Estimation Period on Subjective Video Quality Evaluation for IPTV Service

Sang-yong Ha\*, Chin Chol Kim\* , Woon-Young Jung\*, Jae-young Choi\*\* *Regular Members*,  
Byeong-hee Roh\*\*\*° *Lifelong Member*

### 요 약

영상 품질의 주관적 측정을 위한 다양한 표준 방법들이 제시되었으나, 이들 방법들은 원본 기준 영상이 존재하는 환경에서 10초 내외의 짧은 시간의 시청을 기반으로 한다. 그러나, IPTV 시청 환경에서는 장시간의 시청(예를들어 30분이상)이 이루어지므로, 표준 방법들에서의 10초 시간 주기의 평가는 현실적으로 적절하지 못하다. 본 논문에서는 IPTV 제공 환경에서 주관적 화질 평가를 위하여, 평가 주기 변화가 MOS에 미치는 영향을 분석하고, IPTV의 주관적 화질 평가에 적합한 평가 주기를 선정하기 위한 방안을 제시한다.

**Key Words** : IPTV, QoE (Quality of Experience), Subjective Video Quality, MOS (Mean Opinion Score), ACR (Absolute Category Rating).

### ABSTRACT

There have been considerable standardization works to measure subjective video quality. However, these standardization methods are based on the situation where they have original or reference video sequences, and measure the video quality by looking the sequence during short time duration, e.g. 10 seconds. Since people's TV watching time durations in IPTV services are very long (e.g. more than 30 minutes), the quality measurements by the standards are not practical. In this paper, we analyze the effect of video quality rating period to MOS (Mean Opinion Score), and suggest a proper rating duration of subjective video quality for IPTV services.

### I. 서 론

IPTV 서비스의 보급에 따라 사람이 체감하여 느끼는 품질의 척도를 나타내는 QoE (Quality of Experience)에 대한 관심이 높아지고 있다<sup>[1]</sup>. 영상 QoE에 대한 품질 평가 방법은 사람이 영상 서비스를

시청했을 때 체감하는 품질을 MOS (Mean Opinion Score) 로 정량화하여 나타내어 준다. ITU-R BT.500<sup>[2]</sup>과 ITU-T P.910<sup>[3]</sup>에는 이와 같은 TV화질의 주관적 품질 평가를 위한 다양한 방법들을 제시하고 있다. IPTV 전송중에 발생하는 열화요인인 손실, 지연들이 품질에 미치는 영향에 대한 많은 연구들이 수

※ 본 연구는 방송통신위원회의 출연금으로 수행한 연구개발사업의 결과임

\* 한국정보화진흥원 ({yong, cckim, wyjung}@nia.or.kr), \*\* 성균관대학교 정보통신공학부 (jychoi@ece.skku.ac.kr)

\*\*\* 아주대학교 정보통신전문대학원/정보컴퓨터공학부(bhroh@ajou.ac.kr), (° : 교신저자)

논문번호: KICS2011-02-130, 접수일자: 2011년 2년 28일, 최종논문접수일자: 2011년 3월 29일

행되었다<sup>4,6)</sup>. 그러나 이들 방법들은 품질 평가를 위한 주관적 품질 평가 방법에 대한 구체적인 고려는 제공하고 있지 않다.

국제 표준에서 제시된 화질평가는 크게 원본영상을 필요로 하는 상대평가와 그렇지 않은 절대평가로 나누어 진다. 영상의 객관적 QoE 측정 방법으로서, 원본 영상의 비교 유무로 구분되는 FR (Full Reference) 방식, RR (Reduced Reference) 방식, 그리고 NR (No Reference) 방식들<sup>7)</sup>은 IPTV에서의 화질평가를 위한 운용성이나 실용성 면에서 한계를 갖는다<sup>8)</sup>. 실시간 IPTV 태내 시청환경에서의 화질 평가를 고려할 경우, DSCQS(Double-Stimulus Continuous Quality-Scale method) 나 DSIS (Double-Stimulus Impairment Scale), SAMVIQ (Subjective Assessment Method for Video Quality Evaluation)와 같은 상대평가 방법들<sup>2)</sup>은 원본영상을 필요로 하므로 적용이 불가하나, ACR (Absolute Category Rating)<sup>3)</sup>은 원본 영상 없이 수신 영상으로만 평가하는 절대 평가 방식으로서 IPTV서비스의 영상 QoE 품질평가에 적합한 방식이다. 그러나, ACR은 10초 길이의 영상을 한번만 본 후 10초 이내에 화질의 수준을 5점 평가척도로 평가하는 방식으로서, 매우 잦은 평가 주기로 인해 이용자가 실시간으로 IPTV를 시청하는 실제 환경에 적용하기에는 적합하지 않다.

본 논문에서는 IPTV에서 화질 평가주기 변화가 MOS에 미치는 영향을 분석하고, IPTV 화질 평가에 적합한 평가 주기를 선정하기 위한 방안을 제시한다. 논문에서는 대표적인 주관적 품질 평가 방법들 중의 하나인 ACR과 IPTV 평가 주기에 따른 주관적 품질인 MOS 측정 결과를 비교하여, 이 둘간의 상관성을 분석하여 평가 주기를 선정한다.

본 논문의 구성은 다음과 같다. 제 2장은 ACR 표준 방법론을 기술하고, 제 3장은 본 논문에서의 품질 평가를 위한 방법론을 설명한다. 그리고, 제4장에서는 실험 결과를 보이고, 제 5장에서는 결론을 맺는다.

## II. ACR을 사용한 영상 품질 측정 방법

ACR 방법<sup>3)</sup>은 10초 길이의 영상을 한번만 본 후 10초 이내에서 화질의 수준을 평가하는 절대평가 방식으로, VQEG와 ITU를 중심으로 멀티미디어 화질평가 모델의 개발에 사용되고 있다. 영상을 1회만 보여 줌으로써 짧은 시간 내에 많은 수의 동영상 평가가 가능하다.

ACR 방법의 평가 절차의 한 예를 그림 1에 나타내

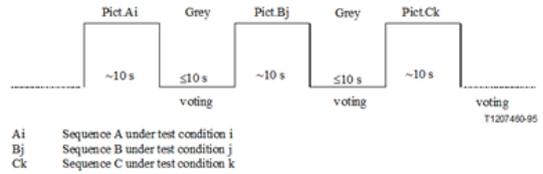


그림 1. ACR의 평가절차

었다. 여기에서는 각 영상 A, B, C에 대하여 조건 i, j, k를 적용하여 제작한 영상들인 Ai, Bj, Ck를 10초 이내로 한 번만 제시하고, 평가자는 10초이내의 voting구간안에 이 영상의 품질을 1(Bad), 2(Poor), 3(Fair), 4(Good), 5(Excellent)의 5개 등급중 하나로 평가하게 된다. 영상을 보여주는 시간은 영상의 콘텐츠(드라마, 스포츠 등)에 따라 영상 길이(10초)를 줄이거나 늘릴 수 있다.

## III. IPTV 화질 평가 방법

### 3.1 실험 영상의 제작

실험을 위한 영상은 배경 화면 및 피사체의 움직임, 화면 변경 등 다양한 영상을 대변할 수 있도록 표 1과 같이 스포츠형, 드라마형, 다큐멘터리형의 총 3개의 대표 영상을 선정하고, 각 실험 영상의 길이는 실시간 시청 환경을 고려하여 30분(다큐멘터리) 60분(드라마), 90분(스포츠)으로 준비하였다.

그리고, 화질 평가를 위하여 원본 영상에 열화 요소(패킷 손실)를 강제로 반영하여 품질이 저하된 영상을 제작하였다. 열화 영상을 제작하는 과정은 그림 2와 같다. 먼저 IPTV에 사용되는 스포츠, 드라마, 다큐멘터리 3가지 종류의 영상을 IPTV 서비스를 통하여 셋탑박스에 수신되는 과정에서 헤드엔드 단에서 스트림 캡처 장치를 사용하여 열화가 없는 원본 영상을 저장하였다. 이 원본 영상에 대하여 열화 영상 발생장치인 허리케인<sup>9)</sup>을 사용하여 열화된 영상을 제작하였다. 열화된 영상을 스트림 발생 장치를 통하여 평가용 TV에

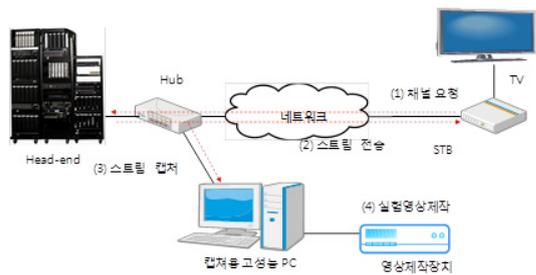


그림 2. 실험 영상 제작 구성도

표 1. 실험을 위한 제작영상 개요

구분	설명
스포츠	축구, 농구 등 넓은 각도의 화면에서 카메라의 움직임이 균일하고, 균일한 색상의 배경 화면위에서 신속하게 이동하는 작은 피사체(선수들)에 의해서 대부분의 움직임이 많은 영상 
드라마	드라마 등 정적인 배경화면에서 움직임이 적고, 주로 눈, 입술 및 얼굴의 표정 등 국부적인 영역에서 움직임이 많은 영상 
다큐멘터리	다큐, 풍경 등 넓은 각도로 촬영되어 움직임이 작고 균일한 각도를 유지하는 영상으로 카메라의 이동에 의해 전체 화면에서 움직임이 있는 영상 

나오게 하고, 이를 평가 단원이 시청하여 ACR 방법과 일반 MOS 측정 방법을 통하여 화질 평가를 수행하였다. ACR과 MOS를 측정하는 환경과 방법은 각각 3.2절과 3.3절에서 설명한다.

3.2 주관적 화질 측정을 위한 실험 환경의 구성

ACR 국제 표준<sup>[3]</sup>에서는 TV 화질 평가를 위한 실험 집단의 구성을 위하여 통계적으로 의미를 갖을수 있도록 15명 이상의 평가자로 구성하도록 권고하고 있다. 본 실험에서는 ACR 평가를 위해서는 총 20명의 평가자를 모집하였으며, IPTV 평가를 위해서는 2분, 5분, 10분 30분 각 주기별로 각각 20명씩 총 80명으로 구성하였다. 그리고 국제표준에서 권고한 평가자의 시청각도 30도를 감안하여 TV 1대당 4명씩으로 배치하였으며, 평가 방법에 대한 오류를 방지하기 위해 평가 척도, 시청 시간, 소요 시간 등 기본적인 방법에 대한 교육을 실시하고, 사전 연습을 10회 이상 반복하였다. ITU-R BT.500-11 권고에 따라, TV는 바닥으로부터 0.5m 이상, 평가자와 TV와의 거리는 2.7m로

하고, 주변 밝기는 조도 200을 기준으로 형광등을 켜 상태에서 시청토록 하였다. 세부 내용은 표 2와 같다.

표 2. 실험환경

구분	내용	비고
TV종류	액정(LCD)	기본모드로 설정
TV크기	32인치	ITU-R BT.500-11의 PVD(preferred viewing distance), 영상크기에 따른 시청거리를 반영
시청높이	TV는 바닥으로부터 높이 0.5m 이상	
시청거리	평가자와 TV와의 시청거리는 2.7m	ITU-R BT.500-11 시청각도 30° 권고 준수
시청각도	시청각도는 정면 30도 이내로 제한	
주변환경	실험실에서 형광등을 켜 상태에서 시청	평가 환경의 주변 밝기(조도 200 기준)
TV당 평가자수	4명	시청각도 30° 이내

3.3 측정 방법

ACR 측정을 위하여, 제작한 실험 영상에서 모든 영상을 10초 간격으로 구분하였고, 10초 동안의 평가 시간을 위해 빈 화면을 삽입하였다. 2분, 5분, 10분, 30분 주기의 IPTV 실험을 위해서는 빈 화면 삽입 없이 평가 중에도 영상이 지속적으로 상영되게 하였다. ACR 실험의 경우, 15분 단위로 10초간 시청, 10초간 평가함에 따라 총 12회(30분-2회, 60분-4회, 90분-6회) 평가시험을 실시하여 1080개의 평가 점수가 도출되었다.

IPTV 실험의 경우, 평가 주기는 2분, 5분, 10분, 30분 4가지 경우 각각에 대해 시청 시간을 주고 시험 횟수는 총 12회로 동일하며 이를 통해 150개의 평가 점수가 도출되었다. 평가척도는 5등급으로 동일하다. 이렇게 측정된 결과는 ACR값과 2분주기, 5분주기, 10분주기, 30분주기 각각에 대한 MOS값을 서로 비교해 보았다.

비교에 있어, 한 평가 주기(2분, 5분, 10분, 30분) 동안에 측정된 MOS값은 1개이지만, ACR의 경우는 각 평가 주기내에 10초길이의 모든 구간에 대하여 측정되므로 여러 개의 값들이 도출되어 직접적인 상호 비교가 곤란하다. 본 논문에서는 이를 위하여, ACR 결과 값은 주기내에서 측정된 10초길이의 평가값에 대한 평균값을 적용하였다. 그러나, 이러한 평균값 산출에 있어서도 ACR 결과 값들의 대부분은 열화가 적용

되지 않은 원본 영상 평가 결과 값들로 구성되게 되어, 평균을 산출하게 될 경우 열화 영상의 저하된 결과 값이 반영되지 못하는 문제점이 있다. 따라서 ACR 측정에서는 열화가 적용되어 나타나는 10초주기들에 대한 ACR값들만 추출하여 이들에 대한 평균치를 사용하였다.

이러한 과정을 2분 평가주기에 대한 그림 3의 예를 들어 설명하기로 한다. 60분 드라마의 경우, 30개의 2분주기 구간이 존재하고, 각 2분 주기 동안에는 모두 12개의 10초 구간이 존재한다. 그림 3에서와 같이 첫 번째 2분주기내의 12개 10초구간중 3개의 구간 (3, 6, 9번째 10초구간)에 열화를 발생시키고, 두번째 2분주기내에서는 2개(3, 9번째 10초구간)의 열화가 발생한 것으로 가정하기로 한다. 첫번째 2분주기에서 MOS 값은 2분 동안의 영상을 모두 시청한후에 한번 수행하여 평가단으로부터 측정된  $MOS_1$  이 된다. 비교를 위한 ACR 측정값은 열화가 발생한 3개의 3, 6, 9 번째 10초구간에서 ACR 에 의하여 측정된  $ACR_{1,3}$ ,  $ACR_{1,6}$ ,  $ACR_{1,9}$  들의 산술평균값으로서  $ACR_1$  이 구해진다. 그리고 첫 번째 2분주기에서는 이들 측정된  $MOS_1$  값과  $ACR_1$  값을 서로 비교하게 된다. 마찬가지로, 두번째 2분 주기에서도 ACR값과 MOS값을 마찬가지로 방법으로 측정해 내고, 이와 같은 방법으로 다른 주기들에서도 측정을 수행한다.

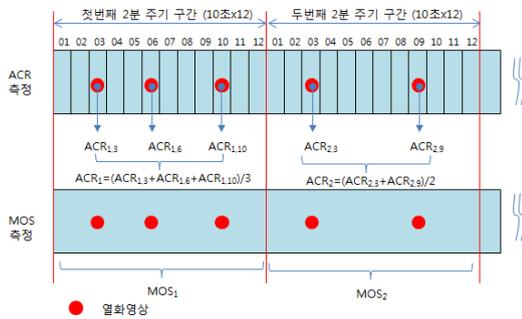


그림 3. 영상 열화 적용 및 이에 대한 ACR과 MOS 측정 방법 (2분주기의 예)

#### IV. 실험결과

3장에서 기술한 실험 방법에 따라, 다큐멘터리, 드라마, 스포츠의 3가지 종류 영상에 대하여 10초 길이의 영상을 대상으로 품질을 측정하는 ACR 방식에 의한 측정값(ACR)과 영상 평가 주기를 2분, 5분, 10분, 30분 길이로 하였을 때 측정된 품질 값 (MOS)들을

구하였다.

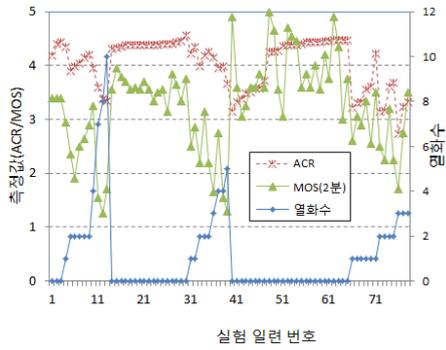
그림 4에는 영상을 10초 단위로 구분하였을 때, 평가주기내의 단위구간들 중에서 열화가 적용된 구간 수(열화수)에 따라 측정된 ACR 값과 MOS값을 비교하여 나타내었다. 그림 4 (a)는 평가주기가 2분인 경우에 대한 결과를 보여준다. 2분의 평가주기에는 10초 단위구간의 개수는 12개가 된다. 10초 단위구간은 ACR에 의한 평가주기임에 주의한다. 그림 4 (a)에서 열화의 개수는 이들 12개의 10초 단위구간들 중에서 열화가 발생한 구간의 개수를 의미한다. 3장에서 설명한 바와 같이, 그림 4 (a)에서 ACR값들은 이들 열화가 발생한 각각의 10초단위 구간들에 대하여 ACR 방법을 사용하여 평가한 품질척도값들을 평균하여 나타낸 값들이다. 그리고, MOS값들은 전체 2분동안의 영상을 완전히 시청한 후에 사용자가 평가한 품질척도 값을 의미한다. 마찬가지로, 각각 평가주기가 5분, 10분과 30분인 경우에 대하여 구한 결과들은 그림 4 (b), 그림 4 (c), 그리고 그림 4 (d)에 나타내었다.

열화수 (열화가 적용된 10초 단위구간의 수)가 0이거나 매우 작은 경우에는 ACR과 MOS 값들간에 큰 차이가 보이지 않는다. 그러나, 평가주기가 2분과 5분인 경우, 열화수가 커질수록 ACR과 MOS값간의 차이가 크게 남을 볼 수 있다. 반면에, 10분 주기의 경우는 열화수의 변화에 무관하게 ACR과 MOS 값이 큰 차이를 보이지 않고 서로 근접한 값을 보여준다. 30분 주기의 경우는 제작영상의 길이의 제한으로 다른 주기들에서와 같은 자세한 결과를 보여주지 않으나, 열화수의 증가에 따른 ACR과 MOS값이 10분주기의 경우보다 다소 크게 나타나고 있다.

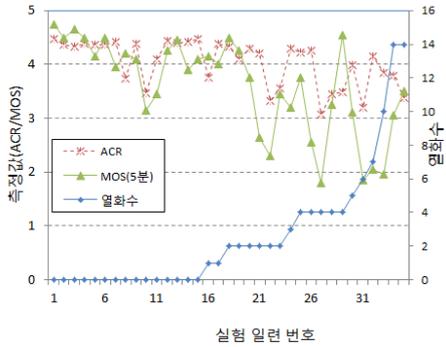
그림 5에는 ACR과 MOS값들간의 상관성 정도를 분석하기 위하여 이들간의 값의 관계를 직접 그래프로 표현하여 나타내었다. 2분과 5분 평가주기에 대한 결과인 그림 5 (a)와 그림 5 (b)는 ACR과 MOS값들이 상호연관성을 갖지 않고 분산되는 경향을 보여준다. 반면에, 그림 5 (c)의 10분주기의 경우는 선형적인 분포를 보여준다. 그림 5 (d)의 30분 주기의 경우는 표본의 부족으로 정확한 분포를 보여주지는 못하나, 집중되지 않고 분산되는 경향을 보여주는 것으로 판단된다.

그림 6에는 그림 5의 결과를 기반으로 한 상관도값을 영상 종류별과 전체에 대한 평균으로 나타내었다. 두 변수 X와 Y간의 상관도는 다음과 같이 정의된다.

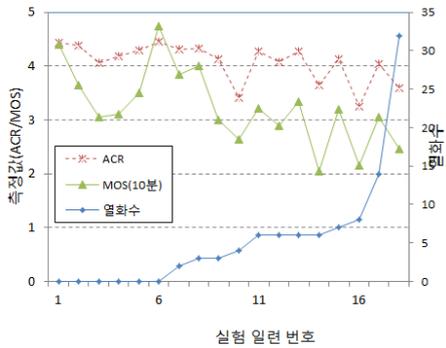
$$corr(X, Y) = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (1)$$



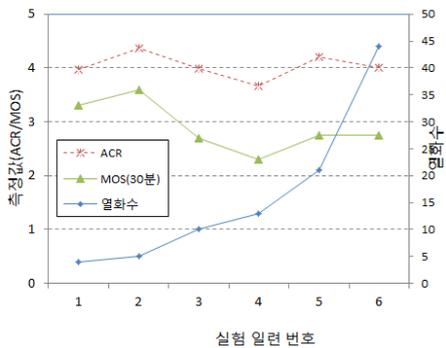
(a) 2분 측정 주기



(b) 5분 측정 주기

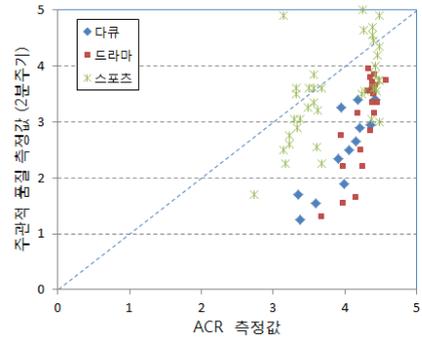


(c) 10분 측정 주기

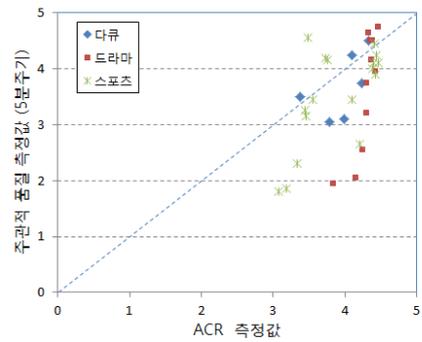


(d) 30분 측정 주기

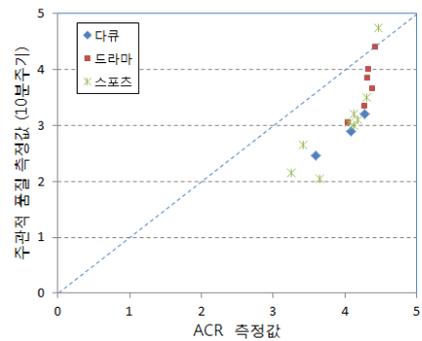
그림 4. 열화 정도에 따른 ACR값과 평가주기별 MOS 변화



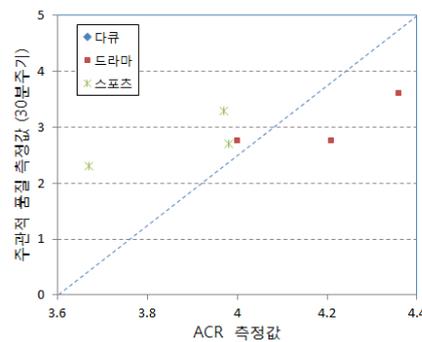
(a) 2분 측정 주기



(b) 5분 측정 주기



(c) 10분 측정 주기



(d) 30분 측정 주기

그림 5. 평가주기에 따른 ACR과 MOS 들간의 상관관계

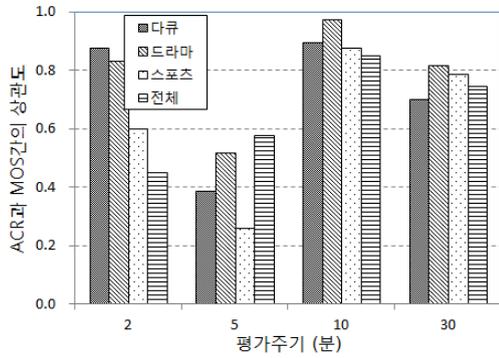


그림 6. 평가주기에 따른 상관도

여기에서,  $\mu_X$ 와  $\mu_Y$ 는 각각 X와 Y의 평균값,  $\sigma_X$ 와  $\sigma_Y$ 는 X와 Y의 표준편차를 의미한다.

전체적으로 10분주기에서 가장 큰 상관도를 보여주는 것을 확인할 수 있다. 그러나, 그림 6에 보인 바와 같이 영상의 종류에 따라 평가주기에 따른 상관도 값이 다소 다르게 남을 볼 수 있다. 드라마 영상의 경우, ACR과 MOS간의 상관도가 10분주기일 때 가장 큰 값을 나타내나, 시간 주기에 크게 영향을 받지 않음을 보여준다. 스포츠 영상의 경우는 2분과 5분주기에서는 매우 낮은 상관도를 보여주고, 10분주기에서 가장 크고, 30분주기에서 다음으로 큰 경향을 보여준다. 이것은 스포츠 영상이 장면변화가 많으므로, 평가주기가 짧을수록 평가 주기별 움직임의 정도에 따라 평가 결과가 크게 달라지고, 시간이 길어지게 되면 이러한 움직임의 영향을 적게 받게 되는 것으로 판단된다. 다큐멘터리 영상의 경우는 2분주기와 10분주기에서 큰 상관도를, 5분주기에서는 가장 작은 상관도를, 30분 주기에서는 작지 않은 상관도를 보여주면서 불규칙한 패턴을 나타낸다. 이것은 다큐멘터리 영상의 경우, 자연 영상뿐만 아니라, 인공적으로 제작 또는 조합한 영상들이 포함되는 복잡한 특성을 갖고 있기 때문으로 판단된다.

## V. 결 론

본 논문에서는 IPTV에서 화질 평가주기 변화가 주관적 영상 품질 평가에 미치는 영향을 분석하였다. 분석을 위한 실험 영상을 다큐멘터리, 드라마, 스포츠의 3가지 카테고리에 구분하여 제작하였고, 표준 방법인 10초주기 평가 방식인 ACR 방법의 결과와 평가 주기를 달리하며 측정된 MOS간의 상관관계를 구하였다. 분석결과는 크게 다음의 세가지로 정리된다.

첫째, 열화 정도가 적거나 거의 없는 경우에는 ACR과 MOS의 결과차이가 큰 차이를 보이지 않으나, 열화 정도가 클수록 평가주기가 짧은 경우에는 ACR과 MOS의 결과 차이가 크게 나타났다.

둘째, 평가주기가 2분, 5분의 경우에는 ACR과 MOS의 상관도 분포가 집중하지 않고 분산되는 반면에, 10분인 경우에는 선형으로 집중하는 경향을 보여 주었다.

셋째, 평가주기의 변화에 따른 상관도는 영상의 종류에 따라 달라지는 경향을 보였다. 즉, 드라마의 경우에는 ACR과 MOS의 결과가 주기에 크게 영향을 받지 않는 것으로 분석되었다. 스포츠의 경우는, 짧은 주기에서는 움직임의 많고 적음에 따른 구간이 주기 내에 포함되는지의 여부에 따라 ACR과 MOS의 결과가 불규칙하게 나오고, 긴 주기에서는 이러한 영향을 덜 받는 것으로 분석되었다. 다큐멘터리 영상의 경우는 자연영상과 인공영상의 조합에 기인하여 전체적으로 불규칙한 특징을 보여주었다. 따라서, 다큐멘터리 영상의 경우는 이러한 자연영상과 인공영상의 조합 정도에 따른 영향도를 추가적으로 더 조사할 필요가 있다고 본다.

IPTV 서비스의 보급에 따라 사람이 체감하여 느끼는 품질의 척도를 나타내는 QoE (Quality of Experience)에 대한 관심이 높아지고 있으나, IPTV 환경에서 영상 QoE 측정을 위한 적절한 방법의 제시는 아직까지 이루어지고 있지 않다. 본 논문의 결과는 여러가지 영상 QoE 측정을 위한 요소들 중에서 평가 주기의 선정과 관련하여 매우 중요한 자료로 활용 가능할 것으로 기대된다.

## 참 고 문 헌

- [1] A. Takahashi, D. Hands, V. Barriac, "Standardization Activities in the ITU for a QoE Assessment of IPTV," IEEE Communications Magazine, Vo.46, No.2, pp.78-84, Feb. 2008
- [2] ITU-R Rec. BT.500-11, "Methodology for the Subjective Assessment of the Quality of Television Pictures," ITU-R, Dec. 2002
- [3] ITU-T Rec. P.910, "Subjective video quality assessment methods for multimedia applications," ITU-T, Sep. 1999
- [4] R. Shumeli, O. Hadar, R. Huber, M. Maltz, M. Huber, "Effects of an Encoding Scheme on



노 병 희 (Byeong-hee Roh)

중신회원



1987년 한양대학교 전자공학과  
(학사)

1989년 한국과학기술원 전기및  
전자공학과 (석사)

1998년 한국과학기술원 전기및  
전자공학과 (박사)

1989년~1994년 한국통신 통  
신망 연구소

1998년~2000년 삼성전자

2000년~현재 아주대학교 정보통신전문대학원 교수

2005년 Stony Brook University 방문교수

<관심분야> 멀티미디어 네트워크 및 응용, 유무선  
네트워크 QoS 및 트래픽 엔지니어링, 유비쿼터스  
센서 네트워크(RFID/USN), 인터넷 보안, 국방전  
술통신네트워크