

EXPERT
KNOW-HOW디지털 영상의 화질평가 ②
ITU 표준권고안의
주관적 화질평가 방법론

+ 정태욱 연세대학교 계산과학공학과 연구교수

① 디지털 영상의 특징

② 주관적 화질평가

③ 객관적 화질평가



들어가며

전자 및 정보통신 기술의 발전으로 다양한 유무선 환경에서 다양한 멀티미디어 서비스가 제공되고 있으며 이러한 환경에서 제공되는 콘텐츠의 품질에 대한 QoS(Quality of Service) 보장 문제가 중요한 이슈로 부각되고 있다. 품질 평가는 다수의 평가자에 의한 주관적 품질 평가로 시행되어야 하지만, 시간과 비용 면에서 많은 제약과 한계를 갖고 있으며 실시간 평가가 불가능하다는 단점이 있다. 최근 ITU를 중심으로 주관적 화질평가와 높은 상관관계를 갖는 객관적 화질평가 방법에 대한 표준화가 진행되고 있다. 주관적 화질평가는 품질을 평가하는 방법 중에서 가장 정확한 방법이다. 주관적 화질이란 보편적 인간이 시각적으로 느끼는 체감품질(QoE: Quality of Experience) 관점에서의 화질을 의미한다. 주관적 평가를 위해서는 측정하고자 하는 항목(화질, 품질, 음질 등)을 결정하고 이에 부합하는 평가 방법론을 결정하여야 한다. 현재까지 주관적 화질평가에는 ITU-R BT.500과 ITU-T P.910에 기술된 방법론이 주로 사용되고 있다. 본 논문에서는 ITU 표준권고안에 기술된 주관적 화질평가 방법론을 살펴보고자 한다.

ITU 표준의 주관적 화질평가 방법론

ITU에서는 표준권고안 ITU-T P.910[1] 및 ITU-R BT.500[2]을 통해 멀티미디어 및 텔레비전 화상에 대한 주관적 품질평가 방법론을 표준화하였다. 이 표준권고안에 기술되어 있는 방법들은 대부분의 멀티미디어 비디오 및 텔

레비전 형식을 포함하지만, HD 영상에 대한 주관적 품질 평가 방법은 별도의 권고안 ITU-R BT.710(HDTV의 주관적 품질평가 방법)[3], ITU-R BT.1210(품질평가에 사용되는 test material)[4] 등에 기술되어 있다.

1. ITU-T Recommendation P.910

(Subjective video quality assessment methods for multimedia applications)

표준권고안 ITU-T P.910은 화상회의, 저장 및 검색 애플리케이션, 원격진료 애플리케이션 등 단방향 서비스의 전반적인 비디오 품질을 측정할 수 있는 방법론을 기술하고 있다. 다음에 소개할 세 가지 주관적 화질평가 방법론 이외에도 원본영상(reference)의 획득 및 저장, 테스트에 사용될 장면의 특성, 실험절차 및 실험환경에 대한 내용이 함께 기술되어 있다. ITU-T P.910에서 기술한 평가 방법들은 원본영상의 명시 유무에 따라 두 가지 범주로 분류할 수 있는데, DCR 방법의 경우 원본영상의 명시되어 있으며, ACR 방법, ACR-HR 방법 그리고 PC 방법의 경우 원본영상의 명시되어 있지 않다.

(a) Absolute Category Rating (ACR)

ACR 방법은 평가할 영상을 각각 한 번씩만 보여주고 재생된 영상의 품질을 범주화(category)하여 측정하는 방식이다. 각 영상의 절대적인 평가 수치를 얻을 수 있는 방법으로 주관적 평가 수치는 평가자의 평가 수치를 평균한 MOS(mean opinion score)를 사용한다. 또한 원본영상과의 화질 차이를 보

기 위해서 재생되는 영상에 원본영상을 숨겨 평가자들이 인지하지 못하고 평가를 하게 하는 방법이 ACR-HR(hidden reference)이다.

이 경우 원본영상과 평가영상간의 화질차이 값인 DMOS(difference MOS)를 사용할 수도 있다. ACR 및 ACR-HR 방법은 실험이 쉽고 빠르다는 장점이 있다. ACR 방법과 비교하여, ACR-HR만의 주요한 장점은 주관적인 점수에서 원본영상의 시각적인 영향을 제거할 수 있다는 점이다. ACR-HR 방법은 모든 원본영상이 최소한 최고의 품질을 제공하므로 대형 실험 환경에 적합하다. [그림 1]은 시간 패턴과 품질 평가의 5가지 범주를 보여주고 있으며, 만약 더 세분화 된 범주가 필요할 경우, 9단계 또는 11단계의 척도를 사용할 수 있다.

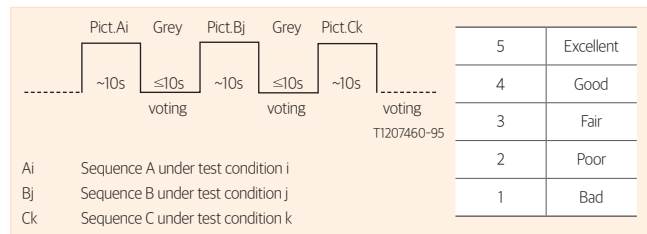


그림 1. ACR 방법의 시간 패턴 및 품질 평가 범주, 출처 : ITU-T P.910

(b) Degradation Category Rating (DCR)

DCR 방법은 원본영상과 평가영상을 시간에 따라 쌍으로 재생하면서 두 영상 간의 화질 차이를 5가지 범주로 묻는 평가방식이다. 즉, 평가자가 원본영상에 대한 처리영상의 열화 정도를 측정하도록 주관적 평가를 구성한다. DCR 방법은 주로 원본영상에 대해 전송된 영상의 정확도(fidelity)를 테스트 하고자 할 때 사용된다. [그림 2]는 DCR의 시간 패턴과 열화 정도를 평가하는 5가지 범주를 보여준다.

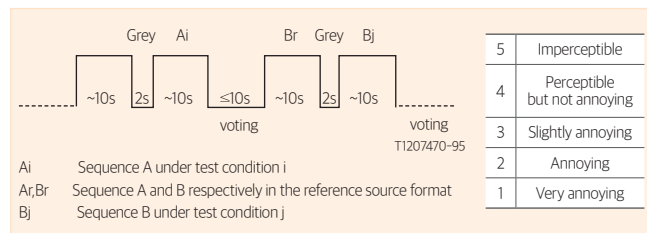


그림 2.

(C) Pair Comparison (PC)

PC 방법은 평가하고자 하는 두 영상을 쌍으로 재생하여 두 영상 중에서 주관

적 선호를 묻는 평가 방법이다. 예를 들면 한 가지 원본영상에 5가지 형태의 열화영상을 가지고 있다면 총 15(=6C2)쌍을 실험하여 각각의 선호를 조사하는 방식이다. 평가영상의 종류가 많다면 모든 경우를 실험하는 것보다 일부의 쌍을 선별해서 하는 방식이 고려된다. PC 방법은 거의 동일한 품질을 나타내는 몇몇 영상에 대해서도 평가할 수 있도록 하는 차별성을 가지고 있다. 아래의 그림은 PC 방법의 시간 패턴을 보여준다.

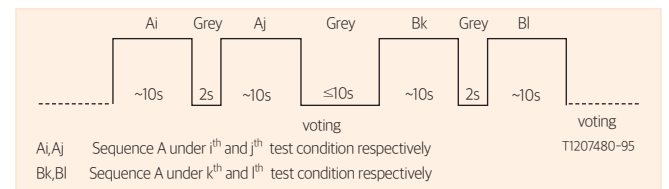


그림 3. PC 방법의 시간 패턴, 출처 : ITU-T P.910

2. ITU-R Recommendation BT.500

(Methodology for the subjective assessment of the quality of television pictures)

ITU-R BT500에 6가지의 주관적 화질평가 방법의 표준권고를 기술하고 있다. 또한 이 표준권고안에는 실험실 환경의 시청조건 및 가정환경의 시청조건, 화면크기에 따른 시청거리, 원본영상의 조건, 평가자 조건 및 실험 세션 구성 등 다양한 내용을 포함하고 있다.

(a) Double stimulus impairment scale (DSIS)

DSIS 방법은 평가자로 하여금 첫 번째는 열화가 없는 원본영상을 보고, 두 번째는 열화가 포함된 동일한 내용의 영상을 보게 한 후, 평가자로 하여금 첫 번째 영상을 고려하여 두 번째 보인 영상의 점수를 측정하게 하는 방식이다. ITU-T P.910에 기술된 DCR과 유사한 방법이고 평가점수는 두 영상의 차이 (impairment)를 나타내는 DCR의 5단계의 범주를 동일하게 사용한다. DCR은 원본영상과 평가영상을 한 번만 보여주고 평가하는 반면 DSIS는 아래와 같이 시간 패턴을 한 번씩 또는 두 번씩 보여주는 방법이 제안되어 있다.

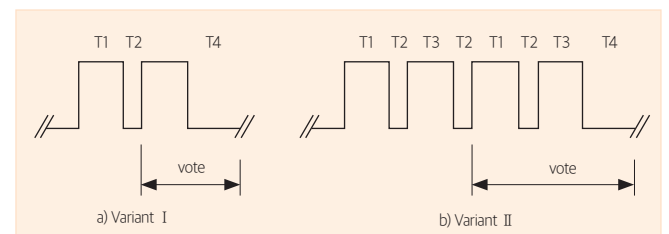


그림 4. DSIS의 시간패턴, 출처 : ITU-R BT.500

(b) Double stimulus continuous quality scale (DSCQS)

DSCQS 방법은 평가영상이 넓은 범위의 화질을 갖지 않을 때 특히 유용하게 사용되는 방법이다. 평가자는 원본영상과 평가영상을 쌍으로 반복적으로 두 번씩 보고 각각의 영상의 평가 점수를 0~100점으로 구성된 수직 눈금자에 표시한다. 한 세션(session)은 30분 정도로 구성하며 무작위로 처리된 영상의 쌍이 보인다. 평가 점수는 원본영상의 점수와 평가영상의 점수의 차이인 DMOS 값을 사용하여 원본에 비해 얼마나 차이가 있는지를 알아보는 평가이다. DSCQS의 평가 점수는 평가영상의 절대적인 점수로 취급되기보다는 원본영상과 평가영상의 점수 차이(DMOS)로 다루어져야 한다.

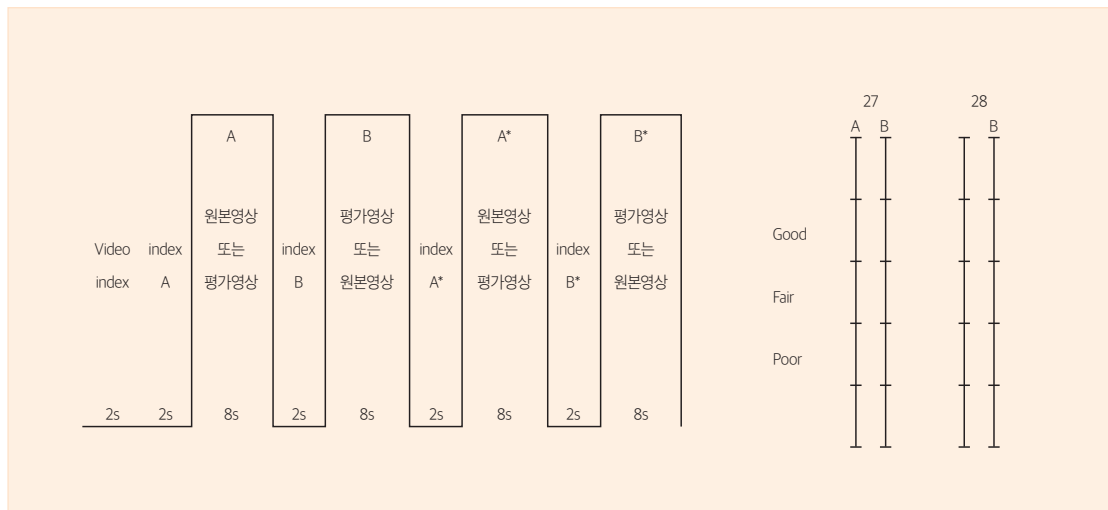


그림 5. DSCQS의 시간패턴 및 평가 스케일, 출처: ITU-R BT.500

(c) Single stimulus (SS)

SS 방법은 재생되는 평가영상에 대하여 평가자가 평가하는 방식으로 ACR과 유사한 방법이다. 평가 스케일은 평가하고자 하는 내용에 따라 ACR의 5가지 범주(Excellent, Good, Fair, Poor, Bad), DCR의 impairment 5가지 범주(Imperceptible, Perceptible but not annoying, Slightly annoying, Annoying, Very annoying), 11-grade numerical categorical scale (SSNCS), 0~100점의 continuous 스케일 등이 사용된다.

(d) Stimulus comparison (SC)

SC 방법은 두 평가영상을 쌍으로 한 번씩만 재생하며, 평가자는 두 영상의 관계에 대한 지수를 평가 점수로 제공하게 된다. SC 방법은 PC 방법과 달리 두 평가영상은 한 모니터에 시간에 따라 순차적으로 재생되거나 두 대의 모니터에 동시에 재생하는 방법이 있다. 평가 수치의 스케일은 7가지의 범주

(Much worse, Worse, Slight worse, The Same, Slightly better, Better, Much better)를 사용하는 방법과 continuous 스케일을 사용한다.

(e) Single stimulus continuous quality evaluation (SSCQE)

SSCQE 방법은 평가영상을 재생하면서 시간에 따라 변화하는 화질을 연속적으로 평가하는 방법이다. 지금까지 소개한 주관적 평가방법은 모두 8~10초 단위의 평가영상 클립(clip)에 대한 전반적인 평가(overall quality evaluation)이며 재생시간에 따라 실시간으로 측정하는 방식은 아니다. SSCQE는 평가영상을 실시간으로 측정하기 위해 10cm의 슬라이드 전자장치를 이용하여 초당 2회씩 샘플링하여 실시간으로 화질을 평가하는 방식이다. 평가 스케일은 DSCQE에서와 같이 5단계(Excellent, Good, Fair, Poor, Bad)로 균등하게 분할된 영역에 슬라이드를 배치시켜 0~100의 수치를 갖는 continuous 스케일을 사용한다.

참고문헌

[1] ITU-T Recommendation P.910, Subjective video quality assessment methods for multimedia applications, 2008.

[2] ITU-R Recommendation BT.500-13, Methodology for the subjective assessment of the quality of television pictures, 2012.

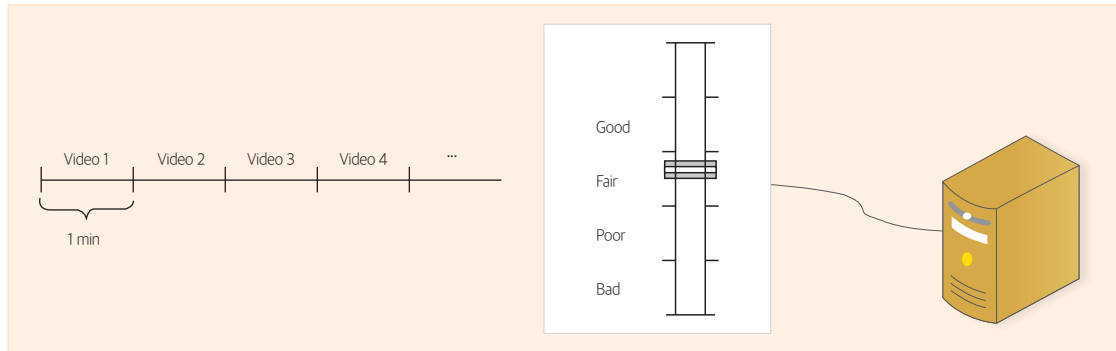


그림 6. SSCQE의 시간패턴 및 평가 스케일(슬라이드)

(f) Simultaneous double stimulus for continuous evaluation (SDSCE)

SDSCE 방법은 원본영상과 평가영상을 동시에 보면서 평가자가 화질의 차이를 평가하는 방법이다. 모니터에 원본영상과 평가영상을 side-by-side로 재생하거나 두 개의 모니터를 동기화하여 재생시키기도 한다. 평가자들은 재생되는 원본영상과 평가영상간의 정확도(fidelity)를 실시간으로 SSCQE의 슬라이드를 이용하여 0~100으로 평가한다. 이때 100점은 perfect fidelity를 나타내고, 0점은 null fidelity를 나타내게 한다.

주관적 화질평가 사례 및 장단점

최근 ITU에서는 주관적 화질평가를 대체할 객관적 화질평가 모델을 표준화하고 있다. 2002년 SDTV에 대한 화질평가를 시작으로 멀티미디어 포맷(QCIF, CIF, VGA), HDTV를 완료하였으며 3DTV 및 UHDTV에 대한 화질평가 표준화를 진행하고 있다. ITU의 화질평가 표준화는 ITU 화질 전문가로 구성된 VQEG(Video Quality Expert Group) 국제 협의회를 통해 진행되고 있다. SDTV-FR(Full Reference) 표준화는 2003년에 진행되었는데, 이때 주관적 화질평가는 DSCQS가 사용되었다. 그러나 DSCQS는 원본영상과 평가영상을 쌍으로 두 번씩 재생하는 특성 때문에 정확한 주관적 점수를 얻을 수 있었으나 시간의 한계로 인해 많은 평가영상을 평가하지 못하고 평가자가 장시간의 평가에 지쳐 정확한 점수를 기대하기 어려운 단점이 있다. 이후 SDTV-RR(Reduced Reference) 표준화에 실시간으로 화질을 평가하기 위해 SSCQE가 제안되었는데, 평가 순간에 슬라이드를 움직이면서 약간의 시간 지연(time delay)이 발생하고, 두 영상의 연결

부분에서 평가 점수가 급격히 변하면서 그 값을 복원하기 어려운 점이 발생하였다. 이로 인해 SDTV-RR(Reduced Reference) 표준화는 잠정 연기되었다.

이후 주관적 평가를 가장 잘 반영하는 DSCQE의 시간상의 단점을 보완하고 그 성능을 대체할 방법으로 ACR 또는 ACR-HR 방법을 표준화에 사용하였다. 멀티미디어(VGA/CIF/QCIF)-FR/RR 영역에 대한 표준화는 2007년 수행되었는데 이때 사용된 주관적 화질평가는 ACR-HR 방법이었다. 2010년의 HD-FR/RR 영역에 대한 표준화 및 2011년 SDTV-RR과 2012년 HD-bitstream 영역에도 ACR-HR 방법이 사용되었다.

맺음말

주관적 화질평가는 평가영상에 대한 주관적인 수치를 도출하는 과정이자 방법이다. ITU-R BT.500 및 ITU-T P.910에서 권고하는 주관적 평가 방법론은 다양한 분야에서 주관적 수치를 도출하고자 표준화된 문서이다. 전반적으로 인간의 주관적 수치를 가장 잘 반영한다고 알려진 DSCQS 방법의 단점을 보완하고자 현재 표준화 과정에서 가장 많이 사용하고 있는 방법이 ACR-HR 방법이다. 그러나 이러한 방법들은 평가영상 클립에 대한 전반적인 주관적 수치 도출에 국한되어 있으며 실시간으로 변화는 화질을 모니터링하는 데는 한계가 있다. 주관적 평가를 실시하고자 할 때, 실험을 통해 얻어야 하는 항목에 대한 전반적인 이해와 이를 가장 잘 반영할 주관적 화질평가 방법론의 선정이 주관적 화질평가 방법의 핵심이라고 할 수 있다. 원본영상의 재생 유무와 비교 여부, 영상의 재생 횟수(single stimulus/double stimulus) 등을 잘 고려하여 주관적 화질평가 방법을 구성하는 것이 중요하다. 📖

참고문헌

- [3] ITU-T Recommendation P.710-4, Subjective assessment methods for image quality in high-definition television, 1998.
- [4] ITU-R Recommendation BT.1210-4, Test materials to be used in assessment of picture quality, 2012