

画質主観評価法小史

その2

三橋 哲雄

A short history of the subjective assessment method of picture quality(2)

MITSUHASHI Tetsuo

Abstract

ITU-R Recommendation BT-500 is the international standard of subjective assessment method of the TV picture quality. The psychological measurement is one of the most important parts in the method. In the first edition of BT.-500 published in 1974, the measurement occupies only one page in total 3.5pages. However, it rapidly increases to 10 pages in the 4th edition in 1990. Selection tables for the measurement were also added. In the 5th edition in 1992, problems on the assessment of digital pictures including program contents effects were described. In the 6th in 1994 and 7th in 1995, it is seen that the measurement of analog pictures is almost completed, and also seen that the important question is the psychological measurement of digital picture quality.

Key Word: Subjective assessment, Picture quality, Recommendation BT.500, Psychological measurement, Digital picture

[要約]

テレビジョン画質の主観評価法の国際的基準である国際電気通信連合無線通信部門（ITU-R）の勧告500の歴史の変遷を、心理学的測定法の面から調査した。1974年の初版500では心理学的測定法は全3.5ページのうち約1ページを占め、1982年の第2版でも同様であった。しかし、1986年の第3版では現用のDSIS法、DSCQS法などの測定法が掲載され、全9ページの内、約4ページへと急増した。さらに1990年の第4版では全ページ数16、心理学的測定法約10ページに増加し、手法選択のための表等も整備された。1992年の第5版ではデジタル画像評価に関する検討も追加された。1994年の第6版、1995年の第7版では記述の整理が行われ、アナログ画像の心理学的測定法はほぼ固まったこと、今後はプログラムの影響を含むデジタル画像の測定法が課題であること、が認められた。

キーワード：主観評価、画質、勧告BT-500、心理学的測定法、デジタル画像

まえがき

テレビジョン画像の画質評価法は関連する内外の学会、産業界諸団体から各種の手法が提案、使用されている。中でも国連の機関であるITU（International Telecommunication Union、国際電気通信連合無線通信部門）による勧告500は最も長い歴史を持ち、国際基準と見なされ、前記各種団体の手法に大きな影響を与えてきた。従って、画質評価法を考える場合、勧告500に関する考察が重要で基本となると考えられる。

勧告500の第1版は1974年に制定され、その後10度の改訂を経て、現在第11版に至っている。その歴史はテレビジョン画質の評価法の歴史であるとともに、そこにはデジタル化で代表される技術の進歩は当然として、ハイビジョン（High Definition Television、HDTV）に代表されるテレビジョンの社会的役割の変遷もまた影響していると考えられる。従って、勧告500の歴史的変遷について検討を進めることは、画質評価法のみならず、今後のテレビジョンの在り方を考察する一つの手がかりを与えるものと考えられる。

このような観点から、勧告500の歴史的変遷について検討を進めており、初版と11版を比較した結果を前回報告した⁽¹⁾。それによれば、記述の量的増加が著しく、デジタル技術の発展が影響していると考えられることが認められた。本稿では、より具体的な評価法の内容、中でも中心となる心理学的測定法について、その変遷を述べる。

1. 主観評価法と心理学的測定法

主観評価の構成を図1に示す。詳細は前報⁽¹⁾に記述されているので、そちらを参照頂きたい。中心となるのは心理学的測定法であり、調整法、単一刺激法、評定尺度法、系列範疇法、一対比較法、シェッフェの方法、等がある。具体的な勧告500の記述を検討するにあたり、以下これらの特徴を簡単に述べておく。詳細は文献（2）等を参照されたい。

調整法：与えられた状態に見えるように評定者が自分で対象を調整するので、手続きが簡単で満足度が高い。

単一刺激法：常に一つだけ提示された対象をカテゴリー（表1参照）で評価するので、所要時間が短く実施が容易である。

評定尺度法：カテゴリー（評定尺度と呼ぶ）に付された点数を回答させそれをそのまま心理尺度値（心理値）とするので定量的結果を容易に得られが、カテゴリー間の等間隔性の保証がないため尺度値間の四則演算の意味が不明である。

系列範疇法：回答カテゴリの分布から心理尺度を構成するため評定尺度の欠点はなく、応用範囲も広いが処理がやや面倒である。

一対比較法：評価対象を対にして相互の比較でyes、noなどを回答させるので判断が容易であるが、組み合わせが多く評定者の負担が大きくなる欠点がある。

シェッフェの方法：一対比較法と同様に比較であるが、違いの程度を回答させるもので、固有のデータ処理法が要求され、処理がやや面倒となる。

ところで、テレビ映像にかかわる心理学的測定法は以下の性質を持つことが望ましい。

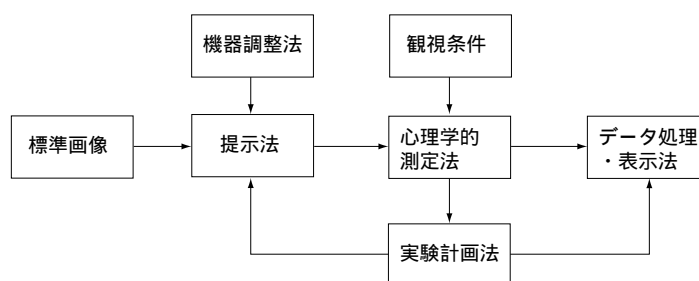
- (a) 評定者が容易に確な判断ができ、誤差が少ないこと
- (b) 結果の処理が容易で分かりやすいこと
- (c) 専門家でなくても容易に実施できること
- (d) パラメーターが独立に制御できない場合でも測定が可能なこと
- (e) 心理尺度構成ができること
- (f) 異種メディア間にわたる評価が可能なこと

これらのうち、(d) (e) は主として新しいシステムの検討を行うために必要であり、(f) はマルチメディアやカラーマネジメントに見られるように、メディアミックスを考えた場合に重要となる。近年のデジタル画像においては特に重要と考えられる。

心理学的測定法は、一対対比較法や恒常法のように刺激（本稿の場合は評価対象となる画像）の提示法やデータ処理法を含むものがあり、場合によっては観視条件まで含んでいるものもある。図1の機能ブロック「提示法」は、これら心理学的測定法固有の提示法を除いた画像の提示方法、すなわち画像の輝度、コントラストなどと、ディスプレイの調整法を含めた画質評価に固有の提示法を意味しており、後述するように、勧告500で規定されたいわゆる観視条件を意味する。また、データ処理・表示法ブロックはいわゆる統計処理にかかわる部分で、提示法同様、心理学的測定法固有のデータ処理を除く統計処理を意味している。

以下、上記を踏まえ、勧告500の心理学的測定法の変遷を検討する。

図1 主観評価の構成



2. 勧告500の主観評価法

本章では、まず基本となる勧告500における主観評価法について述べる。

2.1 勧告500の主観評価法

1974年制定された勧告500は全体で約3.5ページからなり、本文とANNEXからなる。本文は勧告の位置づけであり、主観評価法はANNEXに記載されている。ANNEXは「METHOD FOR LABORATORY ASSESSMENT OF PICTURE QUALITY」(画質の研究所における評価法)

と題されており、

1. Introduction
2. Details of the method

の2章からなる。

2章はさらに

- 2.1 Observers（評定者）
- 2.2 Grading Scales（評価尺度）
- 2.3 Test Pictures（試験画像）
- 2.4 Viewing Conditions（観視条件）
- 2.5 Presentation（提示）

の5節に分かれる。心理学的測定法は2.2節に述べられており、1ページ弱を占めている。

勧告における心理学的測定法の変化をみる意味で、基本となる上記2.2節の大意を原文に沿って下記2.2に記す。なお、我が国では、画像の場合に用いられる quality なる用語は、単なる「品質」と読まれることは少なく、「画質」と読み替えるのが一般的になっている。従って、「picture quality：画像品質」は「画質」と表すのが一般的であるが、本節では原文にできるだけ沿う意味で「品質、画像品質」を用いる。

2.2. 評価尺度

5段階尺度を表1に示す。問題の性質により、品質尺度（quality scale）または妨害尺度（impairment scale）を用いる。

表1 5段階尺度

5段階尺度	
品質	妨害
5 非常に良い	5 分からない
4 良い	4 分かるが気にならない
3 普通	3 やや気になる
2 悪い	2 気になる
1 非常に悪い	1 非常に気になる

局所的な実施条件を満たすため、実験を遂行する際に番号の順序や言葉につけられた番号を変えることは認められるが、結果は上記の形式で表示されるべきである。

使用された尺度、すなわち妨害または品質尺度、は、実験結果と常に一緒に表示しなければならない。

これまで6段階尺度による結果を述べた文書が多数に上っているという観点から、それらの実験結果を利用可能とするため、結果を変換する手段が望まれる。

ある尺度で得られた結果を別の尺度へ変換するに際しては、言語により生じるバイアスに

よる不確実性が生じる。しかし、6段階尺度を用いた実験結果で得られた評点A6を対応する5段階尺度の評点A5へ変換する下記の線形関係を、第一近似として用いることが出来る。

$$A5 = 5.8 - 0.8A6$$

ある種の実験の場合、品質や妨害尺度よりも比較尺度がより便利であり、その場合表2が勧告される。

表2 比較尺度

+3	非常に良い
+2	良い
+1	やや良い
+0	同じ
-1	やや悪い
-2	悪い
-3	非常に悪い

表1、2による評価は、カテゴリー（言語）に対応した1～5または+3～-3の評点で解答が求められ、これに基づく平均、標準偏差等で評価結果は表わされる。従って、心理学的測定法の分類からいえば評定尺度法である。

以上の内容は、500-1（1978制定）において6段階から5段階への変換を本文中から欄外へ移す、従来のANNEXをANNEX とする、等の若干の手直し以外に500-2（1982年制定）まで変化はない。しかし、1986年の第3版、500-3になると、大きな変化を見る。

3 勧告500-3の心理学的測定法～二重刺激法の登場～

1986年の勧告500-3になると、本文は従来と変わらないものの、ANNEX Iに「2.6 Choice of test procedure」（試験手続きの選択）が加えられ、さらにAPPENDIX Iが新たに付加された。従って、全体の量も500-2の3ページ強から500-3の9ページへと3倍増する。これは、後述するが、当時急速に進みつつあった映像のデジタル化とそれに伴う技術基準策定のための画質評価の重要性が、世界的に高まったことにある。

「2.6 Choice of test procedure」においては、異なる機関で実施された主観評価結果の相互比較を容易にすることその趣旨としており、掲載された手法を用いることによりその進展を期待している。特に注目されるのは、従来の表1、2に示した用語を用いる評価法に加え、単一刺激（Single-stimulus）法、二重刺激（Double stimulus）法などの新しい手法が実績のある主観評価法として「Table Preferred procedures and their applications」（望ましい手続きとその応用）に掲載されている点である。手法の概要は、新たに加えられた。「Appendix I CHOICE OF TEST PROCEDURE」中の第8章「Brief details of procedure」（手続きの詳細）中に述べられている。両手法とも提示法からデータ処理法までを含んでおり、現在まで広く用いられてい

る。その意味で、勧告500-3は勧告500の価値を高めた点で重要である。

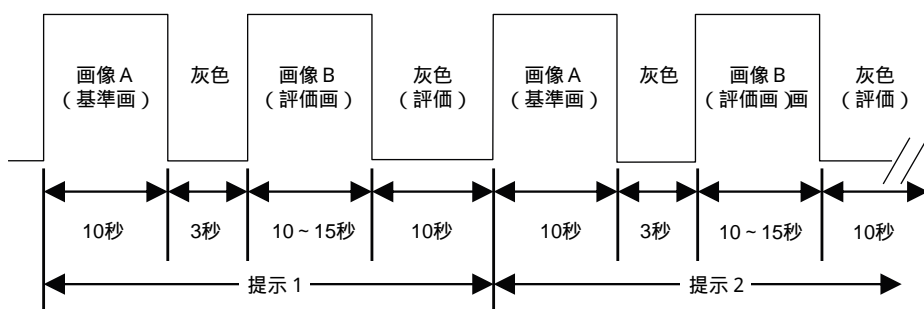
二重刺激法は、その特徴として、微妙な画質変化の評価に適しているとされている。そのため、当時進みつつあった映像のデジタル化における画質変化の評価に有効であることから、現在に至るまで世界的に広く用いられており、MPEG規格の策定に際しても画質評価法として用いられた。以下、簡単に内容を紹介する。

品質評価のための二重刺激法（二重刺激連続品質尺度法、Double Stimulus Continuous Quality-Scale、DSCQS）の評価用紙の例を図2（500-4より引用）に示す。27から30の数字は画像の番号である。アルファベットA、Bは評価したい処理を受けたかどうかを示す。即ちどちらかは常に原画である。例えば、27Aを27番の絵柄で画像処理を受けている画像とすると、27Bは同じ絵柄で処理を受けていない画像（原画）となる。評定者は「非常に良い」から「非常に悪い」までの言葉に対応する縦線の位置をマークする。A、Bどちらが原画であるかは知らせず、結果はAとBの差で表わす。心理学的測定法的に見れば、連続尺度を用いたグラフ法的一种と考えられる。結果は原画と処理画の差で表わすので、評価誤差が少なく微妙な差まで評価が可能とされる。図3（500-4より引用）に提示法を示す。

図2 二重刺激連続品質尺度（DSCQS）評価用紙の例

	27		28		29		30	
	A	B	A	B	A	B	A	B
非常によい								
よい								
普通								
悪い								
非常に悪い								

図3 DSCQSの提示法



4. 勧告 500-4 以後の心理学的測定法

第3版で大きく変化した勧告 500 は、1990 年の第4版でまた大きく変化する。本文が変化しただけでなく、ANNEX、APPENDIX とも記述内容のみならず構成が変更され分量も一層増加する。その理由としては、デジタル化の一層の進展だけでなく、本文中にも記述されているように、新たにハイビジョンに対する対応が世界的課題となったことが挙げられよう。画像提示条件など、より一層詳細な項目が盛り込まれ、量的にも一層の増加をみることになる。さらに、映像の利用分野が従来のテレビジョンのみならず、コンピュータやモバイル機器に代表されるエレクトロニクスの発展に伴い大きく拡大したことも、背景として大きいと考えられる。

4.1. 勧告 500-4 の心理学的測定法

1990 年制定された勧告 500-4 は15 ページと 500-3 と比べ大幅に増加している。まず本文において評価法として可能ならば二重刺激法が望ましいこと、評価試験における機器構成や使用画像、評定者、評価法など実験条件に関する情報を全て記述すること、が新たに付け加えられた。これにより、世界統一画質評価法の具体像を明らかにしたことが、本版の大きな特徴であると考えられる。なお HDTV の主観評価法については別に勧告が作成されており⁽⁴⁾、これについては別途述べることとする。

ANNEX についても本文に対応し、EBU (European Broadcasting Union、ヨーロッパ放送連合) で開発された二重刺激妨害尺度法 (DSIS Method、Double-Stimulus Impairment Quality-Scale method)、および二重刺激 (連続) 品質評価法 (DSCQS Method、Double-Stimulus Continuous-Scale method) にそれぞれ1章を割いて実験系統の構成や提示法、使用画像の性質などの実施法を詳細に記している。また、提示画像や周囲の明るさ、画像コントラストなど、提示条件や観視条件などについても詳細に記述しており、本勧告により実験が具体的に実施可能となるよう配慮されている。二重刺激法以外の以前の版に記載されてきた手法については Alternative method of assessment として記述されている。

APPENDIX も大幅に改定され、APPENDIX I FURTHER GUIDANCE ON THE METHODOL-

OGY、として上記二重刺激以外の手法含め、問題に応じて適切な手法を選択するためのガイドラインが記述され、選択表が掲載されている。表を下記3に示す。

以上のように、本版は単に心理学的測定法のみならず、映像の評価に関わる諸事項を網羅することを意図しており、これによる画質評価法の世界的統一を目指したものと考えられる。

表3 与えられた課題に対し手法を選択するためのガイド

	A	E	N	R	C
Overall quality	X		X	x	
Quality factor			x	x	
Failure Characteristics	x	x	x		
Degradation factor		x	x		
Comparison	x				x

A:二重刺激連続品質尺度法

E:二重刺激妨害尺度法

N: カテゴリー評価法

R: 比率尺度 / 量推定法

C:一対比較法

4.2. 勧告500-5～500-7における心理学的測定法

1992年、500-4をさらに24ページへと大幅に増量した改定版、500-5が制定された。500-4との大きな違いは、500-4がDSIS法とDSCQS法を中心としていたのに対し、500-5では(1) ANNEX1に主観評価法の情報があること、(2) ANNEX2にデジタル化に伴う妨害の主観評価法に関する情報がある、との二つの注を本文末に加えた点にある。また、これに対応し、ANNEX1の冒頭第一章には、上記表3をより詳細にした種々の評価対象に対する手法の選択表や評価用画像の選択基準、観視条件等が配置されている。二重刺激法についてもデータ処理に関する記述など、より詳細になり、実施を容易にしている。

また、ANNEX 1に対するAPPENDIX 1、2が新たに設けられ、デジタル化に際しての問題点を指摘し、画質劣化と画像内容の関係する程度を表すクリティカリティなる概念を提唱するとともに、後者ではデジタル画質劣化が番組内容と伝送条件の複合により生じることを述べ、その測定法について触れている。さらにANNEX2ではテレビジョン信号のデジタル化に際して生じる妨害の評価についての試みが述べられている。本版により従来のアナログ画像の画質評価における心理学的測定法は一定の段階に達し、今後の課題がデジタル画像の評価に適した手法の検討となることを示したものと言えよう。

1994年に500-5は改定され、500-6が制定された。全文は23ページとほぼ同様であり、本文に新たに先行する信号処理プロセスの重要性を示す項目が追加され、ANNEXも改定されたが、改定は試験画像や試験実施時の手続き等の追加と整理が主となっている。500-5で述べられたデジタル画像のクリティカリティや評価実験結果の解析、表示法等に関して新たな記述が加えられた点など、実際の主観評価試験実施に際してはより実用的になったとも言えよう。しかし、心理学的測定法に関しては記述の整理が主で新しい記述は追加されていない。また、

1995年制定の500-7もほぼ同様である。

1998年には500-8が制定された。全文43ページとこれまでで最高の分量であり、記述の整理のみならず、番君内容の影響を含めて時間的に連続な評価を行うための心理学的測定法としてSingle Stimulus Continuous Quality Evaluation (SSCQE)法が記述されており、デジタル画像の評価に対する具体的手法の検討が進められていたことが見受けられる。SSCQEの詳細については、別途予定の、500-8以降の勧告500の変遷に関する報告に譲る。

デジタル画像はテレビジョン以外の画像分野においても広く利用されており、新たな処理法、技術の開発が盛んである。従って、その評価に関しても、ITUのみならず関連諸機関、団体等でも検討が進められ、勧告500の位置づけも変化してゆく。これらについては、上記500-8以降の変遷についての報告時に合わせて報告する予定である。

5. あとがき

画質評価法の中心である心理学的測定法について、国際的基準となっているITUの勧告500における変遷を述べた。その結果、アナログ画像を対象とする心理学的測定法は500-5でほぼ一定の完成に達したことが分かった。しかし、1990年代に急速に進んだテレビジョンのデジタル化は画質評価法に対する新しい課題を生じ、勧告500にもその影響が認められた。

デジタル画像処理において広く用いられているMPEG2等の情報圧縮方式においては、動き情報の削減が柱の一つとなっている。従って、その画質評価に際しては、時間的要素を含んだ動画像を用いる評価が重要となり、対応する新しい心理学的測定法が必要となる。現在の最新版である500-11(2002年制定)では、この要求に対応した連続画質評価なる手法が記載されている。

また、ハイビジョンなど、より高度な表現力を持つ新しい映像システムの開発に伴い、新たな画質評価法の必要性も時を同じくして課題となった。ハイビジョンの画質評価に関しては勧告500とは別の勧告710⁽⁴⁾が扱っており、今回触れることができなかった。上記の時間要素を含む第8版以降の勧告500の変遷と合わせて、次の機会に報告することとしたい。

デジタル化に関わる画質評価に関しては、ITU-R以外の機関の関与も大きい。また、近年、感性や情緒などの重要性が言われており、ハイビジョンの画質評価との関連を含め、これらに対応できる主観評価法も重要な課題と考えられる。今後、勧告500の変遷にこれらがどのように関連してきたか、前記時間的要素と合わせ、調査を進めて行く予定である。

参考文献

- (1) 三橋哲雄：研究ノート・画質主観評価法小史(その1)、尚美学園大学芸術情報学部紀要、第7号、pp.29-39(2006)
- (2) 長谷川敬：心理評価技術、映情学誌、Vol.54, No.9、pp.1259-1264(2000)
- (3) RECOMMENDATION ITU-R BT.500 Method for the subjective assessment of the quality of

television pictures ~ 同 500-7

(4) RECOMMENDATION ITU-R BT.710-4 Subjective assessment methods for image quality in high-definition television