

ユーザ指向ストリーミングサービスのための 映像の主観品質と映像コンテンツの関係を利用した映像分類手法

加藤由花 宮島悠 箱崎勝也
電気通信大学 大学院情報システム学研究所
E-mail: yuka@is.uec.ac.jp

本稿では、映像の意味的内容を考慮したレート制御の実現を目指し、映像のコンテンツと人間が体感する主観品質の関係を明らかにする。ここでは、様々な種類の映像を用いた主観評価実験を行い、人間が共通に持っている品質に対する感覚を明らかにする。さらに、この結果を利用し、様々な映像を主観評価値の似ているグループに分類する手法を提案する。このグループを利用すると、グループ毎にレート制御の方法を変えることにより、ユーザ満足度の高い制御の実現が可能になる。本稿ではさらに、視覚的パラメータのみを考慮した場合の映像分類結果と、主観評価実験の結果得られた分類結果とを比較することにより、映像の意味的パラメータを考慮する必要性を示す。

A Video Classification Method Using Relationship between Video Contents and Subjective Video Quality for User-oriented Streaming Services

YUKA KATO YU MIYAJIMA KATSUYA HAKOZAKI
Graduate School of Information Systems, The University of Electro-Communications
E-mail: yuka@is.uec.ac.jp

This paper analyzes relationship between video contents and subjective video quality for user-oriented streaming services. In this analysis, we conduct subjective assessments by using various kinds of video programs, and propose a method classifying video programs into some groups in which a large majority of users feel the same video quality. The control with a high level of user satisfaction can be done by applying different control method to each group obtained from the proposed method. We also show the necessity of rate control according to video contents by comparing a classification result based on vision parameters with that based on the assessment result.

1 はじめに

近年、アクセスネットワークのプロードバンド化に伴い、インターネット上で映像配信サービスを実現したいという要求が高まってきている。ここでは、ユーザが多様な環境からサービスにアクセスして、映像が伝送されるネットワーク資源を多くのユーザが共有することから、高品質なストリーミングサービスを実現するためには、ネットワーク環境に適したレート制御が不可欠である。このとき、ユーザが体感する映像に対する主観品質は、視聴しているコンテンツやユーザの品質に対する嗜好によって大きく異なると言われている。レート制御においてもこれらを考慮した制御が必要であり、これまで、このような観点から様々な研究が行われてきた。例えば、コンテンツ制作者が映像の意味的重要度をメタデータとしてあらかじめ与えておくことにより、レート配分における優先制御を実現する方式が提案

されている。MPEG-4 コンテンツを対象に、コンテンツ内のオブジェクト毎に重要度の記述を行う手法 [1] や、コンテンツ制作者の主観に基づいて設定された意味的重要度を利用して空間方向と時間方向のレート制御を行う手法 [2] 等がある。これらのメタデータ記述には MPEG-7 [3] 等が利用される。また、シーンチェンジや映像の動きの激しさ等をシステムが自動で抽出し、その特徴に適したレート制御を行う方式が提案されており、例えば、H.263+ を対象としたトランスコーディング手法 [4] 等がある。前者の方式では、コンテンツの制作者によって映像の意味的重要度を指定することができるため、ユーザの主観品質の向上が期待できる。しかし、メタデータ記述にかかる手間は大きく、ライブ放送に対応できない等の問題が存在する。また、指定される重要度はコンテンツ制作者の主観に基づいた重要度であり、映像の内容に従って人間が一般に体感する品質とは必ずしも一致しない。一方、後者の方式では、システ

ムが自動でコンテンツの特徴を抽出するため、メタデータ記述の困難さを克服することができる。しかし、ここで対象としているコンテンツの特徴は、動きの激しさ等の視覚的要因であり、これらのコンテンツの意味的重要度を表現しているわけではない。

このような背景から、本稿では、映像の意味的内容を考慮したレート制御の実現を目指して、映像のコンテンツと人間が体感する主観品質の関係を明らかにする。ここでは、様々な種類の映像を用いた主観評価実験を行い、人間が共通に持っている品質に対する感覚を明らかにする。さらにこの結果を利用し、様々な映像を主観評価値の似ているグループに分類する手法を提案する。このグループを利用すると、グループ毎にレート制御の方法を変えることにより、ユーザ満足度の高い制御が可能になる。

2 前提条件

2.1 特徴記述パラメータ

本稿では、様々な特徴を持つ映像を、ユーザの主観品質という観点から分類する。このとき、映像の特徴をいくつかのパラメータにより記述しておき、このパラメータ値に従ってグループ化を行う。本稿では映像の内容を考慮した制御への適用が目的であるため、一般に映像の特徴を表現する視覚的なパラメータに加えて、映像の内容を表現するパラメータを定義する。以下に分析に用いるパラメータを示す。

- (1) 視覚的パラメータ：平均ビットレート、動きの激しさ、オブジェクトの数
- (2) 意味的パラメータ：ジャンル、映像の重要度、音声の重要度、情報の重要度、娯楽要素

2.2 品質パラメータ

分類された番組は、グループごとに同じ制御が適用されるので、本稿で制御の対象とする品質パラメータを規定する。一般に、ストリーミングサービスにおけるレート制御では、空間方向の情報量（1フレームあたりのビット量）と時間方向の情報量（フレームレート）のどちらか、または両方を変化させることにより、制御を実現する。ここで、ある一定の伝送帯域の下で情報量の削減を行う場合、この空間方向の情報量と時間方向の情報量はトレードオフの関係にある。そのため、ユーザの視聴に対する主観評価が最大になるように、これらの削減量を決定する必要がある。本稿では、ユーザの体感し得るパラメータとして、映像の画質、フレームサイズ（画面の大きさ）、フレームレートの3種類の品質パラメータを規定した。これらのパラメータは、ストリームのコーディング方式に依存しない一般的なパラメータである。

2.3 ユーザの品質に対する嗜好

次に、ユーザの品質に対する嗜好について定量的な定義を行う。一般に、視覚を含むどのような感覚にも、多少とも感情的調子（快-不快、好き-嫌いなど）が伴うことが知られており、例えば、感覚の認識と感情的側面の関係を表1のように序列付けした研究がなされている [5]。この表からは、視覚のような

表1 感覚の認識と感情的側面の関係

	焼灼痛	嗅覚	味覚	聴覚	視覚
Affection	10	8	6	4	2
Cognition	2	4	6	8	10

高等感覚に近くなるほど、知的側面が強くなり対象に規定される度合（客観性）は高まるが、それでもある程度の割合で感情的な面（主観性）を伴うことが読み取れる。そのため、視覚に関しては、基本的には多くの人が共通に持つ嗜好を対象としながら、1人1人の嗜好に配慮することにより、品質に対するユーザの満足度が高まると考えられる。

ここで、本稿では、前項で規定した3種類の品質パラメータを利用してユーザの嗜好を定義する。具体的には、画質、フレームサイズ、フレームレートそれぞれに対する重要度を w_q, w_s, w_f とし、嗜好ベクトル $p(u_i, v_i) = (w_q, w_s, w_f)$ をユーザの嗜好として定義する。ただし u_i はユーザの識別子、 v_i はコンテンツの識別子とする。嗜好ベクトルの各要素は各品質パラメータに対する重みとし、 $w_q + w_s + w_f = 1$ とする。定義からわかるように、この嗜好ベクトルは、ユーザ毎、コンテンツ毎に異なるが、多くのユーザが似通ったベクトルを持っているグループにコンテンツを分類することにより、コンテンツを考慮した制御が可能になる。

3 主観品質とコンテンツの関係

3.1 主観評価実験

3.1.1 実験方法

サービス品質に対するユーザの満足度を評価する方法としては、これまでSD法を用いた研究 [6] などが行われてきた。しかし、音声や映像を対象としたサービスでは、これらの評点をどのように付けるかという判断が容易ではなく、評価の精度が低くなるという問題点があった。このような事象の評価に対して信頼度の高いデータを得るためには、一対比較法が有効であることが確認されている [7]。そのため本稿でも、一対比較法を用いて主観評価実験を行うことにした。ここで、一対比較法とは、 n 個の対象を評定するとしたとき、 $(S_1, S_2), (S_1, S_3), \dots, (S_2, S_n), \dots$ のように2つの刺激の組み合わせを作り、この nC_2 通りの対に対して、どちらがどれだけ優れているかを被験者に評価させる方法である。適切に統計処理を行うことにより、ある刺激が他の全ての刺激に対してどの程度優れているかを決定することができる。比較判断の結果を順位で表すブラッドレーの方法やサーストンの方法と、結果を評点で表すシェッフェの方法の2種類の方法がある [8]。本稿では、比較の結果を点数で表し、その点数に適切な構造を仮定して分析を行うため、シェッフェの一対比較法を用いることにした。また、そのうち、一人の被験者に全組み合わせを受け持たせる中屋の変法 [8] を採用した。

具体的な実験手順を以下に示す。まず、テーマや

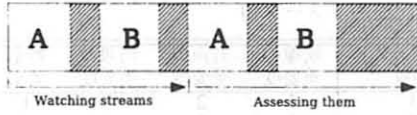


図1 比較の手順

表3 実験で用いた映像品質パラメータ値

	画質	フレームレート	フレームサイズ
基準値	100	30 fps	640×480
高品質	27	10 fps	384×256
中品質	15	3 fps	192×128
低品質	10	1 fps	96×94

スタイル等の番組の特性が異なる映像を30種類用意した。映像コンテンツの一覧を表2に示す。これらの映像は地上波テレビの様々な番組であり、それぞれ映像の長さが異なる。また、これらの映像を始めから終わりまで視聴し評価することは、被験者の大きな負担となる。そのため、本実験では、それぞれの映像の最も特徴的なシーン（ニュース番組ではアナウンサーが一つのニュースを読み終わるまで等）を10秒から20秒程度の映像シーケンスとして切りだし、試験映像を作成した。各試験映像は、2章で規定した記述パラメータにより特徴が記述される。表2には、これらのパラメータ値も合わせて示した。動きの激しさは1～5の5段階（動きが激しいほど値が大きい）で表現し、オブジェクトの数は1～3の3段階（数が多いほど値が大きい）で表現した。また、意味的パラメータにおける重要度は1～3の3段階（重要度が高いほど値が大きい）で表現した。

次に、各々の映像において、2章で規定した3種類の品質パラメータをそれぞれ3段階に変化させた映像を用意した。実験で用いた映像の品質パラメータ値を表3に示す。表中の映像の画質パラメータは、元画像を100としたときの映像の劣化度合い（映像に割り当てられるビット量）であり、実験に利用したエンコーダにおいて設定するパラメータである。

このようにして作成された試験映像の数は、各番組につきそれぞれ10種類となる。本実験では、ある1つのパラメータのみを変化させた2つの映像を一組とし、映像の主観品質としてどちらがどの程度望ましいかを被験者に評価させる。評価は-3～+3の7段階で（評点が低いほど主観評価値も低い）、30種類の番組毎に45組の映像に対して行われる。視聴条件はITU-R勧告BT.500に準拠するように整え、被験者とモニタの距離は60cm、実験室内の光源は1つ、視聴ポイントの照度は20LUXとした。被験者はPCに接続された19インチ液晶モニタ上で映像を視聴する。各映像はあらかじめ作成しておいた同一の映像を利用し、比較を行う映像シーケンスをA、Bとして、図1に示すように比較を行う。AB間には灰色のスクリーンを1秒間挿入し、映像の最後にも3秒間の灰色のスクリーンを挿入する。本実験では、男性6名、女性1名の計7名の被験者から、合計9450

表4 ある被験者の評価結果の例（番組1）

	O	Q1	Q2	Q3	S1	S2	S3	F1	F2	F3
O	-	0	0	-1	-2	-3	-3	0	0	0
Q1	0	-	0	-1	-2	-3	-3	0	0	0
Q2	0	0	-	0	-2	-3	-3	0	0	0
Q3	1	1	0	-	-1	-3	-3	1	1	1
S1	2	2	2	1	-	-2	-3	2	2	2
S2	3	3	3	3	2	-	-1	3	3	3
S3	3	3	3	3	3	1	-	3	3	3
F1	0	0	0	-1	-2	-3	-3	-	0	0
F2	0	0	0	-1	-2	-3	-3	0	-	0
F3	0	0	0	-1	-2	-3	-3	0	0	-
合計	9	9	8	2	-8	-22	-25	9	9	9

O：オリジナル映像

Q1：画質高、Q2：画質中、Q3：画質低

S1：サイズ高、S2：サイズ中、S3：サイズ低

F1：レート高、F2：レート中、F3：レート低

組のデータを取得した。

3.1.2 実験結果

実験結果は、10種類の品質パラメータを縦軸と横軸に取り、評点を要素に持つ表の形にまとめることができる。一例として、ある被験者の番組1に対する評価結果を表4に示す。例えば（S2, Q1）の評点が3であるということは、Q1の映像の方がS2の映像より評点3分だけ良いということの意味している。そのため、Q1の列の合計値によって、Q1の映像が他の映像より評点の合計でどれほど良いと判断されているかを数値化することができる。この例の場合、Q1の映像の評点の合計値は9である。このように番組毎に各パラメータに対する評点を算出した。本稿では、算出された評点の全被験者に対する平均値を各番組ごとの主観品質ベクトルとし、2章で定義した品質パラメータ値で表現する。例えば、番組1の主観評価ベクトルは、 $s(v_1) = (O, Q1, Q2, Q3, S1, S2, S3, F1, F2, F3) = (6.71, 7.00, 3.43, -2.43, -15.29, -24.29, 4.00, 6.86, 7.00)$ と表現される。

3.2 クラスタ分析

次に、前章で算出した主観評価ベクトルの要素を座標軸として10次元空間のユークリッド距離を求め、クラスタ分析を行う。ここでクラスタ分析とは、いくつかのサンプルの中から、それらの間にある何らかの類似度を手がかりに似たものを集め、いくつかの均質な集落（クラスタ）に分類する方法である。大別すると、結果として樹形図が得られる階層的方法と、あらかじめ決められたグループ数に各サンプルを分ける非階層的方法に分けられる。本稿では、最終的なグループ数が不明であるため、階層的方法を利用した。階層的手法では、N個の対象からなるデータが与えられたとき、1個の対象だけを含むN個のクラスタがある初期状態をまず作り、対象 x_1 と x_2 の間の距離 $D(x_1, x_2)$ からクラスタ間の距離 $D(C_1, C_2)$ を計算し、この距離が最も近い二つのクラスタを逐次的に併合していく。距離関数の違いにより、最短距離法、群平均法、ウォード法等が存在するが、本稿では、比較的まとまりの良いクラスタが幾つか得られるウォード法（各対象から、その対象を含むクラスタのセントロイドまでの距離の二乗の総和を最小化する方法）を利用する。図2にクラス

表2 実験に用いた試験映像とその特徴記述パラメータ値

No.	ジャンル	内容	平均 bps	動きの 激しさ	オブジェクト の数	映像の 重要度	音声の 重要度	情報の 重要度	娯楽 要素
1	ニュース	事件	985	1	1	1	2	3	1
2	ニュース	スポーツ	1245	4	3	2	2	3	1
3	ニュース	天気	1014	1	2	1	2	3	1
4	ニュース	一般	1110	1	2	2	2	3	1
5	ドキュメンタリ	壁画	1420	3	3	2	2	3	2
6	バラエティ	バラエティ	1309	3	2	1	2	1	3
7	バラエティ	お笑い	1397	4	2	2	2	1	3
8	バラエティ	クイズ	1283	2	1	1	2	1	3
9	バラエティ	トーク	975	3	3	2	3	1	3
10	ドラマ	ホームドラマ	1087	3	2	2	2	1	3
11	ドラマ	医療	942	2	2	2	2	1	3
12	ドラマ	海外	1020	3	2	2	2	2	3
13	ドラマ	時代劇	1373	5	1	2	2	1	3
14	アニメ	アクション	1063	4	1	3	1	1	3
15	アニメ	一般	1071	2	1	2	2	1	3
16	映画	アクション	987	4	1	3	2	1	3
17	映画	戦争	1181	5	1	3	2	1	3
18	映画	SF	855	1	2	3	2	1	3
19	映画	ミュージカル	779	3	1	3	3	1	3
20	スポーツ	野球	1253	4	2	3	1	1	3
21	スポーツ	マラソン	1331	5	1	2	2	1	3
22	スポーツ	ボクシング	1320	4	3	3	1	1	3
23	スポーツ	ラリー	1300	5	1	3	1	1	3
24	音楽	ポップス	1147	2	3	2	3	1	3
25	音楽	クラシック	962	3	2	1	3	1	2
26	音楽	演歌	1124	2	2	1	3	1	3
27	音楽	ビデオクリップ	1399	5	3	3	3	1	3
28	趣味	料理	1179	3	2	2	2	3	2
29	情報	健康	1179	2	2	2	2	3	2
30	教育	英会話	949	2	2	2	3	3	1

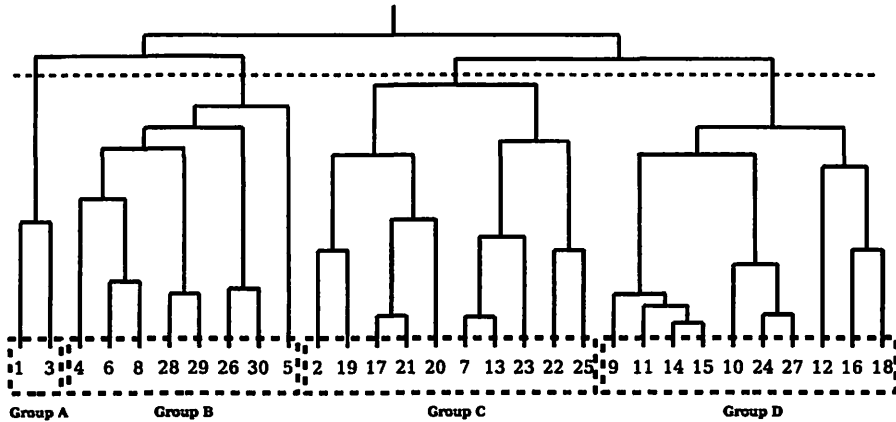


図2 クラスタ分析により得られた樹系図

タ分析により得られた樹系図を示す。

クラスタ数は多く取るほど細かい分析が可能になるが、クラスタ内の要素が少なくなるため要因の抽出が難しくなる。今回、制御の対象となる品質パラメータは3種類であることから3~5程度のグループに分類することを考え、図2中の点線の位置で分割し4つのグループを作成した。左からA、B、C、Dとする。各グループの特徴を以下に記す。

グループA：ニュース（事件、天気）が含まれる。情報の重要度が高く、娯楽要素が低い番組である。映像自体の重要度は高くない。

グループB：ニュース（一般）、ドキュメンタリ（壁画）、バラエティ（バラエティ、クイズ）、趣味（料

理）、情報（健康）、教育（英会話）が含まれる。バラエティ以外は、情報の重要度が高い番組である。映像、音声の重要度は、情報の取得に妨げにならない程度に重要である。バラエティの中でも動きが少ないものが含まれるのは、娯楽要素は高いが、映像や音声の重要度がそれほど高くないためと考えられる。

グループC：ニュース（スポーツ）、バラエティ（お笑い）、ドラマ（時代劇）、映画（戦争、ミュージカル）、スポーツ（野球、マラソン、ボクシング、ラリー）、音楽（クラシック）が含まれる。娯楽要素が高く、映像の重要度が高い番組である。動きの激しい番組が含まれ、スポーツは全て含まれている。

グループD：バラエティ（トーク）、ドラマ（ホーム

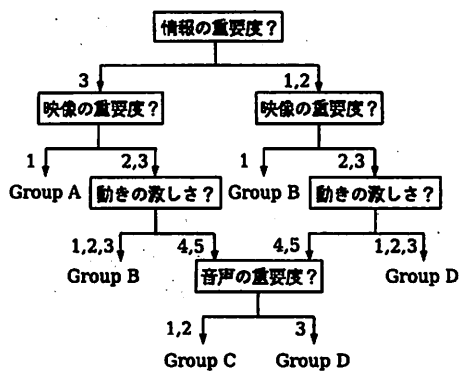


図3 番組の分類方法

ドラマ、医療、海外)、アニメ(アクション、一般)、映画(アクション、SF)、音楽(ポップス、ビデオクリップ)が含まれる。娯楽要素が高いが、映像自体の重要度はグループCほど高くない番組である。映像の重要度が比較的高い音楽番組もこのグループに含まれる。一部例外もあるが、通常のドラマや動きの重要度がそれほど高くない番組が含まれる。

以上の結果より、未知の番組をグループに分ける際の指針を導き出すことができる。ここでの考察対象は、情報の重要度、映像の重要度、音声の重要度、動きの激しさの4点である。分類処理の流れを図3に示す。アニメ(アクション)、映画(アクション、ミュージカル)、音楽(クラシック)の4番組以外は、このルールにより正しく分類が行われる。今回、映像の特徴は表2のように番組ごとに与えられていると仮定したが、これらの特徴はジャンルに従ってある程度自動的に付与することが可能である。

4 考察

本章では、映像の内容を考慮せず、視覚的なパラメータのみを利用して番組のグループ化を行った結果と、前章で示したユーザの主観品質に対する傾向の似ている番組をグループ化した結果とを比較する。ここでは、表2に示した平均bps、動きの激しさ、オブジェクトの数という3種類の視覚的なパラメータを用いて、前章と同様に30の番組に対してクラスタ分析を行った。その結果作成された4つのグループを以下に示す(樹形図は省略)。これは、映像の内容を考慮せずに、視覚的な特徴のみを利用して、似た傾向にある映像を分類した結果である。

グループ1: ニュース(スポーツ)、ドキュメンタリ(壁画)、バラエティ(バラエティ、お笑い、クイズ)、ドラマ(時代劇)、スポーツ(野球、マラソン、ボクシング、ラリー)、音楽(ビデオクリップ)。

グループ2: ニュース(事件、天気)、バラエティ(トーク)、ドラマ(医療、海外)、映画(アクション)、音楽(クラシック)、教育(英会話)。

グループ3: 映画(SF、ミュージカル)。

グループ4: ニュース(一般)、ドラマ(ホームドラ

マ)、アニメ(アクション、一般)、映画(戦争)、音楽(ポップス、演歌)、趣味(料理)、情報(健康)。

グループ1とグループCは、動きの激しい映像という特徴を持っていると考えられ、いくつかの共通点は見られるが、全体的な傾向として、主観評価値に基づくグループ分けの結果とは大きく異なることがわかる。つまり、視覚的なパラメータだけでは主観品質への影響を説明することができず、映像の内容を考慮した分析が必要であることがわかった。

また、視覚的なパラメータは時間ごとに変化していくパラメータであるが、意味的なパラメータは1つの番組の中では変化しないパラメータである点も考慮する必要がある。通常、時々刻々と変化していく視覚パラメータに追従しながら制御を行うことは困難であるとともに負荷の高い処理であるが、番組中変わらない意味的なパラメータを制御に用いることにより主観品質の向上が期待できるのであれば、より効率的な制御の実現が可能になると考えられる。

5 まとめ

本稿では、映像の意味的な内容を考慮したレート制御の実現を目指し、映像のコンテンツと人間が体感する主観品質の関係を、様々な種類の映像を用いた主観評価実験により明らかにした。さらに、この結果を利用して、映像を主観評価値の似ているグループに分類する手法を提案した。本稿ではまた、視覚的なパラメータのみを考慮した場合の映像分類結果と、主観評価実験の結果得られた分類結果とを比較し、映像の意味的なパラメータを考慮する必要性を示した。今後より精度の高い分析を行うために、被験者数を増やした実験を行う予定である。また、今回の実験では我々の主観によって各映像の特徴記述パラメータ値を決定したが、今後これらの値は客観データとして測定していく予定である。

参考文献

- [1] 森亮憲, 勝本道哲. ユーザ要求を反映する MPEG-4 配信 QoS プロトコルの提案. 情報処理学会 DPS ワークショップ論文集, pp. 13-18, 2003.
- [2] 村田嘉利, 増田彰久, 太田賢, 石原進, 水野忠則. モバイルマルチメディアストリーミングサービスのためのコンテンツ指向時空間的解像度制御方式. 情報論, Vol. 43, No. 12, pp. 3838-3847, 2002.
- [3] MPEG-7: Multimedia content description interface. 2002.
- [4] Z. Lei and N. Georganas. Rate Adaptation Transcoding for Pre-coded Video Streams. *ACM Multimedia '02*, pp. 127-136, 2002.
- [5] H. Pierson. *Feeling and emotion*. McGraw-Hill, 1950.
- [6] 山本健司, 伊藤嘉浩, 田坂修二. コンテンツがユーザレベル QoS に及ぼす影響の多次元評価. 信学技報, Vol. CQ2003-12, pp. 13-18, 2003.
- [7] 伊藤嘉浩, 田坂修二. 音声・ビデオ伝送におけるメディア同期を対象としたユーザレベル QoS の定量的評価及びマッピング. 信学論(B), Vol. J86-B, No. 3, pp. 485-498, 2003.
- [8] 日科技連官能検査委員会(編)(編). 官能検査ハンドブック. 日科技連出版社, 1973.