

TR-AC-0048

015

適応的QoS制御のための主観動画品質評価実験

夜久 健一（奈良先端大） 山崎 達也

2000.10. 4

ATR環境適応通信研究所

目 次

あらまし	1
1. 研究の背景	2
2. 主観動画品質評価実験	2
2.1 実験の目的	2
2.2 実験の方法	3
2.3 実験の結果と考察	5
3. まとめ	8
謝辞	9
参考文献	9

適応的 QoS 制御のための主観動画品質評価実験

夜久 健一[†] (keni-ya@is.aist-nara.ac.jp) 山崎 達也[‡]

[†]奈良先端科学技術大学院大学情報科学研究科

計算機アーキテクチャー講座

[‡](株)エイ・ティ・アール環境適応通信研究所

第一研究室

あらまし

マルチメディア通信において、ユーザの利用状況に応じて通信形態や QoS (Quality of Service) パラメータを適応的・自律的に制御することは、通信に対するユーザの利便性あるいは通信リソースの使用効率を高める上で重要な課題である。特にユーザレイヤにおいて汎用的なユーザ QoS モデルを構築することは、ユーザオリエンテッドな技術開発に必要不可欠であると考えられる。本レポートでは動画ストリームを対象に、アプリケーション QoS パラメータの相違がユーザの主観評価に与える影響を測定する実験を行った結果について報告する。ここで報告する実験データは、適応的 QoS 制御を行う上で必要な、ユーザレイヤにおけるユーザ QoS モデルの構築に資する基礎データとなるものと期待される。

Subjective Video Quality Measurement Experiments for Adaptive QoS Control

Kenichi Yaku[†] (keni-ya@is.aist-nara.ac.jp) Tatsuya Yamazaki[‡]

[†]Nara Institute of Science and Technology

Graduate School of Information

Computer Architecture Laboratory

[‡]ATR Adaptive Communications Research Laboratories

Department 1

Abstract

Around 1994-1995, advent of commercial ISP (Internet Service Provider) accelerated the Internet use explosively among people. However, since the current narrowband residential access infrastructure using dialup is too slow, new broadband Internet services are in process of deployment in network infrastructure to satisfy user's demand. This broadband Internet service has potential to deliver various types of multimedia data including video. Even though the network bandwidth becomes broader, the Internet traffic is naturally unpredictable so that QoS (Quality of Service) of multimedia data has to be controlled according to changing communication environment. This report provides experimental data that relate video QoS with user's satisfaction to conduce to the QoS control mechanism.

1. 研究の背景

近年ケーブルテレビによるインターネット接続やNTTのISDNによる定額常時接続サービスフレッツ、xDSLや無線LANなどの新技术を応用したサービスによって、一般家庭にも高速常時接続の基盤が整備され始めている。今後もこうした傾向は加速していくことが見込まれ、インターネットを通じて提供されるサービスも高速常時接続を視野に入れたものが増加している。このようなサービスとして現状で最も普及しているものの1つにネットワーク上での動画のストリーミング配信が挙げられる。しかし、現時点では高速なネットワークインフラというものはまだ発展途上の技術であり、様々な接続形態のネットワークを融合しているインターネットの特性や、今後高速ネットワークの整備が進んでもそれ以上に利用が増加する可能性があることを考えると、インフラ環境に応じて経路を通過するトラフィックを制御する技術というのは不可欠なものと考えられる。また、ユーザの使用する端末の性能も様々であり、必要以上の動画データを受け取っても処理できない場合もある。従ってネットワークや端末の利用可能なリソースに応じて、アプリケーションでサービス品質(Quality of Service: QoS)を変更することが今後必要とされ[1]、その際にユーザの満足度を如何に下げないようにQoS制御を行うかということに関する知見を得ることは重要と考えられる。

本研究では動画ストリームを扱うアプリケーションを対象に、アプリケーションQoSパラメータの相違がユーザの主観評価に与える影響について、主観評価実験により基礎データを取得した。本レポートにおいて実施した主観評価実験の概要と実験結果を述べ、それに関する考察を行う。

2. 主観動画品質評価実験

2.1 実験の目的

本実験の目的はネットワーク上で動画を配信する際に、意図的にコマ落ちを生成してデータを削減した場合、ユーザがどう感じるかを主観的に評価してもらうことである。ネットワークの帯域や端末のCPUリソースに制限がある場合、送受信端末間で転送するデータ量を制限以下に抑える工夫が必要である。動画を配信する仕組みを考えると、一定量のデータを持った画像を動きが再現できる間隔でユーザに届ける必要があるので、データ量を抑制するためには各画像のデータを小さくする方法と一定時間に送信する画像を減らす方法とがある。各画像データを圧縮するためには減色したり画像サイズを縮小したりする方法が考えられるが、今回は送信する画像を間引きしてデータ量を減らす方法に着目した。この場合間引く画像が増えるほど単位時間あたりのデータ量、つまりネットワークの帯域消費量は減少するが、再生される画像はコマ飛びした動きの粗いものになる。ユーザは動きの滑らかなものを嗜好すると仮定して、どの程度のコマ落ちに対してどの程度の不満を感じるものかを主観的に計測することが本実験の目的である。

2.2 実験の方法

主観評価の信頼性を高めデータの交流を図るためには、観視条件、表示画像の内容、表示法、評定者などの標準化が望ましいため、ITU-R から主観評価法全体に関する勧告[2]がなされている。そこで本実験では評価法として二重刺激連続品質尺度法 (Double Stimulus Continuous Quality Scale: DSCQS) [3]を採用し、可能な限り ITU-R の勧告に準拠するよう注意した。

実験システム[4]を図1に示す。システムは、画像ファイルを MPEG 圧縮して送信する送信端末 (Video sender) と中継端末 (Network emulator)、MPEG データを展開して動画再生する受信端末 (Video receiver) の3台のコンピュータを、100BASE-T のネットワークで接続した構成となっている。中継端末上には実際のインターネット環境に近い条件を設定するため、パケットの損失 (ロス) や遅延を模擬する機能も搭載されているが、本実験においては基礎データを収集することを第一としたため、中継端末は送信端末から受信端末へ単純にパケットを転送するルータ機能を提供する役割を行っている。本実験では MPEG-4 シンプルプロファイルを用い、目標伝送レート 64kbps と設定した。

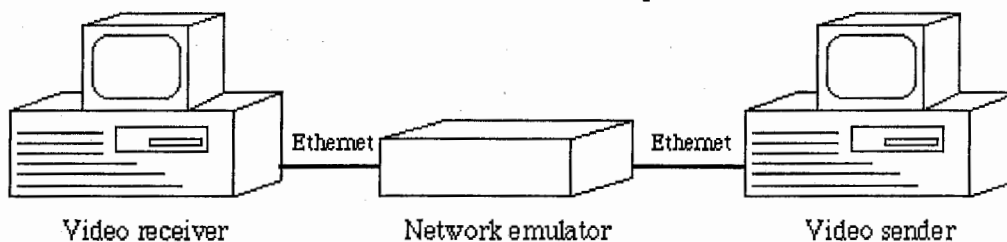


図1 実験システム

今回採用した DSCQS に関して簡単に説明する。DSCQS は基準画像と評価対象となる画像の2種類 (ここでは A、B と表す) を用意し、それらを順次被験者に提示することにより2つの画像の品質を、被験者に評価してもらう方法である。画像の提示は、A→インターバル1→B→インターバル2→A→インターバル3→B の順に行い、インターバル1とインターバル3は3秒間、インターバル2は10秒間に設定した。A、B どちらを基準画像とするかは被験者ごとにランダムに切り替え、被験者には知らせていない。この一連の画像の提示を試行と呼ぶ。被験者は1試行で提示された2種類の画像 A、B に対し、図2に示す評価用紙上の5段階の評価線上に主観的に感じた評価結果をマークすることにより評価を行う。送信端末には比較的動きの少ない風景画像 (以下 Flower) と、動きの激しいスポーツシーン (以下 Football) の画像を用意し、1秒あたりの再生画像枚数 (以下 fps) と画像サイズ (横ピクセル数×縦ピクセル数) を選択できるように構成した。

A、B の画像の組み合わせとしては表1に示す6通りの試行を行っている。なお各ユーザにはこれら6通りの試行の前に、Flower、Football 以外のサンプル画像を用いて実験方法と評価方法の説明を行った。また各試行の間にはおよそ2分半のインターバルを挟み、直前の試行における画像の印象が評価に強い影響を与えないように考慮した。

	1		2		3		4		5		6	
	A	B	A	B	A	B	A	B	A	B	A	B
非常に良い												
良い												
まあ良い												
悪い												
非常に悪い												

図 2 評価用紙

表 1 DSCQS における基準画像と対象画像の組み合わせ

	1 回目	2 回目	3 回目	4 回目	5 回目	6 回目
基準画像	Flower 10fps 360×240	Football 10fps 360×240	Flower 10fps 360×240	Football 10fps 360×240	Flower 10fps 180×120	Football 10fps 180×120
対象画像	Flower 6fps 360×240	Football 6fps 360×240	Flower 3fps 360×240	Football 3fps 360×240	Flower 3fps 180×120	Football 3fps 180×120

2.3 実験の結果と考察

表2 DSCQS 評価結果

No.	性別	経験度	年代	試行回と用いた画像											
				1-基	1-対	2-基	2-対	3-基	3-対	4-基	4-対	5-基	5-対	6-基	6-対
1	男	経験	30	37	34	32	32	42	38	40	25	57	45	37	22
2	男	経験	30	55	48	63	56	64	44	89	37	89	37	89	37
3	男	中間	40	50	50	70	50	80	29	80	25	75	13	72	12
4	男	経験	20	68	78	43	72	61	37	78	35	62	38	50	32
5	女	一般	20	67	44	23	31	77	46	76	51	83	76	73	56
6	女	一般	20	58	49	48	42	71	56	49	42	75	56	80	69
7	男	経験	30	57	55	33	23	46	32	63	42	52	45	29	25
8	女	一般	20	49	44	33	15	54	40	15	20	29	29	15	21
9	男	経験	20	43	64	56	56	78	17	48	4	83	37	63	0
10	男	経験	20	79	61	54	47	90	58	70	36	76	60	30	10
11	女	一般	30	59	42	52	39	59	31	45	31	51	31	26	34
12	女	一般	30	34	33	28	10	37	23	29	12	42	31	35	17
13	女	一般	20	60	80	60	50	90	40	80	40	90	71	80	60
14	男	経験	30	36	25	35	26	44	36	49	45	35	26	44	35
15	男	中間	40	40	36	39	36	44	33	45	31	40	30	39	29
16	女	一般	30	56	37	43	23	60	40	51	20	51	30	40	11
17	男	中間	40	63	69	75	71	45	39	73	52	43	29	53	44
18	男	経験	20	70	54	35	35	82	19	51	1	64	35	37	0
19	男	中間	20	37	31	11	11	49	43	12	16	51	42	23	26
20	男	経験	20	44	35	20	20	70	55	46	35	26	20	24	31
21	男	一般	40	65	55	68	55	69	55	76	56	77	83	90	80
22	男	経験	30	49	44	13	6	66	54	32	25	49	40	42	28
23	男	経験	30	63	42	53	36	70	23	63	33	77	28	77	23
24	男	中間	30	60	56	35	20	47	37	25	17	35	28	14	12
25	男	中間	20	63	49	49	37	68	38	60	32	78	60	25	13
26	男	中間	20	45	44	76	56	84	29	86	39	77	29	70	35
27	男	経験	20	36	43	46	36	44	37	46	33	42	31	36	27
28	男	一般	20	43	63	66	56	63	37	78	24	49	42	37	22
29	女	一般	30	68	68	44	54	85	15	35	5	83	6	37	2

前節で説明した実験方法に基づき、6通りの試行に対する評価をのべ29人のユーザにより行った。DSCQSの評価用紙として長さ10cmの直線を5段階に分割した尺度を用意しているので、各評定者がマークした評価を下端からのmm単位で測定して、それをそれぞれのフレームレートでの再生画像に対する評価とした。評価結果を表2に示す。表2中の「No.」は被験者番号を示し、「経験度」は実験と同時に行ったアンケートに基づき、被験者のインターネット上の映像視聴の経験を実験者が独自に三段階に区分けしたものである。「経験」はインターネット上の映像配信の仕組みをある程度理解し、その視聴の経験のあるレベル、「一般」は映像配信の仕組みの知識も視聴の経験もほとんどないレベル、「中間」は両者の中間に位置付けされるレベルとしてある。また「試行回と用いた画像」の欄の表記で「x-基」および「x-対」は、それぞれ表1中の「x回目の基準画像」「x回目の対象画像」を用いたことを示す。表2中の評価の数値が高いほど再生画像に対する満足度が高いことを示し、50の値は図2の評価用紙の「まあ良い」に相当する。

個人差により評価値をそのまま評定することはできないので、各試行回において基準画像と対象画像の間の評価値の差を求め、アプリケーション QoS の違いがユーザの満足度に如何に影響するかを検定する。評価値の差（基準画像－対象画像）を表3に示し、表3から求められた統計量を表4に示す。また各試行回において性別、経験度別の平均値を表5に示す。

実験の目的である、コマ落ちに対するユーザの主観評価の差を、表4の試行回1と試行回3を比較することによって検討する。試行回1と3はともに同じ画像サイズの Flower を用いているが、試行回1では基準画像と対象画像のfpsの差が4であるのに対し、試行回3ではfpsの差が7である。この表4における平均値の試行回1と3の間の差は、このfpsの差を反映しているものと考えられる。この差にどれだけ有意差があるかを、両試行におけるサンプルが標準偏差が等しいがその値は未知である別々の正規分布から独立に得られたものと仮定し、両者の平均が等しいという仮説をt分布に基づき片側検定した[5],[6]。その結果有意水準0.05%でこの仮説は棄却され、試行回3における評価差は試行回1よりも大きいということが有意水準0.05%のもとで言えるという結果が得られた。また上記の標準偏差が等しいという仮定が成り立たず、2つの正規分布の標準偏差が全く未知である場合は Behrens-Fisher(バーレンス-フィッシャー)問題として知られている[7],[8]。この問題に対する1つの近似検定法として Welch の検定[8]により片側検定を行った結果、この場合も同じく有意水準0.05%で上記の仮説は棄却された。さらに試行回2と4に関しても同様にt分布に基づく片側検定と Welch の(片側)検定を行った結果、有意水準0.05%で試行回4における評価差は試行回2よりも大きいという結果が得られた。すなわち動画像 Flower および Football の両者に対して、10fpsから6fpsに再生画像枚数を変化させた時よりも、10fpsから3fpsに再生画像枚数を変化させた時の方が被験者の評価が落ちることが、本実験の一つの結論として得られた。

表3 基準画像と対象画像の評価値の差

No.	試行回					
	1	2	3	4	5	6
1	3	0	4	15	12	15
2	7	7	20	52	52	52
3	0	20	51	55	62	60
4	-10	-29	24	43	24	18
5	23	-8	31	25	7	17
6	9	6	15	7	19	11
7	2	10	14	21	7	4
8	5	18	14	-5	0	-6
9	-21	0	61	44	46	63
10	18	7	32	34	16	20
11	17	13	28	14	20	-8
12	1	18	14	17	11	18
13	-20	10	50	40	19	20
14	11	9	8	4	9	9
15	4	3	11	14	10	10
16	19	20	20	31	21	29
17	-6	4	6	21	14	9
18	16	0	63	50	29	37
19	6	0	6	-4	9	-3
20	9	0	15	11	6	-7
21	10	13	14	20	-6	10
22	5	7	12	7	9	14
23	21	17	47	30	49	54
24	4	15	10	8	7	2
25	14	12	30	28	18	12
26	1	20	55	47	48	35
27	-7	10	7	13	11	9
28	-20	10	26	54	7	15
29	0	-10	70	30	77	35

表4 基準画像と対象画像の評価値の差の統計量

	試行回					
	1	2	3	4	5	6
平均	4.17	6.97	26.14	25.03	21.14	19.10
分散	133.87	107.96	366.67	296.10	382.53	363.27
標準偏差	11.77	10.57	19.49	17.51	19.90	19.40
最大値	23	20	70	55	77	63
最小値	-21	-29	4	-5	-6	-8
中央値	5	9	20	21	14	15

表5 性別および経験度別の評価値の差の平均値

	試行回					
	1	2	3	4	5	6
平均(男性)	3.19	6.43	24.57	27.00	20.90	20.86
平均(女性)	6.75	8.38	30.25	19.88	21.75	14.50
平均(経験)	4.50	3.17	25.58	27.00	22.50	24.00
平均(中間)	3.29	10.57	24.14	24.14	24.00	17.86
平均(一般)	3.94	9.65	26.53	23.65	20.18	15.65

また各試行回において被験者を男性と女性の2グループに分け、2グループ間の平均値を比較検定したが、t分布に基づく片側検定ならびに Welch の(片側)検定の双方において有意水準15%未満では2グループ間の平均値に大小関係があるとは言えなかった。同様に各試行回において経験度の経験と一般の2グループについても、平均値に関してt分布に基づく片側検定ならびに Welch の(片側)検定を行ったが、有意水準10%未満では2グループ間の平均値に大小関係があるとは言えなかった。経験度に関してはグループ分けする基準自体がアンケートに基づいたものであり、正確なグループ分けの線引きができていなかったことも検定による有意差が出なかった一因と考えられる。

3. まとめ

動画のフレームレートとユーザの評価に関する実験を、ITU-R の勧告に準拠して DSCQS にて行った。結果として2種類の動画像に対し、非常に高い確度で 10fps と 6fps の評価差より 10fps と 3fps の評価差が大きくなることが本実験により明らかになった。一方でその不満の度合いは、画像内容や低下の幅、ユーザ側の様々な要因などに左右され一定ではないという知見が得られた。なお本実験ではもともとの画像サイズの一部を隠蔽す

ることにより画像サイズを変化させているため、実験結果の試行回3と5（あるいは4と6）の間での画像サイズの差による評価差の検定は行わなかった。これらの要因とユーザの満足度との関連を引き続き調査することで、環境に応じた、しかしユーザの満足度を最大に保つ QoS の制御が可能であると考えられる。

謝辞

日頃ご指導頂く、(株)エイ・ティ・アール環境適応通信研究所小宮山牧児社長、蓮池和夫第一研究室室長に感謝致します。また本研究遂行にあたり MPEG-4 符号化復号化器の部分でご協力頂きました沖電気工業(株)の関係各位に深謝致します。

参考文献

- [1] 小菅昌克, 山崎達也, 荻野長生, 松田潤, "マルチエージェントによる適応的 QoS 制御方式," 信学論 B, vol.J82-B, no.5, pp.702-710, May 1999.
- [2] Recommendation ITU-R BT. 500-8, Methodology for Subjective Assessment of the Quality of Television, 1998.
- [3] 熊田純二, "画質評価の最近の動向," 映像情報メディア学会誌, vol.53, no.9, pp.1184-1189, 1999.
- [4] 酒井靖夫, 山崎達也, 蓮池和夫, "ユーザ QoS 測定のための MPEG-4 動画像伝送システム," B-11-8, p.405, 2000.
- [5] Donald F. Morrison, MULTIVARIATE STATISTICAL METHODS, Third Ed., McGraw-Hill, New York, 1990.
- [6] 立花俊一, 田川正賢, 成田清正, エクササイズ確率・統計, 共立出版, 東京, 1996.
- [7] Sheldon M. Ross, INTRODUCTION TO PROBABILITY AND STATISTICS FOR ENGINEERS AND, John Wiley & Sons, New York, 1987.
- [8] 竹内 啓, 確率分布と統計解析, 日本規格協会, 東京, 1979.