

IP ネットワーク上でのマルチメディア遠隔学習の
QoE 評価に関する研究

2011年

深 谷 和 義

目次

第1章	序論	1
1.1	遠隔学習	2
1.1.1	オンデマンド型遠隔学習	3
1.1.2	リアルタイム型遠隔学習	3
1.1.3	本論文で対象とする遠隔学習	4
1.2	IP ネットワーク	4
1.2.1	ネットワーク構造	4
1.2.2	マルチメディア	5
1.2.3	メディア出力制御方式	7
1.3	QoS	9
1.3.1	定義	9
1.3.2	QoS パラメータ	10
1.4	QoE	12
1.4.1	QoE 尺度及び評価方法	12
1.4.2	系列カテゴリー法	15
1.4.3	QoS マッピング	18
1.5	遠隔学習における QoE	18
1.5.1	主観品質	18
1.5.2	正答率	19
1.5.3	返答時間	20
1.6	本論文の目的と構成	21
第2章	音声・ビデオベースの遠隔学習において文字情報が学習効果に及ぼす影響	23
2.1	まえがき	23
2.2	評価実験	24
2.2.1	情報メディア	24
2.2.2	コンテンツ	25
2.2.3	実験システム	26
2.2.4	アプリケーションレベル QoS パラメータ	27
2.2.5	学習効果の評価方法	28
2.3	実験結果と考察	29
2.3.1	アプリケーションレベル QoS 評価結果	29
2.3.2	学習効果の評価結果	29
2.3.3	QoS マッピング結果	33

2.4	むすび	36
第3章	文字情報ベースの遠隔学習において音声・ビデオ品質が学習効果に及ぼす影響	39
3.1	まえがき	39
3.2	遠隔学習実験	40
3.2.1	実験システム	40
3.2.2	コンテンツ	41
3.2.3	学習効果の評価	45
3.3	実験結果と考察	46
3.3.1	正答率の推定値	46
3.3.2	学習効果の評価結果	47
3.4	むすび	49
第4章	多地点 IP 通信によるリアルタイム型遠隔対話環境における QoE	51
4.1	まえがき	51
4.2	遠隔対話環境実験	53
4.2.1	対話内容	53
4.2.2	実験システム	53
4.2.3	評価方法	55
4.3	実験結果と考察	56
4.3.1	主観品質の測定結果	56
4.3.2	返答時間の測定結果	58
4.4	むすび	59
第5章	結論	71
	謝辞	75
	参考文献	77
	発表論文一覧	83

第1章 序論

遠隔学習は、講師と受講者が時間、場所もしくはその両者によって分離された状況での学習を可能にするので、これまでも、教育機関や企業などにおいて盛んに行われている。例えば、印刷メディアを受講者に配布する通信教育は従来からある遠隔学習の代表的なものの一つである。また、テレビやラジオといった放送メディアによる遠隔学習も広く利用されている。

インターネットの普及により、遠隔学習にインターネットを活用する事例が急増している [1]。また、アクセス網の高速化、端末の高性能化により遠隔学習の教材に音声・ビデオ等のマルチメディアを利用できるようになっている。文献 [2] によると、2009年12月1日現在で、大学の学部や研究科におけるインターネット等を活用した遠隔教育の実施率が36.7%にも昇っている。これは、前年比で14.1ポイントの増加である。

インターネットを利用した遠隔学習はオンデマンド型とリアルタイム型の二つのタイプに分類できる [3][4]。オンデマンド型は、受講者が自分のスケジュールで受講する教材（例えば、記録されたビデオ）にどこからでもアクセスできるオンライン配達方式である。受講者同士が同時に一緒にいる必要はない。また、一人の受講者が同じ講義を繰り返し見ることができる。リアルタイム型は、すべての受講者が異なる場所で同時に“出席する”オンライン学習である。このタイプは一般にテレビ会議システムを使い、講師とすべての受講者がスケジュールを合わせて講義に参加する必要がある。

インターネットを利用した遠隔学習が普及したことにより、通信制大学において単位認定される授業として使われるようになった。これは、平成10年度の大学設置基準の改正で認められ、更に、平成13年3月の改正では、“メディアを利用した授業”だけで卒業要件である124単位すべてを取得しても良くなっている [5]。これにより、八洲学園大学やサイバー大学のように、すべての授業をインターネット授業でおこなえる通信課程大学が出来始めている。

遠隔学習における、“メディアを利用した授業”では、全体として面接授業と同等な教育効果があると判断される形態であることが条件とされている [5]。このため、文字、音声、静止画、動画等を用いることで、面接授業により近い環境を確保する必要がある。なお、当初はリアルタイム型しか認められていなかったが、平成13年の改正で、必ずしもリアルタイム型の特徴である同時性・双方向性が確保されていなくても認められるようになった。

しかしながら、インターネット等のIPネットワークは、基本的にベストエフォート型のネットワークであり、通信時に、パケットの欠落、遅延及び遅延揺らぎが生じる。それらは、音声・ビデオ等のメディアの時間的・空間的構造を乱し、その出力品質を劣化させる [6]。その結果、インターネット上での音声・ビデオ伝送におけるQoS（サービ

品質, Quality of Service) は保証されない。そのため, Web アクセスなどの他のトラフィックと混在すると, QoS の劣化に敏感である音声・ビデオストリームの QoE (ユーザ体感品質, Quality of Experience) [7] は容易に劣化する。

実際にネットワークを利用するユーザの QoE には, 様々なものが考えられる。文献 [8] では, QoE を決める属性として, “ 作業効果 (Task effectiveness) ”, “ 作業効率 (Task efficiency) ”, “ ユーザの満足度 (User satisfaction) ”, “ ユーザの快適度 (User enjoyment) ” が挙げられている。遠隔学習においても, 学習による成績の向上, 学習者の学習意欲など, 幅広く QoE を評価することが望まれる。

本論文においては, 2 種類の遠隔学習システムを扱う。一つは学習者がクライアント端末からサーバに蓄積された教材を用いて学習するクライアントサーバシステムである。もう一つは, 一台の講師用端末と複数台の受講者用端末がリアルタイム接続しているテレビ会議システムである。便宜上, 前者を“ クライアントサーバ遠隔学習システム ”, 後者を“ 多地点接続リアルタイム遠隔学習システム ”と呼ぶ。クライアントサーバ遠隔学習システムではオンデマンド型の遠隔学習を扱い, 多地点接続リアルタイム遠隔学習システムではリアルタイム型の遠隔学習を扱う。

本論文では, マルチメディア教材を使用した遠隔学習システムを多面的な QoE で評価する。まず, クライアントサーバ遠隔学習システムを用いて学習した後の成績を学習効果として定量的に調べている。そして, 通信品質が劣化した場合にでも, できるだけ学習効果を高く維持するための方策を検討する。また, 学習者が多地点接続リアルタイム遠隔学習システムを使用して感じる主観品質を扱っている。更に, 同システムにおける返答時間の違いによる学習効率を調べている。そして, 受信バッファリング制御が主観品質と返答時間に及ぼす影響を調べている。

以下では, まず, 本論文に関する議論の前提となる要素技術を説明する。1.1 節では, 遠隔学習の概要を説明する。次に, 1.2 節において, IP ネットワークに触れ, 1.3 節において, QoS を説明する。そして, 1.4 節で QoE を概説し, 1.5 節では, 遠隔学習における QoE を紹介する。最後に, 1.6 節にて本論文の目的と構成を示す。

1.1 遠隔学習

e-Learning と呼ばれることが多い遠隔学習では, 音声・ビデオなどのマルチメディアを用いて, 自宅等の好きな場所で学習可能である。近年は, FTTH (Fiber To The Home) [9] に代表される高速なインターネット環境の発展により, マルチメディア通信を容易に低コストで行えるようになったため, 遠隔学習の需要が増加している。そのため, AACE (Association for the Advancement of Computing in Education) が母体となる ED-MEDIA (World Conference on Educational Multimedia, Hypermedia & Telecommunications) や E-Learn (World Conference on E-Learning) をはじめとして, 遠隔学習に関する研究が世界中で積極的に行われている。

遠隔学習は, 大きく分けてオンデマンド型遠隔学習とリアルタイム型遠隔学習の二つある [10][11]。オンデマンド型遠隔学習は, 事前に収録・編集した講義を教材サーバに蓄積しておき, 学習者が望む時間にそれを視聴するものである。基本的に学習者の意

志で個別学習を行うため、自立学習型遠隔教育 [11] と呼ばれる。一方、リアルタイム型遠隔学習は、テレビ会議システムを用いて、講義者が講義を行っているのと同じ時刻に、遠隔地にいる受講者が学習するものである。これは、通常、複数の学習者が離れた場所で同時に学習するため、仮想教室型遠隔教育 [11] と呼ばれる。以下に、それぞれの遠隔学習の特徴を示す。

1.1.1 オンデマンド型遠隔学習

オンデマンド型遠隔学習には、次のようなメリットとデメリットがある [10][11]。

- メリット

- 学習者の都合だけで、時間的制約なしにいつでも学習することができる。
- 同じ講義を何度でも繰り返し学習することができる。
- 講義者が一度収録した講義を複数年使える。

- デメリット

- 学習上のつまずきなどによって、学習停滞や学習に対するモチベーションの低下が起こりやすい。
- 収録する講義を事前に撮影・編集する必要がある。
- グループ学習やディスカッションなど個別学習以外の学習形態には適さない。

1.1.2 リアルタイム型遠隔学習

リアルタイム型遠隔学習には、次のようなメリットとデメリットがある [10][11]。

- メリット

- 講義者が学習者の反応を見ながら講義を行うため、対面学習に近い環境が実現できる。
- 学習者が講義者に質問したり、学習者同士で議論したりといった双方向コミュニケーションが可能である。
- 学習者は決められた時間に講義を受講することになるので、学習のペースを作りやすく、学習者のモチベーションの低下が起こりにくい。
- 討論型講義、実験、演習、実習など様々な講義形態に対応できる。

- デメリット

- 講義者、学習者に時間的制約がある。
- オンデマンド型遠隔学習に比べて設備費用がかかる。

表 1.1: OSI 参照モデルとインターネットの階層構造

OSI	インターネット
アプリケーション層	アプリケーション層 (HTTP, SMTP, ...)
プレゼンテーション層	
セッション層	
トランスポート層	トランスポート層 (TCP, UDP)
ネットワーク層	ネットワーク層 (IP)
データリンク層	データリンク層 (ノード)
物理層	物理層 (デバイス)

1.1.3 本論文で対象とする遠隔学習

本論文では、2種類の遠隔学習を扱う。一つは、学習者の要求に応じて、教材サーバが学習端末へ教材コンテンツ配信するオンデマンド型遠隔学習である。これは、学習端末と教材サーバが1対1接続している形態を扱う。もう一つは、一台の講師用端末と複数台の受講者用端末が接続しているリアルタイム型遠隔学習である。これは、講師用端末と受講者用端末とがテレビ会議システムを利用して1対多及び多対多接続している形態を扱う。

1.2 IP ネットワーク

1.2.1 ネットワーク構造

IP ネットワークとは、IP (Internet Protocol) [12] を用いて相互に接続されたネットワークのことである。現在普及しているインターネットは、IP ネットワークが世界規模で相互に接続された巨大なネットワークのことを意味している。IP ネットワークの普及によって、遠隔学習においても IP ネットワークを用いることが多くなった。

ISO (国際標準化機構, International Organization for Standardization) と ITU-T (国際電気通信連合電気通信標準化部門, International Telecommunication Union - Telecommunication standardization sector) が定めた国際標準ネットワークアーキテクチャである OSI (Open Systems Interconnection) では、ネットワークの階層構造が定義されている。これを OSI 参照モデルと呼ぶ。一方、インターネットでは、IETF (Internet Engineering Task Force) が多くの RFC (Request For Comments) として階層構造を規定している。表 1.1 に、OSI 参照モデルとインターネットの階層構造を対応させたものを示す。表 1.1 で示すように、IP とは、OSI 基本参照モデルに照らし合わせた場合のネットワーク層に相当するプロトコルである。また、インターネットの階層構造は、OSI 参照モデルの上位3層が、アプリケーション層として考えられている。

IP はコネクションレス型であり、通信を行う前に通信相手との間にコネクションの確立をすることがなく、バッファオーバーフローやデータ誤りがあった場合にも、再送

を行わない。このようなサービスはベストエフォート型サービスと呼ばれ、QoSを保証しない。特に、遠隔講義で扱う音声やビデオは時間構造を持つため、通信品質の劣化の影響を受けやすい。なお、QoSについては1.3節で説明する。

1.2.2 マルチメディア

マルチメディアとは、文字（テキスト）、音声、画像（静止画）、ビデオ（動画）など様々な形態の情報を、コンピュータ上で統合して扱うことである。ここでは、本論文で扱っている文字、音声及びビデオについて述べる。

文字

学習者にとって、文字を読むことによる学習の機会が多い。このことは、面接授業や遠隔授業のいずれにおいても板書型講義が多く行われていることからわかる [13]。

これまでに文字情報が学習者の成績に影響を及ぼすことが広く認められている [14][15]。文字情報による効果として次の二つが挙げられる。

- 補強効果
聴覚に入った音声情報を文字情報として視覚で確認することで補強される。
- オーガナイザー効果
見出しが後続する文字等の情報に対する読み手の理解や記憶に影響を与える [16]。

また、文字情報が与える効果として、次のことが挙げられる [17]。

- 文字と映像が提示された場合は文字に注意を向けやすい。
- 映像に文字を付加することによって注意の分割を強いられ、獲得されるべき情報量が減少する。
- 映像に文字を付加することによって情報量が増加し、かつ刺激が次々に変化するため、二つの情報の十分な統合が図れない。
- 映像に文字を付加することによって多重の処理を強いられ、心理的余裕が減少し、提示された情報が十分に理解できなくなる。

したがって、文字情報が学習者に与える影響は、良い効果が生じる場合と悪い効果が生じる場合の両方の可能性がある。

音声

音声やビデオのデータ量は文字のものに比べて非常に多いが、時間的及び空間的に冗長度が高いため、その冗長部分がある程度削除しても主観的な品質には影響がない。したがって、有限であるネットワーク資源を考慮して、音声・ビデオは圧縮して転送することが望ましい。一方で、近年のネットワークのブロードバンド化に伴い、音声のように必要帯域幅が狭いメディアは、圧縮をせず符号化のみで伝送することも可能となった。

本論文の実験では、音声の符号化方式として、無圧縮であるリニアPCM (Linear Pulse Code Modulation) を用いる。これは、信号を一定時間ごとに数値化して記録する PCM 方式の一種で、サンプリングの間隔が一定で得られたデータを圧縮しない方式である。記録されたデジタルデータの品質は、サンプリング周波数と量子化ビット数で決まる。音楽 CD がこの方式を利用している。

ビデオ

ビデオのデータ量は音声のそれと比べて格段に多いため、大幅に圧縮することが望まれる。しかし、圧縮率を高めるためには、符号化・復号化が複雑になり、処理に多くの時間を必要とする。ビデオの符号化・復号化は送信側・受信側それぞれでリアルタイムに行うことになるため、高性能な機器を必要とする。

本論文の実験では、ビデオの圧縮符号化方式として、H.261 あるいは MPEG-1 を用いる。

H.261 は CCITT (現在の ITU-T) によって策定された動画圧縮の規格で、1990 年に勧告として承認された。デジタル動画像の圧縮符号化方式としては、フレーム間予測、離散コサイン変換 (DCT)、量子化、エントロピー符号化を組み合わせ用いた世界最初の国際標準であり、その後規格化された H.263, H.264, MPEG-1, MPEG-2, MPEG-4 など数多くの動画像圧縮方式の基礎としてとらえることができる。ISDN 上でのテレビ会議に利用することが想定されているため、符号化ビットレートは 64kbps ~ 1.92Mbps の間で 64kbps 刻みで指定することが規格上定められている。

MPEG には、MPEG-1, MPEG-2, MPEG-4 などいくつかの規格がある。その中で、MPEG-1 は 1.5Mbps 程度の比較的低ビットレートの動画像を対象とした符号化方式である。CD, ハードディスクなどのデジタル記録媒体に動画情報を圧縮して蓄積するための国際標準である [18][19]。

MPEG の符号化は、空間的・時間的冗長性を削減することにより行われる。空間的冗長性を削減するために、直交変換を用いており、時間的冗長性を削減するために、予測符号化を用いている。MPEG では、次に示すように、I ピクチャ、P ピクチャ及び B ピクチャという 3 タイプを用いて双方向予測を実現している。なお、MPEG で定められている動画の単位構造を GOP (Group of Pictures) という。

- I (Intra) ピクチャ

その情報だけから符号化された画像で、フレーム間予測を使わずに生成される。

GOP 構造内には、ランダム・アクセスのために最低 1 枚の I ピクチャが必要である。データ容量は他のピクチャよりも大きい。

- P (Predictive) ピクチャ

直前の I ピクチャまたは P ピクチャからの予測を行うことによってできる画像である。一般的に P ピクチャは、イントラ・フレーム (フレーム内予測画像) と、フォワード・インターフレーム (順方向フレーム間予測画像) の両方を含んでいる。

- B (Bidirectionally predictive) ピクチャ

前後の I ピクチャ及び P ピクチャからの双方向予測から符号化を行う画像である。前後の I ピクチャ、P ピクチャに依存して画像を復元しているため、それらが出力されないと、正常に復元できない。データ容量は他のピクチャよりも小さい。

以上の I ピクチャ、P ピクチャ、B ピクチャの構成を表したものをピクチャパターンと呼ぶ。よく用いられる GOP のピクチャパターンの例として、IBBPBBPBBPBBPBB がある。GOP の長さはランダム・アクセスの単位に影響を及ぼし、一般的には 0.5 秒から数秒の範囲の値が選ばれる。また、I ピクチャまたは P ピクチャの周期は、画像の動きに関して符号化の効率を決める。したがって、目的に応じて GOP の設定をする必要がある。

1.2.3 メディア出力制御方式

マルチメディア通信では、ネットワーク遅延の揺らぎにより、連続メディアである音声・ビデオがそのメディアの時間構造を乱されることがある。この時間構造を回復するために、メディア同期制御が必要になる。

メディア同期制御には、メディア内同期制御、メディア間同期制御及び端末間同期制御がある [20]。

メディア内同期は、単一メディア内の MU (メディアユニット, Media Unit) の発生間隔をメディア出力先で維持するためのものである。メディア内同期制御は、各メディアに対して行われる。なお、MU とは、メディア同期のための転送の単位であり、音声サンプルやビデオフレーム等に相当するものである。

メディア間同期は、音声とビデオなど複数のメディアが存在する場合に、それらの間での時間構造を維持するためのものである。マスタメディアのメディア内同期を優先し、スレーブメディアは対応するマスタメディアの時間関係を維持する。

端末間同期は、マルチキャスト送信された MU を、複数のメディア出力先において、同時に出力させるためのものである。端末間の公平さを維持するために必要な制御である。

メディア内同期、メディア間同期及び端末間同期の関係を図 1.1 に示す。

音声・ビデオのパケット通信におけるメディア同期品質を向上させる方法として、受信端末におけるバッファリング制御がある。本論文では、バッファリング制御方式として、簡単なプレイアウトバッファリング方式 [21][22] を用いる。プレイアウトバッファ

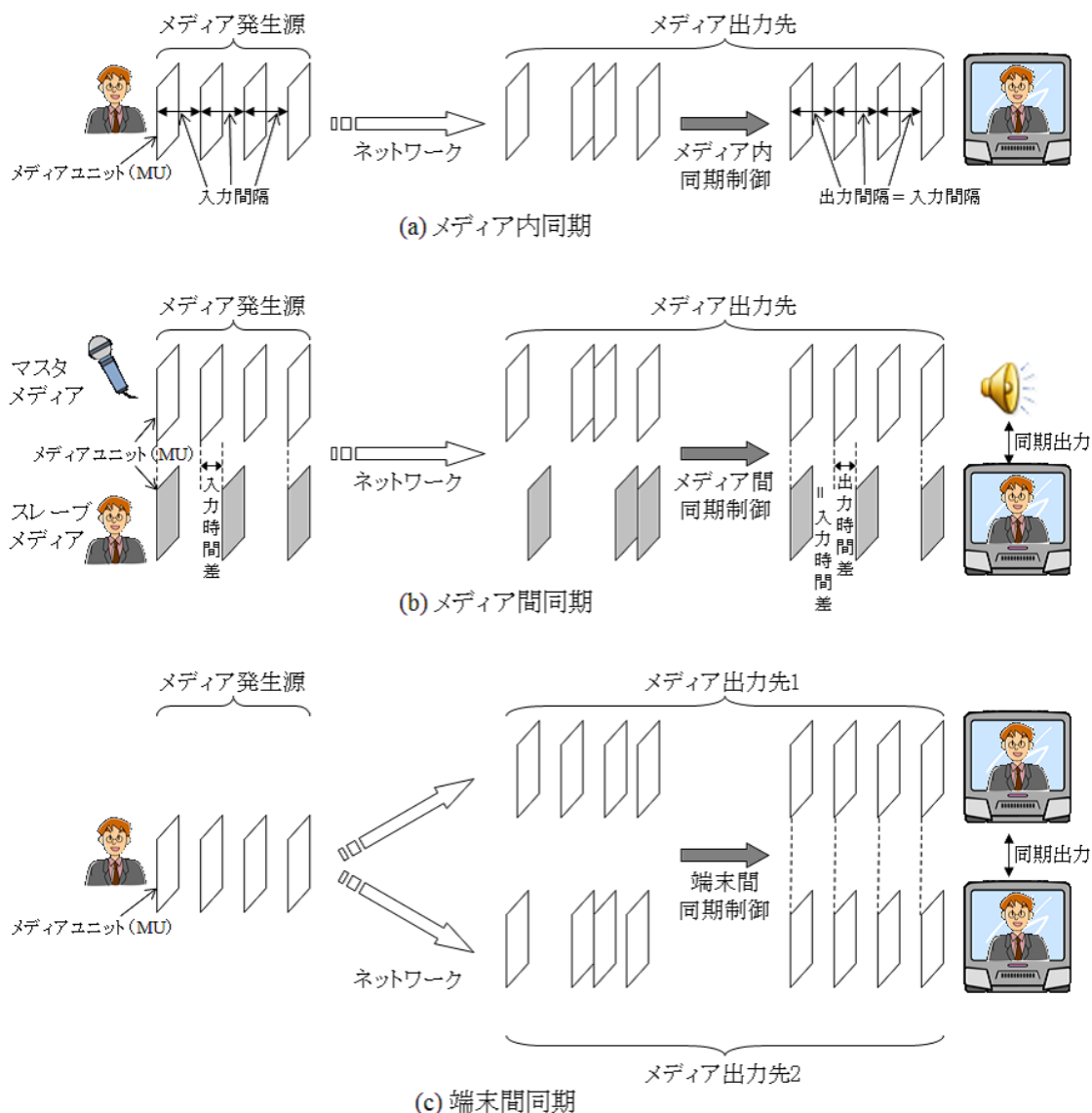


図 1.1: メディア同期

リング方式は、受信側で受け取ったデータを一時的にバッファに格納し、出力すべき時刻に出力する。プレイアウトバッファリング方式の概念図を図 1.2 に示す。

本研究に用いるプレイアウトバッファリング方式のアルゴリズムを以下に示す。最初の MU の出力時刻 t_1 は、最初の MU の到着時刻 A_1 と固定バッファリング時間 B を用いて、式 (1.1) のように表される。

$$t_1 = A_1 + B \quad (1.1)$$

また、2 番目以降に到達した i 番目の MU の目標出力時刻 t_i は、送信元でのタイムスタンプ T_i を用いると、次の式 (1.2) となる。

$$t_i = t_{i-1} + (T_i - T_{i-1}) \quad (1.2)$$

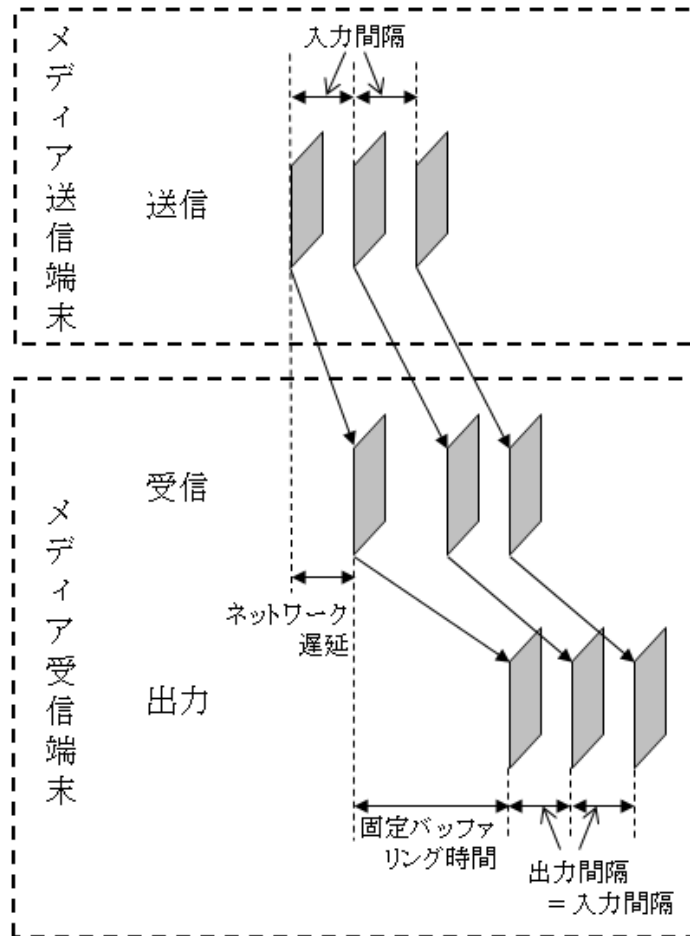


図 1.2: プレイアウトバッファリング方式

受信側での MU の到着時刻が，その MU の目標出力時刻以前であれば，目標出力時刻までバッファリングした後に出力を行う．また，目標出力時刻以降に到着した場合，その MU は破棄される．以上の制御を，音声・ビデオを区別して行う．

1.3 QoS

1.3.1 定義

QoS (サービス品質, Quality of Service) の定義は様々な考え方があり，厳密に定義することは難しい．文献 [23] では“あるべき姿への一致度”と定義されている．“あるべき姿”とは理想的なサービスを意味するが，評価する対象によって理想的とされるサービスは異なる．遠隔学習において理想的なサービスとは，面接授業により直接学習者が講義者の授業を受けるサービスと考えることができる．

1.2.1 項で示したように，IP ネットワークは階層構造を持つ．階層ごとに役割は決まっており， N 層はその下位レベルである $(N - 1)$ 層からサービスを受け，上位レベルであ

る $(N + 1)$ 層にサービスを提供する．従って，QoS もネットワークモデルの各階層に対して定義することができる．文献 [24] では，表 1.2 に示すように，上位層から順に，ユーザレベル QoS，アプリケーションレベル QoS，エンドツーエンドレベル（トランスポートレベル）QoS，ネットワークレベル QoS，ノードレベル（リンクレベル）QoS，物理レベル QoS の六つに分類している．ユーザレベル QoS は，最終的なサービス利用者である人間が感じる品質であり，通信の国際標準化機関である ITU-T では，これを QoE（ユーザ体感品質，Quality of Experience）と定義している [7]．したがって，IP ネットワーク上で音声・ビデオ伝送を行う際には，QoE を定量的に評価できることが重要である．

表 1.2: インターネットの階層構造と QoS の定義

インターネットの各階層	QoS
ユーザ	ユーザレベル QoS, QoE
アプリケーション層	アプリケーションレベル QoS
トランスポート層	エンドツーエンドレベル QoS (トランスポートレベル QoS)
ネットワーク層	ネットワークレベル QoS
データリンク層	ノードレベル QoS (リンクレベル QoS)
物理層	物理レベル QoS

QoE は人間の主観を含んだものであるため，これを直接測定することは時間と労力を必要とする．そこで，測定の容易なアプリケーションレベル以下の階層の QoS から，QoE を推定することが望まれる．

1.3.2 QoS パラメータ

QoS の良し悪しを定量的に表現するための尺度を，QoS パラメータと呼ぶ．インターネットの QoS においては，各階層の QoS ごとに QoS パラメータが存在する．特に，アプリケーションレベル QoS パラメータでは，転送するメディアや目的によっても異なる QoS パラメータを持つ．QoS パラメータの例を表 1.3 に示す [24]．

以下に，本論文で用いる音声・ビデオ伝送に関する主なアプリケーションレベル QoS パラメータを示す．なお，メディア間同期平均二乗誤差以外のパラメータは，音声とビデオごとに測定する．

平均 MU レート（音声： R_a ，ビデオ： R_v ）

単位時間（秒）にどれだけの MU が出力されるかを表す．総出力 MU 数をメディアの総出力時間で割って求める．この値が高いほどメディアは滑らかに表示される．

表 1.3: QoS パラメータの例

QoS のレベル	音声	ビデオ	コンピュータ データ
ユーザ	MOS 心理的尺度		
アプリケーション	PESQ R 値	PSNR フレームレート	スループット 応答時間
エンドツーエンド (トランスポート) ネットワーク ノード(リンク)	スループット PDU 遅延 PDU 遅延揺らぎ PDU 誤り率 PDU 欠落率		
物理	伝送速度 SNR ビット誤り率		

MU 欠落率 (音声: L_a , ビデオ: L_v)

出力されなかった MU 数を送信された MU 数で割った値である。

メディア内同期平均二乗誤差 (音声: E_a , ビデオ: E_v)

MU の出力間隔から発生間隔を引いた値の二乗を平均したものである。出力間隔の変動係数と同様、メディア出力の滑らかさを表す。

メディア間同期平均二乗誤差 (E_{int})

メディア間 (音声・ビデオ間) において、どの程度ずれが生じているかを表す。この値が小さいほどメディア間のずれが少なく、いわゆるリップシンクが保たれている状態と言える。マスタ MU とそれに対応するスレーブ MU の出力時刻の差から、それらの MU 発生時刻の差を引いたものを二乗し、総和をとり、出力 MU 数で割って求める。

出力間隔の変動係数 (音声: C_a , ビデオ: C_v)

これは MU の出力間隔がどれくらい揺らいでいるかを示す尺度である。この値が小さいほど滑らかに表示されていることを意味する。MU の出力間隔の標準偏差を MU の出力間隔の平均で割った数値である。

平均 MU 出力遅延 (音声: D_a , ビデオ: D_v)

バッファリングやネットワーク遅延などにより、MU がどの程度遅れて出力されるかを表す。各出力 MU の出力時刻とその発生時刻 (タイムスタンプ) との差の総計を出力 MU 数で割って求める。値が小さいほどリアルタイム性が高い。

1.4 QoE

1.4.1 QoE 尺度及び評価方法

QoE 尺度

マルチメディア通信において，最終的にサービスを利用するのはユーザである．従って，QoE を保証することが最も重要である．

QoE を定量的に評価するための尺度を，QoE 尺度と呼ぶ．文献 [8] では，QoE 尺度として，“作業効果 (Task effectiveness)”，“作業効率 (Task efficiency)”，“ユーザの満足度 (User satisfaction)”，“ユーザの快適度 (User enjoyment)”を示している．

まず，作業効果を評価する例としては，作業の達成度や作業の精度が挙げられる．次に，作業効率を評価する例としては，作業時間や対話の途切れの回数が挙げられる．更に，ユーザの満足度を評価する例としては，サービスの許容性やコミュニケーションに対する満足度が挙げられる．最後に，ユーザの快適度を評価する例としては，没頭の度合いや面白さの度合いが挙げられる．

測定尺度の分類

文献 [8] で示された QoE 尺度の中で，ユーザの満足度とユーザの快適度は，ユーザが感じる品質であり主観評価になるため，定量的な評価が困難である．しかし，主に心理学で用いられている計量心理学的測定法を用いることで，評価対象に対する主観量を定量化することができる．計量心理学的測定法では，主観量を表す測定尺度は以下の四つに分類される [25][26]．

- 名義尺度 (nominal scale)

事象の分類のために各カテゴリーに与えられた数値による尺度．数値の大きさや順序に意味がなく，カテゴリーの識別だけを行う尺度なので，同一カテゴリーに複数の数値を与えないことと，異なるカテゴリーに同一の数値を与えないことを守れば，いかなる数を用いても差し支えない．

- 順序尺度 (ordinal scale)

相対的な順序関係だけを表す数値による尺度．数値の大小のみが意味をもち，その差は意味を持たない．従って，数値に等間隔性の保証がないため，四則演算が行えない．序数尺度とも呼ばれる．

- 距離尺度 (interval scale)

数値の差 (距離) が意味を持っている尺度．任意の原点と等間隔な単位を持つ尺度である．この尺度では，ほとんどすべての統計的操作が可能である．間隔尺度とも呼ばれる．

- 比例尺度 (ratio scale)

絶対的な原点と等間隔な単位を持った尺度．尺度上での等間隔，等比率が保証され，あらゆる数学的操作を行うことができる．比率尺度とも呼ばれる．

ITU-T 勧告 P. 910[27] は，主観評価方法として，ACR (Absolute Category Rating) 法と DCR (Degradation Category Rating) 法を規定している．ACR 法は単一刺激法 (Single Stimulus Methods) とも呼ばれる評価方法であり，評価映像を表 1.4 に示す 5 段階品質尺度により評価する．一方，DCR 法は，品質評価の基準となるレファレンス映像を見た後に，それを基準として評価映像の品質を表 1.5 に示す 5 段階妨害尺度により評価する．

表 1.4: 5 段階品質尺度

評価値	評価基準
5	非常に良い (Excellent)
4	良い (Good)
3	普通 (Fair)
2	悪い (Poor)
1	非常に悪い (Bad)

表 1.5: 5 段階妨害尺度

評価値	評価基準
5	劣化がわからない (Imperceptible)
4	劣化がわかるが気にならない (Perceptible but not annoying)
3	劣化が気になるが邪魔にならない (Slightly annoying)
2	劣化が邪魔になる (Annoying)
1	劣化が非常に邪魔になる (Very annoying)

一方，ITU-R (International Telecommunication Union-Radiocommunication Sector) 勧告 BT. 500-11[28] は，二重刺激連続品質尺度法 (Double Stimulus Continuous Quality Scale method) 法を規定している．これは，相対評価に分類される評価方法である．二重刺激連続品質尺度法は，評価対象と基準画像とを対にして順に提示する評価方法である．この方法では，評価対象と基準画像とが交互にそれぞれ 2 回ずつ被験者に提示される．そして，被験者は二つの刺激に対して，図 1.3 に示す評価尺度において，任意の場所に横線を引くことで評価を行う．このとき，提示される画像のうち，どちらが基準画像かは被験者には知らせない．

ITU-T や ITU-R では，上述した評価によって得られた結果を平均し，これを評価値としている．特に，5 段階尺度での評価によって得られる結果の平均値である MOS (平

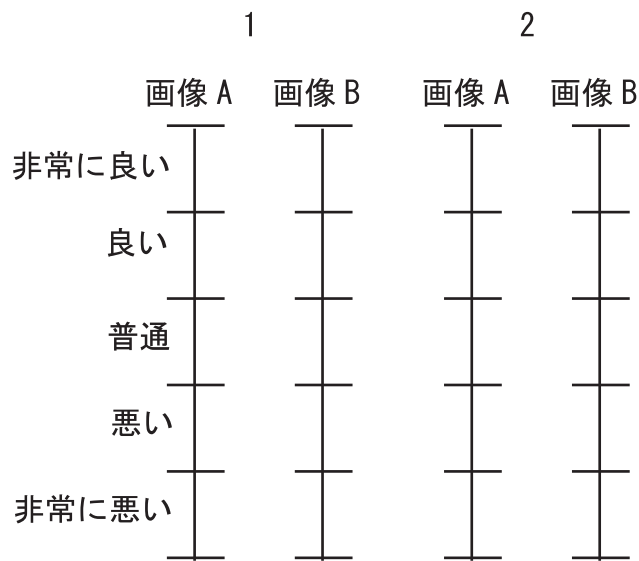


図 1.3: DSCQS における評価尺度

均オピニオン評点, Mean Opinion Score) [29] が主観評価方法の代表的なものである。しかしながら MOS は, 厳密には順序尺度であり, その尺度値は順序関係の情報を表すに過ぎない。そのため, MOS は, 音声のみ, あるいはビデオのみの単一メディアの主観評価には実用上は有用であることが知られているが, マルチメディアに対して有用である保証はない。

音声・ビデオの主観評価方法としては, MOS よりも正確な距離尺度を求めるため, 計量心理学的測定法による心理的尺度を用いたものが提案されている [30]。文献 [31][32] では, このことを実験により, 定量的に示している。そこで, 本論文では, MOS ではなく, 被験者の判断結果の大小関係に加えて主観量の等間隔性も保証する距離尺度である心理的尺度を用いることとする。

測定方法

計量心理学的測定法では, 評価対象を被験者に提示し, それらに対する評価値の測定を行う。以下に, 主な測定方法を示す [25]。

- 極限法

極限法では, 被験者に二つの評価対象を提示する。一方は変化させず, もう一方を適当な間隔で変化させる。被験者はその都度評価対象を比較し, 定められた選択肢から評価対象への評価を選択する。

- 恒常法

恒常法は, 提示する評価対象に関しては極限法と同様だが, 提示順序がランダムである。数段階に変化する評価対象が, それぞれ数十回から百回ほど提示される。被験者は, それぞれの評価対象に対して選択肢から当てはまる項目を選び, その

項目の出現頻度が計算される．そして，その出現頻度に対応して評価値が算出される．

- 一対比較法

一対比較法は，二つの刺激を提示し，その二つを比較させる方法である． n 個の刺激を評価する場合，その全ての組み合わせである ${}_nC_2 (= \frac{n(n-1)}{2})$ 回の比較を行う．この方法は，判断基準が明確で，評価が容易であるが，刺激数が増えることで評価回数が急増する．そのため，被験者の負担を考慮することが必要である．そこで，一対比較法において比較回数が増加することによる問題点を明確にし，それを改善した手法を用いた実験も行われている [33]．一対比較法によって得られる結果は，順序尺度である．しかし，Thurstone の比較判断の法則 [34][35] により，その結果を距離尺度に変換することができる [30]．

- 評定尺度法

評定尺度法とは，実験者が一定個数のカテゴリーを設定しておき，被験者は提示された評価対象をいずれかのカテゴリーに分類する方法である．このとき，実験で用いるカテゴリー数は5または7段階であることが多い．これは，3段階では一つのカテゴリーに該当する範囲が広くなりすぎ，9段階以上では使用されないカテゴリーができるためである．また，評定尺度法による評価では，各カテゴリー間の心理的な間隔は一定とは限らないため，一対比較法と同様に，得られる結果は順序尺度である．そこで，カテゴリー判断の法則 [25][36] により，評定尺度法による結果を距離尺度に変換することが考えられている [37]．

1.4.2 系列カテゴリー法

系列カテゴリー法は，評定尺度法で得られた順序尺度に，カテゴリー判断の法則を適用することで距離尺度を求める手法である [25][31]．この際に得られた距離尺度を心理的尺度と呼ぶ．

カテゴリー判断の法則

カテゴリー判断の法則とは，被験者がある刺激に対して感じる品質の良さは，心理的連続体と呼ばれる脳内のある距離尺度上の数値として生起すると仮定するものである．そして，この値が心理的連続体上にあるカテゴリーの境界値と比較され，その結果，各刺激がカテゴリーに分類される．ただし，刺激に対して生起する値及びカテゴリーの境界は一定ではなく，正規分布に従って変動するものと仮定される．

以上の仮定は，本質的にはカテゴリーの境界も刺激のように振る舞うということであるため，Thurstone の比較判断の法則を導くのと同じ手続きで，以下のようなカテゴリー判断の法則が導き出される．

$$t_g - R_j = Z_{jg} \sqrt{\sigma_j^2 + d_g^2 - 2r_{jg}\sigma_j d_g} \quad (j = 1, 2, \dots, n, \quad g = 1, 2, \dots, m) \quad (1.3)$$

- j : 刺激
- g : カテゴリー
- s_j : 刺激 j の弁別過程
- c_g : カテゴリー g の境界
- t_g : c_g の平均値
- d_g : c_g の標準偏差
- R_j : s_j の平均値
- σ_j : s_j の標準偏差
- p_{jg} : 刺激 j がカテゴリーの境界 g より下であると判断された割合
- Z_{jg} : p_{jg} に対する正規偏差
- r_{jg} : s_j と c_g の相関係数

このモデルを図 1.4 に示す．上記の式は，未知数に対して得られる条件式の数が少ないため，このままの形では解くことができない．そこで，Torgerson は，測定値の繰り返しが同一の被験者の反復によって得られる場合を第 1 類，多数の被験者の各 1 回の試行によって得られる場合を第 2 類，両者の混合された場合を第 3 類と呼び，次の 4 条件を区分している [25] ．

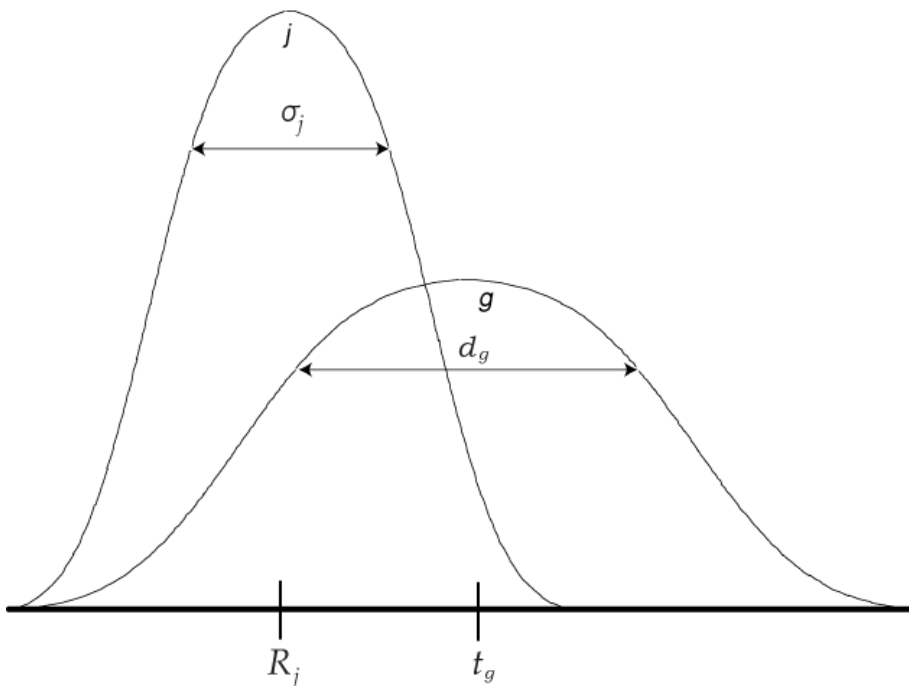


図 1.4: カテゴリー判断の法則のモデル

- 条件 A : $r_{jg}\sigma_j d_g = c$ と仮定される場合 (c は定数)

$$t_g - R_j = Z_{jg} \sqrt{a_j^2 + b_g^2} \quad (1.4)$$

$$a_j = \sqrt{\sigma_j^2 - c}$$

$$b_g = \sqrt{d_g^2 - c}$$

c は定数

- 条件 B : $d_g^2 = c$, $r_{jg} = 0$ と仮定される場合 (c は定数)

$a_j^2 = \sigma_j^2 + c$ とすれば ,

$$t_g - R_j = Z_{jg}a_j \quad (1.5)$$

さらに , この式において , カテゴリーの境界が固定的でされるとある場合 ,

$$t_g - R_j = Z_{jg}\sigma_j \quad (1.6)$$

- 条件 C : $\sigma_j^2 = c$, $r_{jg} = 0$ と仮定される場合 (c は定数)

$b_g^2 = d_g^2 + c$ とおけば ,

$$t_g - R_j = Z_{jg}b_g \quad (1.7)$$

- 条件 D : $d_g^2 = c$, $\sigma_j^2 = c$, $r_{jg} = 0$ と仮定される場合 (c は定数)

$$t_g - R_j = Z_{jg}C \quad (1.8)$$

C は定数

本論文では , 最も仮定が多く , 解法が容易な条件 D を検討する .

適合度検定

尺度値は , 観測された割合 p'_{jg} から求められたものであり , 得られた尺度値から期待される割合 p''_{jg} を算出することができる . 逆に , 得られた尺度値から期待される割合 p''_{jg} を算出できる . この値を用いて , 以下の方法でモデルがデータに適合しているかを調べることができる [38][31] . もし p'_{jg} と p''_{jg} の差があまりにも大きい場合は , モデルはデータに適合していないと考えねばならない .

まず , θ'_{jg} , θ''_{jg} を次のように決める .

$$\theta'_{jg} = \sin^{-1}\sqrt{p'_{jg}} , \quad \theta''_{jg} = \sin^{-1}\sqrt{p''_{jg}} \quad (1.9)$$

次に , n_s を観測者数 , n を評価対象数 , m_j を利用できる θ'_{jg} の数とすると ,

$$\chi_0^2 = \sum_{j=1}^n \sum_{g=1}^{m_j} \frac{(\theta''_{jg} - \theta'_{jg})^2}{821/n_s} \quad (1.10)$$

式 (1.10) の値が , 自由度 $(nm - n - m)$ のカイ二乗分布になることを利用し , 実験結果との検定を行う .

1.4.3 QoS マッピング

前節で得られた QoE 尺度をアプリケーションレベル QoS パラメータにマッピングするため、本論文では重回帰分析を用いる [31]。QoE 尺度を従属変数とし、アプリケーションレベル QoS パラメータを独立変数とすることにより、アプリケーションレベル QoS パラメータを用いて QoE 尺度を表現できるようにする。この重回帰式によって QoE を推定することが可能となる。

この方法を以下に説明する。まず、得られたアプリケーションレベル QoS パラメータに対して重回帰分析を行う。このとき、独立変数の候補としては多くのものが考えられるが、多重共線性の問題より、すべてを独立変数にはできない。そこで、独立変数の候補となるパラメータに対して主成分分析を行う。次に、各主成分の累積寄与率に基づき、候補となるすべてのパラメータを表現するために必要な主成分の数を求める。そして、各主成分の主成分負荷量から、各主成分に基づき、候補となるパラメータをいくつかのカテゴリーに分類する。さらに、各カテゴリーの中から独立変数を一つずつ選ぶ。

求められたすべての独立変数の組み合わせに対して重回帰分析を行い、自由度調整済み重相関係数（重回帰式の精度を表す係数）が最も大きい組み合わせを求める。最後に、求めた重回帰式の偏回帰係数に対して検定を行い、重回帰式に寄与していない独立変数を除外する。

1.5 遠隔学習における QoE

1.5.1 主観品質

教育の効果測定方法として、D. L. Kirkpatrick によって提案されたカークパトリックの4段階評価法（Kirkpatrick's Four Levels of Evaluation Model）が広く知られている [39]。これは、企業内教育を対象とした評価法として提案されたものであるが、遠隔学習の効果測定としてもよく用いられる [40]。

評価方法は、表 1.6 に示す四つのレベルの評価内容から成り立つ。

表 1.6: カークパトリックの評価法

Level	評価内容
レベル 1 反応 (Reaction)	受講者の研修直後の評判, 満足度
レベル 2 学習 (Learning)	受講者の研修内容の理解度
レベル 3 行動 (Behavior)	受講者の行動の変化
レベル 4 成果 (Results)	研修で学んだことからの業績向上, 成果

これらの中で、レベル 1 を測定する方法として、受講者（ユーザ）の満足度を質問紙（アンケート）により調査したり、教師が受講者に対して面接（ヒアリング）により調査したりする。本論文においては、アンケートを行うことで、遠隔学習における QoE

表 1.7: 客観テストの分類

大分類	小分類	問題形式
再生形式	単純再生法	解答を直接求める問題形式
	完成法 (穴埋め再生法)	文章や図式などにあるいくつかの空所を埋めさせる形式
	訂正法	誤った部分を含んだ文章や図などを正しく訂正させる形式
	序列法	無秩序に並べられたものを年代順, 作業順などに正しく並べ替えさせる形式
再認形式	真偽法 (正誤法)	用意された二つの選択肢から一方を選ばせる形式
	多肢選択法	三つ以上の選択肢から正しいものだけを選択させる形式
	組み合わせ法	二つの事象間の中から互いに関係する事項同士を結び付けさせる形式
	選択完成法	多肢選択法と完成法の併用型

評価の一つである主観品質として扱う。また, 1.4.2 項で述べた系列カテゴリー法を用いて定量評価する。

次に, レベル 2 を測定する方法として, 学習直後の理解度を把握するための筆記試験が用いられる。これについては, 次項で詳細を述べる。

なお, レベル 3 及びレベル 4 においては, 定量評価に適さないため, 本論文では扱わない。

1.5.2 正答率

学習後の理解度を調べる客観テストは, 採点者の主観に左右されないため, 信頼性・客観性において最も優れた評価方法である [41]。そのため, 遠隔学習における QoE 評価として, 客観テストは重要な意味を持つ。客観テストには, 表 1.7 に示す再生形式 (Recall-type) と再認形式 (Recognition-type) の 2 種類の方式がよく用いられる [42]–[44]。

次に, 客観テストの分類ごとに特徴を述べる。

まず, 再生形式のテストは, 解答者自らが学習し保持した事項や内容を思い出し (再生して), 指定された条件に当てはまる解答を作り出す形式である。そのため, 再認形式によるテストよりも確かな記憶を解答者に要求することになる。従って, テストの難易度が高くなり, 短い学習時間では正解を得ることが難しい。また, 単純再生法では, 少し複雑な内容を答えさせようとすると, いろいろな答えが可能になってしまい, 客観テストとは言い難くなる。一方, 完成法 (穴埋め再生法) では, 比較的高度な理解力や

関係判断力を確認できるという長所を持ちながら、正解を一つに絞ることが可能である。従って、完成法が客観テストとしてよく用いられる。

次に、再認形式のテストは、与えられた命題が正しいか否かを確認したり、与えられた事象間の関係を再確認したりする形式である [45]。再生形式のテストと異なり、解答者はわからない問題であっても解答を書くことがしやすい。中でも、真偽法（正誤法）は、一つの問題について、“正”-“誤”、“そうです”-“ちがいます”など、用意された2方向の選択肢のうち、いずれか一方を選ばせる方法である。この形式のテストは最も単純な解答方法であるため、偶然により正解を得られることが多い。それに対して、多肢選択法は、一つの問題に対して、例えば、四つから五つ程度の選択肢を与えて、その中から正答と思うものを選ばせる出題形式である。正答は、必ずしも一つとは限らず、複数個あっても構わない。これに答える解答者の心理からみれば、相互に幾重にも比較考慮して、適切なものを選択しなければならないため、判断力・推理力や理解の深さを確認することができる。更に、採点の客観性も確保されるため、今日あらゆる問題形式の中で最も使用される機会が多い。

本論文では、以上のことを踏まえて、これらの客観テストの中からいくつかを学習効果の評価に使う。学習効果の評価尺度としては、正答率を用いる。これは、正答のパーセントとして定義される。

1.5.3 返答時間

リアルタイム型遠隔学習においては、対面学習同様にインタラクティブな学習を可能にする。しかし、IP ネットワークを用いた遠隔学習では、1.2.3 項で述べたように、音声・ビデオがそのメディアの時間構造を乱されることがある。そのため、IP ネットワーク上でのリアルタイム型遠隔対話環境の評価においては、返答時間をいかに定量的に評価するかが重要となる。

今、一人の講師と一人の受講者が IP ネットワークで 1 対 1 接続されている環境を考える。この環境において、講師が受講者に対して発問した場合に、受講者の返答する状況を、一般的に時系列で記述すると次の通りである。

- (1) 講師が発問を始めてから、その問題を言い終わるまで
- (2) 講師が問題を言い終わってから、受講者がその問題を聞き終わるまで
- (3) 受講者が問題を聞き終わってから、返答を開始し始めるまで
- (4) 受講者が返答を始めてから、それを講師が聞き始めるまで
- (5) 講師が返答を聞き始めてから、聞き終わるまで

上記の中で、ネットワーク遅延により影響を受ける時間は、(2) と (4) である。そこで、本論文では、(2) から (4) までの時間の合計を“返答時間”とする。上記の (1) から (5) における発問から返答までの関係と返答時間を図 1.5 に示す。

一方、受講者が複数箇所にいる場合には、多地点を接続することになり、1対多のリアルタイム型遠隔対話環境となる。この場合、IPネットワークを用いた接続では、端末ごとに通信品質の影響で返答時間の違いが生じることが考えられる。そのため、多地点接続の場合、端末によって不公平が生じる可能性がある。そこで、本論文において、複数受講者の返答時間の早遅を QoE 評価尺度の一つとして用いる。

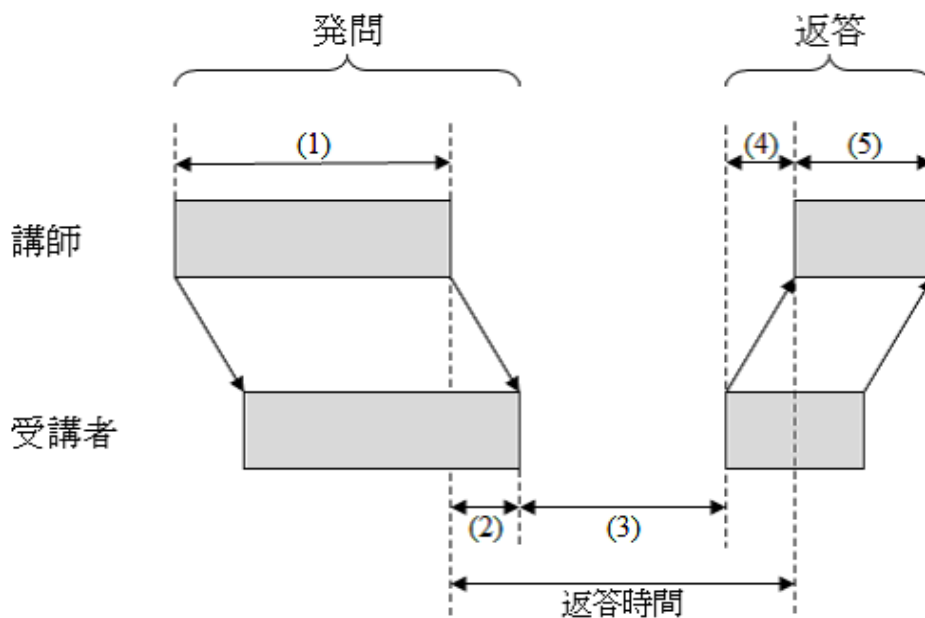


図 1.5: リアルタイム型遠隔学習における返答時間

1.6 本論文の目的と構成

インターネットを利用した遠隔学習においては、面接授業と同等の教育効果が要求されるが、インターネット等の IP ネットワークは、基本的に QoS を保証しない。そのため、Web アクセスやビデオストリーミングなどのトラフィックが混在すると、QoS の劣化に敏感であるマルチメディア遠隔学習の QoE は容易に劣化する。そこで、遠隔学習の QoE を高めるためには、QoE を適切に評価することが必要になる。

本論文では、マルチメディア教材を使用した遠隔学習システムを多面的な QoE で評価する。遠隔学習の QoE として、学習効果、主観評価及び返答時間を扱っている。本論文において扱う遠隔学習システムは 2 種類ある。一つは、一人の講師と一人の受講者が離れた場所にいる 1 対 1 のオンデマンド型遠隔学習システムである。もう一つは、一人の講師と複数の受講者が遠隔地にいる 1 対多のリアルタイム型遠隔学習システムである。1 対 1 オンデマンド型遠隔学習システムにおいては、教材コンテンツで使用する“文字”、“音声”、“ビデオ”の組み合わせの違いが、学習効果に及ぼす影響を定量的に調べる。また、1 対多リアルタイム型遠隔学習システムにおいては、学習者の主観品質と返答時間の違いによる学習効率を調べる。それらの結果から、IP ネットワーク上でのマルチメディア遠隔学習において、遠隔学習システムの開発者、及び教材コンテン

ツの提供者等にとって、遠隔学習者の QoE を向上させるための指針を提供することを、本論文の目的とする。

第2章では、音声・ビデオベースのコンテンツを扱う1対1接続のオンデマンド型遠隔学習システムにおいて、コンテンツ中の文字情報が学習効果に及ぼす影響を検討する。文字情報の提示は出力ディスプレイ上で字幕により行い、字幕を提示しない場合と、音声と同じ言葉の字幕を提示する場合とで学習効果を比較している。被験者がコンテンツを視聴して、覚えた言葉を書き写した解答結果の正答率を学習効果としている。ネットワークに加わる平均負荷量が変化したときに、文字情報が学習効果に与える影響を分析した。その結果、IP ネットワーク上でのコンテンツには、文字情報を加えることが、学習効果の向上に有効であることを示している。

第3章では、テキストベースの教材コンテンツを扱う1対1接続のオンデマンド型遠隔学習システムを用いて、音声・ビデオ品質が学習効果に及ぼす影響を実験で調査する。教材コンテンツとしては、“テキストだけ”、“テキストと音声”、“テキスト、音声とビデオ”の3通りの情報提示方法を用意した。これらの情報提示方法において、通信品質の違いが学習効果（正答率）にどのように影響するのかを評価している。その結果、通信品質が高い場合には、テキストと一緒に送信された音声・ビデオが、テキストだけの場合と比較して学習効果を高くすることがわかった。しかし、通信品質が劣化するにつれて、テキストだけの場合よりも、テキストに音声・ビデオを加えた場合の方が、音声・ビデオが学習効果を低くすることを示した。更に、回帰分析により、通信回線に加えられた負荷トラフィック量に対して最適な情報提示方法を選ぶためのガイドラインを示している。

第4章では、多地点接続のリアルタイム型遠隔学習システムにおいて、受講者用端末がバッファリング制御を行った場合の QoE を評価する。一人の講師と複数の受講者とが、1対多で行う対話を扱う。遠隔学習における QoE として、受講者の主観品質と返答時間とを評価している。音声・ビデオの出力品質を改善するために行われる受信バッファリング制御では、各端末が独立にバッファリング時間を設定すると、端末間同期品質を低下させ端末間の公平さを損なう。本章では、この観点から実験を行うことにより、バッファリング制御が主観品質と返答時間に影響を及ぼす要因を調べている。その結果、総合的な主観品質と返答時間は端末間同期品質に影響されることを明らかにしている。そして、QoE の改善のためには、端末間同期品質を高める必要があることを示している。

第2章 音声・ビデオベースの遠隔学習 において文字情報が学習効果に 及ぼす影響

2.1 まえがき

インターネットに代表される IP ネットワーク上で、音声・ビデオ伝送を扱うアプリケーションの利用が広がっている。この種のアプリケーションには、音声・ビデオに文字情報を含めるものもある。

近年、IP ネットワークを用いて、遠隔学習用コンテンツが使われている。遠隔学習においても、コンテンツに文字情報を含むことが多く見られる。これは、文字の提示で学習者の記憶を増進させ、学習効果を上げることを期待しているからである。

IP ネットワークは、ベストエフォート型のサービスを提供するため、通信時に、パケットの欠落、遅延及び遅延揺らぎが生じる。これらは音声・ビデオの時間構造を乱し、QoS を低下させる要因となる。

文字情報は、IP 伝送されたとき、パケットの欠落、遅延及び遅延揺らぎの影響を受けにくいと考えられる。コンテンツ中の文字は、意味のつながる複数個が同時に表示され、それらを読むのに必要な時間だけ、表示され続けることが一般的である。表示時間が長いので、特定の文字が全く表示されなくなることは少ない。そのため、音声・ビデオに文字情報を加えることで、視聴者に対する情報伝達の効率化が可能になる。

IP ネットワークを用いた遠隔学習において、文字情報を活用することで、学習効果を向上させることが期待できる。本章は、その第一歩として、インフラとしての IP ネットワークを使って、文字情報が学習効果に及ぼす影響を分析する。特に、IP ネットワークのベストエフォート型サービスという欠点を補う手段として、文字情報伝送が有効であることを示す。

IP ネットワークにおいては、様々なアプリケーションが使われ、各種の情報伝達が行われている。そこで、文字情報を加える効果が、アプリケーションの種類によって、どのように異なるかを検討する。

文字情報を含む音声・ビデオ IP 伝送における QoE として、主観品質（ユーザの満足度）と作業効果が挙げられる。主観品質とは、自端末で出力された音声・ビデオを総合し、システム全体に関して感じる良し悪しである。また、作業効果とは、対象とする作業に対する評価を数値などの結果として表したものである。例えば、遠隔学習では学習効果がこれに相当し、リスニングテストの正答率、制限時間内に読み上げた文字数などとなる。

文字情報を含むコンテンツの QoE に関する研究は、既にいくつか報告されている。文献 [46][47] では、ビットレートを変化させて、文字を含んだ画像の劣化度合いを主観評価している。また、文献 [48] では、動画像と文字情報の配置が、学習者の視覚に及ぼす心理的影響を検討している。

学習効果（作業効果）を求めた研究には、次のものがある。文献 [49] では、折り紙を折る工程を示した映像を、被験者にテレビ映像で視聴してもらい、文字情報の有無が、被験者が折り終わるまでの作業時間に及ぼす影響を調査している。文献 [50] で、ISDN 回線を利用した遠隔講義において、文字の大きさの違いによる文字認識率の測定をしている。また、文献 [51] では、スタンドアロンのパソコンを使って、文章を視覚と聴覚で提示した場合の記憶への影響を調べている。

学習用のコンテンツでは、英語字幕を加えることで、英語の学習成績を向上させることを示したものがある [52]–[54]。しかし、これらはビデオや DVD 教材などローカルな環境のみを対象にしており、IP ネットワークによる伝送は考慮していない。

IP ネットワーク環境における研究では、文献 [55] で、負荷を変動させて、スクロール表示文による学習の正答率と主観評価を行っている。しかし、これは文字情報のみの単一メディアにおける作業効果を求めたものである。IP ネットワークにおいて、複数メディアを扱い、文字情報が情報伝達に与える効果を検討した研究は見られない。

そこで、本章では、音声・ビデオ IP 伝送において文字情報が作業効果に及ぼす影響を調べる。特に、遠隔学習研究の第一歩として、学習効果への影響を中心に検討を行う。IP ネットワークに加える負荷量を変化させて、音声・ビデオに文字情報を加えることが学習効果向上に、効果的であることを示す。また、複数のコンテンツを用いて、文字情報が与える影響のコンテンツによる違いを比較する。更に、アプリケーションレベル QoS が学習効果に及ぼす影響を定量的に示す。

以下では、2.2 節で評価方法と実験方法を説明し、2.3 節で実験結果と考察を示す。

2.2 評価実験

2.2.1 情報メディア

本章においては、テレビ放送または DVD ビデオから録画した音声・ビデオを、情報メディアとする。録画した情報メディアは、文字情報を含まないものである。音声と同じ内容を字幕として表示させるように加工したものを、文字情報を含む情報メディアとする。

字幕は、すべて、ディスプレイ内で映像より下部に、黒色の背景に白色の文字で表示する。たとえば、“手を差し伸べてくれる君がいた”といった文字をスクロールなしの字幕で一括表示し、一定時間経過後に一括切り替える。字幕の表示時間は、各字幕を音声で出力するのに必要な時間である。

情報メディアは特定時間分だけ取り出し、それをコンテンツとして扱う。本章では、文字情報を提示していないコンテンツを“字幕なし”、文字情報を提示したコンテンツを“字幕あり”と呼ぶ。字幕ありの場合のイメージを図 2.1 に示す。



図 2.1: コンテンツのイメージ (字幕あり)

2.2.2 コンテンツ

文字情報が視聴者の学習効果へ及ぼす影響は、コンテンツのタイプによって異なる可能性がある。そこで、本実験では、音声重要視されるもの、音声とビデオの両方が重要視されるもの、及びビデオが重要視されるものの3種類のコンテンツタイプを選んだ。音声重要視されるコンテンツタイプの例としてミュージックビデオ、音声・ビデオが重要視されるものの例として漫才、ビデオが重要視されるものの例として競馬を採用した。なお、本章は、遠隔学習研究の第一歩として行ったため、幅広いコンテンツを対象とした。

それぞれのコンテンツタイプでは、以下のような場面を用いる。

- ミュージックビデオ

歌手が早口に歌う場面。比較的聞き取りにくい音楽。字幕ありの場合、歌詞を記述。

- 漫才

漫才を行う二人が互いにやりとりをしている場面。会話の前後の意味がわかりやすい場面。字幕ありの場合、会話を記述。

- 競馬

表 2.1: メディアの仕様

仕様	音声	ビデオ
符号化方式	LinearPCM 48kHz 16bit ステレオ	MPEG-1
表示サイズ [pixel]	-	480 × 360
ピクチャパターン	-	I のみ
再生時間 [sec]	17	17
総 MU 数	340	510
MU レート [MU/s]	20	30
ビットレート [Mb/s]	1.536	4.0

レース中盤，アナウンサーが馬の名前を順番に読み上げる場面．画面上で馬の番号が確認できる場面．字幕ありの場合，馬の番号と名前を記述．

コンテンツは，それぞれのコンテンツタイプにつき 4 個ずつ用意し，計 12 個のコンテンツを用いる．12 個すべてのコンテンツに対して，字幕なし・字幕ありの 2 種類を扱う．

コンテンツタイプ毎の評価を行うために，一つのコンテンツタイプに含まれる 4 個のコンテンツは，上で述べた観点から似たものを選択した．

実験で使用したコンテンツは，次の通りである．ミュージックビデオ，漫才，競馬のコンテンツタイプ毎の平均で，一括表示する文字数は，それぞれ，約 11 文字，約 13 文字，約 9 文字である．同様に，文字の表示時間は，それぞれ，約 2.2 秒，約 1.4 秒，約 2.0 秒である．

ミュージックビデオと漫才では，音声を 100 % 文字化したものを字幕で表示する．競馬では，音声に合わせて，馬の名前と番号のみ字幕で表示する．そのため，平均で音声の約 60 % を文字情報として表示することになった．

以下，ミュージックビデオの 4 個のコンテンツを music1, music2, music3, music4, 漫才の 4 個のコンテンツを，manzai1, manzai2, manzai3, manzai4, 競馬の 4 個のコンテンツを keiba1, keiba2, keiba3, keiba4 と呼ぶ．

2.2.3 実験システム

本実験においては，ns-2 (network simulator-2) [56] を用いて，音声・ビデオ伝送のシミュレーションを行う．その出力結果を用い，アプリケーションレベル QoS を評価する．伝送するメディアの仕様を表 2.1 に示す．表内における MU は，メディアユニット (Media Unit) を表し，アプリケーション層でのメディア同期のための情報単位である．ビデオでは 1 フレームに相当し，音声では一定個数のサンプルに相当する．メディア同期とは情報メディアの時間構造の保持・回復を行うことである [20]．

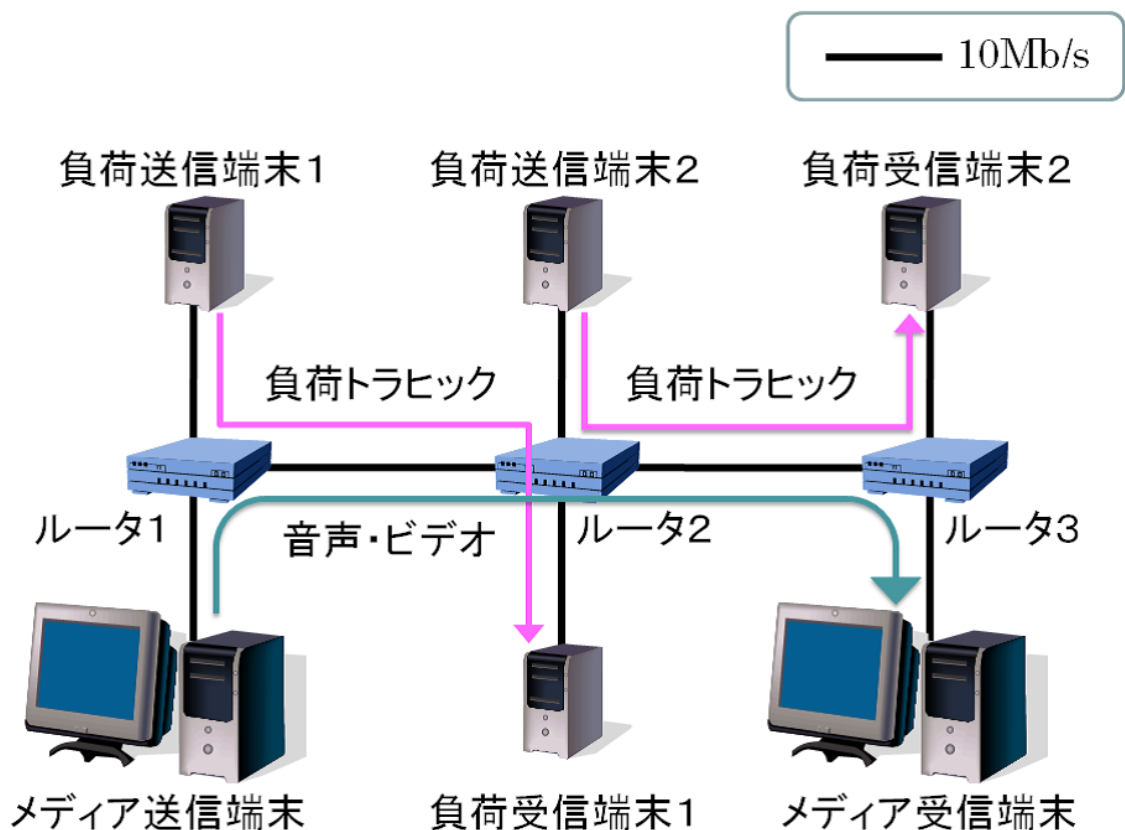


図 2.2: ネットワークトポロジー

シミュレーションで用いるネットワークトポロジーを図 2.2 に示す．図において，すべての端末・ルータ間は，10Mb/s の Ethernet で接続されている．これは，ADSL 相当の一般的な環境を想定している．メディア送信端末からメディア受信端末へ音声・ビデオを伝送する．同時に，負荷送信端末 1，負荷送信端末 2 から，それぞれ，負荷受信端末 1，負荷受信端末 2 へ，負荷トラヒックを伝送する．負荷送信端末 1，負荷送信端末 2 は，負荷トラヒックとして，1480 バイト長の UDP データグラムを指数分布に基づく間隔で生成する．

このシミュレーションにおいて，欠落する MU 数に大きな違いが生じるように，トラヒックの平均負荷量を 4.3Mb/s から 4.7Mb/s まで 0.1Mb/s 刻みで 5 通りに変化させ，音声・ビデオを伝送する．MR で出力された音声・ビデオをファイルとして記録しておく，これを QoE 評価実験の被験者に提示する．

2.2.4 アプリケーションレベル QoS パラメータ

本章では，アプリケーションレベル QoS パラメータとして，次のものを用いる．

第一に，音声・ビデオのメディア内同期の QoS パラメータを考える．メディア内同期とは，単一メディアにおける MU の時間間隔に関するものである．まず，音声とビデオの平均 MU レート R_a と R_v を用いる．ここで用いた下添え字の a は音声を， v はビ

デオを表している．以後に述べるアプリケーションレベル QoS パラメータの説明でも同様である．次に，送信した MU 数に対する欠落した MU 数の割合である MU 欠落率 L_a, L_v を取り上げる．また，音声・ビデオの MU の生成間隔と出力間隔の差の二乗を平均したメディア内同期平均二乗誤差 E_a, E_v と，MU の出力間隔の標準偏差を平均値で割った値である変動係数 C_a, C_v を出力の滑らかさを表すパラメータとして用いる．

次に，メディア間同期品質に関するものとして，ビデオ MU の出力時刻と，これに対応する音声 MU の出力時刻との差から，それぞれのタイムスタンプの差を引いた値の二乗平均であるメディア間同期平均二乗誤差 E_{int} を扱う．音声と唇の動き（ビデオ）を合わせるリップシンクは，メディア間同期の代表例である．

更に，MU が発生してから通信相手の端末で出力されるまでの時刻差の平均値である平均 MU 遅延 D_a, D_v を用いる．

本実験では，10 回のシミュレーションを行い，その結果の平均をアプリケーションレベル QoS パラメータの測定値とする．

2.2.5 学習効果の評価方法

被験者は 20 代の男女 50 人とした．各コンテンツに対する刺激としては，2.2.3 項で示したように，平均負荷量が 4.3Mb/s から 4.7Mb/s の 5 通りを用意した．ここで，刺激とは，心理実験において被験者に何らかの反応を生じさせる要因とするものである．本章においては，特定のコンテンツを用いて，定められた平均負荷量のトラヒックを与えたとき，メディア受信端末で再生されたものを刺激とする．

本実験では作業方法として，再生方式と選択方式を用いる．まず，ミュージックビデオでは，歌声と楽曲が聞こえるコンテンツを用いる．解答用紙には，歌声と同じ内容の言葉を，部分的に空白にして記載したものを用意する．空白文字数は全体文字数の約 42 % である．被験者は，解答用紙の空白部分に，視聴した歌声の言葉を，再生方式で記入する．

次に，漫才では，二人の会話が聞こえるコンテンツを用いる．解答作業は，ミュージックビデオと同様に，再生方式で行う．空白文字数は全体文字数の約 25 % である．

三つめに，競馬では，馬の名前を読み上げる場面のコンテンツを用いる．解答用紙には，馬の番号に対する馬の名前を書く解答欄を記したものを用意する．そのとき，馬の番号は，コンテンツで読み上げられる順に記載しておく．また，別紙で，選択肢として馬の名前を記載した用紙を用意する．被験者は，視聴した馬の名前を，選択方式で解答用紙に記入する．

なお，競馬においては，再生方式だと，被験者が事前に馬の名前を知っているかどうかで，学習効果の差が大きくなる可能性がある．そこで，被験者の予備知識の影響を少なくするために，選択方式を用いることにした．

被験者は，2.2.2 項で示した計 12 個（ミュージックビデオ，漫才，競馬各 4 個）のコンテンツすべてに対する刺激を，字幕なし・字幕ありの 2 種類について，19 インチ液晶ディスプレイで視聴して，作業する．

次に，今回の実験で被験者が行う作業手順を示す．

Step 1) 刺激の再生 (1 回目)

Step 2) 再生直後に解答作業 (1 回目, 制限時間は 30 秒間)

Step 3) 刺激の再生 (2 回目)

Step 4) 再生直後に解答作業 (2 回目, 制限時間は無制限で, 1 回目と同じ用紙に解答)

Step 5) 次の刺激がある場合は, Step 1) に戻る

視聴するコンテンツは, いずれも一度の視聴だけでは記憶が難しいため, 刺激を 2 回与えることにした [57]. 更に, 時間の制約で解答できなくなることを避けるため, 2 回目の解答時間は無制限とした.

被験者は, すべての刺激について, 上記の作業を行う. したがって, 字幕の有無も含めて計 24 回作業を繰り返すことになる.

学習効果を正確に求めるためには, 一人の被験者が同じコンテンツを何度も視聴することで, 作業内容を記憶することがあってはならない. そのため, 被験者はコンテンツ毎にいずれか一つの刺激で作業することにした. したがって, 各刺激に対して, 10 人の学習効果が得られることとなる.

被験者は刺激毎に解答作業を行う. それを採点した正答率を学習効果として扱う.

2.3 実験結果と考察

2.3.1 アプリケーションレベル QoS 評価結果

2.2.4 項で定義したアプリケーションレベル QoS パラメータ測定値の一例として, 漫画における, 平均負荷量に対する音声の MU 欠落率 L_a ビデオの MU 欠落率 L_v を, それぞれ, 図 2.3, 図 2.4 に示す. 95%信頼区間も同時に示している. 図において, 字幕の有無を on, off で示している. 例えば, manzail の場合は字幕ありを manzail-on, 字幕なしを manzail-off で表している. 以降の図においても, 同様に示すこととする.

図 2.3 と図 2.4 から, どの平均負荷量においても, MU 欠落率は, すべてのコンテンツでほぼ同じ値になっていることがわかる. これは, 音声・ビデオのビットレートを一定にしているためである. また, 平均負荷量の増加に伴って, MU 欠落率が増加していることがわかる. これは, ルータ間の帯域幅が 10Mb/s であるため, ルータにおいてパケット欠落が生じるからである.

2.3.2 学習効果の評価結果

平均負荷量に対する正答率を, コンテンツタイプ毎に図 2.5 から図 2.7 に示す. なお, 各測定値は, 10 人の正答率の平均値とし, 95%信頼区間も同時に示している.

また, コンテンツタイプ毎に 4 個のコンテンツの正答率を字幕の有無別で平均したものを表 2.2 と表 2.3 に示す. 更に, 字幕ありの正答率から字幕なしの正答率を引いた値をコンテンツタイプ毎に平均したものを正答率の差として, 表 2.4 に示す.

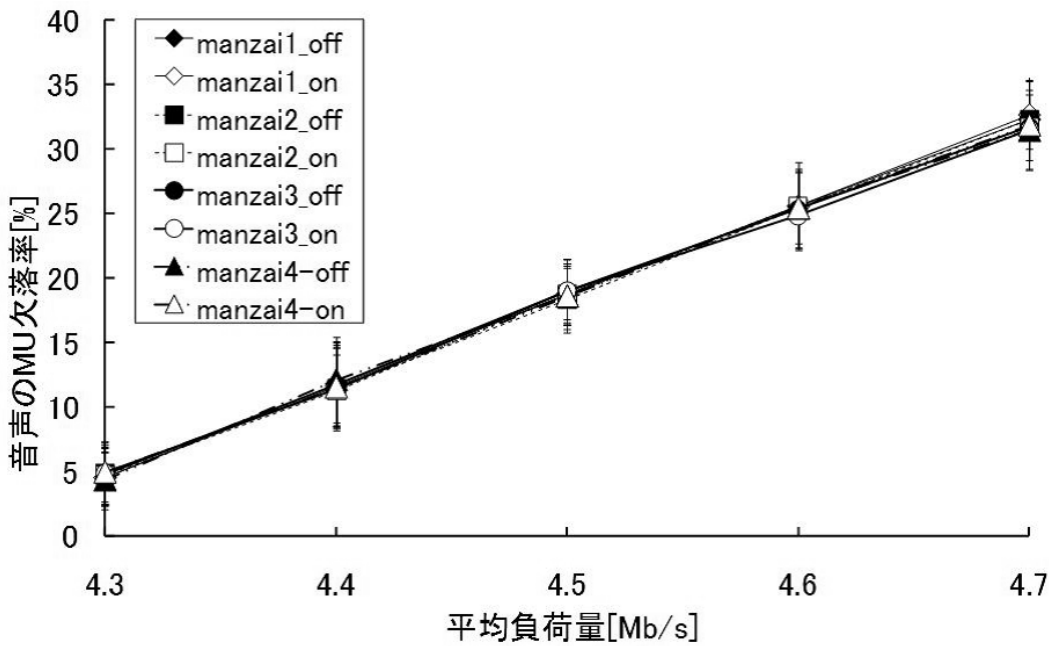


図 2.3: 音声の MU 欠落率 (漫才)

表 2.2: コンテンツタイプ毎の正答率 [%] (字幕なし)

平均負荷量 [Mb/s]	4.3	4.4	4.5	4.6	4.7
ミュージックビデオ	47.2	39.2	32.0	31.3	24.1
漫才	67.7	60.2	51.9	46.6	33.1
競馬	49.1	46.6	41.6	38.8	33.8

図 2.5 から図 2.7 より, すべてのコンテンツにおいて, 字幕なしに比べ, 字幕ありの正答率が高くなっていることがわかる。これは, 字幕が被験者の正答率向上に寄与していることを意味する。また, 字幕ありにおいては, 平均負荷量が増加しても, 全体的に正答率が高いままである傾向が見られる。これは, 映像に含まれる字幕はスクロールしないため, 字幕が映っている間の MU がすべて欠落しない限り, 字幕が映らないことはないためと考えられる。この結果から, ベストエフォート型のサービスを提供する IP ネットワークにおいては, 文字情報が正答率を向上させるのに効果的であるといえる。

次に, コンテンツタイプ間の比較をする。ミュージックビデオと漫才は, いずれも再生方式で評価されている。まず, 字幕なし同士の場合で比較すると, music4 を除き, ミュージックビデオよりも漫才の方が全体的に高い正答率である。これは, 字幕あり同士の場合で比較しても同様である。2.2.2 項で示したように, 漫才はミュージックビデオに比べて文字の提示時間が短くなっている。しかし, 漫才は前後の文脈から解答を推測でき, 記憶しやすい。そのため, 表 2.2 と表 2.3 からわかるように, 漫才の方がミュージックビデオより, 正答率が高くなっている。また, ミュージックビデオでは,

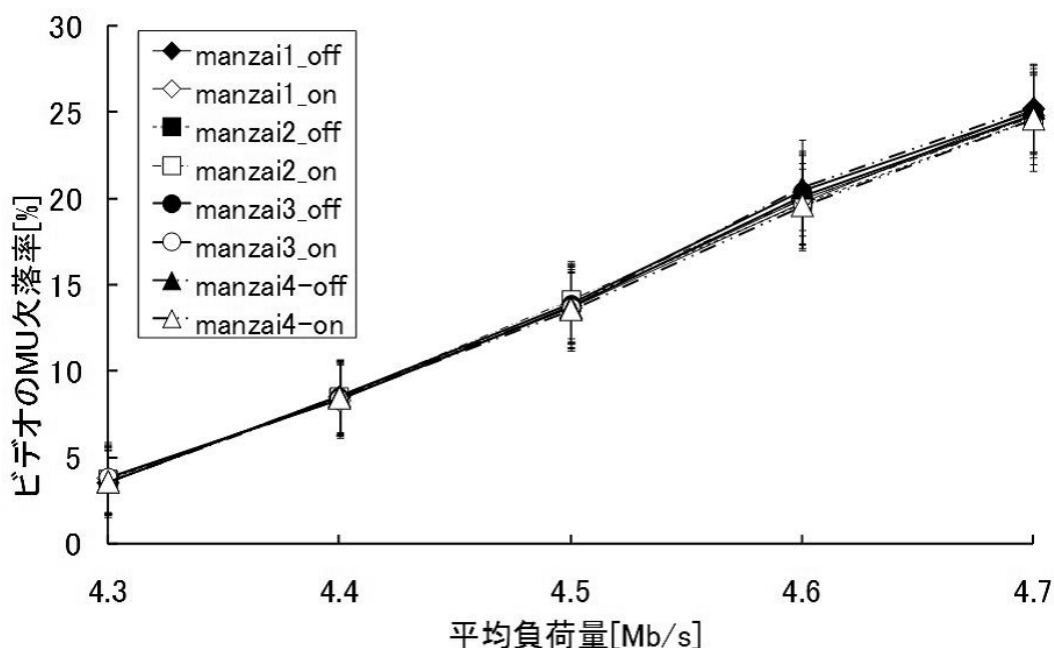


図 2.4: ビデオの MU 欠落率 (漫才)

表 2.3: コンテンツタイプ毎の正答率 [%] (字幕あり)

平均負荷量 [Mb/s]	4.3	4.4	4.5	4.6	4.7
ミュージックビデオ	83.4	78.2	77.9	73.8	71.3
漫才	89.7	89.9	87.9	86.4	85.4
競馬	89.7	88.8	83.8	81.6	81.6

歌声と楽曲の二つが同じ聴覚での処理を必要とする．そのため，正答率の低下に影響を及ぼしたと考えられる．

更に，同じコンテンツタイプ内でコンテンツ間の比較をする．まず，図 2.5 から，ミュージックビデオでは，コンテンツによって正答率の違いが大きいことがわかる．これは，楽曲のテンポ，歌手の歌い方などが異なるためであると考えられる．また，最も楽曲のテンポが速い music2 の方が，最も楽曲のテンポが遅い music4 より，字幕の有無による影響を大きく受ける．これは，音声だけでは記憶処理が間に合わないときほど，字幕によって記憶処理過程が補強されるからだと考えられる．

図 2.6 から，漫才では，平均負荷量が増大しても，字幕ありのときには，正答率があまり低くならないことがわかる．表 2.3 で見ると，平均負荷量の違いによる正答率の最大値と最小値との差は，わずかに 4.5 % しかない．ちなみに，ミュージックビデオでは 12.1 %，競馬では 8.1 % の差がある．この理由は，次のように考えられる．漫才は前述したように，記憶しやすいコンテンツである．平均負荷量が増大して，パケットの欠落，遅延及び遅延揺らぎが生じたとき，音声をうまく聞くことができない場合がある．

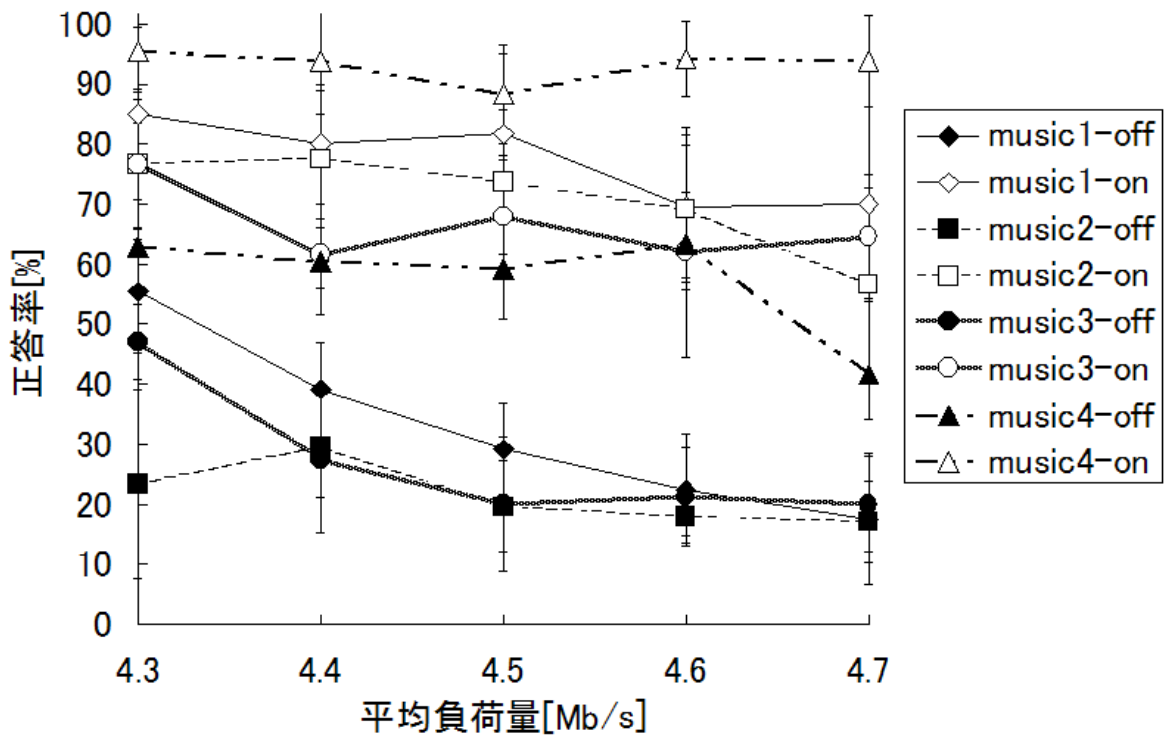


図 2.5: 平均負荷量に対する正答率 (ミュージックビデオ)

表 2.4: コンテンツタイプ毎の正答率の差 [%]

平均負荷量 [Mb/s]	4.3	4.4	4.5	4.6	4.7
ミュージックビデオ	36.3	39.1	45.9	42.5	47.2
漫才	22.0	29.7	36.0	39.8	52.3
競馬	40.6	42.2	42.2	42.8	47.8

しかし、そのときでも、文字を見ることは可能である。そのため、字幕ありの場合は、十分記憶が可能であり、平均負荷量が増大しても、正答率の低下は少ない。

図 2.7 から、競馬では、平均負荷量の増加に対する正答率の低下が、字幕なしのときでも、比較的少ない。表 2.2 で見ると、平均負荷量の違いによる正答率の最大値と最小値との差は、ミュージックビデオでは 23.1%，漫才では 34.6% に対して、競馬では 15.3% である。これは、馬の名前を選択方式で記述するため、一部が聞こえなくても正解できることがあるからだと考えられる。

コンテンツ間の比較から、次の二つのことがいえる。一つは、ミュージックビデオと競馬のように、文字情報がないときに正答率が低いコンテンツほど、正答率向上のために、文字情報を加えることが効果的である。

ミュージックビデオと競馬における正答率の差は、平均負荷量 4.7Mb/s の場合を除いて、漫才のそれより大きくなっていることが、表 2.4 からわかる。もう一つは、IP ネット

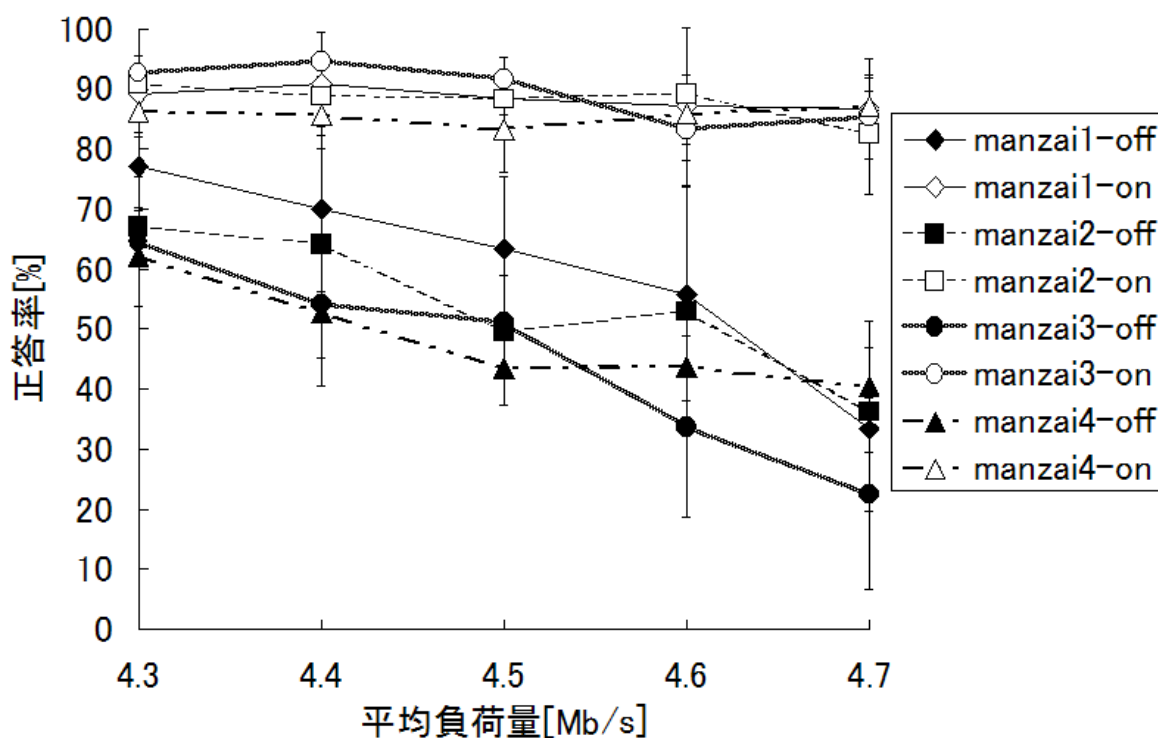


図 2.6: 平均負荷量に対する正答率 (漫才)

トワークにおいて，漫才のように，負荷が少ないときに正答率が高いコンテンツほど，負荷が増加したときに，文字情報を加える効果が大いことである．表 2.4 で平均負荷量が大きくなると，漫才の正答率の差は急激に大きな値になっている．

2.3.3 QoS マッピング結果

重回帰分析を用いて，コンテンツ毎に，アプリケーションレベル QoS パラメータと正答率とのマッピングを行う．これにより，正答率の推定式を求める．字幕の有無を式 (2.1) のようにダミー変数 S によって定義して，独立変数とする．

$$S = \begin{cases} 1 & (\text{字幕あり}) \\ 0 & (\text{字幕なし}) \end{cases} \quad (2.1)$$

2.2.4 項のアプリケーションレベル QoS パラメータとダミー変数を用いて，主成分分析を行った．累積寄与率は第一主成分で 77.5%，第二主成分で 87.2%，第三主成分で 95.1% となった．したがって，第三主成分までを用いれば，本実験で用いる QoS パラメータが保持している情報の 95.1% を表すことができる．よって第三主成分までを用いて QoS マッピングを行う．

QoS マッピングにおいて，従属変数は正答率とする．アプリケーションレベル QoS パラメータとダミー変数は，各主成分との相関の高さでグループ分けをした．グループ分けの結果を表 2.5 に示す．

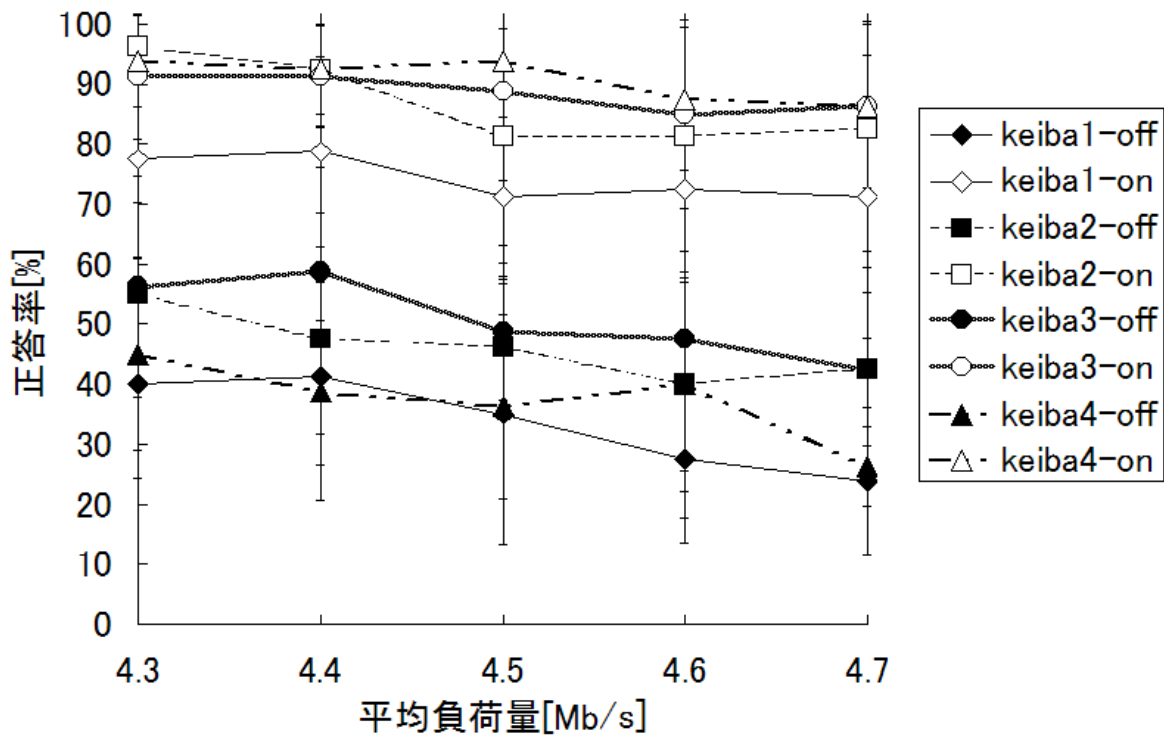


図 2.7: 平均負荷量に対する正答率 (競馬)

同じグループにあるアプリケーションレベル QoS パラメータ同士の組み合わせを除き、すべての組み合わせで重回帰分析を行った。そして、自由度調整済み重相関係数が最も高いパラメータの組み合わせを独立変数として採用した。

すべてのコンテンツに対し、t 検定の結果、重回帰式に寄与していない独立変数は取り除き、再度重回帰分析を行った。得られた重回帰式を、式 (2.2) から式 (2.13) に示す。

式において、 \hat{U} は正答率の推定値である。下添字でコンテンツの種類を表し、ミュージックビデオを $mu(i)$ 、漫才を $ma(i)$ 、競馬を $k(i)$ と記す。ただし、 i はコンテンツ番号である。また、 R^* を自由度調整済み重相関係数とする。

$$\hat{U}_{mu(1)} = 32.899 - 9.733C_a + 44.203S \quad (R^* = 0.973) \quad (2.2)$$

$$\hat{U}_{mu(2)} = 21.409 - 5.486L_v + 49.431S \quad (R^* = 0.987) \quad (2.3)$$

表 2.5: アプリケーションレベル QoS パラメータのグループ分け

グループ	パラメータ
A グループ	$R_a, R_v, L_a, L_v, E_a, C_a, C_v, E_{int}, D_a, D_v$
B グループ	E_v
C グループ	S

$$\hat{U}_{mu(3)} = 27.153 - 7.478D_a + 39.444S \quad (R^* = 0.974) \quad (2.4)$$

$$\hat{U}_{mu(4)} = 57.500 + 35.583S \quad (R^* = 0.941) \quad (2.5)$$

$$\hat{U}_{ma(1)} = 64.503 + 11.837E_a - 9.615E_v + 19.328S \quad (R^* = 0.968) \quad (2.6)$$

$$\hat{U}_{ma(2)} = 53.141 - 7.187E_{int} + 35.635S \quad (R^* = 0.972) \quad (2.7)$$

$$\hat{U}_{ma(3)} = 45.253 - 9.858L_v + 44.287S \quad (R^* = 0.962) \quad (2.8)$$

$$\hat{U}_{ma(4)} = 49.079 + 3.548E_v - 3.368E_{int} + 35.948S \quad (R^* = 0.982) \quad (2.9)$$

$$\hat{U}_{k(1)} = 33.527 - 4.866L_v + 40.695S \quad (R^* = 0.990) \quad (2.10)$$

$$\hat{U}_{k(2)} = 46.278 - 5.661D_a + 40.444S \quad (R^* = 0.993) \quad (2.11)$$

$$\hat{U}_{k(3)} = 50.748 - 4.150L_v + 37.754S \quad (R^* = 0.990) \quad (2.12)$$

$$\hat{U}_{k(4)} = 37.257 - 4.221C_a + 53.486S \quad (R^* = 0.993) \quad (2.13)$$

式(2.2)から式(2.13)のすべてに、正の係数を持つダミー変数 S が含まれている。このことから、字幕が正答率向上の要因になっていることがわかる。以下、コンテンツタイプ毎に考察を行う。

まず、式(2.2)から式(2.5)より、ミュージックビデオの場合を検討する。ミュージックビデオでは、全ての重回帰式において用いられる独立変数が異なっている。これは、コンテンツによって、楽曲のテンポ、歌手の歌い方などが異なるためであると考えられる。各平均負荷量に対する正答率が、他のコンテンツより低い music2 では、独立変数 L_v が要因となっており、画像の乱れが作業に影響を及ぼしている。また、各平均負荷量に対して正答率の変化が小さい music4 では、 S 以外のすべての独立変数が、 t 検定の結果、棄却されることとなった。これは、music4 が他のコンテンツに比べ楽曲のテンポが遅いため、作業を行う上で、負荷の影響が小さかったからだと考えられる。更に、music1, music3 については、独立変数として、それぞれ、 C_a , D_a が用いられている。これは、2.3.2 項で考察したように、歌声と楽曲が音声に含まれているので、音声の乱れが被験者の作業に影響を及ぼすためと考えられる。

次に、式(2.6)から式(2.9)より、漫才の場合を検討する。漫才では、manzai3 を除いたコンテンツにおいて、独立変数 E_v , E_{int} の一方又は両方が用いられている。これは、会話をするコンテンツを用いたことが要因と考えられる。文献[14]では、人間は映像内の動きのある対象に注視しやすく、映像のズームング及び切り替えに対しては、視点が移りやすいと述べられている。これにより、メディアの同期品質が低い場合、画像のぎくしゃく感が増し、被験者が映像に注意を奪われる。そのため、音声を用いて文脈を理解する処理過程が妨害され、作業に影響を及ぼすと考えられる。

三つめに、式(2.10)から式(2.13)より、競馬の場合を検討する。競馬では、独立変数 C_a , D_a と、独立変数 L_v が要因となっている。これらの独立変数が選ばれた理由として、評価方法である選択方式が考えられる。解答用紙には、読み上げられる馬の番号が順番に記してある。そのため、馬の番号を記憶するために、映像に集中して解答を行う場合と、馬の名前を記憶するために、音声に集中して解答を行う場合がある。したがって、ビデオの MU 欠落率、音声の遅延・遅延揺らぎに関する独立変数が選ばれたと考えられる。

平均負荷量に対する正答率の測定値と、式(2.2)から式(2.13)によって得られた推定

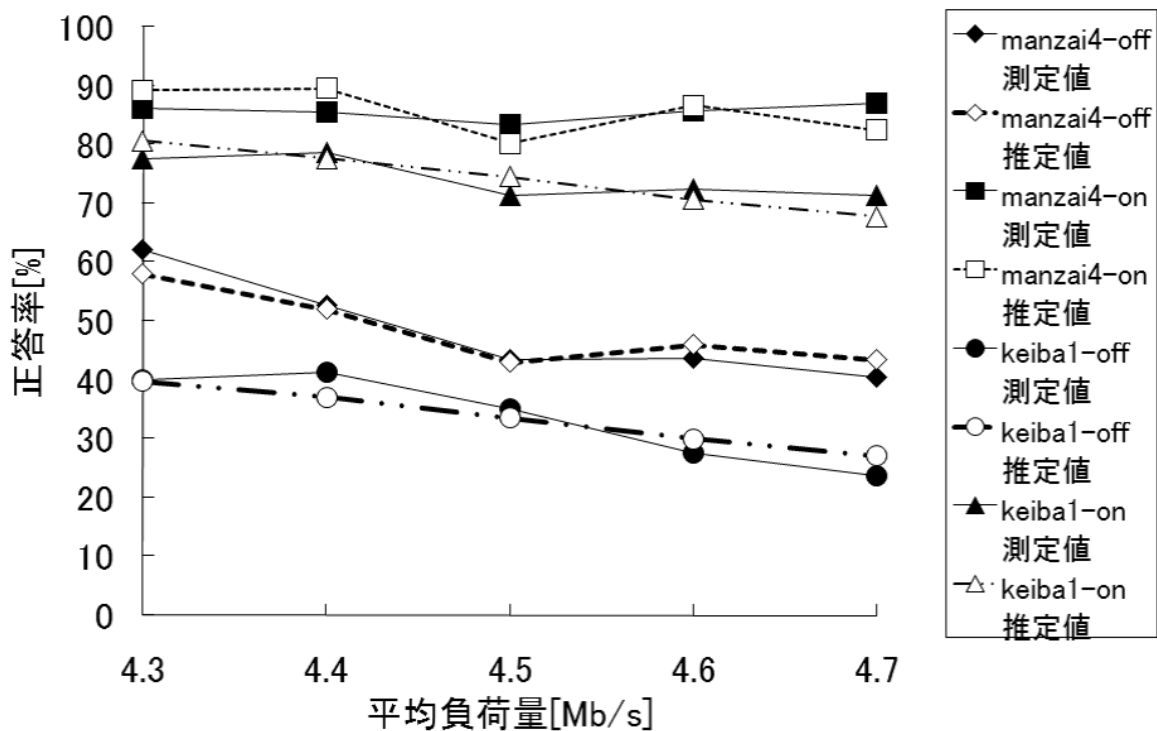


図 2.8: 平均負荷量に対する正答率の測定値と推定値

値の一部を，図 2.8 に示す．各コンテンツにおいて，推定値は測定値に近い値となっており，これからも重回帰式の適合性が確認できる．

以上より，すべてのコンテンツについて，文字情報が正答率向上に有効であることがわかった．また，コンテンツタイプによって，重回帰式に含まれる独立変数の傾向が異なることがわかった．

2.4 むすび

本章では，IP ネットワーク上での音声・ビデオ伝送において，文字情報の有無が視聴者の学習効果（正答率）に及ぼす影響を，実験により調査した．また，アプリケーションレベル QoS との QoS マッピングを行い，正答率の推定式を導いた．その結果，実験したすべてのコンテンツにおいて，文字情報が視聴者の正答率向上に影響を及ぼすことがわかった．また，IP ネットワークにおいて，文字情報を加えることで，平均負荷量が大きくなっても，正答率が低下しにくいことがわかった．特に，負荷が少ないときに正答率が高いコンテンツほど，その傾向が強かった．これらのことから，IP ネットワークにおいては，文字情報を加えることが，情報伝達の効率を向上させることがわかる．したがって，IP ネットワーク上でのコンテンツには，文字を積極的に使うことが有効であるといえる．

また，音声・ビデオ IP 伝送において，効果的に文字情報を加えるためには，次の 2 点に留意すべきである．一つは，欠落や遅延の影響を受けにくくするため，文字情報は

スクロールさせずに表示させる．もう一つは，文献 [52] にもあるように，音声と同じ内容を文字情報として表示させることである．

今後の課題として，IP ネットワーク上における遠隔学習用のコンテンツで，文字情報の有益性を検討することが考えられる．例えば，一度に提示する文字数，文字の提示場所等の違いが学習効果に及ぼす影響を検討することが挙げられる．

第3章 文字情報ベースの遠隔学習において音声・ビデオ品質が学習効果に及ぼす影響

3.1 まえがき

アクセス網の高速化や端末の高性能化の普及で、インターネット上での音声・ビデオを用いた多くのアプリケーションが一般的になっている。遠隔学習アプリケーションもこのカテゴリーに分類される。

遠隔学習において、受講者の環境に応じて、様々なメディアを利用できる [58]。例えば、テキスト、音声やビデオなどのメディアを組み合わせた教材が、Web ベースのアプリケーションで利用されている。1990 年代の終わりのアメリカでは、Web ベースのコースが公立ユニバーシティの 84.1 %，4 年生公立カレッジの 83.3 % で提供されている [59]。

遠隔学習アプリケーションはリアルタイム型とオンデマンド型の二つのタイプに分類できる [3][4]。リアルタイム型は、すべての受講者が異なる場所で同時に“出席する”オンライン学習である。このタイプは一般にテレビ会議システムを使い、講師とすべての受講者がスケジュールを合わせて講義に参加する必要がある。オンデマンド型は、受講者が自分のスケジュールで受講する教材（例えば、記録されたビデオ）にどこからでもアクセスできるオンライン配達方式である。受講者は同時に一緒にいる必要はなく、講義を繰り返し見ることができる。本章では、オンデマンド型の遠隔学習を扱う。

これまでに、多くの研究者が、書かれた文字、読み上げる文字、ビデオのようなマルチメディアを使った情報表現の利点を議論している [60]–[63]。Cowen は、一般的にテキストを読んだ人よりもフィルムを見た人の再生得点の方が高いことを示している [60]。Mayer と Moreno は、学生の理解度は、コンピュータベースのマルチメディア学習環境で、説明をナレーションとアニメーションで示す方が単にナレーションだけで示すよりも良いと述べている [61]。Guan は、動画付きの話し言葉が学習効率上で肯定的効果を与えることを示している [62]。Kim と Gilman は、学習者の英語の語彙学習を増強するために、ウェブベース独学プログラムにおいて、書き言葉、話し言葉、グラフィクスのようなマルチメディア要素の使用を調査しており [63]、研究の結論は、視覚のメディアの使用が語彙習得を援助して、学力得点を増やすのを提案している。ここで注意すべきは、これらの研究はローカル環境を扱っており、ネットワークを考慮していないことである。

文献 [55][64] は、通信品質が遠隔学習における学習効果に影響することを示している。

文献 [55] では、スクロールしている文がネットワーク負荷トラヒックに影響されるとき
のテストにおける正答率を調査している。しかし、これはテキストだけを扱っている。
文献 [64] では、IP ネットワーク上での音声・ビデオ伝送において、ビデオに加えられ
たテキスト情報が学習効果を更に良くすることを示している。ここでは、実験で、通信
品質が劣化しても、学習効果がテキスト情報のおかげで高く保つことを示している。し
かし、この研究は音声・ビデオ伝送を基本としたシステムであり、テキストをオプショ
ンとして用いている。

IP ネットワークにおける通信品質の劣化は受信側での音声・ビデオ品質を低下させ
る。しかし、様々な通信品質の条件において、テキスト、音声、ビデオの組み合わせか
らなる教材の学習効果を調査した研究は見当たらない。

本章では、IP ネットワーク上でのマルチメディア遠隔学習を扱い、様々な通信品質
において教材の最適な情報提示方法を調査する。“テキストだけ”、“テキストと音声”、
“テキスト、音声とビデオ”の3通りの情報提示方法を比較する。これら3通りの情報
提示方法において、遠隔学習における QoE である学習効果を実験により定量的に測定
する。

本章は次のように構成される。3.2 節で実験システムと評価方法を述べる。次に、3.3
節では実験方法と実験結果を示す。

3.2 遠隔学習実験

3.2.1 実験システム

実験システムの構成を図 3.1 に示す。本システムは、二つのルータと四つのパソコン
から構成されている。パソコンは、教材サーバ、学習端末、負荷送信端末、負荷受信
端末として使われる。ルータ 1 には、Cisco Systems 社製の Cisco2691 を用い、ルータ
2 には、同じ Cisco Systems 社製の Cisco7301 を用いる。ルータ 1 とルータ 2 の間の回
線は 10Mb/s の Ethernet であり、他のすべての端末または教材サーバとルータの間は
100Mb/s の Ethernet で接続されている。ルータ間がボトルネックとなるようにこれら
のリンクスピードを選んだ。

本実験では、テキスト、音声及びビデオから構成される教材コンテンツを用いる。コ
ンテンツのベースとしてテキストを使った。事前に文章を読み上げている講師の音声と
ビデオを撮影し、それらの音声・ビデオをコンテンツの部品として使った。実験におけ
るコンテンツの提示方法は次の通りである。

- 文字のみ
- 文字及び講師の音声
- 文字、講師の音声及びビデオ

本実験では、教材サーバ用に Apache2.2.6[65] を実装し、教材の伝送には HTTP1.1 を
用いた。学習者は、学習端末上で、Web ブラウザ (Internet Explorer 7) を利用した。

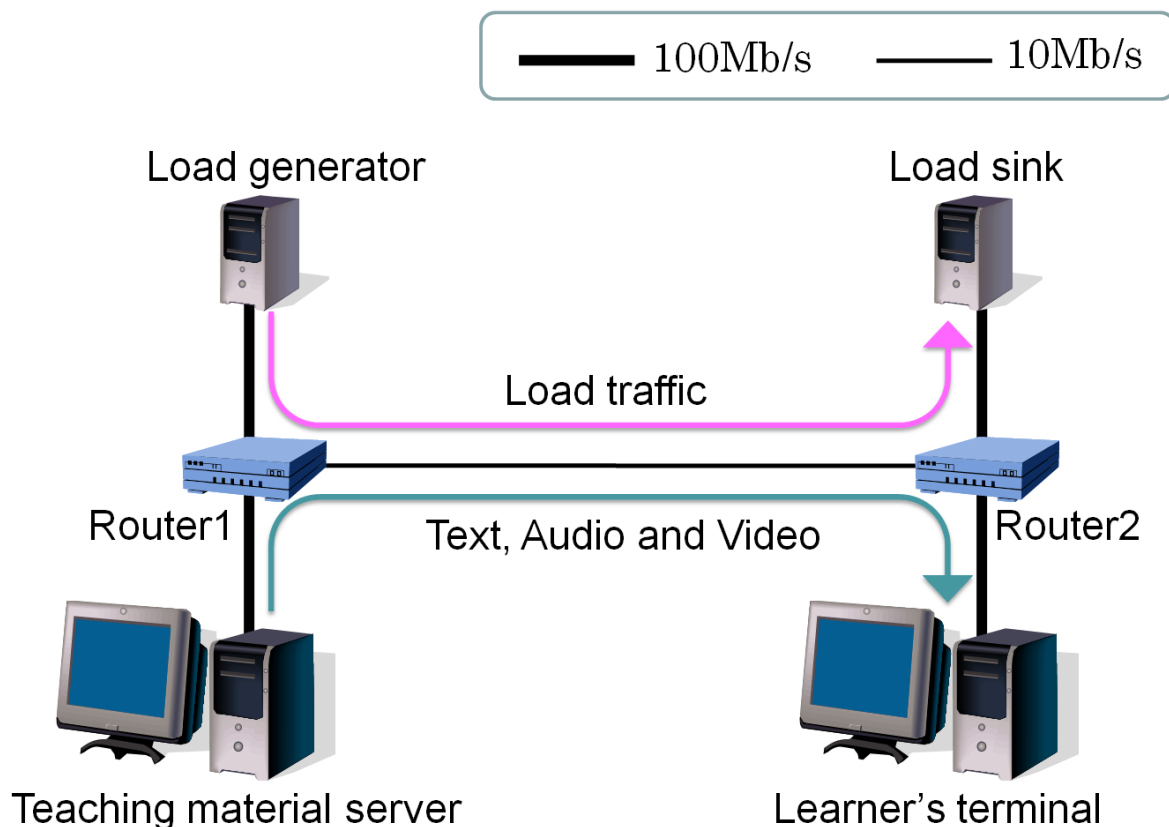


図 3.1: 実験システム構成

学習者の要求に応じて、教材サーバが学習端末へコンテンツを配信する。その際、トランスポートプロトコルとしてTCPを用いている。

学習端末で受け取られたコンテンツの品質を変化させるため、負荷送信端末は負荷受信端末へ負荷トラヒックを送信する。負荷トラヒックは負荷送信端末において、1480バイト固定長のUDPデータグラム（1500バイト長のIPパケット）を指数分布に従う間隔で発生させる。平均負荷量は、5.5Mb/s、6.0Mb/s、6.5Mb/s、7.5Mb/s、8.0Mb/s、8.5Mb/sの6種類とする。

学習端末にはヘッドホンと19インチの液晶ディスプレイが備えられている。液晶ディスプレイから被験者たちまでの距離は特に指定せずに、被験者たちが見やすい位置に調整する（例えば、50cmから1mの間）。

3.2.2 コンテンツ

教材の内容

この実験では、ネットワーク品質の影響が被験者の学習効果を支配できるように設計した。そこで、学習者に理解されやすい内容の教材を2種類選んだ。

本実験では、次の教材が用意された。

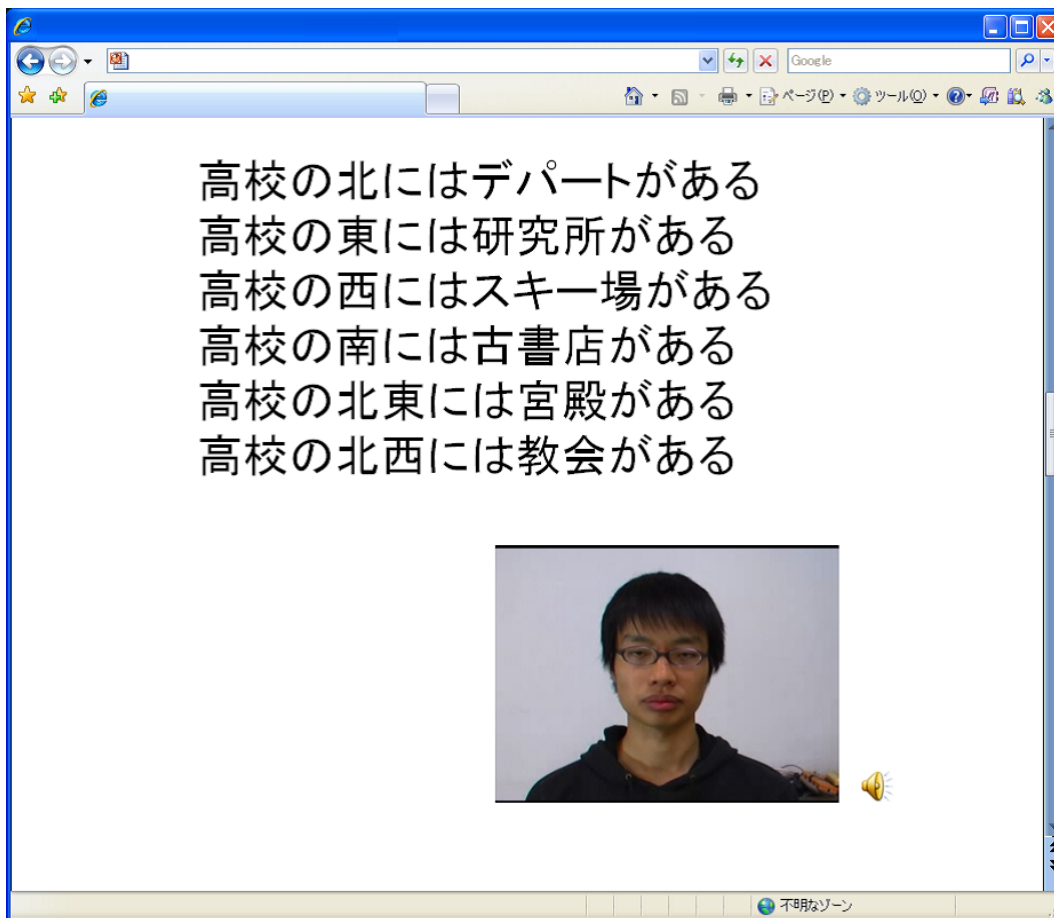


図 3.2: 教材 A の例

- 教材 A :
施設の配置情報を示す箇条書きにされた単純な文
- 教材 B :
随筆の抽出

教材 A のコンテンツの例を，図 3.2 に示す．教材 A において，4 個の短文と 6 個の短文の 2 種類の教材を用意した．便宜上，前者を“教材 A-4”とし，後者を“教材 A-6”とする．

教材 B のコンテンツの例を，図 3.3 に示す．教材 B における文は，大学入試センターが 2007 年に大学入試センター試験で使用した国語の問題文 [66] の一部である．

本実験では，学習者は，情報提示方法に対して様々な通信品質で教材コンテンツを評価する．そこで，被験者の負担を考慮して，文が異なっているコンテンツを 102 個準備した．各学習者はそれら 102 個のコンテンツを評価する．教材 A-4，教材 A-6 のコンテンツの数はそれぞれ 36 である．一方，教材 B は文章が長くて評価時間を長く必要とするため，コンテンツの数を 30 と少なく用意した．

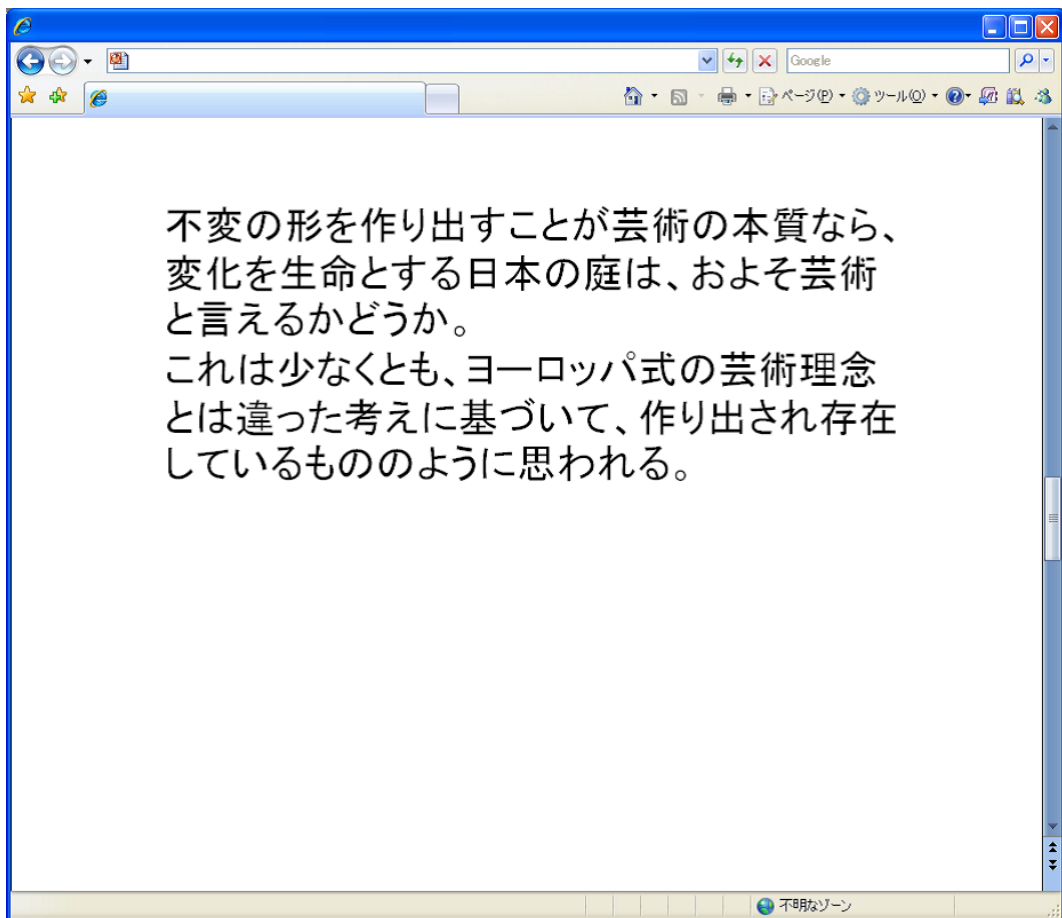


図 3.3: 教材 B の例

被験者に対する実験は、各コンテンツが一人の学習者にちょうど一度だけ提示されるようにデザインした。これは、学習者が前に提示されたコンテンツを覚えるかもしれないからである。

評価方法としては、偶然正解することを避けるため再生方式を選択する。単純再生法は、学習者が一度教材を見せられても、難しすぎて正答できない [57]。そのため、解答用紙にアウトラインが示された穴埋め再生テストを用いる。

正答率を学習効果の評価尺度として用いる。これは、正答のパーセントとして定義される。

一つのコンテンツから一つの穴埋めタイプの問題が作られた。従って、102 個の問題を用意した。各々の問題は、実験で各学習者に提示するために一枚の A4 用紙に印刷された。すべての被験者は、同じ問題を解答する。

図 3.2, 図 3.3 における教材コンテンツの問題を次に示す。

- 教材 A :

高校の北には () がある。

高校の北には () がある。

表 3.1: 音声・ビデオの仕様

仕様	音声	ビデオ
符号化方式	LinearPCM 48kHz 16bit ステレオ	MPEG-1
表示サイズ [pixel]	-	320 × 240
ピクチャパターン	-	IBBPBBPBB PBBPBBPBB
ビットレート [kb/s]	1536	800
フレームレート [frame/s]	-	30

高校の北には () がある .

高校の北には () がある .

高校の北には () がある .

高校の北には () がある .

● 教材 B :

() の () を () ことが () の () なら , () を () とする () の () は , およそ () と言えるかどうか .

これは少なくとも , () の () とは違った () に基づいて , 作りだされ () しているもののように思われる .

メディアの仕様

文字だけの場合の教材コンテンツのファイルサイズは 17~28kbyte である . 文字のフォントはすべて MS Pゴシックとした . また , 文字の大きさは , 教材 A ではすべて 32pt , 教材 B ではすべて 28pt とした . 表 3.1 に文字と一緒に伝送する音声・ビデオの仕様を示す . なお , 音声のビットレートは第 2 章と同じである . 一方 , 本章で扱うビデオは補完的なものであるため , 第 2 章で扱ったものより表示サイズを小さくし , ビットレートを低くしている .

表 3.2 は 3 種類の教材のコンテンツの提示時間を示す . 教材 A のコンテンツはおおよそ同じ提示時間である . 一方 , 教材 B のコンテンツの提示時間は変化する . これは , 教材 A のコンテンツのすべての文章は , おおよそ同じ長さであるが , 教材 B のコンテンツの文章の長さは , コンテンツごとに様々だからである .

表 3.2: 各教材におけるコンテンツの提示時間

教材 A-4	教材 A-6	教材 B
約 14 秒	約 18 秒	約 22 ~ 28 秒

教材の学習方法

今回の実験で，被験者は教材ごとに以下の手順で学習する．

Step 1) 教材のコンテンツを再生させるために，被験者は WEB ページのリンクをクリックする．文字，講師の音声・ビデオは，同時に開始するように設定する．教材コンテンツの文字は自動的に 1 文字ずつ表示する．文字の表示速度は音声・ビデオにおおよそ合わせている．表示された文字はしばらく消えることがない．コンテンツのすべての文字が表示された後，約 1 秒で一斉に消える．文字は画面の上部に表示する．講師のビデオを表示する場合は，それはテキストの右下に置かれる．

Step 2) 教材のコンテンツが消えた後に，被験者は印刷されたそのコンテンツの問題を見る．

Step 3) 被験者は時間の制約なしでそのコンテンツの問題に解答する．

Step 4) 次の教材コンテンツがある場合は，Step 1) に戻る．

便宜上，“文字のみ”を提示する場合を“case 1”，“文字及び講師の音声”を提示する場合を“case 2”，“文字，講師の音声及びビデオ”を提示する場合を“case 3”と呼ぶ．図 3.2 は“case 3”，図 3.3 は，“case 1”に相当する．

3.2.3 学習効果の評価

この実験の被験者数は 20 代の男性 17 名である．各々の被験者（つまり，学習者）が学習端末で再生されたコンテンツを評価する．学習端末で再生されたコンテンツが刺激とみなされる．

教材 A-4，教材 A-6 からどのように刺激を作ったかを説明する．それぞれの教材に対して用意した刺激の数は，648 個（ $=36 \times 18$ ）である．それらの刺激を， $C_{i,j}$ で示す．なお，下付き文字の $i = \{1, 2, \dots, 36\}$ はコンテンツ番号を表し，下付き文字の $j = \{1, 2, \dots, 18\}$ は刺激の作り方を示す番号である．刺激の作り方は 18 通りある．なぜなら，情報提示方法が“case 1”，“case 2”，“case 3”の 3 通りあり，すでに 3.2.1 項で述べたように，メディアトラヒックに影響している平均負荷量が 6 通りだからである．

648 個の刺激の中で，実験で 612（ $=36 \times 17$ ）の刺激，すなわち，学習者ごとに 36 個の刺激を使った．これは，各学習者が一度だけコンテンツを評価する取り決めをしたか

らである．つまり，学習者は，同じコンテンツに対して，異なる情報提示方法や異なる負荷トラヒックで学習することが決してなかった．

学習者が評価した刺激は $C_{1,j}, C_{2,j}, \dots, C_{36,j}$ である． j の値は学習者ごとに異なり，一人目の学習者の場合は， $j = 1, 2, \dots, 18$ であり，二人目の学習者の場合は， $j = 2, 3, \dots, 18, 1$ であり， n 人目の学習者の場合は， $j = n, n+1, \dots, 18, 1, \dots, n-1$ と j の値は周期的な順で変化する．学習者が 17 名なので， $C_{i,18}, C_{i,1}, \dots, C_{i,17}$ の刺激は使わなかった．

教材 B での刺激の作り方も同様である．刺激の数は 450 個 (30×15) である．再び，刺激を $C_{i,j}$ で示す．ここでも，下付き文字の $i = \{1, 2, \dots, 30\}$ はコンテンツ番号を表し，そして，下付き文字の $j = \{1, 2, \dots, 15\}$ は刺激の作り方を示す番号である．教材 B では，平均負荷量を 5.5Mb/s を除いた 5 通りとする．それゆえ，各コンテンツは，15 の刺激ができる．

教材 B の実験において，450 種類の刺激を使った．各々の学習者は， $C_{1,j}, C_{2,j}, \dots, C_{30,j}$ を評価した．一人目の学習者の場合は， $j = 1, 2, \dots, 15$ であり，二人目の学習者の場合は， $j = 2, 3, \dots, 15, 1$ であり， n 人目の学習者の場合は， $j = n, n+1, \dots, 15, 1, \dots, n-1$ である．16 人目と 17 人目の学習者は，それぞれ一人目の学習者，二人目の学習者のための刺激と同じものを評価した．

総評価時間は一名につき約 80 分であった．

3.3 実験結果と考察

3.3.1 正答率の推定値

教材 A-4，教材 A-6，教材 B の結果を図 3.4，図 3.5，図 3.6 にそれぞれ示す．各々の図は“case 1”，“case 2”，“case 3”における平均負荷トラヒックに対する正答率を示している．

実験結果における傾向を明確にするために単回帰分析を使う．単回帰分析では，平均負荷量が独立変数であると考えて，正答率が従属変数であると考えて．

正答率の推定値を $S_{X(Y)}[\%]$ で示し，下付き文字の $X = \{A-4, A-6, B\}$ は，それぞれ，教材 A-4，教材 A-6，教材 B を表す．同様に， $Y = \{case1, case2, case3\}$ は情報提示方法を表す．また， L は平均負荷量 [Mb/s] を示す．式 (3.1) から式 (3.9) は，このようにして得られた単回帰式を示す．

$$S_{A-4(case1)} = 78.6 + 0.420L \quad (3.1)$$

$$S_{A-4(case2)} = 115.7 - 4.78L \quad (3.2)$$

$$S_{A-4(case3)} = 137.1 - 7.83L \quad (3.3)$$

$$S_{A-6(case1)} = 66.0 - 0.770L \quad (3.4)$$

$$S_{A-6(case2)} = 88.6 - 3.92L \quad (3.5)$$

$$S_{A-6(case3)} = 122.4 - 8.93L \quad (3.6)$$

$$S_{B(case1)} = 44.7 - 0.105L \quad (3.7)$$

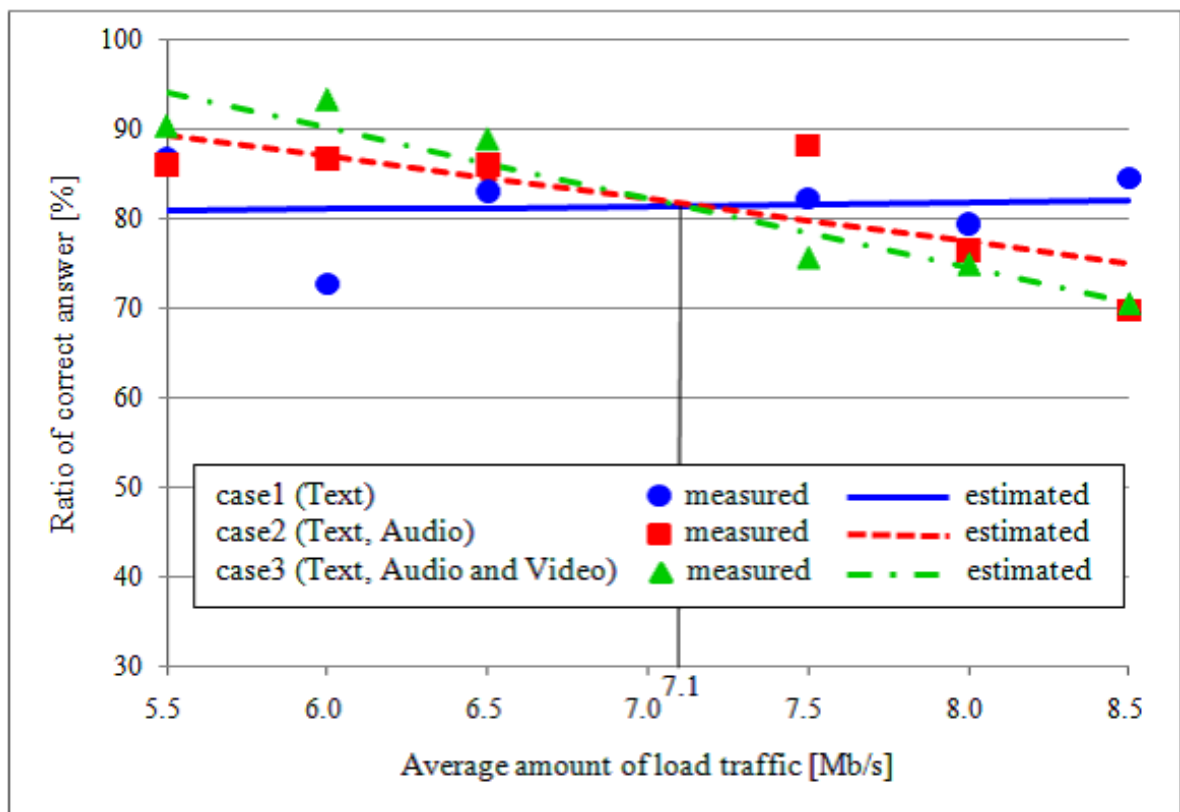


図 3.4: 平均負荷量に対する正答率 (教材 A-4)

$$S_{B(case2)} = 46.4 + 0.062L \quad (3.8)$$

$$S_{B(case3)} = 86.8 - 5.92L \quad (3.9)$$

これらの回帰式を，図 3.4 から図 3.6 に合わせて示す．

3.3.2 学習効果の評価結果

始めに，三つの情報提示方法がいかに正答率に影響するかを検討する．図 3.4 から図 3.6 より，“case 1”では，どの平均負荷量に対しても，正答率がほとんど変わらないことに気づく．これは，式 (3.1)，式 (3.4)，式 (3.7) において， L の回帰係数値が小さいことからわかる．文字のみの教材は，少ないデータ量を送信するので，負荷トラヒックの影響をほとんど受けなかったからである．平均負荷量が低ければ，“case 2”における正答率の値と “case 3” における正答率の値が “case 1” におけるそれよりも高い．これは，講師の音声・ビデオが，文字での記憶を補強するからである [60]–[63]．しかし，平均負荷量が増加すれば，“case 2” と “case 3” では，正答率が減少している．これは，音声品質もしくはビデオ品質の劣化が学習者の集中力と文字の記憶を妨害したためである．

次に，正答率の点から教材を比較する．

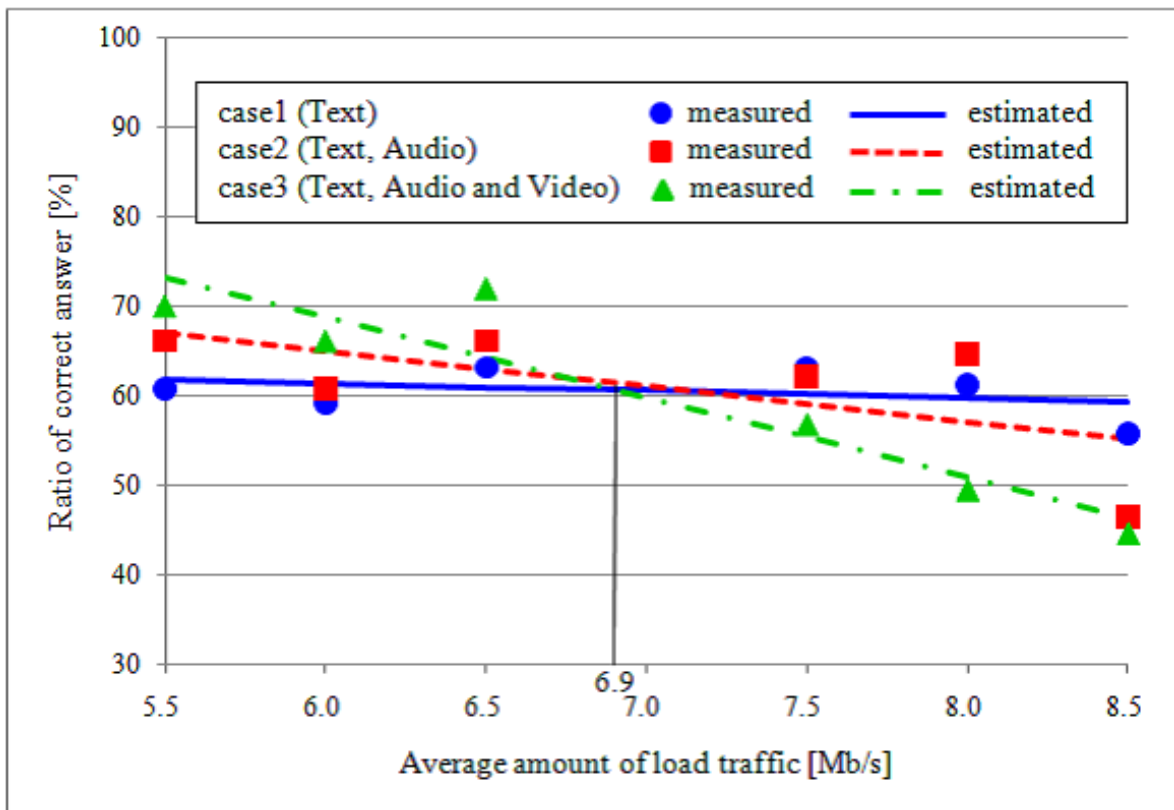


図 3.5: 平均負荷量に対する正答率 (教材 A-6)

図 3.4 と図 3.5 では、教材 A-6 の正答率が教材 A-4 の正答率よりも全体的に低い。このことは、式 (3.1) から式 (3.6) の切片の項からもわかる。これは、被験者は六つの短文からなる教材 A-6 を覚えるのに難しかったからである。特に、平均負荷量が増加すれば、“case 3” では教材 A-6 の正答率は教材 A-4 の正答率に比べて急激に低くなる。このことは、式 (3.3) の L の回帰係数値よりも、式 (3.6) の値が大きいことからわかる。これは、教材 A-6 のコンテンツは教材 A-4 より複雑なため、音声・ビデオ品質の劣化が学習者の集中力と文字の記憶を大きく妨害するためである。

図 3.4 と図 3.6 を参照して、教材 A-4 の正答率と教材 B の正答率とを比較する。そこから教材 B の正答率が教材 A-4 の正答率よりも全体的に低いことを気付く。これは、学習者は文章が長い教材を覚えるのが難しいからである。一方、平均負荷トラフィックの影響による教材 B の正答率の低下率は、教材 A-4 のそれより低い。教材 B のコンテンツは意味につながりがある文章である。そのため、文脈が被験者の文章間のギャップを埋める。

式 (3.1) から式 (3.9) を用いて、実験した教材ごとに、負荷トラフィック量に対して最適な情報提示方法のガイドラインを示す。図 3.4 に見られるように、教材 A-4 では、式 (3.1) から式 (3.3) での回帰式の交点がおおよそ 7.1 Mb/s である。従って、平均負荷量が約 7.1 Mb/s 以下の場合には “case 3” を用いるべきであり、他の場合は “case 1” を用いるべきである。同様に、教材 A-6 では、式 (3.4) と式 (3.6) の交点がおおよそ 6.9 Mb/s

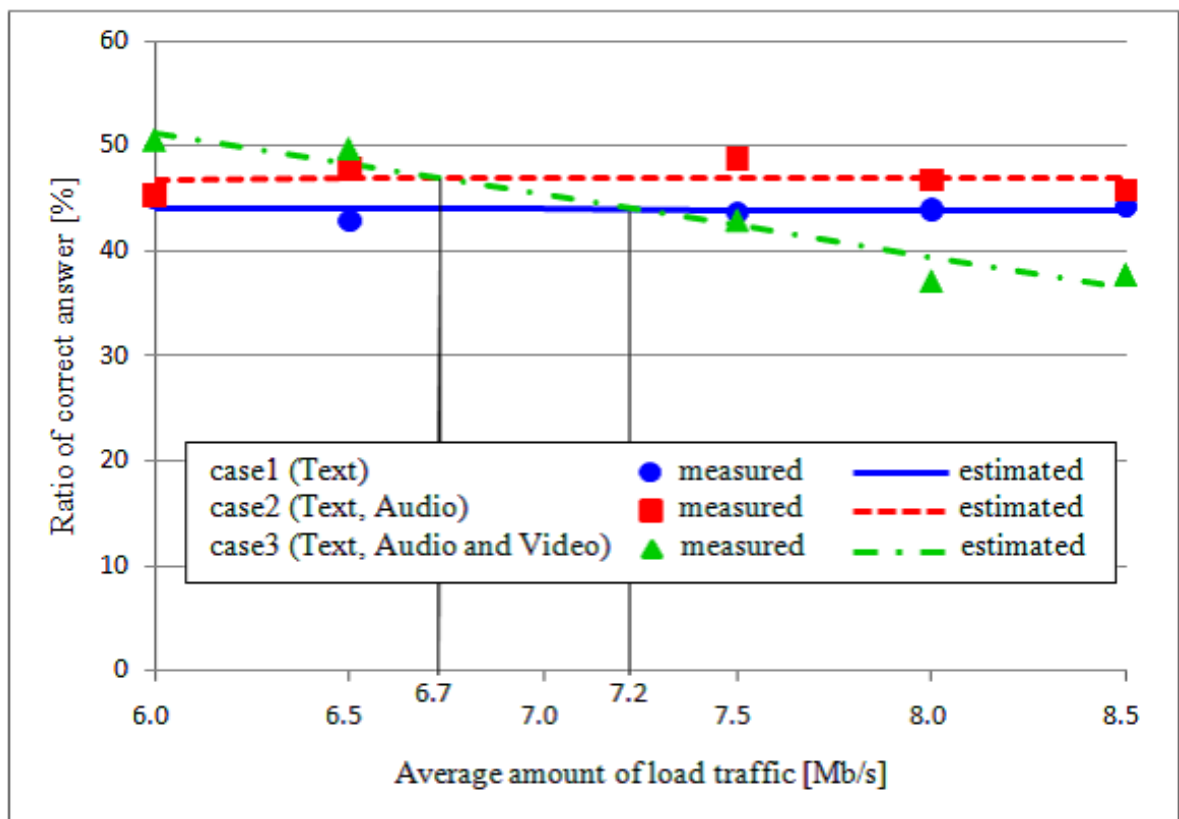


図 3.6: 平均負荷量に対する正答率（教材 B）

を示す．従って，平均負荷量が約 6.9 Mb/s 以下の場合には“case 3”を用いるべきであり，他の場合は“case 1”を用いるべきである．一方，教材 B では，平均負荷量が約 6.7 Mb/s までの場合は“case 3”を用いるべきであり，他の場合は“case 2”を用いるべきである．これらは，あくまで今回の実験環境でのみ示される一例であるが，別の教材に対しても，同様のやり方で最適な情報提示方法のガイドラインを示すことが可能である．

3.4 むすび

IP ネットワーク上でのマルチメディア遠隔学習において，通信品質の変化に対する最適な情報の提示方法を実験により調査した．その結果，次のことがわかった．負荷トラフィックがあるレベルまでの通信品質劣化を引き起こしたとき，文字に加えた音声・ビデオが学習者の学習効果を高める．しかし，劣化が多すぎると，補完的に送られた音声かビデオのいずれかの劣化が，逆に学習効果を低下させる．この傾向は，記憶が難しい教材で目立つ．更に，回帰分析により，負荷トラフィック量に対して最適な情報提示方法のガイドラインを示した．

今後の課題としていくつかの解決すべき問題がある．例えば，本実験では，教材の文字を一文字ずつ表示したが，文字の提示方法が他の方法を扱う必要がある．また，別の教材での学習効果や学習者の学歴や技能が正答率に及ぼす影響を検討するつもりで

ある .

第4章 多地点IP通信によるリアルタイム型遠隔対話環境におけるQoE

4.1 まえがき

遠隔地を双方向に結ぶテレビ会議システムは、移動時間と場所の制約を取り除くので、教育の分野でも利用されている。古くには、1981年に、東京工業大学の二つのキャンパス間を光ケーブルで結んで遠隔講義が行われている [67]。また、1999年には、京都大学とUCLAの間をATM回線で結んで約三カ月にわたって遠隔講義を行った例がある [68][69]。しかし、これらは専用のシステムであるので、多額の設備費用が必要であり、一部の高等教育機関においてのみ行われたものである。

近年、初等中等教育機関においても、ICT環境の整備とともに、テレビ会議システムを利用した遠隔授業が行われるようになってきている。例えば、文献 [70] は、ISDN回線により音声・ビデオをリアルタイムに伝送して、日本と中国における中学生同士の国際交流学習の実践例を示している。また、文献 [71] では、IPネットワークにより高精細動画を伝送して、中学生に対する遠隔授業の実践例を示している。

多地点テレビ会議システムは、複数個所を同時接続できるので、多地点接続の遠隔授業を可能にする。多地点接続の遠隔授業には、共同学習や交流学習などが挙げられる。共同学習は、自分たちとは考え方や習慣の異なる環境の人たちと取り組む場合に特に有効である。例えば、離島の子供たちは普段少人数環境にいるため、テレビ会議システムを用いた複数個所との共同学習は有意義である。文献 [72] では、USBカメラによる簡素なテレビ会議システムを活用して、離島にいる三校の小学生同士をISDN回線で結び、食文化に関する共同学習の実践例を示している。また、文献 [73] では、IPネットワークを利用した大学生同士の国際交流学習において、三地点接続の遠隔授業の方が二地点接続のものより教育効果が高いことを示している。

IPネットワークを用いれば、容易に低コストで多地点接続を実現できるため、テレビ会議システムにIPネットワークを用いることが多くなってきた。しかし、IPネットワークはベストエフォート型のサービスを提供するため、通信時に、パケットの欠落、遅延及び遅延揺らぎが生じる。特に、テレビ会議で扱う音声・ビデオは、時間的な構造をもつメディアであり、IPネットワークにおける通信品質の劣化による影響が大きい。また、通信相手が複数ある多地点接続の場合には、通信対ごとに通信品質が異なる問題がある。これは会議参加者の間に不公平性が生じる原因となる。

パケットの欠落、遅延及び遅延揺らぎによる音声・ビデオの品質劣化を防ぐために、受信端末のアプリケーション層においてバッファリング制御を行うことが有効である。バッファリング制御とは、受信端末が受け取ったデータを一時的にバッファに格納して、

遅延の揺らぎを吸収することによって、乱された時間構造を回復するものである。しかし、バッファリング時間を大きく設定すると、端末間通信の即時応答性が損なわれる。また、各端末が独自にバッファリング時間を設定すると、端末間同期品質 [74] が低下する可能性が生じる。端末間同期とは、マルチキャスト配信するときに、複数端末間におけるメディアの出力時刻を揃えることである。端末間同期品質は、それらの出力時刻差で表される。

IP ネットワーク上での多地点接続では、各通信対において通信品質が異なる場合が生じる。これは、ネットワーク品質の違いや、端末の性能差、音声・ビデオの符号化方式などに起因する。文献 [75] では、通信相手ごとの遅延差とそのパケット欠落率の違いがシステム全体の総合品質に対する影響を調べている。しかし、固定遅延と一定のパケット欠落率がそれぞれ独立にネットワークに与えられており、遅延揺らぎによる影響も考慮されていない。

IP ネットワーク上でのテレビ会議システムの遠隔授業に関する研究はいくつかあるが、その多くは受講者の主観評価による検討を行っているのみである。その中で、文献 [76] では、伝送動画の画質を主観評価するとともに、ビットレートの変化や伝送動画の遅延による影響も検討している。しかし、ここで扱う授業は、通訳を挟むものであったため、遅延の影響が目立たないという結論になっている。

文献 [77] では、多地点 IP 通信によるテレビ会議システムにおいて、受信端末においてバッファリング制御を行った場合の主観品質を評価している。多地点接続による主観品質は、各通信対に関する主観品質の影響を受ける。文献 [77] では、各通信対に関する主観品質の悪い方が総合的な主観品質に強く影響を及ぼすことをマッピングにより明らかにしている。しかし、受信メディアの品質と伝搬遅延のみを取り上げており、端末間同期品質を考慮していない。

テレビ会議システムを利用した遠隔授業では、講義者と受講者、または、受講者同士でリアルタイムに対話のやり取りがある。その際、受信端末がバッファリング制御を行うことで、対話が円滑に行われぬ可能性がある。例えば、バッファリング時間を大きく設定したときに、通信品質が向上して主観品質が良くなっても、対話に要する時間が他の端末よりも悪くなっている。そこで、主観品質の他に返答時間を含めた検討をする必要がある。インターネット上でのテレビ会議システムを用いた遠隔対話環境一般においても、主観品質とともに返答時間を扱った研究はこれまで見られない。

そこで、本章では、多地点 IP 通信によるリアルタイム型遠隔対話環境において、バッファリング制御が QoE に影響を及ぼす要因を調べる。QoE 尺度として、主観品質と返答時間を用いる。そして、多地点接続において効果的なバッファリング制御方法を定量的に議論する。

以下では、4.2 節では実験システムと評価方法を述べる。次に、4.3 節で実験結果と考察を示す。

4.2 遠隔対話環境実験

4.2.1 対話内容

本章では，多地点接続による遠隔対話環境の第一歩として，三地点接続を考える．したがって，三台の端末で実験を行うものとする．これらの端末では，音声・ビデオが生成され，マルチキャスト配信される．また，端末ごとに特定の一人だけが対話する場面を想定して，各端末には，それぞれ一人ずつの被験者を配置する．

本章で扱う対話は，遠隔学習を想定したものである．しかし，返答時間を扱う際，被験者の能力差がこれに影響することは望ましくない．つまり，1.5.3 項における (3) に要する時間が，できるだけ同一になる対話を選ぶ必要がある．また，被験者を主観評価に集中させるため，被験者にかかる負担が小さいことも同時に要求される．したがって，対象とする対話は，質問に対する思考を要するものではなく，即答できる簡単なものとする必要がある．

通信の国際標準化機関である ITU-T は，音声・ビデオのインタラクティブ通信のための作業方法を勧告 P.920[78] で示している．これは，音声・ビデオ通信の遅延品質や画像品質の評価方法を規定している．リアルタイム型遠隔対話環境においては，相手との対話を扱うため音声が重要視されることが多い．そこで，本章においてもこのことを考慮して，音声の遅延が与える影響が重要であり，かつ公平性が必要となる作業を定めた．

対話内容は，次の通りである．被験者は三名で一組のグループとなり，一名が講師，残り二名が受講者となる．講師は講師用端末に，受講者は受講者用端末 1，受講者用端末 2 にそれぞれ配置される．講師は，連続した三つの数字である「1, 2, 3」, 「4, 5, 6」, 「7, 8, 9」の中からランダムに一組を選び読み上げる．受講者は，三つの数字が聞こえ終わった直後に，聞こえたものを復唱して解答とする．講師は，どちらの解答が早く聞こえたのか，あるいは同時に聞こえたのかを各受講者に知らせる．

4.2.2 実験システム

実験システム構成を図 4.1 に示す．実験で用いるネットワークは，三台の端末と一台のマルチキャストルータ，一台の NTP (Network Time Protocol) サーバ，二台の負荷送信端末，二台の負荷受信端末，三台の HUB (10BASE-T) 及び三台のスイッチング HUB (100BASE-TX) から構成されている．三台の端末は，講師用端末一台と受講者用端末が二台である．各端末やルータ間は，Ethernet で接続されている．図 4.1 において，太い線は 100Mb/s，細い線は 10Mb/s で接続されていることを示している．

本実験では，テレビ会議システムのアプリケーションとして，ビデオ伝送用のアプリケーションである vic[79] と音声伝送用のアプリケーションである rat[80] を用いる．

本実験では，返答時間の正確な測定を可能にするため，受講者が講師の質問内容を的確に理解できるようにする．講師用端末に配置される講師の音声・ビデオの品質を良好に保つために，スイッチング HUB を用いる．負荷送信端末 1 に接続されているスイッチング HUB のスイッチング機能により，負荷送信端末 1 から伝送される負荷はマルチ

表 4.1: 音声・ビデオの仕様

仕様	音声	ビデオ
符号化方式	LinearPCM 16bit	H.261
表示サイズ [pixel]	-	325 × 288
平均 MU サイズ [byte]	320	2688
平均 MU レート [MU/s]	100	30
平均ビットレート [kb/s]	256	645

キャストルータへは流れず，HUB を経由して負荷受信端末 1 及び受講者用端末 1 へ流れる．HUB は半二重通信を行うため，受講者用端末 1 から伝送される音声・ビデオの品質は負荷トラヒックとの衝突により劣化する．一方，講師用端末から伝送される音声・ビデオは負荷と同方向に流れるため，衝突が少なく負荷の影響をあまり受けない．同様に，受講者用端末 2 から伝送される音声・ビデオは負荷送信端末 2 から伝送される負荷トラヒックとの衝突により劣化する．

本実験において，各端末に一人の被験者が配置され，前項で規定した対話を行う．各端末には 19 インチ・解像度 SXGA (1280 × 1024) の液晶ディスプレイが搭載されている．各端末に接続されたビデオカメラで被験者の顔を映して，それぞれの対話者にビデオとして伝送する．伝送されたビデオは，図 4.1 の端末部分に示すように，それぞれの対話者のディスプレイにウインドウ形式で表示される．音声はヘッドセットのマイクから入力したものを伝送する．雑音による聞き洩らしを防ぐため，ヘッドセットのヘッドホンで音声を聞く．これらの音声・ビデオの仕様を表 4.1 に示す．なお，本章では，実験の簡単さからこのような仕様を用いる．現在主流となっている H.264 等の符号化方式を用いることは今後の課題である．

本章では，受信バッファリング制御にはプレイアウトバッファリング方式 [81] を用いる．以下に，その概略を説明する．音声・ビデオの転送開始後，端末は，最初に受信した MU (Media Unit) の到着時刻に固定のバッファリング時間を加えたものを，その MU の目標出力時刻とする．MU とは，アプリケーション層での転送単位であり，音声では一定個数のサンプルが，ビデオでは 1 フレームがそれに相当する．その後の MU の目標出力時刻は，最初に受信した MU の目標出力時刻に，二つの MU のタイムスタンプの差を加えたものとする．目標出力時刻までに到着した MU は，目標出力時刻に出力される．一方，目標出力時刻を超えて到着したものは，廃棄される．以上の制御を，MU の送信端末や音声・ビデオを区別して行う．

各端末でのバッファリング時間と負荷の量を種々変更した環境で，被験者は評価を行う．

本章では，多地点 IP 通信によるリアルタイム型遠隔対話環境において，バッファリング制御が主観品質と返答時間に影響を及ぼす要因を調べる．そのため，各端末間でのバッファリング時間が異なるものも評価対象として必要となる．そこで，議論の簡単な

表 4.2: 平均負荷量の組み合わせ

	負荷送信端末 1	負荷送信端末 2
負荷パターン 1	6.5Mb/s	6.5Mb/s
負荷パターン 2	6.5Mb/s	8.0Mb/s
負荷パターン 3	8.0Mb/s	8.0Mb/s

ため、講師用端末では、目標出力時刻の決定に用いる固定バッファリング時間を一定に設定する。バッファリング時間は事前に行った予備実験を参考にして、負荷を加えない場合に比較的良好な音声・ビデオ伝送が可能な 100ms とする。一方、受講者用端末 1 と受講者用端末 2 では、バッファリング時間として 100ms, 200ms, 400ms, 600ms のいずれかを独立に設定する。これらのバッファリング時間も予備実験を参考にして、後述する値の負荷を加えたときに提示されるメディアの品質の違いがわかるように選んだ。したがって、バッファリング時間の組み合わせは $4 \times 4 = 16$ 通りとなる。なお、各端末はそれぞれ NTP サーバに接続されており、1ms 以下の精度で時刻同期を行うことができる。

また、負荷送信端末は、対応する負荷受信端末へ負荷を送信する。負荷として、ペイロード長が 1480 バイトである UDP データグラム (1500 バイト長の IP パケット) を指数分布に従う間隔で発生し、これを送信する。

負荷は、平均負荷量が 6.5Mb/s, 8.0Mb/s の 2 通りとし、それぞれ低負荷、高負荷と定義する。負荷送信端末 1, 負荷送信端末 2 が送信する平均負荷量は、低負荷と高負荷とを組み合わせ、表 4.2 に示す 3 通りの負荷パターンを用いる。なお、音声・ビデオの欠落率の平均値は、どの通信対においても 10% 以下と低くなる値を負荷パターン 1 として定めている。また、音声・ビデオの欠落率の平均値が一番高い通信対において 50% 台と高くなる値を負荷パターン 3 として定めている。

以上より、被験者に提示する環境は、 16 (バッファリング時間の組み合わせ) \times 3 (平均負荷量の組み合わせ) $= 48$ 通りとなる。

4.2.3 評価方法

本章では、主観品質として総合的な主観品質を取り上げる。また、系列カテゴリー法により得られた総合的な主観品質を表す尺度を、総合心理的尺度と呼ぶ。

被験者は、評価前に、負荷及びバッファリング制御を加えない環境で対話を行う。その際、受講者二人の復唱するタイミングが同時になるよう練習を行い、評価時には、習得したタイミングで返答を行うように指示をする。これは、後の実験において、バッファリング時間に対する返答のタイミングの個人差を無視できる大きさにするためである。

返答時間の早遅は、次のように測定する。まず、返答が早かったと講師に判断された受講者を、早返答者とする。各評価において、総質問数に対して受講者が早返答者と

なった回数の割合を、その受講者の早返答率と定義する。また、同時返答率も同様に定義し、これらを返答時間として扱う。被験者は、上述の対話を 30 秒間継続する。

本実験では、受講者のみが主観評価を行うこととする。上記の対話において、早遅や端末間同期品質の影響を主に受けるのは、受講者の主観品質だと考えられるからである^(注1)。したがって、対話後に、受講者は端末から提示された音声・ビデオの総合的な主観品質を評価する。

本章では、多地点接続の遠隔対話環境を扱う。多地点での遠隔対話環境を円滑に進めるためには、複数地点にいる受講者が公平に対話できる必要がある。そのため、主観評価をする際、受講者に公平性を含めてシステム全体を評価するように指示をする。また、返答時間においては、一人の受講者の早返答率が高い場合、他の受講者にとっては、早返答率が低いことになり、公平性に欠ける。一方、同時返答率が高い場合、どの受講者にとっても公平性が高いと言える。そこで、本章では、公平性が高くなるバッファリング制御方法を調査する。

なお、4.2.2 項で述べたように、本実験では講師用端末からの音声・ビデオの品質を良好に保つようにしている。このことにより、受講者は、講師の質問内容を的確に理解できるので、返答時間の正確な測定が可能となる。このとき、バッファリング時間を含めた遅延が、返答時間に影響を及ぼす大きな要因となる。

主観評価は、5 段階妨害尺度を用いた評定尺度法により行われる。評価前の練習で行った環境で提示された音声・ビデオを基準として考えるように説明が与えられる。被験者は、対話中に提示されるメディアに対して、基準からの劣化の度合いを 5 段階のいずれかに分類する。

被験者は、48 通りの環境での対話を終えると、図 4.1 において反時計回りに端末を移動する。これを繰り返し、計三つの端末で対話を行う。したがって、各被験者は、講師を一度、受講者を二度務めることとなる。受講者は、48 通りの環境において、受講者用端末 1、受講者用端末 2 で、それぞれ総合的な品質の評価を行う。したがって、評価対象は、 $48 \times 2 = 96$ 個となった。

被験者は 20 代の男女 39 名である。総評価時間は一名につき約 120 分であった。

4.3 実験結果と考察

4.3.1 主観品質の測定結果

受講者は、96 個の評価対象に対して評定尺度法による主観評価を行った。得られた主観評価結果に対して系列カテゴリー法を適用し、総合心理的尺度を求めた。得られた尺度に対して Mosteller の適合度検定を行ったところ、有意水準 1% において、これらが測定結果に適合するという仮説が棄却されなかった。系列カテゴリー法では距離尺度を求めるため、MOS の場合とは異なり、各カテゴリー境界間の心理的な間隔は一定であるという仮定を置いていない。そこで、カテゴリー 2 の下限値を心理的尺度の原点と

(注1) 講師は、対話中に解答の早遅を用紙に記録する。それと同時に主観評価を行うと、負担が大きくなる。これも、受講者のみが主観評価を行う理由の一つである。

すると、カテゴリ 3、カテゴリ 4、及びカテゴリ 5 の下限値は、それぞれ、0.881、1.843、及び 2.791 となった。各カテゴリの境界値を図 4.2 に示す。

バッファリング時間と総合心理的尺度値との関係を表したものを図 4.3 から図 4.5 に示す。これらの図における横軸は、評価者の端末におけるバッファリング時間と他の受講者の端末におけるバッファリング時間である。4.2.3 項で述べたように、各受講者には公平性を含めてシステム全体を評価させている。そこで、縦軸の総合心理的尺度値は、受講者用端末 1 と受講者用端末 2 のものを平均した値としている。なお、総合心理的尺度値の違いをわかりやすくするため、各図において、最も高い総合心理的尺度値との差が 0.14 未満の場合に総合心理的尺度値が高いものとして、黒色の棒グラフで示す。また、0.14 以上 0.28 未満の場合は灰色の棒グラフで示し、0.28 以上の場合は白色の棒グラフで示す。

図 4.3 から図 4.5 において、負荷パターン 1 より負荷パターン 2 のとき、及び負荷パターン 2 より負荷パターン 3 のときに、どのバッファリング時間においても、総合心理的尺度値は低くなっている^(注 2)。これは、平均負荷量が高負荷になったことで、送信したメディアの品質が劣化する端末が生じるためである。

また、各負荷において総合心理的尺度値が最も高い値となるバッファリング時間に注目すると、図 4.3、図 4.4 においては、各受講者端末でのものが等しいか近い値となっている。各受講者端末でのバッファリング時間を等しく設定すると、公平な対話が実現されることにより、総合心理的尺度値を高く保つことができるのである。

次に、負荷パターンごとに、バッファリング時間に対する総合心理的尺度値を比較する。図 4.3 より、負荷パターン 1 の状況においては、自端末と他の受講者の端末におけるバッファリング時間が短いほど、総合心理的尺度値が高い傾向であることがわかる。特に、いずれかの端末のバッファリング時間が 100ms で両端末のバッファリング時間の差が 100ms 以内の場合に、総合心理的尺度値が高くなっている。この理由を以下に述べる。負荷が低い状況では、バッファリング制御を行わなくてもメディアの品質が良好である。したがって、自端末でのバッファリング時間を大きくすることによるメディア同期品質の改善は得られない。そのため、バッファリング時間の増加は、単に即時応答性の低下を意味する。また、即時応答性は、他の受講者の端末でのバッファリング時間が大きいほど低下する。こうして、低負荷の状況においては、各受講者の端末が短いバッファリング時間を設定すれば、高い総合心理的尺度が得られる。

図 4.4 より、負荷パターン 2 では、総合心理的尺度が高い値となる各端末でのバッファリング時間の組み合わせは、2 通りあることがわかる。一つは、図 4.3 と同様に両端末におけるバッファリング時間が短い場合（実験では、100ms）である。もう一つは、それに比べてバッファリング時間が共に長い場合（同じく、400ms）である。

まず、バッファリング時間が短い場合の理由を示す。負荷が高い状況では、MU の伝送遅延が大きくなる。したがって、バッファリング時間が短いと、目標出力時刻に到着が間に合わず廃棄される MU が増える。そのため、受信されるメディアの品質が悪くなる。しかし、即時応答性が高いので、総合心理的尺度が高くなっている。

(注 2) 負荷パターン 2 と負荷パターン 3 において、自端末のバッファリング時間が 600ms で他の受講者の端末のそれが 400ms のときに、総合心理的尺度値はいずれも 1.40 を示している。これらの表示桁数を増やすと、負荷パターン 2 では 1.404、負荷パターン 3 では 1.395 となっている。

次に、バッファリング時間が長い場合の理由を示す。この場合は、即時応答性は低い
が、廃棄される MU が少なくなるため、メディアの品質が良くなる。また、バッファリ
ング時間を増加することによるメディア同期品質の改善効果が大きい。そのため、今回
の実験では、バッファリング時間が 400ms と長い場合に、総合心理的尺度が高くなる。

図 4.5 より、負荷パターン 3 では、総合心理的尺度値が高い値となるバッファリング
時間に一定の傾向が見られない。これは、負荷パターン 1 及び負荷パターン 2 に比べ
て、全体に総合心理的尺度値が低い値として評価されており、劣化が気になったため明
確な差となって現れなかったからである。

4.3.2 返答時間の測定結果

受講者が 96 個の主観評価を行った際、講師は各受講者の返答時間の早遅を測定した。

バッファリング時間と早返答率との関係を表したものを図 4.6 から図 4.8 に、また、
バッファリング時間と同時返答率との関係を表したものを図 4.9 から図 4.11 に示す。こ
れらの図における横軸は、図 4.3 から図 4.5 と同様に評価者の端末におけるバッファリ
ング時間と他の受講者の端末におけるバッファリング時間である。また、図 4.6 から
図 4.11 の縦軸は、いずれも受講者用端末 1 と受講者用端末 2 の値の平均値を示してい
る。なお、違いをわかりやすくするため、各図において、早返答率または同時返答率が
0.45 以上の場合に早返答率または同時返答率が高いものとして、黒色の棒グラフで示
す。また、0.40 以上 0.45 未満の場合は灰色の棒グラフで示し、0.40 未満の場合は白色
の棒グラフで示す。

図 4.6 から図 4.8 より、どの負荷パターンにおいても、自端末のバッファリング時間
が他の受講者の端末におけるバッファリング時間よりも短いときほど、早返答率が高い
ことがわかる。これは、講師の声が、他の受講者の端末よりも自端末に届くのが早いた
めである。また、図 4.6 から図 4.8 において、すべてのバッファリング時間の組み合わ
せにおいて、負荷パターンの違いによる早返答率の差は、一箇所^{注 3)}を除いて 0.1 以内
と近い値になっている。これは、早返答率が負荷の影響をほとんど受けずに、バッファ
リング時間の影響が支配的であることを意味する。

次に、図 4.9 から図 4.11 より、どの負荷パターンにおいても、自端末のバッファリ
ング時間と他の受講者の端末におけるバッファリング時間が近いときほど、同時返答
率が高い値になっている。これは、講師の声が、二つの受講者用端末に同時に届き、公
平な対話が実現されるためである。また、図 4.9 から図 4.11 において、すべてのバッ
ファリング時間の組み合わせにおいて、負荷パターンの違いによる同時返答率の差は、
ほとんどの場合^{注 4)}に 0.1 以内と近い値になっている。これは、返答時間が受講者用端
末 1 と受講者用端末 2 のバッファリング時間の組み合わせによってのみ決定することを
意味している。

(注 3) 負荷パターン 1 と負荷パターン 3 において、バッファリング時間が自端末で 200ms、他の受講者
端末で 600ms の場合のみ、早返答率の差が 0.13 と大きくなっている。

(注 4) 負荷パターン 1 とそれ以外の負荷パターンにおいて、自端末と他の受講者端末のバッファリ
ング時間が等しい場合にのみ、同時返答率の差が大きい場合がある。

4.3.1項で述べたように、負荷が高い状況の場合、総合心理的尺度が高い値となるバッファリング時間の傾向は明確ではなかった。しかし、早返答率及び同時返答率を見ると、負荷が低い場合と同様の傾向が見られた。すなわち、負荷の高低に関わらず、バッファリング時間が等しい場合に、受講者に対する公平性が高くなるといえる。したがって、公平性を高めるためには、今回実験で用いたすべての負荷パターンにおいて、端末間同期制御を行う必要があることがわかる。

4.4 むすび

本章では、IP ネットワーク上での多地点接続によるリアルタイム型遠隔対話環境において、受講者用端末がバッファリング制御を行った場合の主観品質と返答時間とを検討した。その結果、主観品質が高い値となるバッファリング時間は、各受講者端末でのものが等しいか近い値であることがわかった。特に、今回行った実験の結果においては、受講者用端末が低負荷の影響を受ける場合には、いずれかの端末のバッファリング時間が100msで両端末のバッファリング時間の差が100ms以内の場合であった。また、二つの受講者用端末が、それぞれ低負荷と高負荷の影響を受ける場合には、バッファリング時間が400ms以内で等しい場合に、総合心理的尺度値が高くなった。ただし、受講者用端末が高負荷の影響を受ける場合には、一定の傾向は見られなかった。一方、返答時間の早遅を測定した結果、すべてのバッファリング時間の組み合わせにおいて、負荷パターンの違いによる早返答率や同時返答率の差はほとんどの場合に0.1以内と近い値になっており、バッファリング時間の影響が支配的であることがわかった。また、自端末のバッファリング時間が他の受講者端末のバッファリング時間に近いほど公平性が高く、特に、両端末のバッファリング時間が等しい場合に公平性が高かった。したがって、各端末が独自にバッファリング時間を設定するのではなく、すべての端末で出力時刻が同期するようにバッファリング時間を設定する（すなわち、端末間同期制御を行う）必要があると言える。

今後の課題として、以下のものが考えられる。まず、本章では、返答時間を正確に測定するために、講師からの音声・ビデオの品質を良好に保つ実験環境を設定した。この品質が劣化する場合の検討が必要である。次に、本章の対話は、実際の遠隔授業で用いられるようなものを扱っていない。したがって、現実の遠隔授業に即した対話において、端末間同期品質が及ぼす影響を調査する必要がある。

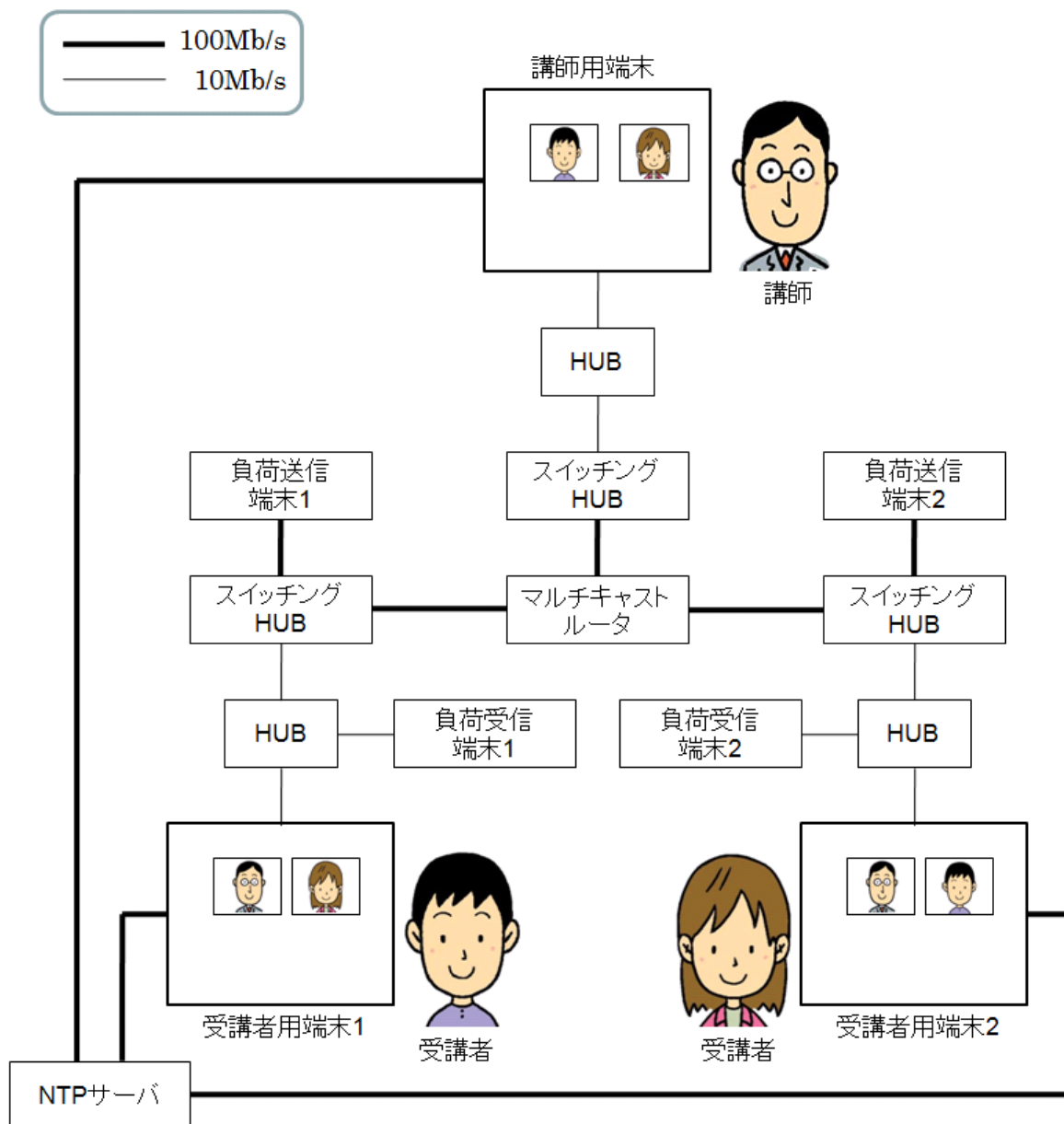


図 4.1: 実験システム構成

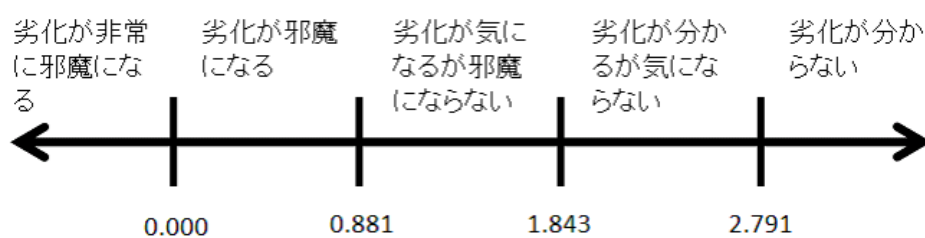


図 4.2: 各カテゴリーの境界値

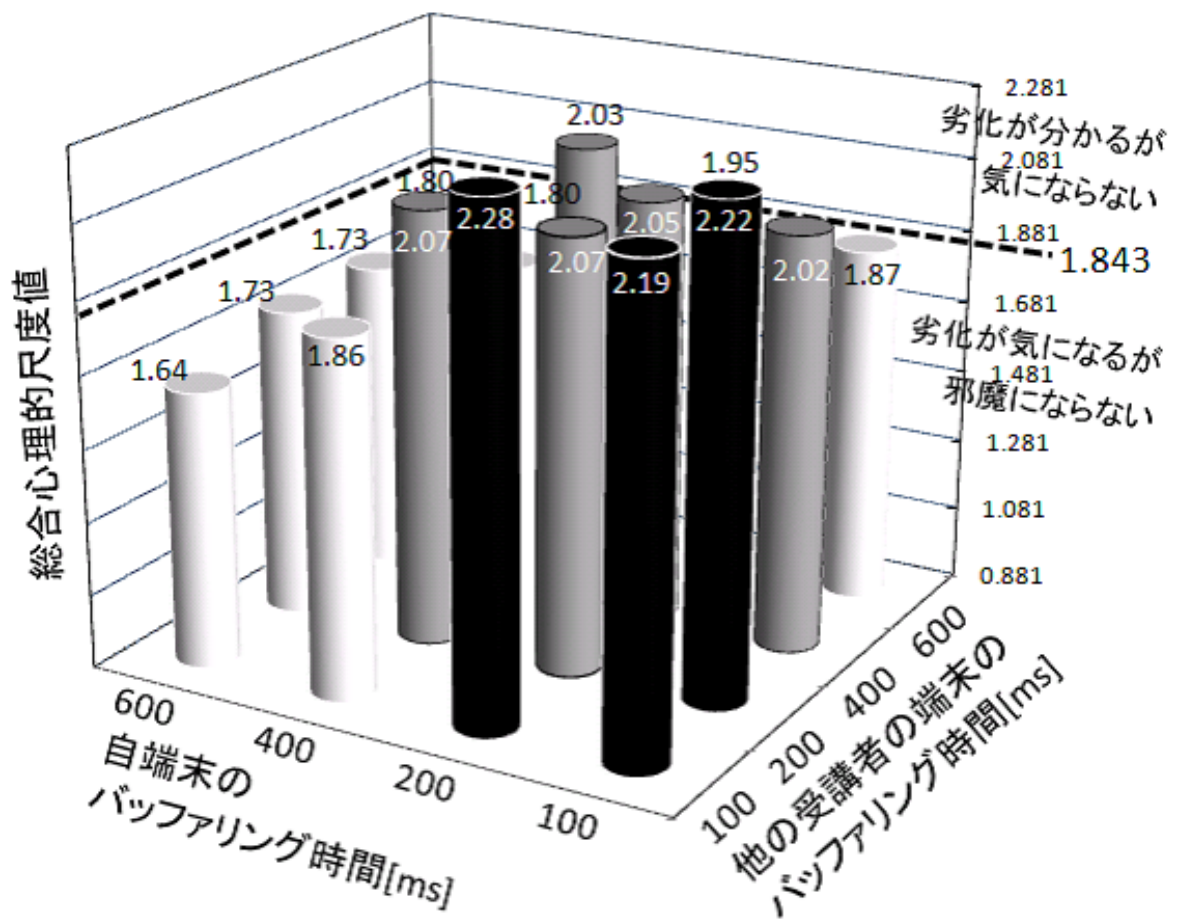


図 4.3: 負荷パターン 1 における総合心理的尺度値

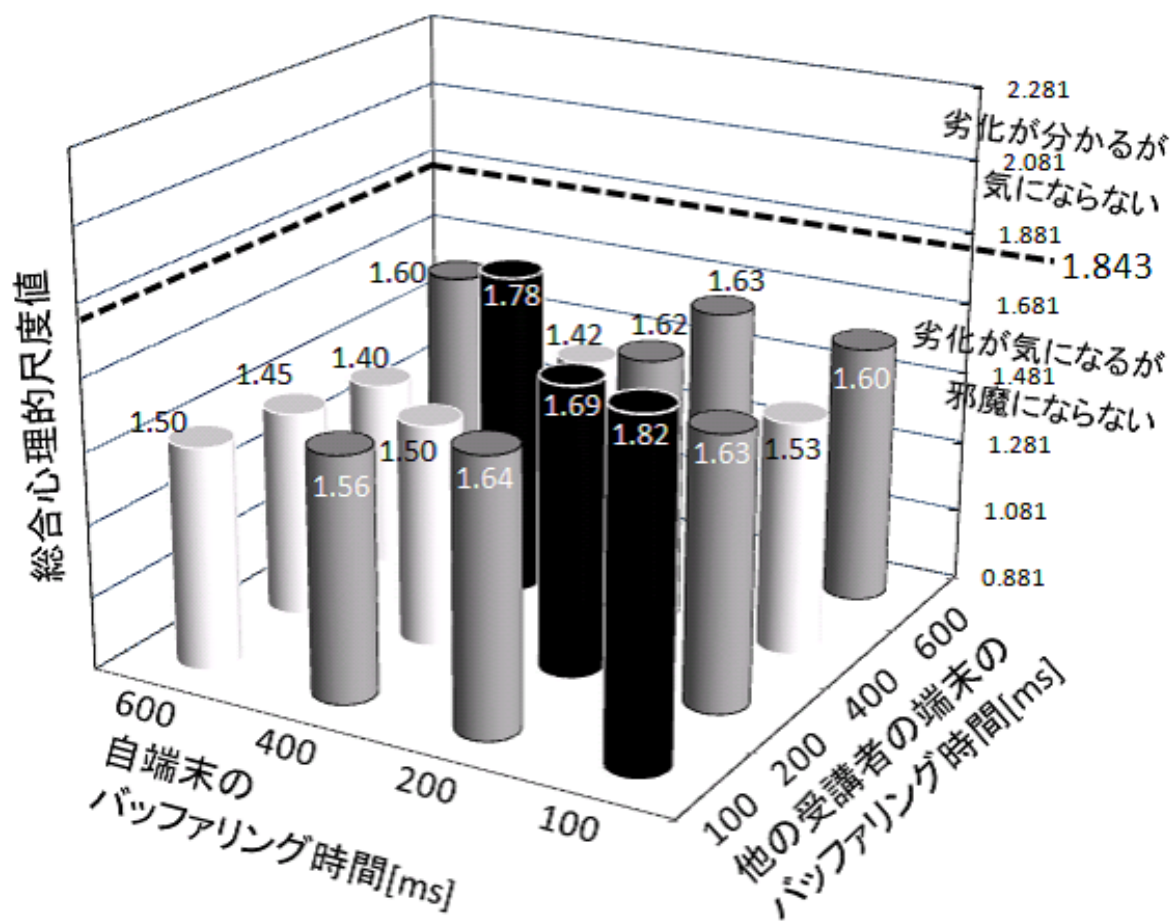


図 4.4: 負荷パターン 2 における総合心理的尺度値

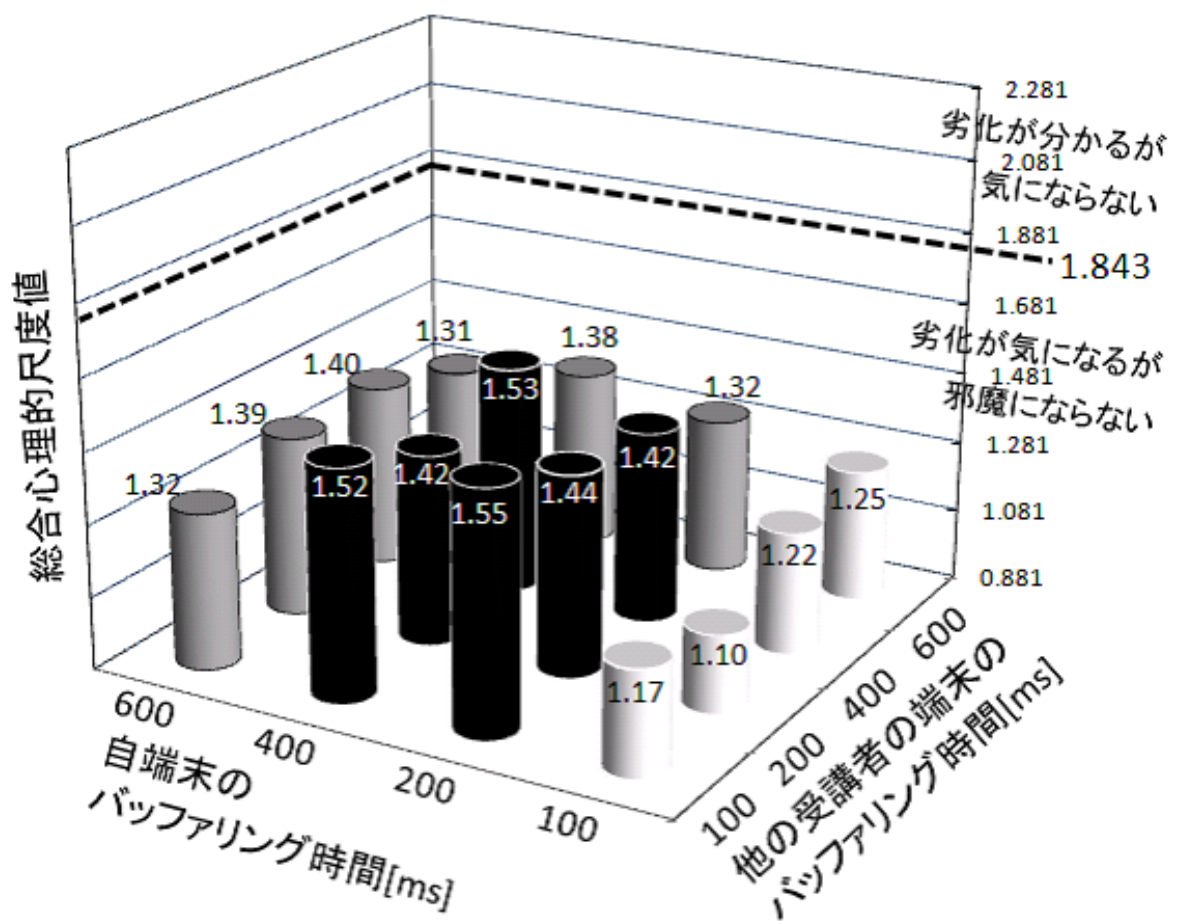


図 4.5: 負荷パターン 3 における総合心理的尺度値

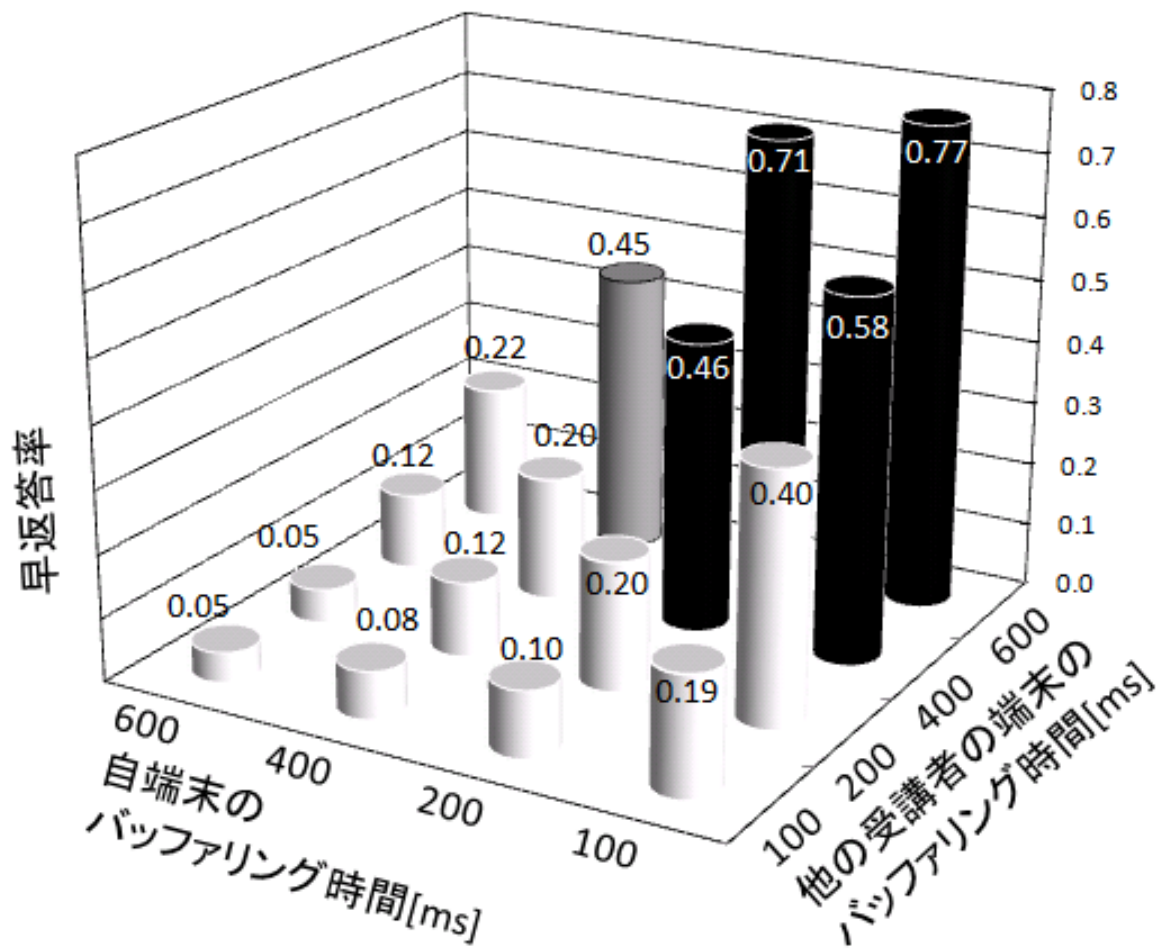


図 4.6: 負荷パターン 1 における早返答率

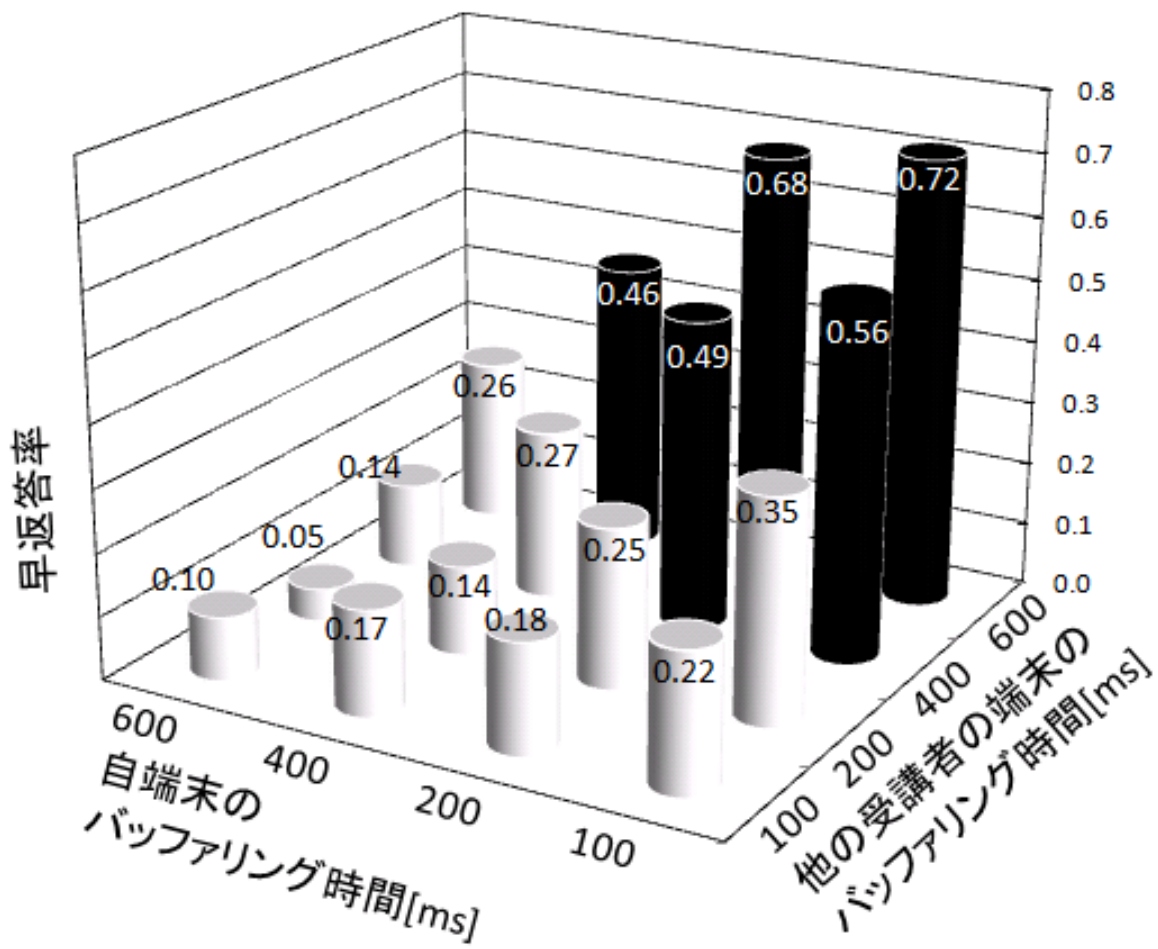


図 4.7: 負荷パターン 2 における早返答率

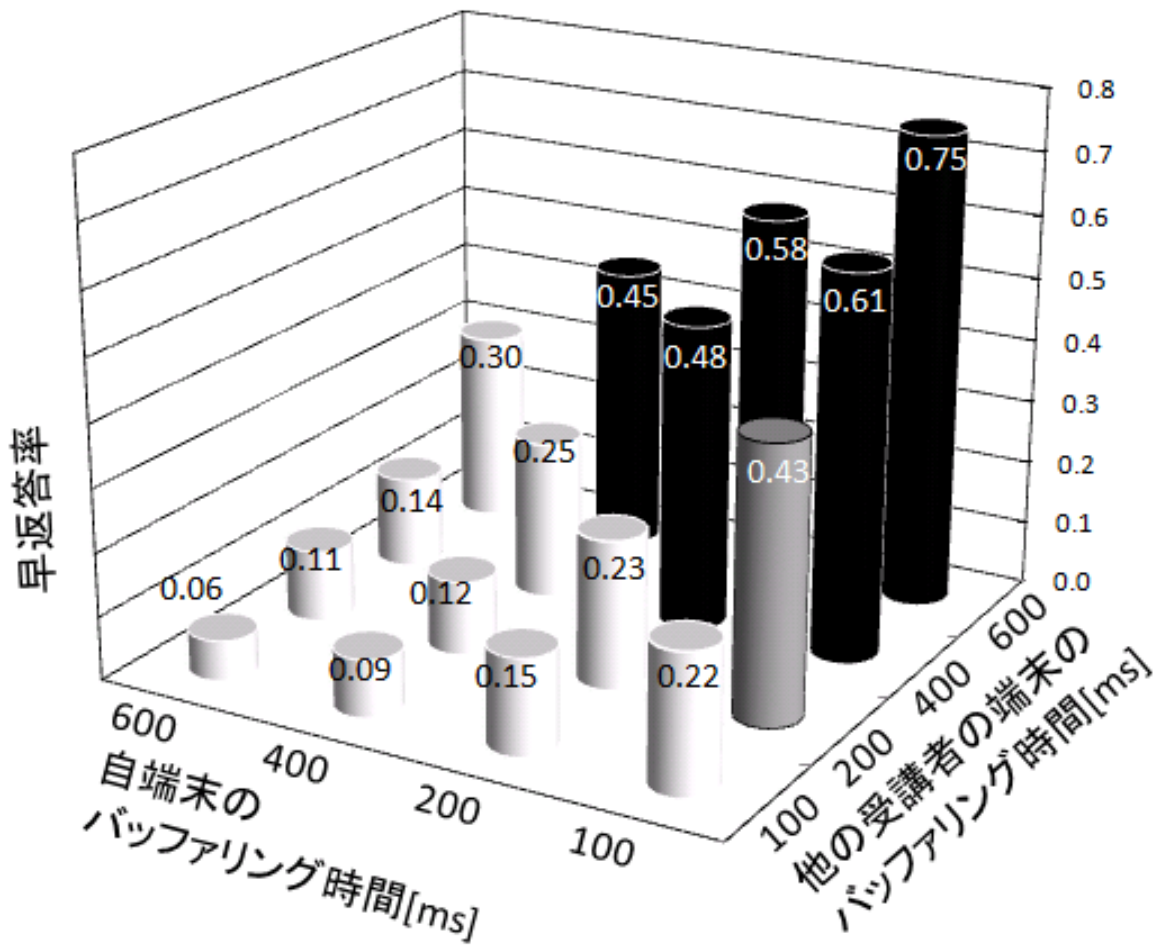


図 4.8: 負荷パターン 3 における早返答率

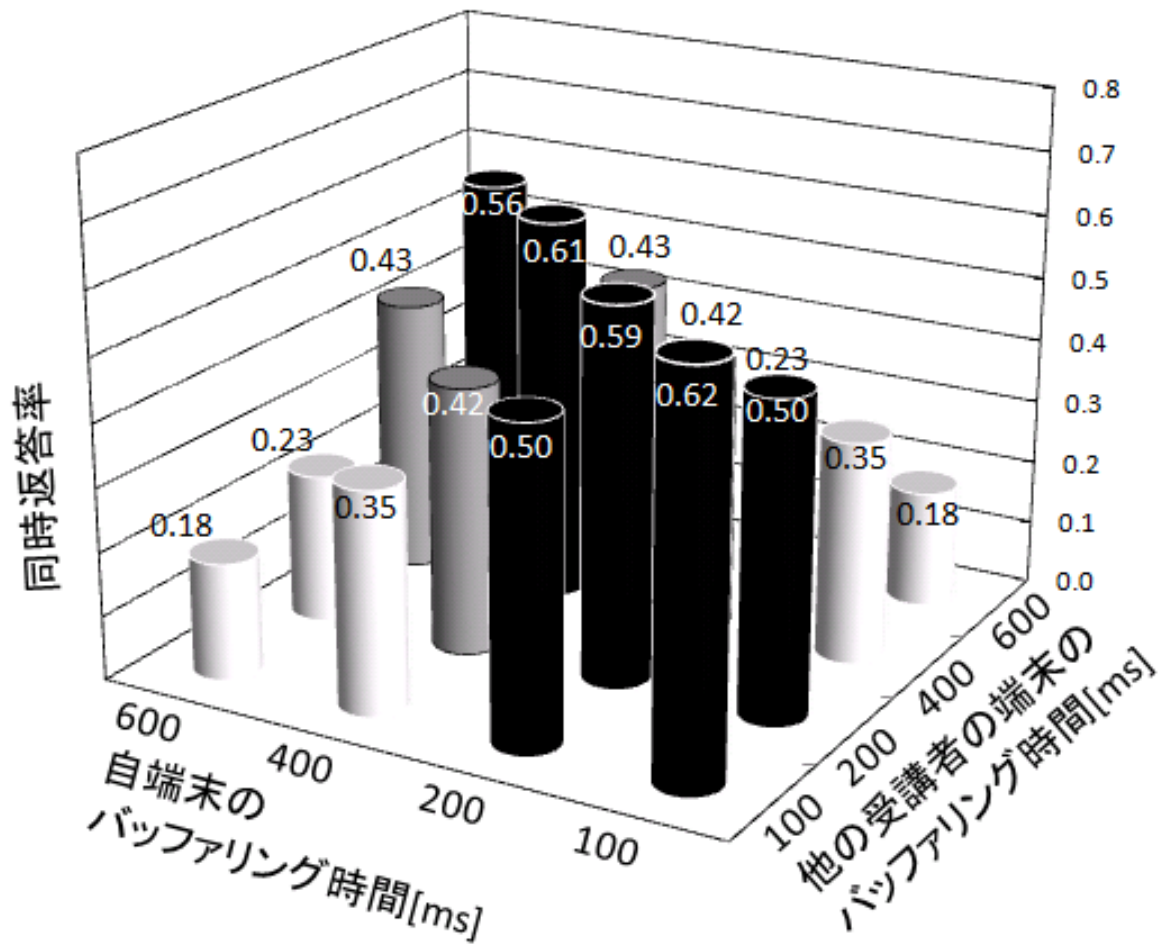


図 4.9: 負荷パターン 1 における同時返答率

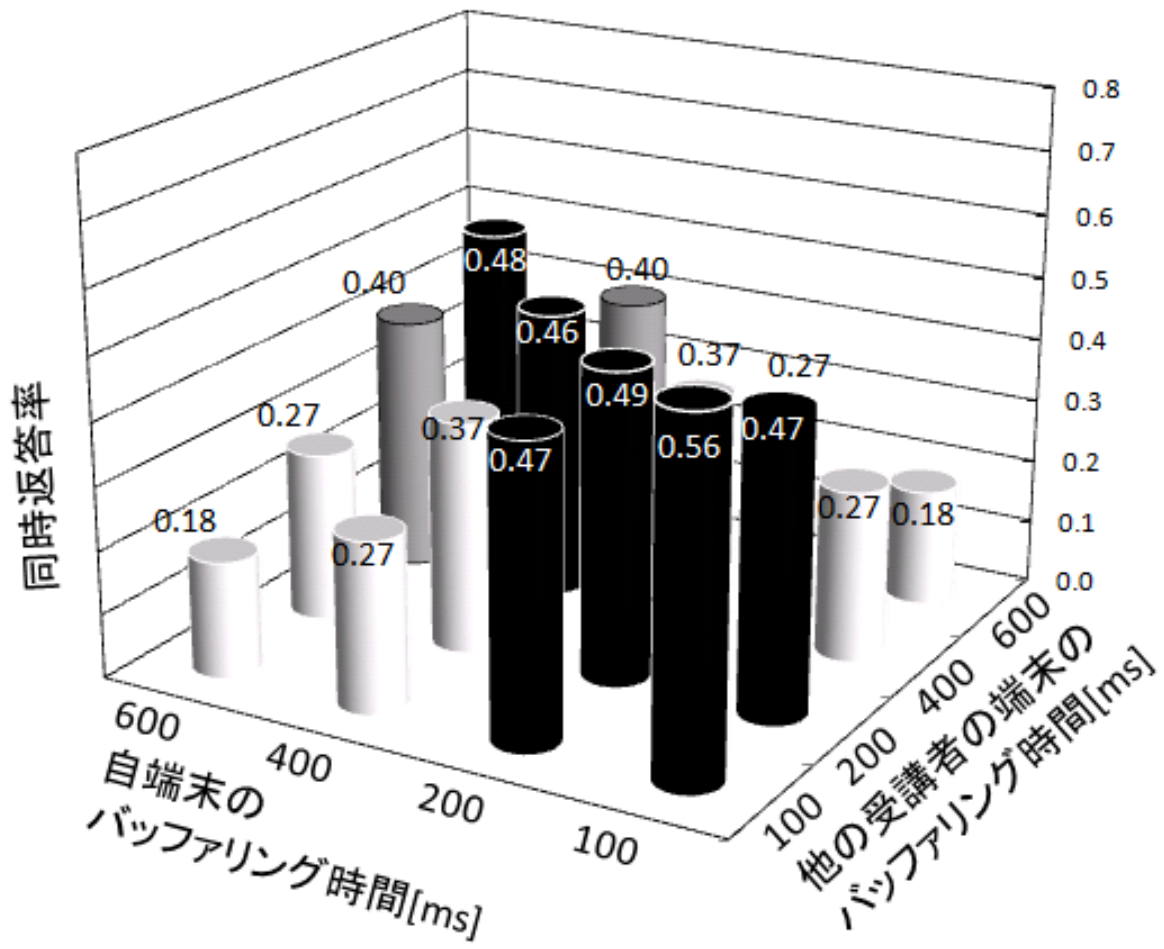


図 4.10: 負荷パターン 2 における同時返答率

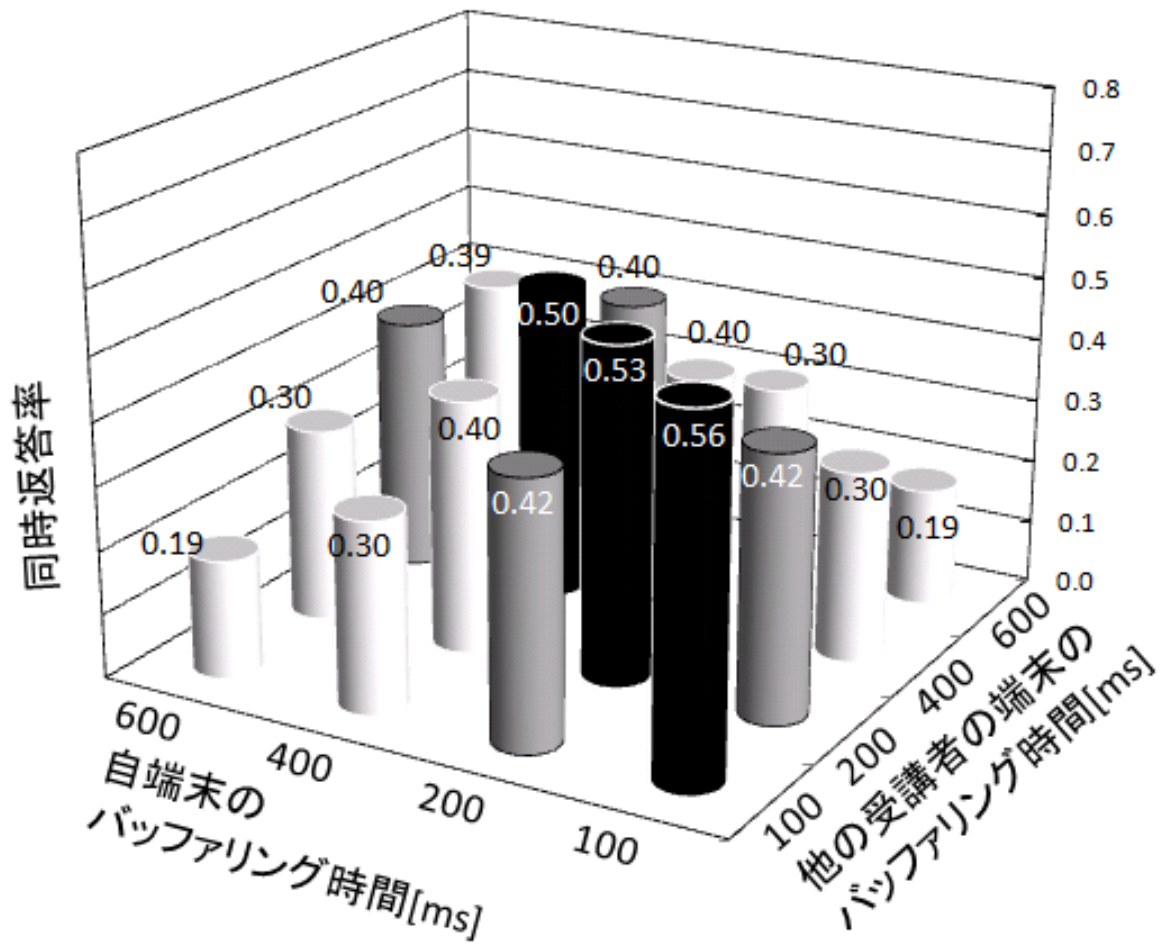


図 4.11: 負荷パターン 3 における同時返答率

第5章 結論

本論文では、2種類のマルチメディア遠隔学習システムを扱い、多面的な QoE 評価を行った。扱った遠隔学習システムは、1対1接続のオンデマンド型遠隔学習システムと1対多接続のリアルタイム型遠隔学習システムである。多面的な QoE 評価として、学習効果、主観評価及び返答時間を扱った。そして、遠隔学習者の QoE を向上させるための指針を示した。

第2章では、IP ネットワーク上での1対1接続オンデマンド型遠隔学習システムを用いた音声・ビデオ伝送において、文字情報の有無が視聴者の学習効果に及ぼす影響を、実験により調査した。学習効果として、問題に対する解答結果の正答率を用いた。また、アプリケーションレベル QoS と正答率との QoS マッピングを行い、正答率の推定式を導いた。その結果、次の結論を得た。

- 実験したすべてのコンテンツにおいて、文字情報が視聴者の正答率向上に影響を及ぼす。
- IP ネットワークにおいて、文字情報を加えることで、平均負荷量が大きくなっても、正答率が低下しにくい。
- 特に、負荷が少ないときに正答率が高いコンテンツほど、文字情報を加えることで、平均負荷量が大きくなっても、正答率が低下しにくい傾向が強い。

第3章では、IP ネットワーク上での1対1接続オンデマンド型遠隔学習システムを用いたマルチメディア遠隔学習において、通信品質の変化に対する最適な情報の提示方法を実験により調査した。本章においても、正答率を学習効果の評価尺度として用いた。その結果、以下の結論を得た。

- 負荷トラヒックがあるレベルまでの通信品質劣化を引き起こしたとき、文字に加えた音声・ビデオが学習者の学習効果を高める。
- 劣化が大きすぎると、補完的に送られた音声かビデオのいずれかの劣化が、逆に学習効果を低下させる。
- 劣化が大きすぎる場合に、補完的に送られた音声かビデオのいずれかの劣化が学習効果を低下させるのは、記憶が難しい教材が目立つ。
- 単回帰式を用いることで、負荷トラヒック量に対して最適な情報提示方法のガイドラインを示すことができる。

第4章では、IP ネットワーク上での1対多接続によるリアルタイム型遠隔対話環境において、受講者用端末がバッファリング制御を行った場合の主観品質と返答時間とを検討した。その結果、次の結論を得た。

- 主観品質が高くなるバッファリング時間は、各受講者端末でのものが等しいか近い値である。特に、今回行った実験の結果においては、受講者用端末が低負荷の影響を受ける場合には、いずれかの端末のバッファリング時間が100msで両端末のバッファリング時間の差が100ms以内の場合であった。
- 二つの受講者用端末が、それぞれ低負荷と高負荷の影響を受ける場合には、バッファリング時間が400ms以内で等しい場合に、総合心理的尺度値が高くなった。ただし、受講者用端末が高負荷の影響を受ける場合には、一定の傾向は見られなかった。
- 返答時間の早遅を測定した結果、すべてのバッファリング時間の組み合わせにおいて、負荷パターンの違いによる早返答率や同時返答率の差はほとんどの場合に0.1以内と近い値になっている。すなわち、バッファリング時間の影響が支配的である。
- 自端末のバッファリング時間が他の受講者端末のバッファリング時間に近いほど公平性が高かった。特に、両端末のバッファリング時間が等しい場合に公平性が高いことがわかった。

以上の結論から、IP ネットワーク上でのマルチメディア遠隔学習において、多面的なQoEにより評価することができた。そして、学習効果を高めるため、通信品質が劣化している場合には、音声・ビデオ主体のコンテンツに対して、文字情報を積極的に利用すると良いことがわかった。また、マルチメディア利用のコンテンツよりも、場合によっては文字情報だけのコンテンツの方が学習効果を高めることがわかった。さらに、多地点接続のリアルタイム型遠隔学習においては、すべての端末で出力時刻が同期するようにバッファリング時間を設定する必要がある。

以上のように、IP ネットワーク上でのマルチメディア遠隔学習において、遠隔学習システムの開発者、及び教材コンテンツの提供者等にとって、遠隔学習者のQoEを向上させるための指針を提供している。

今後、更に研究を進展させるための検討課題は、次のとおりである。

- 学習者がもっと多くの場所から利用する遠隔学習システムにおいて、QoEを評価する。
- 遠隔学習の現実により近い形式の教材コンテンツにおいて学習効果を測定する。
- 情報提示方法を学習者が選択することで、学習効果を高めることができるシステムを検討する。

- 年齢，性別，国籍などの被験者の属性が異なった場合における QoE 評価を行う．
- より多面的な QoE で評価するため，“ユーザの快適度 ”を含めた評価を行う．

謝辞

本研究において，終始熱心に御指導並びに御助言を頂いた田坂修二教授には，深く感謝し，厚く御礼申し上げます．また，本論文をまとめるにあたり，有益な御助言を頂きました石橋豊教授，松尾啓志教授に感謝致します．さらに，日頃から研究の相談に応じてくださった布目敏郎助教に感謝致します．

更に，本研究を進める過程で御協力を頂きました元大学院生の山田祐也氏，野田晃一郎氏，並びに元卒業研究生の小坂耕一氏に感謝致します．

末筆ながら，日頃お世話になりました本研究室の皆様，並びに関係者各位に感謝致します．

参考文献

- [1] 経済産業省, “eラーニング白書 2007/2008年版”, 東京電機大学出版局, 2007.
- [2] 放送大学学園, “2009年度 ICT 活用教育実態国内調査報告書”, 放送大学 ICT 活用・遠隔教育センター研究成果 2009年度報告書, pp. 44–46, 2009.
- [3] 清水康敬, 伊藤秀子, 佐賀啓男, 高比良美詠子, 波多野和彦, 吉田 文, “効果的な遠隔教育/e-Learning 実施の視点”, メディア教育開発センター, NIME 研究報告第1号, 2005.
- [4] S. Hrastinski, “Asynchronous and Synchronous E-Learning,” *Educause quarterly*, vol. 31, no. 4, Dec. 2008.
- [5] 清水康敬, “e-Learning を支える政策と今後の展望”, 情報処理学会誌, vol. 43, no. 4, pp. 421–426, 2002.
- [6] S. Tasaka and Y. Watanabe, “Real-Time Estimation of User-Level QoS in Audio-Video IP Transmission by Using Temporal and Spatial Quality,” in *Conf. Rec. IEEE GLOBECOM2007*, Nov. 2007.
- [7] ITU-T Rec. P. 10/G. 100 Amendment 2, “New definitions for inclusion in Recommendation ITU-T P.10/G.100”, July 2008.
- [8] P. Briiks, B. Hestnes, “User measures of quality of experience: why being objective and quantitative is important,” *IEEE Network*, vol. 24, no. 2, pp. 8–13, Mar./Apr. 2010.
- [9] 三木 哲也, “統括 –アクセスネットワークはどう変わったか, これからどう変わってゆくか?–”, 電子情報通信学会誌, vol. 88, no. 3, pp. 142–149, Mar. 2005.
- [10] 岡本敏雄, 小松秀圀, “eラーニングの理論と実際”, 丸善, 2004.
- [11] 鈴木克夫, “二つの遠隔教育–通信教育から遠隔教育への概念的連続性と不連続性について–”, メディア教育研究, no. 3, pp. 1–12, 1999.
- [12] J. Postel, “Internet protocol,” RFC 791, Sep. 1981.
- [13] 林敏浩, 大谷誠, 田中久治, 秋山和範, 井上真由美, 渡辺健次, 林田行雄, 近藤弘樹, “高精細映像を用いた板書授業方式による遠隔講義”, 電子情報通信学会技術研究報告, ET2000–142, pp. 219–226, Mar. 2001.

- [14] 中島義明, “映像の心理学”, サイエンス社, 1996.
- [15] 中島義明, “メディアにまなぶ心理学”, 有斐閣, 1996.
- [16] 福田充, “活字メディアにおけるオーガナイザー効果に関する実証的研究: 見出し、写真が読み手に与える影響についての認知心理学的再検討”, マス・コミュニケーション研究, no. 47, pp. 127–138, July 1995.
- [17] H. Itoh, “An analysis of eye movements while watching educational TV programs,” *Bulletin of the National Institute of Multimedia Education*, vol. 5, pp. 147–162, 1991.
- [18] 藤原 洋, “最新 MPEG 教科書”, 株式会社アスキー, 1994.
- [19] D. Le Gall, “MPEG: A video compression standard for multimedia applications,” *Commun. of the ACM*, vol. 34, no. 4, pp. 46–58, Apr. 1991.
- [20] 田坂修二, “ネットワーク環境におけるメディア同期”, 電子情報通信学会誌, vol. 84, no. 3, pp. 177–183, Mar. 2001.
- [21] R. Ramjee, J. Kurose, D. Towsley and H. Schulzrinne, “Adaptive playout mechanisms for packetized audio applications in wide-area networks,” in *Proc. of IEEE INFOCOM’94*, pp. 680–688, Apr. 1994.
- [22] K. Fujimoto, S. Ata and M. Murata, “Adaptive Playout Buffer Algorithm for Enhancing Perceived Quality of Streaming Applications,” *Conf. Rec. IEEE GLOBECOM2002*, Nov. 2002.
- [23] 間瀬健一, “マルチメディアネットワークとコミュニケーション品質”, 電子情報通信学会, 1998.
- [24] 田坂修二, “情報ネットワークの基礎”, 数理工学社, 2003.
- [25] 田中良久, “心理学的測定法 第二版”, 東京大学出版, 1977.
- [26] 市川伸一, “心理測定法への招待 測定から見た心理学入門”, サイエンス社, 1991.
- [27] ITU–T Rec. P. 910, “Subjective video quality assessment methods for multimedia applications,” Sep. 1999.
- [28] ITU–R Rec. BT500–11, “Methodology for the subjective assessment of the quality of television pictures,” Jan. 2002.
- [29] ITU–T Rec. P. 800, “Methods for subjective determination of transmission quality,” Aug. 1996.

- [30] 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 の定量的測定及びマッピング”, *電子通信学会論文誌*, B, no. 3, pp. 485–498, Mar. 2003.
- [31] S. Tasaka and Y. Ito, “Psychometric analysis of the mutually compensatory property of multimedia QoS,” *Conf. Rec. IEEE ICC2003*, pp. 1880–1886, May 2003.
- [32] S. Tasaka and Y. Ito, “Real-time estimation of user-level QoS of audio-video transmission over IP networks,” *Conf. Rec. IEEE ICC2006*, June 2006.
- [33] 印東太郎, “鉛筆の硬さの心理的尺度の構成”, *品質管理*, vol. 10, pp. 97–100, 1959.
- [34] L. L. Thurstone, “A law of comparative judgment,” *Psychological Review*, vol. 34, pp. 273–286, Jan. 1927.
- [35] L. L. Thurstone, “Stimulus dispersions in the method of constant stimuli,” *J. Experimental Psychology*, vol. 15, pp. 284–298, 1932.
- [36] 吉田正昭, “心理統計学”, 丸善, 1976.
- [37] 伊藤嘉浩, 田坂修二, “音声・ビデオ伝送における系列カテゴリ法を用いたユーザレベル QoS 評価”, *信学ソ大*, 2002.
- [38] F. Mosteller, “Remarks on the method of paired comparisons: III.a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed,” *Psychometrika*, vol. 16, no. 2, pp. 207–218, June 1951.
- [39] 日本教育工学会, “教育工学事典”, 実教出版, June 2000.
- [40] 土屋洋, “e ラーニング導入ガイド”, 第 6 章効果測定, 東京電機大学出版局, 2004.
- [41] 橋本重治, 辰野千壽, 石田恒好, 長澤俊幸, 宮島邦夫, 村主典英, 海保博之, 桜井茂男, 服部環, “2003 年改訂 教育評価法概説”, 図書文化, 2003.
- [42] H. H. Remmers, N. L. Gage and J. F. Rummel, “A Practical Introduction to Measurement and Evaluation (2nd Ed),” *Education for Living Series*, 1965.
- [43] 橋本重治, 榊原清, “教育評価の技術 – 教師作成テストの作り方と標準検査の使い方”, 図書文化, 1967.
- [44] F. Haist, A. P. Shimamura and L. R. Squire, “On the relationship between recall and recognition memory,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 18, no. 4, pp. 691–702, 1992.
- [45] 中村和男, 富山慶典, “選択の数理”, 東京朝倉書店, 1998.
- [46] 田村武志, 渋井二三男, 菊川健, “デジタル信号圧縮が遠隔学習に及ぼす諸要因”, *電子通信学会論文誌*, A, vol. J75-A, no. 2, pp. 235–243, Feb. 1992.

- [47] 藤木卓, 室田真男, 清水康敬, “DV 動画を用いた遠隔講義における画像劣化の許容範囲”, *教育システム情報学会誌*, vol. 20 no. 3, pp. 266–273, 2003.
- [48] 菅原良, 村木英治, “e ラーニングにおける動画像およびテキスト情報の配置が学習者に及ぼす心理的影響に関する考察”, *コンピュータ&エデュケーション*, vol. 20, pp. 62–67, 2006.
- [49] J. Bobis, J. Sweller and M. Cooper, “Cognitive load effects in a primary-school geometry task”, *Learning and Instruction*, vol. 3, pp. 1–21, 1993.
- [50] 田村武志, “遠隔講義における学習者インタフェースの改善とその評価”, *電子通信学会論文誌, A*, vol. J77–A, no. 3, pp. 494–505, Mar. 1994.
- [51] 滝田亘, 中山実, “視覚と聴覚による文章の提示と記憶への影響”, *日本教育工学会論文誌*, vol. 27 (Suppl.), pp. 81–84, 2003.
- [52] 吉野志保, 野嶋栄一郎, 赤堀侃司, “英語の聞き取り場面における字幕付加の効果”, *日本教育工学会論文誌*, vol. 21 (Suppl.), pp. 29–32, 1997.
- [53] 大倉孝昭, 村尾敏彦, “映画の音声と解答を同期させ相互評価により聴解力を高める CALL システムの開発と学習者評価”, *日本教育工学会論文誌*, vol. 27, no. 3, pp. 315–324, Dec. 2003.
- [54] 植松茂男, “DVD 映画教材利用時の英語字幕が英語学習に与える影響について”, *メディア教育研究*, vol. 1, no. 1, pp. 107–114, 2004.
- [55] 藤木卓, 室田真男, 清水康敬, “低ビットレート用ビデオ会議ツールの授業メディアとしての評価”, *電子情報通信学会技術研究報告*, ET2001–82, pp. 115–122, Dec. 2001.
- [56] “ The Network Simulator – ns-2,” (<http://www.isi.edu/nsnam/ns/>).
- [57] 市川伸一, 伊東裕司, 渡邊正孝, 酒井邦嘉, 安西祐一郎, “記憶と学習”, 岩波書店, 1994.
- [58] M. G. Moore and G. Kearsley, “Distance Education: A Systems View,” *Wadsworth Publishing*, 2004.
- [59] K. C. Green, “The 2001 National Survey of Information Technology in US Higher Education,” *The Campus Computing Project*, Oct. 2001.
- [60] P. S. Cowen, “Film and Text: Order Effects in Recall and Social Inferences,” *Educational Communication and Technology Journal*, vol. 32, no. 3, pp. 131–144, Sep. 1984.
- [61] R. E. Mayer and R. Moreno, “Aids to Computer-Based Multimedia Learning,” *Learning and Instruction*, vol. 12, pp. 107–119, Feb. 2002.

- [62] Y. H. Guan, “The Effects of Multimedia Presentations on the Learning Efficiency of Assembly Instructions,” *AACE ED-MEDIA*, pp. 2541–2547, June 2006.
- [63] D. Kim and D. A. Gilman, “Effects of Text, Audio, and Graphic Aids in Multimedia Instruction for Vocabulary Learning,” *Educational Technology & Society*, vol. 11, no. 3, pp. 114–126, 2008.
- [64] 深谷和義, 小坂耕一, 田坂修二, “音声・ビデオ IP 伝送において文字情報が学習効率に及ぼす影響”, 教育システム情報学会誌, vol. 25, no. 2, pp. 162–172, Sep. 2008.
- [65] The Apache Software Foundation, “Apache HTTP Server Project,” ([http : //www.apache.org/](http://www.apache.org/)) .
- [66] 独立行政法人大学入試センター, “平成 19 年度センター試験 (本試験) 問題 : 国語”, ([http : //www.dnc.ac.jp/center_exam/19exam/mondai_pdf/19kokugo-q.pdf](http://www.dnc.ac.jp/center_exam/19exam/mondai_pdf/19kokugo-q.pdf)).
- [67] 清水康敬, 前迫孝憲, “キャンパス間を結ぶテレビ講義の評価”, 電子通信学会論文誌, A, vol. J69–A, no. 10, pp. 1181–1189, Oct. 1986.
- [68] 村上正行, 八木啓介, 角所考, 美濃導彦, “受講経験・日米受講習慣の影響に注目した遠隔講義システムの評価要因分析”, 電子通信学会論文誌, D-I, vol. J84–D–I, no. 9, pp. 1421–1430, Sep. 2001.
- [69] 村上正行, 田口真奈, 溝上慎一, “日米間遠隔一斉講義における講師・受講生の評価変容の分析”, 日本教育工学雑誌, vol. 25, no. 3, pp. 199–206, Dec. 2001.
- [70] 西田裕一, 松原行宏, 富永浩之, 山崎敏範, 日詰裕雄, 松下文夫, “テレビ会議システムを利用する日中間中学生交流”, 電子情報通信学会技術研究報告, ET2002–112, pp. 103–108, Feb. 2003.
- [71] 藤木卓, 相原玲二, 前田香織, 柳生大輔, 西山敏明, 渡辺健次, “HD 動画伝送を用いた中学校における遠隔授業の実践”, 電子情報通信学会技術研究報告, ET2004–127, pp. 147–152, Mar. 2005.
- [72] 藤木卓, 寺嶋浩介, 園屋高志, 米盛徳市, 仲間正浩, 森田裕介, 関山徹, “三大学の連携による離島の複式学級を結ぶ遠隔共同学習の実践”, 日本教育工学会論文誌, vol. 31 (Suppl.), pp. 137–140, 2004.
- [73] 赤倉貴子, 永岡慶三, 西堀ゆり, “国際間の 3 大学を結ぶ同時双方向遠隔授業が学生に与える効果”, 電子情報通信学会技術研究報告, ET2006–34, pp. 71–76, July 2006.
- [74] Y. Ishibashi, A. Tsuji and S. Tasaka, “A Group Synchronization Mechanism for Stored Media in Multicast Communications,” *Proc. IEEE INFOCOM '97*, pp. 693–701, 1997.
- [75] 栗田孝昭, 六藤雄一, “多地点間映像音声通信サービスに対する品質特性分析”, 電子情報通信学会技術研究報告, CQ2004–07, pp. 17–22, July 2004.

- [76] 藤木卓, 森田裕介, 全炳徳, 李相秀, 渡辺健次, 下川俊彦, 柳生大輔, 上菌恒太郎, 中村千秋, “高精細動画を用いた多地点接続による中学校間日韓遠隔授業の実践と評価”, *日本教育工学会論文誌*, vol. 29, no. 3, pp. 395–404, Feb. 2006.
- [77] Y. Ito and S. Tasaka, “User-level QoS Assessment of a Multipoint-to-Multipoint TV Conferencing Application over IP Networks,” *Conf. Rec. IEEE GLOBE-COM2006*, 2006.
- [78] ITU-T Rec. P. 920, “Interactive Test Methods for Audiovisual Communications,” International Telecommunication Union, 1996.
- [79] S. McCanne and V. Jacobson, “vic: A Flexible Framework for Packet Video,” *in Proc. of ACM Multimedia*, Nov. 1995.
- [80] V. Hardman, A. Sasse, M. Handley and A. Watson, “Reliable Audio for Use over the Internet,” *Proc. INET '95*, Hawaii, Internet Society, Reston, VA, Nov. 1995.
- [81] W. A. Montgomery, “Techniques for Packet Voice Synchronization,” *IEEE J. Select. Areas Commun.*, vol. SAC-1, no. 6, pp. 1022–1028, Dec. 1983.

発表論文一覧

学術論文（査読あり）

1. 深谷 和義, 小坂 耕一, 田坂 修二, “音声・ビデオ IP 伝送において文字情報が学習効率に及ぼす影響”, 教育システム情報学会誌, vol. 25, no. 2, pp. 162–172, Sep. 2008 .
2. 深谷 和義, 山田 祐也, 田坂 修二, “多地点 IP 通信によるリアルタイム型遠隔対話環境において受信バッファリング制御が主観品質と返答時間に及ぼす影響”, 教育システム情報学会誌, vol. 26, no. 2, pp. 161–171, July 2009 .

（参考資料）

3. 田坂 修二, 深谷 和義, “アロハ形予約衛星パケット通信網におけるアクノリッジ方式の解析”, 電子情報通信学会論文誌（B）, vol. J70–B, no. 7, pp. 780–790, July 1987 .
4. 田坂 修二, 深谷 和義, “CSMA/CD ネットワークにおけるウィンドウフロー制御方式の解析”, 電子情報通信学会論文誌（B）, vol. J71–B, no. 12, pp. 1456–1467, Dec. 1988 .

国際会議（査読あり）

1. K. Fukaya, K. Noda, S. Tasaka and T. Nunome, “The Effect of Audio–Video Quality on Learning Effectiveness in Distance Learning over IP Networks,” Proc. IEEE Region 10 Conference (TENCON 2010), pp. 1444–1449, Nov. 2010.

（参考資料）

2. K. Noda, T. Nunome, S. Tasaka and K. Fukaya , “Assessment of Educational Effectiveness in Real–time Distance Lectures over IP Networks,” Proc. IEEE Asia-Pacific Conference on Communications (APCC 2009), pp. 862–865, Oct. 2009.

国内研究会（査読なし）

1. 深谷 和義, 小坂 耕一, 田坂 修二, “文字情報を含む音声・ビデオ IP 伝送における作業成績評価”, 電子情報通信学会技術研究報告, CQ2007-70, OIS2007-60, IE2007-67, pp. 189-194, Sep. 2007 .
2. 深谷 和義, 野田 晃一郎, 田坂 修二, 伊藤嘉浩, “IP ネットワーク上での遠隔学習において通信品質と情報の提示方法が学習効果に及ぼす影響”, 教育システム情報学会研究報告, vol. 23, no. 2, pp. 47-54, July 2008 .
3. 野田 晃一郎, 深谷 和義, 布目敏郎, 田坂 修二, “IP ネットワークでの遠隔講義における符号化ビットレートと受信バッファリング制御が教育効果に及ぼす影響”, 電子情報通信学会技術研究報告, CQ2008-72, pp. 69-74, Jan. 2009 .

国内大会（査読なし）

1. 野田 晃一郎, 深谷 和義, 田坂 修二, 伊藤嘉浩, “IP ネットワーク上での遠隔学習において音声・ビデオの品質劣化が学習効果に及ぼす影響”, 電子情報通信学会 2008 年ソサイエティ大会, B-11-12, Sep. 2008 .
2. 野田 晃一郎, 深谷 和義, 布目敏郎, 田坂 修二, “IP ネットワークでの遠隔講義における受講者の主観的満足度と学習定着度との関係”, 電子情報通信学会 2009 年総合大会, B-11-14, Mar. 2009 .
3. 野田 晃一郎, 布目敏郎, 田坂 修二, 深谷 和義, “IP ネットワークでの遠隔講義における負荷トラヒックと受信バッファリング制御が教育効果に及ぼす影響”, 電子情報通信学会 2009 年ソサイエティ大会, B-11-6, Sep. 2009 .

（参考資料）

4. 深谷 和義, 田坂 修二, 高橋 健一, “アロハ形予約衛星通信網における ACK トラヒックの解析”, 昭和 61 年度電子通信学会通信部門全国大会, 1-97, Sep. 1986 .
5. 田坂 修二, 深谷 和義, “CSMA/CD システムにおけるフロー制御方式の解析”, 電子情報通信学会創立 70 周年記念総合全国大会（昭和 62 年）, 8-29, Sep. 1987 .