JAIST Repository
https://dspace.jaist.ac.jp/

Title	多地点遠隔会議における自然なコミュニケーションの実現に関する研究
Author(s)	小峯，隆宏
Citation	
Issue Date	2005-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/965
Rights	
Description	Supervisor:丹　康雄，情報科学研究科，博士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

博 士 論 文

多地点遠隔会議における
自然なコミュニケーションの実現に関する研究

指導教官　丹 康雄　助教授

北陸先端科学技術大学院大学
情報科学研究科

小 峯　隆 宏

平成１７年３月

要旨

現代社会においては、全参加者が同時刻に 1 箇所に集まって会議を適切な頻度で開催することが困難になってきており、遠隔会議の活用による適切な会議運営が必要となってきている。また、近年のインターネットの普及・広帯域化は、インターネットを活用した遠隔会議システムの研究開発と実用利用の試みの促進に大きく寄与している。一方、会議参加者の一部が遠隔地から参加する多地点遠隔会議において、対面会議と同様の活発な議論を展開するためには、遠隔地間との円滑な会話やアイコンタクトによるスムーズな話者交替の実現等の自然なコミュニケーションを実現できる環境が必要である。しかし、会議参加者が多地点に分散している場合の現在の遠隔会議システムには、音声・映像等の往復伝送遅延時間が大きく遠隔地間との円滑な会話が困難という課題と、遠隔参加者が会話をしたいと思う人の方向に振り向く行為やアイコンタクトによるスムーズな話者交替といった、同一実空間内であれば容易に行える行動が困難という課題が存在し、自然なコミュニケーションの実現を妨げる大きな要因になっている。

本研究では、これらの課題に対し、まず音声・映像の伝送方式や音声ミキシング等のリアルタイム加工処理の改善による往復伝送遅延時間短縮の実現に取り組むことにする。次に、遠隔参加者の分身となって他の会議参加者と一緒に会議に参加する代理人ロボットおよびその遠隔操作システムの導入により、遠隔地にいる参加者の意思表示動作の把握やアイコンタクトによるスムーズな話者交替等の実現に取り組むことにする。

前者の取り組みでは、まず音声・映像等のストリーミング情報の符号化方式として相手の表情まで把握できる映像品質で、かつフレーム間予測符号化を行わず符号化処理時間が小さい、デジタルビデオ（DV）を採用する。そして、ネットワーク途中での複数の DV ストリーミング情報に対する音声ミキシング等のリアルタイム加工処理も、ネットワーク中の IP パケットに搭載されたデータ形式のまま行い、再度 IP パケットに搭載・転送する機構を開発する。また、同機構を実装した遠隔会議模擬システムに対して、定量測定実験および被験者による評価実験を行い、同機構の有効性を実証する。

後者の取り組みは、遠隔参加者の分身を物理的に離れている会議室という実空間内に実際に配置し、それを遠隔操作することで相手の意思表示動作の把握やアイコンタクトによるスムーズな話者交替等を実現するものである。まず遠隔参加者の分身として、目となるビデオカメラ、耳となるマイクロフォン、口となるスピーカおよび顔映像を出力する液晶ディスプレイを一体として、上下左右旋回する機能をもつ代理人ロボットを開発する。次に、ジョイスティックを利用し遠隔参加者が会議への集中度を維持したまま代理人ロボットの遠隔操作が可能なシステムを開発し、遠隔地にいる参加者の意思表示動作の把握やアイコンタクトによるスムーズな話者交替等を実現する。そして、これらの試作システムに対して被験者による評価実験を行い、同システムの有効性を実証する。

最後に、上述の 2 つの機能を組合せた遠隔会議システムが、実際の多地点遠隔会議という環境の中でどの程度有効に機能するかを検証するため、通常の対面会議環境および従来の遠隔会議システムを利用した遠隔会議模擬環境との総合比較評価実験を行う。評価方法の１つの柱として、プロジェクトの進捗や作業のパフォーマンス等の定量評価に用いられる EVM という統合マネジメントの技法を参考に、各システムにおける会議の進捗度を推定する新しい手法を提案し、評価実験を試みる。また、アンケートによる主観的評価や会話回数測定等の行動分析による定量評価という、遠隔会議等の総合評価としてこれまで一般的に用いられている手法も同時に実施する。そして、両者の評価結果を比較することで、今回提案した新しい手法の有効性についても考察する。

目次

第 1 章 序論 9
1.1 研究の背景と目的 9
1.2 論文の構成 10
第 2 章 遠隔会議システム 11
2.1 はじめに 11
2.2 遠隔会議に関するこれまでの取組み 11
2.2.1 遠隔講義の視点からの考察 11
2.2.2 実証実験からの考察 13
2.2.3 遠隔会議に関する考察 15
2.3 遠隔会議での自然なコミュニケーションの実現 15
2.4 現在の遠隔会議システムの課題 16
2.5 提案する遠隔会議システム 18
第 3 章 円滑な会話を実現する遠隔会議システム 20
3.1 はじめに 20
3.2 多地点遠隔会議システム 21
3.3 多地点デジタルビデオ遠隔会議システム 23
3.4 デジタルビデオ・リアルタイム処理機構 25
3.5 評価実験 30
3.5.1 品質劣化度の評価 31
3.5.2 伝送遅延時間の測定評価 32
3.5.3 妨害品質尺度による主観的評価 34
3.6 まとめ 36
第 4 章 実空間共有を実現する遠隔会議システム 37
4.1 はじめに 37
4.2 代理人ロボット 39
4.2.1 システム構成 40
4.2.2 特長 40
4.2.3 試験運用 41
4.3 遠隔操作システム 43
4.3.1 マウスによる遠隔操作システム 43
4.3.2 ジョイスティックによる遠隔操作システム 44
4.4 操作性に関する評価 48
4.4.1 被験者 48

4.4.2 実験構成 48
4.4.3 タスク 48
4.4.4 実験手順 50
4.4.5 デバイス操作時間の評価 51
4.4.6 レスポンス性能の評価 55
4.4.7 テレビ画面への集中度の評価 58
4.4.8 アンケート評価 61
4.4.9 考察 62
4.5 アイコンタクトの効果に関する評価 63
4.5.1 被験者 63
4.5.2 実験構成 63
4.5.3 タスク 63
4.5.4 実験手順 64
4.5.5 被験者の行動分析による評価 65
4.5.6 アンケート評価 68
4.6 まとめ 70
第 5 章 総合評価 71
5.1 はじめに 71
5.2 EVM の概念 71
5.3 提案する定量評価手法 75
5.4 提案する遠隔会議システムの構成 77
5.5 総合評価実験 78
5.5.1 被験者 78
5.5.2 実験構成 78
5.5.3 実験手順 82
5.5.4 提案手法による定量評価 82
5.5.5 被験者の行動分析による定量評価 89
5.5.6 アンケートによる主観的評価 93
5.6 まとめ 94
第 6 章 結論 95
6.1 本論文の総括 95
6.2 おわりに 97
謝辞 99
参考文献 100
付録Ａ 104

付録Ｂ 111
研究業績 115

図目次

図 2.1　遠隔講義の提供形態の視点からの分類.....12
図 3.1　想定する多地点遠隔会議システムの概念図.....21
図 3.2　DIF シーケンス内の DIF ブロック配置順序.....23
図 3.3　3 地点 DV 遠隔会議システムでの会場 3 へ向かう情報の流れ.....24
図 3.4　音声情報の DV 情報への格納手法.....25
図 3.5　DV リアルタイム処理機構を導入時の会場 3 へ向かう情報の流れ.....28
図 3.6　DV リアルタイム処理機構全体の流れ.....28
図 3.7　音声ミキシング処理の流れ.....29
図 3.8　音声ミキシング用加算処理の流れ.....29
図 3.9　入力波形とポイント Y3 での出力波形の比較.....31
図 3.10　片方向伝送遅延時間の測定方法.....32
図 4.1　代理人ロボットを利用した遠隔会議システムの概念図.....38
図 4.2　代理人ロボットのシステム構成.....40
図 4.3　代理人ロボットを利用した遠隔ゼミの様子.....42
図 4.4　マウスによる遠隔操作システムの概要.....44
図 4.5　ジョイスティックによる遠隔操作システムの概要.....45
図 4.6　ジョイスティックでの操作命令割当.....46
図 4.7　ジョイスティックの移動と頭部旋回の関係.....46
図 4.8　操作性に関する評価実験構成.....49
図 4.9　評価実験に用いた英数字配置図.....49
図 4.10　操作性に関する評価実験の様子.....50
図 4.11　デバイス操作時間に関する測定結果.....53
図 4.12　操作経験数によるデバイス操作時間の変化.....54
図 4.13　レスポンス性能に関する測定結果.....56
図 4.14　操作経験数によるレスポンス時間の変化.....57
図 4.15　テレビ画面への集中度に関する測定結果.....59
図 4.16　操作経験数による画面への集中度の変化.....60
図 4.17　操作性に関するアンケート評価結果.....61
図 4.18　アイコンタクトの効果に関する評価実験構成.....64
図 4.19　話者交替時の話者の位置関係の変化.....66
図 4.20　話者交替のトリガーとなった行動の比率.....68
図 4.21　アンケート評価結果.....69

図 5.1　EVM でのプロジェクト進捗状況の把握と完了予測.....74
図 5.2　スケジュール効率指数とコスト効率指数の推移 74
図 5.3　提案手法での会議進捗状況の把握.....76
図 5.4　会議進捗効率指数の推移 76
図 5.5　改良した代理人ロボットの概観図.....77
図 5.6　改良後のジョイスティックの操作命令割当.....78
図 5.7　評価実験のシステム構成 80
図 5.8　対面会議環境での実験風景.....81
図 5.9　従来遠隔会議環境での実験風景 81
図 5.10　提案遠隔会議環境での実験風景.....81
図 5.11　評価実験の組合せと順番 83
図 5.12　各システムでの累積作業時間の推移 84
図 5.13　各システムでの会議進捗効率指数の推移 84
図 5.14　作業形態毎の会議進捗効率指数の推移.....85
図 5.15　作業経験数による作業時間の変化.....87
図 5.16　作業経験数による作業時間変動比率の変化.....88
図 5.17　被験者Ｚから被験者Ｘ,Ｙへのコミュニケーション行動回数.....91
図 5.18　被験者Ｘ,Ｙから被験者Ｚへのコミュニケーション行動回数.....91
図 5.19　被験者Ｘ－Ｙ間のコミュニケーション行動回数.....92
図 5.20　評価実験での会議環境に関する平均オピニオン値の結果 93

表目次

表 2.1　遠隔会議システムの課題 19
表 3.1　主な機器の諸元 30
表 3.2　片方向伝送遅延時間の測定結果 33
表 3.3　5 段階妨害品質尺度 35
表 3.4　妨害品質尺度による主観的評価結果 35
表 3.5　システム構成とオピニオン評点結果に関する分散分析表 36
表 4.1　代理人ロボットの諸元 39
表 4.2　コマンド毎の送信タイミング 47
表 4.3　デバイス操作時間の測定結果に関する分散分析表 53
表 4.4　操作経験数とデバイス操作時間に関する分散分析表 54
表 4.5　レスポンス性能測定結果に関する分散分析表 56
表 4.6　操作経験数とレスポンス時間に関する分散分析表 57
表 4.7　画面への集中度の測定結果に関する分散分析表 59
表 4.8　操作経験数と画面への集中度に関する分散分析表 60
表 4.9　被験者の話者交替時の行動回数（グループ全体） 67
表 5.1　比較対象となる会議環境の特徴 79
表 5.2　課題の内容 83
表 5.3　各グループの被験者のコミュニケーション行動回数 90
表 5.4　評価に用いた 5 段階の絶対範疇尺度 93

第1章　序論

1.1　研究の背景と目的

会議を行う目的は、単独では解決が困難な課題に対して、複数の参加者が様々な視点から活発な議論や情報交換を行うことで課題の解決をはかると共に、参加者全員が会議の内容に関する情報を同時に共有することといえる。このため、課題解決のために必要と考えられる人が 1 名でも欠席した会議は、有効に機能しないことは明らかである。逆に、課題解決のための様々な知識等をもつ人が 1 人でも多く会議に参加することは、深い充実した議論が展開される可能性を高くし、より良い課題解決が期待できる。

しかし、現代社会においては、会議に必要と考えられる参加者が物理的に離れている場所に散在しているケースが多い。このため、その全員を同時刻に 1 箇所に集めて会議を行うためのスケジュール調整が非常に困難となり、適切な頻度での会議開催が望めない場合が発生しているのが現状である。

この解決策の１つとして、通信ネットワークを用いて音声や映像等の情報をリアルタイム伝送することで、物理的に離れた場所にいる相手に対して双方向での議論や情報交換を行う「遠隔会議」の活用がある。古くは、一般電話回線や ISDN 回線等を利用したテレビ電話会議システムが使われている。しかし、伝送される映像品質が悪いため、相手会場の様子や相手の顔の表情を把握できず、本格的な議論を行えるまでの環境は実現できていなかった。

一方、近年のインターネットの普及により、広帯域のインターネット回線を比較的容易に整備できるようになった。このため、インターネットを活用し相手会場の様子や相手の顔の表情までも把握できるほどの映像品質を有する遠隔会議システムが、開発・販売されるような状況にまでなってきている。しかし、複数の会議参加者による活発な議論の実現という面で、これら現在の遠隔会議システムには以下の課題がある。1 つは、ネットワーク伝送処理や音声・映像の符号化処理等により、遠隔地間の音声・映像の伝送遅延時間が大きくなってしまい、円滑な会話を行うことが困難になってしまう点である。もう１つは、遠隔からの会議参加者と相手会場側にいる一般参加者の双方が、それぞれ相手と同一空間上で会議を行っているという一体感をもつことが難しく、疎外感が増大してしまい、お互い受身的な応対になってしまうという点である。

以上のような問題点を踏まえ、本研究では、遠隔会議において複数の参加者による活発な議論を促進するための技術的課題を抽出し、その実現手法の提案および有効性の評価を行う。

1.2　論文の構成

本論文は 6 章から構成される。第 2 章では、遠隔会議における自然なコミュニケーションの重要性、現在の遠隔会議システムの課題を示すと共に、提案する遠隔会議システムの構想について説明する。第 3 章では、デジタルビデオの特性を活用して往復伝送遅延時間を短縮するデジタルビデオ・リアルタイム処理機構の開発・実装と、円滑な会話実現に対する同機構の有効性について述べる。第 4 章では、代理人ロボットおよびその遠隔操作システムの開発・実装と、遠隔参加者の意思表示把握やアイコンタクトの実現等の空間共有に関する同システムの有効性について述べる。第 5 章では、提案する遠隔会議システムが、実際の多地点遠隔会議という環境の中でどの程度有効に機能するかの検証について、新しい評価方法を中心とする評価実験結果より考察する。第 6 章では、本研究の総括を行う。

第2章　遠隔会議システム

2.1　はじめに

遠隔会議において、対面会議と同様の活発な議論を展開するためには、遠隔地間との円滑な会話やアイコンタクトによるスムーズな話者交替の実現等の自然なコミュニケーションを実現することが必要である。本章では、遠隔会議に関するこれまでの取組んできた研究結果を中心に、自然なコミュニケーションを実現するための要件と現在のシステムの課題を整理すると共に、自然なコミュニケーションを実現する遠隔会議システムの構成について述べる。

2.2　遠隔会議に関するこれまでの取組み

2.2.1　遠隔講義の視点からの考察

ここでは、遠隔会議において対面会議と同様な活発な議論を展開するために何が重要かを探るために、遠隔会議の発展形として近年研究および利用が進められている遠隔講義の視点から考察する。講師から受講生への知識の伝授を行うことを主目的とする遠隔講義としては、まず講師と受講生への片方向の情報提供を主とする知識の伝授という片方向通信形態と、受講生からの質疑や講師の回答といった双方向の情報交換等もともなう知識の伝授を行う双方向通信形態、という講師と受講生とのインタラクティブ性という視点がある。また、講師と受講生が同一時間上で知識の伝授を行うリアルタイム形態と、受講生の都合のよいときに知識の伝授を受けることが可能なオンデマンド形態、という講師と受講生との時間の共有性という視点もある。この２つの視点から整理すると、**図 2.1** に示すように大きく 5 つのタイプに分類できる。以下にこれら 5 つのタイプの概要と特徴を記述する。

タイプＡの放送視聴型は、テレビ放送やラジオ放送を利用し、事前に決められている番組プログラムにそって講師による講義を流し、受講者はその時間に講義を視聴する形態である。テレビやラジオによる視聴のため様々な受講生にとって利用しやすい利点を持っているが、一般に講義中の質疑応答が実現できないため講義が単調になりやすいという欠点もある。

タイプＢの教材提供独習型は、ビデオテープや CD-ROM の教材を受講者に提供し、受講者が都合のよいときに教材を使って視聴・学習する形態である。受講者個人に都合のいい時間・ペースで学習できるという利便性を持っているが、講義内容の更新や訂正が即座にできないため、内容が普遍的なものであればいいが情報に新鮮さが必要な内容には向いていない。

タイプＣのオンデマンド講義型は、受講者がインターネット上の Web から PowerPoint

形式等のプレゼンテーションファイル、RealPlayer 形式等の音声・映像ストリーミングを中心とする講義コンテンツをダウンロードし、単独聴講する形態である。タイプＢと同様に受講者個人に都合のいい時間・ペースで学習できる利便性を持っているが、質疑応答は電子メールや電子掲示板等を利用することになるため、受講者からの質問等がいつ届くか予測できず講師の負担が大きくなりやすいという欠点もある。

タイプＤのライブ講義型は、通常の講義と同じように特定の時間帯に講師が講義を実施して、その同じ時間帯に遠隔地にいる受講者がリアルタイムで視聴し質疑応答等の会話も行う形態である。同じ講義を他の受講者と一緒に受けている、時間を共有している、といった一体感を持たせることが比較的容易であり、この一体感の実現は講師・受講者双方に適度な緊張感の維持と講義・受講意欲の向上に寄与する可能性を持っている。しかし、これらを実現するためには広帯域通信回線やシステムを確保する必要があるため、多地点に受講生が分散しているような場合はコストが高くなる傾向にある。

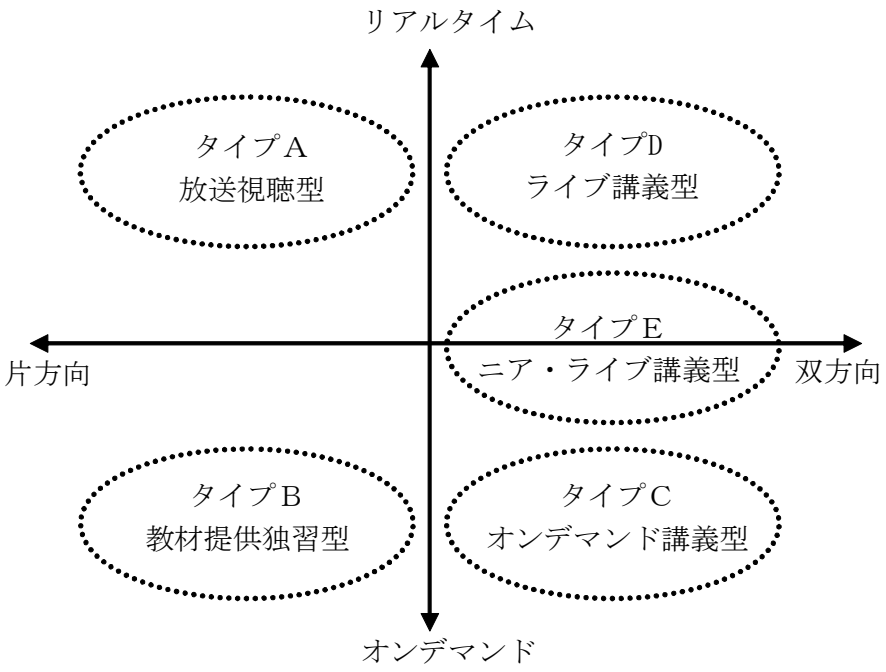


図 2.1　遠隔講義の提供形態の視点からの分類

タイプＥのニア・ライブ講義型は、タイプＣのオンデマンド講義のスタイルではあるが、Web 上の講義コンテンツの掲載期間を極めて短くし、かつ質疑応答等を時間限定のリアルタイム電子掲示板等にすることで、タイプＤのライブ講義と似たリアルタイム性を追求する形態である。期間・時間を限定するため、講師はその時間帯にまとめて質疑応答に対応でき、受講者側も皆で一緒に受講しているという一体感をある程度感じとるこ

とができる。また、受講者個人に掲載期間内ではあるが都合のいい時間・ペースで学習できるというある程度の利便性を持っている。

このように、それぞれの遠隔講義形態に一長一短があり、万能な遠隔講義システムはないことが分かる。わずかに空いた時間を見つけマイペースでじっくりと独自学習をしていくような場合には、タイプＢやＣのようなオンデマンド型が適している。一方、大学の一般講義や研修等のように短期間の間に集中して知識や技能を身につけるような場合には、ある程度の緊張感と集中力を保ち、その講義の内容に対する学習意欲を維持するような環境を講師・受講者側双方に提供することが重要であり、必然的にダイプＤのライブ講義型での実現を目指すことになる。

ここで、対面会議と同様な活発な議論を展開することを目標とする遠隔会議に目を向けると、遠隔会議はダイプＤのライブ講義型に相当することがわかり、双方向からの円滑な会話を実現できることが重要であるといえる。

2.2.2　実証実験からの考察

次に、筆者が過去実施した、ATM 通信実験ネットワーク活用による遠隔会議実証実験の結果から、遠隔会議において対面会議と同様な活発な議論を展開するために何が重要かを考察する。以下では、遠隔ディベート対戦、日本カナダ脳外科遠隔ワークショップ、そして 2000 年に実施された北陸先端科学技術大学院大学（以下 JAIST と記述）の遠隔講義それぞれの実施概要と技術的課題を述べる。

遠隔ディベート対戦の実証実験[1]では、本来ディベーター2 チームと審判が同一会場内でディベート対戦を行うところを、ATM実験ネットワークで3会場間を結び、NTSC-MPEG2符号化装置によるリアルタイム双方向通信を実現することで、遠隔ディベート対戦に実施した。ディベート対戦は、2 つのチームがある議題について賛成・反対を主張する立場に分かれ議論し合い、その議論内容の優劣を審判が判定するものであり、審判や対戦チームへの積極的なアピールや、対戦チームの細かな表情変化に対応した戦術の変更等が重要なポイントになる。このため、本実験では複数の会議参加者間で活発な議論や情報交換を行い課題の解決していく遠隔会議よりも厳しい会議環境が要求され、遠隔会議に必要な技術要素を解析するのに適している。実験は 2 回実施し、実験後被験者へのアンケート等も実施した。その結果、映像に関しては相手の表情まで瞬時に把握するために NTSC 放送品質クラスが必要であること、音声に関してはエコーバックにより自分の声が遅れて戻ってきてしまう現象が対戦チームとの議論に大きな障害であることが判明した。また、このエコーバックの大幅な低減化が必要であることや議論を交わす際には視線の一致つまりアイコンタクトが重要であることも判明した。

日本カナダ脳外科遠隔ワークショップ[2]は、日本・カナダの間を衛星通信回線と光ネットワークを組合せた ATM 通信実験ネットワークで結び、通信総合研究所（現在、独立

行政法人情報通信研究機構：以下 NICT と記述）が開発した HDTV-MPEG2 符号化装置によるリアルタイム双方向通信を実現することで、遠隔ワークショップを実施した。本実験の主なねらいは、高精細映像の遠隔地へのリアルタイム伝送によって、どの程度高臨場感を達成できるかを評価することにある。実験は 2 回実施し、実験後被験者へのアンケート等も実施した。その結果、映像に関しては手術用の糸まで判別可能な品質の映像品質は医学分野においても十分実用可能であるという評価を得た。しかし、衛星通信回線の帯域制限から HDTV-MPEG2 符号化装置において約 25Mbps 帯域にまで符号化圧縮をする必要があり、画像符号化処理遅延は約 700ms、ネットワーク伝送遅延時間も含めると往復伝送遅延時間は約 2 秒となってしまい、会話が途絶える場面が多くなっていた。音声に関しては単一指向性マイクの使用等である程度のエコーバックは低減させることができたが、遠隔地との会話を中心とするやりとりについての違和感はまだ残ってしまった。

2000 年 9 月に実施した JAIST の遠隔講義[3]では、JAIST と NICT 間を研究開発ギガビットネットワーク（Japan Gigabit Network：以下 JGN と記述）で結び、ATM-IEEE1394 Link Unit による NTSC 放送品質クラスのデジタルビデオ（Digital Video：以下 DV と記述）情報のリアルタイム双方向通信を実現した。そして、NICT にいる講師が JAIST にいる複数の受講生に対して情報科学研究科連携講座「超高速ネットワーク構成学」の遠隔講義を実施した。その結果、映像に関しては講師側および受講生側会場にそれぞれ 2 台ずつのカメラで、NICT 側のスタッフによる講師の追尾映像と講義資料を撮影する固定映像、JAIST 側のスタッフによる複数の受講生を撮影する映像と講義室後方から講義室全体を撮影する固定映像、計 4 種類のリアルタイム映像伝送を実現することで、講師・受講生双方から高い評価を受けた。一方、音声に関しては講師の音声が受講生側会場を経由して講師側会場に戻ってきてしまうエコーバックの課題がここでも生じた。エコーバックを解消する方法として、音声の双方向同時通信を諦め講師が講演している間は、受講生側会場の音声出力をすべて手動操作でオフにする手法を試みた。しかし、この場合講師が受講生側会場の映像のみを見ながら講演することになり、臨場感が著しく欠落した状態になった。更に、受講生との会話・質疑応答等がスムーズに行われないため、特に講師が大きなストレスを感じる結果になった。また、上記の手法を止め市販のエコーキャンセリング装置の利用を試みたが当時は正常に機能させることができず、エコーバックがほとんど低減されない状態か、低減されてエコーバックではない正常音声が途切れる等不安定になってしまった。一方、この遠隔講義では受講生側会場全体に発生している音を常に集音して講師側会場に伝送し、その音を講師がイヤホンによって聴き取ることにより受講生側会場の雰囲気や受講生の反応を受け取りやすくする環境も試みた。しかし、講師自身の発する音声が大きく遅延しかなりの音量で戻ってくるため、講師はストレスを感じることになり、大きな成果はあげられなかった。

以上、上述の 3 つの実証実験結果をまとめると、遠隔会議で対面会議と同様な活発な

議論を展開するためには、音声・映像の往復伝送遅延時間が短く円滑な会話が可能であること､相手の表情や意思表示を的確に把握できるNTSC放送品質以上の映像であること、遠隔地にいる参加者とのアイコンタクトによるスムーズな話者交替を実現すること、エコーキャンセリング装置等を機能させてエコーバックの限りなく小さくすること、が重要と考察できる。

なお、最後のエコーキャンセリング装置に関しては、エコーキャンセリングに関するサーベイ（**付録A**参照）を行った結果、スピーカからマイクに回り込む音の音量がある一定範囲内にないとエコーキャンセリング機能は正確に動作しないことが分かった。このため、エコーバックの対象入力音量がある一定範囲内に収まるようにマイクとスピーカの位置と音量を手動で調整するか、オートゲインコントロール装置をエコーキャンセリング装置の手前で用いることで、自動調整する必要があることが判明した。

2.2.3　遠隔会議に関する考察

遠隔講義の視点からの考察と、これまでの実証実験からの考察より、遠隔会議で対面会議と同様な活発な議論を展開するためには、音声・映像の往復伝送遅延時間が短く円滑な会話が可能であること、相手の表情や意思表示を的確に把握できるNTSC放送品質以上の映像であること、遠隔地にいる参加者とのアイコンタクトによるスムーズな話者交替を実現すること、が重要であると考察できる。この 3 点を本論文では「自然なコミュニケーションの実現」として以下議論を行うことにする。

2.3　遠隔会議での自然なコミュニケーションの実現

まず、遠隔会議において自然なコミュニケーションを実現するための要件を、対面会議を基準として整理する。

(1) 遠隔地にいる参加者の表情の把握

対面会議では、各参加者は同室内にいる他の参加者の表情を直接見ながら、会議を円滑に進める。たとえば、相手が内容を理解しているのか、疑問点があるのか、等を相手の表情から即座に判断し、補足説明や質問をするといった行動がこれに相当する。

遠隔会議でこのような行動を行うためには、遠隔地にいる参加者の表情まで把握できるような鮮明な映像がリアルタイムに近い状態で受信できることと、把握したい参加者の映像を容易に選択できることが必要となる。

(2) 円滑な会話の実現

対面会議では、全参加者が同室内にいるため、他の参加者同士で行われている音声会話を直接聞きつつ、自らもその会話の合間を見計らって必要に応じた発話を行い、会話

を発展させるといった行動を繰り返すことで、円滑な会話を実現している。

遠隔会議でこのような行動を行うためには、遠隔地間でやりとりされる音声の品質がはっきり聞き取れるレベルの音声品質であることのほかに、遠隔地にいる参加者の発話をマイク等で収集し、ネットワークを介してその音声情報がスピーカで再現されるまでに要する音声伝送遅延時間が、遠隔地間で会話のリズムを妨げないレベルにまで小さくする必要がある。特に遠隔地間での円滑な会話に許容される音声・映像の伝送遅延時間に関しては、研究報告[4][5]から、会話の目的によって許容時間は異なるものの、円滑な会話実現にはより小さい伝送遅延時間の実現が最も有効であることが確認できる。

(3) 遠隔地にいる参加者とのスムーズな話者交替の実現

対面会議では、各参加者の意思表示を自らの発声、ジェスチャ、アイコンタクト等によって、他の参加者に知らせることで行われる。たとえば、ある参加者に質問をしたい場合は、その人の方向を向きながらの発話や挙手、そしてアイコンタクトをとることで、会議を混乱させることなく話者交替をすることが可能になる。これにより、資料のある部分に注目してもらいたい場合は、該当部分を指示棒やレーザポインタ等で指し示したりすることで実現する。

遠隔会議でこのような行動を行うためには、リアルタイムでの音声伝達のほか、話をしたい人の方向に振り向きアイコンタクトをとる行動や、ある部分を指し示すといった行動も何らかの手法を用いて実現できる必要がある。特に、人と会話するという行為では、相手とのアイコンタクトをとりながら行うことが自然であり、前述した円滑な会話の成立条件の１つにもなる。

2.4　現在の遠隔会議システムの課題

遠隔会議システムは、実写映像を使ったシステムと、実写映像ではなくバーチャルリアリティ技術を活用したシステム、また両者を融合したシステムの大きく 3 つに分類できる。

まず、実写映像を使ったシステムの課題を整理する。実写映像の利用でも衛星回線を活用したシステム[6]は、多数の遠隔会議室の確保が容易になり、地理的制約もなくなるという利点はあるが、衛星経由の音声・映像等の伝送遅延時間がどうしても大きくなってしまうため、円滑な会話の実現という面で大きな問題点がある。これに対し、実写映像と高速ネットワークを活用した場合の遠隔会議システムについては、近年のインターネット広帯域化の進展にともない、このインターネットを活用し相手会場の様子や相手の顔の表情までも把握できるほどの音声・映像を提供できる遠隔会議システムが開発されるようになった。また、質疑応答支援システムやプレゼンテーション手法の改善等、対面会議環境に近づけることを目的とした多くの研究開発[7]-[13]が行われている。し

かし、これらのシステムにおいても、映像の圧縮方式がフレーム間の相関性を利用した時間方向の圧縮を行うフレーム間圧縮方式や高い圧縮率に設定することで、品質の高い映像の送受信を実現している。このため、相手会場の様子や相手の顔の表情までも把握できるほどの映像を提供する場合は、符号化処理遅延時間が大きくなってしまい、円滑な会話に支障が出る。また、この符号化処理遅延時間を小さくする設定にすると、送受信される音声・映像の品質が低下してしまい、この場合は遠隔地にいる参加者の表情の把握という面で大きな問題点がある。

次に、実写映像ではなくバーチャルリアリティ技術を活用したシステム[14][15]の課題を整理する。このシステムは、コンピュータ等が創り出す仮想空間上にアバタと呼ばれる各参加者の分身を配置し、このアバタが相手のアバタの方向を振り向く等の意思表示動作を行うと共に音声による円滑な会話を実現するものである。この場合、時間と空間を共有することで遠隔地間にいる参加者同士の一体感が増強され、かつアバタの擬似顔画像等を変化させることで大まかな表情の表現は可能となる。しかし、相手の細かな表情の変化までをこの実写を使わないアバタで表現することは困難である。

最後に、実写映像とバーチャルリアリティ技術の両者を融合した遠隔会議システム[16][17]の課題について整理する。このシステムとしては、上述のアバタの顔部分に実写の顔映像をはめ込むビデオアバタを利用する形態や、各会場の実写映像を合成したり立体表示したりすることで、1つの大きな実写映像仮想空間を表示する形態が存在する。これにより、相手の細かな表情や意思表示の把握が、対面会議に近い環境で実現できると。しかし現時点では、音声・映像および仮想空間をリアルタイムに表現するためには、大型スクリーンや特殊な伝送・表示装置を使う必要がある等、汎用性が低いという課題がある。

自然なコミュニケーション実現に関する現在の遠隔会議システムの課題を、**表 2.1** にまとめる。さらに、本研究では、2 地点間接続という特殊な環境ではなく 3 地点間以上を結ぶ環境での遠隔会議システムの開発を目指しており、その場合音声ミキシングや映像の切り替えといったリアルタイム編集機能もシステムに組み入れる必要がある。このため、上述のシステムでは音声・映像の伝送遅延時間に、ネットワーク上を流れる際のデジタル信号、編集処理で扱われるアナログ信号間の変換処理時間、リアルタイム編集処理時間が加わることになる点も課題として検討する必要がある。

2.5　提案する遠隔会議システム

本研究では、これらの課題に対し、以下に示す研究開発で解決を図り、その有効性の評価を行う。

1つ目は、まず音声・映像等のストリーミング情報に対する符号化方式の符号化処理時間が小さく品質も高い方式であるデジタルビデオ（Digital Video[18][19]：以下 DV と記述）を採用する。そして、ネットワーク途中での複数の DV ストリーミング情報に対する音声ミキシング等のリアルタイム加工処理も、ネットワーク中の IP パケットに搭載されたデータ形式のまま行い、再度 IP パケットに搭載・転送する機構を開発する。この機構に、DVTS（Digital Video Transport System[20][21][22]）というインターネットを介して DV ストリーミング情報のリアルタイム伝送を行えるアプリケーションを併用することで、NTSC テレビ放送クラスの映像と 48-kHz サンプリングの高品質な音声を用いた遠隔会議システムが実現可能となる。そして本研究では、同機構を実装した遠隔会議模擬システムに対して、定量測定実験および被験者による評価実験を行い、同機構の有効性を実証する。

2つ目は、遠隔参加者の分身を物理的に離れている会議室という実空間内に用意し、それを遠隔操作することで、相手の意思表示動作の把握やアイコンタクト等を実現するものである。まず遠隔参加者の分身として、目となるビデオカメラ、耳となるマイクロフォン、口となるスピーカおよび顔映像を出力する液晶ディスプレイを一体として、上下左右旋回する機能をもつ代理人ロボットを開発する。次に、ジョイスティックを利用し、遠隔参加者が会議への集中度を維持したまま代理人ロボットの遠隔操作が可能なシステムを開発し、遠隔地にいる参加者の意思表示動作の把握やアイコンタクト等を実現する。これにより、代理人ロボットを介して会議に参加する人は、遠隔地映像が表示されている画面に集中した状態で代理人ロボットを遠隔操作でき、自分の見たい方向を見ることが可能となる。一方、代理人ロボットのいる会議室内の参加者は、遠隔参加者の細かな表情を把握できるだけでなく、遠隔参加者が今どこに注目しているのか、誰に話し掛けているのか、といった情報も同時に認識可能となる。そして、これらの試作システムに対して被験者による評価実験を行い、同システムの有効性を実証する。

最後に、上述の 2 つの機能を組合せた遠隔会議システムが実際の多地点遠隔会議という環境の中でどの程度有効に機能するかを検証するため、通常の対面会議環境および従来の遠隔会議システムを利用した遠隔会議模擬環境との総合比較評価実験を行う。評価方法の１つの柱として、プロジェクトの進捗や作業のパフォーマンス等の定量評価に用いられる EVM という統合マネジメントの技法を参考に、各システムにおける会議の進捗度を推定する新しい手法を提案し、評価実験を試みる。また、アンケートによる主観的評価や会話回数測定等の定量評価という、遠隔会議等の総合評価としてこれまで一般的に用いられている手法も同時に実施する。そして、両者の評価結果を比較することで今

回提案した新しい手法の有効性についても考察する。

表 2.1　遠隔会議システムの課題

要件	課題
(1) 遠隔地にいる参加者の表情の把握	・実写映像を使わないと顔の細かな表情までは再現が難しい。 ・実写映像でも、MPEG-2等では要件(2)を満たすとやはり細かな表情までは再現できない。
(2) 円滑な会話の実現	・要件(1)を満たす音声・映像を提供すると、音声・映像の圧縮処理時間が増大する。 ・3地点以上の遠隔会議ではさらにアナログ音声ミキシング等の処理時間が増大する。
(3) 遠隔地にいる参加者の意思表示の把握	・相手と視線を合わせて会話する等の動作を実現させるためには特殊な装置や環境が必要になる。

第3章　円滑な会話を実現する遠隔会議システム

3.1　はじめに

本章では、多地点に分散している会議参加者が、伝送遅延時間の少ないインターネットを活用し、音声と実写映像を中心とするストリーミング情報をリアルタイムに送受信することで議論を行う遠隔会議の形態に焦点をあて、自然なコミュニケーションの実現手法[23]に基づいたシステムに関して議論する。

実空間内での円滑な会話の環境実現に関しては、テレビ電話等の遠隔コミュニケーションでの自由会話で許容される音声・映像の伝送遅延時間が、往復 400ms であるという研究報告[4]がある。更に、遠隔教育・学習の分野においては、音声・映像の伝送遅延時間が往復 300ms の場合でも、心理的な影響により学習効率が低下するという研究報告[5]もある。このように、会話の目的によって許容時間は異なるものの、円滑な会話実現には、より小さい伝送遅延時間の実現が最も有効であることが確認できる。よって、本研究では、多地点遠隔会議における音声・映像の往復伝送遅延時間の低減化に着目したシステムの提案と評価を中心に行う。また、本研究では、多地点遠隔会議を対象とするため、音声・映像等のストリーミング情報の符号化処理時間の短縮と共に、音声ミキシング等の加工処理時間の短縮も考慮したシステム構成を検討する必要がある。

一方、伝送遅延時間が小さくても、伝送される音声・映像等のストリーミング情報の品質が低ければ、相手の話し声が聞き取りにくい、相手の細かな表情や動作が把握しにくい、といった状況が発生し、結果として自然なコミュニケーションの実現は困難になる。このため、音声・映像等のストリーミング情報の符号化処理時間が低く、かつ品質が高い符号化方式の採用が必要である。

本研究では、多地点遠隔会議の場合での円滑な会話の実現手法として、音声・映像等のストリーミング情報の符号化方式の符号化処理時間が小さく品質も高い方式を採用する。そして、符号化されたストリーミング情報を IP パケットに搭載・転送し、音声ミキシング等の複数のストリーミング情報に対するリアルタイム加工処理も、ネットワーク中の IP パケットに搭載されたデータ形式のまま行い、再度 IP パケットに搭載・転送する、という手法を提案する。音声・映像等のストリーミング情報の符号化方式として、符号化処理時間が短くかつ符号化後の音声・映像等のストリーミング情報の扱いが容易な DV が､現時点では上記の手法を適用可能な符号化方式の１つであると考える｡そして、インターネット環境において、複数の遠隔参加者が参加し、自然コミュニケーションを実現する DV 遠隔会議システムを構築し、性能評価を行った。

3.2　多地点遠隔会議システム

まず、本論文で想定する多地点遠隔会議システム環境を**図 3.1** に示す。3 地点以上の地理的に離れた会場間が、インターネットで接続されている環境で、各会場にある音声・映像等のストリーミング情報の符号化・送信を行うエンコーダと、受信・復号化を行うデコーダにより、双方向のリアルタイム遠隔会議を行うことを想定する。同環境としては、各会場間をすべて双方向接続するフルメッシュ接続を行い、映像の選択や発言者の声量に合わせて行う音量レベル調整等の操作を、各会場で実施する形態も存在する。しかし、この形態では会場数の増大にともない、操作機材と操作者数も増大するため、遠隔会議の運用効率が低くなる。このため、本研究では、会場数に関係なく 1 箇所において、各会場に配信する映像の選択や発言者の声量に合わせて行う音量レベル調整等の操作を、会議参加者以外の第三者が、音声ミキサ等のリアルタイム加工処理装置を活用して実施する形態の、多地点遠隔会議環境を想定する。なお、図 3.1 では、同機能をインターネット上にある中間ノード内に設置する場合を示しているが、同機能がどこか１つの会場内に存在している場合も基本的な情報の流れは変化しない。

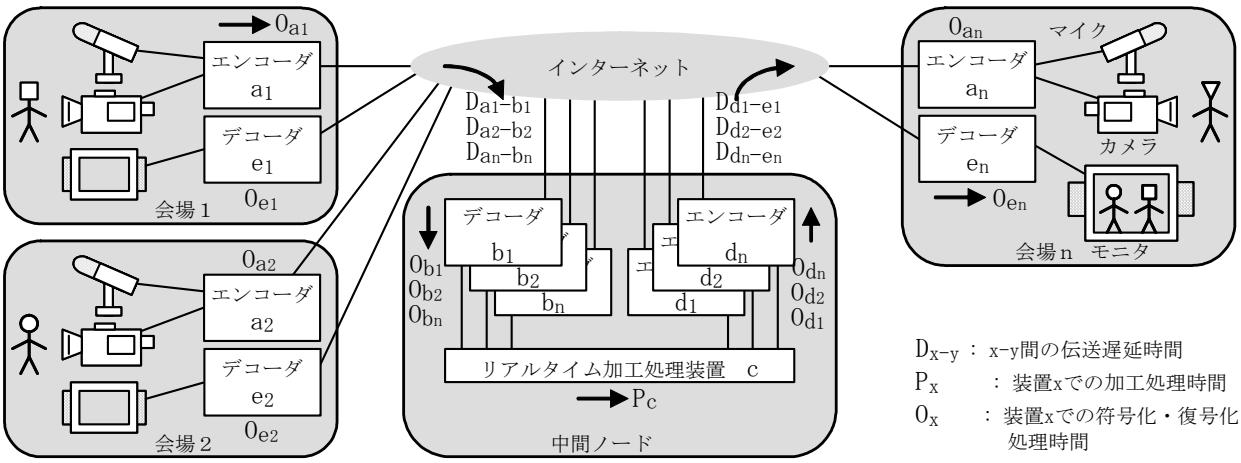


図 3.1　想定する多地点遠隔会議システムの概念図

このような想定環境において、我々の目指す自然なコミュニケーションを可能とする遠隔会議の実現のためには、多くの技術的課題がある。本論文では、相手の話し声が明確に聞き取れ、相手の細かな表情や動作が把握できる品質をもった音声・映像符号化方式の選定と、円滑な会話実現のために重要な音声・映像等の伝送遅延時間の短縮実現に絞って議論する。以下では、音声・映像等の伝送遅延時間に関する技術的課題について、図 3.1 を参照しながら伝送遅延時間短縮の可能性を項目別に説明する。図 3.1 内の英字

記号は、それぞれ装置 x から装置 y 間のネットワーク伝送遅延時間 D_{x-y}、装置 x における音声・映像等のリアルタイム加工処理時間 P_x、および符号化・復号化処理時間 O_x を表現しており、会場１でエンコーダ a_1 に入力された音声・映像等の情報が会場ｎのデコーダ e_n から出力されるまでの所要時間（片方向伝送遅延時間）T_{a1-en} は

$$T_{a1-en} = O_{a1} + D_{a1-b1} + O_{b1} + P_c + O_{dn} + D_{dn-en} + O_{en} \quad \cdots\cdots\cdots(式１)$$

となる。

ネットワーク伝送遅延時間は、伝送距離や伝送装置の構成等に依存するものであり、基本的には本研究で扱う遠隔会議システムの範囲外である。なお、インターネット伝送遅延時間に関しては、研究開発用ギガビットネットワークを活用した、ギガビットイーサネットによる高速 IP 実験網の性能評価実験が実施されており、IPv4 の ping コマンドによる往復伝送遅延時間(Round Trip Time:以下 RTT と記述)が、東京・大阪間で 13.2ms、東京・沖縄間で 46.8ms、という研究報告[24]がある。本論文ではこれらの値を１つの参考値として議論する。

加工処理時間は、デコーダから出力される複数のストリーミング情報に対して、音声ミキシング等の加工処理を行うための所要時間である。音声・映像情報に関しては、現在のところデコーダから出力されるアナログ情報をリアルタイム加工処理するため、ほぼ一定の所要時間となっており、単独での処理時間短縮は困難である。しかし、ネットワーク上を流れているデジタル情報を、デジタル状態のままリアルタイム加工可能な機構または装置が存在する場合は、中間ノード内のデコーダでのデジタル信号からアナログ信号へ変換する D/A 変換処理、およびエンコーダでのアナログ信号からデジタル信号へ変換する A/D 変換処理を省略することができるため、この加工処置時間の短縮は可能である。

符号化処理時間については、音声・映像等の符号化に要する処理時間が小さいエンコーダおよびデコーダを採用することで、遅延時間短縮が可能である。ただし、自然なコミュニケーションの実現にはある程度の音声・映像品質が必要であり、また符号化処理時間と品質はトレードオフの関係になる傾向が強いため、符号化方式の選定から行う必要がある。更に、符号化後のデジタル情報に対する加工処理が容易な符号化方式を採用した場合には、前述のとおり加工処理時間の短縮が可能であることも考慮する。

本研究では、音声・映像の符号化方式として、符号化後のデジタル情報に対しても編集・加工処理が容易な DV を採用する。そして、同情報をインターネット上でリアルタイム伝送する DVTS をエンコーダおよびデコーダとした双方向遠隔会議システムと、複数の DV ストリーミング情報をデジタル情報のままリアルタイム加工処理する機構を開発することで、自然なコミュニケーションの実現を試みる。

3.3　多地点デジタルビデオ遠隔会議システム

DV は、NTSC テレビ放送クラスの映像と 48-kHz サンプリング 16bits 線形量子化 PCM の音声を提供する。また、DV はフレーム間予測符号化を行わずフレーム内符号化のみを行いその符号化の圧縮率も低いため、MPEG-2[25]や MPEG-4[26]と比較して符号化処理時間が小さくかつ提供される音声・映像品質が高くなっている。DV 情報は、80 バイト長サイズの DIF ブロックを最小構成単位として、DIF ブロック 150 個のまとまりが DIF シーケンス、DIF シーケンス 10 個のまとまりがフレーム（30 フレームで 1 秒に相当）という構成になっている。**図 3.2** は 1DIF シーケンス内での DIF ブロックの配置構成を示しており、音声情報が格納される音声 DIF ブロック、映像情報が格納される映像 DIF ブロック、タイムコード等が格納されるサブコード DIF ブロック、映像の属性情報が格納される映像属性 DIF ブロック等から構成され、その格納位置は固定となっている。また、DV の符号化処理は 1 フレーム単位で行われており、DIF シーケンス 10 個（シーケンス番号 0～9）中に 1 フレーム分の音声情報および映像情報が格納される構成になっている。このため複数の DV 情報のデジタル状態でのリアルタイム加工処理が容易と考えられ、従来のアナログ状態への変換をともなったアナログ情報のリアルタイム加工処理を省略可能である。

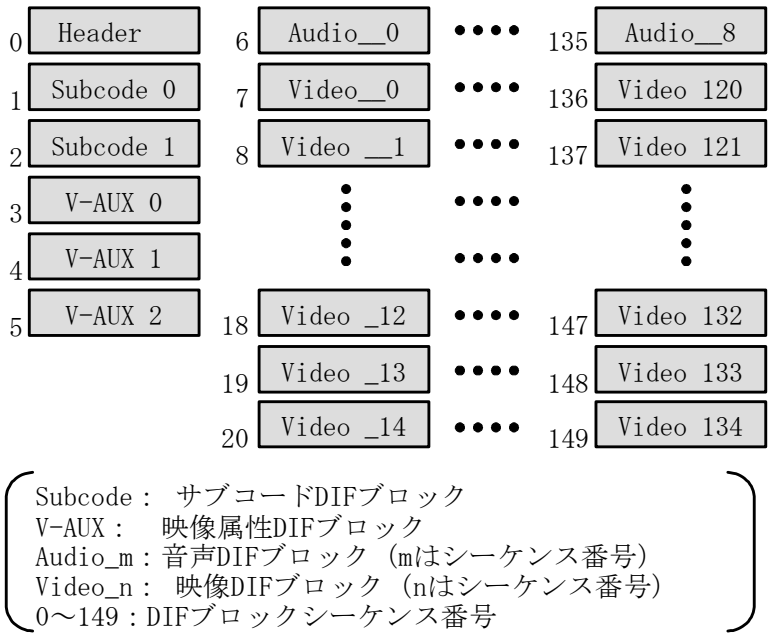


図 3.2　DIF シーケンス内の DIF ブロック配置順序

DVTS は、IEEE1394 ケーブル上を流れてくる DV 情報を適当な数の DIF ブロック一塊として RTP/UDP/IP パケットに搭載し、インターネット上にリアルタイム伝送する DVTS 送信機能と、インターネット上を流れてくる IP パケットから DV 情報を再構築し、IEEE1394

ケーブル上へリアルタイム伝送する DVTS 受信機能を持っている。DVTS は様々な OS 環境の通常仕様の PC で稼動し、特に Windows OS 環境では PC のモニタ画面上での映像出力およびスピーカへの音声出力が可能であり、DV 情報の IEEE1394 規格への格納作業が必要なくなる。このため、FreeBSD OS、Linux OS で稼動する DVTS 受信 PC と比較して、最終出力までに要する処理時間が約 1 フレーム分短縮され、結果往復伝送遅延時間の短縮が実現できる。

図 3.3 は、DVTS を活用した 3 地点 DV 遠隔会議での、各会場から会場３へ向けた片方向の音声・映像等のストリーミング情報の流れを示している。まず、アナログ音声・映像情報が、各会場の音声・映像信号測定ポイント X1、X2、X3 から入力され、A/D 変換後に DVTS 送信 PC から中間ノードに向けて伝送される。中間ノードでは、DVTS 受信 PC で同情報を受信し、D/A 変換装置、アナログ音声・映像加工処理装置、A/D 変換装置を経由して DVTS 送信 PC から会場３に向けて伝送される。そして、会場３の DVTS 受信 PC で同情報を受信し、音声・映像信号測定ポイント Y3 から DV 情報が出力される。なお、会場１および２での DVTS 受信 PC においても、同様の音声・映像情報の流れがある。この多地点 DV 遠隔会議システムは、MPEG-2 や MPEG-4 を活用した場合と比較して符号化処理時間が小さいため、ある程度の往復伝送遅延時間の短縮は期待できる。しかし、この遠隔会議システム構成には、中間ノードでの A/D および D/A 変換をともなうアナログ音声・映像加工処理装置が、伝送遅延時間の増大と音声・映像等の品質劣化を生じさせるという技術的課題がまだ残っている。このため、自然なコミュニケーション実現には不十分であり、これを改善することで更なる往復伝送遅延時間の短縮を図る必要がある。

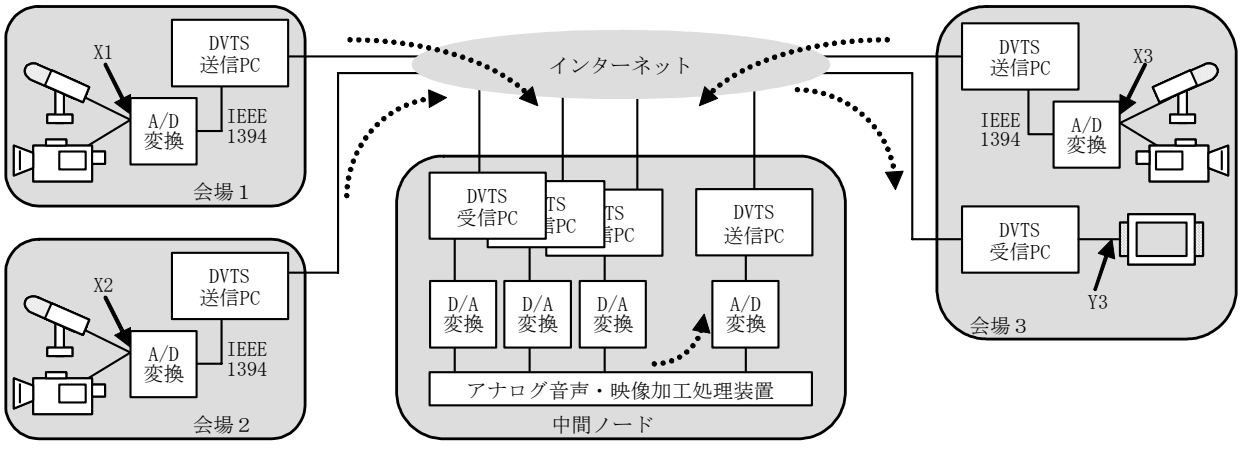


図 3.3　3 地点 DV 遠隔会議システムでの会場３へ向かう情報の流れ

3.4　デジタルビデオ・リアルタイム処理機構

本研究では、中間ノードでの A/D および D/A 変換をともなう、アナログ音声・映像加工処理による伝送遅延時間の増大と音声・映像等の品質劣化に関する技術的課題を解決するため、DV リアルタイム処理機構を開発した。先で述べたように、DV の符号化処理は 1 フレーム単位で行われている。特に 48-kHz サンプリング 16bits 線形量子化 PCM の音声情報に関しては、**図 3.4** に示すように 2 チャネル（図中の ch-0、ch-1）の PCM 音声情報各 1 フレーム分が 8 バイト長のヘッダー情報と、16bits サンプリング音声データ群である 72 バイト長のデータ領域から構成される音声 DIF ブロック 90 個に分割格納される。また、この 90 個の音声 DIF ブロックは、1,500 個の DIF ブロックで構成される 1 フレーム分の DV 情報の中の固定格納位置に配置される。これは、複数の会場から伝送されてくる DV 情報各 1 フレーム分に対して、同一の DIF シーケンスのシーケンス番号（0～9）で、かつ同一の DIF ブロックのシーケンス番号（0～149）をもつ複数の音声 DIF ブロックを取り出すことで、容易にデジタル加工処理が可能であることを示している。ここで、複数の線形量子化 PCM 音声データに対する音声ミキシングは、音声サンプリングデータ単位の加算演算で実現されるため、図 3.4 の構成である複数の音声情報に対する音声ミキシングは、同一の格納位置に相当する音声 DIF ブロック内のサンプリングデータ単位の加算演算で実現できる。一方、DV の映像情報に関しては 1 フレーム内で完結しているフレーム内符号化処理であるため、複数の会場から伝送されてくる DV 情報各 1 フレーム分に対して、1 フレーム分の映像 DIF ブロック 1,350 個すべてを配信したい映像情報の DIF ブロックで上書きすることで配信映像の選択も実現できる。

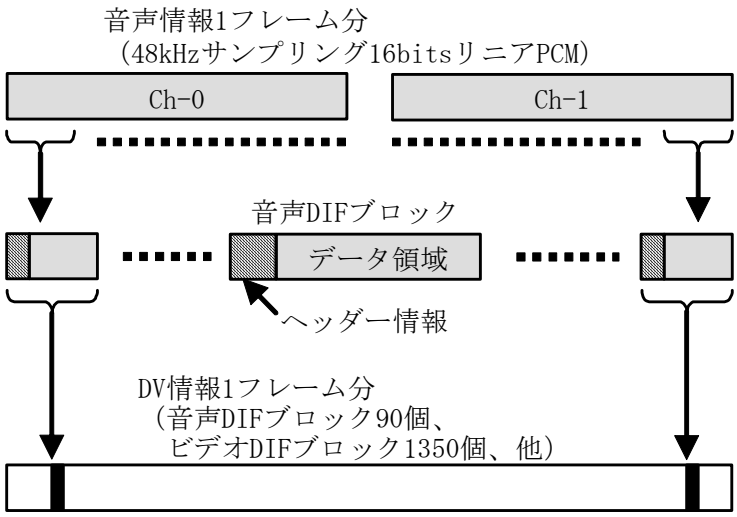


図 3.4　音声情報の DV 情報への格納手法

本研究では、上述した DV 情報の特性を活用することで、多地点から到着する DV 情報のデジタル状態での音声ミキシング処理、および全地点への配信映像の選択等を行う、DV リアルタイム処理機構を実現した。**図 3.5** は、図 3.3 の構成に、DV リアルタイム処理機構を導入した場合の 3 地点遠隔会議システム構成と、各会場から会場３へ向けた片方向の音声・映像等のストリーミング情報の流れを示している。DV リアルタイム処理機構は、各会場から伝送される DV 情報を受信する DV 受信機能 R1、R2、R3、1 フレーム分の DV 情報を一時格納する共有メモリ、共有メモリ内の DV 情報に対する加工処理を行うリアルタイム DV 処理機能、各会場に向けて DV 情報を送信する DV 送信機能 S1、S2、S3 から構成される。なお、会場１および２での DVTS 受信 PC においても、同様の音声・映像情報の流れがあり、その場合は図 3.5 内の DV 送信機能 S3 が、それぞれ DV 送信機能 S1 および S2 に置き換わる。本機構は、図 3.3 の中間ノードの機器構成と比較して A/D および D/A 変換が省略され、複数の PC で実行していた DV 送受信機能とアナログ音声・映像加工処理装置で実施していたリアルタイム加工処理機能を 1 台の PC 上のアプリケーションプログラムとして稼動する。

図 3.6 は、3 地点遠隔会議システムでの DV リアルタイム処理機構全体の流れを示している。各会場１、２、３から伝送されてくる DV 情報は、それぞれ DV 受信機能 R1、R2、R3 において RTP/UDP/IP パケットから取り出される。このうち配信映像として選択されている DV 情報のみが配信 DV 情報用共有メモリに格納される。そして、格納された順番で随時共有メモリから DIF シーケンスのシーケンス番号 p、DIF ブロックのシーケンス番号 q の DIF ブロックが読み出され、DV 送信機能 S1、S2、S3 から各会場の DVTS 受信 PC を目的地とした RTP/UDP/IP パケットに搭載され、転送される。ただし、共有メモリから読み出した DIF ブロックが音声 DIF ブロックである場合は、各会場から伝送されてくる音声の音量調整を行う音声ミキシング処理が追加される。

図 3.7 は、DV リアルタイム処理機構内での音声ミキシング処理の流れを示している。まず、3 地点遠隔会議システムの各会場１、２、３から伝送されてくる DV 情報のうち音声 DIF ブロックのみが、音声 DIF ブロック用共有メモリ M1、M2、M3 内の所定位置にそれぞれ格納される。図 3.6 の配信 DV 情報用共有メモリから音声 DIF ブロックが読み出された場合、そのDIF シーケンスのシーケンス番号 p および DIF ブロックのシーケンス番号 q が同一である音声 DIF ブロックを、音声 DIF ブロック用共有メモリ M1、M2、M3 からそれぞれ取り出す。そして、各会場向けに設定された音声ミキシング加算演算処理を行い、元の音声 DIF ブロックと入れ替えることで音声ミキシングをリアルタイムで実現する。

図 3.8 は、上述の音声ミキシング処理の中核である、音声ミキシング用加算演算処理の仕組みを示しており、ミキシング演算用に取り出された音声 DIF ブロック内の音声データがすべて 16bit 線形量子化 PCM データであるとして、その流れを説明する。この場合、会場１の音声データは 16bit 単位の線形量子化 PCM データ群[A1, A2, ･･･]となり、

同様に会場 2 の音声データは[Ｂ1，Ｂ2，･･･]、会場 3 の音声データは[Ｃ1，Ｃ2，･･･]と表現できる。すると、会場ｎ（n=1,2,3）へ送信する音声情報として、会場１、２、３それぞれの音声の音量をそれぞれα倍、β倍、γ倍にして音声ミキシングする場合は、会場ｎへの送信音声データ群[Ｎ1，Ｎ2，･･･]は、

$$Nx = \alpha \cdot Ax + \beta \cdot Bx + \gamma \cdot Cx \qquad \cdots\cdots\cdots(式2)$$

（ただし、x は 1 から最大 36 までの整数）

という単純加算演算で表現できる。そして、各会場の音声のボリューム調整は DV リアルタイム処理機構において自由に変更可能である。

なお、本 DV リアルタイム処理機構のプログラム開発を開始するにあたり、DV での音声情報がどのように配置されているのかについてサーベイ（**付録Ｂ**参照）した結果、一部の民生用 DV 機器では、音声情報が映像情報と同期していない機種が存在していることが判明した。その場合、DIF ブロック内の PCM データの総数が、同一シーケンス番号（ｐ，ｑ）の DIF ブロック同士であっても、異なるケースが稀ではあるが発生する。その結果、PCM データ 1 サンプル分ではあるが、正常な音声ミキシングはできず、場合によっては雑音が挿入することになる。このため、複数の DV 情報から PCM 音声データサンプルをそれぞれ取り出し加算演算することで、音声ミキシングを行う DV リアルタイム処理機構を安定的に機能させるためには、音声情報が映像情報と同期する機能をもつ DV 機器を用いることが必要である。

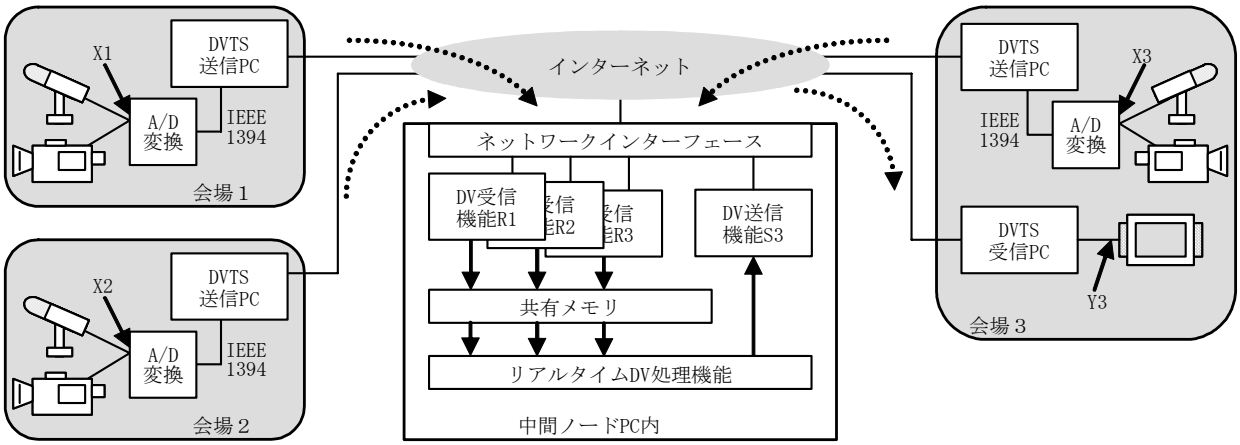


図 3.5　DV リアルタイム処理機構を導入時の会場３へ向かう情報の流れ

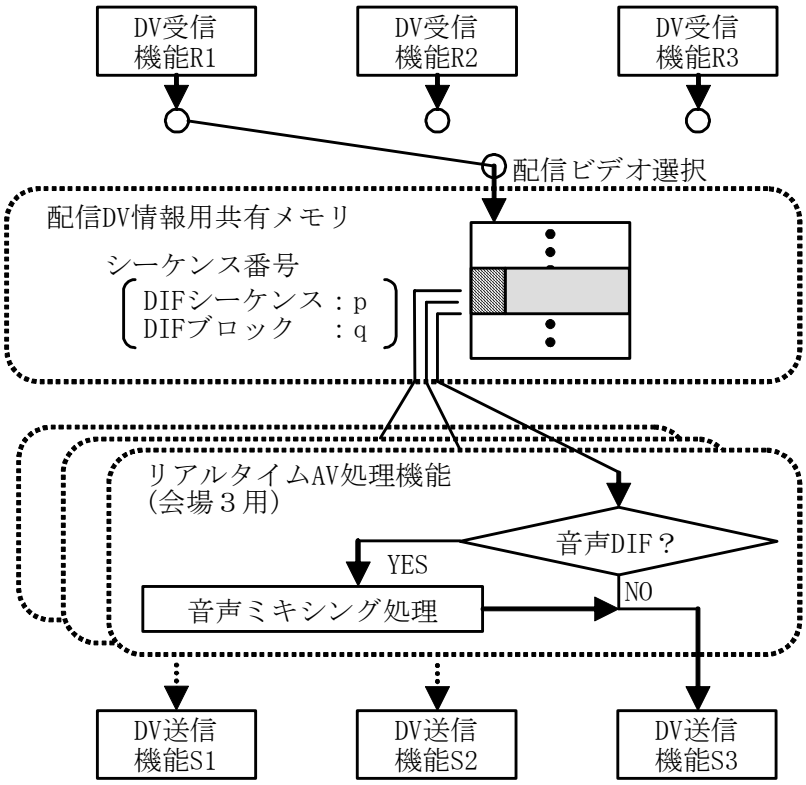


図 3.6　DV リアルタイム処理機構全体の流れ

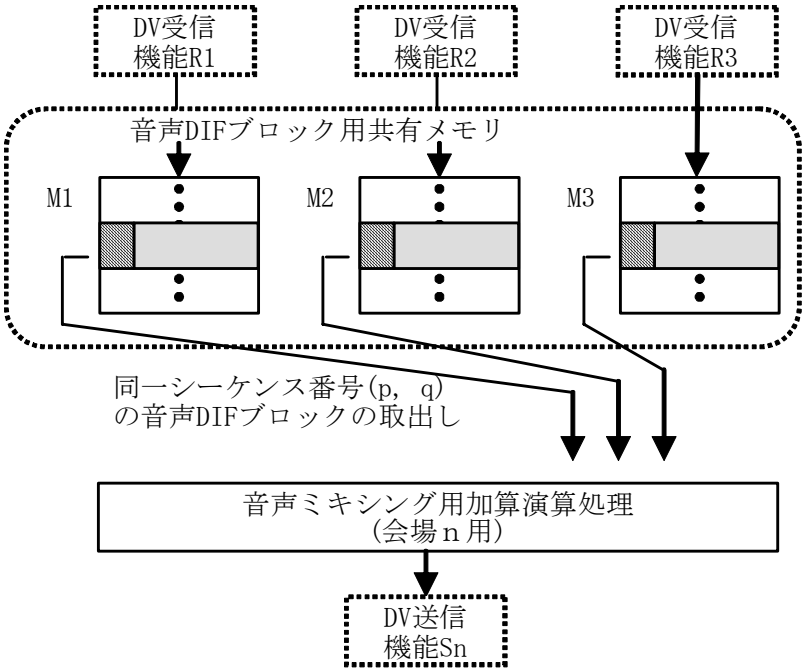


図 3.7　音声ミキシング処理の流れ

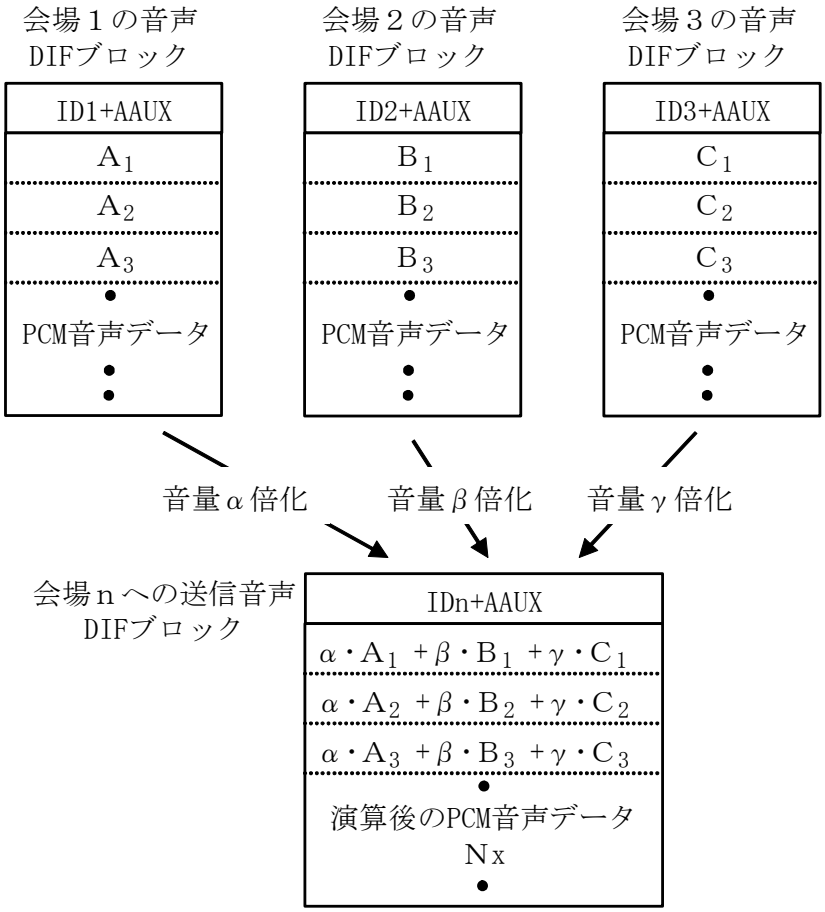


図 3.8　音声ミキシング用加算処理の流れ

3.5　評価実験

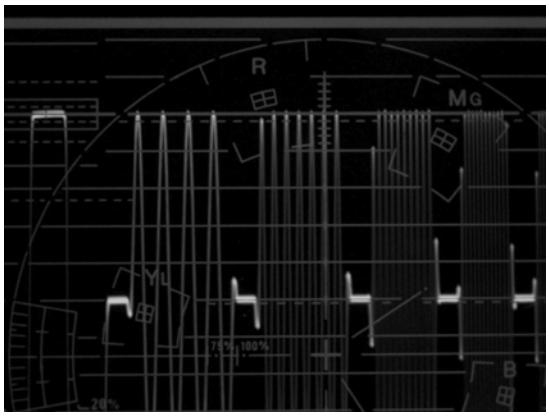
図 3.3 に示すようなアナログ音声・映像加工処理装置を用いた DV 遠隔会議システム構成（以後、構成 A と記述)、我々の提案する図 3.5 に示すような DV リアルタイム処理機構を用いた DV 遠隔会議システム構成（以後、構成 B と記述)、図 3.5 の構成で中間ノード PC を用いず DVTS 送信 PC から DVTS 受信 PC への直通伝送を行う 2 地点間 DV 遠隔会議システム構成（以後、構成 C と記述）の 3 つのシステム構成を対象にして、比較性能評価を行う。評価項目は、伝送する音声・映像情報の品質劣化度の評価、伝送遅延時間の測定評価、妨害品質尺度による主観的評価である。今回の比較性能評価に用いた主な機器の諸元を**表 3.1** に示す。

表 3.1　主な機器の諸元

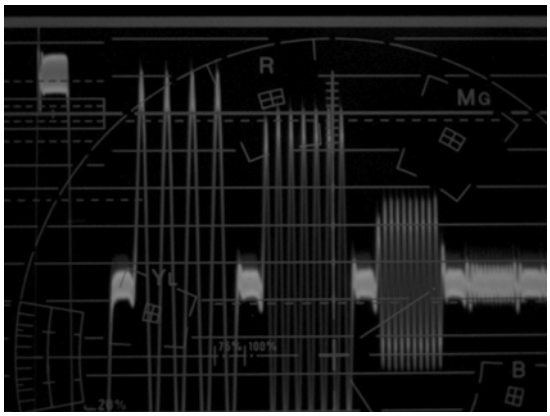
機器名称	諸元
DVTS送信PC	DVTS version 1.0b OS : FreeBSD 4.4 CPU : Pentium-3 1GHz Memory : 256MB
DVTS受信PC	DVTS for WinXP OS : Windows XP CPU : Pentium-M 1.6GHz Memory : 512MB
アナログAV 編集装置	FUTEC MXProDV (アナログAVポート使用)
A/D、D/A変換	Canopus ADVC-100
中間ノードPC	OS : FreeBSD 4.4 CPU : Pentium-3 1GHz Memory : 256MB
DVTS受信PC (中間ノード内)	DVTS version 1.0b OS : FreeBSD 4.4 CPU : Pentium-3 1GHz Memory : 256MB
DVTS送信PC (中間ノード内)	DVTS for WinXP OS : Windows XP CPU : Pentium-M 1.6GHz Memory : 512MB

3.5.1　品質劣化度の評価

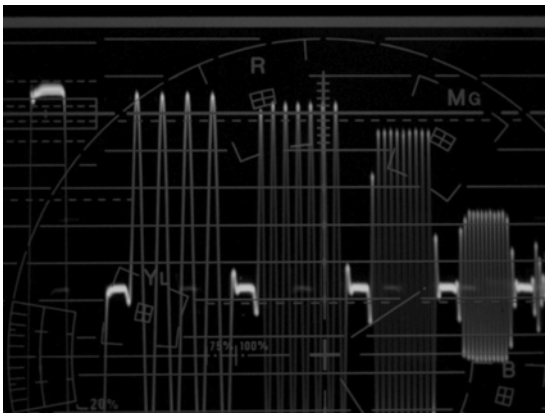
品質劣化度の評価では、まず構成 A、B、C それぞれの音声・映像信号測定ポイント X1 に特定のテストパターン信号を入力し、音声・映像信号測定ポイント Y3 から出力信号を取り出す。そして、入力されたテストパターン信号と出力信号をオシロスコープ表示画面で比較し、出力表示波形の差異から評価を行った。なお、本評価でのインターネット環境は、スイッチングハブ 1 台のみの構成で実施している。**図 3.9** は、テストパターン信号として NTSC ジェネレータで生成された 100 RE マルチバースト信号のポイント X1 での入力波形と、構成 A、B、C それぞれのポイント Y3 での出力波形を示している。この図から、入力波形に対する構成 B での出力波形の歪の度合いは、構成 A での出力波形の歪の度合いと比較して小さくなっており、構成 C の出力波形とほぼ一致していることが分かる。このため、本研究で提案する DV リアルタイム処理機構を用いた DV 遠隔会議システム構成は、アナログ音声・映像加工処理装置を用いた場合と比較して品質劣化を低く抑えることができ、かつ DV リアルタイム処理機構での品質劣化はないといえる。



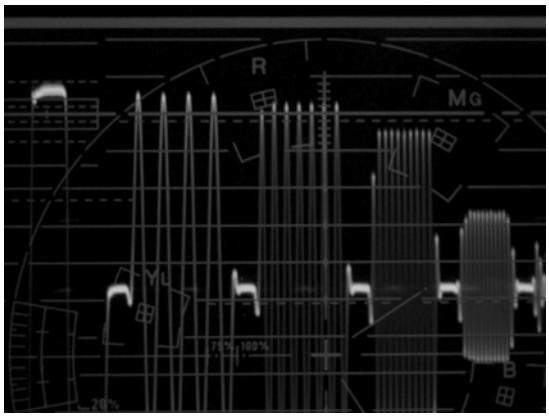
入力波形：100 RE マルチバースト信号



構成A（アナログ音声・
映像編集加工処理装置を使用）



構成B（DVリアルタイム処理機構を使用）



構成C（2地点DV直通伝送）

図 3.9　入力波形とポイント Y3 での出力波形の比較

3.5.2　伝送遅延時間の測定評価

伝送遅延時間の測定評価では、まずNTSCジェネレータにおいて、ある特定のテストパターン画像から別のテストパターン画像に切り替える一連の動作を行い、このNTSCジェネレータからの出力信号を、構成A、B、Cそれぞれの音声・映像信号測定ポイントX1に入力する。そして、音声・映像信号測定ポイントY3からの出力信号と、ポイントX1の入力信号（NTSCジェネレータからの出力信号）をオシロスコープ表示画面にて比較し、両者の信号波形の時間軸方向のずれの大きさから、2地点間の伝送遅延時間の測定を行った。**図3.10**に、ポイントX1からポイントY3までの被測定系区間での片方向伝送遅延時間の測定方法を示す。本測定評価では、オシロスコープ上に表示される信号波形が明らかに異なる2種類のテストパターン信号が効果的であるため、今回は画面すべてが赤色のテストパターン信号と画面すべてが青色のテストパターン信号を用いた。なお、この測定方法では映像信号の伝送遅延時間を測定することになっているが、DVでの音声情報と映像情報は同期しているため、音声信号と映像信号の伝送遅延時間は等しく、今回の測定結果はDVストリーミング情報の伝送遅延時間といえる。また、本測定評価でのインターネット環境は、スイッチングハブ1台のみの構成で実施しているため、今回の測定結果にはネットワーク伝送遅延時間は含まれない。

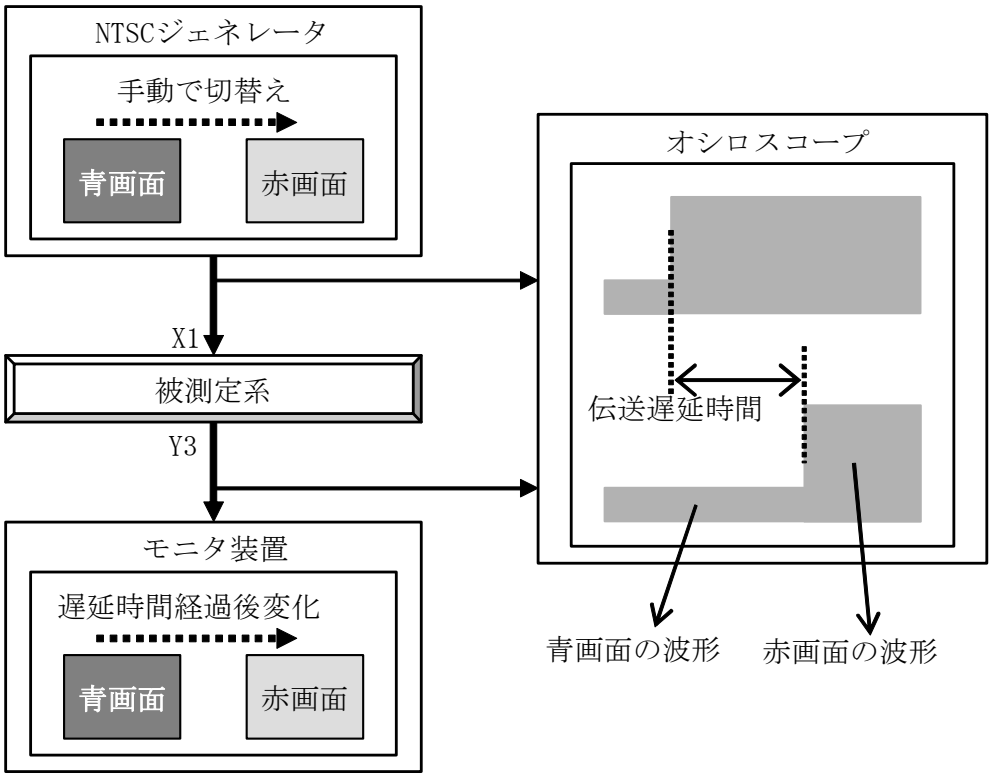


図 3.10　片方向伝送遅延時間の測定方法

表 3.2　片方向伝送遅延時間の測定結果

システム構成	遅 延 時 間 （平均値）	（標準偏差）
アナログ音声・映像編集加工処理装置を使用（構成Ａ）	447.8 ms	4.5 ms
DVリアルタイム処理機構を使用（構成Ｂ）	191.0 ms	2.5 ms
2地点DV直通伝送（構成Ｃ）	156.8 ms	1.9 ms

表 3.2 に、各システム構成での音声・映像信号測定ポイント X1 から Y3 までの片方向の伝送遅延時間を、10 回ずつ測定した平均値と標準偏差を示す。この測定結果から、DV リアルタイム処理機構を導入した構成 B の伝送遅延時間は、構成 A の場合と比較して大幅に短縮されていることが分かる。また、構成 B と構成 C の違いは基本的に中間ノード PC の有無のみであることから、DV リアルタイム処理機構での加工処理時間は約 34ms と推定できる。同機構のプログラムでは、DV の音声情報が 1 フレーム単位で処理を行う必要から 1 フレーム分のバッファを利用しているので、加工処理時間のうち 33ms はバッファリング時間と考えられ、プログラムの改良による短縮可能時間は最大で 1ms 程度と推定できる。

次に、この測定結果から推定される、DV ストリーミング情報の往復伝送遅延時間の考察を行う。ネットワーク伝送遅延時間を考慮しない場合の往復伝送遅延時間は、従来のアナログ音声・映像加工処理装置を使用した構成Ａでは 895.6ms になるのに対し、本研究で提案する DV リアルタイム処理機構を使用した構成Ｂは 382.0ms となっている。これは、DV リアルタイム処理機構の導入だけで、往復伝送遅延時間が約 57%短縮できていることを示している。

ここで、DV リアルタイム処理機構を利用した多地点遠隔会議での、片方向伝送遅延時間の理論限界値について考察する。インターネット上の伝送を考慮しない場合、基本的にはカメラ等からのアナログ音声･映像信号から 1 フレーム分の DV 情報に変換を行う A/D 変換装置内、DV リアルタイム処理機構における DV 音声情報ミキシング用、そして D/A 変換装置で 1 フレーム分の DV 情報を受信してアナログ音声・映像に変換する D/A 変換装置内で､それぞれ 1 フレーム分バッファリングが必要であり､理論最小値は計 90ms となる。しかし、インターネット上の伝送を含めて考えた場合、ネットワーク伝送遅延時間のほかに、DVTS 送信 PC での IP パケットの送信バッファと、DVTS 受信 PC での受信バッファが現在必要であり、そのバッファ容量は PC の処理能力とインターネット回線の伝送能力に大きく依存する。インターネットの特性を考えると、これらのバッファ容量を０にするのは困難であるが、それでもこれらのバッファリング時間の理論最小値はほぼ０にな

る可能性はある。よって、DV リアルタイム処理機構を利用した多地点遠隔会議での片方向伝送遅延時間の理論最小値は、90ms と物理的に離れている地点間の信号伝送に要する時間の合計値となり、これは片方向のネットワーク伝送遅延時間が 110ms 以内であれば多地点遠隔会議において円滑な会話が実現できる可能性があることを示している。

更に、現在の符号化技術や情報伝送方式の制約を考えない場合の理論限界値についても考察する。本質的な映像情報の流れは、カメラの CCD で受光する光信号をデジタル電気信号に変換し、その電気信号をネットワーク上に伝送し、適切な情報加工処理を行った後、モニタ画面に再度光信号として表示する、というステップになる。たとえば、カメラの CCD で変換したデジタル電気信号を変換・圧縮することなく直接ネットワーク上に伝送するシステムが開発できれば、大幅な処理遅延時間短縮につながる。また、ネットワーク部分も、インターネットのようなパケット転送方式ではなく、高速な同期通信ネットワークシステムが開発できれば、情報の送受信機能部分において 1 フレーム分等の巨大なバッファを用意する必要がなくなる。途中の情報加工処理についてはその加工処理内容によって大きさは変動するものの、ある程度のバッファが必要になると考える。このため、本質的にかかってしまう処理時間は、カメラの CCD で受光する光信号からデジタル電気信号に変換処理時間、同信号の送受信処理時間、ネットワーク上を流れる信号の伝搬遅延時間、モニタ画面への再表示処理時間、そして途中の情報加工処理時間の 5 項目になる。このうち、前者 4 項目は 5 番目の情報加工処理時間と比較すると極端に小さくほぼ 0 と考えてよい。以上から、本質的にかかってしまう処理時間は、途中の情報加工処理時間に等しいと言え、たとえば 1 フレーム分の複数の映像情報を一括で加工処理するような場合はこの 1 フレーム分の時間が理論限界値となる。

3.5.3　妨害品質尺度による主観的評価

文献[4]の評価手法を参考に、構成 A（アナログ音声・映像加工処理装置を用いた DV 遠隔会議システム構成)、構成 B（DV リアルタイム処理機構を用いた DV 遠隔会議システム構成)、構成 C（2 地点間 DV 遠隔会議システム構成）という 3 種類の遠隔会議模擬環境それぞれに対し、2 つの会場（会場１と会場３）に被験者 1 名ずつ配置し、自由会話を行ってもらう形態での主観的評価を実施した。

被験者は、NICT の事務系職員 20 名で、男性 12 名女性 8 名、遠隔会議の経験者 10 名未経験者 10 名という内訳である。

評価手法は、会場１および３にいる被験者 1 名ずつによる 1 分間の自由会話後に、**表 3.3** に示す 5 段階妨害尺度によるオピニオン評点評価を行ってもらう実験を、評価基準となる往復伝送遅延時間がゼロの場合と、評価対象の遠隔会議システム構成の場合を交互に行った。

表 3.4 は、評価した 5 種類の遠隔会議システム構成の内訳と評価結果を示している。

本実験では、遅延発生装置を用いることでネットワーク伝送遅延時間を模擬しており、特に本研究で提案するDVリアルタイム処理機構を活用した構成Bに関しては、東京・大阪間のインターネット環境を想定した片方向 7ms の遅延のほかに、東京・沖縄間を想定した片方向24msの遅延と参考用の片方向100msの遅延の場合も評価実験を行った。表3.4から、評点が 4 以上つまり「自由会話をしていて遅延が気にならない」と評価した被験者は、東京・大阪間を想定したインターネット環境での構成Aでは10%であるのに対し、同インターネット環境での構成BおよびCでは90%であった。なお、今回のオピニオン評点の結果の違いが、システム構成の違いによるものなのかを判定するため、分散分析[27]を実施し、その結果を**表 3.5** に示す。その結果、棄却率5%（判定結果の信頼度95%に相当）で、オピニオン評点結果はシステム構成によって差異があると判定された。また、東京・沖縄間を想定したインターネット環境での構成 B の場合も、70%の被験者が「自由会話をしていて遅延が気にならない」と評価している。このため、DV リアルタイム処理機構を用いたDV遠隔会議システムを利用することで、遠隔地間での自由会話が実現され、遠隔会議での自然なコミュニケーション実現に大きく近づいたといえる。

表 3.3　5段階妨害品質尺度

評点	品質尺度
5	遅延が全く感じられない
4	遅延がわずかに感じられるが、気にならない
3	遅延がわずかに気になる
2	遅延が気になる
1	遅延が非常に気になる

表 3.4　妨害品質尺度による主観的評価結果

システム構成	ネットワーク模擬遅延時間	推定される往復遅延時間	オピニオン評点の分布 [5]	[4]	[3]	[2]	[1]	オピニオン評点平均値
アナログ音声・映像編集加工処理装置を使用（構成A）	7 ms	909.6 ms	0	2	7	11	0	2.6
DVリアルタイム処理機を使用（構成B）	100 ms	582.0 ms	0	4	14	2	0	3.1
同上	24 ms	430.0 ms	2	12	6	0	0	3.8
同上	7 ms	396.0 ms	2	16	2	0	0	4.0
2地点DV直通伝送（構成C）	7 ms	327.6 ms	12	6	2	0	0	4.5

表 3.5　システム構成とオピニオン評点結果に関する分散分析表

変動要因	変動	自由度	分散	分散比	P-値	F 境界値 (棄却域5%)
システム構成	41.03	2	20.52	53.28	8.97E-14	3.16
繰り返し誤差	21.95	57	0.39	-	-	-
合計	62.98	59	-	-	-	-

3.6　まとめ

本章では、多地点に分散している会議参加者が、地上のインターネットを活用し、音声と実写映像を中心とするストリーミング情報をリアルタイムに送受信することで会議を行う、双方向リアルタイム遠隔会議に焦点をあて、自然なコミュニケーションの実現手法に基づいたシステムに関して議論を行った。そして、多地点遠隔会議の場合での円滑な会話の実現手法として、符号化処理時間が短くかつ符号化後の音声・映像等のストリーミング情報の扱いが容易な DV を採用し、インターネット上を流れる IP パケットに搭載されているデジタル状態の複数の DV ストリーミング情報に対して、中間ノードで直接加工処理を行う DV リアルタイム処理機構を提案した。そして、本機構と DVTS を併用することで、NTSC テレビ放送クラスの音声・映像を用いた多地点遠隔会議の模擬システムを構築し、定量測定実験および被験者による評価実験を行った。評価の結果、DV リアルタイム処理機構を用いた多地点遠隔会議システムは、従来の遠隔会議システムと比較して、相手の表情まで把握できる NTSC 放送品質クラスの映像を提供し、かつ往復伝送遅延時間を 57%低減化することに成功し、円滑な会話の実現可能性を示した。また、多地点遠隔会議環境での同機構の利用は、ストリーミング情報に対するアナログ・デジタル間信号変換処理回数の削減につながり、結果ストリーミング情報の品質劣化抑制という面においても有効に機能していることがわかった。

なお、中間ノードでの複数ストリーミング情報に対するリアルタイム加工処理機能は、DV 情報に限定された機能ではなく、また音声ミキシングや配信映像の選択以外のデジタル加工処理にも適応可能である。このため、同機能は今後別の形態で有効利用される可能性があると考えられる。

第4章　実空間共有を実現する遠隔会議システム

4.1　はじめに

インターネットの広帯域化、および H.323 等の遠隔会議システム標準プロトコルの普及により、近年、遠隔会議の実用利用の機会が増加してきている。特に H.323 を活用したシステムでは、インターネット等を経由して、遠隔参加者がカメラを操作することが容易に実現できるため、遠隔参加者が会議室内の注目したい映像を自分の意思で選ぶ等、会議への積極的な行動がある程度可能になる。しかし、会議室側の参加者には小さな遠隔カメラが動作するのを確認できるものの、遠隔参加者の映像は固定のままであるため、今どこに注目しているのか等の、同一実空間内にいる参加者同士ではアイコンタクトや顔の方向によって瞬時に判断できる情報が把握できない。また、遠隔操作インタフェースが、リモコンのボタン押下やマウスクリック等によるメニュー選択という場合も、会議への集中度を維持したまま遠隔操作するという視点では、まだ改善の余地があると考える。このように、現在の遠隔操作システムは、単一会議室内で行われる通常の対面会議では可能な、自然なコミュニケーションの実現を妨げる要因が多いといえる。

遠隔会議システムにおいて、自然なコミュニケーションを実現するためには、遠隔参加者が会話をしたいと思う相手の方向に顔を向け、アイコンタクトをおこなうことのほかに、会議室内の参加者が遠隔参加者の視線方向や表情を把握できることが重要である。そして、これは複数の物理的に離れている会場空間であっても、同一会場空間の場合と同じような空間共有を体感できる仕組みの実現で解決できる。このような空間共有実現には、仮想空間上にアバタと呼ばれる自分の分身を配置して、アバタ同士による自然なコミュニケーションを実現させようとする手法[14]-[17]や、遠隔参加者の分身として他の参加者と並んで、同一実空間上の会場内に筐体（以後、代理人ロボットと記述する）を置き、遠隔参加者が遠隔操作することで自然なコミュニケーションを実現させようとする手法[28][29]が存在する。両手法とも遠隔会議への適用例は多くはないが、顔の向き・視線の制御やジェスチャ機能を活用することで、自然なコミュニケーションを実現する可能性を十分有しているといえる。

本研究では、現時点ではまだアバタのリアルタイム表示処理よりも実写映像の表示処理の方が比較的に容易であること、仮想空間を表現する大型スクリーンや特殊な表示装置を使わずに既存の会議室等に導入しやすい、といった理由から、今回、実空間上に代理人ロボットを設置し、実写映像を用いる手法で空間共有実現を試みることにした。ただし、同手法を遠隔会議に用いる場合、遠隔参加者が何らかの入力デバイスで遠隔操作を自ら行う必要があるため、自然なコミュニケーションの実現には遠隔操作が簡便で、かつ会議への集中力を維持できる工夫が必要である。遠隔カメラの簡便な遠隔操作を実

現するための入力デバイスと遠隔制御手法に関しては、これまでも研究[30][31]されている。しかし、遠隔操作によって遠隔会議映像が表示されている画面から目を離す等の会議への集中力低下の軽微化という視点で研究は少ない。本論文では、この部分に焦点をあてた議論と考察を行う。

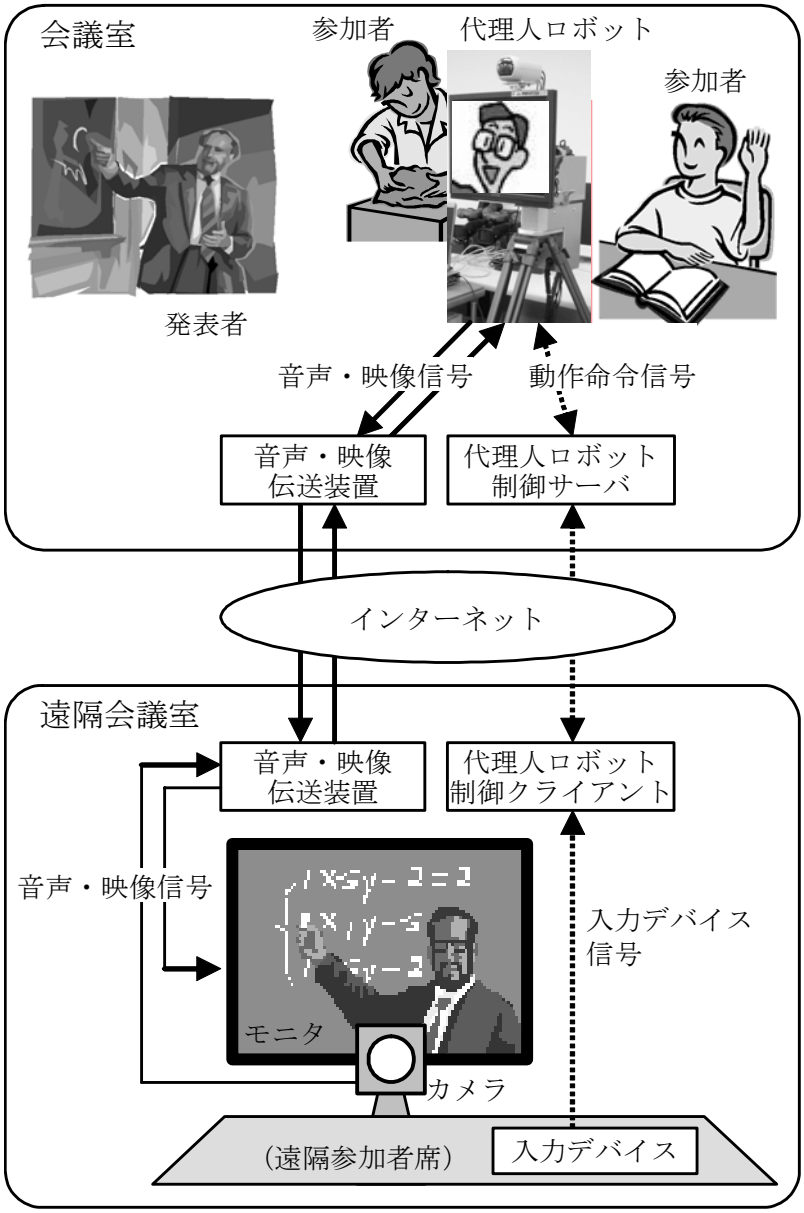


図 4.1　代理人ロボットを利用した遠隔会議システムの概念図

本章では、まず遠隔参加者の目となるビデオカメラ、耳となるマイクロフォン、口となるスピーカおよび顔映像を出力する液晶ディスプレイを一体として、上下左右旋回する機能をもつ代理人ロボットを開発した。次に、同装置を遠隔操作するための入力デバ

イスとして、マウスおよびジョイスティックをそれぞれ用いる 2 種類の遠隔操作システムを開発し、遠隔操作が簡便でかつ会議への集中力を維持できる遠隔操作システムに関する比較評価実験を行った。その結果から、遠隔参加者が会議への集中度を維持したまま遠隔操作が可能な遠隔操作システムについて考察する。また、代理人ロボットとその遠隔操作システムが、遠隔地間の参加者同士のアイコンタクトおよび話者交替の実現の面で有効に機能していかどうかの評価を、遠隔会議模擬実験での被験者の行動分析とアンケート評価から試みる。

4.2 代理人ロボット

本論文で想定する代理人ロボットを利用した遠隔会議システム環境を図 4.1 に示す。会議室の参加者と遠隔会議室にいる遠隔参加者は、インターネットを介して会議を行う。会議室には、参加者のほかに遠隔参加者の分身として、他の参加者と並んで実際に会議に出席する代理人ロボットが設置される。音声・映像伝送装置により代理人ロボットのモニタ画面上に遠隔参加者の顔の実写映像が表示されると共に、頭部カメラにより会議室の映像が遠隔会議室内のモニタに表示される。また、代理人ロボット制御クライアントおよびサーバを介することで、遠隔会議室内の遠隔参加者は、入力デバイス操作によって代理人ロボットの頭部旋回動作等の遠隔操作を行うと想定する。

表 4.1 代理人ロボットの諸元

項目	データ
全長 重量 最大消費電流	155 cm 25 kg 2 A （全て移動キャスター装着状態で遠隔カメラ及び液晶モニタ非装着状態）
頭部旋回角度	左：約90° 右：約90° 下：約60° 上：約30° （全てモニタ正面位置を基準）

4.2.1　システム構成

遠隔会議室にいる参加者と会議室内の参加者が、共に同一空間で一緒に議論しているというような空間共有感覚を高めることを目的とし、代理人ロボットを試作した。**表 4.1** および**図 4.2** に、試作した代理人ロボットの諸元とシステムの概要を示す。代理人ロボットの本体は、遠隔カメラ、マイクロフォン、モニタ、スピーカから構成される頭部と、その頭部を旋回動作させる駆動部とキャスター付の大型三脚から構成される胴体部分にわけられる。また、遠隔参加者の目に相当する遠隔カメラでは、旋回動作やズーム機能によって、上述の頭部旋回では調整できない微小な視線移動や視野の拡大・縮小という行為が可能である。代理人ロボット制御サーバで稼動する制御プログラムは、インターネットを経由して届く遠隔会議室での遠隔操作の情報を取得して、遠隔カメラや頭部旋回制御装置に対応する動作命令を送信する。

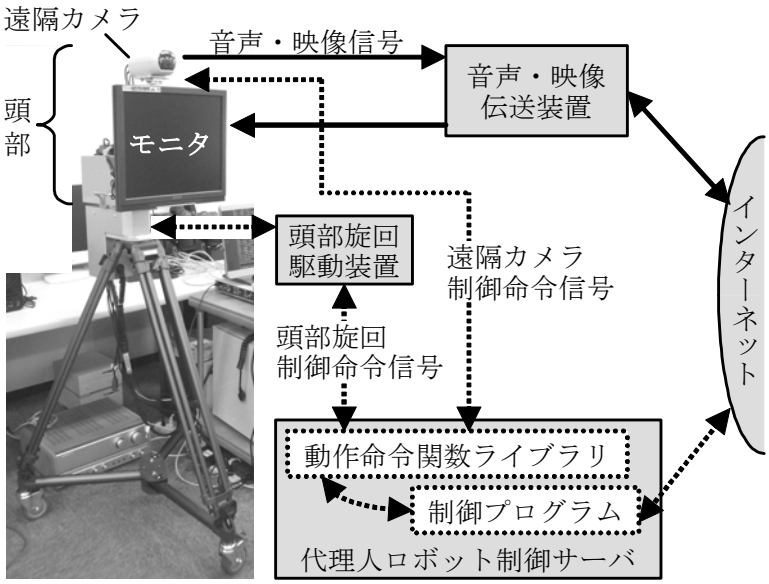


図 4.2　代理人ロボットのシステム構成

4.2.2　特長

試作した代理人ロボットには 3 つの面で大きな特長があり、以下で説明する。

1 点目は、代理人ロボット本体の全長が等身大スケールであり、かつ液晶モニタに表示される遠隔参加者の顔映像を実物大の大きさに調整して表示することが可能というものである。人は普段同じ実空間上にいる実物の人同士で会話をすることに慣れているため、音声と映像を用いた遠隔での会話を行う場合、表示される顔映像が実物に比べて小さいとその存在感が低下してしまい、逆に実物に比べて大きいと威圧感や違和感が増加[32]してしまう。これは、円滑な会話の実現にとって大きな妨害要因になる。このため試作した代理人ロボットは、会議室内にいる他の参加者の代理人ロボットに対する違和感を

減少させ、会議室内での遠隔参加者の存在感を高めるため、等身大スケールの全長および顔寸法にするよう設計した。

2 点目は、遠隔参加者の目と耳の役割をする遠隔カメラと、顔映像表示と口の役割をする液晶モニタを一体とした頭部を、遠隔操作により上下左右旋回等の動作を行うというものである。これにより、遠隔参加者は会話をしたいと思う相手の方向に顔を向け、アイコンタクトをとることができると共に、会議室内の他の参加者からは、遠隔参加者の視線方向や表情を把握できる。

3 点目は、遠隔カメラと液晶モニタの取り付けマウント部分の標準規格対応、頭部旋回動作や遠隔カメラ制御等の動作命令関数のライブラリ化等による汎用性の確保である。まず、代理人ロボット頭部の遠隔カメラ用の取り付けマウントは DIN 規格、液晶モニタの取り付けマウントは VESA 規格に対応しているため、遠隔カメラと液晶モニタの選択面での汎用性は高い。また、代理人ロボットへの動作命令を関数ライブラリ化とすると共に、機種毎に異なる遠隔カメラの制御命令パラメータを記述したプロファイル情報を関数ライブラリが参照することで、動作命令の一元化を実現した。これにより、遠隔カメラの機種に依存しない制御プログラムで動作し、新機種の遠隔カメラに交換する場合も、プロファイル情報にパラメータ情報等を追加するだけで稼動可能となる。

以上 3 点の特長を述べたが、このうち 1 点目と 2 点目に関しては、遠隔会議室にいる参加者と会議室内の他の参加者が、共に同一空間で一緒に参加しているというような空間共有感覚の向上という面で、大きな効果が期待できる。また、複数の代理人ロボットが参加する遠隔会議の場合も、実空間上に代理人ロボットを置く形態のため、対面会議と同じような座席位置構成を実現できる。このため、遠隔参加者が多くなるほど既存の遠隔会議と比較して、代理人ロボットを利用した遠隔会議システムの利便性は高くなると考える。一方、代理人ロボットでの参加者が増大した場合、隣の座席位置にいる代理人ロボットが自分と反対方向の参加者と会話をしている状況等においては、隣の遠隔参加者の顔映像表示画面を完全には視認することができなくなる可能性が出てくる。しかし、そういった状況は対面会議でも起こり得るものであり、対面会議と同様に会話をしているどちらかの参加者の表情や会話音声から状況を把握し、必要であれば自ら発声することで話者交替を実現できると考える。

4.2.3 試験運用

研究開発用ギガビットネットワーク[33]を活用して、石川県内の大学内ゼミ室と東京都内の遠隔講義室の間を接続したネットワーク環境において、筆者自身が遠隔参加者となって、この代理人ロボットを利用した遠隔ゼミを約 1 年間試験運用した。上記期間のうち前半は、後術するマウスによる遠隔操作システムを利用し、後半は後述するジョイスティックによる遠隔操作システムを利用した。ジョイスティックによる遠隔操作イン

タフェースに関しては、当初は飛行機の操縦系をベースにスティックを手前方向に傾けると、代理人ロボットの頭部は上方向に旋回する仕様で運用した。しかし、上下方向に頭部を旋回させる場合は問題なかったが、右上方向といった頭部の斜め旋回を行おうとした際には、スティックを右前方向に傾ける行為が自然であることが判明した。このため、手前方向に傾けると代理人ロボットの頭部は下方向に旋回する仕様に変更する等、遠隔操作システムの改良をこの段階で行っていた。**図 4.3** はこの遠隔ゼミの様子である。液晶モニタに映っている遠隔参加者は、次節で述べる遠隔操作システムを利用して代理人ロボットを遠隔操作しながら、他の参加者と並んだ状態でゼミに参加していた。

この結果、従来の遠隔会議システムの利用と比較して、自然なコミュニケーションをとりやすい遠隔ゼミはある程度達成していた。しかし、大学内ゼミ室空間内に対して特定の場所や方向を指し示すといった遠隔指示行為ができないという課題が、まず浮上した。更に、遠隔参加者によるプレゼンテーションの際には、遠隔参加者が大型プロジェクタに表示したプレゼンテーション画面の前に立って発表する形態を実施したが、その映像を代理人ロボットのモニタ画面から大学内ゼミ室内の大型スクリーンへの切り替え表示手法が非効率的であるという課題も発見した。また、この遠隔ゼミでは、遠隔参加者が同研究分野を専門とする学生であり、音声・映像のリアルタイム伝送として、DVTS に比べて往復伝送遅延時間の少ない ATM-IEEE1394 Link Unit を主に利用していた。このため、今回の試験運用だけでは代理人ロボットの有効性を論じることはできないと考える。そこで、本研究では、一般の大学生を被験者とした評価実験を実施し、遠隔操作システムに関する定量評価を行うことにした。

図 4.3　代理人ロボットを利用した遠隔ゼミの様子

4.3　遠隔操作システム

代理人ロボットを活用した遠隔会議システムには、遠隔参加者が何らかの入力デバイスで遠隔操作するシステムが必要である。また、遠隔会議での自然なコミュニケーションの実現のためには、遠隔操作が簡便でかつ会議への集中力を維持できる遠隔操作インタフェースを有するシステムであることが望ましい。そこで、どのような入力デバイスが遠隔会議での自然なコミュニケーションの実現に有効であるかを議論するため、本研究では入力デバイスとして代表的なマウスとジョイスティックを利用した 2 種類の遠隔操作システムを試作した。

マウスによる操作は、マウスカーソルを視認しながらのマウスの移動による画面内の目標地点への移動やメニュー項目の選択、少数のボタンを利用したメニュー画面の表示やカーソル上のコマンド実行、等の操作画面を視認しながらの操作形態が一般的である。また、マウスによる代理人ロボット旋回動作の座標系については、操作画面を視認しながらの絶対座標系、操作画面を見ずに完全なブラインド操作を行う相対座標系のどちらも実現できる。しかし、遠隔カメラのズーム・パン・チルト等の数多くの遠隔操作も同一のマウスで実現する場合、メニュー画面での項目選択・実行という操作画面を視認しながらの形態を併用する必要がある。

ジョイスティックによる操作は、スティック上の多くのボタン操作とスティックの傾け操作等すべての行為に対して、操作画面を見ずにブラインド操作する形態に適している。ジョイスティックによる代理人ロボット旋回動作の座標系については、スティックの可動範囲がマウスと比較して小さいため、絶対座標系では旋回精度が粗くなってしまい正確な旋回動作が困難である。一方、相対座標系ではスティックのセンター位置へのオートフィードバック機能を利用することで容易に実現できる。

以上の特性から、本研究では、マウスによる絶対座標系での旋回動作を行う遠隔操作システムと、ジョイスティックによる相対座標系での旋回動作を行う遠隔操作システムを試作し、性能評価を行うことにした。

4.3.1　マウスによる遠隔操作システム

今回試作したマウスによる遠隔操作システムの概要を**図 4.4** に示す。マウス操作による遠隔カメラ制御に関しては、市販品サーバが存在するため同システムを使うこととし、頭部旋回制御のみを CGI プログラムで作成した。本システムでは、制御クライアントから制御サーバと遠隔カメラ用市販品サーバにアクセスし、各 CGI プログラムによって制御クライアントの画面上に頭部旋回制御画面と遠隔カメラ制御画面を表示する。頭部旋回制御画面内の長方形エリアは、代理人ロボット頭部の旋回可能なすべてのエリアに対応しており、同エリア内においてマウスクリックをすると十字マーカーで表示されている現在位置からクリック地点まで頭部が旋回動作する。遠隔カメラ制御画面も、遠隔カ

メラの上下左右旋回制御とズーム制御を画面上のパン・チルト・ズーム機能のスライドバーをマウス操作することで行う。なお、両方の制御ともプリセット操作として頭部を正面位置に戻すボタンと遠隔カメラを正面位置に戻すボタンが各制御画面内にある。

マウスは、入力デバイスとしては近年頻繁に使用されるインタフェースの 1 つであるため、マウス操作自体は簡便といえる。しかし、操作時にマウスカーソルを追いかける動作が必要となるため会議映像等が表示されている画面から一瞬目を離す動作が発生する可能性がある。

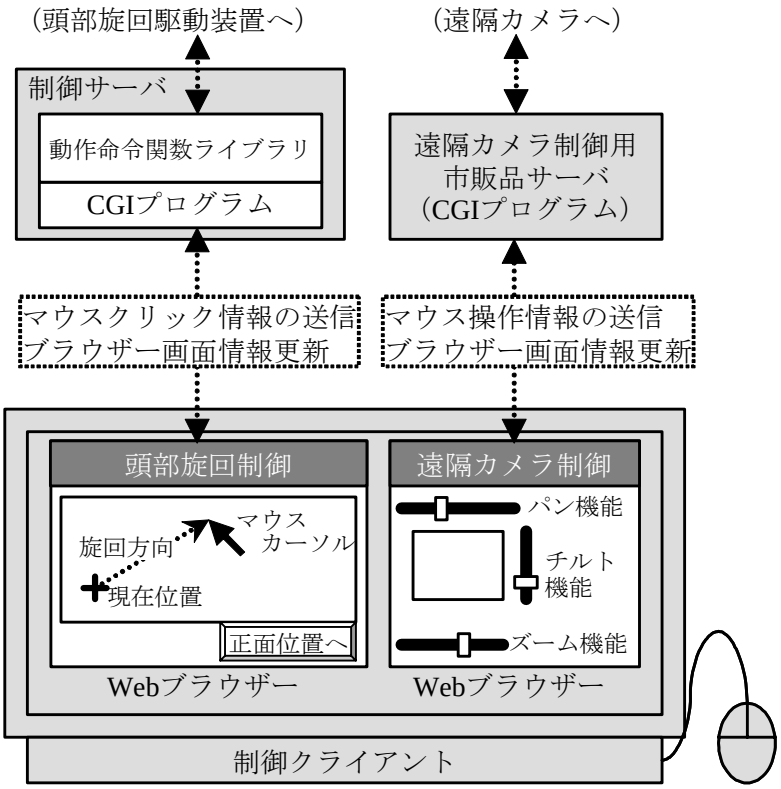


図 4.4　マウスによる遠隔操作システムの概要

4.3.2　ジョイスティックによる遠隔操作システム

今回試作したジョイスティックによる遠隔操作システムの概要を**図 4.5** に示す。**図 4.6** に示すように、頭部旋回制御と遠隔カメラ制御をジョイスティックのみで行うため、1 組のサーバ・クライアント型プログラムで両制御を行う。本システムでは、制御クライアントにおいて、ジョイスティック情報を 100ms 間隔のポーリングで取得し、その情報が頭部旋回制御命令もしくは遠隔カメラ制御命令で、かつ後述する送信タイミングを満たしている場合のみ、その情報を送信する。一方、制御サーバでは受信したジョイスティ

ック情報に対応する動作命令関数を呼び出すことで、代理人ロボットを動作させる。

頭部旋回制御は、**図 4.7** に示すように現在の頭部の位置を基準とする相対移動とし、上・下・左・右の各旋回動作は、ジョイスティックのスティック部分を前方・手前・左・右にそれぞれ傾けることで動作する。これにより、左上方向への旋回動作はジョイスティックを左前方への傾けることで実現できる。なお、スティックを握る手のわずかな動きにも反応し、自分の意志とは関係なく動作するのを防止するため、スティックのセンター位置からある一定傾き角度以内を、無反応領域に設定した。これにより、同領域内でのスティックの傾きではコマンドを送信せず、それ以上の傾きとなる移動感知領域にスティックがある場合にのみ、スティックのセンターからの相対座標（ｘ、ｙ）に対応した頭部の相対旋回動作コマンドを送信する。また、ジョイスティックのフィードバック機能は有効にしてあり、スティックから手を離すと中心位置に戻るようにした。ただし、過度の先行コマンド送信を防ぐため、1 回の頭部旋回動作の平均所要時間から、移動感知領域内にスティックが連続して傾いたままの場合は、約 1 秒間隔で頭部の相対旋回動作コマンドを送信することとした。また、1 回の頭部旋回動作時間がなるべく均等でかつ滑らかな挙動になるように、頭部の旋回移動量、旋回速度、旋回加速度はスティックの傾き角度の大きさに比例するように設定した。

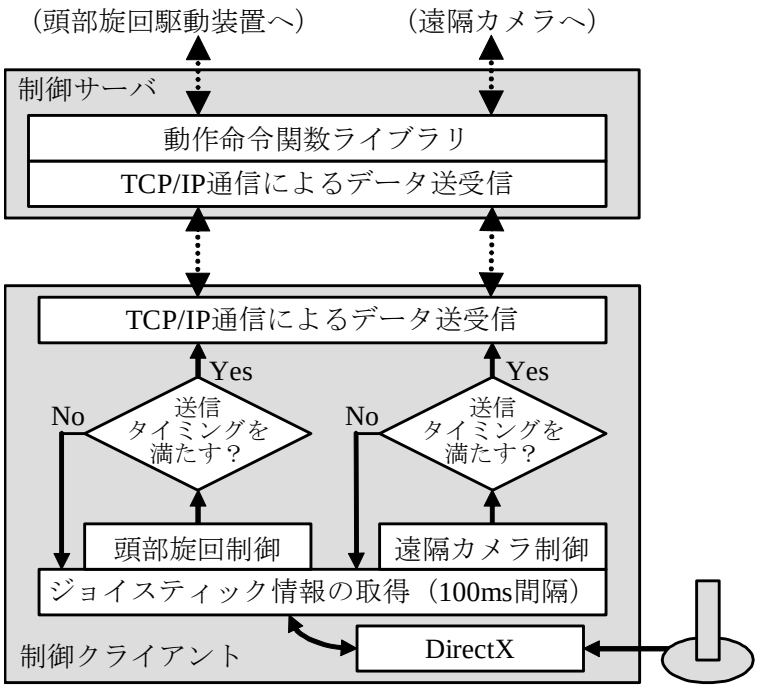


図 4.5　ジョイスティックによる遠隔操作システムの概要

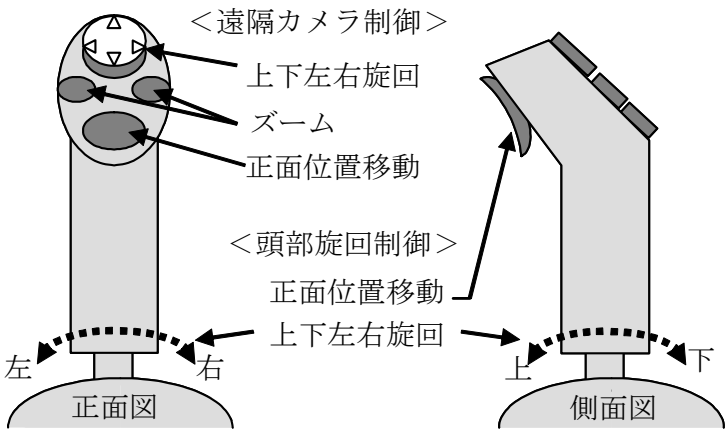


図 4.6　ジョイスティックでの操作命令割当

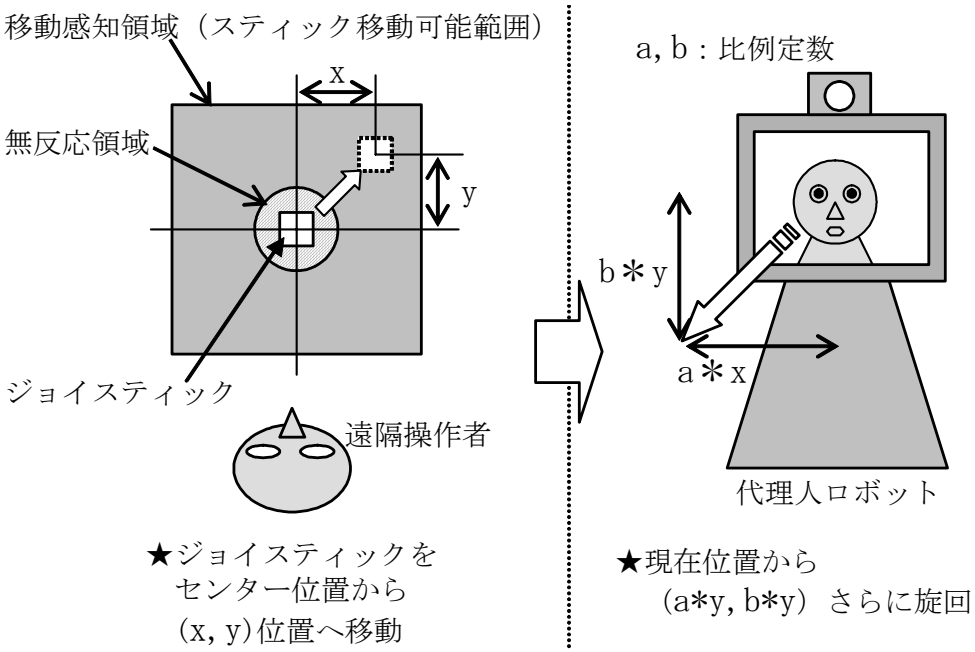


図 4.7　ジョイスティックの移動と頭部旋回の関係

遠隔カメラのズームおよび旋回制御は、代理人ロボット制御サーバの動作命令関数ライブラリに、何種類かの市販の遠隔カメラの旋回制御コマンドおよびパラメータ特性を登録しておくことで、登録済の市販の遠隔カメラであれば同じジョイスティック操作で統一した遠隔操作を実現できるようにした。こちらも過度の先行コマンド送信を防ぐため、1 回の遠隔操作での動作所要時間から同じボタンを連続押下している場合は、約 0.5 秒間隔で同じ遠隔操作コマンドを送信する設定にした。

表 4.2 に、上述したコマンド毎の送信タイミングについてまとめている。

ジョイスティックは、操作時に会議映像等が表示されている画面を見ながら行うことが可能であるが、入力デバイスとしては一般的に利用される機会は少なくジョイスティック操作がスムーズに行われない可能性がある。

表 4.2　コマンド毎の送信タイミング

ポーリング（100ms間隔）時のスティック状態		送信タイミングの判断
ボタン押下なし （頭部の旋回コマンド）	無反応領域内の移動	送信しない
	無反応領域→ 移動感知領域への移動	次回（100ms後）のとき スティック座標値を送信
	移動感知領域内の移動	10回（1秒）に1回の割合で スティック座標値を送信
ボタン「１」押下 （頭部の正面位置 への旋回コマンド）	同じボタン連続押下	送信しない
	上記以外	即時送信
その他のボタン押下 （遠隔カメラの 操作コマンド）	同じボタン連続押下	5回（0.5秒）に1回の割合で 操作コマンドを送信
	上記以外	即時送信

4.4　操作性に関する評価

マウスおよびジョイスティックをそれぞれ用いる2種類の遠隔操作システムに対する、代理人ロボットの遠隔操作に関する基本性能評価実験を実施し、遠隔操作が簡便でかつ会議への集中力を維持できる遠隔操作システムに関する比較評価を行う。その結果から、遠隔参加者が会議への集中度を維持したまま、遠隔操作が可能な遠隔操作システムについて議論する。

4.4.1　被験者

被験者は、JAIST の大学院生 10 名である。被験者は全員、正常視力を有しており、マウスによるコンピュータ利用を日常的に行っているが、ジョイスティックの利用経験はほとんどない。

4.4.2　実験構成

本評価実験のシステム構成を図 4.8 に示す。代理人ロボットは被験者から直視できない場所に設置し、被験者は評価用テレビに表示される遠隔カメラからの映像を見ながら、入力デバイスを操作して代理人ロボットの頭部旋回制御を行う。今回の実験では入力デバイスの操作手法の個人差を小さくするため、遠隔カメラ制御は行わず頭部旋回制御のみの遠隔操作手法で実施した。また、遠隔カメラ映像が評価テレビ画面へ表示されるまでの経路は遅延時間がない経路Ａと音声・映像伝送装置を経由して遅延時間がある経路Ｂを用意し、遠隔操作システムの種類と合わせて計 4 パターンの評価実験を各被験者に対して実施した。本評価実験での経路Ｂの場合の遅延時間は 7 フレーム分（約 231ms）であった。なお、後述する評価解析を行うために、記録カメラ 1、2、3 による被験者の顔、入力デバイスを操作する手、評価用テレビ画面の中心エリアの各映像と、評価用テレビに表示される画面の計 4 つをマルチ合成映像として評価解析用記録テープに録画した。

4.4.3　タスク

今回被験者に与えたタスクは、評価用テレビ画面に表示されている図 4.9 に示すような英数字配置画面を見ながら、試験官の読上げる英数字を画面内の四角枠の中央エリア内に収める行為を、遠隔操作システムによって実行するというものである。1 回のタスクでの試験官が英数字を読上げる回数は 18 回で、英数字の種類と順番は固定で「３→６→５→２→１→５→６→８→５→Ｃ→Ｅ→５→Ｅ→Ｇ→５→Ｂ→Ａ→５」とした。そして、読み上げた英数字が画面中央エリア内に入った状態で、旋回動作が終了したと確定し、次の英数字を読み上げることとした。これにより、上・下・右・左・左下・右上の 6 方向への、長距離／短距離移動の計 12 種類の移動パターンを、均等に計測する。図 4.9 の

英数字は、図 4.8 内のホワイトボードに掲示してあり、指示する各英数字の大きさ（縦 α・ 横α）は評価用テレビ画面表示領域の大きさ（縦 15α・横 20α）と整合するように調整し、中央エリアは評価用テレビ画面上に大きさ（縦 2α・ 横 2α）の枠として直接固定掲示した。なお、実験中は可能な限り評価用テレビ画面に意識を集中し、必要な時以外はテレビ画面から目をそらさないよう指示をした。

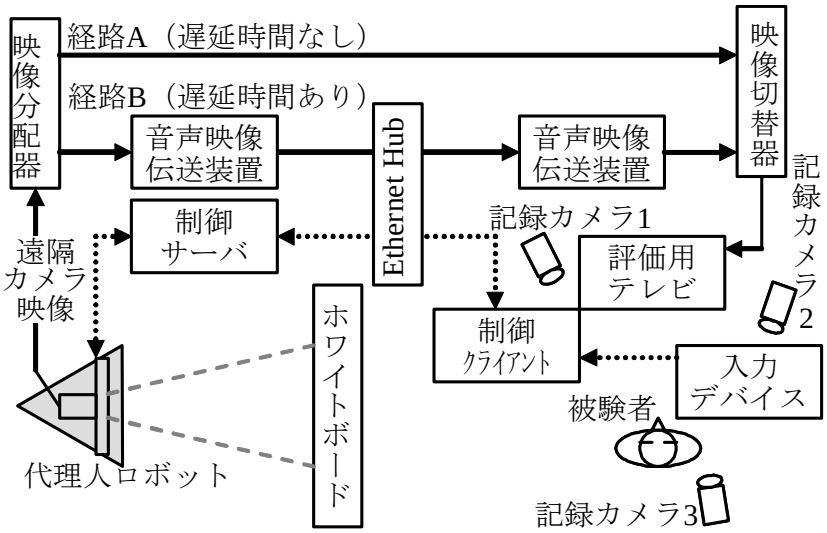


図 4.8　操作性に関する評価実験構成

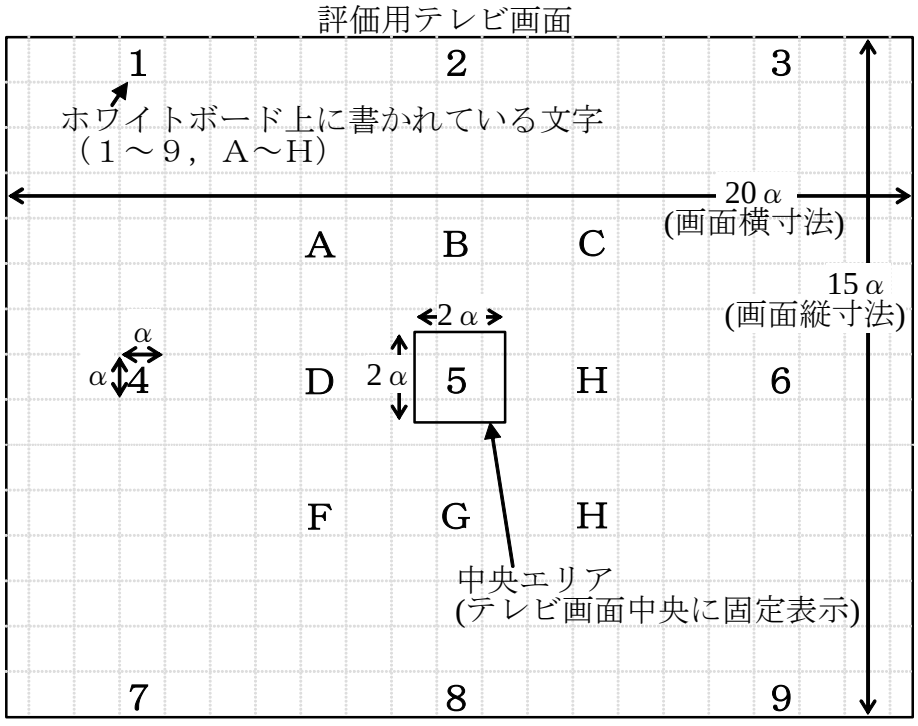


図 4.9　評価実験に用いた英数字配置図

4.4.4　実験手順

被験者 1 名に対して、マウス／ジョイスティックによる遠隔操作システム、遠隔カメラ映像の遅延なし／ありの 4 通りの組合せパターンを実施し、各パターンにおいて上述のタスク 1 回を行った。**図 4.10** は、マウスによる遠隔操作・遠隔カメラ映像の遅延なしのパターンでの評価実験の様子である。なお、最初のマウス操作およびジョイスティック操作を行うタスクの際は、その入力デバイスによる遠隔操作を 1 分間自由で行ってもらい、操作方法を学習してもらった。また、同様のタスクを連続して行うことによる測定順番の差異を少なくするため、奇数人目の被験者には遠隔カメラ映像の遅延なしの場合を先に実施し、偶数人目の被験者はその逆順に実施し、また 5 人目までの被験者にはマウスによる遠隔操作システムでの実験を先に実施し、6 人目以降の被験者にはジョイスティックによる遠隔操作システムでの実験を先に実施した。

図 4.10　操作性に関する評価実験の様子

4.4.5　デバイス操作時間の評価

ここでは入力デバイスの操作時間に関する評価を行う。評価解析用記録テープから、試験官が目標地点となる英数字を読上げてから、被験者の遠隔操作により中央エリア内に該当する英数字が収まるまでの所要時間を、記録されているフレームカウンターから算出した。今回、DVCAM による記録を行ったため、1 フレームは 1/30 秒に相当することになる。

図 4.11 は、全被験者のデバイス操作時間に関する測定結果のうち、「３→６」および「５→２」という代理人ロボット頭部の上下方向旋回の場合を、マウス／ジョイスティックによる遠隔操作システム（図中では「マウス」と「ジョイスティック」と表記)、遠隔カメラ映像の遅延なし／あり（図中では「遅延なし」と「遅延あり」と表記）の 4 通りの組合せパターンに分類して、プロットしたグラフである。このグラフのみでは、デバイス操作時間に関して２種類の遠隔操作システムで差があるかどうかおよび遠隔カメラ映像の遅延の有無で差があるかを判定することはできない。よって、この測定結果に対して、分散分析を実施し差があるかどうかの判定を行う。本評価実験では、同じ被験者が、２つの要因の組合せ環境下それぞれにおいて実験を行った観測値を対象とする、「２つとも対応のあるニ要因配置」形態の実験観測値に相当するため、2 つとも対応のある二要因分散分析[34][35]を実施して、分散分析表を作成する。そして、この分散分析表から、2 つの要因間の相関の有無と、一要因内にあるいくつかの水準のうちどれかに差があるかどうかを判定することができる。なお、本評価実験では、遠隔操作システムという要因内にはマウスとジョイスティックという 2 つの水準、遠隔カメラ映像の遅延という要因内には遅延なしと遅延ありという 2 つの水準、で構成されているため、分散分析表から直接、２種類の遠隔操作システムで差があるかどうかおよび遠隔カメラ映像の遅延の有無で差があるかを判定することが可能である。

表 4.3 に、棄却率 5%（判定結果の信頼度 95%に相当）とした場合の、図 4.11 にプロットされているデバイス操作時間の測定結果に関する分散分析表を示す。表中の「Ｆ境界値」はＦ分布の 5%棄却域境界点の値であり、各変動要因の「分散比」の値がこの「Ｆ境界値」より大きい場合は同要因内の水準間には差があると判定できる。以上より、デバイス操作時間の測定結果に関しては、遠隔カメラ映像の遅延の有無では差がなく、2 種類の遠隔操作システムには棄却域 5%で有意差があり、遠隔カメラ映像の遅延の有無と 2 種類の遠隔操作システムの間には交互作用はないと判定できる。この判定結果と図 4.11 のグラフから、ジョイスティックによる遠隔操作システムの方がデバイス操作に費やしている時間が短いと判定できる。

次に、本評価実験中の 2 種類の遠隔操作システムの操作に関する熟練度の効果について評価を行う。**図 4.12** は、実験前半である 1 回目の操作時と実験後半である 13 回目の

デバイス操作時間の変化を被験者毎にプロットしたグラフを、2 種類の遠隔操作システムに分類して示したものである。操作内容は共に「５→３」への代理人ロボット頭部の斜め上方向旋回である。このグラフのみでは、本評価実験中の 2 種類の遠隔操作システムの操作に関する熟練度による差があるかどうかは判定できないため、やはり分散分析を実施し差があるかどうかの判定を行う。

表 4.4 に、棄却率 5%（判定結果の信頼度 95%に相当）とした場合の、図 4.12 にプロットされているデバイス操作時間の測定結果に関する分散分析表を示す。この結果より、操作経験数の違いつまり本評価実験中の遠隔操作システムの操作熟練度による差はなく、2 種類の遠隔操作システムには棄却域 5%で有意差があり、操作経験数の違いと 2 種類の遠隔操作システムの間には交互作用はないと判定できる。

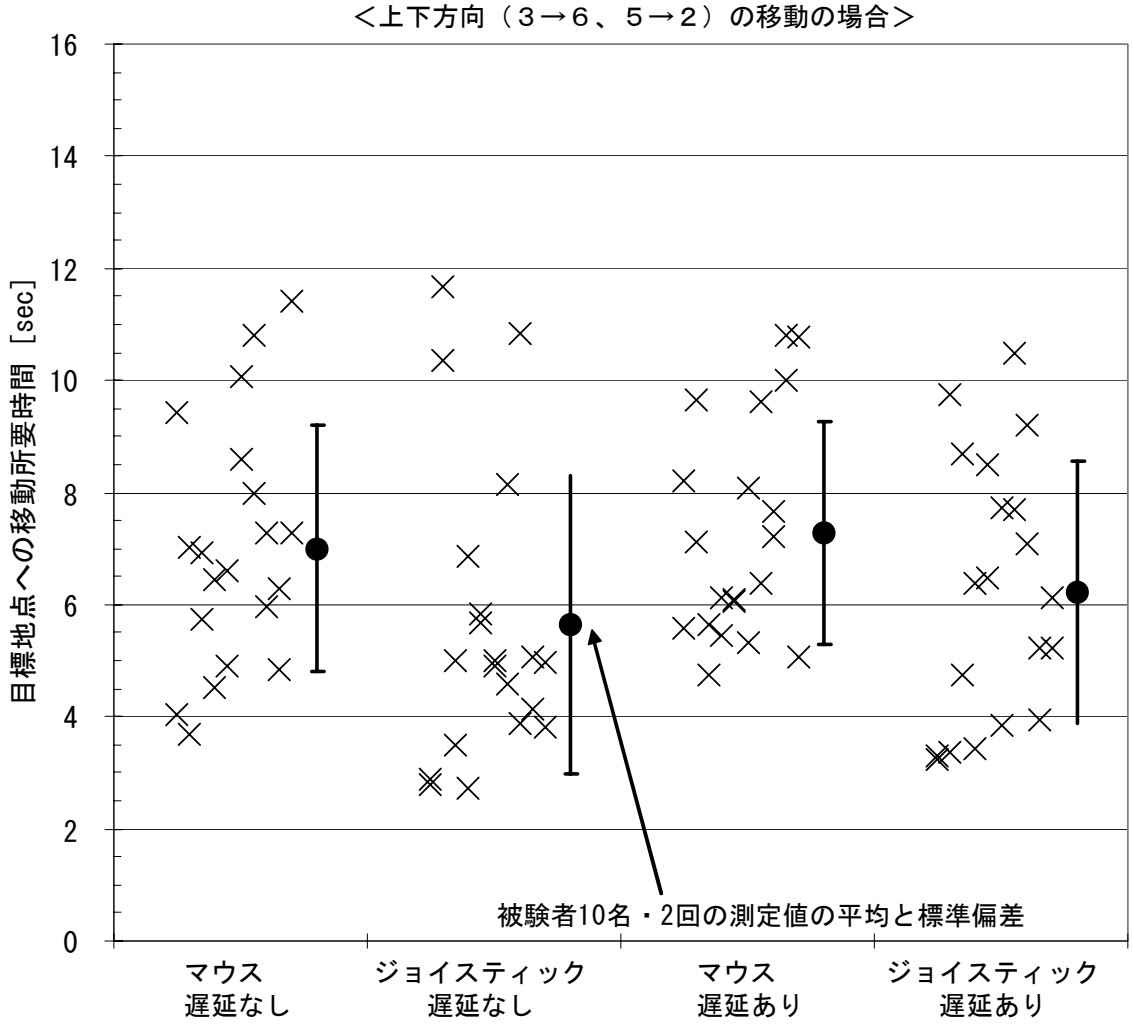


図 4.11　デバイス操作時間に関する測定結果

表 4.3　デバイス操作時間の測定結果に関する分散分析表

変動要因	変動	自由度	分散	分散比	P-値	F 境界値 (棄却域5%)
遠隔操作システム	2.93E+01	1	2.93E+01	5.51E+00	2.16E-02	3.97
映像の遅延	3.80E+00	1	3.80E+00	7.14E-01	4.01E-01	3.97
交互作用	4.46E-01	1	4.46E-01	8.38E-02	7.73E-01	3.97
繰り返し誤差	4.04E+02	76	5.32E+00	−	−	−
合計	4.38E+02	79	−	−	−	−

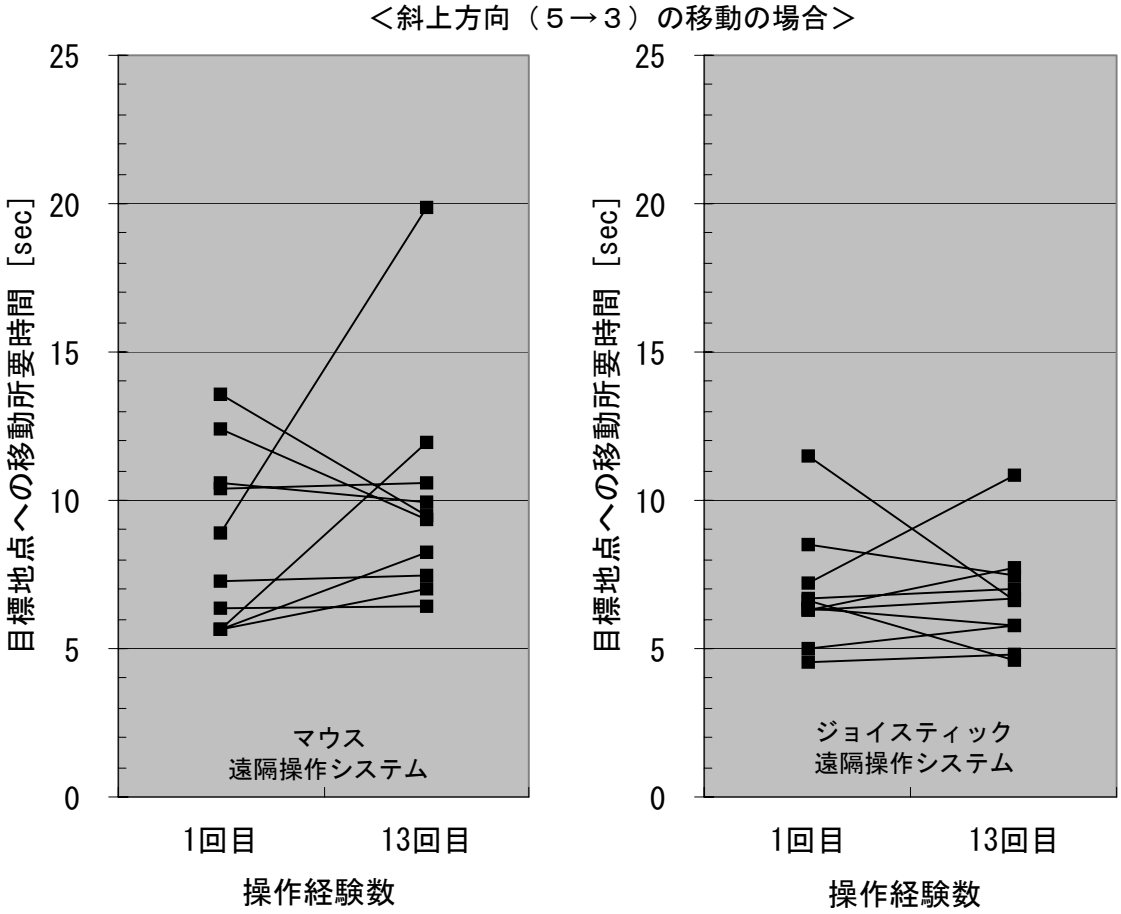


図 4.12　操作経験数によるデバイス操作時間の変化

表 4.4　操作経験数とデバイス操作時間に関する分散分析表

変動要因	変動	自由度	分散	分散比	P-値	F 境界値 (棄却域5%)
遠隔操作システム	6.32E+01	1	6.32E+01	8.25E+00	6.78E-03	4.11
操作経験数	3.75E+00	1	3.75E+00	4.89E-01	4.89E-01	4.11
交互作用	6.08E+00	1	6.08E+00	7.95E-01	3.79E-01	4.11
繰り返し誤差	2.76E+02	36	7.66E+00	−	−	−
合計	3.49E+02	39	−	−	−	−

4.4.6　レスポンス性能の評価

ここでは、被験者が動かしたいと考えてから実際の旋回動作が開始するまでの時間、つまりレスポンス性能に関する評価を行う。評価解析用記録テープから、試験官が目標地点となる英数字を読上げてから、被験者の遠隔操作により実際の旋回動作が開始するまでの所要時間を、記録されているフレームカウンターから算出した。

図 4.13 は、全被験者のレスポンス性能に関する測定結果のうち、「３→６」および「５→２」という代理人ロボット頭部の上下方向旋回の場合を、マウス／ジョイスティックによる遠隔操作システム（図中では「マウス」と「ジョイスティック」と表記)、遠隔カメラ映像の遅延なし／あり（図中では「遅延なし」と「遅延あり」と表記）の 4 通りの組合せパターンに分類して、プロットしたグラフである。前節と同様に、この測定結果に対して、分散分析を実施し差があるかどうかの判定を行う。

表 4.5 に、棄却率 5%（判定結果の信頼度 95%に相当）とした場合の、図 4.13 にプロットされているレスポンス性能の測定結果に関する分散分析表を示す。この結果、レスポンス性能の測定結果に関しては、遠隔カメラ映像の遅延の有無では差がなく、2 種類の遠隔操作システムには棄却域 5%で有意差があり、遠隔カメラ映像の遅延の有無と 2 種類の遠隔操作システムの間には交互作用はないと判定できる。この判定結果と図 4.13 のグラフから、ジョイスティックによる遠隔操作の方が動かしたいと考えてから、旋回動作が開始するまでの時間がかなり短くかつバラツキが小さいことが分かる。

次に、本評価実験中の 2 種類の遠隔操作システムの操作に関する、熟練度の効果について評価を行う。**図 4.14** は、実験前半である 1 回目の操作時と実験後半である 13 回目のレスポンス時間の変化を、被験者毎にプロットしたグラフを、2 種類の遠隔操作システムに分類して示したものである。操作内容は、共に「５→３」への代理人ロボット頭部の斜め上方向旋回である実験前半である。これに対しても分散分析を実施し差があるかどうかの判定を行う。

表 4.6 に、棄却率 5%（判定結果の信頼度 95%に相当）とした場合の、図 4.14 にプロットされているレスポンス性能の測定結果に関する分散分析表を示す。この結果より、操作経験数の違い、つまり本評価実験中の遠隔操作システムの操作熟練度による差はなく、2 種類の遠隔操作システムには棄却域 5%で有意差があり、操作経験数の違いと 2 種類の遠隔操作システムの間には交互作用はないと判定できる。

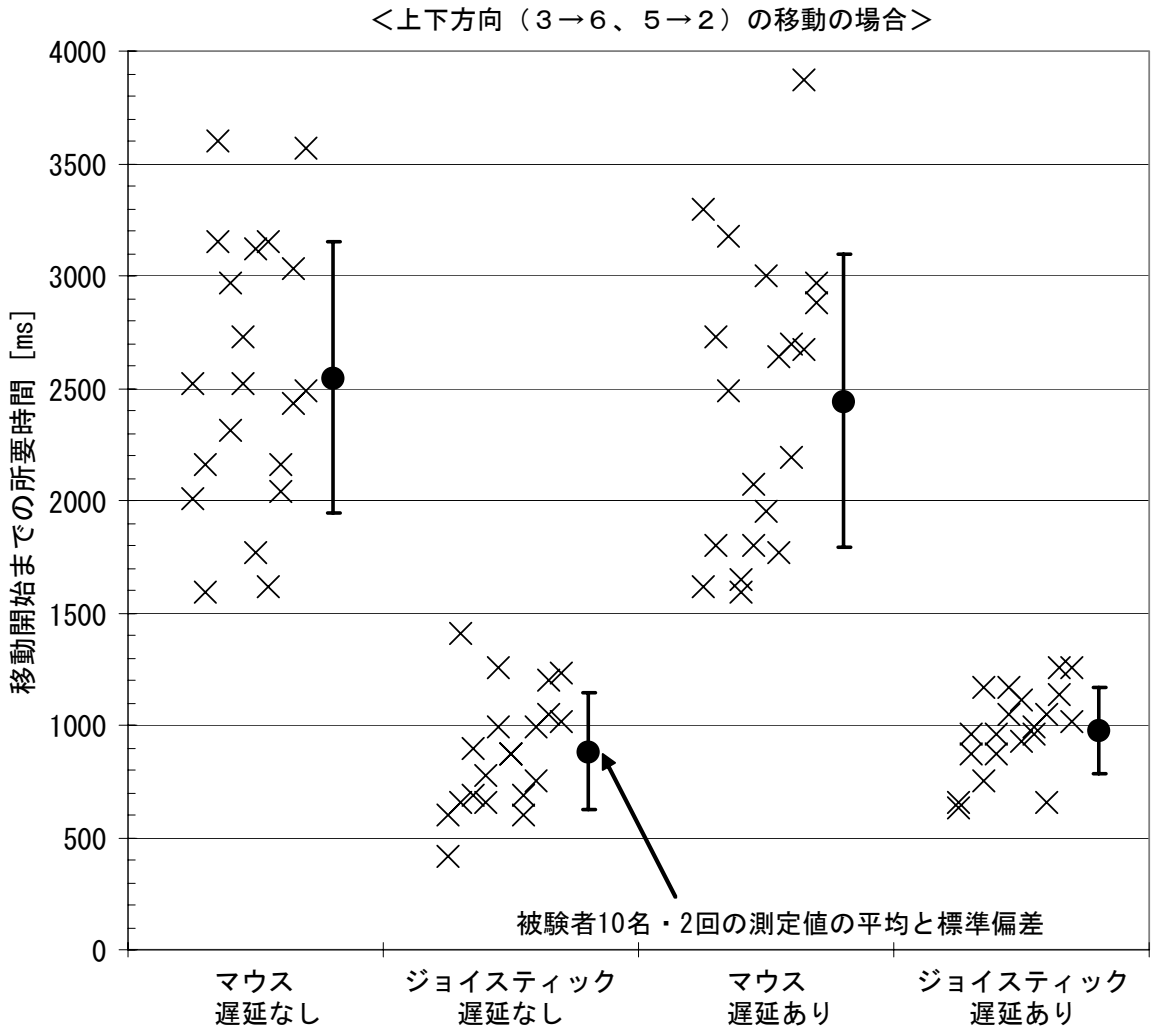


図 4.13　レスポンス性能に関する測定結果

表 4.5　レスポンス性能測定結果に関する分散分析表

変動要因	変動	自由度	分散	分散比	P-値	F 境界値 (棄却域5%)
遠隔操作システム	4.91E+07	1	4.91E+07	2.19E+02	4.18E-24	3.97
映像の遅延	7.20E+02	1	7.20E+02	3.21E-03	9.55E-01	3.97
交互作用	1.90E+05	1	1.90E+05	8.49E-01	3.60E-01	3.97
繰り返し誤差	1.70E+07	76	2.24E+05	−	−	−
合計	6.64E+07	79	−	−	−	−

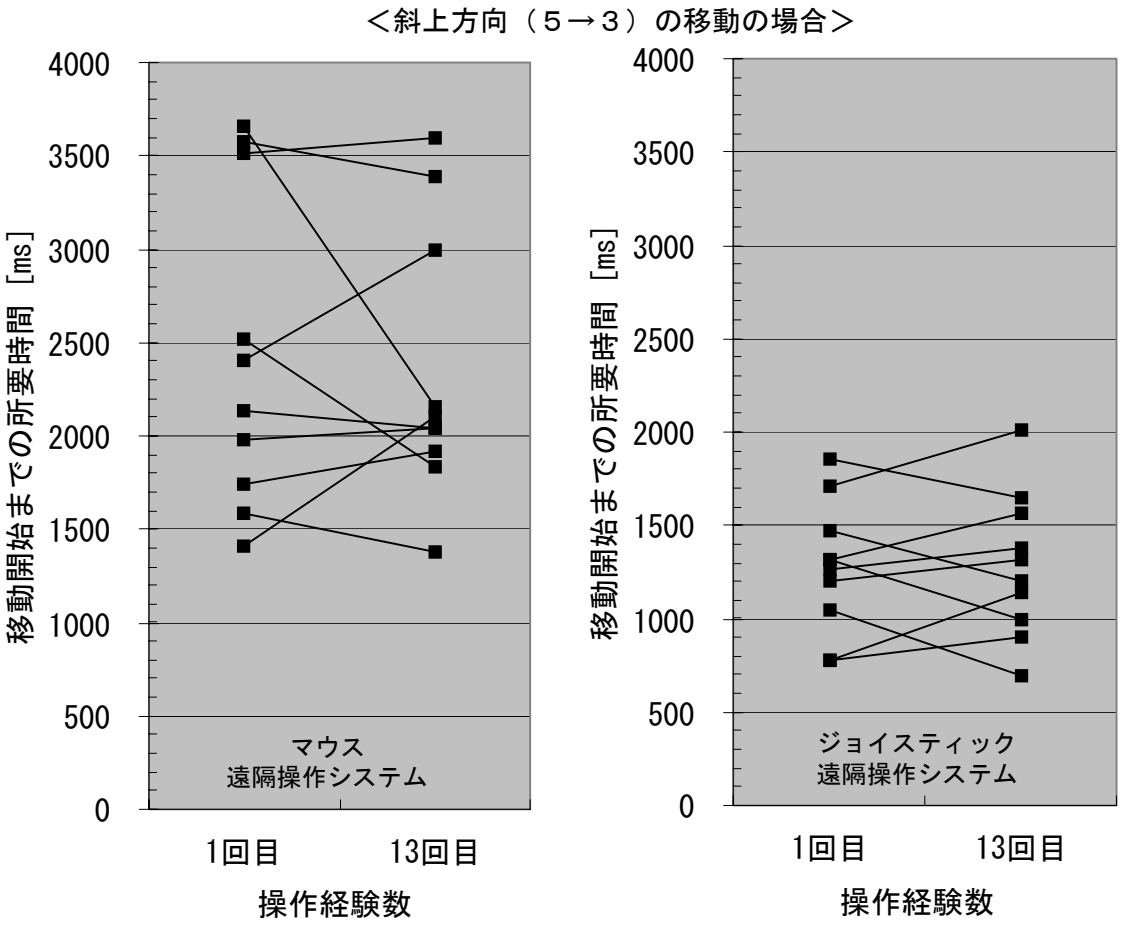


図 4.14　操作経験数によるレスポンス時間の変化

表 4.6　操作経験数とレスポンス時間に関する分散分析表

変動要因	変動	自由度	分散	分散比	P-値	F 境界値 (棄却域5%)
遠隔操作システム	1.25E+07	1	1.25E+07	3.28E+01	1.60E-06	4.11
操作経験数	2.30E+04	1	2.30E+04	6.04E-02	8.07E-01	4.11
交互作用	3.25E+04	1	3.25E+04	8.52E-02	7.72E-01	4.11
繰り返し誤差	1.37E+07	36	3.81E+05	−	−	−
合計	2.63E+07	39	−	−	−	−

4.4.7　テレビ画面への集中度の評価

ここでは、被験者がテレビ画面を見ている時間の割合、つまりテレビ画面への集中度に関する評価を行う。評価解析用記録テープから、試験官が目標地点となる英数字を読上げてから、被験者の遠隔操作のためにテレビ画面から視線をそらしている時間と、画面を見ている時間それぞれの累計時間を、記録されているフレームカウンターから算出した。

図 4.15 は、全被験者のテレビ画面への集中度に関する測定結果のうち、「３→６」および「５→２」という代理人ロボット頭部の上下方向旋回の場合を、マウス／ジョイスティックによる遠隔操作システム（図中では「マウス」と「ジョイスティック」と表記）、遠隔カメラ映像の遅延なし／あり（図中では「遅延なし」と「遅延あり」と表記）の 4 通りの組合せパターンに分類してプロットしたグラフである。前節と同様に、この測定結果に対して、分散分析を実施し差があるかどうかの判定を行う。

表 4.7 に、棄却率 5%（判定結果の信頼度 95%に相当）とした場合の、図 4.15 にプロットされているテレビ画面への集中度の測定結果に関する分散分析表を示す。この結果、テレビ画面への集中度の測定結果に関しては、遠隔カメラ映像の遅延の有無では差がなく、2 種類の遠隔操作システムには棄却域 5%で有意差があり、遠隔カメラ映像の遅延の有無と 2 種類の遠隔操作システムの間には交互作用はないと判定できる。この判定結果と図 4.15 のグラフから、ジョイスティックによる遠隔操作の場合は 100%画面に集中できているのに対し、マウスによる遠隔操作の場合は画面に集中している割合は平均 55%と低くなっていることが分かる。

次に、本評価実験中のテレビ画面への集中度に関する、熟練度の効果について評価を行う。**図 4.16** は、実験前半である 1 回目の操作時と実験後半である 13 回目の画面への集中度の変化を被験者毎にプロットしたグラフを、2 種類の遠隔操作システムに分類して示したものである。操作内容は共に「５→３」への代理人ロボット頭部の斜め上方向旋回である実験前半である。これに対しても分散分析を実施し差があるかどうかの判定を行う。

表 4.8 に、棄却率 5%（判定結果の信頼度 95%に相当）とした場合の、図 4.16 にプロットされているテレビ画面への集中度の測定結果に関する分散分析表を示す。この結果より、操作経験数の違い、つまり本評価実験中の熟練度による差はなく、2 種類の遠隔操作システムには棄却域 5%で有意差があり、操作経験数の違いと 2 種類の遠隔操作システムの間には交互作用はないと判定できる。

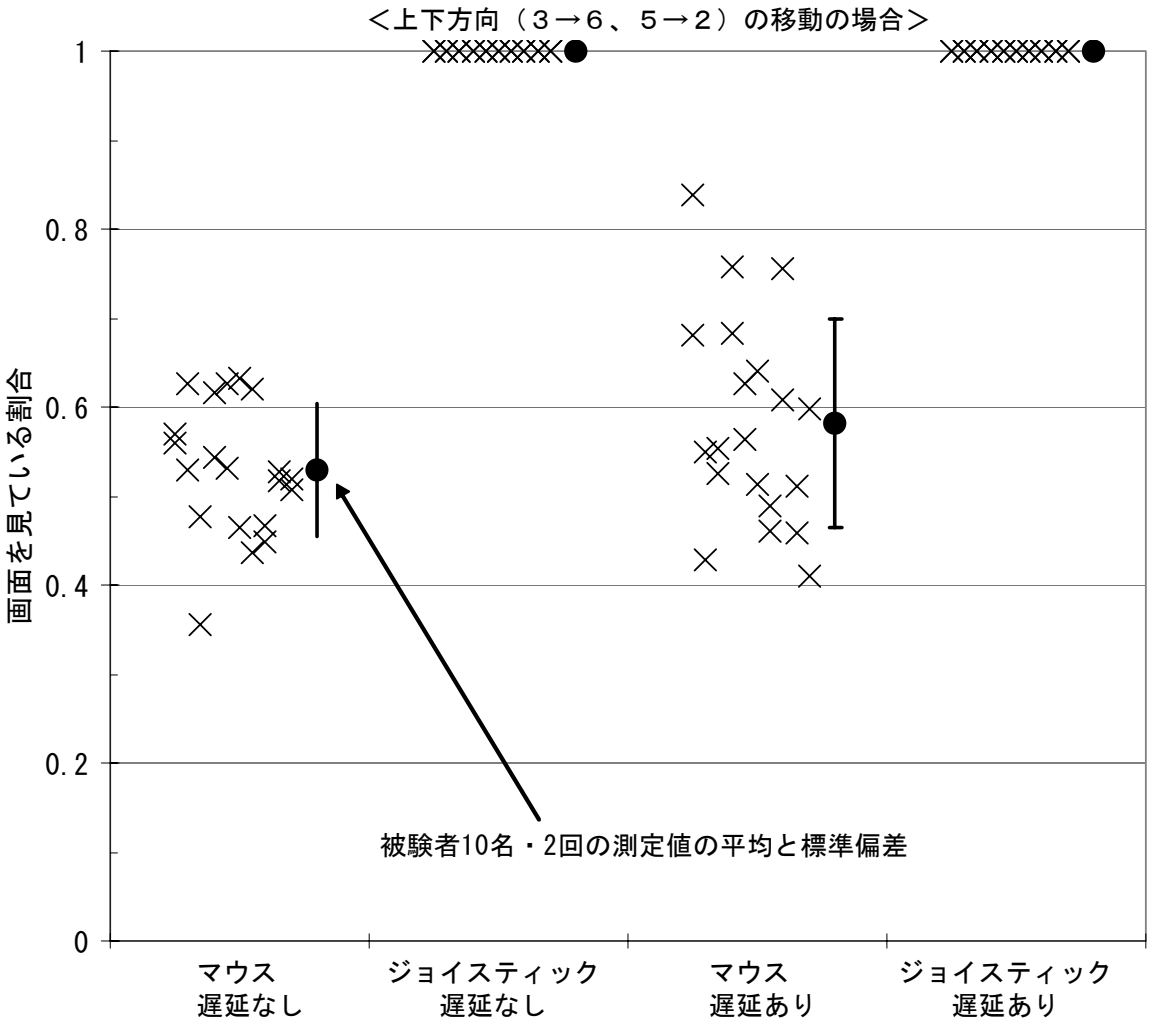


図 4.15　テレビ画面への集中度に関する測定結果

表 4.7　画面への集中度の測定結果に関する分散分析表

変動要因	変動	自由度	分散	分散比	P-値	F 境界値 (棄却域5%)
遠隔操作システム	3.95E+00	1	3.95E+00	8.22E+02	1.69E-42	3.97
映像の遅延	1.44E-02	1	1.44E-02	2.99E+00	8.80E-02	3.97
交互作用	1.44E-02	1	1.44E-02	2.99E+00	8.80E-02	3.97
繰り返し誤差	3.65E-01	76	4.81E-03	−	−	−
合計	4.34E+00	79	−	−	−	−

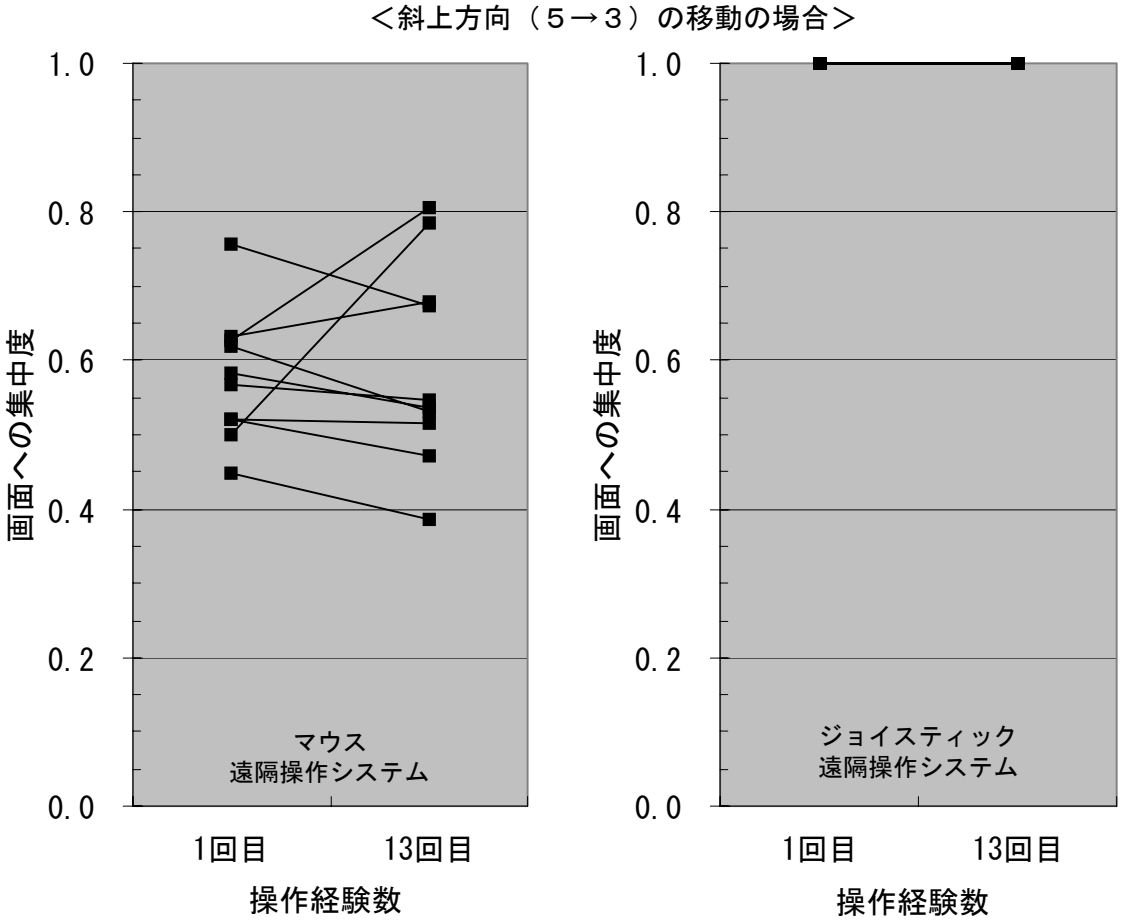


図 4.16　操作経験数による画面への集中度の変化

表 4.8　操作経験数と画面への集中度に関する分散分析表

変動要因	変動	自由度	分散	分散比	P-値	F 境界値 (棄却域5%)
遠隔操作システム	1.72E+00	1	1.72E+00	2.61E+02	4.42E-18	4.11
操作経験数	5.76E-04	1	5.76E-04	8.76E-02	7.69E-01	4.11
交互作用	5.76E-04	1	5.76E-04	8.76E-02	7.69E-01	4.11
繰り返し誤差	2.37E-01	36	6.58E-03	−	−	−
合計	1.96E+00	39	−	−	−	−

4.4.8　アンケート評価

評価実験終了直後、被験者に対してアンケートを実施した。**図 4.17** は、アンケートのうち 2 つの設問に関する評価結果のグラフである。遠隔操作の簡便さについては、ジョイスティックによる遠隔操作システムの方が、マウスによる遠隔操作システムに比べて若干よい評価となり、画面への集中度についてはジョイスティックによる遠隔操作システムの方が、マウスによる遠隔操作システムに比べてかなりよい評価をうけていることが分かる。また、設問 3 として「遠隔会議で実際に使用するとした場合どちらを使ってみたいですか？」というアンケートをとったところ、8 名がジョイスティックによる遠隔操作システム、2 名がマウスによる遠隔操作システムとなり、どちらでもないという回答者はゼロであった。

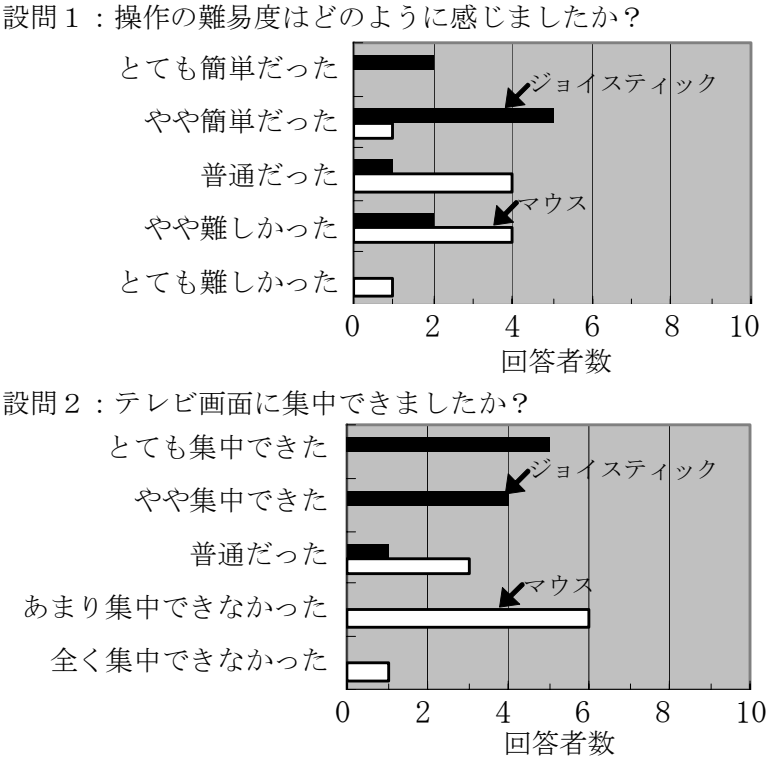


図 4.17　操作性に関するアンケート評価結果

4.4.9　考察

まず、今回試作開発したマウス遠隔操作インタフェースとジョイスティック遠隔操作インタフェースに関して、どちらが遠隔会議での自然なコミュニケーションの実現に有効であるか考察する。遠隔操作の簡便については、4.4.5 および 4.4.8 の評価結果から、ジョイスティックによる遠隔操作システムの方が、マウスの場合に比べて簡便な操作であるといえる。会議への集中力の維持については、4.4.7 および 4.4.8 の評価結果から、ジョイスティックによる遠隔操作システムの方が、マウスの場合に比べて、会議映像等が表示される画面への集中度が高く維持できるといえる。また、4.4.6 の評価結果から、代理人ロボットの旋回動作開始までのレスポンス性能も、ジョイスティックの場合の方が高いことがわかった。以上から、今回使用したマウスによる遠隔操作システムとジョイスティックによる遠隔操作システムを比較した場合は、ジョイスティックによる遠隔操作システムの方が適していると考える。ただし、この結果はジョイスティック操作がマウス操作より適しているというものではなく、会議映像画面に集中した状態のままでもブラインド操作による遠隔操作が可能なシステムが、遠隔会議には適していることを示しているに過ぎない。よって、マウスを利用した遠隔操作システムでも、たとえば「ブラインド操作でのマウスの上下左右移動によって、代理人ロボットの頭部旋回動作が行われるようにし、遠隔カメラの制御ほかの遠隔操作については、マウス右クリックで会議映像が表示されているモニタ画面上にメニューが表示され、希望の動作のメニューを選択する」といったような遠隔操作インタフェースであれば、今回のジョイスティックの遠隔操作インタフェースと同等の効果を得られる可能性は十分にあると考える。

次に、マウスとジョイスティック以外の遠隔操作インタフェースの適用可能性を考察する。上記以外の遠隔操作インタフェースとしては、視線による制御[36]、手指のジェスチャによる制御[37]のほか、視線とマウスの併用による制御[38]等が挙げられる。ただ、現時点のこれらの遠隔操作インタフェースは、視線や手指の動きを検出するための装置を体の一部に装着した状態を続ける必要や誤動作を抑えるために、一定時間の凝視行為を行う等の視線や手指の特定動作を意識し続ける必要がある。このため、そのまま遠隔会議での代理人ロボットの遠隔操作インタフェースとして活用するには、使用者の負担が大きくなる可能性が高い。しかし、たとえば文献[38]のシステムをベースにして、非接触型の視線検出装置を使い、モニタに表示されている会議画面の更に右側を見たいときは、モニタの右側部分に視線を移動させた状態でマウスクリック等の確定行為をすることで、代理人端末頭部が右旋回する、等の直感的な遠隔操作インタフェースを実現できれば上記の課題が解決される可能性は十分ある。よって、今回評価の高かったジョイスティックとどちらが有効であるか今後検討する必要がある。

4.5 アイコンタクトの効果に関する評価

ここでは、代理人ロボットとその遠隔操作システムが、遠隔地間の参加者同士のアイコンタクトおよび話者交替の実現の面で、有効に機能しているかどうかの評価を行う。本評価は、対象となるシステムのユーザビリティの評価[39]の一形態と考えることができる。被験者がタスクを実行する過程を観察し、被験者の行動や発話等からユーザインタフェース上の問題点を発見するユーザビリティテストを参考にして、遠隔会議模擬実験での被験者の行動分析[40]とアンケート評価を実施した。

4.5.1 被験者

被験者は、NICT の事務系職員 20 名である。被験者は全員、正常視力を有している。今回の実験では被験者 4 名ずつの計 5 グループに分けて評価実験を実施した。

4.5.2 実験構成

本評価実験のシステム構成を**図 4.18** に示す。本評価実験での人員配置は、代理人ロボットと 3 名の被験者W、Ｘ、Ｙを会議室内の 1 つの机を囲む形で配置し、被験者Ｚは遠隔会議室から 4.3.2 で提示したジョイスティックによる遠隔操作システムを利用して、代理人ロボットを介して遠隔参加する。被験者Ｚは、必要に応じて会議室内の代理人ロボットの旋回動作、遠隔カメラのズームや旋回動作等をジョイスティック操作により実行可能とする。遠隔会議システムとしては、第 3 章で提案した DVTS による音声・映像リアルタイム伝送機能と、DV リアルタイムミキシング機構を活用した多地点遠隔会議模擬システムを、表 3.1 と同じ機器構成で構築した。ネットワーク想定環境としては、遅延発生器を用いて、第 3 章の評価実験において東京・大阪間のインターネット環境を模擬したときと同じ、片道 7ms に設定した。

なお、後述する評価解析を行うために、4 台のカメラ映像をマルチ合成映像として、評価解析用記録テープに録画した。

4.5.3 タスク

今回被験者に与えたタスクは、会議室内にいる 3 名の被験者と、代理人ロボットを介して別室から遠隔参加する 1 名の被験者の計 4 名に、「日本でのサマータイム制導入の是非について」というテーマについて議論してもらい、会議室内の被験者Wの近くに置いてあるホワイトボード上に要点をまとめてもらう作業とした。議論開始からまとめるまでの時間は 30 分とした。議論を進行させ会議室内のホワイトボード上に要点をまとめさせる作業を担当する司会役は遠隔会議室から参加する被験者Ｚとし、実際にホワイトボード上に書き込む作業を行う書記役は会議室内にいる被験者Wとした。

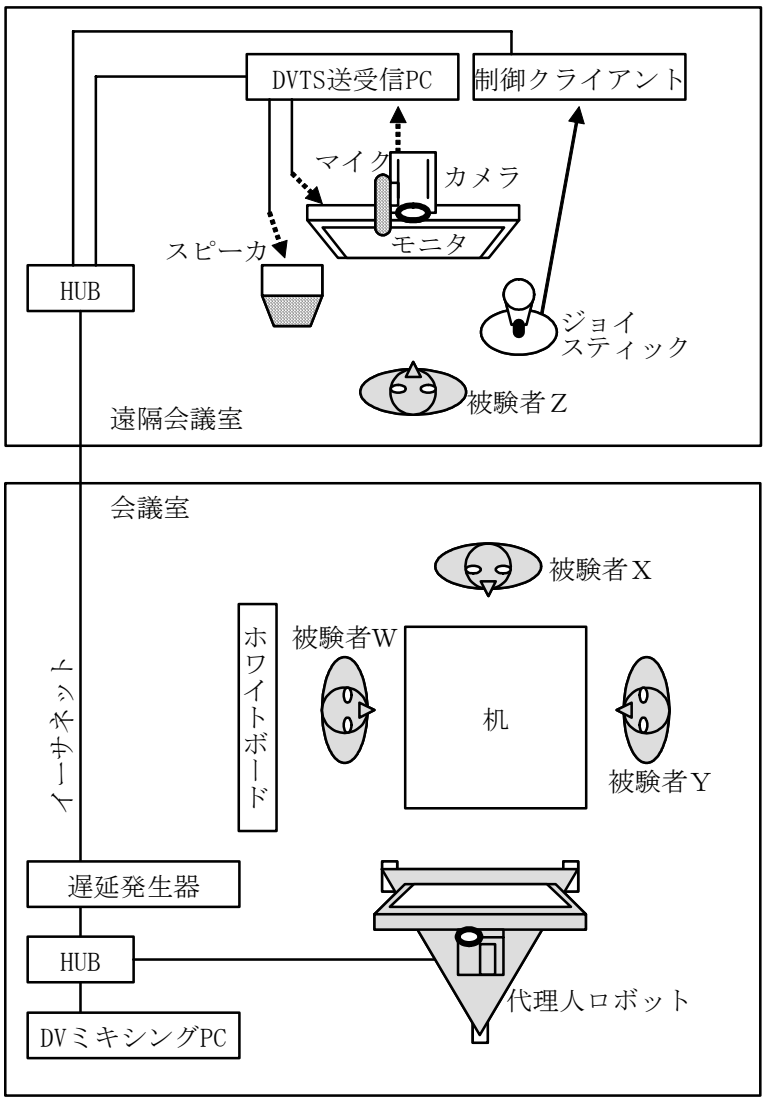


図 4.18　アイコンタクトの効果に関する評価実験構成

4.5.4　実験手順

各グループの被験者 4 名に対して、まず抽選で被験者W，X，Y，Zの役割を決めてもらい、各自所定の位置に着席後、代理人ロボットの遠隔操作方法の説明を行った。そして、これから取り組んでもらうタスクについてテーマ名のみ伏せた形で説明した後、3分間システムを自由に利用してもらった。

その後、会議室内において、テーマ内容を提示し議論を開始してもらい、タスク終了後各被験者に対してアンケートを実施した。

4.5.5　被験者の行動分析による評価

ここでは、被験者がタスクを実行する過程のうち、会話をしている被験者 2 名のうち 1 名が別の被験者と入れ替わる「話者交替」過程に着目した。話者交替が行われるトリガー、およびそのトリガーから実際に話者交替が完了するまでの被験者もしくは代理人ロボットの行動回数を分析することで、定量評価を行った。

話者交替時の行動としては、遠隔会議室から参加している被験者Ｚの遠隔操作による代理人ロボットの旋回動作（以下、「旋回動作」と記述）、これから会話を開始しようとする 2 名の被験者間のアイコンタクト（以下、「アイコンタクト」と記述）、呼びかけやジェスチャ等アイコンタクト以外で話者交替のトリガーとなり得る行動（以下、「発声ほか」と記述）の大きく 3 つの行動に分類した。本評価の目的は、代理人ロボットの行動が、遠隔地間の参加者同士のアイコンタクトおよび話者交替の実現の面で有効に機能していかどうかの評価である。このため、**図 4.19** に示すように話者交替時の話者の位置関係の変化により、以下の 4 種類に分類して行動分析を行った。代理人ロボットの行動が関与しないものとして、会議室内にいる 3 名の被験者間での話者交替（以下、「会議室内→会議室内」と記述）と、代理人ロボットと会議室内のある被験者の会話から会議室内の会話への話者交替（以下、「遠隔地間→会議室内」と記述）がある。一方、代理人ロボット又はその遠隔操作者である被験者Ｚの行動が関与するものとして、会議室内の 2 名の会話から代理人ロボットと会議室内のある被験者との会話への話者交替（以下、「会議室内→遠隔地間」と記述）と、代理人ロボットと会話する会議室内の被験者の交替（以下、「遠隔地間→遠隔地間」と記述）がある。

表 4.9 は、各グループのタスク実行中での話者交替時の行動回数とその統計結果を、話者交替時の行動と話者の位置関係の変化で分類し、5 つのグループ全体でまとめた行動分析結果である。本評価実験では、話者交替時の行動を、トリガーとなる最初の行動から実際に話者交替が完了するまでの行動内容として、同表にある 15 パターンの行動に分類した。そして、評価実験録画記録を再生・解析することで、それぞれの行動回数をカウントした。なお、同表で数値が入っていない部分は代理人ロボットの行動が関与しない場合は無効となる行動パターンであることを示している。この統計結果から、まず話者交替時の話者の位置関係の変化の違いにより、話者交替数が大きく異なっていることがわかる。これは司会を務める遠隔会議室にいる被験者Ｚと会議室内にいる被験者とのやりとりが多くなっているためと考えられる。このため、本評価実験での評価は行動回数の絶対値ではなく、比率により行うこととする。話者交替時のアイコンタクト実施率については、会議室内の被験者間での話者交替時で 95%、代理人ロボットが関与する遠隔地間での話者交替時で 97～98%という極めて高い実施率を示している。代理人ロボットの有無に関わらず、話者交替におけるアイコンタクトの重要性が非常に高いことが確

認できる。また、遠隔会議での代理人ロボット利用と話者交替の実現との関係には、話者交替時の代理人ロボット旋回動作実施率が約 90%を示していることから、代理人ロボットの利用は、話者交替の実現の面で有効に機能していると考える。更に、遠隔会議での代理人ロボット利用とアイコンタクト実現との関係についても、旋回動作後のアイコンタクト実施率が 80%近い値を示していることから、代理人ロボットの利用はアイコンタクト実現の面でも有効に機能していると考える。

図 4.20 は、各グループのタスク実行中での話者交替において、トリガーとなった最初の行動の比率を示している。同図より、代理人ロボットの行動が関与している「会議室内→遠隔地間」および「遠隔地間→遠隔地間」での話者交替では、代理人ロボットの旋回動作がトリガーとなっているケースは平均で 40%を超える高い比率であることが分かる。これは、代理人ロボットを介して遠隔参加している被験者Ｚが、話者交替のために積極的に代理人ロボットを遠隔操作していることを示している。

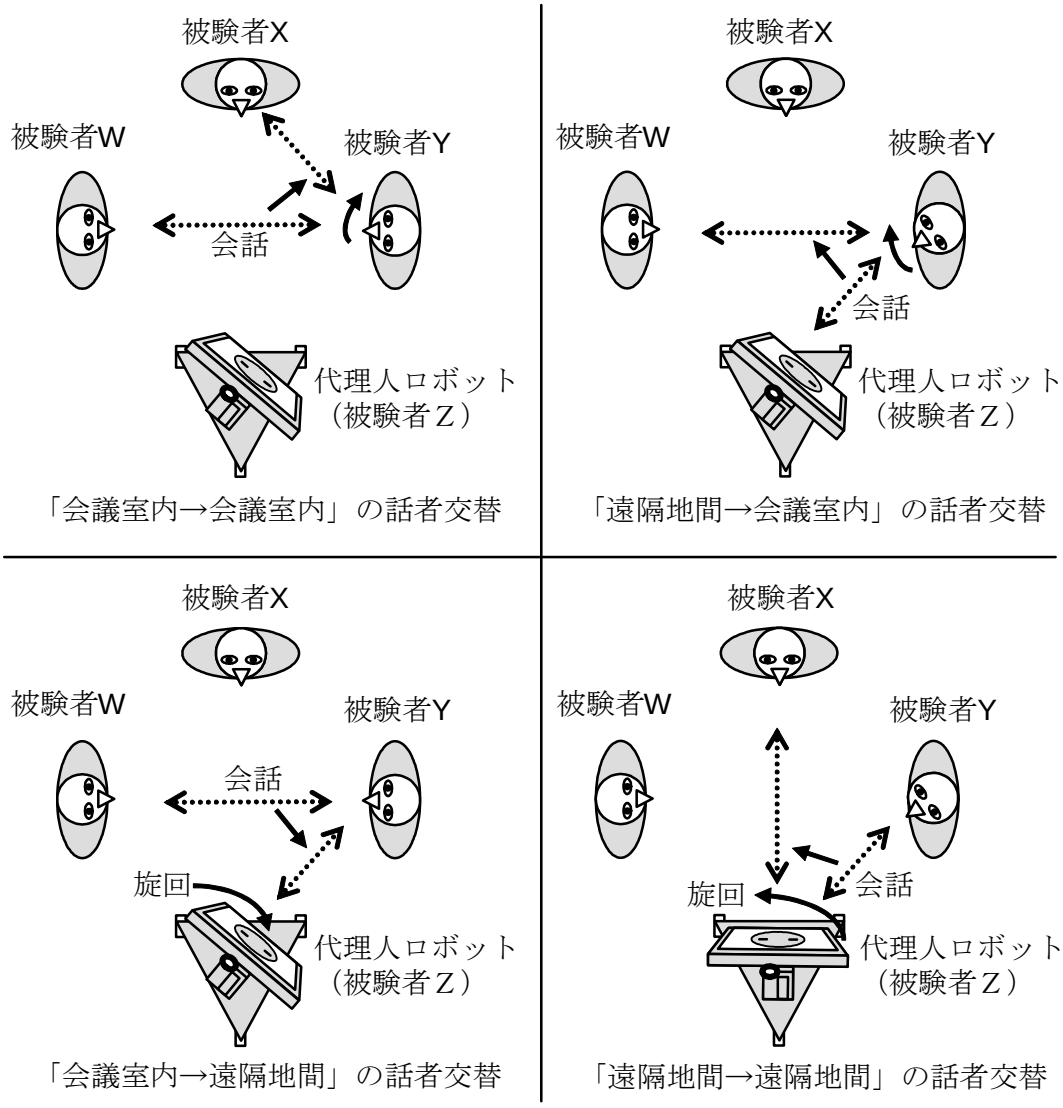


図 4.19　話者交替時の話者の位置関係の変化

表 4.9　被験者の話者交替時の行動回数（グループ全体）

	（5グループの総計）	話者交替時の話者の位置関係の変化			
		会議室内 ↓ 会議室内	遠隔地間 ↓ 会議室内	会議室内 ↓ 遠隔地間	遠隔地間 ↓ 遠隔地間
話者交替時の行動パターン	旋回動作 → アイコンタクト → 発声ほか	−	−	10	13
	旋回動作 → 発声ほか → アイコンタクト	−	−	44	100
	旋回動作 → アイコンタクト	−	−	0	11
	旋回動作 → 発声ほか	−	−	0	4
	旋回動作	−	−	0	0
	アイコンタクト → 旋回動作 発声ほか	−	−	10	1
	アイコンタクト → 発声ほか 旋回動作	−	−	3	1
	アイコンタクト → 旋回動作	−	−	0	2
	アイコンタクト → 発声ほか	51	13	4	12
	アイコンタクト	0	0	0	0
	発声ほか → 旋回動作 アイコンタクト	−	−	36	86
	発声ほか → アイコンタクト 旋回動作	−	−	4	24
	発声ほか → 旋回動作	−	−	4	0
	発声ほか → アイコンタクト	88	41	10	7
	発声ほか	8	3	0	0
統計値	話者交替数：T	147	57	125	261
	話者交替時のアイコンタクト実施数：I	139	54	121	257
	話者交替時の旋回動作実施数：R	−	−	111	242
	旋回動作→アイコンタクト実施数：α	−	−	90	210
	話者交替時のアイコンタクト実施率：I／T	0.95	0.95	0.97	0.98
	話者交替時の旋回動作実施率：R／T	−	−	0.89	0.93
	旋回動作後のアイコンタクト実施率：α／I	−	−	0.74	0.82

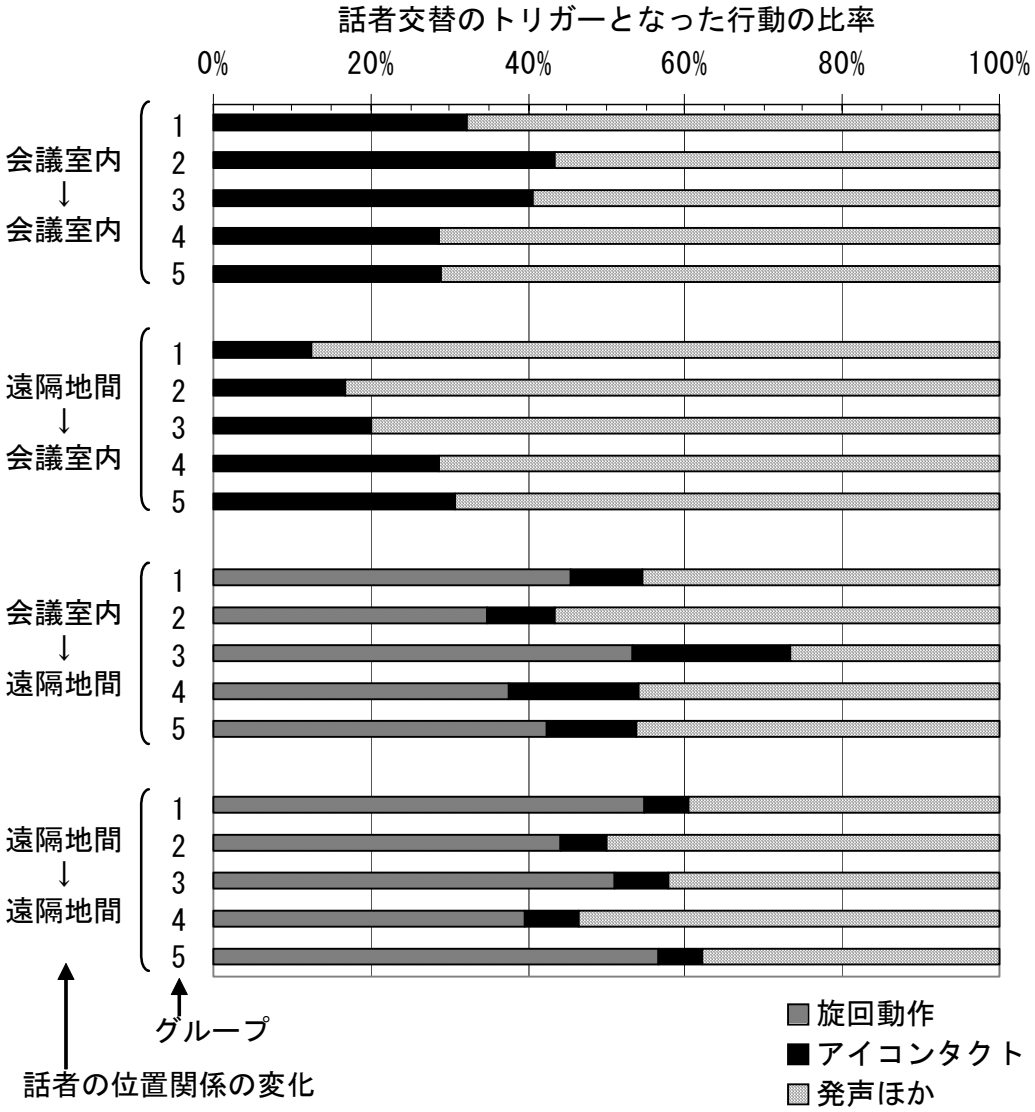


図 4.20　話者交替のトリガーとなった行動の比率

4.5.6　アンケート評価

図 4.21 は、各被験者に対して実施したアンケートの評価結果を、遠隔会議室から参加する被験者Ｚと、会議室で参加した 3 名の被験者に分類してグラフにしたものである。遠隔会議室から参加する被験者Ｚに関しては、3 つの設問すべてに対して高い評価結果を得た。これは、代理人ロボットを介して遠隔参加する側には、代理人ロボットの利用による遠隔地との会話は容易になり、会話のための負担も少ないことを示している。一方、会議室で参加した 3 名の被験者に関しては、設問 1 と 3 では高い評価結果を示しているものの、設問 2 は少し低い評価結果を示している。本アンケートでは設問毎に自由にコメントを書き込む欄を設けており、この設問 2 に関しては「振向く動作が遅い」、「そこ

に人がいるという存在感がまだ小さい」というコメントがあり、これらは設問 2 において「やや負担を感じる」と回答した被験者のものであった。以上から、代理人ロボットと同席し会議を行う参加者側には、代理人ロボットの利用による遠隔地との会話は容易になりシステム全体としての有効性も確認できるが、代理人ロボットを介しての会話に対する負担が残っていることが分かる。この負担軽減のついては今後の課題である。

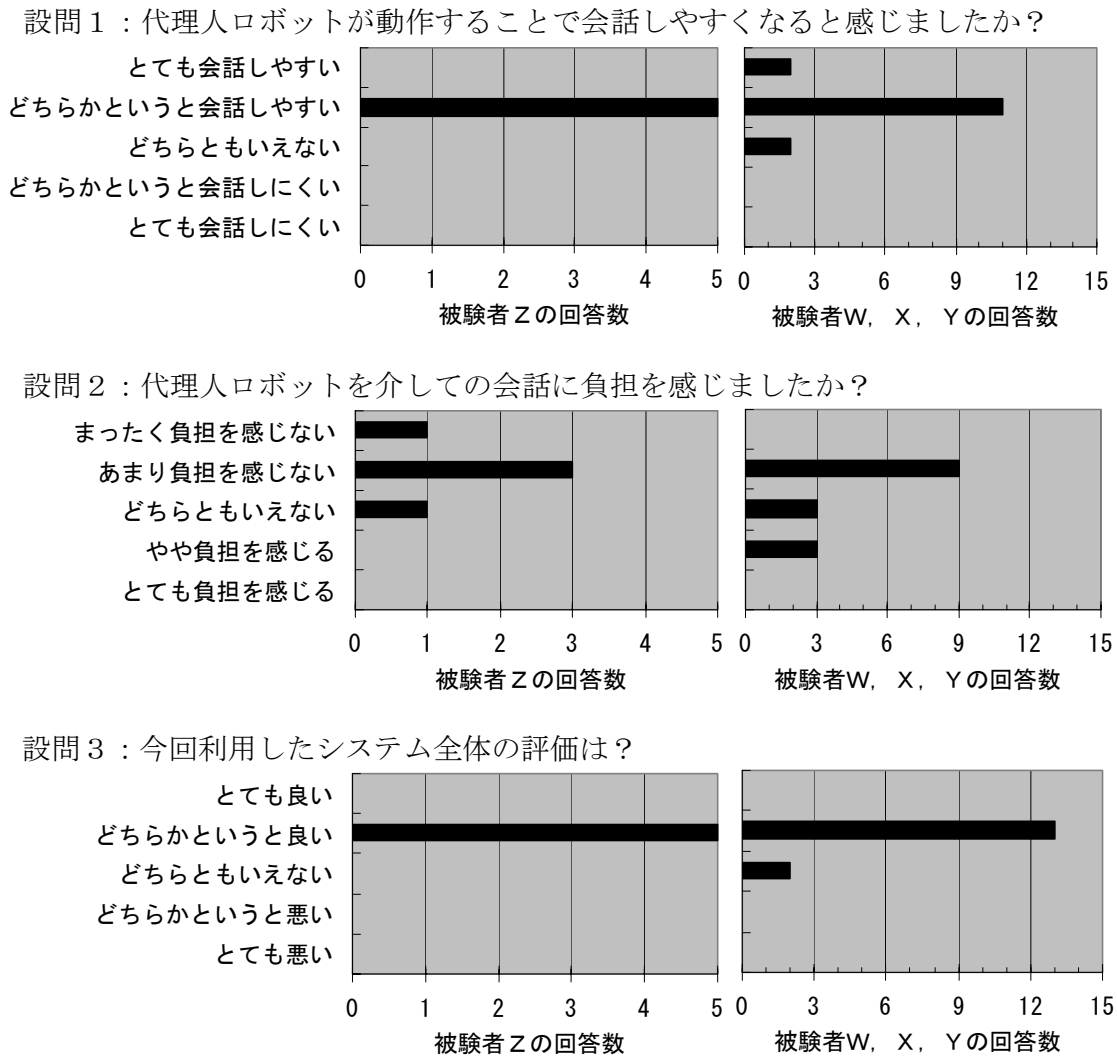


図 4.21　アンケート評価結果

4.6 まとめ

本章では、リアルタイム遠隔会議システムにおいて自然なコミュニケーションを実現するための手法として、複数の物理的に離れている会場空間であっても、同一会場空間の場合と同じような空間共有を体感できるシステムについて議論した。そして、他の参加者と実際に並んで会議に参加することを目的として、ほぼ実物大の遠隔参加者の実写顔映像が表示されるモニタや、目となる遠隔カメラ等が一体となっている頭部が、遠隔操作により実際に上下左右に振り向く動作が可能な代理人ロボットを開発した。

次にマウスとジョイスティックという代表的な入力デバイスによる 2 種類の異なる遠隔操作システムを試作し、被験者による操作性に関する評価実験を行い、どのような遠隔操作環境が使用者に負担が少なく会議に集中できるかという、自然なコミュニケーション実現に重要な要件について考察した。評価結果としては、ジョイスティックの傾き量に応じて、現在位置から自動的に傾き方向へ相対旋回動作を行う遠隔操作試作システムのほうが評価は高いことがわかった。そして、代理人ロボットを利用した遠隔会議では、ブラインド操作が簡便で会議映像画面に集中した状態のまま遠隔操作が可能な手法が、自然なコミュニケーション実現に重要な要件であるという考察結果を得た。

また、代理人ロボットとその遠隔操作システムが、遠隔地間の参加者同士のアイコンタクトおよび話者交替の実現の面で有効に機能しているかどうかの評価を、遠隔会議模擬実験での被験者の行動分析とアンケート評価から試みた。行動分析評価の結果、代理人ロボットの有無に関わらず、話者交替におけるアイコンタクトの重要性が非常に高いことが確認されると共に、遠隔会議での代理人ロボットの旋回動作が、アイコンタクトおよび話者交替と強い相関性を持っていることが分かった。このため、代理人ロボットの旋回動作が、遠隔地間の参加者同士のアイコンタクトおよび話者交替の実現の面で有効に機能しているという考察結果を得た。また、アンケート評価の結果、代理人ロボットを介して遠隔参加する側には、代理人ロボット利用による遠隔地との会話は容易になり会話のための負担も少ないことが分かった。一方、代理人ロボットと同席し会議を行う参加者側には、代理人ロボットを介した会話に対する負担の低減化という検討課題があることも明らかになった。

第5章　総合評価

5.1　はじめに

本研究の目的は、3 地点以上の地理的に離れた会場間が、インターネットで接続されている環境での遠隔会議において、遠隔地間との円滑な会話やアイコンタクトによるスムーズな話者交替の実現等の自然なコミュニケーションを実現し、活発な議論が可能となる遠隔会議システムの提案および有効性の評価である。これを実現させるための機能として、DV リアルタイム処理機構、および代理人ロボットとその遠隔操作システムの提案を行い、それぞれについて自然なコミュニケーションに実現という面で有効に機能することを示した。しかし、上述の 2 つの機能を適用した遠隔会議システムが、実際の遠隔会議という環境でどの程度有効に機能するかは検証していない。

本章では、DV リアルタイム処理機構、および代理人ロボットとその遠隔操作システムを適用した遠隔会議模擬環境を構築し、通常の対面会議環境および従来の遠隔会議システムを利用した遠隔会議模擬環境との総合比較評価を行う。このような総合評価については、各システムを利用した被験者に対してアンケートを行う主観的評価や、遠隔参加者との会話回数測定等の定量評価[17][28]が一般的である。しかし、本研究では、評価方法の１つの柱として、プロジェクトの進捗や作業のパフォーマンス等の定量評価に用いられる EVM (Earned Value Management[41]) という統合マネジメントの技法を参考に、各システムにおける会議の進捗度を推定する新しい手法を提案し、評価実験での評価を試みる。そして、アンケートによる主観的評価や会話回数測定等の定量評価というこれまで一般的に用いられている評価方法も同時に実施し、両者の結果を比較することで今回提案した新しい手法の有効性についても考察する。

5.2　EVM の概念

EVM は、プロジェクトの人件費や作業時間等の累積作業コストを定量化することで、計画段階からの作業コストの超過状況やスケジュールの遅れ状況を把握すると共に、プロジェクトの最終的な作業コストと終了時期を予測することを目的とした、統合マネジメント技法である。EVM は、米国防省が欧州の国防関連政府機関と協力して開発し、これまで公的機関の IT 調達マネジメント等の大規模で複雑なプロジェクトマネジメントに適用されてきている。しかし、EVM の概念自体は難しいものではなく、様々な分野での容易に利用可能な技法である。たとえば、建設関連のプロジェクトのほとんどには、コスト・エンジニアリングと呼ばれる作業がある。担当の管理者は、業者への支払の前に請求書に対応する作業内容の完了確認を行うことで、実際の達成作業に対する実コストという重要な関係に注目しており、これは簡単な形式の EVM の利用である。

EVM の利用の流れは、プロジェクトのスケジュールと各作業のコストの見積り、プロジェクト開始後の定期的な作業コスト等の測定、測定結果からのプロジェクトの進捗状況の把握、最終的なスケジュールとコストの予測、に大きく分類でき、**図 5.1** を参照しながらそれぞれの概要を説明する。

プロジェクトの各作業のコストやスケジュールの見積りには、一般に WBS（Work Break down Structure）と呼ばれる手法を利用する。WBS では、プロジェクトの各作業を徐々に小さい単位に分割し、作業を完了するために必要な作業人数・作業時間・経費等の見積りが算出可能なレベルのタスクにまで分解する。プロジェクト管理者はこれらの見積りを積み上げることで、プロジェクトのスケジュールとそれに沿った累積予定コストにまとめることが可能になる。図 5.1 では、累積予定コスト（Planed Value：以下 PV と記述）のグラフに相当し、完了予定時期のコスト値はプロジェクトの総コストつまり総予算になる。

プロジェクト開始後の作業コスト等の測定は、四半期等の一定期間毎やプロジェクトの節目等の段階において実施され、測定される値は達成価値（Earned Value：以下 EV と記述）とコスト実績値（Actual Cost：以下 AC と記述）である。EV は、その時点で実際に完了している作業内容を PV と同じ基準でコストに変換した値であり、プロジェクト終了時には PV と同じ値になる。AC は、その時点までに実際に費やしたコストの実績値である。図 5.1 では、測定時期の EV 値および AC 値の 2 点に相当する。

プロジェクトの進捗状況は、図 5.1 のグラフより、まず計画スケジュールに対する差異が EV-PV として算出され、この値が負つまり EV が PV よりも下にある場合は計画よりスケジュールが遅れていることが把握できる。また、計画コストに対する差異が EV-AC として算出され、この値が負つまり EV が AC よりも下にある場合は計画よりコストが超過していることが把握できる。なお、これらの傾向は継続して測定・把握することが重要であるため、実際に達成した作業と当初の計画していた予定作業との関係を示すスケジュール効率指数（Schedule Performance Index：以下 SPI と記述)、および実際に達成した作業とその実施に費やしたコストの関係を示すコスト効率指数（Cost Performance Index：以下 CPI と記述）をそれぞれ、

$$SPI = EV / PV \qquad \cdots\cdots\cdots(式２)$$

$$CPI = EV / AC \qquad \cdots\cdots\cdots(式３)$$

で定義し、**図 5.2** に示すような推移グラフの形式で把握する。この場合、SPI の値が 1.0 より小さくなるほどスケジュールの遅れが大きく、また CPI の値が 1.0 より小さくなるほどコスト超過が大きいことを意味する。

最後に、最終的なスケジュールとコストの予測について説明する。一般にプロジェク

トの進行の遅れは、一般に作業人員の増加等のコスト追加により調整が可能であるため、測定時以後のこれ以上のスケジュールの遅れはないと仮定でき、最小で当初の完了予定時期になるという予測を行う。一方、最大で「累積予定コスト PV のグラフ上で測定時の EV と同じ値になる時期と測定時期の期間差」で表現される測定時のスケジュールの遅れの分だけ、プロジェクトの完了時期が遅れるという予測を行う。最終的なコストの統計的予測については、測定時でのコスト効率指数 CPI を改善することは統計的に困難でるという傾向にあり、測定時以後も同じ値が継続すると想定する。よって、最終的なコストの最善値となる下限予測値 C_B は総予算を TB として、

$$C_B = AC + (TB - EV) / CPI \qquad \cdots\cdots\cdots(式４)$$

となり、スケジュールの遅れを取り戻すためにコストを追加する場合での最終的なコストの最悪値となる上限予測値 C_T は、

$$C_T = AC + (TB - EV) / (CPI \times SPI) \qquad \cdots\cdots\cdots(式５)$$

と予測する。図 5.1 では、最終的なコストおよびスケジュールの予測範囲を斜線エリアで示している。

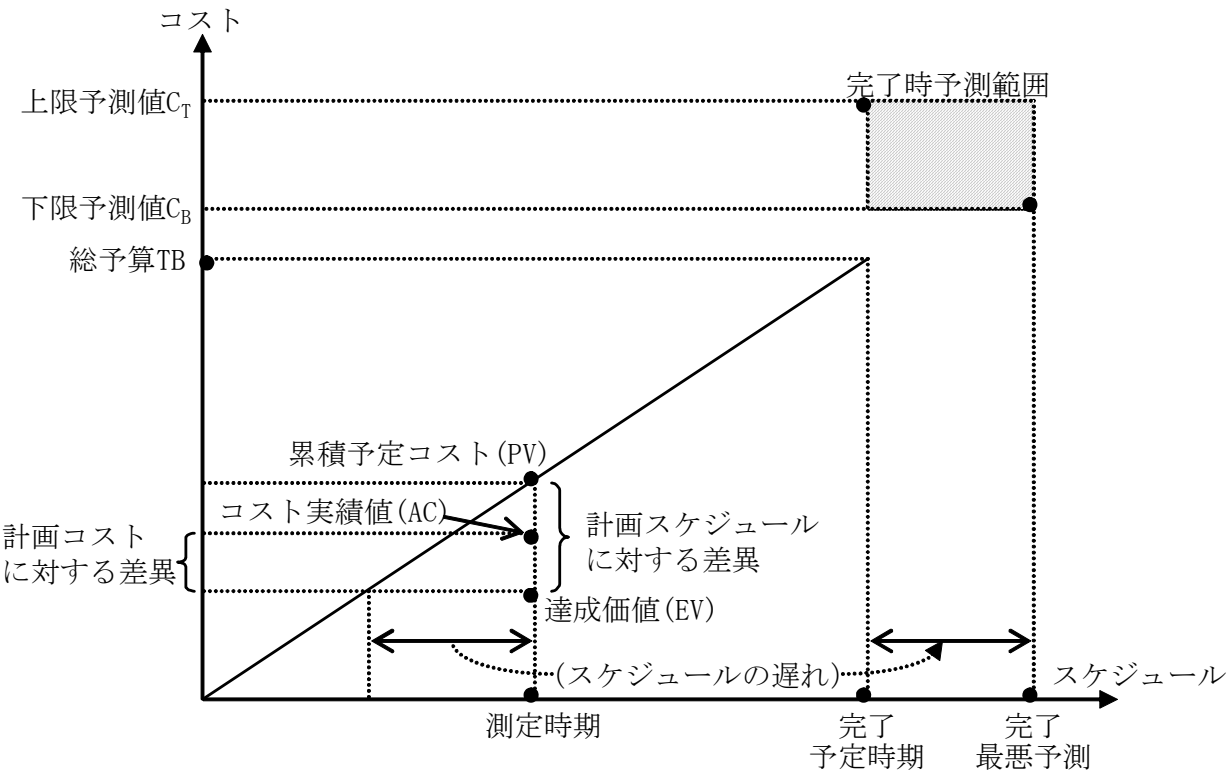


図 5.1　EVM でのプロジェクト進捗状況の把握と完了予測

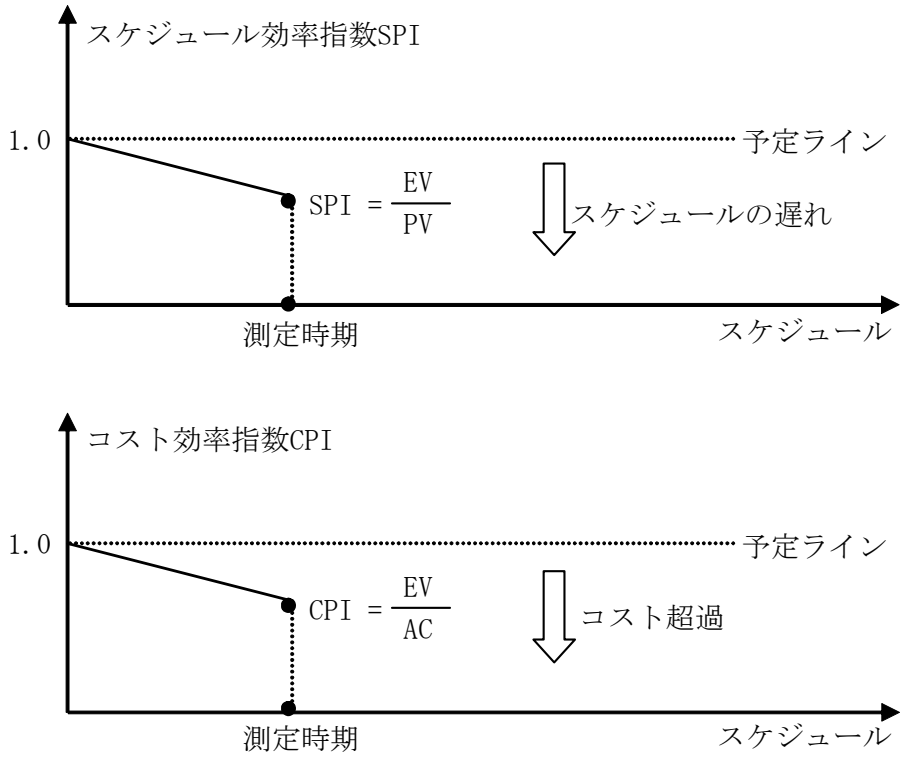


図 5.2　スケジュール効率指数とコスト効率指数の推移

5.3　提案する定量評価手法

本章では、ある遠隔会議システムが、通常の対面会議と比較して実際どの程度機能しているかという面での定量評価手法として、前節で説明した EVM という統合マネジメント技法の一部を参考にし、対象となる遠隔会議システムの会議進捗度を対面会議での会議進捗度との比較測定を行う新しい定量評価手法を提案する。提案する定量評価手法では、EVM での各作業のコストやスケジュールの見積り部分の省略と、EVM での作業コスト等の測定部分および測定結果からのプロジェクトの進捗状況の把握部分の転用を行っており、以下に同手法の利用の流れを説明する。

今回提案する定量評価手法では、対面会議でのある測定値を基準とし対象となるシステムを利用した遠隔会議での測定値と比較することで、対象となるシステムの会議進捗度を推定することを考えている。このため、対面会議での測定値がそのまま EVM での累積予定コスト PV のグラフ（図 5.1 を参照）に相当することになり、WBS 等による事前の詳細な見積りが必要なくなる。

次に、会議進捗度の測定について、**図 5.3** の測定グラフを参照しながら説明する。まず、対面会議および対象となる各会議での環境とスケジュールに関しては、参加者数を同数とし取り組む議題も、複数個の課題から構成される同一のものとする。これにより、測定時期を各課題が解決した時点等に定めることができ、かつ測定時の対面会議での PV 値および各会議での実際の達成価値 EV は一致する。一方、測定するコストは、会議開始から該当する課題が解決するまでに要した作業時間と参加者数の積とし、各会議の参加者数が一定である場合は、単純に会議開始から該当する課題が解決するまでに要した作業時間になる。この測定値が、基準となる対面会議の場合は、EVM での PV 値に相当し、比較対象である他の会議の場合は、基準となる対面会議では EVM でのコスト実績値 AC に相当する。

最後に、上述の測定結果から、対象となるシステムの会議進捗度と対面会議での会議進捗度とで比較評価する。まず、同一測定時期における対面会議での累計作業時間と、比較対象となる遠隔会議システムを利用した会議での累計作業時間の関係を示す会議進捗効率指数（Progress Performance Index：以下 PPC と記述）を定義する。この PPI は、EVM でのコスト効率指数 CPI の計算式（式３）を転用することで、

PPI =（対面会議での累計作業時間）/（対象となる会議での累計作業時間）　・・(式５)

になる。**図 5.4** は、この PPI の推移グラフを示しており、PPI の値が 1.0 より小さい場合は、対象となる遠隔会議システムを利用した会議の進捗度が、基準となる対面会議と比較して遅れていることを示す。また、同グラフ結果を会議の議題内容や作業形態の違い

等で分類し、それぞれについて再度グラフ化することで、対象となる遠隔会議システムを利用した会議の進捗度と議題内容との関係も把握することが可能となる。

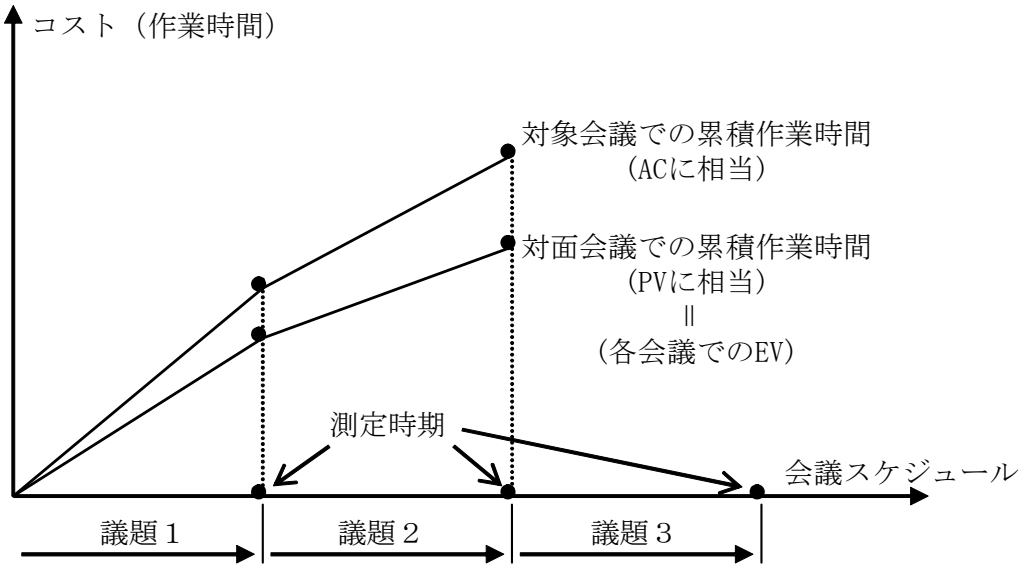


図 5.3　提案手法での会議進捗状況の把握

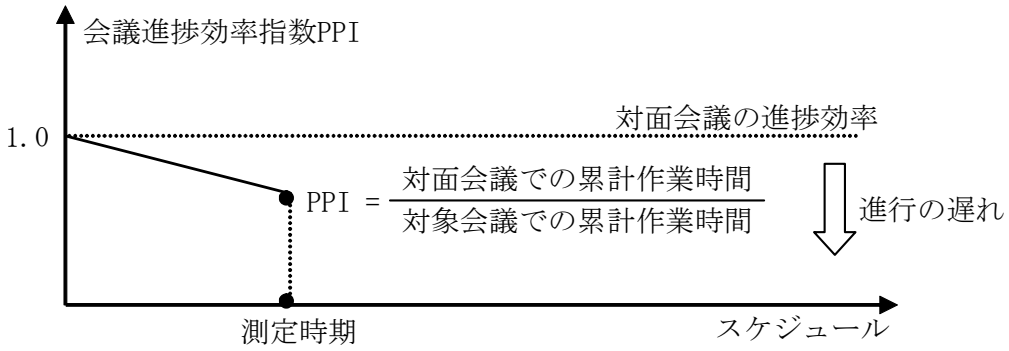


図 5.4　会議進捗効率指数の推移

5.4　提案する遠隔会議システムの構成

総合評価実験を行うにあたり、提案する遠隔会議システムの構成について整理する。

音声・映像のリアルタイム伝送手法に関しては、第 3 章で提案した DVTS と DV リアルタイム処理機構を活用する。これにより、円滑な会話を実現するために必要な音声・映像の品質および往復伝送遅延時間が達成できる。

図 5.5　改良した代理人ロボットの概観図

相手の意思表示動作の把握やアイコンタクト等のコミュニケーション支援手法に関しては、第 4 章で提案した代理人ロボットとジョイスティックによる遠隔操作システムを基本的に活用する。ただし、本章での総合評価を行うにあたり、代理人ロボットの改良を行った。第一の改良点としては、ある部分を指し示すというコミュニケーション支援動作を遠隔参加者も行えるようにするため、**図 5.5** に示すように上下左右旋回の遠隔操

作が可能な遠隔カメラの上部にレーザポインタを装着することで、レーザポインタを活用した、遠隔参加者による遠隔指示行為を可能とした。また同時に、ジョイスティックの操作命令割当についても、図 5.6 に示すようにレーザポインタの電源 ON/OFF を切り替えるボタンを用意した。第二の改良点としては、代理人ロボットの筐体内にすべての機器を収めることで Ethernet ケーブルと電源ケーブルがはずれない範囲内ではあるが、遠隔会議中に参加者 1 名による代理人ロボットの移動を可能とした。

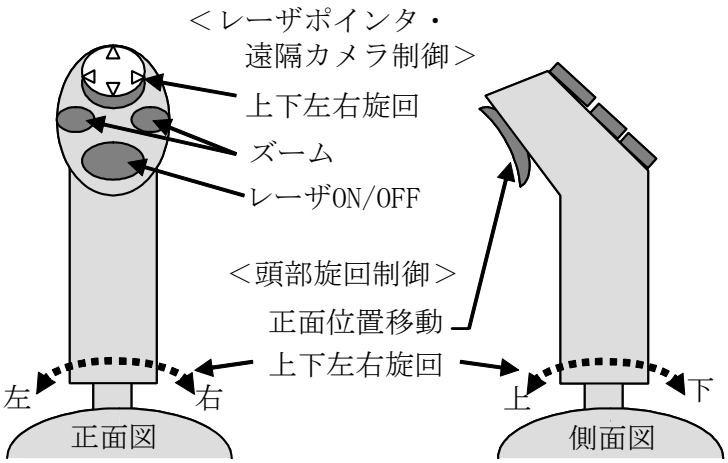


図 5.6　改良後のジョイスティックの操作命令割当

5.5　総合評価実験

5.5.1　被験者

被験者は、JAIST の大学院生 9 名である。被験者は全員、正常視力を有している。今回の実験では被験者 3 名からなるグループＡ、Ｂ、Ｃの３つに分けて評価実験を実施した。

5.5.2　実験構成

まず、比較対象となる会議環境は、被験者 3 名が同部屋内で課題に取り組む「対面会議」環境、被験者のうち 1 名が市販のテレビ会議システムを介して遠隔会議室から参加し他の 2 名と共に課題に取り組む「従来の遠隔会議」環境、そして同じ被験者 1 名が代理人ロボットを介して同様に取り組む「提案する遠隔会議」環境、の 3 種類とする。表 5.1 にそれぞれの会議環境の特徴を示す。

次に、各会議環境での総合評価実験のシステム構成を図 5.7 に従って説明する。

対面会議環境では、会議室内に被験者 3 名が集まり課題に取り組む。従来の遠隔会議環境では、市販の遠隔会議システムを利用して会議室と遠隔会議室間で遠隔会議を実施できる環境を用意し、被験者Ｚは遠隔会議室から会議室内の会議に遠隔参加し、一方会

議室内にいる被験者Ｘ、Ｙは、大型スクリーンに表示されている被験者Ｚと共に課題に取り組む。また、被験者Ｚは必要に応じて会議室に設置されているテレビ会議システムのカメラの遠隔操作をリモコンにより実行可能とする。提案する遠隔会議環境では、5.4で説明した代理人ロボットとその遠隔操作システムをそれぞれ会議室内と遠隔講義室内に設置し、DVTS による音声・映像リアルタイム伝送機能も利用することで会議室と遠隔会議室間で遠隔会議を実施できる環境を用意する。本評価実験では 2 地点遠隔会議の形態になっているが、多地点遠隔会議の形態にも対応できるように、第 3 章で提案した DV リアルタイムミキシング機構も活用した遠隔会議環境を模擬した。被験者Ｚは、遠隔会議室から会議室内の会議に遠隔参加し、会議室内にいる被験者Ｘ、Ｙは被験者Ｚの顔映像が表示されている代理人ロボットと共に課題に取り組む。また、被験者Ｚは必要に応じて会議室内の代理人ロボットの旋回動作、遠隔カメラのズーム操作、レーザポインタによる遠隔指示等をジョイスティック操作により実行可能とする。なお、上記２つの遠隔会議環境における会議室と遠隔会議室間のネットワーク伝送遅延時間は、遅延発生器を用いて第 3 章の評価実験において東京・大阪間のインターネット環境を模擬したときと同じ片道 7ms に設定した。

後術する定量評価分析を行うため、被験者Ｚの様子と会議室内の作業の様子を、評価分析用記録テープに録画した。**図 5.8**、**図 5.9**、**図 5.10** は、それぞれ対面会議環境、従来遠隔会議環境、そして提案遠隔会議環境での評価実験風景である。

表 5.1　比較対象となる会議環境の特徴

会議の種別	会議環境の特徴
対面会議	・被験者3名が同部屋内において、課題に取り組む。
従来の遠隔会議 （従来遠隔）	・市販のテレビ会議システムの利用。 →遠隔会議室からリモコンによる遠隔カメラ旋回、ズーム操作。 →会議室内の大型スクリーンに遠隔会議室の被験者映像を表示。
提案する遠隔会議 （提案遠隔）	・代理人ロボットとその遠隔操作システムの利用。 →遠隔会議室からジョイスティックによる遠隔操作。 →顔が表示されるモニタ等から構成される頭部の遠隔旋回。 →レーザポインタによる遠隔指示。 →等身大スケールの全長と表示される顔の大きさ。 →会議室内の被験者による代理人ロボットの移動。 ・DVTSとDVリアルタイム処理機構の利用。 →MPEG-4等に比べて低遅延の伝送を実現。

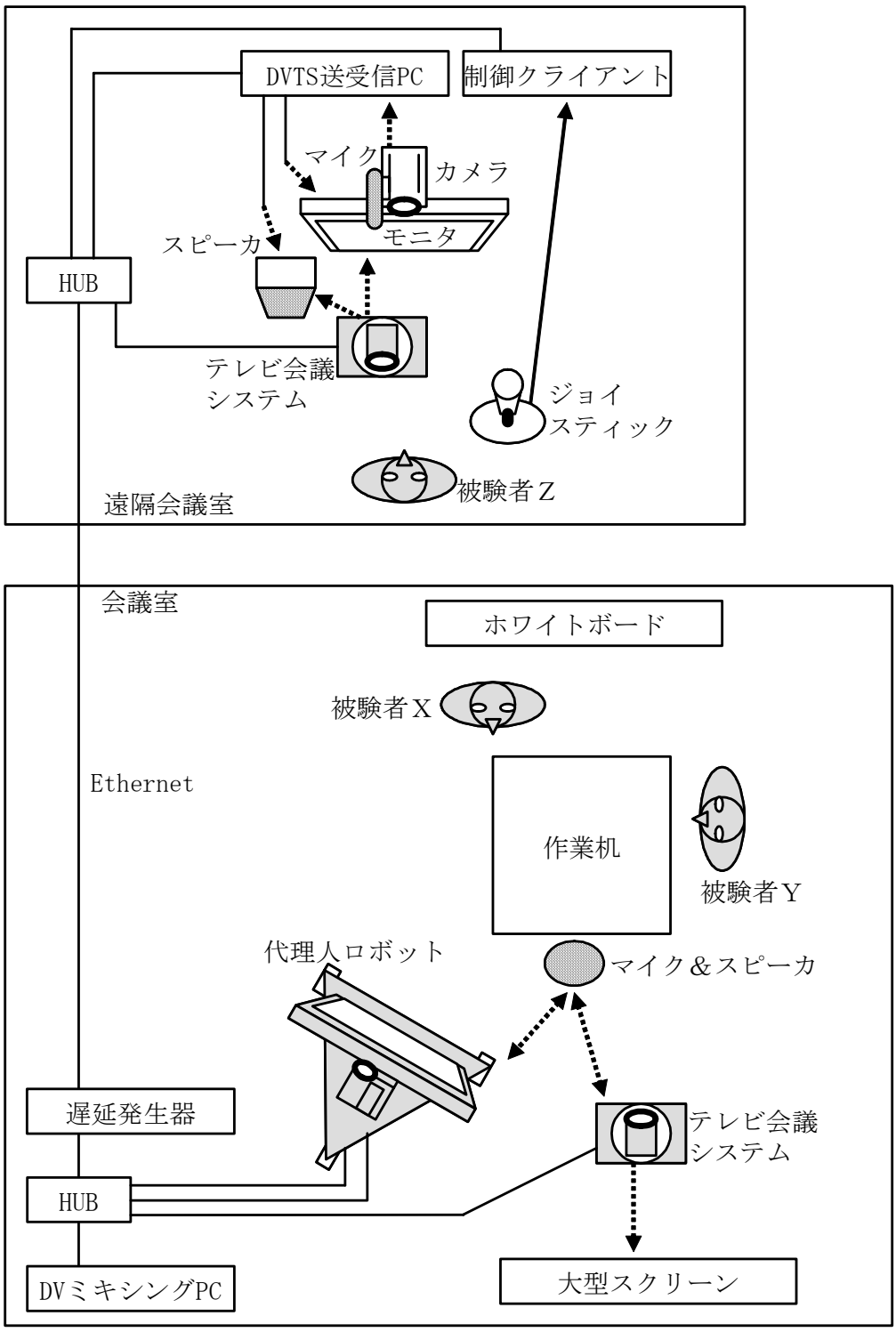


図 5.7　評価実験のシステム構成

図 5.8　対面会議環境での実験風景

図 5.9　従来遠隔会議環境での実験風景

図 5.10　提案遠隔会議環境での実験風景

5.5.3　実験手順

被験者 3 名からなる各グループには、**表 5.2** に示す 3 つの課題すべてに順次取り組んでもらった。各課題の取り組み手順は、まずホワイトボードを利用したシステム設計図を作成する「設計作業」を実施してもらい、設計作業終了後は実際の組立と検証を行う「組立作業」を実施してもらう。組立作業に必要な機器類はすべて会議室内の作業机周辺にあらかじめ用意してあり、各被験者が自由に利用することとした。

図 5.11 は、各グループに取り組んでもらった課題と会議環境の組合せと順番を示している。この組合せで実施した延べ 9 課題の実験結果を会議環境毎にまとめて分析することで、各グループ間の個体差や課題の難易度差を小さくし 3 つの会議環境の差のみを把握できるようにした。

5.5.4　提案手法による定量評価

5.3 で提案した新しい定量評価手法により、各システムにおける会議の進捗度を推定する比較定量評価を行う。本評価実験において測定時期は、各課題の設計作業および組立作業終了時の計 6 回になり、測定する値は各会議での累積作業時間とした。会議の進捗状況を把握するグラフ（図 5.3 を参照）とその結果から作成する会議進捗効率指数 PPI の推移グラフ（図 5.4 を参照）から、各システムにおける会議の進捗度の推定と評価を行う。

図 5.12 は、図 5.11 に従って行われた評価実験結果を対面会議、従来遠隔会議、提案遠隔会議の 3 種類に区分して各測定時期での累積作業時間を図 5.3 の形態に従いプロットしてグラフである。**図 5.13** は、同測定結果から算出した会議進捗効率指数 PPI を、図 5.4 の形態に従ってプロットしたグラフである。同図より、提案遠隔会議は、従来遠隔会議よりは会議が効率よく進捗する傾向にあり、対面会議環境により近い環境になっていると推定できる。

本評価実験では同じ系統の課題に選び、同課題内では作業形態の異なる設計作業と組立作業に分割して取り組んでもらった。このため、次に設計作業形態、組立作業形態それぞれに対する各システムにおける会議の進捗度の推定と評価を行う。**図 5.14** は図 5.13 のうちの設計作業形態と組立作業形態に分類して、それぞれの形態での会議進捗効率指数 PPI の推移をグラフ化したものである。このグラフ結果からは、設計作業に関しては提案遠隔会議が対面会議とほぼ同等の環境を提供できているが、組立作業に関しては一転して、提案遠隔会議の進捗度合いが従来遠隔会議とほぼ同じにまで低くなっていることが分かる。この原因は、組立作業で実際に機器の移動や結線といった作業を行う人数が、体面会議では 3 名であるのに対し遠隔会議環境では 2 名となってしまうため、パフォーマンスが低下していることが大きな要因になっていると考えられる。

表 5.2　課題の内容

課題の種別	内容
課題１	3地点遠隔会議において以下を実現する模擬システムを構築してください。 ・音声ミキサーを利用して他の2会場の音声のみ聞こえるようにする。
課題２	2地点遠隔会議において以下を実現する模擬システムを構築してください。 ・エコーキャンセラー2台を利用してエコーバックを極力小さくする。
課題３	以下を実現する映像編集システムを構築してください。 ・6つのカメラ映像を1つのモニタ画面内に4分割、7分割等で表示する。 ・各映像の表示位置は任意に選択可能にする。
共通事項	・まず設計作業として、機器構成や結線等のシステム設計図を ホワイトボード上に作成する。 ・設計作業が終了したら、組立作業に移行し検証まで行う。 ・各作業は被験者3名で協力して取り組む。

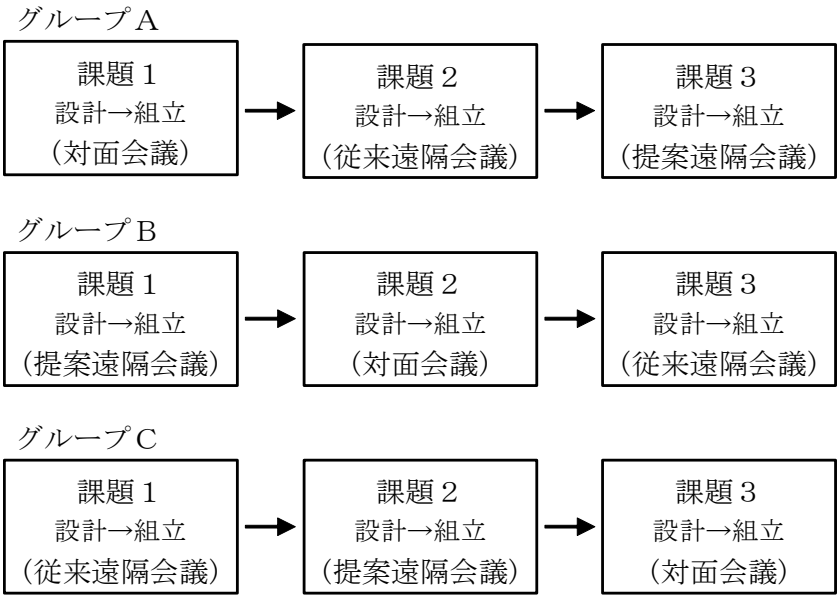


図 5.11　評価実験の組合せと順番

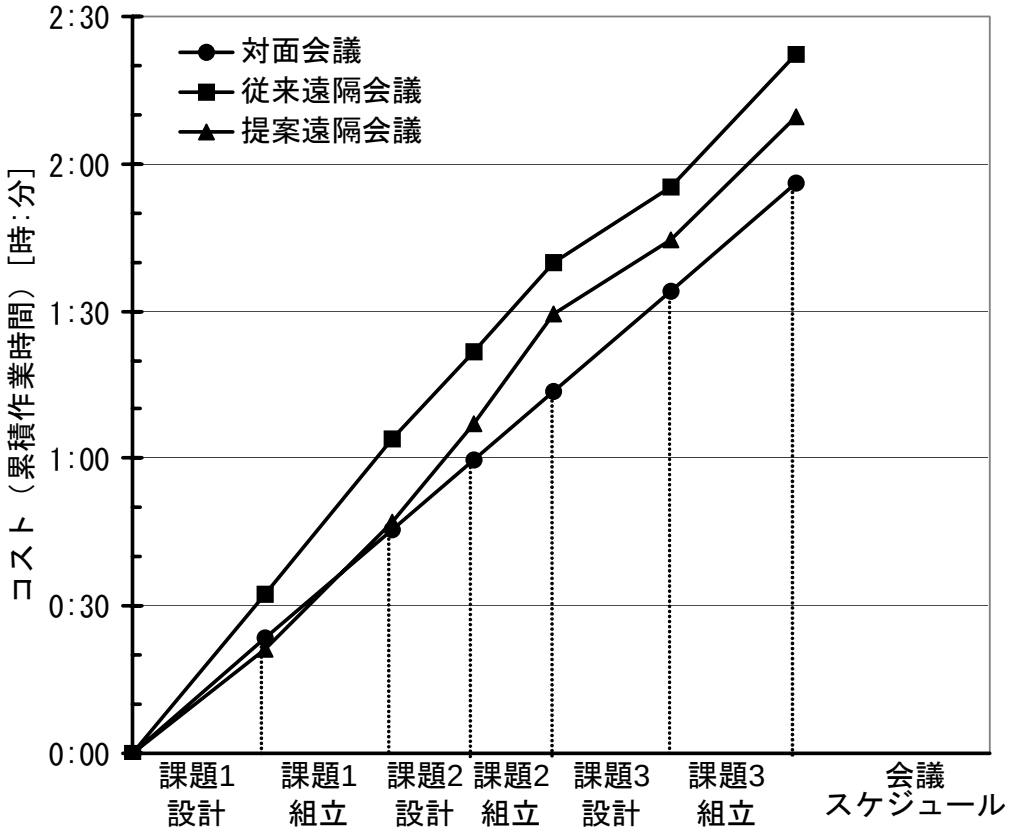


図 5.12　各システムでの累積作業時間の推移

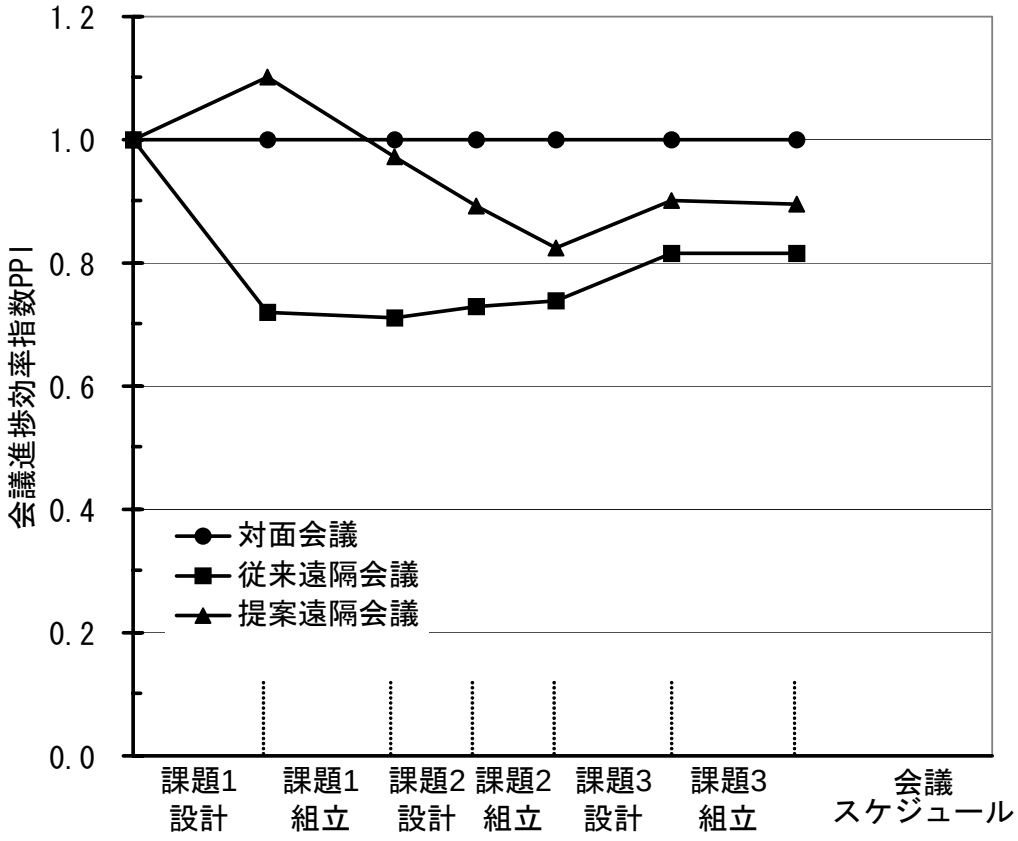


図 5.13　各システムでの会議進捗効率指数の推移

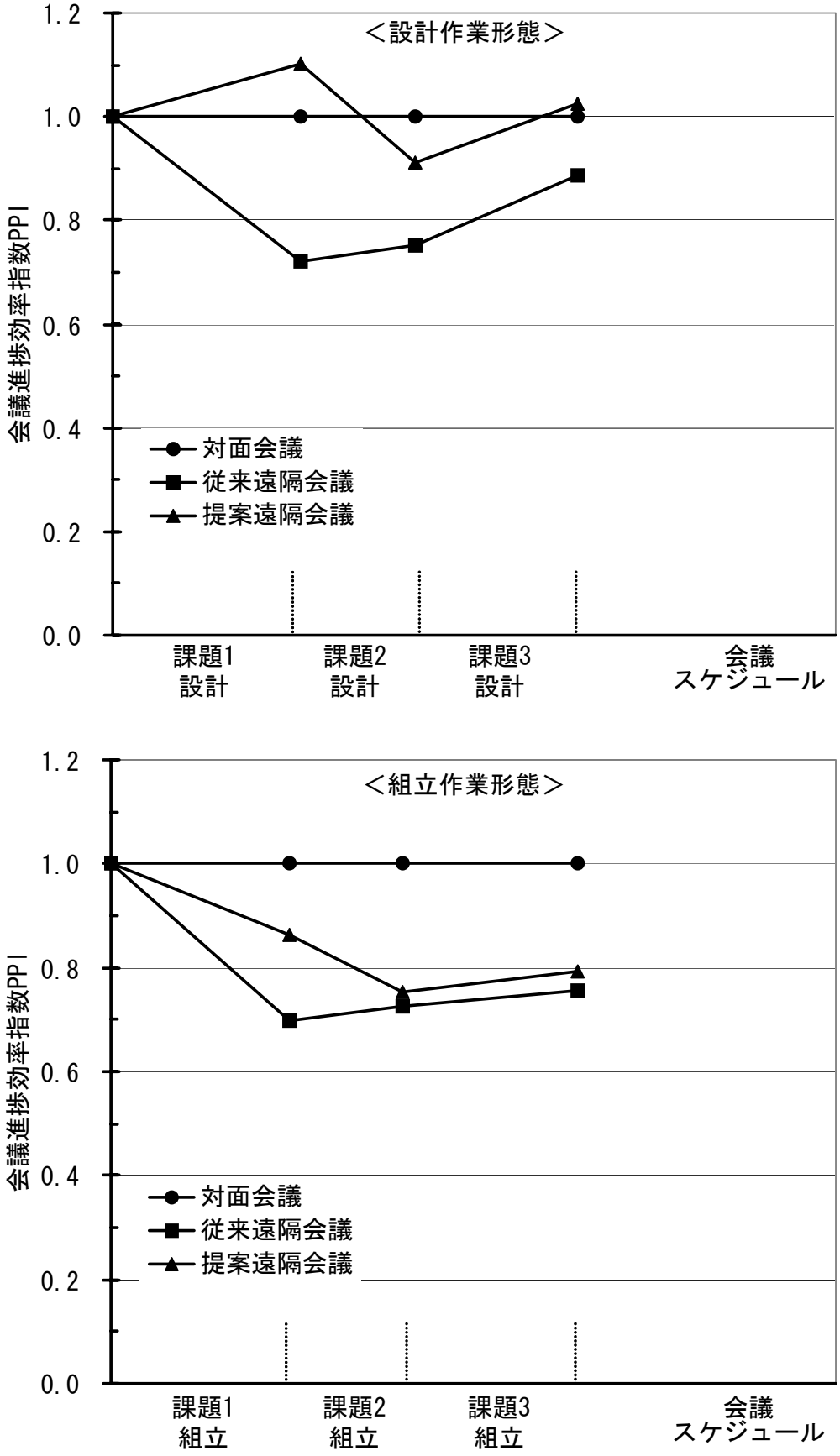


図 5.14　作業形態毎の会議進捗効率指数の推移

次に、本評価実験中での設計作業および組立作業に関する熟練度の効果について評価を行う。**図 5.15** は、設計作業と組立作業のそれぞれ 1 回目（課題 1)、2 回目（課題 2)、3 回目（課題 3）と作業経験を積むに従い作業時間がどのように変化しているかを示すグラフである。**図 5.16** は、図 5.15 を 1 回目の作業時間を基準とした作業時間変動比率に変換したグラフである。これらのグラフより、設計作業に関しては、各グループとも設計作業経験数の増大にともない、作業時間は減少していく傾向にあることはわかり、特に設計作業 2 回目の作業時間は 1 回目から約 30%も減少している。この作業時間の変化は、課題の難易度によるものと、作業熟練度の効果によるものの大きく 2 つの要因が考えられるため、このすべてが作業熟練度の効果によるものとはいえないが、本評価実験中の設計作業に関しては、ある程度の熟練度の効果があると考える。一方、組立作業に関しては、グループ毎に作業時間の変化が異なっており、本評価実験中の組立作業に関しては作業熟練度の効果がないと考えられる。

以上の結果より、手足を使った実作業をともなわない設計作業等が中心の会議内容の場合には、提案遠隔会議システムは有効に機能し、会議進捗度合いという面では対面会議とほぼ同等のパフォーマンスを実現できると推定される。ただし、今回は提案評価手法の有効性を従来の評価手法結果との比較から推定するため、設計・組立という極めてシンプルで類似した課題を 3 グループに 3 つずつ取り組んでもらうという試験的な環境で評価を行った。しかし、提案評価手法は本来 EVM と同様に実際の会議において長期的に測定することでその真価を発揮し、長期的な会議の途中段階における精度の高い会議進捗度の評価が可能となると考える。

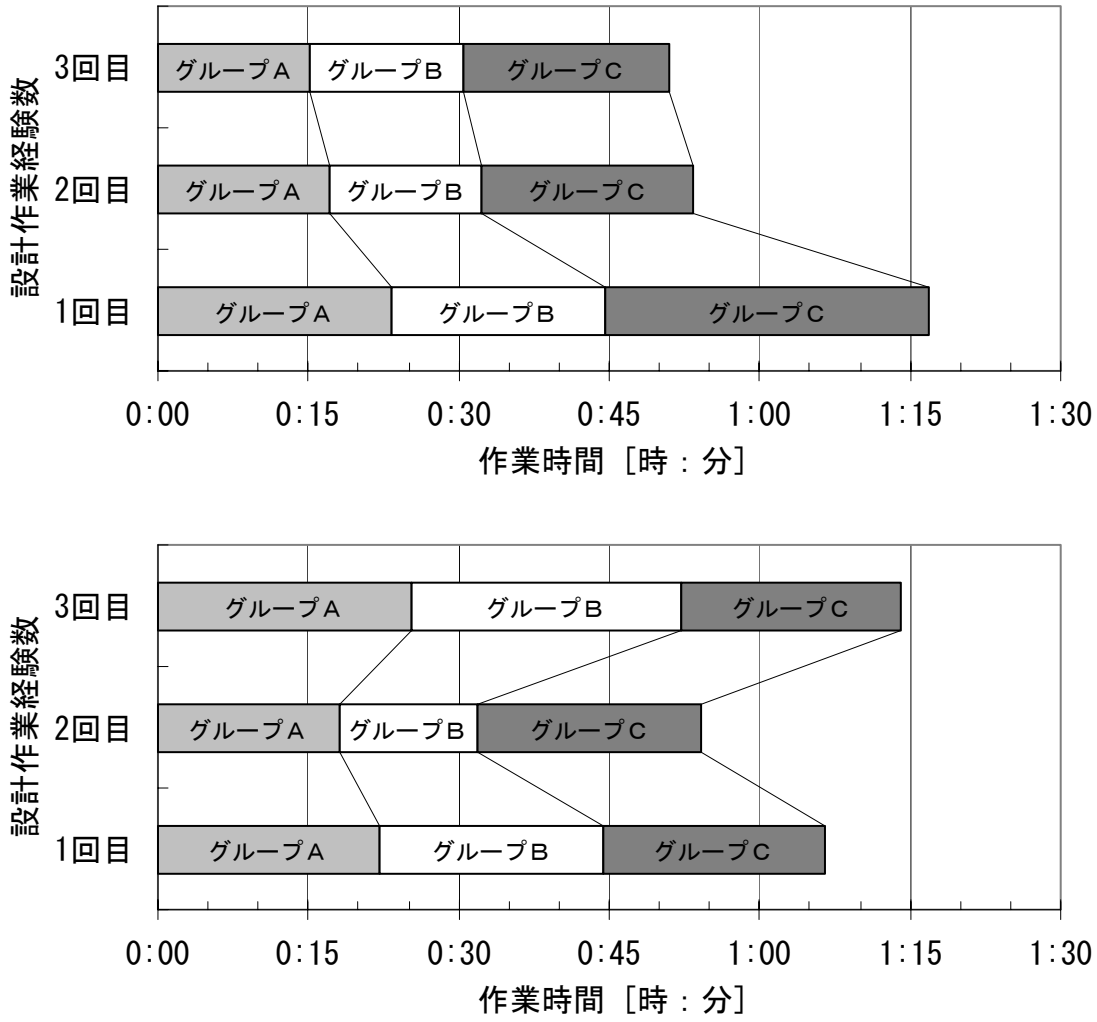
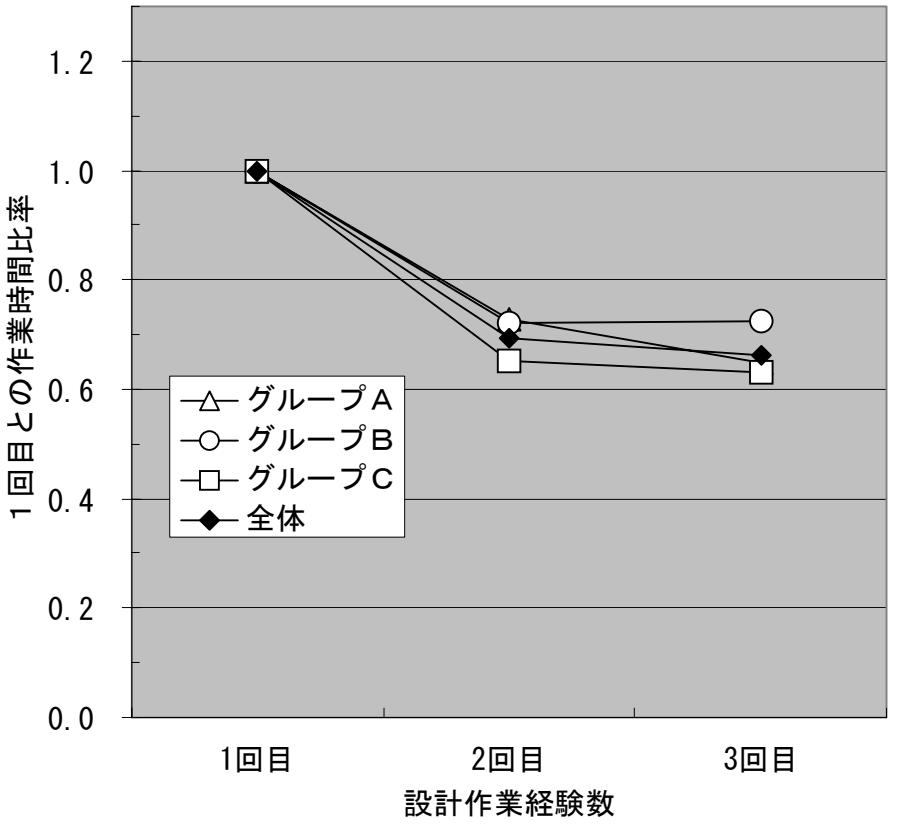


図 5.15　作業経験数による作業時間の変化



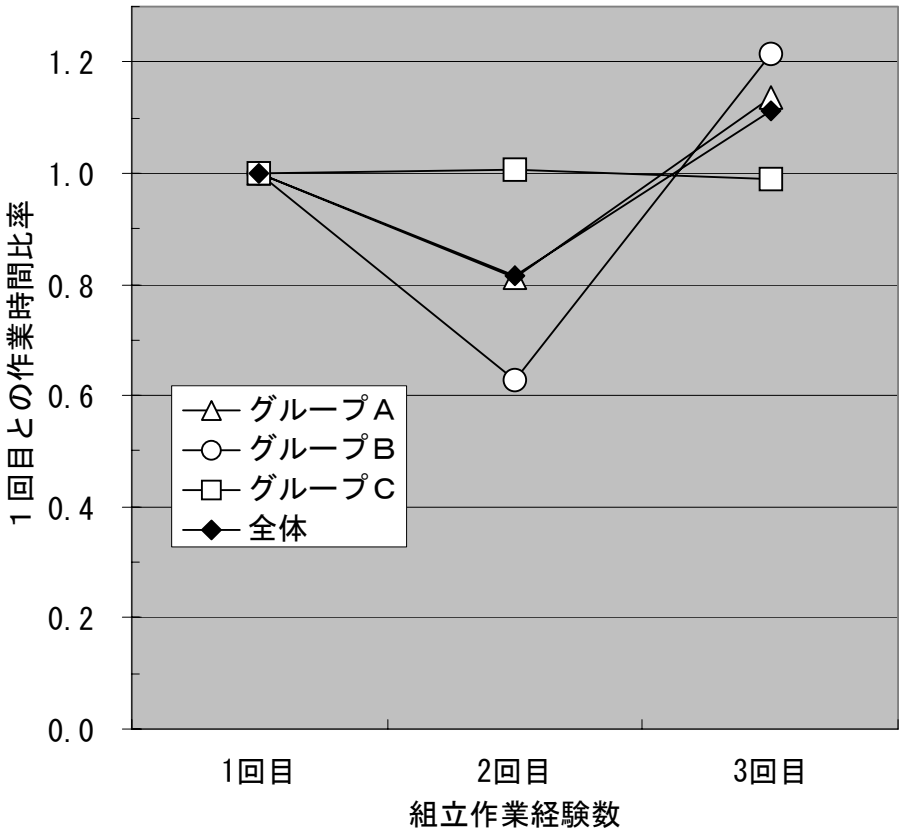


図 5.16　作業経験数による作業時間変動比率の変化

5.5.5　被験者の行動分析による定量評価

会議室または遠隔会議室で作業を行う被験者Ｘ、Ｙ、Ｚがそれぞれコミュニケーションをとるためにとった行動回数等の行動分析[40]による定量評価を行った。行動方向としては、被験者Ｚから被験者Ｘ、Ｙに向けて行われる行動（遠隔会議では遠隔会議室から会議室へのコミュニケーション行動)、被験者Ｘ、Ｙから被験者Ｚに向けて行われる行動（遠隔会議では会議室から遠隔会議室へのコミュニケーション行動)、被験者ＸとＹの 2 者間のコミュニケーション行動に区分できる。コミュニケーション行動内容としては、相手に向かっての発話行動、ジェスチャ行動（発話をともなっての行動も含む）のほか、遠隔会議室にいる被験者Ｚの場合は、自分の顔映像が表示されている代理人ロボット頭部を遠隔旋回させてある方向を振り向く行動（発話をともなっての行動も含む)、代理人ロボットに搭載されているレーザポインタを利用した遠隔指示行動（発話をともなっての行動も含む）が挙げられる。

表 5.3 は、図 5.11 に従って行われた各グループの評価実験における各被験者のコミュニケーション行動回数を示している。ただし、代理人ロボット頭部の遠隔旋回行動（表中では顔表示モニタ旋回と表記）については、1 回の旋回行動後 3 秒以内に再度旋回する行動は一連の行動とみなしカウントしていない。また、相手に向かっての発話行動は、ある被験者の発話がトリガーとなって開始される、被験者 3 名による一連の会話をまとめて１つとカウントしている。なお、一連の会話の区切りの目安として今回は 3 秒以上の無会話状態が続いた時点と判定している。

図 5.17、図 5.18、図 5.19 は、表 5.3 を各グループの行動回数を累計し、それぞれ被験者Ｚから被験者Ｘ、Ｙに向けて行われるコミュニケーション行動回数、被験者Ｘ、Ｙから被験者Ｚに向けて行われるコミュニケーション行動回数、被験者ＸとＹの 2 者間のコミュニケーション行動回数に分類して、グラフ化したものである。

まず、図 5.17 および図 5.18 より、遠隔地間のコミュニケーション行動回数に関して考察を行う。設計作業の場合には、提案する遠隔会議形態は対面会議形態とほぼ同等の行動回数を実現していることが分かる。また、従来の遠隔会議形態はかなり少なくなっており、特に遠隔会議室から参加する被験者Ｚからのコミュニケーション行動回数が著しく減少しほぼ静観している状態であることが分かる。一方、組立作業の場合には、提案する遠隔会議形態は、従来の遠隔会議形態ほど悪くはないものの、対面会議の場合と比較するとコミュニケーション行動が 40%程度少なくなっており、その内容に関しても発言が著しく少ないことが分かる。また、代理人ロボット頭部を遠隔旋回させることによる作業状況の把握行動の占める割合が高いことも判明した。

次に、図 5.19 も含めた会議形態毎のコミュニケーション行動回数に関して考察する。従来の遠隔会議形態の場合、対面会議形態と比較して設計作業においても会議室内の 2

名の被験者間のコミュニケーション行動回数が遠隔地間のコミュニケーション行動回数を補完する形でかなり増大していることが分かる。これは、遠隔参加者だけでなく会議室内の参加者にとっても、遠隔地間のコミュニケーションが会議室内でのコミュニケーションに比べて負担が大きいことを示している。また、代理人ロボットを利用した提案遠隔会議形態においても若干ではあるがこの傾向が確認できる。

表 5.3　各グループの被験者のコミュニケーション行動回数

			対面会議		従来遠隔		提案遠隔	
			課題１		課題２		課題３	
	行動方向	行動	設計	組立	設計	組立	設計	組立
グループＡ	被験者Ｚ→被験者Ｘ，Ｙ	発話	7	12	3	1	4	2
		ジェスチャー＊	5	3	0	0	0	0
		顔表示モニタ旋回＊	－	－	－	－	2	5
		レーザポインタON＊	－	－	－	－	2	1
	被験者Ｘ，Ｙ→被験者Ｚ	発話	12	14	9	3	14	10
		ジェスチャー＊	3	3	0	0	0	0
	被験者Ｘ－Ｙ間	発話	19	16	20	19	20	22
		ジェスチャー＊	4	8	5	7	6	4

（＊：発話しながらの行動も含む）

			提案遠隔		対面会議		従来遠隔	
			課題１		課題２		課題３	
	行動方向	行動	設計	組立	設計	組立	設計	組立
グループＢ	被験者Ｚ→被験者Ｘ，Ｙ	発話	4	4	7	13	2	1
		ジェスチャー＊	0	0	3	5	0	0
		顔表示モニタ旋回＊	3	5	－	－	－	－
		レーザポインタON＊	3	0	－	－	－	－
	被験者Ｘ，Ｙ→被験者Ｚ	発話	13	9	15	16	7	4
		ジェスチャー＊	2	0	2	0	0	0
	被験者Ｘ－Ｙ間	発話	23	13	18	15	23	20
		ジェスチャー＊	4	9	6	7	8	8

（＊：発話しながらの行動も含む）

			従来遠隔		提案遠隔		対面会議	
			課題１		課題２		課題３	
	行動方向	行動	設計	組立	設計	組立	設計	組立
グループＣ	被験者Ｚ→被験者Ｘ，Ｙ	発話	1	0	2	2	3	7
		ジェスチャー＊	0	0	0	0	3	6
		顔表示モニタ旋回＊	－	－	5	11	－	－
		レーザポインタON＊	－	－	2	0	－	－
	被験者Ｘ，Ｙ→被験者Ｚ	発話	9	5	14	10	15	18
		ジェスチャー＊	0	0	1	0	2	1
	被験者Ｘ－Ｙ間	発話	32	28	24	16	24	18
		ジェスチャー＊	5	9	5	7	2	6

（＊：発話しながらの行動も含む）

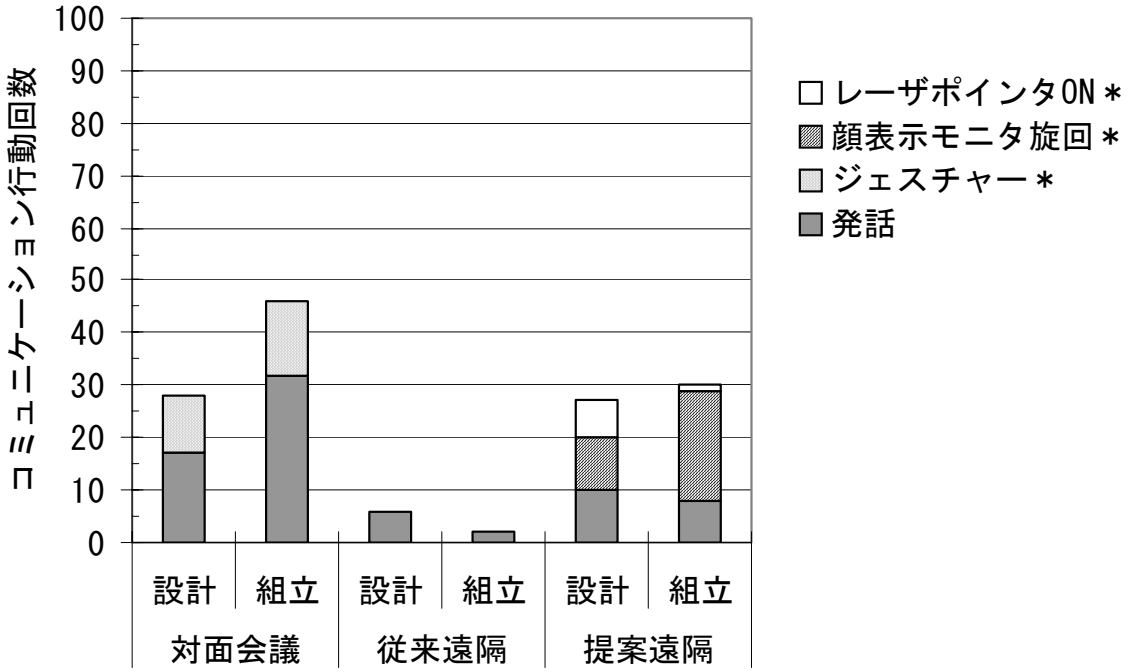


図 5.17　被験者Ｚから被験者Ｘ，Ｙへのコミュニケーション行動回数

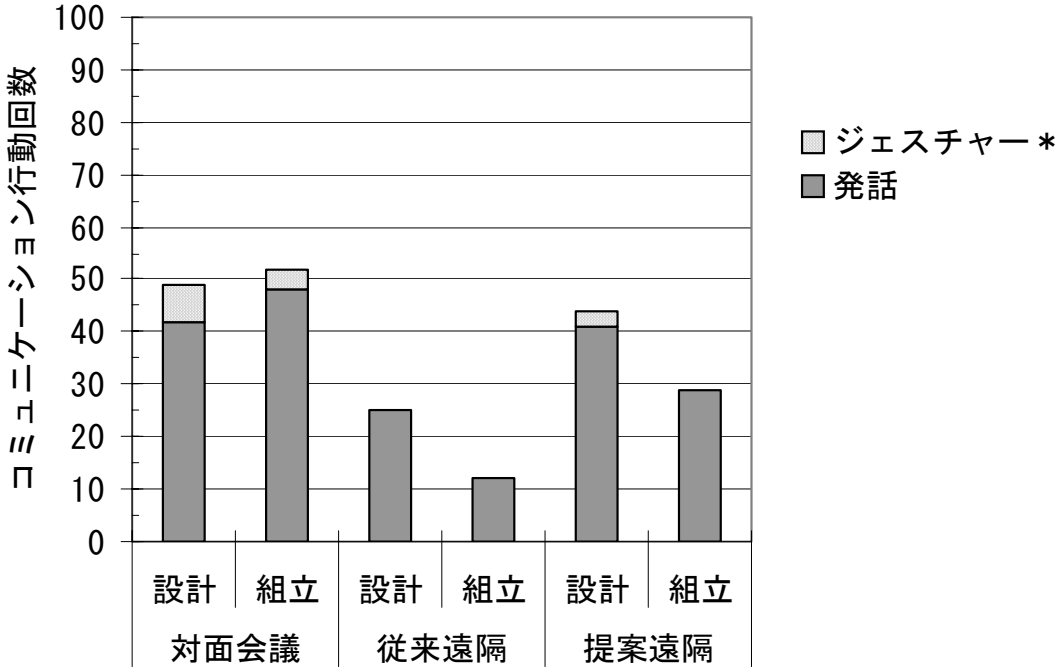


図 5.18　被験者Ｘ，Ｙから被験者Ｚへのコミュニケーション行動回数

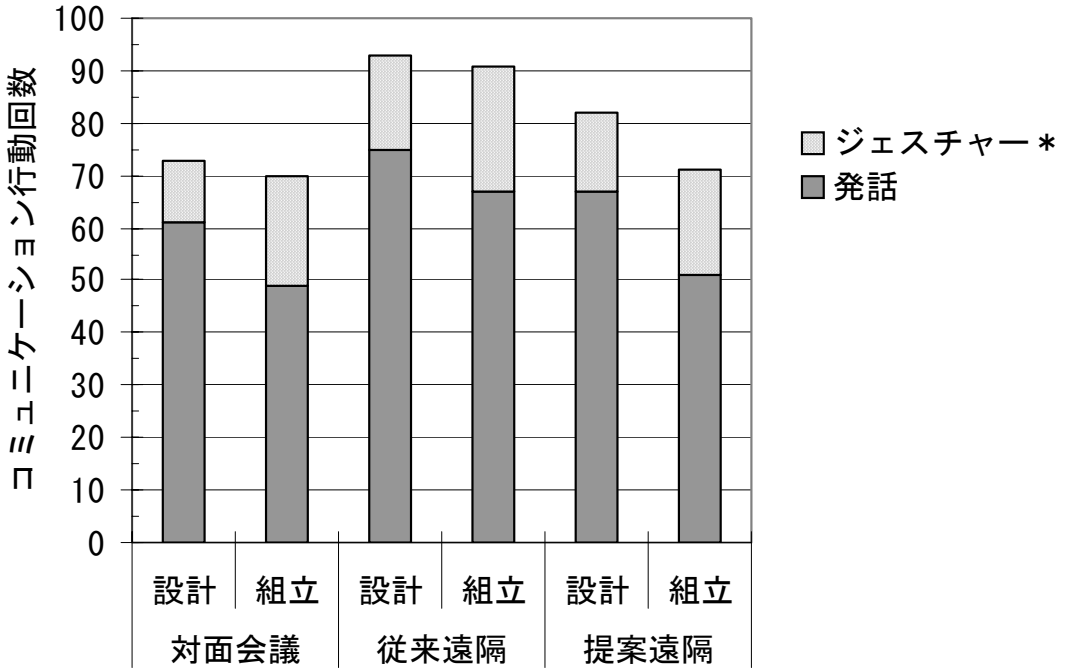


図 5.19　被験者Ｘ－Ｙ間のコミュニケーション行動回数

ここで、上述した被験者の行動分析による定量評価データと、5.5.4 の提案手法による定量評価データとの間に相関関係があるかどうかを考察する。「グループ C・課題２・提案遠隔会議」でのデータに着目すると、図 5.12 に示した提案手法による各システムでの累積作業時間の推移に関しては、他の条件下に比べて作業時間が長くなっていることが分かる。一方、表 5.3 に示した被験者のコミュニケーション行動回数に関しても、同条件下のデータは、やはり他の条件下に比べて代理人ロボットの顔表示モニタの旋回動作回数が多いことが分かる。すべての記録ビデオを再度分析した結果、この相関関係は遠隔参加者の旋回動作に反応して、会議室内の被験者が議論や作業を一時中断するという行動パターンが、グループ C で頻繁に発生していることが分かった。つまり、この現象はグループの個体差が主要因で発生したと考えられる。この現象は、代理人ロボットの顔表示モニタの旋回動作が同じ会議室内にいる参加者の行動に大きな影響を与えることを示しており、無駄な旋回動作は会議の進行を妨害する大きな要因になるということも示している。

5.5.6　アンケートによる主観的評価

評価実験終了時に被験者に対してアンケートを実施した。アンケートの評価尺度としては、一般的に絶対評価より相対評価の方が個々の差が出やすいが、今回のように 3 つ以上の比較の場合は比較の次元が異なり矛盾が起きやすい。このため、今回は**表 5.4** に示す 5 段階の絶対範疇尺度（Absolute Category Rating：以後 ACR と記述)によるオピニオン評価を採用した。**図 5.20** に、今回の評価実験での会議環境に関する平均オピニオン値（Mean Opinion Score：MOS）の結果を示す。この結果は被験者 9 名という少数によるものであるため、この結果のみで明確な考察をすることはできないが、傾向として提案遠隔会議は従来遠隔会議よりも会話および共同作業という面で改善されているが、対面会議環境には及ばない状況にあるといえる。

表 5.4　評価に用いた 5 段階の絶対範疇尺度

評点	評価尺度
5	非常によい
4	よい
3	普通
2	あまりよくない
1	全然よくない

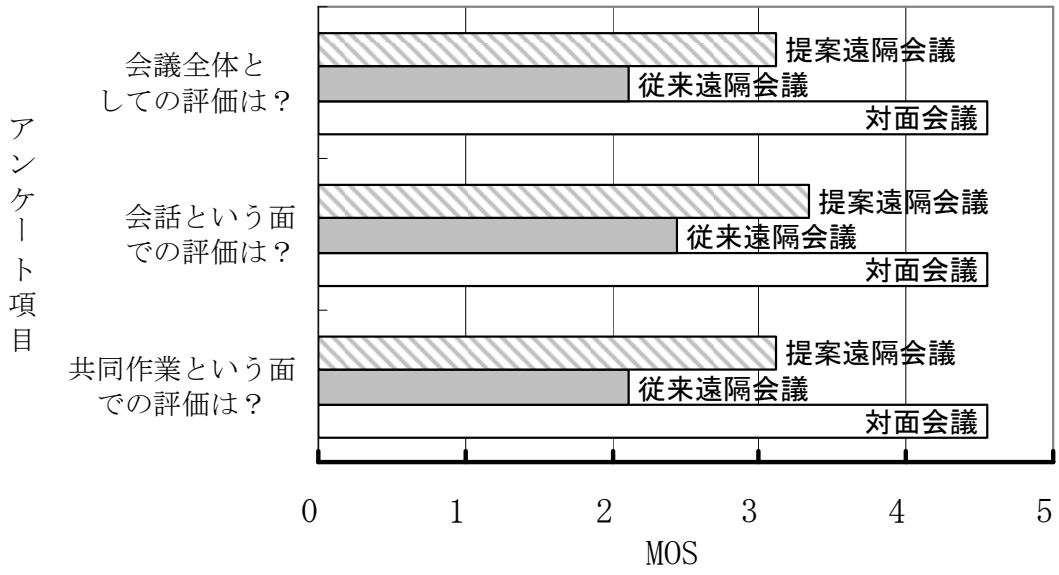


図 5.20　評価実験での会議環境に関する平均オピニオン値の結果

5.6 まとめ

本章では、DV リアルタイム処理機構と代理人ロボットを活用する提案遠隔会議システムが、実際の多地点遠隔会議という環境の中でどの程度有効に機能するかを検証するため、上述の 2 つの機能を組合せた遠隔会議模擬環境を構築し、通常の対面会議環境および従来の遠隔会議システムを利用した遠隔会議模擬環境との総合比較評価を行った。評価方法の１つの柱として、プロジェクトの進捗や作業のパフォーマンス等の定量評価に用いられる EVM という統合マネジメントの技法を参考に、各システムにおける会議の進捗度を推定する新しい手法を提案し、評価実験での評価を試みた。その結果、手足を使った実作業をともなわない設計作業等が中心の会議内容の場合には、提案遠隔会議システムは有効に機能し、会議進捗度合いという面では対面会議とほぼ同等のパフォーマンスを実現できるという定量評価結果を得た。また、遠隔参加者がコミュニケーションをとるためにとったと思われる行動回数に関する定量評価結果と、各システムを利用した被験者に対してアンケートを行う主観的評価結果は双方とも、今回提案した新しい定量評価手法による結果と同じ傾向を示していることが分かった。

以上、多地点遠隔会議において自然なコミュニケーションを実現し活発な議論が可能となる多地点遠隔会議システムとして、本論文で提案した DV リアルタイム処理機構、および代理人ロボットとその遠隔操作試作システムの適用は、手足を使った実作業をともなわない設計作業等を中心とする会議形態においては有効に機能し対面会議に近いパフォーマンスを実現する可能性を持っていると結論する。

第6章　結論

6.1　本論文の総括

本研究では、3 地点以上の地理的に離れた会場間が、インターネットで接続されている環境で行われる多地点遠隔会議において、遠隔地間との円滑な会話やアイコンタクトによるスムーズな話者交替の実現等の自然なコミュニケーションを実現し、活発な議論が可能となる多地点遠隔会議システムの提案および有効性の評価を行った。以下に本論文の総括を行う。

第 2 章では、まず遠隔会議に関するこれまでの取組んできた調査および実証実験結果を中心に、遠隔会議において対面会議と同様な活発な議論を展開し、自然なコミュニケーションを実現するための要件と現在の遠隔会議システムの課題を整理した。その結果、遠隔会議において自然なコミュニケーションを実現するための要件として、「遠隔地にいる参加者の表情の把握」、「円滑な会話の実現」、「遠隔地にいる参加者とのスムーズな話者交替の実現」が重要であることを示した。本論文では、音声・映像の伝送方式や音声ミキシング等のリアルタイム加工処理の改善による往復伝送遅延時間短縮の実現と、遠隔参加者の分身となる代理人ロボットとその遠隔操作システムの導入による遠隔地にいる参加者の意思表示動作の把握やアイコンタクト等の実現、の２つの研究開発を中心にして課題解決をはかる遠隔会議システムの構想を述べた。

第 3 章では、多地点に分散している会議参加者が、地上のインターネットを活用し、音声と実写映像を中心とするストリーミング情報をリアルタイムに送受信することで、会議を行う双方向リアルタイム遠隔会議に焦点をあて、自然なコミュニケーションの実現手法に基づいたシステムに関して議論を行った。そして、多地点遠隔会議の場合での円滑な会話の実現手法として、符号化処理時間が短くかつ符号化後の音声・映像等のストリーミング情報の扱いが容易な DV を採用し、インターネット上を流れる IP パケットに搭載されているデジタル状態の複数の DV ストリーミング情報に対して、中間ノードで直接加工処理を行う DV リアルタイム処理機構を提案した。そして、本機構と DVTS を併用することで NTSC テレビ放送クラスの音声・映像を用いた多地点遠隔会議の模擬システムを構築し、定量測定実験および被験者による評価実験を行った。評価の結果、DV リアルタイム処理機構を用いた多地点遠隔会議システムは、従来の遠隔会議システムと比較して、相手の表情まで把握できる NTSC 放送品質クラスの映像を提供し、かつ往復伝送遅延時間を 57%低減化することに成功し、円滑な会話の実現可能性を示した。また、多地点遠隔会議環境での同機構の利用は、ストリーミング情報に対するアナログ・デジタル

間信号変換処理回数の削減につながり、結果ストリーミング情報の品質劣化抑制という面においても有効に機能していることがわかった。

第 4 章では、リアルタイム遠隔会議システムにおいて自然なコミュニケーションを実現するための手法として、複数の物理的に離れている会場空間であっても、同一会場空間の場合と同じような空間共有を体感できるシステムについて議論した。そして、他の参加者と実際に並んで会議に参加することを目的として、ほぼ実物大の遠隔参加者の実写顔映像が表示されるモニタや、目となる遠隔カメラ等が一体となっている頭部が、遠隔操作により実際に上下左右に振り向く動作が可能な代理人ロボットを開発した。

次にマウスとジョイスティックという代表的な入力デバイスによる 2 種類の異なる遠隔操作システムを試作し、被験者による操作性に関する評価実験を行い、どのような遠隔操作環境が使用者に負担が少なく会議に集中できるかという、自然なコミュニケーション実現に重要な要件について考察した。評価結果としては、ジョイスティックの傾き量に応じて、現在位置から自動的に傾き方向へ相対旋回動作を行う遠隔操作試作システムのほうが評価は高いことがわかった。そして、代理人ロボットを利用した遠隔会議では、ブラインド操作が簡便で会議映像画面に集中した状態のまま遠隔操作が可能な手法が、自然なコミュニケーション実現に重要な要件であるという考察結果を得た。

また、代理人ロボットとその遠隔操作システムが、遠隔地間の参加者同士のアイコンタクトおよび話者交替の実現の面で有効に機能しているかどうかの評価を、遠隔会議模擬実験での被験者の行動分析とアンケート評価から試みた。行動分析評価の結果、代理人ロボットの有無に関わらず、話者交替におけるアイコンタクトの重要性が非常に高いことが確認されると共に、遠隔会議での代理人ロボットの旋回動作が、アイコンタクトおよび話者交替と強い相関性を持っていることが分かった。このため、代理人ロボットの旋回動作が、遠隔地間の参加者同士のアイコンタクトおよび話者交替の実現の面で有効に機能しているという考察結果を得た。また、アンケート評価の結果、代理人ロボットを介して遠隔参加する側には、代理人ロボット利用による遠隔地との会話は容易になり会話のための負担も少ないことが分かった。一方、代理人ロボットと同席し会議を行う参加者側には、代理人ロボットを介した会話に対する負担の低減化という検討課題があることも明らかになった。

第 5 章では、DV リアルタイム処理機構と代理人ロボットを活用する提案遠隔会議システムが、実際の多地点遠隔会議という環境の中でどの程度有効に機能するかを検証するため、上述の 2 つの機能を組合せた遠隔会議模擬環境を構築し、通常の対面会議環境および従来の遠隔会議システムを利用した遠隔会議模擬環境との総合比較評価を行った。評価方法の１つの柱として、プロジェクトの進捗や作業のパフォーマンス等の定量評価

に用いられる EVM という統合マネジメントの技法を参考に、各システムにおける会議の進捗度を推定する新しい手法を提案し、評価実験での評価を試みた。その結果、手足を使った実作業をともなわない設計作業等が中心の会議内容の場合には、提案遠隔会議システムは有効に機能し、会議進捗度合いという面では対面会議とほぼ同等のパフォーマンスを実現できるという定量評価結果を得た。また、遠隔参加者がコミュニケーションをとるためにとったと思われる行動回数に関する定量評価結果と、各システムを利用した被験者に対してアンケートを行う主観的評価結果は双方とも、今回提案した新しい定量評価手法による結果と同じ傾向を示していることが分かった。

以上、多地点遠隔会議において自然なコミュニケーションを実現し活発な議論が可能となる多地点遠隔会議システムとして、本論文で提案した DV リアルタイム処理機構、および代理人ロボットとその遠隔操作試作システムの適用は、手足を使った実作業をともなわない設計作業等を中心とする会議形態においては有効に機能し対面会議に近いパフォーマンスを実現する可能性を持っていると結論する。

6.2 おわりに

遠隔会議において活発な議論と円滑な会議進行を実施するためには、音声・映像等の低遅延伝送による円滑な会話の実現と、アイコンタクトによるスムーズな話者交代の実現という二つの技術的課題を解決することが重要である。その解決を目指した研究開発はこれまでの数多く行われてきているが、2 地点間という特定環境での実現やバーチャルリアリティ技術の利用等特殊な伝送・表示装置による実現という例が多い。

本研究では、これに対し、システムの汎用化と上述の技術的課題の本質部分の抽出およびそれに対する簡便な手法での解決を試みた。音声・映像等の低遅延伝送による円滑な会話の実現に関しては、ネットワーク転送中の複数のストリーミング情報に対する音声ミキシング等の編集加工処理時間の短縮が課題の本質部分である。よって、ネットワーク転送中の複数のデジタル情報をアナログ情報への変換・アナログ編集加工処理・デジタル情報への再変換といった複雑な処理を行うのではなく、ネットワーク転送中のデジタル情報そのものに対する簡便な演算処理だけでデジタル編集加工処理を実現し処理時間の短縮を図ることにした。一方、アイコンタクトによるスムーズな話者交代の実現に関しては、等身大で表示される人の顔映像が実際に振り向けることが課題の本質部分である。よって、遠隔参加者の顔を表示する液晶モニタと目の役割を果たすカメラが一体となって旋回可能な代理人ロボットと遠隔操作システムを開発すると共に、液晶モニタやカメラの取り付けマウントの標準規格の採用を行った。

そして、上述の二つの研究開発成果を実装した多地点遠隔会議模擬システムを構築し、評価実験を実施した結果、共同実作業をともなわない打合せ等であれば対面会議に近い

パフォーマンスを示すという評価結果を得た。この結果は、多地点遠隔会議での活発な議論と円滑な会議進行を実現するためには、必ずしも特殊な伝送・表示装置の活用が必須ということではなく、汎用機器とシンプルな解決手法の組合せであっても対面会議に近い遠隔会議実現の可能性があることが示している。

今後は、デジタル情報のリアルタイム処理機構を利用した多地点遠隔会議の運用制御や代理人ロボットの遠隔操作に対してSIP(Session Initiation Protocol）やH.281等の国際標準規格を採用もしくは新たな標準規格の開発により汎用性を更に高めていくことを検討する予定である。これにより、近い将来、今回提案したシステムのコンセプトを踏襲した遠隔会議システムの利用が促進されていくと考える。

謝辞

本研究の遂行、ならびに本論文の作成にあたり、懇切なるご指導を賜りました北陸先端科学技術大学院大学情報科学研究科の丹康雄助教授に謹んで感謝の意を表すと共に御礼申し上げます。

本論文をまとめるにあたり、有益なご助言とご教示を賜りました北陸先端科学技術大学院大学　篠田陽一教授、松澤照男教授、日比野靖教授、独立行政法人メディア教育開発センター　仁科エミ助教授に深く御礼申し上げます。

本研究の遂行にあたり、有益なご助言とご教示を賜りました独立行政法人情報通信研究機構情報通信部門インターネットアプリケーショングループ　勝本道哲グループリーダーに深く御礼申し上げます。

本研究全般にわたり、ご配慮を賜りました独立行政法人情報通信研究機構　塩見正理事、福地総合企画部長、同機構情報通信部門　松島裕一部門長、久保田文人研究主管、中川晋一主任研究員に御礼申し上げます。

本研究の遂行にあたり、ご支援頂きました北陸先端科学技術大学院大学情報科学研究科の丹研究室の皆様、文部科学省「IT プログラム」VizGrid プロジェクトの皆様、独立行政法人情報通信研究機構の皆様に感謝致します。

最後に、日頃より研究活動を支えてくれた、父母　小峯勇・サチ江、義父母　貫薗純一・マツ子、そして家族である小峯千加子・彩加にこの場をかりて感謝します。

参考文献

[1] 伊藤球，檜垣泰彦，小峯隆宏，永岡慶三：ディベーターにとって快適な遠隔ディベート，教育システム情報学会誌，Vol.14，No.3，pp.105-109（1997）.

[2] Komine T.，Hamada G.，Suzuki R.，Tsang E. and Kubota F.，GIBN Multimedia Network Experiments－ATM Satellite Communication Experiments between Japan and Canada－，20th Annual Pacific Telecommunications Conference，U.S.A.（1998）.

[3] Komine，T.，Machizawa，A.，Nakagawa，S.，Kubota，F. and Tan，Y.：Development of “high presence” video communication system -Trial experiment of the Next Generation real-time remote lecture -，Proceedings of The 16th International Conference on Information Networking，Vol.2，4C-4，Korea（2002）.

[4] 栗田孝昭，井合知，北脇信彦：オーディオビジュアル通信における伝搬遅延の影響，電子情報通信学会論文誌B-I，Vol.J76-B-I，No.4，pp.331-339（1993）.

[5] 大西仁，山崎聡，望月要，中村直人 ，結城皖曠：潜在学習を利用した伝送遅延が与える心理的影響の測定，電子情報通信学会技術報告，CQ2003-42（2003）.

[6] 田中健二,近藤喜美夫：大学間衛星ネットワーク（スペース・コラボレーション・システム）の構成，電子情報通信学会論文誌D-I，Vol.J82-D-I，No.4，pp.￥581-588（1999）.

[7] 吉野孝ほか：インターネットを介したパーソナルコンピュータによる遠隔授業支援システムの開発と適用，情報処理学会論文誌，Vol.39，No.10，pp.2788-2801（1998）.

[8] 重野寛,間下直晃,楢原常宣,松下温：講義イベントに着目した XML ベース遠隔教育システム，情報処理学会論文誌，Vol.42，No.9，pp.2319-2327（2001）.

[9] He，A.ほか：Tele-Seminar Room：ギガビットネットワークを利用した多地点リアルタイム双方向通信に基づく遠隔ゼミ支援システム，情報処理学会論文誌，Vol.43，No.2，pp.647-660（2002）.

[10] 若原俊彦: ATM-PVC 網を利用した遠隔会議システムの構成と特性, 電子情報通信学会論文誌 B-I, Vol.J81-B-I, No.8, pp.494-506 (1998).

[11] 香川修見,片山薫,上林弥彦: 遠隔教育システムにおける回答支援のための質問選択機構, 電子情報通信学会論文誌 D-I,I Vol.J80-D-II, No.7, pp.1878-1886 (1997).

[12] 前田香織,相原玲二,大槻説乎: 遠隔講義のためのマルチメディア教材提示, 情報処理学会論文誌, Vol.40, No.1, pp.161-167 (1999).

[13] He, A.ほか: RIDEE-SPS : リアルタイム双方向遠隔教育環境のプレゼンテーションシステム, 情報処理学会論文誌, Vol.44, No.3, pp.700-708 (2003).

[14] 白戸仁博,佐々木整,竹谷誠: バーチャルリアリティ技術を用いた遠隔教育システムの開発と適用, 電子情報通信学会論文誌 D-I, Vol.J83-D-I, No.6, pp.619-626 (2000).

[15] 伊藤京子,神月匡規,石井裕剛,吉川榮和: キャラクタエージェントをアバタとナビゲータとして利用したネットワークコミュニティの実験, 情報処理学会論文誌, Vol.44, No.7, pp.1812--1827 (2003).

[16] 小木哲朗,山田俊郎,玉川憲,廣瀬通孝: 共有没空間におけるビデオアバタを用いた位置関係表現, 電子情報通信学会論文誌 D-I, Vol.J82-D-II, No.10, pp.1826-1834 (1999).

[17] 福井健太郎,喜多野美鈴,岡田謙一: 仮想空間を使った多地点遠隔会議システム : e-MulCS, 情報処理学会論文誌, Vol.43, No.11, pp.3375-3384 (2002).

[18] IEC 61834: Helical-scan digital video cassette recording system using 6,35 mm magnetic tape for consumer use (525-60, 625-50, 1125-60 and 1250-50 systems), International Electrotechnical Commission.

[19] 久保田幸雄: 図解デジタルビデオ読本, オーム社 (1995).

[20] Ogawa, A., Kobayashi, K., Sugiura, K., Nakamura, O. and Murai, J.: Design and implementation of DV based video over RTP, Packet Video 2000, pp.140-146 (2000).

[21] Sugiura, K., Sakurada, T. and Ogawa, A.: Demonstration of high quality media transport system using Internet, IPSJ Journal, Vol.41, No.12, pp.1321-1326 (2000).

[22] Digital Video Transport System. http://www.sfc.wide.ad.jp/DVTS/

[23] Komine, T., Tsuchiya, T., Katsumoto, M. and Tan, Y.: Development of the real-time processing proxy node for multimedia applications on interactive telecommunications, Proceedings of The International Symposium on Towards Peta-Bit Ultra-networks, pp.156-161 (2003).

[24] 町澤朗彦ほか: GbE によるネイティブ IPv6 広域実験網の設計と構築, マルチメディア,分散,協調とモバイルシンポジウム, pp.85-90 (2001).

[25] 藤原洋: ポイント図解式最新 MPEG 教科書, アスキー(1994).

[26] 大久保榮,川島正久: 要点チェック式 H.323/MPEG-4 教科書, IE インスティテュート (2001).

[27] 市原清志: バイオサイエンスの統計学, 南江堂 (1990).

[28] 上坂純一,葛岡英明,小山慎哉,山崎敬一: 遠隔作業指示支援ロボットの操作インタフェースがロボットの志向表現に与える影響の研究, 情報処理学会論文誌, Vol.45, No.1, pp.168--177 (2004).

[29] 小山慎哉,葛岡英明,上坂純一,山崎敬一: 実空間を対象とした遠隔コミュニケーション支援システムの設計要件提案と開発, 情報処理学会論文誌, Vol.45, No.1, pp.178-187 (2004).

[30] 郷健太郎,伊藤雅広,今宮淳美: ズーム情報を利用した適応型遠隔カメラ制御法, 情報処理学会論文誌, Vol.43, No.2, pp.585-592 (2002).

[31] 村田和義ほか: 遅延時間の影響を軽減する遠隔カメラ操作インタフェース, ヒューマンインタフェース学会論文誌, Vol.5, No.1, pp.123-133 (2003).

[32] Reeves, B., Nass, C. : The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places, The University of Chicago Press (1996).

[33] Japan Gigabit Network WWW page. http://www.jgn.nict.go.jp/index.html

[34] 渡辺洋: 心理統計の技法, 福村出版(2002).

[35] 酒井良幸: Excel で学ぶ統計解析, ナツメ社(2003).

[36] 大野健彦: 視線を用いた高速なメニュー選択作業, 情報処理学会論文誌, Vol.40, No.2, pp.602-612 (1999).

[37] 塚田浩二,安村通晃:Ubi-Finger:モバイル指向ジェスチャ入力デバイスの研究, 情報処理学会論文誌, Vol.43, No.12, pp.3675-3684 (2002).

[38] 大和正武, 門田暁人, 松本健一, 井上克郎, 鳥居宏次: 一般的な GUI に適した視線・マウス併用型ターゲット選択方式, 情報処理学会論文誌, Vol.42, No.6, pp.1320-1329 (2001).

[39] 黒須正明: ユーザビリティテスティング, 共立出版(2003).

[40] 海保博之, 原田悦子: プロトコル分析入門, 新曜社(2001).

[41] クウォンティン フレミング, ジョエル コッペルマン: アーンド・バリューによるプロジェクトマネジメント, 日本能率協会マネジメントセンター(2004).

付録Ａ

音響エコーキャンセリングに関するサーベイ

A.1　概要

人と人とのコミュニケーションにおける情報伝達手段は音声を介した言語が中心である。遠隔地にいる人とコミュニケーションを行う場合も同様であり、遠隔地の相手に自分の発した音声を正確に伝達することが重要である。遠隔地にいる人とのコミュニケーションを実現する遠隔会議システム（以後、遠隔コミュニケーションシステム）は、通常、マイクロフォンとスピーカを用いた拡声通話系を用いており、多人数が一同に会して会話できる、通話をしながら両手が自由に使える、円滑な会話が実現しやすい等の利点がある。しかし、その反面音響エコーが発生するという問題点が生じる。音響エコーとは、遠端話者の音声が通信回線および近端話者側の音響エコー経路（スピーカ→室内音響伝播経路→マイクロフォンという経路）を経由して遠端話者側に戻り、こだまのように知覚される現象である。音響エコーが発生すると、話者は自分の声を少し遅れて自分で聞くことになるため大変話しづらくなり、通話品質が大きく劣化することになる。また、音響エコーが回線の両端で発生すると、通信回線を介して音響エコーの閉ループが形成され、ループ利得が１を超え位相も一致している場合には発振現象（ハウリング）も発生し通話不能状態になる[a1]。音響エコーキャンセリングはこのような問題を解決する技術であり、過去に様々な研究・開発が行われている。

音響エコーキャンセリングは、音響エコー経路のインパルス応答を推定して、これに基づいて擬似音響エコーを作成し、マイクロフォンに入った信号からこの擬似音響エコーを差し引くことで音響エコーを消去するというものである。音響エコー経路のインパルス応答は、マイクロフォンの位置の変化や人の移動、ドアの開閉等で容易に特性が変化するため、エコーキャンセリングを常に正確に機能させるためにはこのインパルス応答の特性変化に追随しなければならない。これを実現するためには、変化する前に求めたインパルス応答を、変化により生じた誤差が最小になるように逐次修正して、新たに正確なインパルス応答を得るという、適応フィルタの考え方が通常用いられる[a2]-[a5]。このため音響エコーキャンセリングには、常に正確なインパルス応答を推定し続けられる収束特性のよい適応アルゴリズムが必要になる。

２章では、これまで多くの研究者によって研究されてきた様々な適応アルゴリズムを大きく４つに分類し、それぞれの特徴についてまとめることにする。3 章では、現状での音響エコーキャンセラへの適用について述べる。

A.2　適応アルゴリズムの分類と特徴

これまでに提案されている様々な適応アルゴリズムの基本となる構造は、以下の４つに分類される。

・LMS アルゴリズム（Least Mean Square Algorithm）

・学習同定法（Learning Identification Method）

・射影アルゴリズム（Projection Algorithm）

・RLS アルゴリズム（Recursive Least Square Algorithm）

以下では、この４つの適応アルゴリズムの特徴について述べる。表 a.1 に各適応アルゴリズムの特徴をまとめている。

A.2.1　LMS アルゴリズム（Least Mean Square Algorithm）[a6]-[a9]

LMS アルゴリズムは、1960 年に Widrow と Hoff によって最初に導かれたもので、伝統的に Stochastic gradient algorithm と呼ばれることもある。LMS アルゴリズムは、広い意味で、二乗平均誤差を最急降下法に基づいて最小にする方式で、安定性があり、演算量が少ないという利点をもち、現在でも代表的な適応アルゴリズムとしての地位を占めている。しかし、入力信号の大きさが変化する場合は収束速度が低下し、人の音声のような有色性を持つ非定常な信号（以下、有色信号）入力の場合には収束速度が著しく低下するという欠点を持っている。

A.2.2　学習同定法（Learning Identification Method）[a10]-[a12]

学習同定法は、LMS アルゴリズムに比べてやや複雑で演算量が多くなるが、収束速度が入力信号の大きさに依存しないという特徴を持っている。学習同定法は、入力信号が学習同定法は、別名 NLMS アルゴリズム（Normalized Least Mean Square Algorithm）とも呼ばれており、LMS アルゴリズムの係数修正項を適応フィルタの状態ベクトルノルムで正規化した形となっている。このため、学習同定法も、有色信号入力の場合には収束速度が著しく低下するという欠点を持っている。

本グループに分類できるアルゴリズムとしては他に、室内における音波の伝播の特徴と音声の性質に着目することにより収束速度を向上させる ES アルゴリズム（Exponentially weighted Stepsize Algorithm）も挙げられる。

A.2.3　射影アルゴリズム（Projection Algorithm）[a13]-[a17]

射影アルゴリズムは別名、アフィン射影アルゴリズム（Affine Projection Algorithm）とも呼ばれており、適応フィルタのタップ数（用語解説参照）を L として、最適フィルタの過去ｐ点（ただし L 以下）の入出力関係を利用することにより高速に適応を行う。このため、一般的に次数ｐが大きくなるほど（特に有色信号入力に対しての）収束速度

が向上するが演算量も増加する。実時間処理を行うシステムでは、p=2 とする応用例が多い。なお、学習同定法は次数ｐ=1 の射影アルゴリズム、このあと述べる RLS アルゴリズムは過去全点 p=L に対する射影アルゴリズムと解釈でき、射影アルゴリズムは学習同定法と RLS アルゴリズムの中間的な性能をもつといえる。

本グループに分類できるアルゴリズムとしては他に、ES アルゴリズムと射影アルゴリズムを融合した ES 射影アルゴリズムも開発されており、学習同定法とほぼ同じ演算量で、収束速度が約 4 倍を実現している。

A.2.4 RLS アルゴリズム (Recursive Least Square Algorithm) [a18][a19]

RLS アルゴリズムは、適応過程で必要とされる逆行列計算を逐次的に行い二乗誤差を最小にする方式で、有色信号入力の場合、他のアルゴリズムに比べて収束速度の劣化が少ないという特徴を持っている。

しかし、他のアルゴリズムの乗算回数が L に比例するだけであるのに対して、RLS アルゴリズムは L の 2 乗に比例して増加するという欠点を持っている。そのため、演算量の削減を図った高速 RLS アルゴリズムも検討されているがそれでも 7L 以上の倍精度浮動小数点の積和演算が必要であり、かつ、安定性もよくないという問題がある。また、インパルス応答が時間的に変化する場合は収束速度が学習同定法程度まで低下してしまう。

本グループに分類できるアルゴリズムとしては他に、ES アルゴリズムと RLS アルゴリズムを融合した ES-RLS アルゴリズムが開発されており、演算量を考えなければ収束速度の最も高いアルゴリズムである。

A.３．音響エコーキャンセラへの適用

図 a.1 に各適応アルゴリズムの性能をまとめた。横軸は、学習同定法の収束速度（有色信号入力時）を 1 とした場合の相対値を示している。収束速度は高いということはエコーを消去するまでにかかる時間が短く、高速追随性を有していることになる。縦軸は、適応フィルタのタップ数 L での演算量を示している。演算量が多いと実時間処理をするためのコストが高くなるため、当然演算量が少ないほうが実用的ということになる。図１で示したように、高速追随性と低演算量を兼ね備えている ES-射影アルゴリズムを音響エコーキャンセラに適用するのが、現状では最もよい解であるといえる。

音響エコーキャンセラにおける入力信号が遠端話者の音声という有色性をもつ非定常な信号であるため、その音声信号を一般のスピーカで出力すると近端話者側の音響エコー経路に非線形性が生じてエコーが十分消去されないケースが考えられる。特に入力音声信号の大きさが極端に小さかったり大きかったりした場合は、高い非線形性を生じる[a20][a21]。

このため、入力信号の大きさを非線形性が生じない範囲におさめる工夫が必要であり、

オートゲインコントロール（AGC）を行う装置を音響エコーキャンセラの手前で用いることで、入力信号の大きさを適正な範囲内に自動調整することが一般に行われている。

＜用語解説＞

タップ数：エコーキャンセラの規模を表す１つの尺度である。エコーキャンセラは通常FIR（Finite Impulse Response）型ディジタルフィルタで構成されており、タップ数はフィルタ係数（乗算器）の数に相当し、タップ数の多いものは長いエコーに対しても消去可能になるが、ハードウェアの規模も大きくなる。音響エコーキャンセラでは 800 から 2000 程度の値が一般的である。

表 a.1　各適応アルゴリズムの特徴

アルゴリズム	特徴	演算量（L：適応フィルタのタップ数）
LMS	○演算量が少ない ×入力信号の大きさが変化→　収束速度が低下 ×有色信号入力の場合→　収束速度が著しく低下	L
学習同定法	○収束速度が入力信号の大きさに依存しない ×有色信号入力の場合→　収束速度が著しく低下	2×L
射影	○有色信号入力の場合→　LMS および学習同定法より収束速度が向上する	2×p×L （p：次数）
RLS	○収束速度が最も高い ×インパルス応答が時間的に変化する場合→収束速度が学習同定法程度に低下 ×演算量が膨大	（2×L）の 2 乗

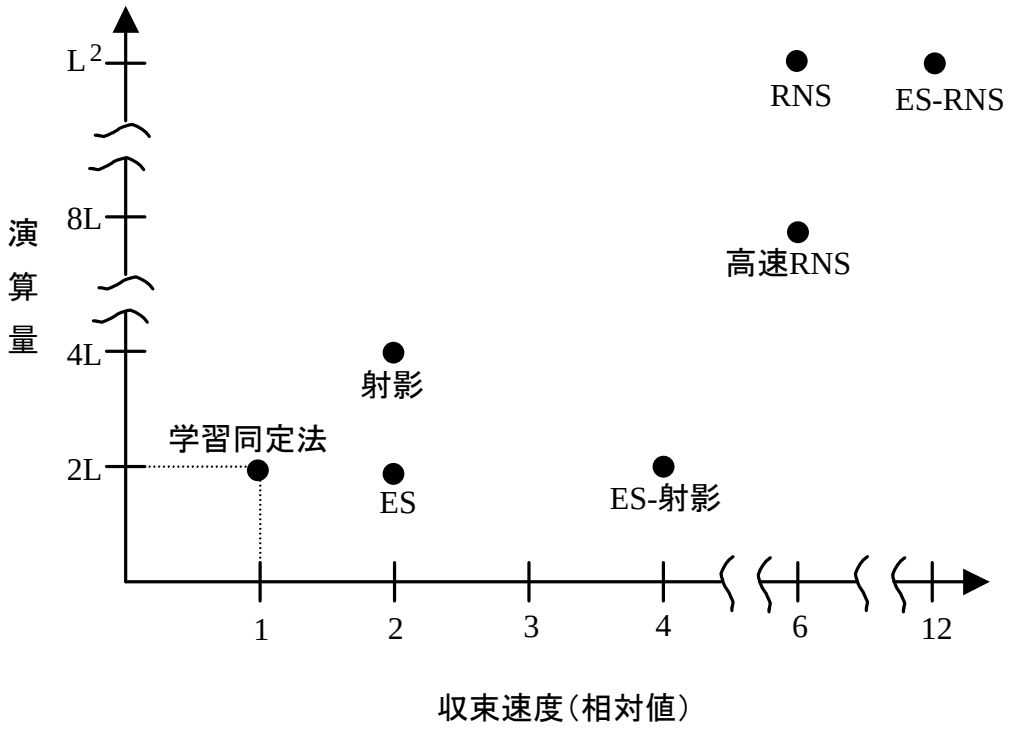


図 a.1　各適応アルゴリズムの性能

付録 A 参考文献

[a1] 大賀寿郎, ハウリング, 音響学会誌, Vol.56, No.2, pp.115-120 (2000).

[a2] 金田豊, 適応フィルタの概要, 音響学会誌, Vol.48, No.7, pp.489-492 (1992).

[a3] 古川博基, 適応フィルタの実用技術(3) - 音響エコーキャンセラへの応用 - , 音響学会誌, Vol.48, No.7, pp.513-516 (1992).

[a4] 山森和彦, 情報流通におけるエコーキャンセラ技術の重要性とその展望, NTT R&D, Vol.50, No.4, pp.226-230 (2001).

[a5] 北脇信彦,小島順治, シームレスな音響環境の実現を目指して, NTT R&D, Vol.44, No.1, pp.33-38 (1995).

[a6] Brenard Widrow, John M. Mccool, Michael G. Larimore, C. Richard Johnson, Stationary and Nonstationary Learnig Characteristics of the LMS Adaptive Filter, Proc. IEEE, Vol.64, No.8, pp.1151-1161 (1976).

[a7] 浜田晴夫, アダプティブフィルタの基礎 (その 1) , 音響学会誌, Vol.45, No.8, pp.624-630 (1989).

[a8] 浜田晴夫, アダプティブフィルタの基礎 (その 2) , 音響学会誌, Vol.45, No.9, pp.731-738 (1989).

[a9] Harris, R., Chabries, D. and Bishop, F., A Variable Step (VS) Adaptive Filter Algorithm, IEEE Trans. Accoust., Speech, Signal Processing, Vol.ASSP-34, pp.309-315 (1986).

[a10] Nagumo, J. and Noda, A., A Learning Method for Syetem Identification, IEEE Trans. Automatic Control, Vol.AC-12, No.3, pp.282-287 (1967).

[a11] 武藤憲司, 柴山秀雄, 島田一雄, 鈴木陽一, 曽根敏夫, ファジィ制御を用いた学習同定法のステップゲイン修正法, 音響学会誌, Vol.53, No.12, pp.941-948 (1997).

[a12] Makino, S., Kaneda, Y. and Koizumi, N., Exponentially Weighted Stepsize NMLS Adaptive Filter Based on the Statistics of a Room Impulse Response, IEEE Trans. Speech, Audio Processing, Vol.1, No.1, pp.101-108 (1993).

[a13] 尾関和彦, 梅田哲夫, アフィン部分空間への直交射影を用いた適応フィルタ・アルゴリズムとその性質, 信学論(A), Vol.J67-A, No.2, pp.126-132 (1984).

[a14] 辻川剛範, 梶川嘉延, 野村康雄, 音響エコーキャンセラにおける射影法の有効性について, 信学技報, DSP99-128, pp.119-126 (1999).

[a15] 田中雅史, 山森和彦, 高速アフィンブロック射影法, NTT R&D, No.50, No.4, pp.231-239 (2001).

[a16] 小島順治, 牧野昭二, 羽田陽一, 島内末廣, 高性能音響エコーキャンセラの開発, NTT R&D, Vol.44, No.1, pp.39-44 (1995).

[a17] 牧野昭二, 金田豊, 音響エコーキャンセラ用 ES 射影アルゴリズム, NTT R&D, Vol.44, No.1, pp.45-52 (1995).

[a18] 酒井英昭, 最近の適応アルゴリズムの動向 -RLS 法を中心として-, 音響学会誌, Vol.48, No.7, pp.493-500 (1992).

[a19] 牧野昭二, 金田豊, 音響エコー経路の変動特性を反映させた RLS 適応アルゴリズム, 音響学会誌, Vol.50, No.1, pp.32-39 (1994).

[a20] 梶川嘉延, 適応 Volterra フィルタの現状と展望, 信学論(A), Vol.J82-A, No.6, pp.759-768 (1999).

[a21] 蘆原郁, 桐生昭吾, 周波数掃引帯域除去フィルタと同期加算法を用いたスピーカの非線形歪測定法, 音響学会誌, Vol.56, No.2, pp.69-77 (2000).

付録Ｂ

DV での音声情報配置に関するサーベイ

B.1　概要

2 つ以上の DV 情報に対してデジタル信号状態のまま音声ミキシング等の処理を行う DV リアルタイム処理機構のプログラム開発を開始するにあたり、DV での音声情報がどのように配置されているのかについて実測調査と文献調査を行い、その結果を以下に示す。

B.2　DV での音声情報の配置形態

DV の規格は現在、民生用 DV 機器用に開発された民生用 DV、圧縮方式等は民生用 DV と同等だが音声が映像に同期する機能をもつ SONY が業務用に開発した DVCAM、同じく PANASONIC が業務用に開発した DVCPRO、の 3 種類が存在するが、基本的な規格は共通であり本サーベイではこの共通部分の規格については DV 規格と記述することにする。

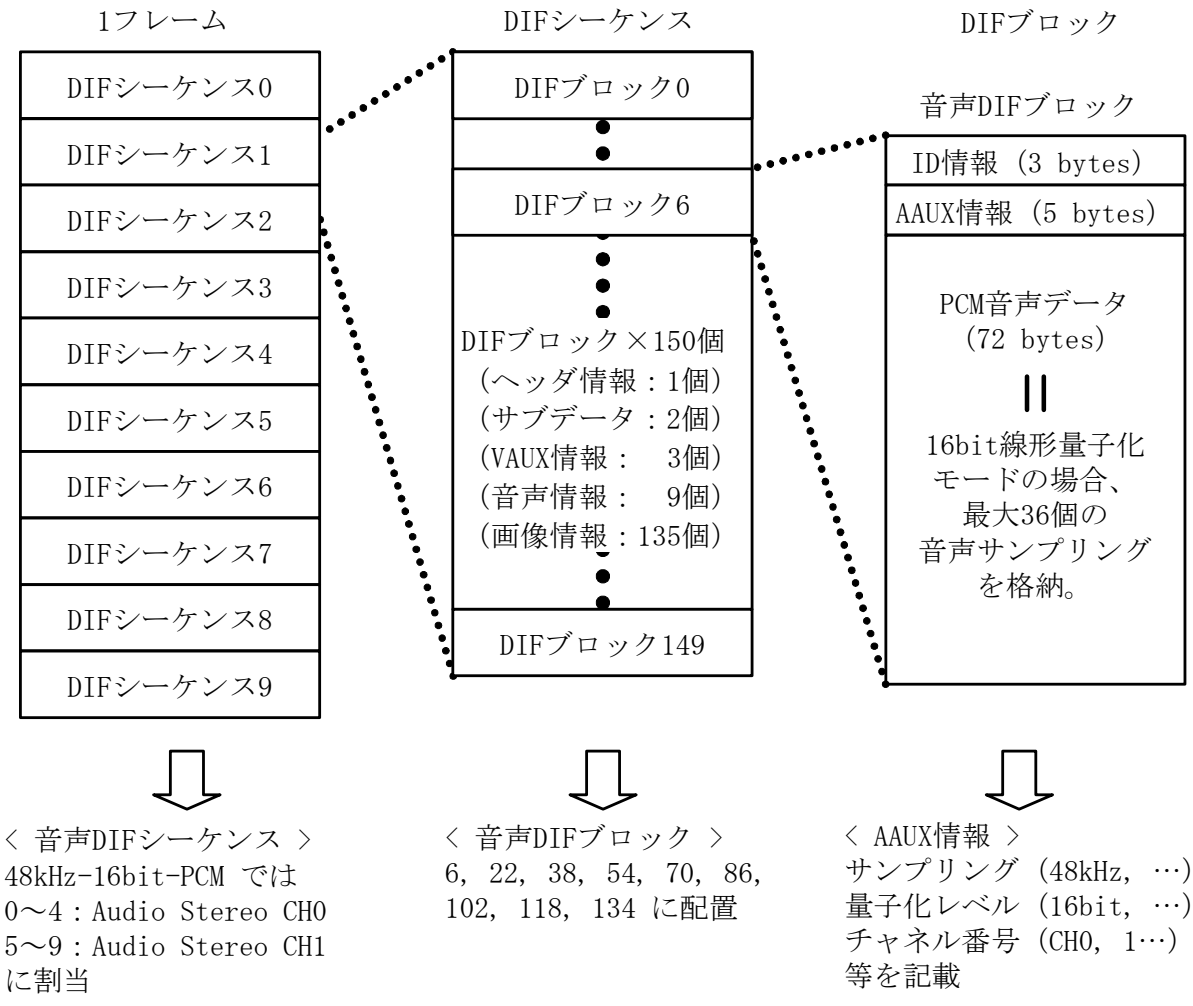


図 b.1　DV 規格での音声情報の配置構成

図 b.1 図に DV 規格での音声情報の配置構成を示す。DV 規格では、1 フレーム分つまり 1/30 秒分の音声情報が最小単位として扱われ、各 DIF シーケンスの内の所定の DIF ブロックに PCM 音声データが格納される。音声 DIF ブロックには、ID 情報と PCM 音声データ以外に、音声符号化の際のパラメータとなるサンプリング周波数、量子化レベルや音声チャネル数等の音声属性情報が格納される AAUX（Audio Auxiliary）情報も格納される。また、DV リアルタイム処理機構で扱う予定の 48kHz サンプリング 16bit 線形量子化 PCM 符号化音声データの場合は、1 フレーム分の DIF シーケンス 0～4 内にステレオ音声のチャネル０のデータが格納され、DIF シーケンス 5～9 内にチャネル１のデータが格納されるため、リアルタイムでの音声ミキシングを行うためには 1 フレーム分の音声情報バッファが必要であることが分かる。

B.3　音声 DIF ブロック内の音声データ格納数について

まず、音声 DIF ブロック内の PCM 音声データサンプル数の情報は AAUX 情報内に記述される仕様になっているため､実際の DV 情報をキャプチャし実測解析を行った｡図 b.2 は、民生用 DV での AAUX 情報部の値の実測結果の抜粋である。DV 規格の仕様と比較解析した結果、図 b.2 中の□枠部分の下位 6bits が、現 1 フレーム分の DV 情報に格納されている 1 チャネルあたりの PCM 音声データのサンプル総数を示す「AF Size」に該当し、本実測では「011001」または「011010」のどちらかの値をとり、フレーム単位で値が変動する場合があることがわかった。しかし、300 フレーム分を解析した結果、周期的な規則は見つからなかった。

次に、「AF Size」について考察する。図 b.3 は、NTSC 方式の場合の「AF Size」と 1 フレームあたりの PCM 音声データサンプル数の関係を示したものである。前述の実測結果は 48kHz サンプリング環境での実測であり､1 フレームあたりの PCM 音声データサンプル数は 1601 個または 1602 個のどちらかであることが分かる。この「AF Size」について更にサーベイした結果、DVCAM と DVCPRO の場合は、「Locked Audio」という音声情報クロックが映像情報クロックに同期する方式をとっており、「AF Size」は一定値をとるように調整する機構が組み込まれているが、民生用 DV では複雑な同機構を省き「Unlocked Audio」方式をとっており「AF Size」は上述のとおり周期的な規則がない可変値を許していることが判明した。

音声情報クロックと映像情報クロックがずれてしまう原因は、NTSC の映像フレームレート 29.97 と DV のフレームレート 30.0 の違いから発生する。以下 8kHz－16bit 線形量子化－PCM 符号化音声データの場合を例に説明する。MTSC 映像の 1 秒間のフレーム数は 29.97 で、対応する PCM 音声データサンプル数は 1 秒間あたり 48000 個であるため映像 1 フレーム分の音声データサンプル数は 1601.60･･･　という非整数値になってしまい、結

果フレーム毎に実際に格納される音声データサンプル数が変動してしまうことになる。これは、前述の実測結果においても、30 フレーム毎の音声データサンプル数は 48048±1 という値をとっており 1 フレームあたり平均 1601.6 という統計結果と一致する。

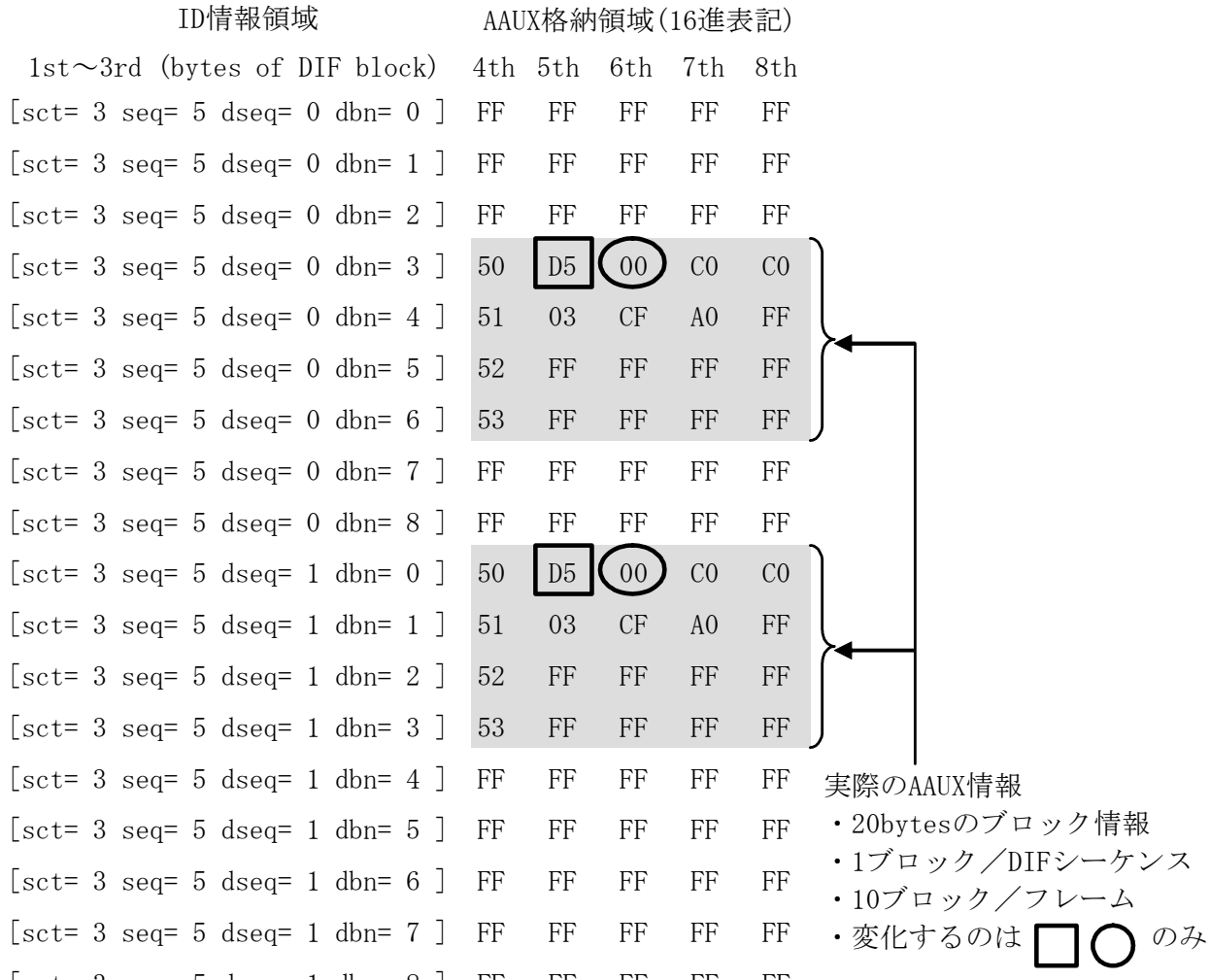


図 b.2　民生用 DV での AAUX 情報の実測結果（抜粋）

これまでの調査結果を踏まえて何種類かの DV 機器での「AF Size」と実際の音声データサンプル数の関係を実測解析した結果を図 b.4 に示す。機種毎に細かな違いは見られるものの「Locked Audio」に対応する機種では「AF Size」の値が 5 フレーム単位で一定値をとり、5 フレーム毎の音声データサンプル数が 8008 という固定値をとることが分かる。

この結果、複数の DV 情報から PCM 音声データサンプルをそれぞれ取り出し加算演算することで音声ミキシングを行う DV リアルタイム処理機構を安定的に機能させるためには、「Locked Audio」に対応する DV 機器を用いることが必要であることがわかった。

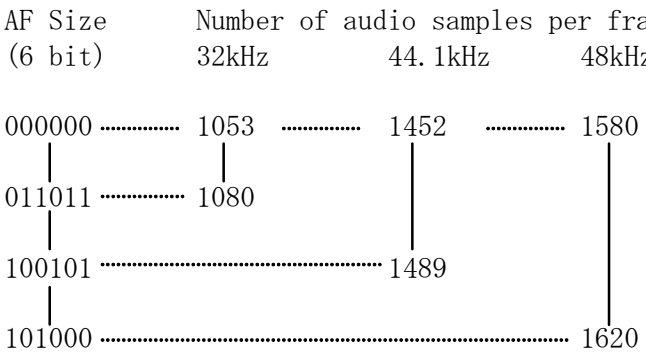


図 b.3　「AF Size」と PCM 音声データサンプル数の関係

	SONY DSR-11 (DVCAM記録テープ再生)	Canopus ADVC500		SONY DVMC-MS1
		(DVCAM / DV) Locked Audio	(DV) Unlocked Audio	
AF_Sizeの値 1フレーム当りの音声データ格納数に相当 in AAUX情報	1602 1602 1602 1602 1600 繰り返し	1602 1602 1602 1602 1600 繰り返し	1601 1602 1602 1601 1602 繰り返し	1602 / 1601 のどちらか (周期性なし)
データ格納エリアにある有効と思われる音声データ数	AF_Sizeの値と同数のデータが格納。 残りのエリアはNULL。	AF_Sizeの値と同数のデータが格納。 残りのエリアにも非固定値が格納される	AF_Sizeの値と同数のデータが格納。 残りのエリアにも非固定値が格納される	AF_Sizeの値と同数のデータが格納。 残りのエリアはNULL。
5フレーム毎の音声データ数	8008 (一定)	8008 (一定)	8008 (一定)	変動
測定条件	・全て48kHz-16bit-線形量子化音声モードで解析。 ・全ての入力ソースはDSR-11のテープ再生ビデオ。			

図 b.4　機種毎の音声データサンプル数の実測結果

研究業績

査読付き論文

1. 小峯隆宏, 勝本道哲, 丹康雄:
   多地点遠隔講義で自然なコミュニケーションを実現する
   DV リアルタイム処理機構の開発,
   情報処理学会論文誌, Vol.46, No.2, （採録済）.

2. 小峯隆宏, 勝本道哲, 丹康雄:
   遠隔講義での代理人ロボット活用による空間共有実現への試み,
   電気学会論文誌 C 分冊, (投稿中).

査読付き国際会議発表論文

1. Komine T., Kubota F., Nakagawa S. and Amemiya A.:
   Development of Japan Gigabit Network,
   Proceedings of Joint IEEE ATM Workshop 2000 and 3rd International Conference on ATM, pp.543-548, Germany (2000).

2. Komine, T., Machizawa, A., Nakagawa, S., Kubota, F. and Tan, Y.:
   Development of ”high presence” video communication system
   -Trial experiment of the Next Generation real-time remote lecture -,
   Proceedings of The 16th International Conference
   on Information Networking, Vol.2, 4C-4, Korea (2002).

3. Komine, T., Tsuchiya, T., Katsumoto, M. and Tan, Y.:
   Development of the real-time processing proxy node for multimedia applications on interactive telecommunications,
   Proceedings of The International Symposium on Towards Peta-Bit Ultra-networks, pp.156-161, Japan (2003).

学会発表

1. 小峯隆宏, 町澤朗彦, 中川晋一, 久保田文人, 丹康雄:
次世代リアルタイム遠隔講義システムの実証実験,
情報処理学会マルチメディア通信と分散処理研究会, DPS 107-25, pp.145-149 (2002).

2. 小峯隆宏, 勝本道哲, 丹康雄:
Proposal for the digital audio processing mechanism
in the real-time multipoint telecommunication,
電子情報通信学会インターネットアーキテクチャー研究会, IA 2002-36,
pp.103-107 (2002).

共著発表論文

1. Tsuchiya, T., Komine, T. and Tan, Y.:
A Network Node for Synchronizing Multiple Streaming Data Based
on Timestamp Information,
Proceedings of The International Symposium on Towards Peta-Bit Ultra-networks,
pp.96-102, Japan (2003).