

広帯域ネットワークを用いた遠隔同時音楽演奏の実験とその考察

茂木俊一*, 一岡義宏*, 田中健二**, 西永望**, 勝本道哲**

*株式会社 大林組

〒108-8502 東京都港区港南 2-15-2 品川インターシティ B 棟

**独立行政法人 通信総合研究所

〒184-8795 東京都小金井市貫井北町 4-2-1

あらまし: 近年、大容量ネットワークの普及により、短時間に高精細な映像を伝送できるだけでなく、TV 会議システム等のコミュニケーションツールを用いて遠隔地を結び、高精細な映像を双方向、多方向に送受信することが可能となってきた。このため、パーソナルな目的への応用にとどまらず、教育、医療等の分野でのリアルタイム遠隔コラボレーションへの応用に期待が高まって来ている。リアルタイム遠隔コラボレーションでは、遠隔地に居ながらあたかも同一空間を共有しているような臨場感を伝えることが大変重要である。しかしながら、ネットワークによるデータ伝送の遅延がその大きな障害となっている。そこで、筆者等は、同時性あるいは同調性が重要視され、その乱れが演奏者の心地良さに顕著に影響を与える音楽合奏に着目し、遠隔での同時音楽演奏実験において画像、音声の遅延が演奏に及ぼす影響を検証することにより、伝送遅延と臨場感の関係を捉える試みを行なった。本稿では、実験手法ならびに検証結果を報告するとともに、ブロードバンド通信時代に向けたリアルタイム遠隔コラボレーションの一手法を考察したので報告する。

A Study of Real-time Ensemble Collaboration over Broad-band Network

Shunichi Mogi*, Yoshihiro Ichioka*,
Kenji Tanaka**, Nozomu Nishinaga**, Michiaki Katsumoto**

*Obayashi Corporation

Inter City B 2-15-2 Konan Minatoku Tokyo 108-8502

**Communication Research Laboratory

4-2-1 Nukuikitamachi Koganeishi Tokyo 184-8795

Abstract: At present, popularization of broad-band network enables to not only transfer high-quality video and image data in short time, but also send and receive them on two-way or multi-way communication with a tool like TV conference system. Expectation of applying real-time network collaboration to educational assistance, medical treatment, as well as personal objectives, is getting higher and higher. It is very important for real-time network collaboration to transfer live performance feeling to share existing atmosphere among the persons in distance. However, network delay has been a great barrier. Therefore, authors focused on ensemble to realize the relationship between network delay and live performance feeling on real-time network collaboration, because players' confusion of simultaneous or synchronous playing is caused by network delay easily affects their comfort. This paper addresses the way of experiment and its result, and the idea of real-time network collaboration toward Broad-band Network Age.

1. はじめに

近年、ブロードバンド通信の普及に伴い、インターネット上で短時間に音声や動画像のデータを、双方向もしくは多方向に伝送したり、遠隔地とリアルタイムにコラボレーションしたりすることも可能となった。今日では、ブロードバンド通信を使った遠隔会議、遠隔医療、遠隔教育、ゲーム対戦、TV 電話などの実験とともに、それらサービスの実用化が進められている。

リアルタイム遠隔コラボレーションでは、遠隔地に

居ながらにして、あたかも同じ空間上にいるかのような臨場感ある空間を造ることが重要である。しかし、ネットワークによるデータ伝送の遅延がその大きな障害となっている。

そこで、筆者らは同時性あるいは同調性が重要視され、その乱れが演奏者の心地良さに顕著に影響を与える音楽合奏に着目した。

以下、2章では、広帯域ネットワークの普及していなかった4年前に、長野オリンピックで行われた5大

陸同時音楽合唱について、当時どのように遠隔で音楽コラボレーションがおこなわれたか、システムの概要を紹介する。

3、4章では、高速広帯域ネットワークを利用した同時音楽演奏実験で、画像、音声の遅延が演奏に及ぼす影響から、伝送遅延と臨場感との関係を捉える試みについて述べる。

5章では、ブロードバンド通信時代に向けたリアルタイム遠隔コラボレーションの手法についての考察を報告する。

2. 遠隔同時音楽演奏の調査

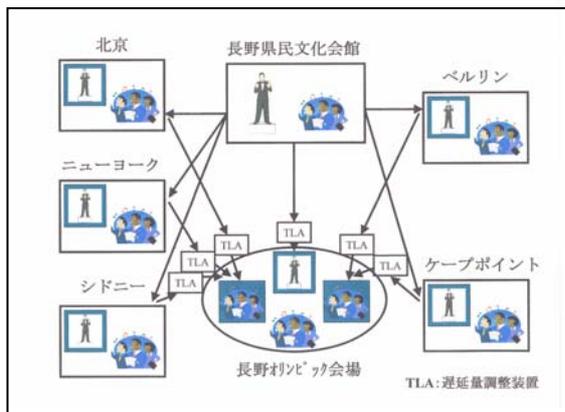


図1. 長野五輪「第九」五大陸中継

1998年2月7日に長野オリンピックの閉会式にて、小澤征爾氏指揮の下、5大陸を結んだ世界大合唱が、長野南運動公園スタジアムのスクリーンに同期して映し出された[1]。

本デモンストレーションのシステム概要を、図1に示し、手順を以下に示す。

1. 小澤氏のいる長野県民文化会館から五大陸へ送信される映像及び音声の伝送時間を、あらかじめ測定する。
2. 小澤氏指揮の映像が5大陸会場に送信される。
3. その映像を見て合唱した映像が、更に長野五輪会場へ送信される。
4. 1であらかじめ測定した伝送時間を用い、各会場からの映像及び音声、同じ遅延時間(約2s)となるように遅延量調整装置を使って調整し、全ての会場の映像及び音声を同期させる。
5. 五輪会場にその同期した映像・音声を放映する。

このイベントでは、五輪会場にいた観客は、同期した映像・音声を見て、あたかも五大陸同時に合唱しているかのような風景を見ることができた。しかし、長野県民文化会館の合唱者は、あくまでもその場にいる小澤氏の指揮に合わせて合唱をするだけであり、また

各大陸の合唱者も、映像に映し出された小澤氏の指揮に合わせて合唱するだけであった。そのため、お互いの合唱を聴きながら自らの合唱をおこなっておらず、双方向での同時合唱はおこなわれていなかった。

筆者らは、この時行われたコラボレーションに比べ、映像及び音声遅延量が短くなるようなシステムを構築することにより、遅延による演奏への影響を小さくし、お互いの演奏を聴きながら、双方向のコラボレーションができるか、その可能性について追求する。

3. MPEG over ATM 遠隔実験

高速通信回線を使った遠隔同時音楽演奏の可能性について検証するために、遠隔2会場で同時に音楽演奏を試みる。まず本実験では、遅延時間測定を測定し、次に同時音楽演奏に対する演奏者の主観評価をおこない、さらに遅延に対する主観評価実験をおこなった[2]。

3.1 ネットワーク構成

図2にネットワーク構成を示す。伝送路としては、通信・放送機構が運営する研究開発用ギガビットネットワークを用い、20MbpsのPVC(Permanent Virtual Connection)をCBR(Constant Bit Rate)にて確保し、けいはんなと京都間を、つくば経由と直接の2種類の方法によって接続した。けいはんな及び京都の地点にはTV会議システムを設置し、環境を準備した。機器の仕様は、表1のとおりである。

表1. 機器仕様

コーデック装置	NTT エレクトロニクス製 RB2000/121B	
符号化モード	映像	MPEG2
	音声	MPEG 1レイヤー-2
DV カメラ	SONY 製 DCR-VX9000	

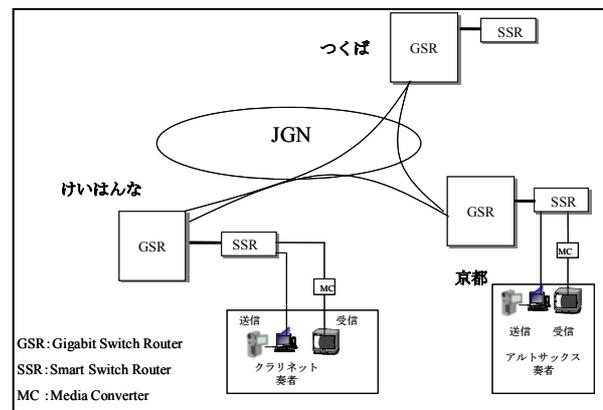


図2. ネットワーク構成

3. 2 実験方法

図2の実験環境において、けいはんなー京都間の音声、映像及びpingの往復伝送遅延時間の測定をおこない、さらにそれぞれの地点に演奏者A、Bを配置し、遠隔同時音楽演奏を行った。

(1) 遅延測定

(ア) 音声遅延測定

けいはんな側にデジタルオシロスコープ及びCDプレーヤーを配置した。CDにて125Hzのトーンバーストを発生させ、デジタルオシロスコープの画面に、直接入力した音と、ネットワークを介して戻ってきた音との波形を同時に表示し、その差分時間を計測することにより、往復音声遅延時間を算出した。

(イ) 映像遅延測定

けいはんな側にVTRを配置し、1秒間隔に1~30の数字が入っている映像評価測定用ビデオテープを再生し、直接表示させた映像と、ネットワークを介して戻ってきた映像とを、TV画面に合わせて表示した。次に、TV画面をデジタルカメラにて撮影し、その映し出された数字の差から、往復映像遅延時間を算出した。

(ウ) pingによる遅延測定

けいはんなから京都にあるPCのIPアドレスに向けてpingを打つことにより、パケットの往復伝送時間を測定した。

バッファサイズの設定を変更し、各往復遅延時間の測定した結果を表2に示す。

表2. 各測定における往復遅延時間(ms)

バッファサイズ	けいはんなー京都		けいはんなー京都つくば経由		
	音声	映像	音声	映像	ping
4kbyte(映像無)	190		360		80
4kbyte	360	500	360	500	80
48kbyte	470	530	590	670	80
64kbyte	520	530	600	670	80

表2から、バッファサイズを4kbyteから64kbyteへと大きくすると、音声遅延時間は、けいはんなー京都間は190ms~520ms、けいはんなー京都つくば経由間では360ms~600msとそれぞれ長くなっていることがわかる。一方映像遅延時間は、バッファサイズが48kbyteと64kbyteにおいて、けいはんなー京都間は530ms、けいはんなー京都つくば経由間では670msとそれぞれ同じとなった。これは、映像の測定分解能が、1/30sであることより、同時間になったものと考えられる。また、同じバッファサイズでも映像と音声で遅延時間が異なるのがわかる。これは、音声はMPEG1、

映像がMPEG2と両者それぞれ異なる符号化方式を採用していることに起因する。

(2) 遠隔同時音楽演奏

けいはんな側にクラリネット奏者A、京都側にアルトサックス奏者Bを分かれて配置する。そして、お互いに相手の音声及び映像を確認できるようにして、以下の確認をおこなった。

(ア) テンポをt=120とし、通常通り演奏をおこなう。

(イ) 一方がテンポをとり、相手がそれに合わせて演奏をおこなう。

(ア)の場合、通常通り演奏を行うと、遅延が大きすぎ、演奏者は混乱してしまった。次に、(イ)の一方の場合、合わせて演奏する演奏者は、問題なく演奏できたが、テンポを取る演奏者は、合わせて演奏する演奏者の音が戻ってきた時、自分の演奏と遅延が生じ、混乱を生じてしまった。

(3) マイクコードによるローカル実験

(2)で計測した最も音声遅延時間の短い190ms以下での演奏者の主観評価をおこなうため、以下のマイクコードによるローカル実験をおこなった。

音が完全に隔絶された2つの近接した部屋に、マイク及びスピーカーをそれぞれ用意し、2本のマイクコード10mで自分の部屋と相手の部屋にあるマイクとスピーカーとを結び、お互いに相手の部屋の音声を聞くことのできる環境を構築した。

次に、演奏者A、Bをそれぞれの部屋に配置し、同時音楽演奏を試み、主観評価をおこない、その時の音声遅延時間をデジタルオシロスコープで測定した。

その結果、演奏者は遅延時間を全く気にすることなくお互いに演奏することができた。またこの時の音声遅延測定時間は8msとなり、理論上の通信伝播遅延時間である $20/(3 \times 10^8)$ msよりも大きくなった。これは、デジタルオシロスコープで波形を読み取った際に、誤差が生じたものと考えられる。

(4) 遅延に対する主観評価実験

(2)及び(3)を踏まえ、片方の演奏者が相手の演奏する遅れた音に対し、どのくらいの遅延時間であれば、許容して演奏することが可能であるかを調べるため、以下の主観評価実験をおこなった。

演奏者A、B2人が背中合わせとなり、お互い離れながら演奏し、その遅延時間の許容できる距離を測定することにより、音速からその許容時間を推測した。

このとき、許容できる距離は往復約20mとなり、音速340m/sから、許容遅延時間は約60msとなった。

音声遅延に対する演奏者の主観評価実験結果を、図3に示す。

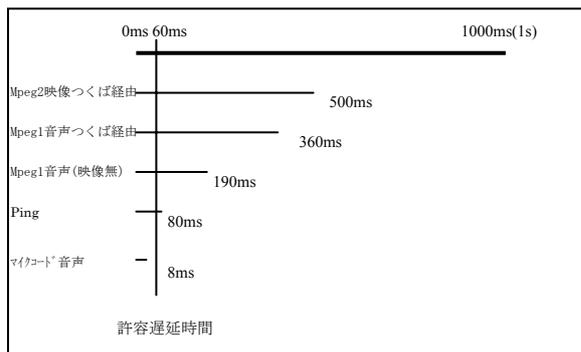


図3. 実験結果

図4から、ping 遅延時間は80msとなり、既に許容遅延時間である60msを超えていることがわかる。伝送遅延時間を除いた機器の処理遅延時間も、映像音声遅延時間の190~500msから80msを除いた110~420msとなり、許容遅延時間を超えている。許容遅延時間を超えていないものは、マイクロコードで接続した場合のみであることがわかる。

ここで行われている処理手順を以下に示す。

1. DVカメラからのNTSC信号をMPEG2データにエンコード
2. CLADにてATMセル化
3. 20MbpsのCBR帯域のパス通過
4. ATMセルの組立
5. MPEG2データをデコード
6. 動画/音声の再生

ネットワーク上の伝送時間を表す3以外の1~6の処理作業で、既に許容遅延時間を超えていることがわかる。この理由として、パケット交換、符号化などによる遅延を吸収するためにデータバッファリングを行うことが必要なため、このバッファリングによってデータ到達に大きな遅延が発生していると思われる。

この実験では、ネットワークにATM網を用いたが、次に、IP網を使い、同様の実験をおこなった。

4. DV over IP 遠隔実験

本実験では、日米間を高速広帯域国際通信回線で結び、UDP/IPを用いたDV(Digital Video)技術[3][4]による接続で、遠隔同時音楽演奏実験をおこない、映像及び音声遅延測定及び、演奏者による主観評価をおこなった。なおプロトコルには、IPv4を用いた。

4. 1 システム構成

図4に、システム構成図を示す。実験網としては、独立行政法人通信総合研究所(小金井市)と米国カリフォルニア州パロアルト市にあるNTTコミュニケーションズMCL(Media Communication Lab)とを結んだ。機器仕様は、表3のとおりである。

表3. 機器仕様

米国側	送	DELL 製	プレジジョン 340
	受	CPU	Pentium4 1.8GHz
	信	MEMORY	512MB PC800チップセット 850
	PC	OS	WindowsXP
	DVカメラ	SONY 製 DCR-VX1000	
日本側	受	EPSON 製	
	信	CPU	Pentium4 2.0GHz
	PC	MEMORY	512MB PC133チップセット 845
	送	OS	WindowsXP
	送	DELL 製	デイモンジョン 4300
日本側	信	CPU	Pentium4 1.7GHz
	PC	MEMORY	512MB PC133チップセット 845
	送	OS	WindowsXP
	DVカメラ	SONY 製 DCR-VX1000	

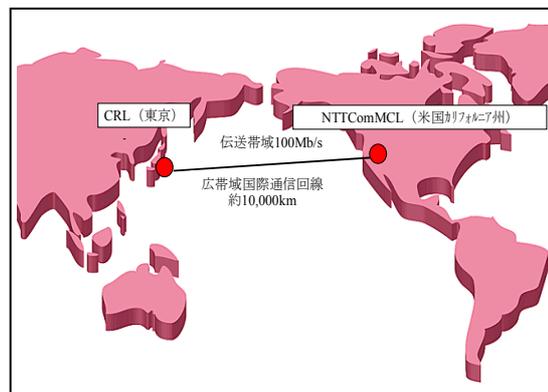


図4. システム構成

4. 2 実験方法

(1) 遅延測定

3章でおこなった実験同様、音声・映像・pingの遅延時間を測定した。

図5に映像遅延測定でのモニター画面例を示す。



図5. 映像遅延測定
(左: ループバック, 右: 送信元)

音声遅延は 465ms、映像遅延は、図 5 より 2つの数字の差から、 $(12+30-21)/30 \times 1000 = 700\text{ms}$ となった。

表 4 に各往復遅延時間を示す。

表 4. 各往復遅延時間 (ms)

ping	音声	映像
108	465	700

ここで行われている処理手順を以下に示す。

1. DV カメラからの NTSC 信号を IEEE1394 にカプセル化した DV パケットストリームに変換し、PC に転送する。
2. DV パケットを UDP にカプセル化する。
3. IP 網を通り、受信 PC へ送信する。
4. UDP にカプセル化された DV パケットを脱カプセル化する。
5. DV パケットから、音声は PC の内部サウンドカードへ出力し、映像は directshow にてデコードした映像を画面出力する。

ネットワーク上の伝送時間を表す 3 以外の 1～5 の処理作業で、既に許容遅延時間を超えていることがわかる。

また、表 4 より、映像と音声の遅延時間に大きな開きがあったが、これは、受信側での脱カプセル化された DV パケットのバッファリング処理が映像と音声で、前者が directshow、後者がサウンドボードと異なっているためと推察できる。

(2) 遠隔同時音楽演奏

CRL 側に、前回実験をおこなったクラリネット奏者 A、MCL 側にアルトサクソ奏者 C を分かれて配置する。

お互いに相手の音声及び映像を確認できるようにして、次の手順で実験を行った。

(ア) 一方がテンポを $t=120$ とし、相手がそれに合わせるようにして演奏を行う。

(イ) 次に $t=90$ とし同様に進行。

(ウ) 片側に指揮者が入り、画像は指揮者を映し、相手側はそれに合わせて演奏する。

(エ) 輪唱をおこなう。方法としては、テンポを取る側に合わせて、同時に相手側と一緒に合唱し、戻ってきた音を聴きながら、テンポを取る側が輪唱をおこなうことができるか試みる。

(ア) の場合、相手のテンポに合わせた演奏者は前回同様問題なく演奏できたが、テンポを取る演奏者は、合わせて演奏した人の音が戻ってきた時、表 4 からわかる通り、音声遅延時間 465ms と前回の実験で測定した許容遅延時間 60ms を大きく上回り、混乱を生じてしまった。

(イ) $t=90$ とすると、(ア) の時に比べて演奏しやすく感じた。

(ウ) の場合、指揮者がいる地点は、指揮者に合わせて演奏をおこない、指揮者がいない地点は、指揮者を映した映像を見ながら、その映像に合わせて演奏を行ったが、両地点とも問題なく演奏できた。また、映像と音声の遅延時間のずれがあったが、殆ど気にせず、相手の音に合わせて演奏することができた。

(エ) の場合、遅延時間と一拍分の長さが、偶然、同時間となり、テンポを取る側は、相手の音がタイミングよく聞こえ、混乱することなく輪唱することができた。

(3) 遅延量調整装置によるローカル実験

演奏した同じ曲を CD にて再生し、遅延量調整装置を介して遅延時間を持たせた音と、CD から直接聴いた遅延のない音とを同時に聴き、どのくらいの遅延時間まで許容して聴くことができるか実験をおこなった。なお演奏者 C を被験者とし、最初のテンポは $t=120$ とした。

その結果、やはり遅延時間が 60ms を超えたあたりから演奏が聴きづらくなり、3 章で述べた実験と同様の結果が得られた。また、テンポを遅くすると許容遅延時間は 60ms より大きくなった。

5. 考察

これらの結果を総合し、以下のことが推察できる。

(1) テンポと演奏者の心地よさとの関係

4 章 (2) の実験結果より、 $t=90$ の時、 $t=120$ に比べて演奏しやすいという結果となった。

これは、 $t=90$ の時に、 $t=120$ と比べ、一拍分の音が長くなることにより、遅延拍数が逆に小さくなり、演奏しやすくなったと考えられる。

また、4 章 (3) の実験結果である、テンポを遅くすると許容遅延時間が大きくなったことを勘案すると、演奏者の心地よさは、絶対遅延時間より、遅延拍数に影響を大きく受けられると思われる。

(2) 映像・音声のずれと演奏スキル

3 章 (2) (ウ) の実験結果より、指揮者の画像を見ながら演奏を行う際、演奏者にとって、演奏上画像が音声より重要視され、画像と音声がずれているにも関わらず演奏できたと考えられる。

また、これは通常コンサート会場でオーケストラが演奏する場合に、舞台上で自分から遠くにいる人の音が遅れて聞こえてくるが、それを聞かないようにして、指揮者を見て演奏している時と同様の演奏方法を演奏者がおこなっていたと考えられる。

(3) リアルタイムコラボレーション手法

今回、遅延時間と、輪唱のタイミングがたまたま合い、片側では同時合唱、もう片方では輪唱形式での遠隔コラボレーションを行うことが確認できた。このことから、遅延を前提とした演奏は可能であり、その一手法として、クラシックのフーガ技法を用いた演奏が考えられる。

具体的には、テンポを合わせる側は、基本旋律を転調させた旋律を、テンポを取る側の音に合わせて同時演奏する。一方テンポを取る側は、転調された相手の演奏を遅延量調整装置にてタイミングを調整し、その音を聴きながら、基本旋律を演奏する。この手法により、心地よさを感じながら、お互いの演奏者は演奏をすることが可能であると考えられる。

また、小節遅れを織り込みながらのケチャ¹、ミニマルミュージック²などの演奏も考えられる。

6. まとめと今後の課題

本論文では、2つの異なるシステムにおいて、広帯域ネットワークを使った遠隔同時演奏実験をおこない、遅延時間の測定および演奏者の主観評価をおこなった。

今回の実験環境では、いずれも遠隔での同時演奏は可能であったが、臨場感をもって同期し、ハーモナイズした音楽演奏をおこなうことはできなかった。

結論として、遠隔同時演奏をおこなうには、今回の実験の場合、往復伝送遅延時間が約 60ms 以下であることが望ましいこと、また、もしくは遅延をポジティブに利用した従来にない、新しい音楽コラボレーションを考える必要があることがわかった。今後は、フーガ技法を用いた演奏をおこない、実際に心地よい演奏ができるか検証をおこないたいと考えている。

今回の測定結果は、あくまでも被験者がクラリネット奏者A及びアルトサクソ奏者B、Cの3人のみのデータであり、今後は様々な音楽ジャンルや演奏者の経験年数などによって、同じ遅延条件下で演奏のし易さがどのように異なるか統計的なデータを取りたいと考えている。またその際、演奏者の主観評価手法についても事前に検討したいと考えている。

DV over IP 実験時に起こった映像と音声の遅延時間のずれについては、今後ローカル実験をおこない、映像と音声に遅延時間の差があるのか検証し、その原因について究明したいと考えている。

¹インドネシア・バリ島の民族音楽。

²1960年代中頃から現れた短い旋律断片やリズムを繰り返すことを基調とした音楽。

謝辞

本研究を進めるにあたり、大林組田北博司氏をはじめ、演奏に協力頂いた晴山大作氏、佐藤和宏氏、通信放送機構けいはんなセンター伊藤正也氏、通信放送機構京都センター梶原聡氏、川勝正洋氏、NTTコミュニケーションズ株式会社 MCL 伊藤正樹氏、宮川晋氏と多くの方々に、貴重な助言を頂きました。ここに感謝を表します。

本研究のため、映像音声遅延量調整装置を試用させて頂いたエレテックス株式会社に、感謝いたします。

最後に、本研究の機会を頂いたメディア教育開発センター永岡慶三教授に、深く感謝の意を表します。

参考文献

[1]山北淳,古川賢信,須賀川豊,渡辺義典: TLA が結んだ5大陸長野オリンピック開会式「五大陸第九合唱」,映像情報メディア学会 1998.5 発表会

[2]松原慶祐: 超高速通信網を用いた遠隔バンドビジョン

<http://www.jaist.ac.jp/~ksuke-m/reserch/subtheme.html>

[3]杉浦一徳、小川晃通、中村修、村井純: 民生用 DV を用いたインターネットビデオ会議システム, 情報処理学会誌, Vol. 40, No.7,1999

[4] 杉浦一徳、小川晃通、中村修、村井純、中川晋一: インターネットにおける TCP 協調型の DV 転送技術, 情報通信学会コミュニケーションクオリティ研究会 2000.2