

顔のエッジ表現を用いたコミュニケーションシステム

大石貴也¹⁾ 徳永幸生¹⁾ 米村俊一²⁾ 大谷淳³⁾

- 1) 芝浦工業大学工学部情報工学科 〒337-8570 埼玉県さいたま市見沼区深作 307
2) 日本電信電話 NTT サイバーソリューション研究所 3) 早稲田大学大学院 国際情報通信研究科
Tel:048-687-5066 E-mail:m103176@sic.shibaura-it.ac.jp

人間同士のコミュニケーションにおいて顔情報は重要である。一方、顔をリアルに表示する TV 電話には抵抗感のあることが種々報告されている。本報告では顔のエッジ表現を用いたコミュニケーションシステムを構築し、コミュニケーションにおける顔情報の役割を探るため、顔のエッジ情報量を変えることで会話特性がどのように変化するか実験考察した。具体的には、顔画像に対して、エッジ抽出処理、2 値化処理、細線化処理を順次施すとともに、画像をサブブロックに分割する。顔のサンプリング率（分割画像をサンプリングする確率）をパラメータとしてエッジを表示する割合を変化させて被験者に呈示し、会話に及ぼす影響について分析・評価した結果を報告する。

Visual Communication System Based on Edge Abstraction

Takaya Oishi¹⁾ Yukio Tokunaga¹⁾ Shunichi Yonemura²⁾ Jun Oya³⁾

- 1) Department of Information Science and Engineering, Faculty of Engineering, Shibaura Institute of Technology 307, Fukasaku, Minuma-ku, Saitama-shi, Saitama, JAPAN 337-8570
2) Cyber Solutions Laboratories, NTT 3) GITS, Waseda University

A facial expression plays an important role in human communication because it represents some states of mind. On the other hand, it is generally known that a feeling of shyness exists in the communication using TV telephone that displays a face with reality. Our study focuses on the functional characteristics of facial expressions, the feeling of shyness in particular, in the communication using TV telephone. This paper addresses three key challenges: developing a visual communication system based on edge abstraction, introducing the edge abstraction system into TV telephone aiming at reducing a feeling of shyness, and examining how the feeling of shyness changes according to the sampling rate of edge abstraction.

1. 研究背景・目的

人間同士のコミュニケーションでは、Face-to-Face の対面対話が重要である。人間の会話では、メッセージのわずか 7%が言語で伝達される一方、55%は顔の表情によって伝達されるという[1]。このように、顔の表情等のノンバーバルコミュニケーションは人間にとって不可欠である。顔情報に代表されるノンバーバルメッセージを遠隔地でやり取り可能なシステムとして、TV 電話システムがある。従来の固定 TV 電話に加え、最近では携帯電話で顔画像を送受できるモバイル TV 電話サービスが登場した。しかし、顔をリアルに表示する TV 電話には抵抗感のあることが種々報告されている[2]。例えば、顔の余計なしわや肌荒れ、会話場所の背景等は相手に見せたくない場合がある。視線を合わせることに不慣れた人には緊張

感や恥ずかしさが生まれる可能性もある。実際、TV 電話を通じての対話は、利用者に違和感や居心地の悪さを与え、好ましくないという実験結果が報告されている[3]。また、アバターを使ったコミュニケーションシステムも提案されている[4]が、喜怒哀楽の顔表情が人工的に作られたものであるため、表情の面での自分らしさが出ないという問題点がある。例えば、話者が同じキャラクターを使うと、顔の表情は誰が使っても同一のものになってしまうので、その人特有の個性的な顔表情を相手に伝えることはできない。

そこで本研究では顔のエッジ表現を用いたコミュニケーションシステムを構築し、コミュニケーションにおける顔情報の役割を探るため、顔のエッジ情報を種々変えた時、会話特性がどのように変化するかを実験・評価した。

2. 顔画像のエッジ抽出アルゴリズム

2.1 エッジ抽出とランダムサンプリング画像

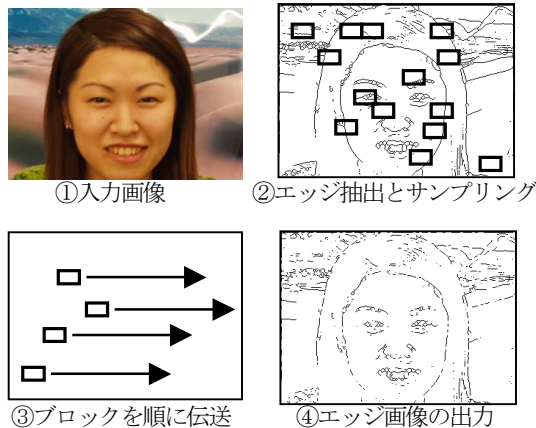


図 2.1 画像入力からエッジ画像の出力までの手順

図 2.1 は画像の入力から、サンプリングされたエッジ情報による顔の表現までの流れを示したもので、顔画像 (①) に対して、エッジ抽出処理, 2 値化処理, 細線化処理を順次施すとともに、画像をサブブロックに分割する。その後、分割したサブブロックをランダムにサンプリング (②) し、サンプリングしたブロックを伝送 (③) する。その際、顔のサンプリング率 (分割画像をサンプリングする確率) をパラメータとしてエッジを表示する割合を変化させて顔画像を表現 (④) する。エッジ情報のサンプリング法に関しては 2.2 節で述べる。

2.2 エッジ情報のサンプリング

TV 電話のように画面に互いの顔を表示して対面対話すると、表示される顔は表情のみならず、顔の位置も絶えず動いていることが分かる。画面全体の中で顔が大半を占めているため、顔の動きがよく見えるからである。絶えず動く顔のエッジをサンプリングし表示すると、ある時刻に顔の一部を構成するエッジであったものが、次の時刻ではその顔の正しい一部とはなっていない。しかし、そのエッジをある一定以上の確率でランダムにサンプリングして表示すると、画像全体の中で顔の表情や動きを十分に把握することが可能である。

本コミュニケーションではこの特性を利用する。ランダムサンプリングの手順は以下の通りである。

1. 入力画像から (上記の手順で) エッジ画像を作る。
2. このエッジで表現された画像を複数のブロック (例えば 80×60) に分割する。
3. 各ブロックの左上の画素を標本点にし、出現確率をパラメータとする random 関数を利用し 1 か 0 を出力する。
4. 標本点が 1 のブロックのみ、相手端末に送信し、その標本点を含むブロックを表示する。

図 2.2 左側に示すようにサンプリングされたブロックデータは、受信側では右側に示すように表示される。



図 2.2 サンプリングしたエッジ情報の出力

3. システムの構築

3.1 DirectX の概要

実験システムでは DirectX を用いた [5] [6]。DirectX は、アプリケーションソフトで直接ハードウェアを制御することが可能なため、高度なマルチメディア処理を PC 上で実現することができる。DirectX の主要な機能のコンポーネントを表 3.1 に示す。

表 3.1 DirectX の主要なコンポーネント

名前	役割
DirectX Graphics	グラフィックスプログラミングのための API を提供 DirectX 7 以前の DirectDraw と DirectX3D を統合
DirectX Audio	サウンドプログラミングのための API を提供 MIDI等のサウンドファイルの再生、キャプチャーが可能
DirectInput	キーボード、マウス等の入力装置に関する API を提供
DirectPlay	ネット上のコンピュータ同士の通信を行うコンポーネント
DirectShow	高品質なマルチメディア再生、キャプチャーが可能
DirectSetup	DirectX コンポーネントのワン・コールインストールを提供

3.2 DirectShow と GraphEdit の概要

本システムは映像と音声をリアルタイムで通信する必要があり、DirectShow [7] を用いて実現した。DirectShow を構成する基本要素は、フィルタと呼ばれるソフトウェアコンポーネントである。フィルタは、通常、マルチメディアストリームに対して単一の操作を実行する。

指定されたタスクを実行するために、アプリケーションでは複数のフィルタを接続し、1つのフィルタからの出力を別のフィルタの入力とする。接続されたフィルタの集合は、フィルタグラフと呼ばれる。Web カメラから入力した映像をウインドウ上に再生するフィルタグラフの例を図 3.1 に示す。なお、フィルタグラフを構築するために、GraphEdit を利用した。



図 3.1 フィルタグラフの例

実験システムは、顔画像のエッジ抽出などを行う画像処理部分を作成し、フィルタとして登録後、フィルタグラフにロードすることで実現した。

3.3 実験システム

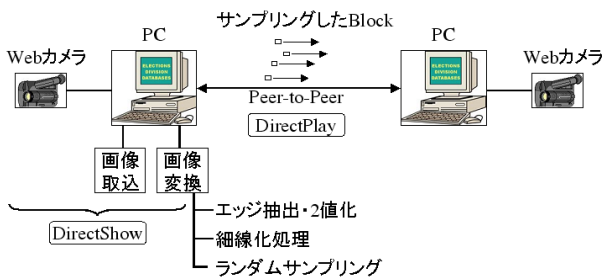


図 3.2 実験システムの概要

図 3.2 は実験システムの概要である。まず、Web カメラから入力された映像を画像データとして取り込む。その映像から 1 フレーム毎に取り出して RGB データにし、各フレームに対してエッジ抽出、細線化、ランダムサンプリング等の画像処理を施す。図 3.2 に示すように、画像を取り込みからランダムサンプリングまでを DirectShow で表現した。次に、ランダムサンプリングしたブロックを Peer-to-Peer でお互いに送受信する。この Peer-to-Peer 通信は、ネットワーク上のコンピュータ同士の通信を行うコンポーネントである DirectPlay を用いて実現した。

実験システムでは、画像取込部と画像変換部を経て、ランダムサンプリングされた顔エッジが PC 上にリアルタイムで表示されるが、この処理を行うフィルタグラフを図 3.3 に示す。

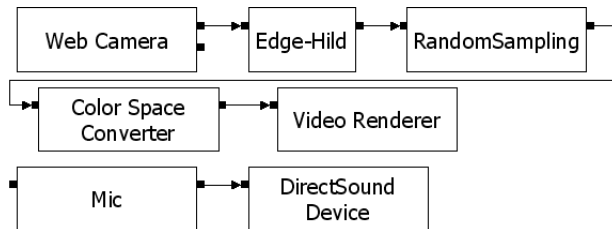


図 3.3 実験システムのフィルタグラフ

まず、Web Camera フィルタにより Web カメラから入力された画像データに対して Edge-Hild フィルタでエッジ抽出処理、2 値化処理、細線化処理 (Hilditch の方法) を施す。その後、RandomSampling フィルタで画像をブロックに分割してランダムサンプリングを行い、Color Space Converter フィルタで適切な色空間の変換を行った後、Video Renderer フィルタでウインドウ上に表示する。同時に、Mic フィルタで Web カメラ内蔵のマイクから入力された音声を DirectSound Device フィルタで出力する。これらのフィルタの処理により、顔のエッジ情報と音声情報をリアルタイムで送受信可能なコミュニケーションシステムを実現した。

4. 実験

3 章で述べた実験システムを用い、コミュニケーションにおける顔の役割を探るため、顔のエッジ情報量を変えることで会話特性がどの様に変わるか実験評価する。

4.1 実験画像の選択

エッジ抽出のパラメータとしては、画像の分割数、分割されたブロックをサンプリングする割合であるサンプリング率、エッジ抽出後の 2 値化を行う際の閾値の 3 つがあるが、このうち前者 2 つが本システム特有のパラメータである。そこで、縦軸を分割数、横軸をサンプリング率にしてエッジで表現された顔画像を比較する (図 4.1)。

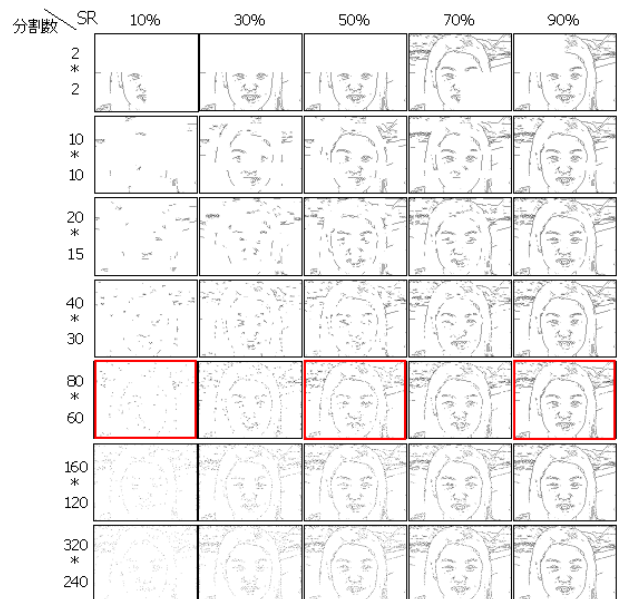


図 4.1 エッジで表現された種々の顔画像

顔のエッジ表現を比較すると、 2×2 分割や 10×10 分割の場合には、ブロックサイズが大きすぎ、ちらつきが大きくなる。逆に 320×240 分割の場合には、ブロックサイズは 1 画素になり、ちらつきがほとんどなく、顔エッジの見え方に大きな変化が見られない。これらの点を考慮し、ちらつきが適度な 80×60 分割の画像を選択し、サンプリング率は 10%、50%、90% の 3 つを選択した。以上により、実験に使用する顔画像は通常の TV 電話の画像を含めた以下の 4 つとした。

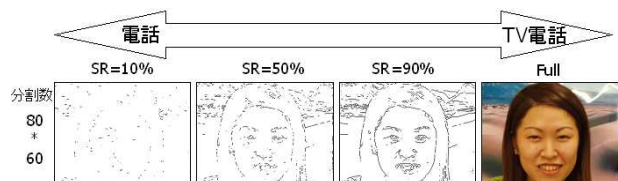


図 4.2 実験で使用する画像

図 4.2 で、最も情報量の少ない 10%の画像はほとんど顔が映らず、電話に近いと言える。そして 50%, 90%, Full (通常の TV 電話) とリニアに顔の情報量を変化させていった時、会話特性がどのように変化するのかを実験する。

4.2 実験環境

図 4.3 に示すように実験者と被験者が壁で仕切られた別々の部屋に入り、直接コミュニケーションが取れない状況の中で、本システムを使って会話をする。

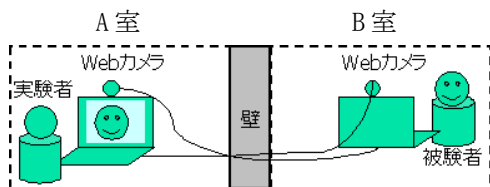


図 4.3 実験環境

被験者の会話特性（行動特性及び心理特性）を調べるために、会話中はビデオカメラで常に撮影する。また、本システムでは実験終了後に会話中の注視対象を分析するため、被験者映像のエッジ抽出を行う処理と並列して被験者の顔映像をファイル保存する。これを実現するフィルタグラフを図 4.4 に示す。

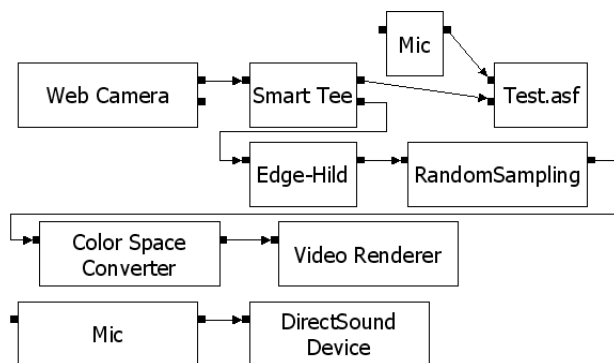


図 4.4 実験に使用するフィルタグラフ

図 3.4 のフィルタグラフとの違いは、Smart Tee フィルタを使うことによって、カメラから入力された映像を Capture ピン (Smart Tee フィルタの上部の黒点) と Preview ピン (Smart Tee フィルタの下部の黒点) に分けてそれぞれ処理を行うことである。Capture ピンから出力された映像は Video Input ピンを通じて Test.asf ファイルに保存される。同時に、Mic フィルタから出力された音声も Audio Input ピンを通じて Test.asf ファイルに保存される。これにより、被験者の映像と会話内容が保存可能となる。Preview ピンを通じた処理と音声の出力は図 3.4 のフィルタグラフと同じである。

4.3 実験の手順

1名の実験者と大学生 12名を被験者として以下の会話実験を行う。なお、実験者と被験者は全員初対面である。

実験の手順：設定したテーマに基づき、実験者と被験者の 1対1で会話実験を行う。テーマは、初対面でも会話がスムーズに進行することを考慮し、タレント写真の一覧を見ながら好みのタレントについて話し合うという課題を設定した。各実験で条件がなるべく同じになるように、実験者が被験者に対してする質問はあらかじめ決めておく。例えば、「この写真一覧の中で、最も好きなタレントは誰か？そしてその理由は何か？」というように質問する。被験者を図 4.2 で選定した 4つのシステム毎に 4グループ (各グループ 3人) に分けて実験を行い、グループ毎の測定データの平均値を取って評価する。

- 第 1 グループ：サンプリング率 10%の本システムを使用し、一人につき 3分間会話する。
- 第 2 グループ：サンプリング率 50%の本システムを使用し、一人につき 3分間会話する。
- 第 3 グループ：サンプリング率 90%の本システムを使用し、一人につき 3分間会話する。
- 第 4 グループ：通常の TV 電話を使用し、一人につき 3分間会話する。

測定項目：行動特性と心理特性を測定する。

行動指標特性の評価項目として、顔のエッジ情報量を変えることで会話時の注視行動がどのように変化するかを調べる。本来であれば、円滑なコミュニケーションを行う上で最も重要である相手の目を見るという行動の発生頻度を測定すべきであるが、TV 電話を通じてアイコンタクトの頻度を測定することは困難であるため、ここでは相手の目を見ている時間による簡易な方法を用いる。

心理特性の評価指標としては、映像によるコミュニケーションのしやすさ、恥ずかしさ、違和感等の 5項目について、5段階の評定尺度による質問紙を用いて測定した。ただし、コミュニケーション時に感じる「恥ずかしさ」といった心理特性は、被験者の資質による影響を受けると考えられるため、シャイネス指標を用いて被験者の事前スクリーニングを行った。

シャイネスの測定では、以下のような項目に答えてもらう形式を取った[8]。

1. 私は、新しい友達がすぐにできる。
2. 私は、人がいるところでは気後れしてしまう。
3. 私は、引っ込み思案である。

全部で 16項目あり、あてはまる (5点)、ややあてはまる (4点)、どちらともいえない (3点)、ややあてはまらない (2点)、あてはまらない (1点) の 5段階評価を行う。質問 1 は逆転項目であり、点数を集計するときに評点を逆転させる。このように 16項目全ての評点の平均値をとる。評点が高く 5点に近ければその人はシャイであり、評点が低ければシャイでないと判定する。

人によって相手の顔を見る時間は違うが、似た気質 (シャイネスの値が同じ) を持つ人が相手の顔を見る時間には相関があると考えられる。したがって、各グループで

同じシステムを使用するなら、顔を見る時間に大きな差が出ないと考えられる。逆に、各グループで違うシステムを使用して顔を見る時間に差が出たなら、それは使用システムの影響であると考えられる。つまり、第1グループから第4グループで、顔を見る時間に差が出たなら、それは顔の情報量の違いが被験者に何らかの影響を及ぼした結果と考えられる。

5. 実験結果と考察

5.1 実験結果

グループ毎のシャイネスの平均値は第1グループ（サンプリング率10%）から第4グループ（通常のTV電話）まで順に、3.27, 2.80, 2.53, 2.77となった。3.0が中間値なので、2.5から3.5の範囲に収まるよう被験者をグループ分けし、実験を行った。

次に、被験者が実験者の顔を見ている時間のグループ毎の平均は以下ようになった。

- 第1グループ（サンプリング率10%）：42.3秒
- 第2グループ（サンプリング率50%）：81.3秒
- 第3グループ（サンプリング率90%）：45.0秒
- 第4グループ（通常のTV電話）：58.3秒

これをグラフにしたものが図5.1である。

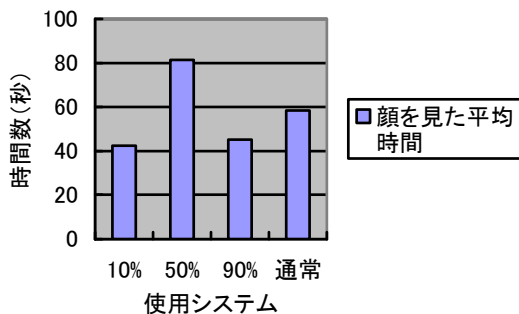


図 5.1 グループ毎の平均時間

また、実験終了後に行ったアンケートでは、顔の見え方と会話時の恥ずかしさについて、以下の項目を5段階評価してもらった。この結果を図5.2に示す。

- ① 相手の表情がわかったか？
(見えた[5点] ←→ 見えない[1点])
- ② 自分の映りが気になったか？
(気になる[5点] ←→ 気にならない[1点])
- ③ 相手の顔を見るのに恥ずかしさを感じたか？
(感じた[5点] ←→ 感じない[1点])
- ④ このTV電話は話しやすかったか？
(話しやすい[5点] ←→ 話しにくい[1点])
- ⑤ 明るい感じで話ができただか？
(できた[5点] ←→ できない[1点])

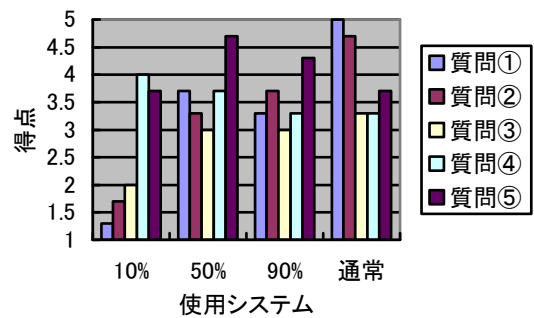


図 5.2 グループ毎の得点平均

さらに、同アンケートで「初対面の人と会話をする時、通常のTV電話と白黒線画のTV電話（本システム）ではどちらがどの程度、相手の顔を見ることに恥ずかしさを感じるか？」という質問を行ったところ、被験者12人全員が通常のTV電話の方が恥ずかしいと答えた。どの程度恥ずかしいかに関して5段階評価してもらった結果を図5.3に示す（5点を最も恥ずかしいとする）。

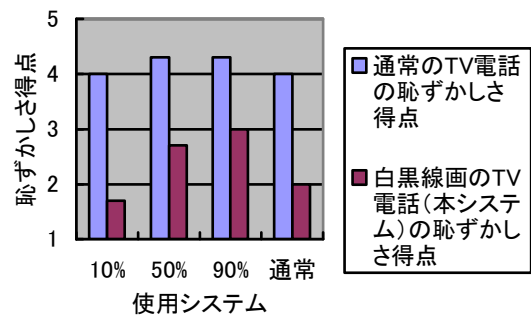


図 5.3 グループ毎の恥ずかしさ得点平均

4グループ全ての恥ずかしさ得点平均は、通常のTV電話の恥ずかしさが4.2点で、白黒線画のTV電話（本システム）の恥ずかしさが2.3となった。

5.2 考察

斉藤孝氏によると、コミュニケーションを円滑にするための基本原則として、以下の4つを提唱している[9]。

- ① 目を見る：目を見ることで、場の空気が暖まり、コミュニケーションが円滑になる。しかし、日本では目を見つめていると気持ちが伝わりすぎるようで恥ずかしいということもある。
- ② 微笑む：微笑むということは相手を受け入れているというサイン。受け入れられていると分かれば、話しやすくなり、会話は盛り上がる。
- ③ 頷く：頷きは、相手の意見に同意・同調する傾向を示すが、その関係は絶対的ではない。
- ④ 相槌を打つ：話のテンポや流れをよくすることが相槌の主たる狙い。

話をする時に、この4つを確実にこなすことで、会話

の雰囲気は暖かくなると斉藤氏は唱える。

しかし、実際にTV電話を通じて会話をしてみると、抵抗感や恥ずかしさといった感覚を経験することがある。特に、①の目を見るということに関しては、通常のTV電話では映像が鮮明すぎるため、ありのまま見られているという心理状態から緊張感や恥ずかしさが生まれ、相手の顔や目を見づらいということが考えられる。

恥ずかしさの緩和

5.1節で示した実験結果より、顔のエッジ表現を用いたコミュニケーションシステムにより、相手の顔を見る恥ずかしさが軽減され、コミュニケーションがより円滑になることが推定される。

以後、この推定に基づき、顔のエッジ表現と恥ずかしさの緩和との関係について考察する。

まず、図5.1で、第4グループ(通常のTV電話)では相手の顔を見ていた平均時間が58.3秒であるのに対し、第2グループ(サンプリング率50%)では全グループ中最高の81.3秒であった。その差は23秒であるが、被験者が1回に顔を見る時間は平均で2.3秒であるので、第2グループは回数にして10回も多く顔を見たと言える。つまり、顔のエッジが半分程度見える状態(本システムのサンプリング率50%)の時に、相手の顔を見る恥ずかしさが最も軽減された可能性がある。また、図5.2を見ると、①②③のグラフが共にS字カーブを描きながら右上がりになっている。すなわち、10%点から50%/90%まで大きく上り、さらに実映像で上るというグラフである。これは、10%点では顔映像の十分な知覚ができなかったため、顔映像に対する被験者の注意が向かなかったが、50%-90%で顔映像としての知覚が成されたため注意が顔映像に向いたと考えられる。このことは、第1グループ(サンプリング率10%)の被験者における「何が映っているのか分からなかった」、「普通の電話に感じた」等のアンケート結果からも裏付けられた。質問項目②では、エッジ映像の情報密度が上るのに伴って評定値も上昇していることから、被験者は相手の顔がより鮮明に見えてくるに従って、自分の顔の映りが気になったといえる。一方、恥ずかしさに関する評定③は、50%-90%では差は無く、実映像において評定が上昇したことから、恥ずかしさの程度はエッジのサンプリング密度にはあまり影響を受けない。さらに、図5.3では、被験者12人全員がエッジ映像より実映像の方が恥ずかしいと答えた。その代表的な理由は、実映像はリアルで鮮明すぎて初対面だと恥ずかしいが、エッジ映像だと細かい表情が分からず、自分の映りをあまり気にせず話せるからというものだった。以上より、恥ずかしさの程度はエッジ映像か実映像かといった違いに強く左右され、特にサンプリング率50%のエッジ映像の場合に恥ずかしさが最も小さい。

コミュニケーションの円滑さ

次に、コミュニケーションの円滑さについて考察する。図5.2において、質問④の「話しやすさ」のグラフは右下がり、また質問⑤「明るい感じで話ができただか？」は50%点を頂点とする山形グラフとなった。このことから、被験者は50%点のエッジ映像による通信において会話をリラックスして楽しめたと考えられる。以上より、被験者はサンプリング率50%のエッジ映像通信において恥ずかしさが最も小さく、リラックスして会話を楽しめることが示された。

6. まとめと今後の課題

本報告では、顔のエッジ表現を用いたコミュニケーションシステムを構築し、顔のエッジ情報密度を変化させた時、会話特性がどの様に変化するかを実験的に評価した。その結果、被験者はサンプリング率50%のエッジ映像通信において恥ずかしさが最も小さく、リラックスして会話を楽しめることが示された。今回の実験では、顔のエッジ情報密度を変化させた場合に人間の会話特性がどのような影響を受けるかを大まかに調べることを目的として、小規模に探索的な研究を行った。

今回得られた結果をもとに、今後は会話特性をより詳細に、また、定量的に評価することが必要である。

<参考文献>

- [1] W.フォン・ラフラー・エンゲル編著、本名 信之訳、「ノンバーバルコミュニケーション」、大修館書店、1981。
- [2] 犬童 早苗、小磯 花絵、下嶋 篤、岡田 美智雄、片桐 恭弘、「映像を介したコミュニケーションの特徴分析」、信学技報HCS96-42(1997-01)、pp.27-34、1997。
- [3] 原田 悦子、「人の視点からみた人工物研究」、共立出版株式会社、1997。
- [4] 黒川 隆夫、「ノンバーバルインタフェースと今後の展開」、信学技報HIP2000-33(2000-10)、pp.49-54、2000。
- [5] sonson@Picture&Software
<http://www.geocities.jp/yuichiy2001/>
- [6] DirectX
<http://www.microsoft.com/>
- [7] DirectShow
http://www.microsoft.com/japan/msdn/library/default.asp?url=/japan/msdn/library/ja/jpdx8_c/ds/default.asp
- [8] 堀 洋道、「心理測定尺度集(1)~(3)」、サイエンス社、2001。
- [9] 斉藤 孝、「コミュニケーション力」、岩波新書、2004。