

顔の輪郭画像を用いた会話システムにおける視覚特性

吉田亨[†] 米村俊一[‡] 徳永幸生[†] 杉山精[†] 大谷淳^{†‡}

[†]芝浦工業大学

[‡]NTT サイバーソリューション研究所

^{†‡}早稲田大学

1. はじめに

コミュニケーションにおいて顔、とりわけその顔の表情は極めて重要な役割を果たしている。相手の表情から、性別、年齢などの個人情報や、更に喜怒哀楽などの心理的の内面さえも把握することができるからである。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が、一方で 55% は顔の表情によって伝達されるという報告 [1] もある。

しかし、顔をリアルに表示する TV 電話には抵抗感があるといった報告 [2] や、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 [3] も報告されている。

本稿では、顔映像を用いた映像通信の抵抗感軽減を目的として、顔の輪郭映像を用いたコミュニケーションシステムを構築し、伝送映像の情報量と会話特性の関係を実験的に検証する。今回は、会話時における発言時・聴取時での目の注視率の違いに注目し、コミュニケーション時に使用する映像による会話特性の違いを計測、評価を行った。

2. 輪郭映像通信システムの概要

図 1 は、原画像の入力からサンプリングされたエッジ表現による輪郭映像の生成までの流れを示したものである。Web カメラから入力された映像を取り込み、1 フレーム毎に取り出した原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。その後、画像をサブブロックに分割し、分割したサブブロックをランダムにサンプリングして、出力した映像が輪郭映像となる。

図 2 に、輪郭映像通信システムの概要を示す。Web カメラを装備した PC が、ルータを介してネットワーク接続されている。Web カメラからの入力映像に対し、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し、相手側 PC に映像を送信する。そして、映像の通信を行うことで、送信側 PC には相手側 PC から送信された映像が表示される。

画像の取り込み、輪郭映像の生成、映像の通信は、Microsoft 社の DirectX を用いて構築した。システムは DirectX に付属する GraphEdit のプレビューの機能を用い、輪郭映像の生成は、DirectShow 上で動作するエッジ抽出フィルタを開発し、GraphEdit 上で動作させた。音声はマイク内蔵のスピーカを双方向で直結することで、音声チャネルを確保した。

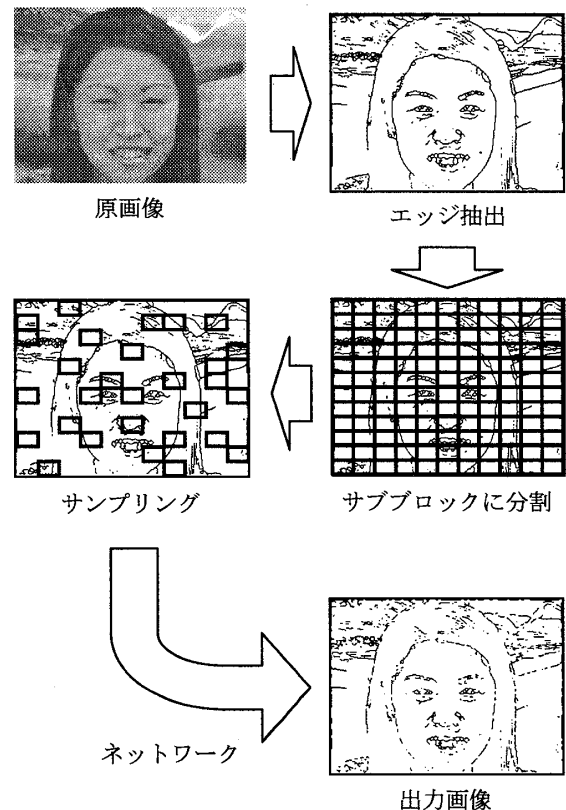


図 1 輪郭映像生成処理の概要

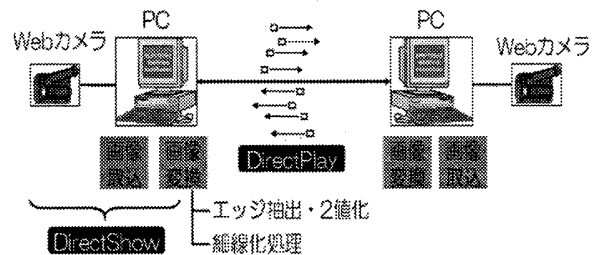


図 2 実験システムの概要

3. 会話実験

会話時の人間の挙動は様々な要因によって変化するため、会話行為を観測することによってシステム評価を行う場合、それらの要因を統制する必要がある。本実験では、コミュニケーションスタイルを一意的に統制するために、次のように会話状況を設定した。

話し相手：実験者と日常的に会話を交わす顔見知り

会話内容：相手に依頼する説得的な内容

会話場面：非対面の遠隔通信

Visual Characteristics of Visual Communication System Based on Edge Abstraction

[†]Tohru YOSHIDA (m106110@sic.shibaura-it.ac.jp)

[‡]Shunichi YONEMURA (yonemura.syunichi@lab.ntt.co.jp)

[†]Yukio TOKUNAGA (tokunaga@sic.shibaura-it.ac.jp)

[†]Kiyoshi SUGIYAMA (sugiyama@sic.shibaura-it.ac.jp)

^{†‡}Jun OHYA (ohya@waseda.jp)

[†]Shibaura Institute of Technology

[‡]Cyber Solutions Laboratories, NTT ^{†‡}Waseda University

3.1 被験者

実験者と顔見知りの大学生 3 名を被験者として用いた。また、日常的に TV 電話を利用している被験者はおらず、輪郭映像を用いた映像通信システムの利用経験者もない。

3.2 使用する映像

会話実験は、リアル映像と伝送映像の情報量を落としたり輪郭映像との 2 条件下で行った。なお、輪郭映像のサンプリング率は 100%とした。図 3 に、使用映像の比較例を示す。

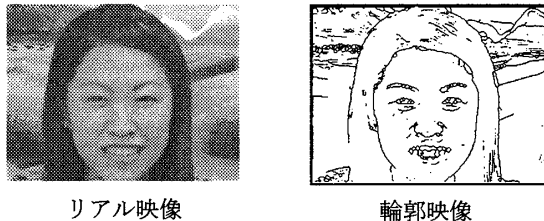


図 3 実験で使用する映像

3.3 実験装置

注視点の抽出・注視時間の測定には、アイマークレコーダを用いた。アイマークレコーダは、帽子に視線観測装置を搭載した簡易（被験者への負担が小さい）タイプである nac 社の EMR-8B を使用した。

アイマークレコーダには視界映像を撮影するカメラと、瞳の動きを観測するカメラがついていて、視界映像に瞳の動きから測定されたアイマークを付加した映像を出力する。アイマークが記録された視界映像はデジタル化して保存し、実験後の解析で使用した。

3.4 実験手続き

被験者にタスク内容を記した紙を見せ、実験者と遠隔会話を行なわせた。タスク内容には現在自分が置かれている状況と、説得する内容が書かれている。今回タスクは 2 パターン用意した。実験を行う前には、システムに慣れてもらうため、2~3 分程度の会話練習を行った。

使用する映像の 2 条件とタスクの 2 条件は被験者ごとにランダムに組み合わせ、会話を行った。

4. 実験結果

4.1 注視時間の測定

目の注視時間を測定する際、アイマークが瞳を中心として周囲約 1cm の範囲にフォーカスされた場合を注視したと判断した。今回、実験者は眼鏡をかけているため、眼鏡の上にアイマークがフォーカスされた場合、目を見たと判断した。

4.2 平均注視時間

被験者によって発言時間、聴取時間に個人差が存在するため、目の注視時間をプロットするには絶対時間を使用せず、 $(\text{目の注視時間}) \div (\text{発言時間}, \text{聴取時間})$ により正規化した値を用いた。図 4 に、被験者の平均注視率（相手の目を見た時間率）を示す。図 4 のグラフから、リアル映像、輪郭映像共に、発言時よりも聴取時で注視率が高くなっていることが確認できる。

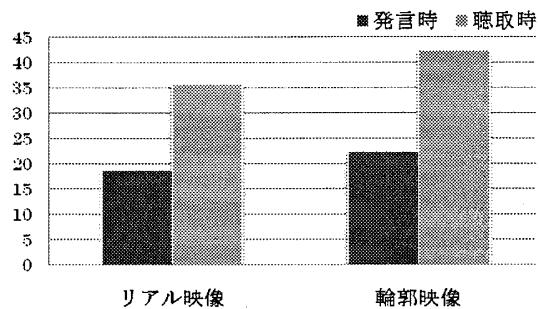


図 4 システム毎の平均注視率

5. 考察

アガイル[4]は対面会話において一般的な話題の会話の場合、一方が相手に向ける視線量は聞きながら相手を見ている場合に 75%、話しながらの場合に 40%という一種の基準値を示しており、会話時に、相手の話を聞いている際には自分が発言している際よりも視線量が多いことを挙げている。

今回の実験ではサンプル数が少なく、有意差検定を行っていないが、発言時と聴取時を比較してみると、リアル映像、輪郭映像どちらに関しても、発言時よりも聴取時において注視率は高いことから、伝送映像の情報量に関係なく、発言時よりも、聴取時のほうが目の注視率は高いことがわかる。これは、発言時には考えを表現することに集中し、また、発言時に相手に目を向けると、発言が終わったと思われる可能性があるため、相手に視線を向けられないものと考えられる。

また、今回測定した注視率はアガイル[4]が示したもののよりも注視率の値が低いことがわかる。これは日本人が外国に比べ、相手を直視することに抵抗を持っていることが表れていると考えられる。

6. まとめ

本稿では、顔の輪郭映像を用いた通信システムを構築し、会話時における発言時と聴取時とでの目の注視率に着目し、コミュニケーション時に使用する映像による違いの計測、評価を行った。その結果、映像を問わず、発言時より聴取時のほうが目の注視率は高くなり、また、発言時、聴取時間問わず、輪郭映像のほうが視線量が高くなることが示唆された。

しかし、本稿で行った実験は、ある特定の会話状況であり、今回使用した会話状況以外でも同様な結果が得られるかどうか検討する必要がある。また今後は、目以外の場所の注視箇所や、視線を送るタイミングのパターンにも着目し、考察していきたいと考えている。

参考文献

- [1] W. フォン・ラフラー、エンゲル 編著、本名 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、小磯、下嶋、岡田、片桐：映像を介したコミュニケーションの特性分析、信学技報 HCS96-42, pp22-34(1997)。
- [3] 原田：人の視点からみた人工物研究、共立出版株式会社(1997)。
- [4] 大坊：しぐさのコミュニケーション、サイエンス社(1998)。