

J-003

## 人の目の動きに含まれる「心の動き」情報の分析

小田 英介<sup>†</sup> 王 磊<sup>†</sup> 井上 真紀子<sup>†</sup> 伊藤 昭<sup>†</sup> 寺田 和憲<sup>†</sup>  
岐阜大学大学院工学研究科<sup>†</sup>

### 1 はじめに

人はコミュニケーションにおいて言語情報だけでなく、多くの非言語情報を利用している。目の動きは、そのような非言語情報の一つであるが、そのような情報を抽出する最初の一步として、人が対話中の顔を見てどのような「心の動き」を読み取るのかの実験を行った。具体的な手法としては、人が対話している時の顔を撮影し、対話中の人の顔動画像を作成した。次に、その中から典型的と思われる(数秒程度の)シーンを切り出し、切り出された音声を含まない映像のみのビデオクリップから、対話の状況を知らない被験者がどのような「心の動き」を読み取るのかを実験した。

### 2 動画像の作成

人が顔からどのような「心の動き」を読み取るかを調べるために用いる顔動画像は、不自然にならないようにする必要がある。けれども、人は初対面の人と会話するときは緊張するだろうし、カメラで撮影されていることに気がつけばカメラを意識してしまうことが多々ある。そのため自然に表出される顔動画像を撮影するために、次のような実験を行った。実験中のカメラなどの配置は図1のとおりである。

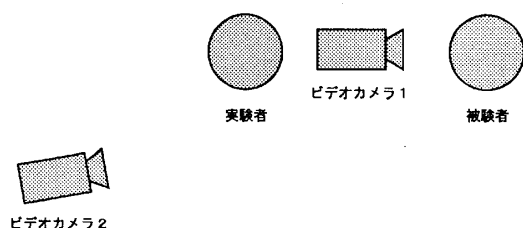


図1: ビデオカメラと被験者、実験者の配置

実験に先だって、被験者には視線による移動物体

Analysis of the information on the state of mind contained in the human gaze movement

<sup>†</sup>Oda Eisuke, Wang Lei, Inoue Makiko, Akira Ito, Kazunori Terada, Gifu Univ.

追視実験を行い、その時の顔の様子をビデオカメラ1を使って撮影すること、その後口頭でアンケートに答えてもらうことなどを伝えた。追視実験の終了後にビデオカメラ1を被験者の目の前から取り除き、アンケートの項目以外は自由に談話を行った。この間、被験者の知らないビデオカメラ2を用いて被験者の顔を撮影し続けた。なお実験後のアンケートによれば、実験中にビデオカメラ2で撮影を行っていることに気がついた被験者はいなかった。このことから追視実験終了後の被験者の顔情報を自然に表出されたなものであると判断し、以後の評価実験用データとして用いた。

### 3 顔情報評価実験

人が顔からどのような情報を得ているのか/得ることができるのかを調べるため、次のような顔情報の主観評価実験を行った。

#### 評価対象:

前節で述べた方法で作成した自然な顔動画像から、目領域の判別が容易な1人(20代前半の女性)の動画(約50分)を取り出し、この中から1~5秒の長さの動画クリップ(640×480pixels,30frame/sec)14枚を切り出した。この動画クリップを切り出す際には、わかりやすい目の動きを含んでいるできるだけ多様なシーンを採用するように心がけた。なお、動画クリップには音声が含まれていないため、評価者は対話文脈からの判断が出来ないようになっている。

#### 評価方法:

動画を評価する形容文20個を用意し、被験者には動画を1枚見る毎に「非常にそう思う」「少しそう思う」「どちらとも言えない」「少しそう思わない」「全然そう思わない」の5つから、当てはまるものを選んでもらった。その結果は、「非常にそう思う」の5点から「全然そう思わない」の1点までに数値化して表現する。

評定に使った20の形容文を一語で表したものが表1の変数である。実際に用いられた形容文は、例

例えば「期待」は「何かを期待している」であり、「自信」は「自信をもっている」などである。評価者は20代前半の学生10名(男性8名, 女性2名)であり, 14個のビデオクリップについて, 20個の項目を評価するのに要した時間はおおよそ15-30分であった。

#### 4 因子分析

実験からは, 14クリップ×被験者10名=140個の評価ベクトルが得られた。この20次元ベクトル140個を用いて因子分析を行い, ここで扱った顔動画に含まれている「心情報」の抽出を試みた。その結果, 固有値1以上という条件で5つの因子を抽出し, 累積寄与率56.34%となった。抽出した因子のバリマックス回転後の因子負荷量を表1に示す。次

表1: 因子負荷量(バリマックス回転後)

変数\因子	第1	第2	第3	第4	第5
期待	<b>0.679</b>	-0.088	0.080	-0.051	0.010
自信	<b>0.630</b>	-0.269	-0.109	-0.123	-0.065
質問意思	<b>0.620</b>	0.322	0.028	0.057	-0.042
楽しみ	<b>0.616</b>	-0.506	-0.174	-0.414	-0.001
照れ	<b>0.613</b>	-0.124	0.174	-0.291	-0.080
伝達意思	<b>0.541</b>	0.074	-0.336	0.406	-0.143
思い出し	<b>0.526</b>	-0.09	0.070	0.445	0.043
様子伺い	<b>0.374</b>	0.224	0.070	0.082	0.109
不安	-0.092	<b>0.737</b>	0.277	0.065	-0.132
発言否定	-0.119	<b>0.610</b>	0.065	0.043	-0.068
落胆	-0.185	<b>0.551</b>	0.259	-0.006	-0.150
理解不能	0.352	<b>0.515</b>	0.298	-0.166	0.078
困惑	0.113	<b>0.509</b>	0.087	-0.033	0.085
退屈	-0.022	0.255	<b>0.852</b>	0.040	-0.008
興味無し	0.003	0.228	<b>0.843</b>	-0.029	-0.156
他に傾注	0.224	0.302	<b>0.663</b>	-0.113	-0.142
考え中	0.071	0.095	-0.044	<b>0.808</b>	0.094
驚き	0.237	0.084	0.030	<b>-0.682</b>	-0.053
傾聴	-0.053	-0.033	-0.047	0.015	<b>0.870</b>
理解	-0.001	-0.070	-0.161	0.122	<b>0.796</b>

に各因子の意味を考えるために, この因子を用いて実験で用いた14個のクリップの意味による分類を試みた。まず, それぞれのクリップに対する10人の被験者の回答の平均値をそのクリップの評価値と考え, その5因子への射影をそのクリップの「意味」と考えた。

次に, それぞれのクリップに対して2人の実験者が簡単なキーワードによるラベル付けを行い, このラベルとともに14個のクリップを第一因子と第二因子からなる2次元空間に配置した。結果を図1に示す。図で, 5次元の射影空間でのユークリッド距離が一定(1.5)以下のものを線で囲んで同一グループとして表現した。

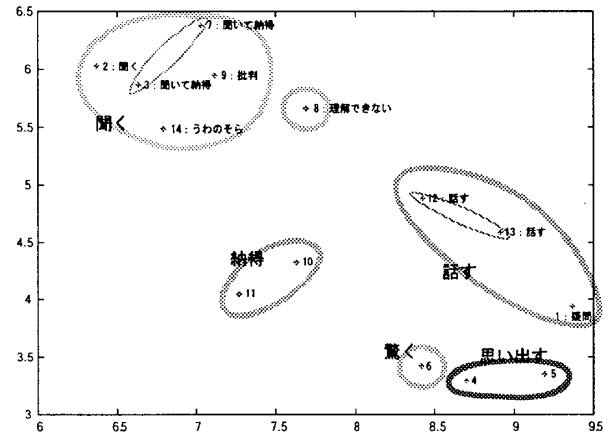


図2: ビデオクリップの分類

図を見ると, 話すに係わる状況が右下に, また聞くことに係わる状況が左上に来ており, 第1因子は情報発信に, 第2因子は情報受信に対応していると考えられる。また対応する形容文から第3因子は外部情報無視を, 第4因子は内部情報処理を, 最後に第5因子は外部情報への注意に対応していると考えられる。このように, 因子分析の結果抽出された因子が, 話し相手に対するスタンスを表現しているように見えるのは興味深い結果である。

#### 5 ロボットとの対話—表情の少ない動画の場合

これまでに述べたものは, 人と人との対話であったため表情の表出があり, 目の動き情報とはいえないものであった。我々は, より目の動きに情報が集中している動画を取得する目的で, 人と頭部ロボットとの対話実験を行った。このとき映像は, 頭部ロボットの目の位置に備え付けられた二つのカメラからの映像を用いる。現在これらを解析中であり, 我々は両者を比較することで表情の効果を見るとともに, 次のステップとして目領域だけの映像での実験を計画している。