

K-14 対話者の顔観測に基づく対面コミュニケーションの状況分析
 Understanding Atmosphere of Communication from Observation of Human Faces

伊藤 淳子[†]
 Junko Itou

角所 考[‡]
 Koh Kakusho

美濃 導彦[‡]
 Michihiko Minoh

1. はじめに

人間は、会議、講義、あるいは日常会話などの対人コミュニケーションを行う際、意識的あるいは無意識のうちにコミュニケーションの状況を把握・分析しながら自身の次の行動を決定している。このようなコミュニケーションの状況を計算機による観察から客観的に分析することができれば、インタビューの様子を撮影した映像資料から観察者にとって興味のある状況を抜き出す際、何らかの指標を与えることにより膨大な資料の中から注目箇所を抜き出すことが可能となり、検索の手間を省くことができる。

対人コミュニケーションの状況は「セッティング」「外見」「情動」などの要因から構成される [1][2]。これらの要因の中で、同一のコミュニケーションでありながら時間と共に変化する要因は、参加者個人の一時的な特徴である気分や情動といった「感情」のみである。したがって、この感情の変化によってコミュニケーションの状況が変化するという。

コミュニケーションでは、言語だけでは伝達が困難な様々なノンバーバル情報がやりとりされる。これらノンバーバル情報の中でも、顔は、その輪郭や肌の様子、目、口といった顔器官の造形や配置によって個人の印象や風貌を伝えたり、時間と共に変化する表情によって心理状態を伝えるなど、重要な役割を担っている [3]。特に表情は感情を表出し、伝達する最も重要なノンバーバル情報として知られている [4]。このことから、顔はノンバーバル情報を伝達するためのメディアとみなすことができる。そこで、本稿ではこれを「顔メディア」と呼び、これに注目して感情を推定し、コミュニケーションの状況を分析する。

2. コミュニケーション状況の構造と顔メディア

2.1 状況の構成図

Brownら [1]は、コミュニケーション状況を構成する要因について、図1に示すような構造化を行った。

本研究で想定しているコミュニケーションは、ミーティングスペースなどに設置された椅子に座った状態で行われるものとする。この場合、図1の要因のうち、場面、参加者間の関係はコミュニケーションが行われている間変化しない。参加者個人に関しても、人種などは変化せず、外見などの特徴も安定しているため、あるコミュニケーションにおいて時間とともに変化する要因は「一時的な特徴」のみとなる。すなわち、一時的な特徴の変化によってコミュニケーションの状況が変化すると考えられる。

2.2 顔メディアと感情の表出

前述のように、感情は顔メディアに注目することによって推測することが可能である。顔メディアによって伝達されるノンバーバル情報としては、顔表情、視線、頷きがある。また、発話の有無は音声だけでなく口の動きの観察によっても認識できる。本稿ではこれらを特に Facial Signal (FS) と呼び、FS の典型的な組み合わせのパターンを感情と定義

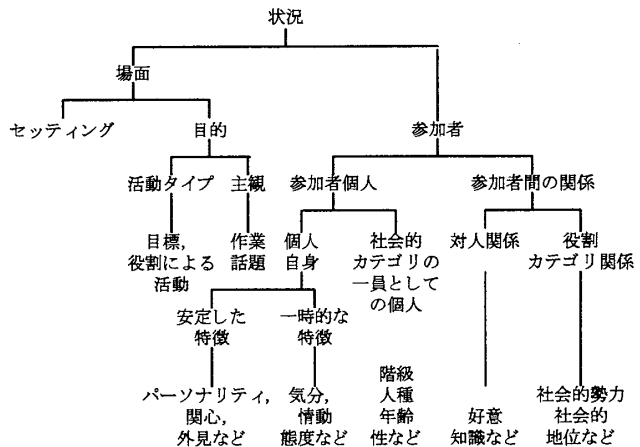


図1: 状況の構成図

する。図2にコミュニケーションにおいて扱われる情報の種類およびFSの占める位置を示す。

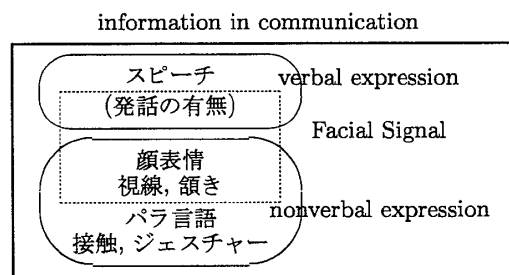


図2: コミュニケーションにおけるシグナル

コミュニケーションに参加している各人のFSを認識し、感情を抽出することによって、コミュニケーションの過程を感情の系列で記述し、これを状況とする。これにより、コミュニケーションの状況を顔メディアによる情報のパターンに基づいて分類できると共に、それらのパターンをクエリとして利用することで、会議等のコミュニケーションを記録した映像データベースから、特定のパターンを示している部分を即座に検索することも可能となる。

2.3 Facial Signal による感情の分類

対話において、話し手のスピーチを聞き手が聞いているとき、以下の行動が見られる [5]。

- ▷ 話し手の方を向く
- ▷ 話し手と視線をあわせようとする
- ▷ 時々頷く

逆に、話に興味のない場合には以下の行動が見られる。

- ▷ 微笑みが見られない
- ▷ 話し手の方を見ることがない
- ▷ 頷きが見られない

視線方向を顔の向きとみなすことによって、これらの行動はFSに基づいて次のように解釈することができる。まず、話し手の方を向く、向かない、話し手と視線をあわせるといった行為は、FSの視線方向の推定により判断できる。

[†] 京都大学大学院 情報学研究科

[‡] 京都大学 学術情報メディアセンター

頷き行為に関しては、顔の向きの動きにより抽出することが可能である。また、微笑が見られるか否かはFSの1つである顔表情から推定することができる。したがって、上記の各項目は、FSの値の組み合わせによって記述することのできる、ある感情が反映された行為であるといえる。

聞き手のスピーチを聞いている際に生じる感情は、さらに顔表情によって細分類可能である。たとえば、笑顔は相手の心を和ませ、くつろがせる効果があるといわれる[5]。すなわち、話を聞いている際の聞き手の表情によって、くつろぎを感じながら聞き、およびくつろぎを感じてはいないが聞いている状態への分類が可能である。

以上の議論に基づき、3つの感情とそれを構成するFSの値の組み合わせを表1に示す。表にないFSの値の組み合わせは、次の感情への遷移途中と考える。

表1: FSに基づく感情推定

FS	感情		
	くつろぎ	くつろぎ感じず	聞いている
顔表情	笑顔	笑顔以外	聞いている
視線		話し手	聞いている
頷き		あり	聞いている
発話の有無		なし	聞いている

対人コミュニケーションの形式には、会議のような多対多の対話、講義のような多対1の対話、対談のような1対1の対話などがあるが、本研究では、このうち最も基本的と考えられる2人の人物による1対1の対話を対象とし、表1に示した感情の組み合わせによりコミュニケーションの状況を分類することを目標とする。よって、コミュニケーションの状況は、2人の人物の3つの感情の組み合わせにより、計9つに分類される。

3. 感情推定の検証

3.1 対話実験

本研究で対象とするような1対1の対面コミュニケーションにおいて、前節で述べたような感情が発生し、かつそれらを顔メディアに注目することにより推定することが可能であることを調べるため、実際の対話の様子を観察した。

互いに顔見知りである普段からよく会話をしている被験者を2人1組とし、向かい合って座った状態で自由に会話してもらい、その様子を2台のビデオカメラで撮影した。被験者は同性、撮影時間は約15分程度で、会話の内容は共通の知人のこと、よく行く店のことなど日常的な事柄であった。

撮影にあたっては、自然な対話が行われるよう、図3に示すように各組ともカメラを被験者の斜め後方の充分離れた位置に設置し、可能な限り制約を加えず自由に会話をしてもらった。

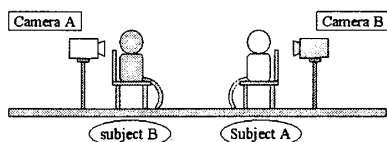


図3: 撮影環境

3.2 観察結果

撮影映像のうち1組から切り出した5分間について、5人の観察者が5秒おきに静止画からFSの値を推定し、同

時に観察者がその映像からどのような態度、感情が感じられるかをたずねた。

まず、FSにより聞き手が“聞いている”と判断される状態のうち、実際に観察者が聞いていると感じた状態は99.4%であり、この指標は必要条件を満たしていることがわかる。これに対し、FSにより“聞いている”と判断される状態のうち、実際に聞いていないと感じた状態はごくわずかであり、逆に、観察者が聞いていないと感じられた状態のうち97.3%が表のFS値と一致している。したがって、聞いていないことを判断するための指標は誤りではないがさらに詳細な条件を付け加える必要があることがわかった。

続いて、FSにより“聞いている”と判断される状態について議論する。“くつろぎを感じていない”と判断される状態における感情推定結果のうち、2番目に多かった感情は“真剣”に話を聞いているというものであり、20.2%を占めた。最も多かったのは、和やかでも真剣でもなく、特に何も感じずに話している、という解釈であった。“くつろぎ”と判断された状態のうち観察者が実際に和やかに聞いていると判断したものは84.2%と高い割合を示した。しかし、FS値が異なるにも関わらず和やかに話していると観察者が考えた状態も相当数存在した。これらは、特にくつろぎと判断された状態の前後に集中しており、時間あるいは状態の遷移も含めた感情推定を行う必要があることを示唆しているものと考えられる。

以上から、FSにより話を聞いているか否かの大きな判断はほぼ可能であり、表1に基づく分類は適当であると考えられる。今後、この結果をもとにさらに詳細な分類と調査を行うと同時に、計算機による自動化を目指す。

4. むすび

本稿では、まず、対面コミュニケーションの状況を構成する要因のうち、コミュニケーションの状況にダイナミックに影響する主な要因が感情であることを述べた。次に、顔メディアに注目することによって感情を推定することが可能であることを示し、実際に対話中に見られる対話者の行動に基づき、対話者の顔メディア間に見られるFSの値によってコミュニケーションの状況を分類できる可能性について議論した。実際に対話が行われている様子を撮影した映像を、FSに基づいて複数人が観察した結果、この分類は大まかな推定には適しているがさらに詳しいFS値の分類と実際の感情との関連性を調べる必要があることがわかった。今後は、本稿で得られた結果をもとに、さらに詳細な状況分類を行うと共に、映像からコミュニケーション参加者の表情および頭部の向きを計算機により抽出する処理について検討する。

参考文献

- [1] Brown, P., & Fraser, C.: "Speech as a marker of situation" In K.R. Scerif and H. Giles (Eds.) Social markers in speech" Cambridge University Press, 1979.
- [2] 大坊 都夫: "対人コミュニケーションの社会性" 対人社会心理学研究, No.1, pp.1-16, 2001
- [3] 吉川 左紀子: "顔・表情認知研究の最前線" 映像情報メディア学会誌, Vol.54, No.9, pp.1245-1251, 2000.
- [4] 黒川 隆夫 著, 電子情報通信学会 編: "ノンバーバルインタフェース" オーム社, 1994.
- [5] 工藤 力 著: "しぐさと表情の心理分析" 福村出版, 1999.