

会話参加者の行動を自動記録する行動観察ソフトウェアの開発

蛇穴 祐稀[†] Prima Oky Dicky Ardiansyah[†] 伊藤 久祥[†]

[†]岩手県立大学ソフトウェア情報学部

1. はじめに

対人コミュニケーションにおいて、言語的と非言語的コミュニケーションとの両立が重要な役割を担っている。このようなコミュニケーションを定量的に分析するために、会話参加者の行動（表情、発話、頷き、顔向けなど）をコーディングする必要がある。しかしながら、観察者によるコーディングは直感的および主観的判断が働くという課題が指摘されており、信頼の高いコーディング結果を得ることが困難である^[1]。そこで本研究では全方位カメラを利用して、会話場面の映像から会話参加者の行動を自動的にコーディングするソフトウェアの開発を試みる。

2. 行動観察ソフトウェアの提案

提案の行動観察ソフトウェア（図 1）は会話参加者の行動のうち、発話や表情、頷き、顔向けをコーディングの対象とする。撮影は、双曲面ミラー型全方位レンズを装着したデジタルビデオカメラで撮影を行う。得られた全方位画像をパノラマ画像に展開^[2]し、その画像から会話参加者の行動をコーディングする。手続きを以下に示す。

2.1 顔特徴の検出

一連のパノラマ画像（以後、これを「パノラマ動画」と呼ぶ）からの顔特徴検出手順を以下に示す。

- パノラマ動画に顔認識を適用する。
- 認識した顔領域を各々切り出し、ROI (region of interest) 画像として格納する。
- 各 ROI 画像に対して、Saragih ら (2011) ^[3] による顔特徴を検出・追跡し、出力する。ここで、顔の検出ができなければ、上記 a. に戻り、ROI を初期化する。
- 顔の位置の変化にともない、ROI を更新する。
- パノラマ動画の最後のフレームまで処理が完了したとき、この一連の処理を終了する。

2.2 発話の検出

複数フレーム間における口の開閉の変動に対

Automated Behavior Analysis during Conversation
Yuki JAANA[†], PRIMA Oky Dicky Ardiansyah[†], Hisayoshi ITOH[†],

[†] Faculty of Software and Information Science, Iwate Prefectural University

し、閾値処理を行うことで発話を検出する。口の 12 特徴点の y 座標に対する標準偏差の平均は、

$$\sigma_{speak} = \frac{1}{n} \sum_{i=1}^n \sqrt{\frac{\sum_{j=1}^{12} (y_j - \bar{y})^2}{12}}$$

で求められる。本研究では、フレーム数 n を 20, σ_{speak} の閾値を各会話参加者 0.73, 0.85, 0.82 画素とする。

2.3 表情の検出

上記 2.1 で検出した顔特徴と、表情データベースの顔画像から検出した顔特徴の照合を行い、表情の検出を行う。ここで、照合の評価式をユークリッドとする。表情データベースは、JAFEE (Japanese Female Facial Expression) を利用した。JAFEE には、日本人女性 10 人による 7 表情（怒り、嫌悪、恐れ、幸福感、悲しみ、驚き、無表情）を含む 213 枚の画像が収録されている。

2.4 うなずきと顔向けの検出

上記 2.1 で検出した顔特徴を 3D 顔モデルの特徴点に当てはめ、各点に対応する奥行き情報を取得する。これに POSIT (Pose from Orthography and Scaling) を適用することで、顔の 6 自由度 ($pitch, yaw, roll, x, y, z$) を求めることができる。うなずきの検出は、 $pitch$ 角度の標準偏差

$$\sigma_{head nod} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\varphi_i - \bar{\varphi})^2}$$

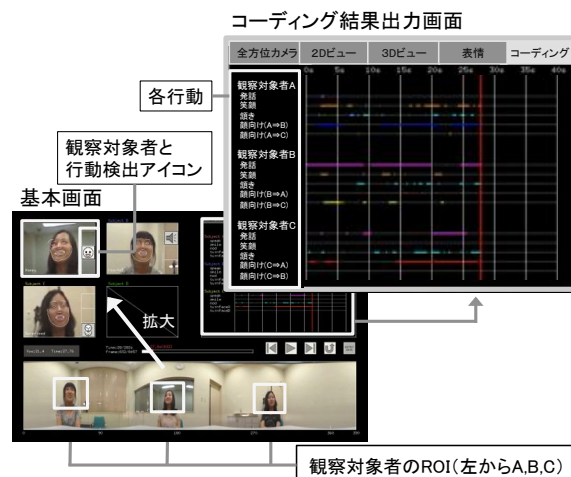


図 1 提案する行動観察ソフトウェアの画面

表 1 手動観察 1, 手動観察 2, 自動観察の各コーディング結果の合計平均時間

行動	合計時間 (秒)		
	手動観察1	手動観察2	自動観察
A 発話	115.4	122.9	111.6
A 幸福感	189.6	97.3	158.4
A 頷き	12.2	7.7	6.6
B 発話	104.0	101.5	147.5
B 幸福感	174.2	125.3	91.2
B 頷き	30.7	21.1	17.7
C 発話	86.4	70.8	144.3
C 幸福感	87.0	82.2	129.2
C 頷き	32.0	28.5	49.6
顔向け (A⇒B)	133.8	59.4	3.2
顔向け (A⇒C)	84.0	57.2	95.5
顔向け (B⇒A)	174.3	159.3	67.8
顔向け (B⇒C)	103.0	96.1	31.1
顔向け (C⇒A)	116.9	110.8	42.3
顔向け (C⇒B)	150.7	135.6	56.6

で求められる。本研究ではフレーム数 n を 50, $\sigma_{head\ nod}$ の閾値を 1.3° とする。会話参加者 (A) から会話参加者 (B) への顔向けは,

$$\theta = \arccos\left(\frac{a \cdot b}{|a||b|}\right) \quad (0^\circ \leq \theta \leq 180^\circ)$$

の値から判定できる。本研究では, $\theta > 150^\circ$ が満たされたとき A から B への顔向けが行われたと判断する。

3. 自動コーディングの評価実験

提案ソフトウェアによる自動コーディングと、観察者による手動コーディングを行い、それらの結果を比較する。

3.1 実験内容

観察対象の動画として, 3 人 (それぞれの会話参加者を「A」, 「B」, 「C」とする) での会話場面を撮影した。撮影には Sony 製 Bloggie (MHS-PM5K) と, 通常のビデオカメラ 2 台を使用した。その動画をもとに, 提案ソフトウェアによるコーディング (自動観察), 通常ビデオカメラの映像からの観察者によるコーディング (手動観察 1), 全方位カメラの映像からの観察者によるコーディング (手動観察 2) を行い, それらの結果の差異を分析した。ここで, 短期間において同一行動を繰り返した場合, その一連の行動を 1 回と数えることとした。

手動観察 1 では, 行動コーディングシステム PTS-113 型 (株式会社ディケイエイチ) を使用しコーディングを行い, 観察者として, 行動観察の経験がある心理コースの学生 3 名を採用した。手動観察 2 では, 観察者として 17 名のメディア

表 2 発話, 幸福感, 頷きのコーディング結果

標本	相関係数	p 値
手動観察1 - 手動観察2	0.87**	0.002
手動観察1 - 自動観察	0.73*	0.024
手動観察2 - 自動観察	0.76*	0.015

** : $p < 0.01$ * : $p < 0.05$

表 3 各コーディング結果の顔向けの関係 (A⇒B, A⇒C を除く)

標本	相関係数	p 値
手動観察1 - 自動観察	0.99**	0.006
手動観察2 - 自動観察	0.99**	0.004

** : $p < 0.01$

システムコースの学生を採用した。

3.2 実験結果

表 1 は, 実験結果を示す。手動観察 1 と自動観察, 手動観察 2 と自動観察によるコーディング結果は, 発話, 表情 (幸福感), 頷きの行動時間において 5% の有意水準で強い相関が認められた (表 2)。一方, 顔向けにおいては, 相関が認められなかったが, A から B の顔向けと A から C への顔向けの結果を除けば, 手動観察 1 と自動観察, 手動観察 2 と自動観察によるコーディング結果は, 1% の有意水準で強い相関が認められた (表 3)。

4. おわりに

本研究では, 会話中の行動を自動観察できるソフトウェアを開発し, 手動での行動観察の結果と比較した。その結果, 発話, 表情 (幸福感), 頷きについて, 手動と自動による観察結果との間に強い相関が認められており, 本システムが有効であることを確認できた。今後, 本システムをさらに改良することで, 面接場面などでのコミュニケーションスキルのトレーニングなどに有効活用できるものと考えている。

参考文献

- [1] 西郡仁朗, “自然会話データ「偶然の初対面の会話」 - その方法論について -”, 人文学報, Vol.330, pp.1-18, 2002.
- [2] 山澤一誠, 八木康史, 谷内田正彦, “移動ロボットのナビゲーションのための全方位視覚系 HyperOmni Vision の提案”, 電子情報通信学会論文誌 D-II, J79-D-II, No.5, pp.698-707, 1996.
- [3] Saragih, J., Lucey, and S., Cohn, J., “Deformable Model Fitting by Regularized Landmark Mean-Shifts”, International Journal of Computer Vision, Vol. 91, No. 2, pp.200-215, 2011.