

全方位カメラを用いた同調現象の観測に基づく人間関係の推定

Capturing Entrainment-based Interpersonal Communication Using Omnidirectional Camera

清水敏雄*1 松村真宏*2 谷内田正彦*3
 Toshio Shimizu Naohiro Matsumura Masahiko Yachida

*1 情報科学芸術大学院大学 *2 大阪大学大学院経済学研究科
 Institute of Advanced Media Arts and Sciences Graduate School of Economics, Osaka University

*3 大阪大学大学院基礎工学研究科
 Graduate School of Engineering Science, Osaka University

In this paper, we integrate "entrainment" phenomenon with image processing as a new and unique approach to capture interpersonal communication among participants in a meeting. We use omnidirectional camera to capture 360-degree view at a time, and process the panoramic images to estimate the relations. We investigated the performance of this approach from various point of view using supervised dataset. The results show that the proposed approach has a promising possibility of automatically capturing interpersonal communication.

1. はじめに

会話においては、言葉だけでなく目線や表情、ジェスチャーなど様々な情報によって意思を伝え合う。言葉による情報を言語情報、言葉以外の目線やジェスチャー、表情などによる情報を非言語情報と呼ぶ。会話者の間で、非言語情報に類似や同期が見られることがあり、その現象は同調現象と呼ばれる [1]。一例として、2 者間会話における発話時間の類似がある [2]。これは、一方の発話時間が増減すると、もう一方の発話時間もそれに対応して増減する現象である。このほかにも声の調子、表情、しぐさ等に同調現象が見られる。同調現象の表れ方は、会話者同士の共感や親密さ、会話の質に応じて変化する。すなわち、同調現象を調べることによって会話者間の人間関係が推定できる。そのため従来から、同調現象に関する研究が広く行われてきた。

会話をはじめとする多人数でのやり取りを研究する際には、会話の様子を記録した映像から分析するやり方がある。映像を人手によって調べたり [3]、映像に画像処理を行うなどして分析する方法である [4]。しかしこの方法は、複数のカメラを用いるなど大掛かりな機材を必要とする場合があった。近年、その問題を解決する機材として全方位カメラが注目されている。これは、360 度の視界を持つ全方位画像の撮影が可能なカメラである。この全方位カメラ 1 台で、会話に参加している人物全員の様子を 1 度に記録することができる。全方位カメラの特性を活かして様々な研究がなされている [5][6][7]。

本研究では、全方位カメラを用いた同調現象の観測に基づく人間関係の推定手法を提案する。これは、これまで人文科学の分野で研究されてきた同調現象を、画像処理の技術を用いて分析しようとする試みである。会話の様子を全方位カメラで記録し、得られた映像に対して画像処理を行うことで同調現象を観測する。そして、観測された同調現象を実際の会話の様子と比較することで、提案手法による人間関係が持つ特徴を明らかにする。

2. 同調現象の観測

全方位カメラを用いた同調現象の観測法について説明する。本研究では、図 1 に示す全方位カメラを用いて会話の様子を記録する。図中の全方位ミラーに周囲 360 度の視界を持つ像が映り、その像がビデオカメラによって撮影されることで全方位映像が得られる仕組みである。全方位カメラにより記録された映像に対して画像処理を行い、人物の動きを取り出す。そして取り出された動きから同調現象を観測する。以下でプロセスの詳細について述べる。



図 1: 全方位カメラ

2.1 人物の動きの抽出

本節の処理では、時間軸に沿って各人物にいつ動きがあったかを調べる。

参加者は全方位カメラの周囲に位置し、会話を行う。その様子を全方位カメラによって記録することで図 2 に示すような映像を得る。記録された映像はフレームごとの画像に分解し、全てのフレームをパノラマ展開しておく。その後、以下の処理を行う。

総フレーム数を N とする。

連絡先: 清水敏雄,
 情報科学芸術大学院大学, 岐阜県不破郡垂井町綾戸 1202-5,
 090-1109-6833, clearwater0411@yahoo.co.jp



図 2: 全方位カメラ映像

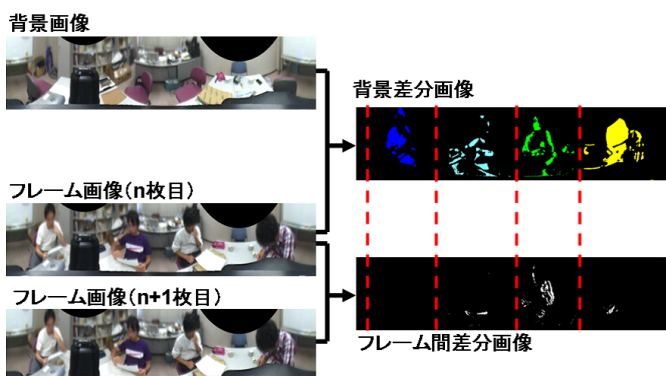


図 3: 動き量の求め方の流れ

Step1: n 枚目のフレームに対して背景差分を行う。背景の画像には、環境中に人物がいなかったときの画像を用意しておく、それを用いる。

Step2: Step1 で得た背景差分画像中の領域を、k-means 法によって人物ごとに振り分ける。k-means 法とは、個体をあらかじめ定められた数の集合に分類する手法である。

Step3: 領域の分類をもとに、各人物の位置を特定する。

Step4: n 枚目のフレームと $n+1$ 枚目のフレームとの間で差分を取る。

Step5: Step3 の人物位置に基づいて、Step4 で得たフレーム間差分画像中の領域がどの人物のものかを調べ、人物ごとに領域の面積を合計する。面積の合計が各人の動きの量となる。

以上の処理を $n = 1, 2, \dots, N$ について行うことにより、各フレームごとの人物の動きの量が得られる。

次に各人物ごとに適切な閾値を定め、各フレームにおける動きの量を、0 か 1 で表現された 2 値のデータに変換する。あるフレームにおける動きの量が閾値よりも大きければ値を 1 に、そうでなければ 0 に変換する。さらに、ノイズを除くため、2 値化したデータに対して平滑化を行う。

以上の処理により、時系列に沿って各人物にいつ動きがあったかを表したデータを得る。

2.2 同調の観測

同調関係にある人物の間では、動作に同期・類似が見られる。そこで本研究では、ある 2 人の人物が同時に動作していた時間を計測する。

前節で得た、人物にいつ動きがあったかを表すデータを用いて計測する。いま、人物 A に動きのあったフレームの集合を N_A 、人物 B に動きのあったフレームの集合を N_B とすると、人物 A と人物 B が同時に動いていた時間の総和は $|N_A \cap N_B|$ となる。この時間を同調時間と呼ぶ。この同調時間を各人物の組において計測する。

3. 実験と評価

3.1 収録環境

本研究では 2 つの会話の事例を記録し、実験を行った。それぞれを事例 1、事例 2 と呼ぶ。会話の内容は両事例とも、教官と学生数名による研究についてのミーティングであった。ある部屋に机を用意し、机の中央に全方位カメラを設置した。参加者はその机の周囲に配置された椅子に着席し、ミーティングを行った。

ビデオカメラには SONY の DSR-PD150 を用いた。撮影された映像は、サイズが 640×480 ピクセル、フレームレートは 30fps であった。事例 1、事例 2 からそれぞれ 5 分間の映像を処理した。

また実際の会話の様子を把握するため、通常のビデオカメラによっても会話を記録した。そしてその映像から人手によって動き・発話・視線の情報を調べた。時間軸に沿って各人物がいつ動き、いつ発話し、いつ視線配布を行っていたかを調べた。視線についてはその対象も記録した。これらの情報を正解データと呼ぶ。

事例 1、事例 2 における会話中の人物の様子を表 1 に示す。また、正解データから得た、各人物に動きがあった時間の累計を表 2 に示す。単位はフレームである。

表 1: 各人物の様子

	人物	各人物の様子
事例 1	A	<ul style="list-style-type: none"> 大半はプリントを読んでいる 発話は少ない。B と数回会話 プリントをめくる動作が多い
	B	<ul style="list-style-type: none"> 自分の研究について報告 発話が多い。C と頻繁に会話 身振り手振り、またはメモを取りながら会話
	C	<ul style="list-style-type: none"> まとめ役的存在 発話が多い。B に対して助言 身振り手振りを交えて会話
	D	<ul style="list-style-type: none"> 会話には参加せず、発話者を見ている 発話は皆無。まれに笑う 発話者の動きに連動して動くことが多い
事例 2	A	<ul style="list-style-type: none"> 大半はプリントを読んでいる 発話は数回。B や C と数回会話 プリントをめくる動作が多い
	B	<ul style="list-style-type: none"> 自分の研究について報告 発話が多い。皆からの質問に答える 身振り手振りを交えて会話
	C	<ul style="list-style-type: none"> B に対して助言 発話はやや多い。A や D と数回会話 身振り手振りを交えて会話
	D	<ul style="list-style-type: none"> 会話には参加せず、その時々発話者を見ている 発話は B とわずかに言葉を交わす程度 発話者のほうへ視線を向ける動作が目立つ
	E	<ul style="list-style-type: none"> B と会話 中程度の発話量 話を聞きながらよく体を揺らす

表 2: 動きのある時間の累計

	人物	動きのある時間の累計
事例 1	A	2031
	B	3560
	C	3228
	D	2109
事例 2	A	4514
	B	4840
	C	3568
	D	4955
	E	3415

3.2 動き抽出の評価

2.1 節の処理によって得た、時間軸に沿って各人にいつ動きがあったかを表すデータの正確さを、正解データとの比較により評価する。

評価の方法を説明する。各人物について 2 値化・平滑化を施したデータのうち、動きがあるフレームの集合を N とする。また動きの正解データによる、動きがあるフレームの集合を C とする。これらより式 (1), (2) を用いて適合率と再現率を求める。

$$precision(\text{適合率}) = \frac{|N \cap C|}{|N|} \quad (1)$$

$$recall(\text{再現率}) = \frac{|N \cap C|}{|C|} \quad (2)$$

適合率と再現率は一般にトレードオフの関係にあるため、この 2 者から式 (3) により F 値を算出し、その値によって評価する。

$$F \text{ 値} = \frac{2 \times precision \times recall}{precision + recall} \quad (3)$$

各人物について、2.1 節の 2 値化・平滑化のパラメータを変化させながら F 値を計算し、 F 値が最大となる時のパラメータを調べる。 F 値が最大となったときの結果を表 3 に示す (各事例の上段)。また表 3 中のベースラインは、 $|N \cap C|$ の期待値から算出した F 値を示している。比較として、各人の動き量の平均値を 2 値化の閾値とし、かつ平滑化を行わないときの F 値も計算した (各事例の下段)。

表 3: 動き抽出の評価

		$ N $	$ C $	$ N \cap C $	F 値	ベースライン
事例 1	A	1864	2035	1009	0.518	0.216
	B	4746	3558	2702	0.651	0.452
	C	3305	3228	2296	0.703	0.363
	D	2052	2114	1340	0.643	0.231
	A	1426	2035	589	0.340	0.186
	B	1344	3558	985	0.402	0.217
	C	984	3228	852	0.405	0.167
	D	1182	2114	735	0.446	0.168
事例 2	A	4600	4514	3961	0.869	0.506
	B	5049	4840	3983	0.806	0.549
	C	4900	3567	2704	0.639	0.459
	D	6242	4954	4212	0.752	0.614
	E	3375	3418	2443	0.719	0.377
	A	1753	4514	1690	0.539	0.281
	B	1526	4840	1409	0.443	0.258
	C	869	3567	721	0.325	0.155
	D	870	4954	780	0.268	0.164
	E	935	3418	871	0.400	0.163

結果から分かるように、最適なパラメータを用いて処理を行ったときには、人間関係推定を行うのに十分な精度で動きが抽出された。ただし、正解データから最適なパラメータを求め

るこのやり方は、実用化にあたっては不向きである。パラメータの自動設定は今後の課題である。

また提案手法によって得られた、動きのあるフレームの累計 ($|N|$) と、正解データから求めた動きのあるフレームの累計 ($|C|$) との相関係数を計算した。その結果、事例 1 では 0.960、事例 2 では 0.753 という強い正の相関が示された。この結果から、提案手法により抽出された各人の動きは、実際の人物の動きと高い類似性を持っていることが分かる。

3.3 同調時間の評価

2.2 節で述べたように、本研究では 2 人の人物が同時に動作していた同調時間を計測する。各事例について、各人物の組ごとに同調時間の累計を計測した。その結果は表 4 中の $|N|$ に示す通りである。時間の単位はフレームである。

表 4 中の $|C|$ は、正解データから計測した同調時間を表している。さらに $|N \cap C|$ を求め、それをを用いて F 値を算出した。結果として、最適なパラメータにより処理したときには F 値は 0.267 ~ 0.693 の値を示した。 $|N|$ と $|C|$ の相関係数を求めたところ、事例 1 では 0.878、事例 2 では 0.751 という強い正の相関が示された。この結果から、提案手法により人間関係の推定が可能だといえる。

表 4: 同調時間の評価

		$ N $	$ C $	$ N \cap C $	F 値	ベースライン
事例 1	A-B	1258	805	340	0.330	0.097
	A-C	1016	803	380	0.418	0.082
	A-D	539	503	139	0.267	0.050
	B-C	2003	1356	830	0.494	0.169
	B-D	1097	652	258	0.295	0.090
	C-D	953	984	550	0.568	0.094
	事例 2	A-B	2493	2449	1691	0.684
A-C		2386	1965	1354	0.622	0.244
A-D		3554	2756	2185	0.693	0.328
A-E		1297	1306	656	0.504	0.165
B-C		2943	1964	1196	0.487	0.254
B-D		3471	2576	1825	0.604	0.330
B-E		2047	1767	1067	0.560	0.203
C-D		3870	1990	1371	0.468	0.279
C-E		1863	1081	627	0.426	0.151
D-E		2285	1767	1016	0.501	0.224

3.4 視線配布との比較

得られた同調時間を、会話中の視線配布の状況と比較する。会話中に交わされた視線を、視線を向けた人の役割・向けられた人の役割に応じて、以下のように分類する。

All: 役割に関係なく、会話中に交わされた視線全て

S L: 話し手から聞き手への視線

L S: 聞き手から話し手への視線

O S: 傍観者から話し手への視線

O L: 傍観者から聞き手への視線

話し手 (Speaker) とは、発話と視線配布を同時に行っている人物のことである。聞き手 (Listener) は、話し手から視線を向けられている人物のことである。また傍観者 (Observer) とは、会話に加わっていない人物のことである。

各人物について、視線の対象、視線の種類、その発生時間の累計を調べた。各視線の発生累計時間と同調時間の相関係数を算出した結果が表 5、表 6 である。時間の単位はフレームである。この結果から、事例 1 では All, S L, L S において

強い正の相関が示されている。このことから、事例1においては話し手と聞き手の間で同調が顕著であったことが分かる。また事例2では、O Sにおいて強い正の相関が示されている。このことから事例2においては、話し手と傍観者の間で同調が顕著であったことが分かる。以上のように、提案手法においては会話参加者のみでなく傍観者も含めた人間関係を得ることが可能である。

表 5: 各視線との相関係数 (事例 1)

人物の組	同調時間	All	S L	L S	O S	O L
A-B	1258	2491	1164	436	0	57
A-C	1016	2199	152	0	167	688
A-D	539	221	0	0	70	6
B-C	2003	4261	1292	536	113	13
B-D	1097	886	0	0	441	93
C-D	953	1735	0	0	269	49
同調時間との相関係数	1.000	0.927	0.816	0.833	-0.102	-0.137

表 6: 各視線との相関係数 (事例 2)

人物の組	同調時間	All	S L	L S	O S	O L
A-B	2493	850	53	53	3	6
A-C	2386	1378	228	189	178	78
A-D	3554	385	114	0	49	7
A-E	1297	204	18	0	44	40
B-C	2943	2151	684	279	73	31
B-D	3471	2312	170	3	353	103
B-E	2047	1625	667	48	0	38
C-D	3870	1227	139	101	170	11
C-E	1863	257	0	0	9	0
D-E	2285	517	26	0	0	9
同調時間との相関係数	1.000	0.445	0.089	0.196	0.560	0.083

4. 考察

提案手法の利点と欠点について議論する。本研究では、会話における人間関係の分析について、新たな可能性を模索した。従来研究においては、会話や視線配布の分析が主であったのに対し、本研究では計算機を用いた同調現象の観測による分析を試みた。提案手法が持つ利点と欠点について考察する。

利点として、まず参加者と傍観者を含んだ人間関係の推定が可能となる点が挙げられる。例えば、これまで行われてきた発話のみからの分析では、発話しない人物についての関係を知ることはできなかった。しかし提案手法による分析では、参加者だけでなく傍観者も含めた人間関係を得ることができる。次に機材の簡易さが挙げられる。これまでも人間関係推定の自動化を試みた研究は存在したが、それらは大掛かりな機材を必要とする場合があった。それに対して、提案手法が必要とする機材は全方位カメラ 1 台のみであり、非常に手軽である。そのため様々な場所・状態での実験が可能となる。最後に、システムがシンプルな技術から構成されている点が挙げられる。用いられている技術は、画像間で差分を取ったり、クラスタリングを行ったりするなど、比較的シンプルなものである。そのため、カメラと基本的な画像処理のできる計算機さえあればシステムが実現可能であり、また実験環境の違いにも柔軟に対応できる。

一方で提案手法の欠点として、画像処理に用いるパラメータをいかにして設定するかという問題がある。2 値化・平滑化のパラメータは、処理の結果を大きく左右する重要な要因であ

る。本研究では人手によって各人物の動きを調べ、最もそれに近い処理結果の得られるパラメータを設定した。しかし、実用化を目指すにあたってこの方法は現実的ではなく、パラメータの自動設定の方法が必要となる。

5. おわりに

本研究では、全方位カメラによる同調現象の観測に基づく人間関係の推定手法を提案した。全方位カメラによって会話の様子を記録し、得られた全方位映像から画像処理によって人物の動きを取り出した。そして取り出された動きから同調現象を観測することで人間関係を抽出した。処理の精度を評価した結果、提案手法によって十分に人間関係の推定が可能であることが示された。さらに得られた人間関係を実際の会話と比較することで、提案手法による人間関係の特徴を分析した。その結果、提案手法による人間関係には、参加者・傍観者を含む全員の関係が反映されていることが示された。

人文科学の分野においては、人間同士でのやり取りを同調現象の観点から分析することが多く行われてきた。本研究の特徴は、人文科学の分野で行われてきた同調現象の分析を、画像処理技術の観点から実現しようと試みたことにある。また、提案手法はシンプルな技術の組合せで実現されており、技術そのものよりもその使い方に新しさがある。提案手法はまだ初期段階であり、今後さらなる改良を加えてシステムを向上させていくことで、より有効な同調現象の観測が可能となりうる。

人間同士の会話について研究・分析することは、より効率的な組織運営を手助けし、より創造的な活動の実現に寄与する。本研究がそれらコミュニケーションの研究に貢献することを期待している。

参考文献

- [1] 長岡千賀: 対人コミュニケーションにおける非言語行動の 2 者相互影響 (2003), (博士論文, 大阪大学大学院人間科学研究科).
- [2] 大坊郁夫: しぐさのコミュニケーション 人は親しみをどう伝えあうか (1998), (サイエンス社).
- [3] 坊農真弓: 対面コミュニケーションにおける視点概念 (2004), (博士論文, 神戸大学).
- [4] 大塚和弘・竹前嘉修・大和淳司・村瀬洋: 複数人物間の対面会話を対象としたマルコフ切り替えモデルに基づく会話構造 (2006), (情報処理学会論文誌).
- [5] Stiefelbogen, R., Chen, X., Yang, J.: Capturing Interaction in Meeting with Omnidirectional Camera (2005), (International Journal of Distance Education Technoloies).
- [6] Rui, Y., Gupta, A., Cadiz, J. J.: Viewing Meeting Captured by an Omni-Directional Camera (2001), (Proceedings fo ACM CHI2001).
- [7] Sun, X., Roote, J., Kimber, D. J., Manjunath: Panoramic Video Capturing and Compressed Domain Virtual Camera Control (2001), (ACM Multimedia).