

体動を用いた非記号的な対話特徴量の検討

井上 敬斗[†]近藤 一晃[‡]下西 慶[‡]中村 裕一[‡][†] 京都大学工学部[‡] 京都大学学術情報メディアセンター

1. はじめに

非言語のやり取りが対話において大きな役割を担っていることは良く知られている。うなずきや顔の方向、ジェスチャなどの体の姿勢に基づいた振る舞いはその典型例である。これらは姿勢という物理量の中から、特有の意味や機能が見出されたパターンを記号として取り出したものである。しかし、我々がコミュニケーションを取る際には、他人のちょっとした動きやしぐさからも意図や思惑を感じる。記号として分類されていない弱いパターンにも、対話の重要な構成要素が含まれていると考えられる。もしそのようなパターンを取り出すことができれば、対話状況の自動認識や定量評価、遠隔対話時におけるインタラクションチャンネルのデザインなどへ応用できる。

そこで本研究では、対話参加者の姿勢を記号化することなく、単純な量（体動）として用いた非記号的な特徴量「対話参加者間の体動同期度」を提案する。実対話における体動の生起や、体動同期度がどのような対話状況を反映するかについて分析し、体動を用いることの有効性について検証する。

2. 体動同期度

2.1 設計指針

対話とは人物間でメッセージをやり取りする行為であり、どのようなメッセージをどのような時系列で送り・受け取りあうのかによって構成される。記号化された対話特徴であるかどうかに関わらず、この原理が同様に働いていると考えるのが自然である。このようなメッセージやり取りに着目した研究は過去にも数多く存在する。例えば [1] では発話や身振りのタイミングが対話にもたらす影響について報告している。しかし記号的な特徴を取り出してから分析を行っており、本研究のように非記号的な特徴を用いてはいない。単純な量である体動は必ずしも明確な意味を持たないため、メッセージの内容には言及せず、人物間における時間関係のみに着目することが望ましい。本研究では以下のような対話参加者間の体動の時間近接に着目し、まとめて「体動の同期」と呼ぶ。

(1) ある対話参加者 A の体動に続いて他の参加者 B の体動が現れた場合。(2) 対話参加者 A と B の体動が同時に現れた場合。

(1) は、A が発したメッセージを B が受け取った、またその応答として B がメッセージを発した、ことの表出である可能性が高い。また (2) は、A と B が同時にメッセージを送りあった、もしくは直前に受け取ったメッセージに対して同時に反応した、ことの表出である可能性が高い。すなわち、体動の同期度合いは、対話におけるメッセージのやり取りの密度、言い換えれば対話エンゲージメントを反映した特徴量になることが期待される。

2.2 抽出方法

体動同期度の抽出は以下のステップで構成される。

1. 画像認識を用いた人体関節位置の推定

対話参加者をカメラで撮影した画像から関節位置推定を行う。画像認識を採用した理由は、体動という大まかな量を扱うのであれば、加速度センサ等を用いた正確な計測は必ずしも必要ないと考えたからである。また対話状況の自動認識や遠隔会議への応用の幅が広がるという副次的効果も望める。関節位置には画像上での二次元位置 (u, v) を用いる。三次元位置 (x, y, z) の推定値は奥行き方向に不安定な上に、身体基準点が隠れてしまったときに著しく精度が落ちるからである。なお二次元の関節位置でも無視できない量の高周波ノイズがどうしても含まれてしまう。そこで人間の動作周波数には上限があることを利用し、ガウス平滑化により低周波成分のみを取り出してノイズを軽減する。

2. 関節位置の移動に基づいた体動量の導出

時間隣接フレーム間における関節位置 $(u(t), v(t))$ の移動量を体動量 $m(t)$ と定義する。

$$m(t) = \sqrt{(u(t+1) - u(t))^2 + (v(t+1) - v(t))^2} \quad (1)$$

できるだけ意味を排除するために、移動方向については考慮せずに、単純な移動量を計算している。また上式には含まれていないが、カメラ-人物間の距離に依存した値にならないよう、人物肩幅の時間平均値を用いて正規化する。

3. 対話参加者間における体動量の時間同期度の計算

対話参加者 A, B 間における体動の同期度 $S_{AB}(t)$ を、A, B の時系列体動量 $m_A(t)$ と $m_B(t)$ のコサイン類似度、および体動量の大きさをを用いて以下のように定義する。

$$S_{AB}(t) = \max_{\Delta t} (C \sqrt{V_A V_B}) \quad (2)$$

$$C = \frac{\sum_{k=t-N/2}^{t+N/2} m_A(k) m_B(k + \Delta t)}{\sqrt{\sum_{k=t-N/2}^{t+N/2} m_A^2(k)} \sqrt{\sum_{k=t-N/2}^{t+N/2} m_B^2(k + \Delta t)}} \quad (3)$$

$$V_A = \max(0, \frac{\sum_{k=t-N/2}^{t+N/2} (m_A(k))}{N+1} - e_A) \quad (4)$$

$$V_B = \max(0, \frac{\sum_{k=t-N/2}^{t+N/2} (m_B(k + \Delta t))}{N+1} - e_B) \quad (5)$$

N は体動同期度を計算するときに着目する時間幅を、 Δt は $m_A(t)$ と $m_B(t)$ の間で時間連続して発生する体動を捉えるための時間シフト量を示す。 e_A, e_B は微小な体動量をノイズとみなし有効な値を取り出すための閾値である。これらの値より小さな体動量は 0 とみなされる。相互性を考慮した定量化を行うため、人物 AB の体動がどちらも十分に大きいときに体動同期度も大きくなるよう、 m_A および m_B の強さを積の形で導入している。

3. 実験

3.1 一対一対話を対象とした体動同期度の抽出

体動を用いることの有効性を分析するために、実際の対話シーンを対象に検証を行った。大学生 2 名に 4 つの話題を与え、それぞれについて 5 分程度で意見交換する

A non-symbolic dialogue feature using body movement K. Inoue, K. Kondo, K. Shimomishi, Y. Nakamura (Kyoto Univ.)

よう指示するとともに、その様子を正面に設置したカメラで撮影した。関節位置推定には media pipe[2] を使用し、関節位置が安定して取得できていた両肩の midpoint を体動量の計算に用いた。

二名の体動量および二名間の体動同期度を抽出した結果を図1に示す。690秒付近の区間(図1の右橙色枠)では二名の体動量がほぼ同時(約0.4秒差)で大きな値を示している。また630秒付近の区間(図1の左橙色枠)では、被験者Bの体動が発生してからおよそ1秒後に被験者Aの体動が発生している。これらの区間に対する体動同期度は共に高い値を示していることがわかる。一方で605秒付近の区間(図1の左黄色枠)では被験者ABの体動はともに非常に小さい。655秒付近の区間(図1の右黄色枠)では被験者Aの体動は有意に発生しているものの被験者Bではほとんど見られない。このような区間では体動同期度は非常に小さな値として現れている。以上のことから、2.1節で述べた設計指針に沿った体動同期度が、(2-5)式によって抽出できることが確認された。

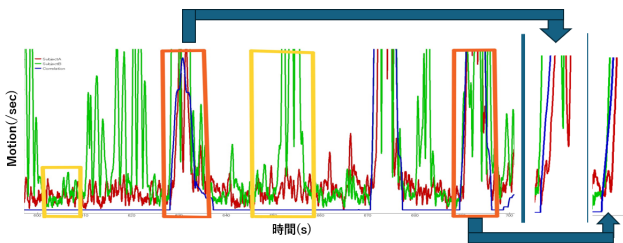


図1: 2人の被験者の体動量とその同期度(左)と青線引き上がり部分の拡大図2枚(横軸: 時間, 縦軸: 体動量)(被験者の体動量: 赤と緑, 同期度: 青)

3.2 体動同期度と対話エンゲージメント

体動同期の度合いを数値化することで対話に対する注意の強さ、言い換えれば対話エンゲージメントを反映した特徴量になると考え、抽出された体動同期度と対話エンゲージメントについて分析した。図1の左オレンジ枠および右オレンジ枠にあたる、630秒付近および690秒付近における被験者ABの様子を図2に示す。

図2(左)のシーンでは、困惑した笑いととも疑問を述べる人物B(左話者)に対して、人物A(右話者)がすぐさま苦笑いで応えていた。人物Bから人物Aへのメッセージ送信、およびそれを受信した人物Aによる人物Bへの応答メッセージが明確に現れており、強いやり取りが見て取れた。図2(右)のシーンでは、人物Bの直前の発話に対して、両人物がほぼ同時に笑いだしていた。文脈が共有されていることが伺え、ここでもやり取りの強さが見て取れた。また人物BとAの間におけるメッセージ送受信・応答と、人物Bの自己とのやり取りが並行して発生している興味深い状況でもある。上記以外の場面でも、体動同期度が高い場面では強いメッセージのやり取りが行われており、強い対話エンゲージメントの傾向が見られた。これらの結果から体動同期度によって対話エンゲージメントの強さを定量化できる可能性が確認された。



図2: 左:629s及び630sでの被験者の様子 右690sでの被験者の様子

3.3 体動量と記号特徴との相違

既存の記号的な特徴に対して体動が有効な追加特徴になっているかを調査した。図3左に頷きの有無で区間を場合分けしたときの体動量ヒストグラムを示す。頷きの有無でヒストグラムに大きな差はないため、体動は頷きとほぼ無相関であったことがわかる。図3右に表情変化の有無で区間を場合分けしたときの体動量ヒストグラムを示す。表情変化が生じているときにやや体動が大きく現れる傾向があるが、表情変化の有無で体動量分布を明確に分離できるほどではない。以上の結果から、頷きや表情変化といった既存の記号的な特徴で体動特徴を置き換えることは難しいことが示された。

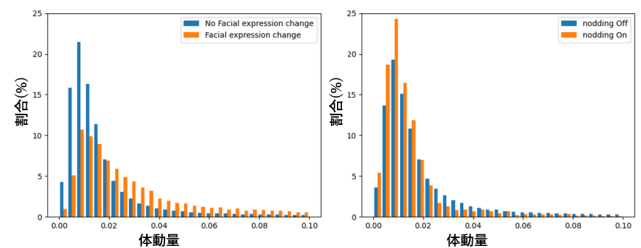


図3: 左: 表情変化の有無での体動ヒストグラム 右: 頷きの有無での体動ヒストグラム (赤: 有り, 青: 無し)

4. おわりに

本研究では、非言語コミュニケーションにおける非記号的な特徴として体動に着目し、対話参加者間での同期度を新たな特徴量として提案した。二名対話から体動同期度を抽出した結果、対話エンゲージメントを示す指標としての有効性が確認された。また体動は頷きや表情変化といった記号的な特徴では置き換えられないことも確認された。今後は、様々な対話設定、例えば多人数対話や遠隔対話における対話エンゲージメントとの関連を調査する予定である。

参考文献

- [1] 高杉将司・山本知仁・武藤ゆみ子・阿部浩幸・三宅美博 コミュニケーションロボットとの対話をういた発話と身振りのタイミング機構の分析 計測自動制御学会論文誌 Vol.45, No.4, pages 215-223, 2009.
- [2] media pipe <https://developers.google.com/mediapipe>