映像からの歩行者の意図推定に必要な要因の検討

山添 大丈[1,a)]　満上 育久[1]　八木 康史[1]

概要：我々は，人の歩き方 (歩容) からその人の意図を推定することを目指し，研究を進めている．本稿では，歩き方からの意図推定手法の実現可能性について検討するため，人は，歩く様子を撮影した映像から，その人の意図を推定できるのかどうかを調査する．また，歩き方から意図を推定するにあたって，どういった部分に着目しているかについても分析を行う．３種類の意図（「ついていく」，「向かう」，「逃げる」）を含む映像列を用いて被験者実験を行い，a) 歩き方からの意図推定は可能，b) 約 70% 程度の正解率，c) 歩き方のうち，形状 (シルエット) やテクスチャが意図推定に重要，d) 頭部以外にも意図推定に有用な情報が含まれることが分かった．これらの結果は，我々の目指す，歩き方からの意図推定手法の実現可能性を示すものといえる．

1. はじめに

我々は，人の歩き方 (歩容) から意図や興味といったその人の内的状態を推定することを目指し，研究を進めている．意図や興味などを推定するにあたっては，頭部方向や注視方向が重要と考えられるため，現在は，歩き方から頭部方向・注視方向を推定する手法について，検討を進めているところである [1], [2], [3].

一方で，意図や興味の推定において，一般的には，頭部方向や注視方向が重要と考えられているが，歩き方からの意図・興味推定において，必ずしも，頭部方向や注視方向が重要とは限らない．

そこで，本稿では，特に歩き方からの意図推定に着目し，人が映像を観察して意図を判断する被験者実験を通じて，人は他者の振る舞い (歩行) を撮影した映像から，その人の意図推定が可能であるか，可能であるとすれば，どういった部分に着目して意図の推定を行っているか，について検討する．

行動の観測による人の意図推定については，様々な研究が行われている [4], [5], [6], [7], [8], [9]．これらの研究では，人が他人の動きから意図を推定する際のメカニズムを調べることを主な目的としている．そのため，人の動きを映像として提示するだけでなく，バイオロジカルモーションのような人の動作を簡略化した形で提示することで，どういった要因が意図推定に重要かを調査している [8], [9].

例えば，McAleer ら [8] は，２人の人物が６種類の意図 (Chasing, Fighting, Flirting, Following, Gurading, Playing) を演じているシーンについて，各人物の動きを「シルエット」もしくは「重心位置」として提示した場合の意図推定の正解率について分析を行っている．結果から，人は６種類の意図を推定できること (正解率約 70%)，「シルエット」と比較すると正解率は低下するものの，「重心位置」として提示した場合でも，ある程度 (40-60%) は意図が推定できることが示されている．また，この研究では，視点についても検討されており，上からの観測 (overhead view) に比べて、横からの観測 (side view) では，意図の推定精度が低下する，といった結果が得られている．

これらの従来研究では，人が他者の意図を推定するためのメカニズムについて調査・検討することが主な目的であった．そのため，これまでは，映像からの撮影対象の意図推定手法に向けた議論はなされてこなかった．

そこで，本稿では，特に歩き方からの意図推定に着目し，人が他者の振る舞い (歩行) を撮影した映像から，その人の意図推定が可能どうか，また，どういった部分に着目して意図推定を行っているかについて評価するとともに，意図推定手法を実現に向けて，どういった部位を観測し，どういった特徴を取得すべきかについて，実験結果をもとに考察する．

2. 実験の概要

まず本稿で扱う意図と実験に使用する映像列について説明する．本稿の実験で用いる映像列は，図１に示すような１０台のプロジェクタ・スクリーンからなる約２０ｍ×４ｍの環境において，ゲームをしている様子をスクリー

[1] 大阪大学
Osaka University, Toyonaka, Osaka, 670-0043, Japan
[a)] yamazoe@osipp.osaka-u.ac.jp

図 1 意図映像の撮影環境

ン上に設置された Microsoft Kinect により撮影したものである．

このゲームでは，プレイヤーはスクリーン上に表示されるクマについていくように指示されている．ハチが出てきたときには，プレイヤーは逃げるように指示されている．また，クマが移動するスクリーン中の環境には，いくつかの入口があり (図 1 のスクリーン上の黒い部分)，クマは時々入口に入ったり，別の入口から出てくる，といった振る舞いを行っている．そのため，ゲーム中，プレイヤーはクマについて行く，入口に入って消えたクマを探し，遠くに現れたクマに向かう，ハチから逃げる，といった振る舞いをすることになる．ここで，ゲーム中のプレイヤーの意図のうち，本稿の実験で扱う意図を以下の 3 種類とする．

ついていく (F): クマを横に見ながら，クマについていっている状態
向かう (R): クマがプレイヤーから遠い位置に現れ，クマに近づいている状態
逃げる (E): ハチから逃げている状態

本稿では，これら 3 種類の意図を，プレイヤーの様子を撮影した映像列から推定できるかどうか，推定するにあたってはどういった要因が重要かについて検討を行う．

2.1 本稿で検討する要因

映像から意図を推定するにあたっては，様々な要因が影響すると考えられるが，本稿では，以下の 3 種類の要因について検討する．

1) 時間的要因（動作開始時点の有無）

一つ目の要因としては，歩行の開始から歩行中までの一連の歩行動作のうち，どの区間が意図推定に重要かについて検討する．それとも，ある区間の映像が得られれば十分なのかを検討する．

具体的には，一連の歩行動作を「歩行開始時」,「歩行中」の 2 つの段階に分割し，以下の 3 条件について比較する (実験 1).

1-a) 歩行開始時のみを提示（「歩行開始時」条件）
1-b) 歩行中のみを提示（「歩行中」条件）
1-c) 歩行動作全体を提示（「歩行全体」条件）

2) テクスチャ・形状情報

2 つ目の要因としては，意図推定におけるテクスチャや形状情報の重要性について検討する．これらを検討するため，以下の 3 条件について比較実験を行う (実験 2).

2-a) 映像列をそのまま提示（「ビデオ」条件）
2-b) 人物の動作をシルエットとして提示（「シルエット」条件）
2-c) 人物の動作を重心点のみで提示（「重心」条件）

ここで,「シルエット」条件のビデオについては，Kinect で撮影されるカラー画像，デプス画像の両方を用いて，背景差分手法により，ビデオを作成している．また,「重心」条件のビデオについては,「シルエット」条件のビデオのシルエット領域の重心点を計算し，x 座標のみを用いて白い丸を描画している (y 座標は画像の真ん中).

3) 頭部情報の有無

三つ目の要因としては，意図推定における頭部方向の有無について検討する．これらの検討のため，以下の 3 条件について比較実験を行う（実験 3).

3-a) 映像列をそのまま提示（「体全体」条件）
3-b) 人物の頭部のみを提示（「頭部のみ」条件）
3-c) 人物の頭部を隠して提示（「頭部なし」条件）

3. 実験

前節で述べた 3 種類の要因について，映像からの意図推定における重要性について検討するため，以下の 3 種類の実験を行った．

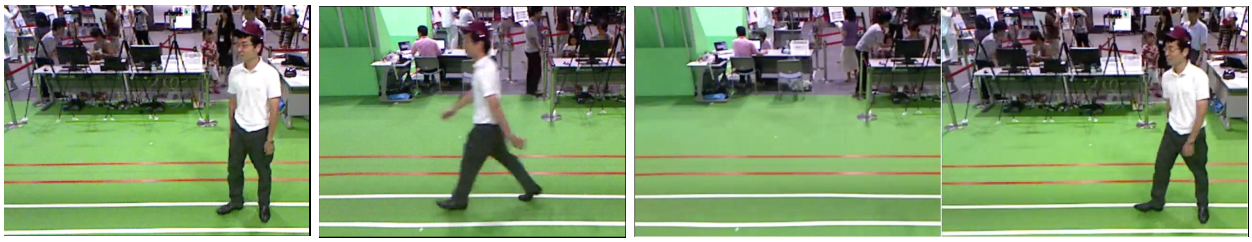
3.1 実験手順

まず，全実験に共通する実験手順について説明する．被験者はゲーム中のプレイヤーの歩行の様子を撮影した１０秒程度の映像 (図 2-4) を見てから，その映像中のプレイヤーが「ついていく」,「向かう」,「逃げる」の 3 種類の意図のうちどれであるかを選択する．意図の選択にあたっては，被験者には，直観的に意図を選択するように指示するとともに，映像提示終了後１０秒間だけ，回答画面が表示され，その後自動的に次のビデオに移るようになっている (時間切れの場合は未回答として分析からは除外).

6 名のプレイヤーの 3 種類の意図の映像について，図 2-4 に示すビデオパターン (計 15 種類) を準備した ($6 \times 3 \times 15 =$ 計 270 パターン). ビデオの提示順は，被験者ごとにカウンターバランスを取っており，9 名の被験者 (21–33 歳の男性) で実験を行った．

3.2 実験 1 と結果

実験 1 においては，以下の仮説について検証するため，



「歩行開始時」条件)　「歩行中」条件　「歩行全体」条件

図 2　実験 1 用の映像列

「ビデオ」条件　「シルエット」条件　「重心」条件

図 3　実験 2 用の映像列

図 2 に示した 3 種類のビデオパターンにおける意図の推定精度について比較，評価する．

仮説 1： 歩行動作のうち，どの区間 (動き始め，歩行中) を観察しても，映像からの意図の推定精度は変化しない．

結果: 表 1-3 と図 5 に「歩行開始時」,「歩行中」,「歩行全体」条件についての全被験者の回答結果と正解率を示す．結果より，全ての意図について，期待値 (33%) 以上の正解率が得られていることがわかる．

次に，の正解率について,「意図の種類」と「時間的要因」に関して 2 要因分散分析を行った．結果 (表 4) と事後検定 (Tukey 法) の結果より，意図の種類について正解率に有意差があり,「ついていく」の正解率 (91.3%) が有意に高いことが分かった．さらに，提示区間については有意傾向 (p=0.059) があり,「歩行中」に比べて,「歩行動作全体」の正解率が有意に高いことが分かった (pj0.05)．一方で,「歩行開始時」と「歩行動作全体」の比較では，正解率が有意差がなかった (p=0.74)．

以上より，仮説 1 は棄却され，観測できる歩行映像の区間によって，映像からの意図の推定精度 (正解率) が変化することが分かった．

3.3　実験 2 と結果

実験 2 においては，以下の仮説を検証するため，図 3 に示した 3 種類のビデオ（「ビデオ」,「シルエット」,「重心」条件）における意図の推定精度について比較，評価する．

仮説 2： 歩行動作の提示において，テクスチャ・形状情報をなくしても映像からの意図の推定精度は変化しない．

表 1　「歩行開始時」の回答結果

	ついていく	向かう	逃げる
ついていく	0.89	0.07	0.04
向かう	0.15	0.78	0.07
逃げる	0.07	0.11	0.81

表 2　「歩行時」の回答結果

	ついていく	向かう	逃げる
ついていく	0.91	0.06	0.04
向かう	0.20	0.76	0.04
逃げる	0.06	0.33	0.61

表 3　「歩行区間全体」の回答結果

	ついていく	向かう	逃げる
ついていく	0.94	0.04	0.02
向かう	0.09	0.81	0.09
逃げる	0.07	0.11	0.81

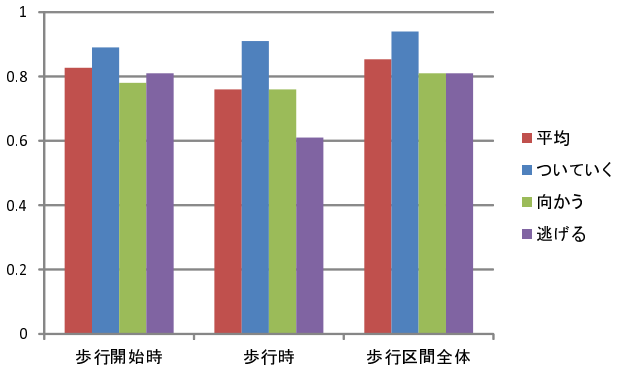
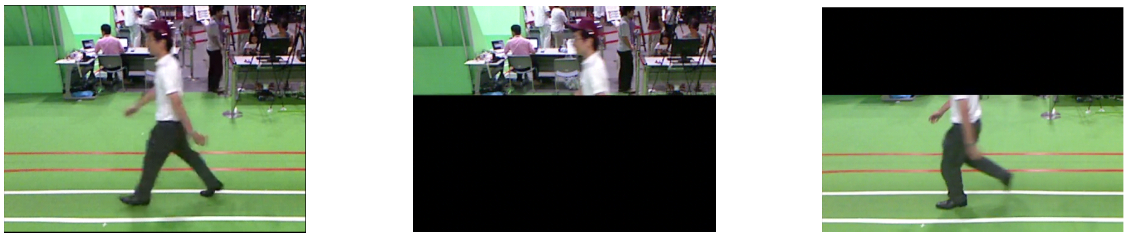


図 5　条件ごとの正解率 (実験 1)

結果: 表 5-7 と図 6 に「ビデオ」,「シルエット」,「重心」条件についての全被験者の回答結果と正解率を示す．



3-a) 3-b) 3-c)

図 4 実験 3 用の映像列

表 4 2 要因分散分析結果 (実験 1)

	F	p	
意図の種類	8.56	<0.01	**
時間的要因	2.85	0.059	*
交互作用	1.45	0.217	

表 5 「ビデオ」条件の回答結果

	ついていく	向かう	逃げる
ついていく	0.90	0.06	0.04
向かう	0.18	0.77	0.06
逃げる	0.06	0.22	0.71

表 6 「シルエット」条件の回答結果

	ついていく	向かう	逃げる
ついていく	0.76	0.17	0.07
向かう	0.18	0.75	0.07
逃げる	0.17	0.30	0.54

表 7 「重心」条件の回答結果

	ついていく	向かう	逃げる
ついていく	0.75	0.17	0.08
向かう	0.27	0.52	0.20
逃げる	0.28	0.44	0.29

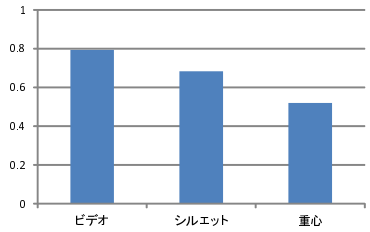


図 6 条件ごとの正解率 (実験 2)

正解率について，「テクスチャ・形状情報 (ビデオ・シルエット・重心点)」に関して 1 要因分散分析を行った．結果 (表 8) と事後検定 (Tukey 法) の結果より，「テクスチャ・形状情報」について正解率に有意差があり，3 条件間の全てで正解率に有意差があることが分かった ($p<0.01$)．

以上より，仮説 2 は棄却され，歩行動作の提示方法によって，映像からの意図の推定精度 (正解率) が変化することが確認された．

表 8 1 要因分散分析結果 (実験 2)

	F	p	
Display	29.34	<0.001	**

3.4 実験 3 と結果

実験 3 においては，以下の仮説を検証するため，図 4 に示した 3 種類のビデオ (「体全体」,「頭部のみ」,「頭部なし」条件) における意図の推定精度について比較，評価する．

ここで，図 4 では，「ビデオ」条件のみの例を示しているが，「シルエット」条件についても実験を行っており，結果として，「頭部情報の有無」と「テクスチャ情報 (「ビデオ」と「シルエット」のみ)」の 2 要因について分析を行った．

仮説 3: 歩行動作の映像の提示において，頭部のみ，頭部以外だけを提示しても映像からの意図の推定精度は変化しない．

結果: 表 9-11 と図 7 に「体全体」,「頭部のみ」,「頭部なし」条件についての全被験者の回答結果と正解率を示す．

次に，正解率について，「頭部情報の有無」と「テクスチャ情報 (「ビデオ」と「シルエット」のみ)」に関して 2 要因分散分析を行った．結果 (表 12) と事後検定 (Tukey 法) の結果より，「頭部情報の有無」,「テクスチャ情報」ともに正解率に有意差があり，「頭部情報の有無」,「テクスチャ情報」についての交互作用が見られた．

「頭部情報の有無」については，「体全体」と「頭部なし」,「頭部のみ」と「頭部なし」間で正解率に有意差があり ($p<0.001$),「頭部なし」の場合に正解率が有意に低下することが分かった．また，「頭部情報の有無」,「テクスチャ情報」の交互作用については，「ビデオ・体全体」と「ビデオ・頭部のみ」では有意差がないものの，「シルエット・体全体」と「シルエット・頭部のみ」では正解率に有意差があり，シルエット条件のほうがより正解率が低下していることが確認された．

このことから，仮説 3 は棄却され，観測できる人物の領域（人物全体・頭部のみ・頭部なし）によって，映像からの意図の推定精度 (正解率) が変化することが分かった．

表 9 「全体」条件の回答結果

	「ビデオ」			「シルエット」		
	ついていく	向かう	逃げる	ついていく	向かう	逃げる
ついていく	0.90	0.06	0.04	0.76	0.17	0.07
向かう	0.18	0.77	0.06	0.18	0.75	0.07
逃げる	0.06	0.22	0.71	0.17	0.30	0.54

表 10 「頭部のみ」条件の回答結果

	「ビデオ」			「シルエット」		
	ついていく	向かう	逃げる	ついていく	向かう	逃げる
ついていく	0.94	0.05	0.02	0.57	0.30	0.12
向かう	0.09	0.80	0.11	0.18	0.73	0.09
逃げる	0.07	0.22	0.70	0.12	0.33	0.55

表 11 「頭部なし」条件の回答結果

	「ビデオ」			「シルエット」		
	ついていく	向かう	逃げる	ついていく	向かう	逃げる
ついていく	0.80	0.15	0.06	0.74	0.19	0.06
向かう	0.11	0.76	0.13	0.15	0.80	0.06
逃げる	0.23	0.37	0.40	0.32	0.44	0.23

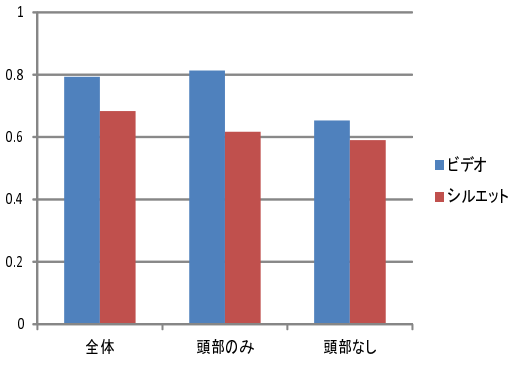


図 7 条件ごとの正解率 (実験 3)

表 12 2 要因分散分析結果 (実験 3)

	F	p	
頭部情報	8.56	<0.01	**
テクスチャ	2.85	0.059	*
交互作用	1.45	0.217	

4. 考察

4.1 実験に関する考察

まず，実験 1 の結果より，人は歩行中の人物を撮影した映像列からその人の意図を推定できることがわかった．[8] では，同様の条件での正解率が約 70% と報告されており，推定する意図の種類が異なるため，単純には比較できないものの，同等の結果が得られたといえる．

3 種類の意図ごとの正解率の比較より，「ついていく」，「向かう」と比べて，「逃げる」を推定することが難しいことがわかった．この原因としては，「逃げる」はプレイヤによって逃げ方が異なり，あるプレイヤは後ろ (ハチ) を確認しながら逃げるのに対し，別のプレイヤはハチが出てきたのを確認した後は，後ろを振り向くことなく逃げていた．「歩行時」条件の回答結果 (表 2) において，「逃げる」が「向かう」に判断されがちなことからも，以上の原因の可能性を示すものといえる．

「時間的要因」における重要度については，実験 1 では，「区間全体」，「歩行開始時」と比べて，「歩行時」条件での正解率の低下傾向 ($p < 0.1$) が見られており，「歩行開始時」が意図推定において重要である可能性が示された．ただし，「歩行時」条件においても，平均で 7 割程度の正解率が得られており，基本的には，いずれの区間を観測したとしても，意図の推定が可能といえる．

次に，実験 2 の結果について考察する．結果より，「ビデオ」，「シルエット」，「重心」の順に正解率が低下しており，映像からの意図推定において，テクスチャや形状は重要であるといえる．また，「シルエット」と「重心」については，[8] でも比較・実験がなされている．推定する意図の種類が異なるため，単純には比較できないものの，「シルエット」，「重心」条件でそれぞれ約 70%, 約 40% の正解率となっており，同様の結果となっていることがわかる．

実験 3 では，頭部の重要性について実験した．「頭部の有無」については，「体全体」と「頭部なし」，「頭部のみ」と「頭部なし」の間で正解率に有意差があり，「頭部なし」の場合に正解率が有意に低下することが分かった．このことは，意図推定における頭部領域の重要性を示すものと言える

一方で，「頭部情報の有無」，「テクスチャ情報」については交互作用も見られている．「ビデオ・全体」と「ビデオ・頭部のみ」では有意差がないものの，「シルエット・体全

体」と「シルエット・頭部のみ」では正解率に有意差があり，「シルエット」条件においては，「頭部のみ」条件で正解率が低下している．

このことから，頭部は意図推定に重要であるとともに，頭部以外の動きも，意図推定において意味があるといえる．「ビデオ」条件では，テクスチャを含む頭部領域が観測できているため，頭部の観測だけからでも詳細な頭部運動が推定できる．頭部領域に意図推定に十分な情報が含んでいるため，頭部以外の動きの影響がなくなっていると考えられる．これに対し，「シルエット」条件では，頭部領域が観測できていても，詳細な頭部運動を推定することは難しいため，頭部だけでは意図推定に十分な情報が得られず，そのため，頭部以外が観測できることで，正解率が向上していると考えている．

4.2 映像からの意図推定に向けた考察

ここでは，映像からの意図推定手法を目指す観点から，改めて実験結果について考察する．

まず，実験１では，人は歩き方から約 80% 程度の精度で，その人の意図を推定できることを確認した.．また，歩き始め時点での観測が望ましいものの，歩行中の観測が得られれば，約 70% 程度では，意図推定が可能であることも確認した．これらの結果は，人が歩く様子をカメラで観測することによる映像からの意図推定手法の実現可能性を示すものといえる．

次に，実験２では，人が意図を推定する際のシルエット形状やテクスチャの重要性について確認した．特に，歩行の様子がシルエットとして提示されたとしても，約 70% 程度で，意図推定が可能なことがわかっている．我々は，現在のところ，歩き方をシルエット画像列として処理することを考えているが [2]，以上の結果は，シルエット画像を利用する妥当性を示すものといえる．

5. まとめ

本稿では，人が歩く様子を撮影した映像列からその人の意図を推定することができるか，どういった部分に着目して意図を推定しているかについて，被験者実験を行い，その結果について述べた．被験者実験では，クマについていったり，ハチから逃げるゲームにおいて撮影された映像を用いて，３種類の意図（「ついていく」，「向かう」，「逃げる」）と３種類の要因（「時間的要因」，「テクスチャ・形状情報」，「頭部情報の有無」）について，評価を行った．実験結果より，３種類の意図について，a) 歩き方からその人の意図推定が可能であること，b) 意図推定において，歩行開始時点の映像が得られる方が望ましいが，どの区間を観測しても，約 70% 程度の正解率が得られること，c) 形状 (シルエット) やテクスチャが意図推定に重要であること，そして，d) 頭部領域に含まれる情報は多いものの，頭部以外にも意図推定に有用な情報が含まれることを確認した．

これらの結果は，我々が目指している，人の歩き方の観測からの意図推定の実現可能性を示すものと考えている．

今後は，今回検討した３種類の意図以外の意図についても評価・実験を行うとともに，映像からの意図推定において必要となる情報・要因について，さらに検討を進める．また，歩行の様子を撮影した映像からの意図推定手法の実装方法についても，検討していく予定である．

謝辞 本研究は，科学技術振興機構 (JST) 戦略的創造研究推進事業 (CREST) の支援により実施した．

参考文献

[1] Nakazawa M., Mitsugami I., Yamazoe H., Yagi Y.: *Observation of Gait Changes Associated with Human Intentions*, 2nd Joint World Congress of ISPGR and Gait and Mental Function, 2013.

[2] 中澤満，満上育久，山添大丈，八木康史: 歩容特徴による歩行者の側方注視の有無推定, MIRU2013, SS6-10, 2013.

[3] 岡田典，山添大丈，満上育久，八木康史: 没入型歩行環境を用いた注視と歩行の解析 MIRU2013, SS3-15, 2013.

[4] Csibra G. Gergely G, Biro S., Koos O. and Brockbank M.: *Goal attribution without agency cues: The perception of "pure reason" in infancy*, Cognition, Vol 72, pp.237–267,1999.

[5] Scholl B.J.,and Tremoulet P.D.: *Perceptual causality and animacy*, Trends in Cognitive Sciences, Vol 4, pp.299–309, 2000.

[6] McAleer P., Mazzarino B., Volpe G., Camurri A., Paterson S. and Pollick F.E.: *Perceiving animacy and arousal in transformed displays of human interaction*, Proc. Int'l Sympo. Measurement, Analysis and Modeling of Human Functions, pp.67-71, 2004.

[7] Barrett H.C., Todd P.M., and Miller G.E. and Blythe P.W.: *Accurate judgement of intention from morion cues alone: A cross cultural study*, Evolution and Human Behavior, Vol 26, pp.313–331, 2005.

[8] McAleer P. and Pollick F.E.: *Understanding intention from minimal displays of human activity*, Behavior Research Methods, Vol 40, No. 3, pp.830–839, 2008.

[9] Manera V., Schouten B., Becchio C., Bara B.G., Verfaillie K.: *Inferring intentions from biological motion: A stimulus set of point-light communicative interactions*, Behavior Research Methods, Vol 42, No. 1, pp.168–178, 2010.