

子ども発達研究のためのマルチモーダル映像記述フレームワーク

Multimodal video description framework for understanding child's development

長尾 貴正*1
Takamasa NAGAO

瀬戸 淳也*1
Junya SETO

石川 翔吾*2
Shogo ISHIKAWA

桐山 伸也*2
Shinya KIRIYAMA

竹林 洋一*3
Yoichi TAKEBAYASHI

*1静岡大学情報学研究科

Graduate School of Informatics, Shizuoka University

*2静岡大学情報学部

Faculty of Informatics, Shizuoka University

*3静岡大学創造科学技術大学院

Graduate School of Science and Technology, Shizuoka University

This study aims to construct a multimodal video description framework which analyzes child's self-action in the real world. It also aims to develop an observation-supporting tool utilizing behavior description language. The framework and the tool have a characteristic of successive procedure of example analysis, hypothesis generation and comparative examination. We have accumulated and analyzed the behavioral data of the same child for four years. Based on the result of the analysis, the/our study shows the usefulness and effectiveness of the framework and the tool.

1. はじめに

子どもは両親や社会とインタラクションを行い、多様なスキルや知識の獲得に必要な根源的な知識(常識)を身につける。また、子どもは大人と比べてナイーブであり、思考過程が行動へと表出しやすい傾向にある[竹林 09]。

子どもの発達を分析する手法として、子どもの行動や音声にアノテーションし、発達を分析する研究がある。筆者らも同様にして映像や音声を基にアノテーションを行い、月齢ごとのデータを分析し発達を差分で表現してきた。しかし、アノテーションにおける行動分析では、ダイナミックな行為をタグラベルで表現するのは難しいこと、自由記述であることから記述者によって同一語彙であっても違う意味で扱われるという問題点があった。

そこで、人間の行動をコンピュータで扱えるプログラムとして表現することにより、ダイナミックな行動の分析と項目間の関係性の記述を可能にすることで、上記の問題点を解決する。観察研究の立場から人間の行動・思考をプログラマ的に分析・記述し、プログラム記述を行ったデータから子どもの行動発達を分析した結果を示し、マルチモーダル映像記述フレームワークの構築を提案する。

2. マルチモーダル映像記述フレームワーク

2.1 関連研究

人工知能の研究分野におけるインタラクション分析は人間の行動や思考を理解する主要なアプローチの1つである。人間の発話音声や映像、その他センサーを使って多種多様な情報を取得し、取得したデータに対して意味付けを行ない、人間の行動を理解する研究である。インタラクション分析の活用方法には(1)人間の行動理解における新たな発見(2)評価・学習用データ取得のため(3)コーパス構築・公開するための3つがある[坊農 07]。

映像・音声・センサーデータを収集・閲覧するだけでなく、各々のデータに注釈を付与することで「そこで何が起きている

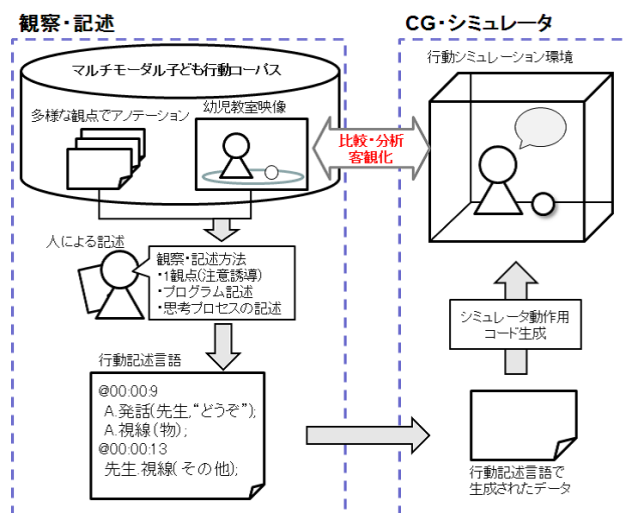


図 1: マルチモーダル映像記述フレームワーク

のか」、「どんな意味があるのか」記述し、把握する。現在では、多様なデータを分析するために多くの分析ツールの開発が進められている。その中の1つである ANVIL[Kipp 12]では、音声を波形として捉え、映像と同期させ、時間を軸にタイムラインにそって項目ごとの記述が行える。更に、項目の記述だけではなく、項目間の記述も階層的に記述できる構造を構築している。

記述されたデータは主にコーパスの深化に用いられるか、統計的手法による行動分析が行われている。そこで、行動分析が記述に留まっている事を問題点とし、記述データの新たな利用法について検討したマルチモーダル映像記述フレームワークについて検討した。

2.2 フレームワークの構成

マルチモーダル映像記述フレームワークでは、人間の行動をプログラムで表現することを軸として観察・記述を行い、記述したデータをシミュレーションを通して可視化することを目的とする。本フレームワークの全体の構成を図1に示す。シミュ

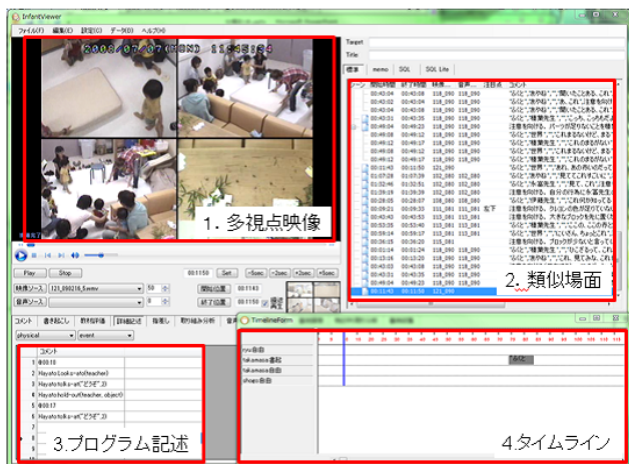


図 2: 行動分析ツール: 1. 多視点映像 2. 類似場面の蓄積 3. プログラム記述 4. タイムライン記述

レータで動作することを想定した行動記述を行うために、人間の行動をプログラムの様に記述をする行動記述言語の開発を行う。行動をプログラム形式に記述する理由は 2 つある。表出した特徴のある行動のモデル化を行うこと、シミュレータでの動作のためのコンピュータに理解できる言語の開発の 2 つである。構築された行動記述言語を用いて行動分析を行うことで、シミュレータは行動記述言語で記述されたデータを基にシミュレータ用の動作コードを生成する。シミュレータは生成されたコードにより記述通りに動作する。実世界の子どもの映像とシミュレータで動作した行動の比較・分析ができる基盤の構築を行う。

本論文では、マルチモーダル映像記述フレームワークの観察・記述に軸をおき、プログラムの様に行動を記述する手法と記述された記述データの分析について考察する。観察・記述 (図 1 左) では、観察の基盤としてマルチモーダル子ども行動コーパスを利用し、コーパスのデータを基に実世界の映像・音声をエビデンスに行動分析ツールを用いて、多面的に行動をプログラムとして分析する。

2.3 行動分析ツール

実世界の子どもの自発的な行動を対象に行動分析を行い、外面的特徴や内面的特徴に多様な注釈を付与し、マルチモーダル子ども行動コーパスを構築してきた [石川 09]。コーパスには 4 年 8 ヶ月開催した幼児教室の映像・音声収録されており、同一幼児の経年的分析が可能である。コーパスの構築には CODOMO-Viewer [石川 09] を用いて、マルチアングルの映像記述と音声を利用し、多様な観点から注釈を付与してきた。

しかし、既存の行動分析ツールでは、マルチモーダル映像記述フレームワークの構築を行えないため、既存のツールを拡張し、新たに記述を支援するシステムを構成した。図 2 に、本フレームワークを構築するために開発したプログラム記述用の行動分析ツールを示す。行動分析ツールではプログラム記述に対して観察の敷居が高くなるという問題点から、タイムラインの記述を導入した。観察時、タイムライン記述に記述することで、プログラム表現へと変換される仕組みを構築した。

2.4 行動記述言語

マルチモーダル映像記述フレームワークの基軸となるプログラム記述について、人と世界を切り分けた記述方法について示す。

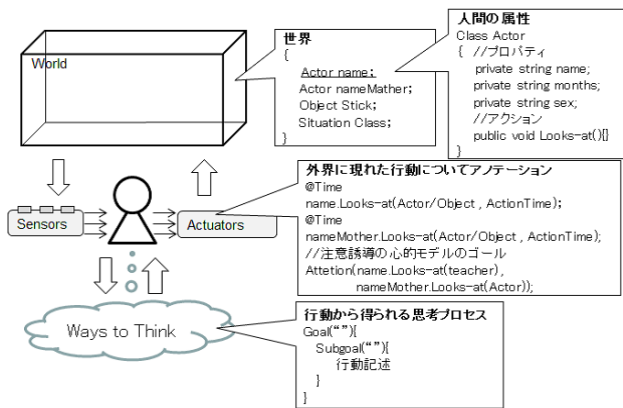


図 3: 外界と人のインタラクション

人は世界とインタラクションを行う。耳や目といった Sensor から情報を取得し、その情報を基に思考し、過去の記憶を参照する。その後、思考の 1 プロセスが決定され、Actuator で世界にアクションを起こす (図 3)。行動分析では世界に表出した Actuator の行動について、行動記述言語で記述を行い、他者や先生、物といった世界で扱うデータを world に切り分けて記述を行う。

世界

世界に記述すべき事柄は、その世界に出てくる行為者・他者・物・状況 (授業/クラス) を対象とする (表 1)。

表 1: 行動記述言語を用いた分析: 世界の記述

世界の記述	説明
Actor 行為者名;	行為者
Actor 他者名;	他者
Actor Teacher;	先生
Object Name;	物
Situation [Class/Free];	授業時間 or 自由時間

人・物の属性

表 2 に人の属性の記述を示す。世界で人を宣言し、人が誰なのか属性を記述する。物については物の名前だけ記述することにする。

表 2: 行動記述言語を用いた分析: 人の記述

人の属性	説明
name	名前
sex	性別
month	月齢
emotion	感情
imprimer	インプリマ

外面的特徴の記述

全ての状況をプログラムで記述することは不可能であるため、マルチモーダル子ども行動コーパスに存在する "注意誘導" を対象とした場面について着目する。"注意誘導" を記述する語彙として、指示表現の心的発達モデルを構築した際に記述の構造化を行ったものを活用する [石川 09]。

内面的特徴の記述

内面的特徴は外面的特徴をエビデンスに段階的に記述を

表 3: 行動記述言語を用いた分析：行動の記述

観察項目	動作名	プログラム記述
発話 視線 ジェスチャー	移動する 指差し 差し出す 受け取る 観察する	Actor.talks-to(Actor,Word,Power) Actor.looks-at(Actor/Object) Actor.gesture.moves-to(Actor/Object) Actor.gesture.point-at(Actor/Object) Actor.gesture.hold-out(Actor,Object) Actor.gesture.recieve(Actor,Object) Actor.gesture.observes(Actor)

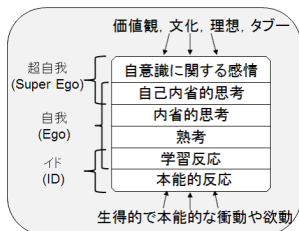


図 4: 6 階層思考モデル

行う。Minsky の 6 階層思考モデル [Minsky 06] 図 4 を利用し、内面的特徴について記述を検討する。注意誘導の場面を対象に、「行動目標」、「感情」という側面から内面的特徴の記述を行う (表 4)。例えば、注意誘導を行う場面というのは、目標に「相手に物を渡したい」や「相手の人に物をとって欲しい」といった具体的な目標が存在する。人間の行動は連続的であるため、プログラムで表現した場合でも連続的に行動した意味を記述するために用いる。

表 4: 目標のプログラム記述

内部項目	プログラム記述
行動目標	Actor.goal("目標") { }
下位目標	Actor.subgoal("下位の目標") { }

3. 注意誘導の行動モデル構築

3.1 注意誘導

基盤となる注意誘導のデータは幼児の月齢 14ヶ月～23ヶ月に亘る 10ヶ月 240 場面に対して、定義された構造 (発話・韻律・視線・ジェスチャー・他者の行動) により記述され、複数人の観察者によって分析・カンファレンスを行いデータの妥当性を評価している。

注意誘導は、自分の注目している物と他者が注目しているものが同じである場合に達成される。子どもは他者に対して他者の状況を認知し、他者の注意を共有する状態になることで、注意誘導を達成する。そのため、他者の注意優先対象という項目を構築して、注意誘導を把握する。

注意誘導が現れる多くの場面では相手に物を渡すことや相手と何か話したい時など、行動の達成目標の下にあり、相手の注意を自分に向けるといった下位目標の位置づけになる。

3.2 注意誘導のラベリング

注意誘導の記述を行う上で対象とした場面の説明と、ある時間区間で区切った映像・音声に対して記述したデータを図 5 に記す。注意誘導の状態を記述するため次の記述構造を利用した。

K君のゴール：先生に持っているものを渡す				
映像	(1)	(2)	(3)	(4)
項目				
状況	物を先生に渡そうとする	もう一度チャレンジする	さらにチャレンジする	物を先生に渡す
発話	どうぞ	どうぞ	どうぞ	どうぞ
韻律	平坦型	強調型	強調型	平坦型
視線	対象物	対象物→先生	先生	先生
しぐさ	差し出す	差し出す	差し出す	差し出す
認知項目	先生の状況	先生の状況	先生の状況	先生の状況
先生の反応	気づいてない	気づいてない	気づいてない	K君を見た
注意誘導	<対象物, その他> <先生, 他者>		<先生, その他> <先生, A君>	

図 5: K 君がもっている物を先生に渡す際に注意誘導が現れる場面の記述例。(1)K 君が先生に物を渡そうとする,(2,3)先生に気づいてもらえないが何度か試みる,(4)先生が気づくまで待ち、物を渡すことに成功する。

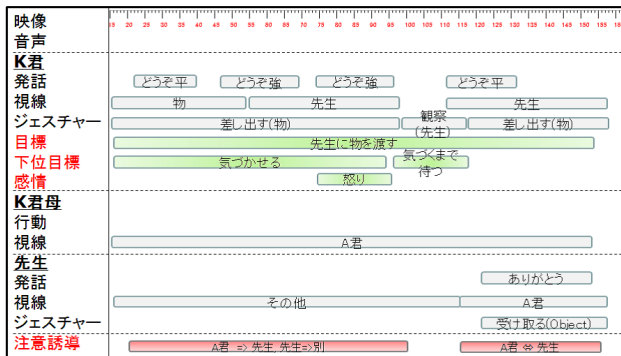


図 6: K 君がもっている物を先生に渡す際に注意誘導が現れる場面：タイムラインで記述を行った例

[発話：子どもの発話を記述] [韻律：平坦，疑問，強調，その他] [視線：他者，物，キョロキョロ，別の方向] [ジェスチャー：指差し，差し出す，動く，叩く，示す，掴む] [他者：子ども，先生，母親] [他者の働きかけ：反応なし，同意する]。

注意誘導の行動モデルでは、行為者の視線・しぐさ・他者の状況から、行為者の注目を引いている物や箇所，他者の注視している物や箇所について記述する注意誘導という項目を構築した。注意誘導を <行為者の注意，他者の注意> という形式で記述を行なっている。

今回対象とした場面では、4つに区切りそれぞれ記述したものを表にまとめた。ラベリングでは分析者の観点によって、特徴となる場面を独自に切り出し、ある行動の区切りが良いところで分けられる。従って、1つ1つの区切られた映像・音声を記述した後、データを表のようにまとめ各区間の記述を比較しながら、注意誘導という行動について着目して分析を行う必要がある。

3.3 注意誘導のプログラム記述

ラベリングで得られたデータを利用し、区切られた場面ごとの記述をタイムライン形式に変換し、そこから行動記述言語を用いたプログラムへと表現したものを示す。図 6 にタイム

表 5: タイムラインからプログラム記述

	プログラム記述
1.	time(55:10) {
2.	K.looks-at(object);
3.	K.gesture.hold-out(teacher, object)
4.	}
5.	time(55:20) {
6.	K.talks-to(teacher,"どうぞ","平坦")
7.	}
8.	time(55:23) {
9.	K.looks-at(teacher);
10.	K.talks-at(teacher,"どうぞ","平坦")
11.	}
12.	time(55:26) {
13.	K.talks-at(teacher,"どうぞ","強調")
14.	}
15.	time(55:33) {
16.	K.looks-at(teacher);
17.	teacher.looks-at(K);
18.	time(55:34) {
19.	K.talks-at(teacher,"どうぞ","平坦");
20.	K.gesture.hold-out(teacher,object);
21.	teacher.gesture.recv(K,object);
22.	teacher.talks-to(K,"ありがとう","平坦")
23.	}

ラインで記述したデータについて示し、表 5 ではプログラムで記述されたデータについて示す。表 6 にサブゴールの記述について示す。サブゴールは現れた行動が何を目標としたアプローチであるのかまとめるものであるため、表 5 とは別途に記述する必要がある。

ラベリングと同様に注意誘導の行動モデルを構築する。行動目標である「相手に物を渡す」ことは、初めに「相手に気づいてもらう」ことが必要であり、相手の注意を自分に向けることが必要である。故に、「他者が自分に気づく」までが注意誘導であり、下位目標で括られた行動は注意誘導を行って行く過程である。ラベリングと同様に、注意誘導は<行為者の注意, 他者の注意>で表現すると、達成条件は K.looks-at(teacher), teacher.looks-at(K) である。

また、今回の場面では注意誘導は他者に行為者の意図を伝える時に現れる行動と考え、物を渡す時に現れる注意誘導の行動モデルとなると推測される。従って、視線だけではなく発話を行うことにより相手の注意を誘導している場面であると説明できる。タイムラインを縦軸の時間で注意誘導達成のための行動を取り出した時、K.looks-at(object), K.talks-to(teacher,"どうぞ","平坦") や K.looks-at(teacher), K.talks-at(teacher,"どうぞ","強調") が目標達成のためのアプローチとなる。

表 6: サブゴールの記述

	プログラム記述
1.	K.subgoal("気づいてもらう") {
2.	time(55:10) {
3.	K.looks-at(object);
4.	K.gesture.hold-out(teacher, object)
5.	}
6.	time(55:20) {
7.	K.talks-to(teacher,"どうぞ","平坦")
8.	}
9.	return false;
10.	}

3.4 考察

注意誘導の場面を記述する記述構造を決めるためには類似場面をいくつも分析・検討を行なうことが必要である。ゆえに、ラベリングでは1つ1つの場面に状況の説明を自由記述で加えることで類似場面から記述構造を考えることができるが、プログラム記述では決められた構造に準じた場面の記述しか行えない。また、プログラム記述では与えられた文法でしか場面を表現することができないため、先にどんな記述を行うのか検討する必要がある。

場面を区切って記述したラベリングと一連の場面を記述したタイムライン記述では、タイムライン記述のほうが観察時に行う記述の敷居が高くなる。なぜならば、タイムライン記述では行動の開始時間や終了時間を見直す必要があること、項目の記述を行わなければならないことにある。しかし、記述後の分析はタイムライン記述を用いることでラベリングよりも関係性について分析を行いやすくなる。

注意誘導で構築した内部モデルでは、プログラム記述を用いることで特定の行動が現れた時注意誘導を行なっているが自動検知できることや、注意誘導というキーワードで行動を集めた際に、月齢で比較することで行動の変化が読み取れる。

4. 結論

実世界における子どもの行動をプログラムで表現する手法と記述データを基にシミュレータによって可視化するマルチモーダル映像記述フレームワークについて提案した。行動をプログラムで記述することによりダイナミックな行動の表現を可能にし、新たな行動分析の手法を示した。注意誘導の場面では、注意誘導を達成するために現れた行動が行動セットとしてまとめることが可能になった。プログラム記述は特徴のある行動を行動セットとして保持することで構造を構築することができ、目標を成し遂げるアプローチを示す有効な手がかりとなることが示せた。

今後は、行動分析ツールの拡張と、ツールを用いた更なる事例分析、事例分析から内面的特徴を発見・構築を行う。

参考文献

- [Minsky 06] Marvin Minsky (著), 竹林 洋一 (訳): ミンスキー博士の脳の探検 - 常識・感情・自己とは - (原題: The Emotion Machine), 共立出版, (2009)
- [坊農 07] 坊農真弓ほか: 多人数インタラクション研究には何が必要?: インタラクション研究の国内外の動向と現状 (連載チュートリアル, 多人数インタラクションの分析方法【第1回】) 人工知能学会誌 22(5), 703-710, (2007)
- [竹林 09] 竹林洋一ほか: 工学的視点からの幼児の行動観察とコーパス構築 - 認知・行動モデルの深化がもたらすもの -, 日本音響学会誌 65(10), pp.544-549(2009)
- [石川 09] 石川翔吾ほか: 心的発達分析に基づく指示表現モデルの構築, 人工知能全国大会第 23 回, 1A1-1, (2009)
- [Kipp 12] Kipp, M. (to appear) ANVIL: A Universal Video Research Tool. In: J. Durand, U. Gut, G. Kristoffer-son (Hrsg.) Handbook of Corpus Phonology, Oxford University Press. (2012)