社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2003－ICS－134 (1)
2003／10／8

対話的進化ロボティクスにおけるアクティブ教示法
Active Teaching for an Interactive Learning Robot

片上 大輔[*1] 山田 誠二[*2]
Katagami Daisuke Yamada Seiji

[*1]東京工業大学大
Tokyo Institute of Thechnology
[*2]国立情報学研究所
National Institute of Informatics

We have proposed a fast learning method that enables a mobile robot to acquires autonomous behaviors from interaction between human and robot. In this research we develop a behavior learning method ICS (Interactive Classifier System) using interactive evolutionary computation considering an operator's teaching cost. As a result, a mobile robot is able to quickly learn rules by directly teaching from an operator. ICS is a novel evolutionary robotics approach using classifier system. In this paper, we investigate teacher's physical and mental load and proposed a teaching method based on timing of instruction using ICS.

1. はじめに

人間が活動する環境内において行動する自律ロボットでは，未知環境における行動，予期しない人間からのインタラクションなど，行動に必要な知識を事前に用意しておく事が難しい状況でタスクをこなすことが要求される．そこで，自律行動の獲得のための学習や環境への適用が必要となってくる．

近年においては，学習や適応の枠組みとして強化学習や進化計算手法を用いてロボットに自律的に制御器を獲得させる研究が注目を集めてきた．これらの手法の目的の一つは，身体性や環境との相互作用ダイナミクスを制御器の構築に反映させる際に設計者による不適切・不必要なバイアスを排除することである．そのため，従来では知識を強化学習の枠組みにいれずにエージェントに試行錯誤させて学習することが前提とされてきた．しかし，実環境問題に適用するにあたってその実行速度が問題となっている．

そこで，同じ環境に存在する人間とのインタラクションを利用する研究が行われてきている．特に，アプリオリな知識を持たないロボットや初期段階の試行錯誤のロボットにおいては，人間からの教示は有効な自律行動の獲得手法であると言える．しかし，ある程度の自律性を持ったロボットにおいては，人間からの教示は常に必要であるわけではない．教示が必要でない時は，人間に負担をかけることなく，それまでの人間とのインタラクションにより蓄えられた教示情報を元にして自律的に行動するべきである．このように，人間とロボットのインタラクションを通じて，ロボットの自律性を獲得する手法が重要視されてきている．

このようなヒューマン・ロボットインタラクションの研究としては現在多くの研究が行われている．石黒らは，移動ロボットの状態空間の構築 [4] を強化学習により行っているが，始めに人間が教示した行動をお手本にして学習を行っており，リアルタイムな人間とのインタラクションは行われていない．麻生らは，人間と音声会話によるコミュニケーションを行う事情通ロボットによって，未知環境の地図情報を構築する枠組み [1] を提案している．人間とロボットとの対話によりロボットの行動を獲得しているわけではない．堀口らは，人間とロボットのインタラクションの設計として相互主導型インラクションの概念を用い，力覚フィードバックを利用した移動ロボットの自動化プロセスと人間の操作の協調行動を実現 [3] しているが，その学習結果をロボットの行動獲得には反映してはいない．稲邑らは，ユーザとの対話に基づい

連絡先: 片上 大輔，東京工業大学
〒226-8502 神奈川県横浜市緑区長津田町 4259 R1 棟-521,
Tel: 045-924-5218, Fax: 045-924-5218,
E-mail: katagami@ntt.dis.titech.ac.jp

てBayesian Networkを用いて確率的にロボットの行動獲得[10]を行っているが，進化計算手法により段階的に行動獲得を行う我々の手法とは方法論的に大きく異なる．

これらの研究に対して我々は，ロボットが動作する際に人間から適切な行為としての教示情報を受け取って，タスクを解決しうる状態認識・行為ルールを進化的に獲得し，ロボットの自律性を実現することを目的とする．我々はこのような枠組みを対話的進化ロボティクスInteractive Evolutionary Robotics (IER)[5]と呼ぶ．本研究では，IERの枠組みにおいて教示者の認知的負荷を考慮したタイミングで教示を行うActive Teaching法を提案し，従来の教示法とシミュレーションにより比較実験を行い，検証する．

2. 教示による学習

2.1 教師の負荷

教師の負荷には，肉体的な疲労と心理的な疲労の2種類ががあると考えられる．本研究では，教師の負荷を簡単に計るために肉体的負荷と認知的負荷にわける．これにより，本枠組みにおいては認知的負荷を教示のタイミング，肉体的負荷を教示の回数と考えることができる．

一般に，対話型学習の場合，教示を行えば行うほどいいパフォーマンスが得られることが言えるが，人間の労力は無限ではない．教示コストは低ければ低いほどいいというトレードオフの関係になっていることは明らかである．人間が疲れをしらない機械と協調して，世代毎に多くの個体を比較評価し，評価値を入力するには限界がある．これが実用上の大きな問題になっている．また，第2の問題点として，比較評価の際の肉体的および心理的疲労軽減のため，個体数と探索世代数を，通常のEC探索に比較して非常に少なくせざるを得ないことがある．これは，収束悪化につながる．結果として，教示回数（肉体的負荷）を少なくすることは難しい．

IERでは，これに対して入力装置によるロボットの直接操作を行い，その操作とその時の環境情報からルールの自動生成を行うことで教示とする．この枠組みにおいては，来の対話型手法のように多くの個体を比較評価または評価値の入力を行う必要はなく，肉体的および心理的疲労が大幅に軽減されることが期待される．

また，対話型学習を実ロボット環境に適用する場合，ロボットは教示指令が無い場合は自律的に学習を行い，教示者が用意した入力装置から直感的にロボットを操作することで自動的にルールが作成され教示を行う方法が考えられる．この方法により，システムが自律的に学習を行う点，直感的な教示により自動的にルールが作成される点，および，任意の時に追加学習を行える点においてこの疲労問題が軽減されると考えられる．

2.2 教示のタイミング

前述の通り，教示数を少なくすることは非常に難しい．そこで，本研究では教示のタイミングに注目した．

教示の行われるタイミングは，前述の教示者の負担と大きく関わってくる．一般的には，あらかじめ教示を行ったり(Off-line Teachingと呼ぶ)，システムの要求時に教示を行う（Passive Teachingと呼ぶ）ことが多いが，これらの手法は教示の行うタイミングをシステム側に委ねているため，教示を行うために人間側が待機しなければならず，シミュレーションにおける実験はもとより実験時間がかかる実環境学習においてはさらに教示者の負担が増大する．

そこで、我々は以下のようなActive Teaching法を提案する。そして，従来のOff-line Teaching法，および，Passive Teaching法と認知的負荷を計る実験を行い，心理学的評価により比較検証する．それぞれを以下に説明する

2.21 Active Teaching

この教示法では，教示者はロボットが自律行動を行うのをみながら好きなタイミングでロボットを動かしタスクを達成させる．これにより，教示者は教示をしていることを意識せずに，また学習者の挙動を全て把握した上で教示をするかしないかを悩むことなく教示を行うことができる．ただ，システム側にその様な仕様を組み込むのが難しいことが言える．

2.22 Off-line Teaching

Off-line teachingは，あらかじめ決められた回数の教示ステップにて教示を行い．その後自律行動ステップにて自律学習を行う方法である．

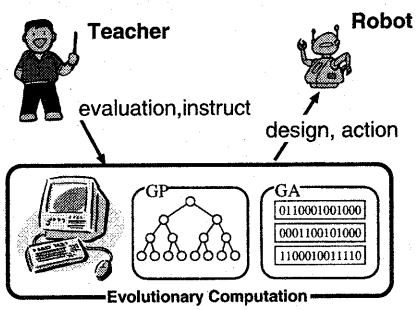

図 1　対話的進化ロボティクス

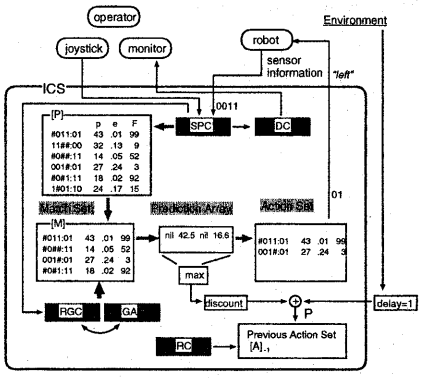

図 2　対話的クラシファイアシステム

2.23 Passive Teaching

システムの要求時に教示を行う，もしくは教示を行うようにユーザに要求する方法を passive teaching 法と呼ぶ．Mishima, Asada らは，教示者と学習者の間に生じる環境認識のずれ（Cross Perceptual Aliasing）による教示学習の効率悪化を Passive Teaching の教示方法により解決している [7]．Passive Teaching は学習効率において教示の無駄が少なくいい方法と考えられる。しかし，教示者としては，システムが行動を要求するまでは監視していなくてはならない。またそのタイミングがいつくるかわからないため，教示数の割に精神的な負担が大きくなると考えられる．

3. 対話的進化ロボティクスに基づく教示

3.1 対話的進化ロボティクス

対話的進化ロボティクス (IER) は，人間を評価系に組み込み進化的に探索を行う対話型進化計算法 (*Interactive Evolutionary Computation*(IEC)) の評価能力を用いて効率のよい実環境ロボット学習を行うことを目的とした枠組みである．またこれは，遺伝的アルゴリズム，遺伝的プログラミング，進化戦略などの進化的計算手法を用いて，対話的にロボットを設計するプローチであるとも言える．図 1 に **IER** の概要図を示す．

我々は，この枠組みによるアプローチが未知な環境に対する学習，特に試行錯誤の初期段階の学習 (今後，初期学習と呼ぶ) に効果があると考える．また，人間とロボットのインタラクションにより人間だけでは解けないような局所解に対して解を得るといった結果も期待できる．さらに，本手法は学習アルゴリズムに依存しないため，進化ロボティクス一般に広く適用できることが言える．

3.2 対話的クラシファイアシステム

ICS は IER の枠組みに基づき，学習分類子システム (Learning Classifier System: LCS) に IEC の対話機能を組み込むことにより，自律的な学習に加え教示による学習も行うことができるロボット学習モデルである．学習アルゴリズムである LCS には Wilson が提案した XCS[8] を使用している．XCS は ZCS[9] を改良したもので，精度 (accuracy) と呼ばれるパラメータを追加したものである．構築したシステムの概要図を図 2 に示す．

本研究で開発したシステムは，操作者の教示情報をもとにクラシファイアを作成するルール生成部 (RGC)，ロボットのセンサ情報を処理するセンサ処理部 (SPC)，GUI インタフェース等の表示部 (DC)，学習を行う強化学習部 (RC) からなる．

【RGC】 Rule Generation Component は，教示によるルールの作成を行う．教示者はロボットをインタフェースに表示される情報を見ながら，入力装置を用いて操作し，そこでの操作履歴とその時のロボットのセンサ情報をセンサ処理部 (SPC) が受け取り，それより RGC が新しくルールを作成しルールリストに加える．ルールの作成手続きは，主に XCS を基本に教示情報 (operator がロボットを操作した行動情報) からルールを

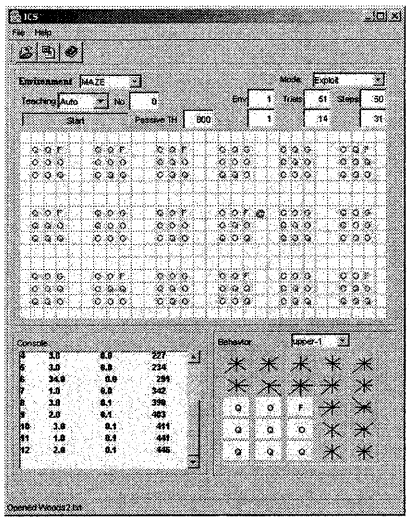

図 3　ユーザインタフェース

作成できる様に改良した．

1. システムはロボットのセンサ情報 X と教示情報 a_t を SPC から受け取る．
2. 集団 [P] から X にマッチしたクラシファイアがマッチセット [M] に移され，システムは [M] で表された各々の行為 a_i を支持するクラシファイアの Prediction 値を Fitness 値で正規化して $P(a_i)$ を作成する．$P(a_i)$ の値は Prediction Array に置かれ，行為が選択される．行為選択は，決定論的行為選択もしくは，ルーレットホイール選択により行われる．
3. 行為選択により選ばれた行為 a_j と教示により得られた行為 a_t を比較し，$a_j \neq a_t$ ならば，[M] の中で行動部に a_j を持つルールの行動部を a_t に書き換える．$a_j = a_t$ ならば，変更はしない．
4. 選ばれた行為 a_j を支持する [M] の中のクラシファイアからなる行動セット [A] を作成する．行為 a_j は効果器に送られ，a_t の入力があった場合は，すぐに報酬 r_{teach} が与えられる．a_t の入力が無い場合も報酬 r_{imm} が環境から返される（返されない場合もある）．

【RC】Reinforcement Component は，クラシファイアシステムにおける強化学習部である．前のステップのクラシファイアのパラメータを更新することで学習を行う．教示者の操作がないときは，ロボットはそれまでに作成されたルールから自律的に行動を行うことができる．

【DC】Display Component は，SPC により処理されたデータの表示を受け持つ．開発したインタフェイスを図 3 に示す．

【SPC】Sensor Processing Component は，ロボットの各種センサの処理および教示情報の処理を行う．処理されたデータは DC および RGC に送られ表示とルールの作成を行う．

3.3　学習の手続き

ICS では教示モードと自律行動モードの 2 つのモードを交互に行うことで学習を進める．2 つのモードの手続きを以下に示す．

教示モード

1. ロボットのセンサ情報から，その状態空間（条件部）を用意する．
2. 3 種類の教示のタイミングの手続きのいづれかにより教示を行う．
3. 操作者の指示情報とその時の環境情報によりルールを作成する．
4. 同じクラスタに属するルールがなければ，新しくルールとして追加する．
5. 同じクラスタに属するルールがあれば，報酬として強化値をあげる．

自律行動モード

1. Rule List に蓄えられたルールに従って行動する．
2. マッチセットにおける一つ前の GA からのタイムステップ数の平均が閾値を越えるならば，GA がそのマッチセットに対して走る．（ニッチ GA）

教示モードと，自律行動モードでの概要図を図 4 に示す．

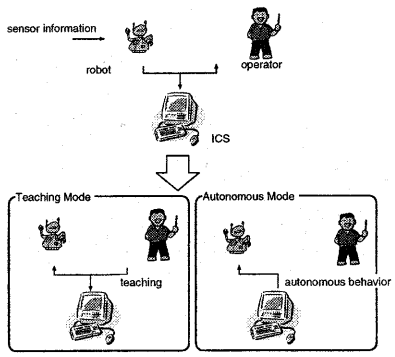

図 4　教示モードと自律行動モード

3.4　教示のタイミングの手続き

教示モードの手続きの 2 において実行される教示のタイミングとしては，2.2 章の 3 つが挙げられる．それぞれの手続きを以下に示す．

off-line teaching

1. ロボットのセンサ情報から，その状態空間（条件部）を用意する．
2. 教示回数の閾値 i 以下なら，ユーザに指示をリクエストする．
3. 教示回数の閾値 i よりも大きければ，ロボットは行動選択を決定論的行い，それによって選ばれた行為を実行する

passive teaching

1. ロボットのセンサ情報から，その状態空間（条件部）を用意する．
2. 状態空間に対して有効な行為があればそれを実行する
3. 有効な行為がなく，さらに教示回数の閾値 i 以下ならば，ユーザに指示をリクエストする．
4. それ以外ならば，ロボットは行動選択を決定論的行い，それによって選ばれた行為を実行する

active teaching

1. 教示回数の閾値 i 以下で，状態空間に対してユーザからの指示があれば実行する．
2. なければ，ロボットは行動選択を決定論的行い，それによって選ばれた行為を実行する．

4.　実験

4.1　実験設定

教示と教示者の負荷の関係を調べるために，まずシミュレーションにおける実験を行った．本研究の有効性を示すために，シミュレーションによる予備実験を行った．実験環境としては，クラシファイアシステムの研究でテストベッドとして使われている Woods 環境 [8] の一つである Woods2 環境を用いた．この環境は教示者と学習者（エージェント）の間の認識のずれの問題，いわゆる相互知覚行動不整合問題（Cross Perceptual Aliasing）は発生しないと考えられる [6]．図 5 に Wood2 環境を示す．Woods 環境はマルコフ環境であり，その上端と下端および左端と右端は連続したトーラス状の空間となっている．また環境は 30 × 15 のセルからなり，図 5 の 5 × 5 の領域が横に 6 つ，縦に 3 つ並んでいる．図 5 の"∗" は自律エージェントであり，Animat と呼ばれている．Animat は始めにランダムに Blank セルに配置され，1 マスずつ動いて目標地点であるセンサコードの異なる 2 種類の Food F または G（それぞれ，110,111）にでるだけ早く到達することを学習する．Animat は自分の現在いる位置の周囲 8 方向の情報を知覚し，それに基づいて行動を決定する．ただし，障害物である Object O または Q（010, 011）の方向には移動することができない．Blanks は". " で 000

1 つのクラシファイアは条件部と行動部からなる 24 ビットの文字列で表示される．条件部は {0,1,#} から，行動部は 0〜7 からなる整数で表される．「#」は don't care シンボルと呼ばれ，ルールの一般化を行うのに用いられる．例えば#000#100000###100#0000:3 で表されるクラシファイアは，条件部が左の 24 ビットの文字列で表されており，それぞれ 3bit ずつ animat の回りに装備されたセンサの反応を北から時計回りに 8 つの方向のオブジェクトのセンサコードを表している．行動部は 0 が北方向で時計回りにそれぞれ表し，7 が北西方向を表す．

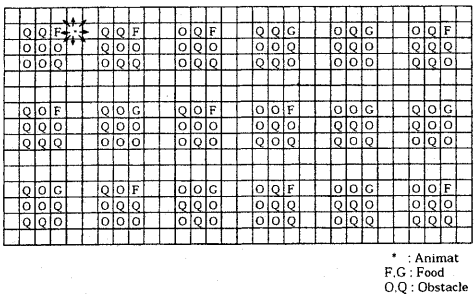

図 5　Woods2 環境

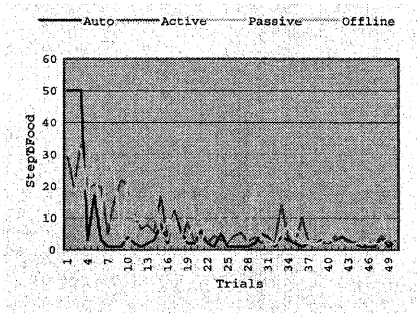

図 6　Food までのステップ数

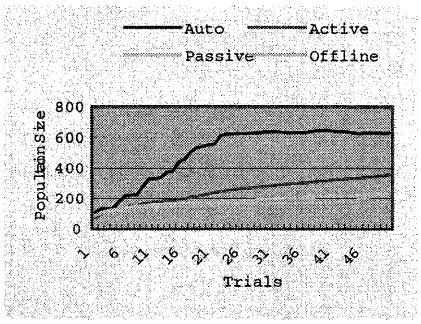

図 7　集団サイズ

4.2　実験の概要

Active Teaching と Passive Teaching または Off-line Teaching との比較実験を行った．ゴールにたどり着くか，50step 移動すると 1 試行．50 試行を 1 実験として，7 人の大学院生を被験者にした．

本研究のように人間の認知的負荷を含む実験において，その負荷を調べるために，別のタスクを用意する方法が行われている．例えば，実験の参加者は本来のタスクを行いながら，別のタスクを行うように頼む [2] といった方法が行われている．本研究では，実験本来のタスクは，エージェントが Food にたどり着くようにユーザがロボットに指示を与えることであるが，その教示における認知的負荷を計測するために，エージェントの教示タスクを行いながら 2 桁の足し算の問題を行わなくてはならないものとする．

4.3　教示効果の考察

本実験では，Animat が初期位置から Food にたどり着くまでの期間，あるいは設定した最大ステップ数を消費するまでの期間 (Step to Food) と，集団サイズ (Population Size) を求めた．ここでの集団サイズは，マクロクラシファイア（条件部と行動部が同じクラシファイアは同じものとし集団サイズには数えない）の数とする．Food までの Step 数を図 6 に示す．また，集団サイズを図 7 に示す．

教示効果については，各手法においてあまり差はでなかった．もともと学習については同じものなので，教示のタイミングによってユーザの教示ミスが変化するかによるが，各手法によって特に差はなかった．初期段階では各ユーザがまだ教示環境に慣れておらず，パフォーマンスが低くなっているが，30 試行ぐらいで慣れてきているのがわかる．

集団サイズをみると，自律行動の場合は約 600 クラシファイアで収束している．教示を用いた各手法は，約 200 クラシファイアで収束している．これは，人間の判断により効率だけではなく，意図を重要視したルールが作成されるためであり，効率が同じで幾通りの解法がある場合にも，バリエーションに富んだルールを作成することは難しいことを示している．ここで，Active 法のみ集団サイズが 400 クラシファイアを越えている．これは，Active 法がユーザの好きなタイミングで教示を行うため，ある瞬間毎の教示を行うことになるため，同じ条件でも違う教示を行ってしまうことになると考えられるが，今後ルールを詳しく解析するなどして検討する必要がある．

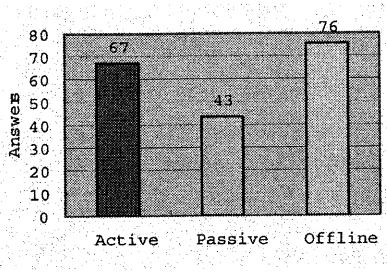

図 8　解答数

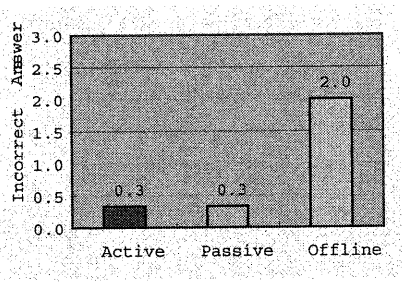

図 9　誤回答数

4.4　教示の負荷の考察

教示の負荷を計測するために，ユーザにはセカンダリタスクとして2桁の足し算の問題を解いてもらい，その解答数と誤答数を計測した．図8に，ユーザがプライマリタスクを実行中に解いた2桁の足し算のの解答数を示す．Active 法においては，各被験者の平均解答数が67問であり，Offline で教示をするのと同じぐらいの余裕があると考えられる．それに対し Passive 法では，解答数が低く，教示のタイミングをシステム側が決定することで，教示者に認知的負荷を与えていることがわかる．

図9に，ユーザがプライマリタスク実験中に解いたセカンダリタスクの解答の間違いの数を示す．Offline 教示に平均で約80問中で2問の間違いが発生した．Offline 教示のようにつづけて教示をすることで，タスク達成のスピードがあがるが，同時に正確性が失われることがわかる．

図10に，認知的負荷のアンケートの結果を示す．これはユーザビリティの評価を参考にし，3つの教示法について相対的な評価で以下の有

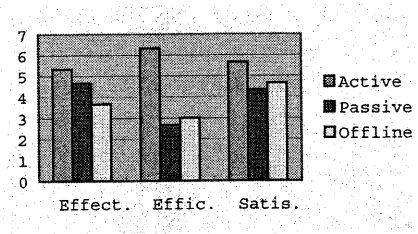

図 10　ユーザビリティ

効性，効率性，満足度について7点法でそれぞれ評価してもらった．

- Effectiveness （有効性）：
  ユーザーが指定された目標を達成する上での正確さ、完全性
- Efficiency（効率性）：
  ユーザーが目標を達成する際に、正確さと完全性に費やした資源
- Satisfaction（満足度）：
  製品を使用する際の、不快感のなさ、及び肯定的な態度

有効性と満足度については有意な差がなかったが，効率性について有意な差 ($p < .05$) がみられた．Passive 法はユーザの自由記述のアンケートによると，「教示するタイミングが制限（指定）されるためセカンダリタスクに集中できない」「確認作業が多くてタスクに集中できない」など，タイミングをシステム側が決定しているため認知的名負荷が高まっていることがわかる．セカンダリタスクの結果としては，認知的負荷の効果は見られているが，誤回答数など，ユーザがそれを認識していないこともあり，満足度や有効性までには結び付かなかったと考えられる．しかし，Active 法においては，「エージェントの動きを観察する必要があり時間がかかった」「O,Q,F,G など文字なので一瞬迷う」など，注意を自分のタイミングで移動することによってその瞬間のエージェントの動きの把握をする必要があり，その判断が正確に行われるようにすることで，これらについては改善されると考えられる．

5.　結論

本研究では、IER の枠組みにおいて教示者の認知的負荷を考慮したタイミングで教示を行う

Active 教示法を提案し，従来の教示法とシミュレーションにより比較実験を行い検証した．教示者の認知的負荷を計測するために，ユーザにはセカンダリタスクとして2桁の足し算の問題を解いてもらい，その解答数と誤答数を計測した．また，ユーザビリティの評価を参考にした認知的負荷のアンケートによりその効果を調べた．タスクのパフォーマンス（Food までのステップ数）は変わらずに，認知的負荷に関して Active 教示法に有効な結果がみられた．

参考文献

[1] H. Asoh and Y. Motomura and I. Hara and S. Akaho and S. Hayamizu and T. Matsui: Combining probabilistic map and dialog for robust life-long offifce navigation; *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 807–812 (1996)

[2] J. Crandall and M. A. Goodrich: Characterizing Efficiency of Human Robot Interaction: A Case Study of Shared-Control Teleoperation; *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1290–1295 (2002)

[3] Y. Horiguchi and T. Sawaragi and G. Akashi: Naturalistic Human-Robot Collaboration Based upon Mixed-Initiative Interactions in Teleoperating Environment; *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 876–881 (2000)

[4] H. Ishiguro and R. Sato and T. Ishida: Robot Oriented State Space Construction; *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1496–1501 (1996)

[5] D. Katagami and S. Yamada: Interactive Classifier System for Real Robot Learning; *IEEE International Workshop on Robot and Human Interaction*, pp. 258–263 (2000)

[6] D. Katagami and S. Yamada: Interactive Evolutionary Robotics from Different Viewpoints of Observation; *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1108–1113 (2002)

[7] C. Mishima and M. Asada: Active Learning from Cross Perceptual Aliasing Caoused by Direct Teaching; *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1420–1425 (1999)

[8] S. W. Wilson: Classifier fitness based on accuracy; *Evolutionary Computation*, Vol. 3, No. 2, pp. 149–175 (1995)

[9] S. W. Wilson: ZCS: a zeroth order classifier system; *Evolutionary Computation*, Vol. 2, pp. 1–18 (1994)

[10] 稲邑 哲也, 稲葉 雅幸, 井上 博允: ユーザとの対話に基づく段階的な行動決定モデルの獲得; 日本ロボット学会誌, Vol. 19, No. 8, pp. 983–990 (2001)