2I3-3

# アンビエント環境における経験型強化学習を用いたインタラクティブデバイスの制御

Interactive device control by using reinforcement learning in the ambient network environment

中瀬 絢哉[*1] Junya Nakase
森山 甲一[*2] Koichi Moriyama
清川 清[*1] Kiyoshi Kiyokawa
沼尾 正行[*2] Masayuki Numao
栗原 聡[*2*3] Satoshi Kurihara

[*1]大阪大学大学院情報科学研究科
School of Information Science and Technology, Osaka University

[*2]大阪大学産業科学研究所
Institute of Scientific and Industrial Research, Osaka University

[*3]JST CREST
JST CREST

The ambient networks need a function to select proper interactions in order to satisfy users. In this research, we propose a method to find the proper interactions to sleepy or sleeping users in a working space (e.g., lab and office) by reinforcement learning. This method controls a lamp, an aroma device and a music device and finds an interaction sequence that lets the users awake comfortably. We conducted an experiment to verify the validity of this method.

## 1. はじめに

近年，日常生活での事故が増加している．高齢化社会の訪れに伴い，多くの高齢者が小さな段差で躓くような，家庭内に潜む危険によって怪我をしている．また，高齢者だけでなく，乳幼児の全事故の大半が家庭内で発生しており．このことも日常の中にある危険を把握し切れてない結果である．そして恒常的に工事などが行われている工場やプラントなどにおいても，情報伝達が徹底されていないが故の事故が発生している．このような人の気づかない部分や対処しきれない部分をサポートする環境の整備が急務となっている．ここで注目されているのが，アンビエント情報社会基盤である [1].

アンビエント情報社会基盤とは，センサが取り付けられた環境において，センサからの情報により実世界の状況を知覚し，その状況に応じた働きかけを能動的に実行する高機能な環境である．このアンビエント情報社会基盤が実現すれば，人が気づかないような小さな危険がその人の身に降りかかりそうな時，危険を警告したり，排除したりすることができる．また，人の状態をセンサの情報から検出することによって，その人の望んでいることを実行することが可能である．例えば，図 1 のように環境がセンサの情報からユーザが疲れていると判断し，リラックスできるように空調を涼しくしたり，照明を暗くしたりするなどユーザが快適になるように調整することも可能である．このように，アンビエント情報社会基盤の実現は人々に大きな利益をもたらす．このアンビエント情報基盤の実現で，人の作業補助，身の回りの危険予知といったことが可能になると期待されている．

このアンビエント情報基盤を構築するには以下の 3 つが必要である．

- 実世界の情報を取得するためのセンサネットワークの構築
  環境の状況を知るための情報を取得するセンサで構築されたネットワークの構築が必要である．ユビキタスネットワークの発達に伴い，PC や携帯電話などのネットワーク端末の普及やネットワーク技術の進歩のため，人が社会のあらゆるものとネットワークによってつながることが可能になった．これら身近な電子機器（PC，家電製品）にセンサを加えることによって，環境や人の情報を獲得することが可能であるし，センサを設置する場合においても，センサからの情報を集約，処理するネットワークはユビキタスネットワークを利用できる．

- 実世界の情報からの人の行動や状態の推定
  センサからの情報をそのまま使用するだけでは，複数のセンシング対象が存在する環境において行動を把握し切れなかったり，行動を把握するために膨大な数のセンサを設置しなければならない．さらに，ユーザの状態を知ることもユーザに適した働きかけをする上で重要である．そこで，センサの情報を解析し，ユーザの行動や状態を抽出する技術が必要である．

- 状況に適した働きかけの選択
  不確定な行動をとる人間に対して，条件的に決められた働きかけをするのでは適切な働きかけをすることはできない．したがって，適切なモデル化をするなど不確定な状況にあった働きかけを選択する適応力のある技術が必要である．

本研究では 3 つ目に着目する．具体的には，日常のワークスペースを対象として，疲労などから起こる眠気をセンシングにより察知し，環境からのインタラクションによって，心地よく覚醒させる環境の実現を目指す．それぞれの人にとっての最適なインタラクション系列は同一ではないため，環境が個人に適応できることが重要である．今回は，眠気を催している，または眠っている状態において，人を快適に覚醒させるインタラクションチャネルとして，照明・香り・音楽の 3 種類を利用する．そして，個々人における最適なインタラクション系列を獲得することを目指す．しかし，照明，香り，音楽にはいろいろなバリエーションがあるため，すべての組み合わせをユーザに対して実行して，その中から最適なものを選択するという方法では，組み合わせ爆発を起こすことから非効率かつ非現実的である．そこで本研究では前提知識を必要とせず，試行錯誤的に解を探索する強化学習 [3]，特に，少ない試行によって解を探

連絡先: 中瀬 絢哉，大阪大学産業科学研究所沼尾研究室，大阪府茨木市美穂ケ丘 8-1，Tel:06-6879-8426，Fax:06-6879-8428，nakase@ai.sanken.osaka-u.ac.jp

索できる Profit sharing[4] を用いることにより，最適なインタラクション系列の獲得を目指す．

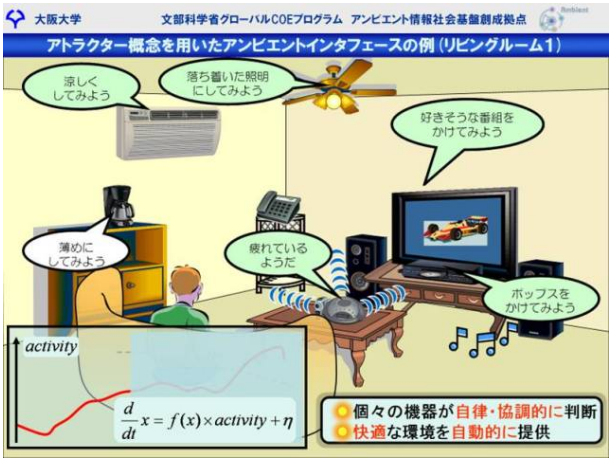


図 1: アンビエント環境の例 [2]

## 2. 関連研究

センシングの分野では赤外線センサや圧力センサといった物理センサの進歩は目覚しく，安価で非常に利用しやすくなっており，また web カメラなどの画像取得端末も高性能なものが安価で利用できる．加えて，web カメラの画像処理によって瞬きを検出する技術 [5] のように単純なセンサから高度な情報を取得する技術の研究も進んでいる．このような技術は今回の研究のように人の状態を検出する必要のある場合では非常に重要である．

また，これらのセンサを用いた人の行動や状態を抽出する技術の研究も進んでいる．人の習慣的な行動の抽出についての研究では，多数のセンサが設置された環境における人の行動を，センサからの情報によって追跡する研究 [6] が進められている．人の状態を抽出する研究では，瞬きや体の動きといった身体情報をセンシングすることで，その人の眠気を推定する研究 [7][8] が進められている．この技術は本研究でのインタラクションを実行するタイミングを獲得するために活用できる．

人に適応したインタラクションの選択の研究も進んでいる．家庭内での多くの家電機器がユーザを取り巻く状況において，環境の状態（室温，湿度，照度など）と機器の特性をモデル化することによってユーザの状況にあったインタラクションを実行する研究が進められている [9]．

## 3. 想定するアンビエント環境

本研究において，想定する環境はオフィスや研究室といったデスクワークを基本とする場所である．デスクワークをこなしているうちに疲労がたまることによって，眠気を催したり，または居眠りをしてしまう．そのような時に，環境から人を良い状態にする働きかけをするという状況を想定する．このような状況では「人が眠い，眠っていることを検知すること」と「眠っている人を快適に目覚めさせること」をする機能が環境に求められる．

本研究では，人を快適に目覚めさせるため，人にインタラクションを実行する．人を起こすためのインタラクションとして「におい」,「光」,「音楽」の 3 つを用意した．また，これらのインタラクションは次のデバイスを用いて制御する．

- 光：光の照度は覚醒度に大きく影響を及ぼし，照度が高くなると覚醒度が高くなること [10] が知られている．また，光の色温度が高くなることで覚醒度が高くなることが知られている [11]．このことから色温度を変更できる照明を用いた．
- 香り：香料が人のストレスを軽減させること [12] が知られており，人を快適にするには有効な手段である．香りを拡散させるためにアロマ加湿器を用いた．
- 音楽：BGM はコンピュータを用いる仕事において，ストレスを軽減する作用があり [13]，人を快適にする効果が期待できる．そこで，音楽を流すスピーカーを用意した．

飛行機の機内では，乗客を起こす際に音楽を流し，その後暗い照明・明るい照明の順につける工夫をしている例がある．この例からインタラクションを実行する順番も快適さや覚醒度に影響すると考えられるため，3 回のインタラクションからなる系列を，人を目覚めさせるために用いた．快適な目覚め方は個人により異なると考えられることから，その人に適したインタラクション系列の獲得が重要であり，強化学習を用いてそれらの獲得を目指す．

## 4. 提案手法

強化学習によって最適なインタラクション系列を獲得するため，被験者からリアルタイムに得られる覚醒度・快適度を報酬とする Profit sharing によって学習を行った．

強化学習は大きく，Q 学習 [3] などを代表とする環境同定型と Profit sharing[4] を代表とする経験強化型の 2 つに分けられる．環境同定型は，環境の正確な同定が可能であれば，最適な解を得ることが保証される，すなわち最適性をもつ．ただし，その環境の同定には膨大な試行回数が必要となる．加えて，環境の動的な変化に大きく影響を受けてしまう．対して，経験強化型は環境の同定は不要で，行った試行ごとでの評価をするため，少ない試行で学習効果が得られる．また，環境の動的変化に対して，影響を受けた試行のみが影響を受けるため，環境の緩やかな変化にも対応できる．本研究の目的から考えると，人が眠ってしまう回数から考えて，試行回数は少ないものが好ましく，また，人の感情のようなゆらぎのあるものを対象にするため，多少の変化に強いものが求められる．以上より，本研究では経験型強化学習の Profit sharing を用いる．

また，人の状態の指標として，人の快適さと覚醒の程度を数値化したものを使用し，これらの積を報酬として学習を行う．

手法の内容は，与えられた報酬を実行したインタラクション系列に分配し，その選択経路の強化を行う．学習を繰り返すことによって，より報酬を得られる経路が強化されていき，最も強化された経路が最も報酬を獲得できる系列，つまり人を快適に目覚めさせるインタラクション系列となる．図 2 の例では，インタラクション系列 (2，1，3) を選択した時に，報酬 24 が獲得されたとすると，その選択した経路にその報酬が分配され強化される．

本研究で対象にしている状況では，学習回数が少なく，また，日常で使用しながら学習することが考えられるため，最悪なインタラクションは避けるべきである．そこで今回は方策としてボルツマン分布を用いたソフトマックス法を用いる．行動

選択の確率は次式（1）のようになる．

$$P = \frac{e^{\alpha_{(x,y)}/\tau}}{\sum_Y e^{\alpha_{(x,y)}/\tau}} \quad (1)$$

式（1）の $\alpha$ の部分には $a$, $b$, $c$ が入り，$x$ は前回選んだ行動（$a$ の時はなし），$y$ は選択できる行動の集合 $Y$ の中から選んだ行動が入る．

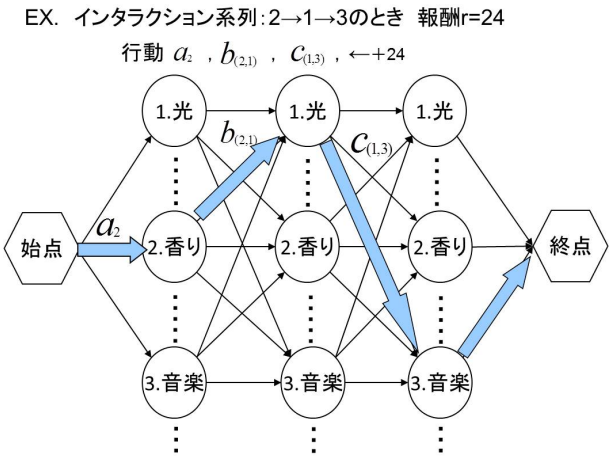


図 2: エージェントの状態図

## 5. 検証実験

提案手法の検証のために，図 3 の 3 つのデバイスを使用して，以下の実験を行った．これらのデバイスからそれぞれ光（白色，黄色），香り（ミント，コーヒー），音楽（穏やか，明るい）を発生させ，計 6 種類のインタラクションを用意した．したがって，エージェントの状態数は 6 × 3 である．

まず，エージェントは 6 種類のインタラクションの中から 1 つを選び，実行する．インタラクションが 10 秒行われたら，次のインタラクションを選択する．そして，同様に 10 秒後に次のインタラクションを選択する．最後に 10 秒経過すると全てのインタラクションを停止する．このとき，一度実行したインタラクションは最後にインタラクションを停止するまで実行し続ける．ただし，同じ種類のインタラクション（ミントの香りとコーヒーの香りなど）は同時に実行されず，前に実行していたインタラクションから次に選ばれたインタラクションに切り替わり，同じインタラクションであればそのまま実行し続ける．この一連の流れをインタラクション系列とし，これによって人を目覚めさせる．

次に，被験者を擬似的に眠い状態に近づけるため，簡単な反復計算（100 マス計算）をしてもらう．単純な作業は人の覚醒度を低下させる効果が示されている [14]．その計算をしてもらった後，決定されたインタラクティブデバイス系列を被験者に実行した．そして，被験者にインタラクション終了時の自身の状態（覚醒度，快適度）を入力してもらう．入力してもらう状態の覚醒度，快適度はともに 1〜7 の 7 段階で評価してもらい，それぞれ目覚めていれば 7，眠いと 1，快適であれば 7，不快であれば 1 と評価してもらった。図 4 に学習段階の流れを示す．

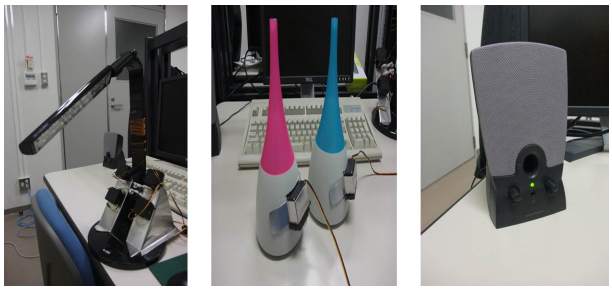

図 3: 色温度可変照明・アロマ加湿器・スピーカー

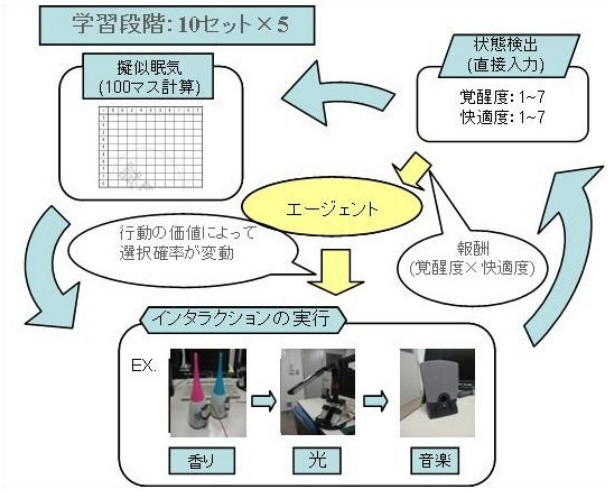


図 4: 学習段階の流れ

図 4 の過程を 10 回繰り返し，時間を空けてから同じ過程を更に繰り返すことで，計 50 回（10 × 5 回）の学習を行い，決定されたインタラクション系列の評価を行った．今回の学習で選ばれたインタラクション系列と比較するために，50 回の試行のインタラクションの中から「報酬が最も大きいもの」「報酬が最も小さいもの」「報酬の中央値を与えられたもの」を選び出した．ただし，それぞれ当てはまるものが複数存在した場合，その中から重複するものは除いて，無作為に選び出した．あらかじめ被験者に，「目が覚めるもの」「心地よいもの」「眠った時に起こされたいもの」の 3 つの条件において，インタラクションの順位付けをしてもらうことを伝えた上で，これら合計 4 つのインタラクション系列を被験者に連続して実行した．実行後，被験者に眠い状況を想定してもらい，よいと思う順番に 4 つのインタラクション系列を順位付けてもらった．

## 6. 実験結果

実験の結果，被験者それぞれで学習結果が異なり，8 名の被験者において 3 名が学習結果を最も好むと評価するなど，強化学習によって，各人の嗜好を学習できる可能性を確認することができた．最も良いと評価された学習結果を表 1 にまとめた．

また，被験者からの報酬値の推移のグラフ（図 5）に見られるように，10 セットごとの報酬値にばらつきがある．これは，実験を 10 セットごとに別の日や時刻に行ったため，それぞれにおける被験者の状態に差が生じ，その状態の差が報酬に影響したためと考えられる．インタラクションを実行する前にも被

験者の状態を入力してもらい，インタラクション前後での状態の差分をとることによって，この問題を回避することが可能であると考えられる．

| | 1 回目 | 2 回目 | 3 回目 |
|---|---|---|---|
| A さん | 音楽（明るい） | 香り（コーヒー） | 香り（ミント） |
| B さん | 香り（コーヒー） | 光（黄色） | 光（白色） |
| G さん | 光（白色） | 音楽（明るい） | 光（白色） |

表 1: 学習結果例

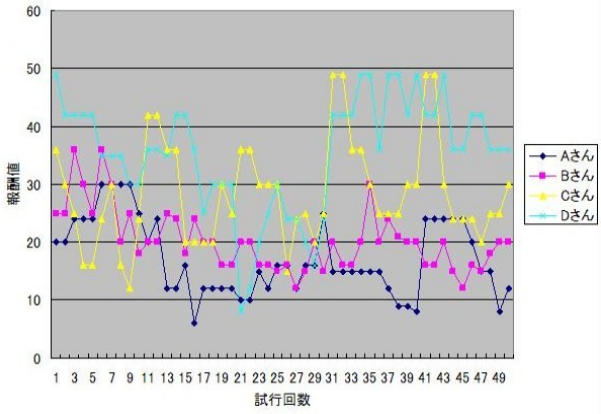


図 5: 被験者からの報酬値の推移

## 7. まとめ

本研究では，デスクワークの中で眠気を催したり，眠ってしまった状況を想定し，その人を快適に目覚めさせるために，個人に適したインタラクションの実行系列を獲得しようと試みた．全ての組み合わせを実行することは試行回数が多くなり，非効率である．そこで，人の状態（覚醒度，快適度）を報酬とした経験型強化学習の Profit sharing を用いて，少ない試行回数で最適なインタラクション系列を獲得する手法を提案した．この提案手法の有効性を調べるために，被験者 8 人に対して検証実験を行った．

まず，眠気を催す状況を擬似的に作るために 100 マス計算をさせた．その後，用意した光（白色，黄色），香り（ミント，コーヒー），音楽（穏やか，明るい）の 6 種類のインタラクションの中から提案手法によって選択されたインタラクション系列を実行し，終了後に覚醒度と快適度を入力してもらう．この一連の流れを繰り返し，入力してもらった覚醒度と快適度の積を報酬として学習を繰り返した．

被験者 8 人に対して行った提案手法の検証実験において，3 人が学習結果であるインタラクション系列を最も良いと評価した．このことから Profit sharing を用いることで，個人ごとの適切なインタラクション系列を獲得できる可能性を確認した．

本研究ではインタラクションを提供するタイミングの検出を，人を 100 マス計算によって擬似的に疲労させることで代用したが，今後の方針としては瞬きや体動，脈拍などの生態信号をセンサによって獲得することによって，その情報から眠気や寝ていることを検知することで，インタラクションを実行するタイミングを検出することが考えられる．また，報酬のばらつきや試行を重ねるごとに報酬値が減っていることから，人の状態によって報酬が正しくインタラクションの評価になっていない可能性が見つかった．このことから，報酬の設定はユーザーの状態の差分を是正する必要がある．

## 8. 謝辞

本研究の一部は，文部科学省グローバル COE プログラム「アンビエント情報社会基盤創成拠点」の支援を受けて行ったものである．

## 参考文献

[1] 大阪大学グローバル COE プログラムアンビエント情報社会基盤創成拠点,http://www.ist.osaka-u.ac.jp/GlobalCOE

[2] 大阪大学グローバルＣＯＥプログラムアンビエント情報社会基盤創成拠点アンビエントインターフェイス領域, 会議資料

[3] Richard S. Sutton, Andrew G. Barto, “強化学習”, 森北出版, 2000.

[4] 宮崎 和光, 木村 元, 小林 重信, “Profit Sharing に基づく強化学習の理論と応用”, 人工知能学会誌, 14(5), pp.800-807, 1999.

[5] Michael Chau, Margrit Betke, “Real Time Eye Tracking and Blink Detection with USB Cameras”, Boston University Computer Science Technical Report, 2005.

[6] 本田 誠一, 福井 健一, 森山 甲一, 栗原 聡, 沼尾 正行, “赤外線センサーネットワークによる人物追跡”, 人工知能学会全国大会 (第 20 回), 2006.

[7] 浜田 尊裕, 白井 了, 小林 史和, 伊藤 丈裕, 足立 和正, 中野 倫明, 山本 新, 井東 道昌, “ドライバの運転状態の検知-個人差に対応した閉眼時間変化からの意識低下レベルの検知”, 画像センシングシンポジウム講演論文集, Vol.9, pp.177-182.

[8] 永作 浩, 屋所 健司, 稲垣 敏之, 古川 宏, 伊藤 誠, “体動情報に基づくドライバの漫然運転リアルタイム検出”, ヒューマンインタフェースシンポジウム論文集, 2005 号, pp.351-356.

[9] 長江 洋子, 山田 松江, 井垣 宏, 青山 幹雄, “連続的アンビエントサービスシステムとホームネットワーク環境による評価”, 情報処理学会研究報告, ソフトウェア工学研究会報告, 2007(33), pp.127-134, 2007.

[10] 道盛 章弘, 荒木 和典, 萩原 啓, 坂口 敏彦, “照度の覚醒度、自律神経活動に及ぼす影響：AAT と R-R 間隔のスペクトル解析による評価”, 照明学会全国大会講演論文集, 30, p.185, 1997.

[11] 道盛 章弘, 荒木 和典, 萩原 啓, 井邊 浩行, 坂口 敏彦, “色温度が覚醒度に与える影響 : 生理指標、心理指標、行動指標による評価”, 照明学会全国大会講演論文集, 31, p.220, 1998.

[12] 細井 純一, 井上 かおり, 庄司 健, 谷田 正弘, 土屋 徹, 津久井 一平, 土師 信一郎, 福井 寛, 福本 正勝, 堀井 和泉, 三浦 靖彦, “香りのストレス緩和効果の血中および唾液中コンチゾールを指標とした評価”, 自律神経誌, pp.260-264, 2002.

[13] 菊田 文夫, “コンピュータを用いた作業に起因する精神的ストレスを軽減させる BGM の効果について”, 聖路加看護大学紀要, 2010.

[14] 橋本邦衛, “安全人間工学”, 中央労働災害防止協会, 1984.