

## ユーザの適応を促す相互適応的インターフェイスシステム

A mutual adaptive interface which can induce a user's adaptation

小松 孝徳\*<sup>1</sup>  
Takanori Komatsu宇都宮 淳\*<sup>1</sup>  
Atsushi Utsunomiya鈴木 健太郎\*<sup>1</sup>  
Kentaro Suzuki植田 一博\*<sup>1</sup>  
Kazuhiro Ueda開 一夫\*<sup>1</sup>  
Kazuo Hiraki岡 夏樹\*<sup>2</sup>  
Natsuki Oka\*<sup>1</sup> 東京大学 大学院総合文化研究科  
Department of System Sciences, The University of Tokyo\*<sup>2</sup> 松下電器産業株式会社  
Matsushita Electric Industrial, Co., Ltd.

The purpose of this paper is to construct a meaning acquisition model as a basic technology for a mutual adaptive speech interface which can communicate smoothly with an everyday user. We assumed that this model should have competences that enable to induce a user's adaptation and utilize this induced adaptation for its meaning acquisition processes. As a result of a testing experiment, we could confirm that our constructed model had the competences that enable to recognize the given everyday users' instructions when appropriate instructions were given to users from an experimenter.

## 1. はじめに

本研究の最終的な目的は、ユーザとの相互適応的な関係を通じて、自然なコミュニケーションを成立させることのできる適応的な音声インターフェイスを構築することである。ここで「相互適応的な関係」[小松 02]とは、ユーザがインターフェイスに対して適応学習し、その適応に基づいてインターフェイスもユーザに適応学習するというプロセスを繰り返すこと、つまり、両者が相手のことを学習し、その学習結果に応じて、自分も相手に対して適応していくことを繰り返すといった「適応のループ」が実現されている関係のことである。このような関係は、コミュニケーションを成立させることのできる二者間(例えば、「人間の親子」や「ペットとその飼い主」など)において共通に見られる現象の一つであると考えられ、また自然なコミュニケーションを成立させることが望まれているユーザと適応的インターフェイスとの間にも同様に存在すべき関係であると考えられる。本稿では、そのような相互適応的なプロセスを成立させることのできる適応的音声インターフェイスの実現に向けた基礎技術の確立を目的とした。具体的には、ユーザの適応を促進し、その適応を利用した意味獲得モデルがユーザの発話意図を学習していくという「適応のループ」の一部分をユーザとモデルの間に実現させることで、ユーザの発話意図を理解する意味獲得モデルを構築した。

また、本研究では、従来の音声インターフェイス研究の多くが注目している音韻情報ではなく、韻律情報に焦点を当てて意味獲得モデルの構築を行なうことも目的とした。音韻情報を用いて発話処理を行なう場合、発話中の音韻情報の意味をあらかじめ辞書のようなデータベースとして用意しておく必要があり、そのため、ユーザはあらかじめ決められたコマンドしか使用することができない。しかし、ユーザと音声インターフェイスが相互適応によってコミュニケーションを構築していくような状況では、あらかじめ決められたコマンドを用いるばかりでなく、この二者が状況に応じて自由にコマンドを設定できた方が良い場合があると考えられる。また、発話コミュニケーション

ンにおける韻律情報の役割に関する研究は近年盛んに行なわれており[林 98, Pirrehumbert 90], ある特定の韻律パターンが聞き手に対して普遍的な解釈をされていることが明らかにされている。さらに、韻律情報に注目することで、音韻情報を扱う場合と比べ少ない計算量で発話を処理することが可能となり、加えて、韻律情報には話者の感情が直接的に反映されやすい性質があるため、音声インターフェイスにおけるインタラクティブな入力手段として有用であると考えられる。よって、本研究は、このような韻律情報に注目した意味獲得モデルを構築することで、現在まで行なわれている音韻情報に注目した音声インターフェイス研究を補完し、かつ、相互適応的なインターフェイスを実現するために重要な役割を果たすと考えられる。

具体的には、以下のようなアプローチによって、所望の意味獲得モデルを開発することとした。

1. 人間を適応する存在として捉え、その認知的な特性に注目する。そしてこの人間の適応能力を具体的に観察、分析する。
2. ユーザの自然な適応を促し、その適応を利用することでユーザの発話意図を理解する意味獲得モデルを構築する。
3. そのような意味獲得モデルの構築において、本研究では韻律情報に注目した開発を行うが、その際に「韻律的特徴と行動との結び付けを学習することによる発話意図の理解」、「韻律情報に含まれる普遍的な情報の利用」という、従来の韻律情報を用いたインターフェイスが注目していなかった事柄を実現する。

## 2. コミュニケーション実験

人間の適応能力を利用しながら発話意味を理解する「学習アルゴリズムの構築」に向けた知見を得るため、人間同士がコミュニケーションを成立する際に、相手に対してお互いがどのように適応しているのかを観察するコミュニケーション実験を行った。本研究では、音韻情報ではなく韻律情報に焦点を当てているため、この実験では、相手が何かを話していることはわかるがその意味はわからないような状況、つまり発話中の音

Contact: 小松孝徳 (公立はこだて未来大学 情報アーキテクチャ学科), 〒 041-8655 北海道函館市亀田中野町 116-2, mailto: komatsu@fun.ac.jp, URL: <http://www.fun.ac.jp/~komatsu>

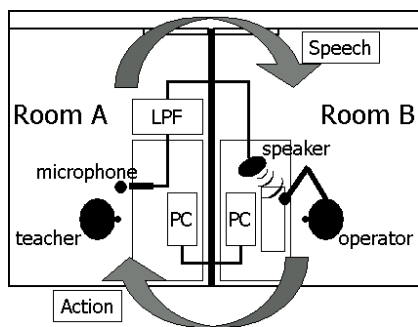


図 1: コミュニケーション実験環境

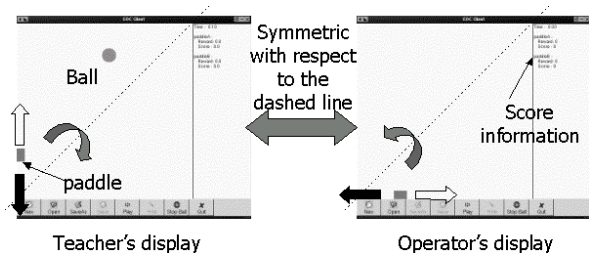


図 2: TV ゲームの画面

韻情報を利用できないような環境を設定し<sup>\*1</sup>、そのような状況において話し手（ゲームの教示者）の発話意味を聞き手（ゲームの操作者）がどのように理解していくのかを観察した。その結果、与えられる未知の音声の意味を獲得し、コミュニケーションを成立させることができた被験者ペアでは、お互いが相手のことを学習しながら相手に徐々に適応していくという相互適応的なプロセスが観察された<sup>\*2</sup>。具体的には、以下のような行動がそれぞれの被験者に観察された。

● 教示者

1. 使用する教示の種類を減少させ、
2. 警告韻律を用いることで操作者の行動に注意を与えていた。

● 操作者

1. 発話を与えられてから行動することで、現在の教示に対する理解度を教示者に示し、
2. 異なる種類の発話に対しては異なる行動をとり、
3. 与えられた警告韻律（突然教示者が声を荒げた時に観察される韻律的特徴）を利用して自分の行動を修正していた（教示者が怒ったような声を出すと「自分の行動は間違っていたようだ」と推測して、自らの行動を修正していた）。

このように、コミュニケーションを成立させることができたペアにおいては、操作者が教示者の自然な適応を促し、同時にその適応を利用していくことで、教示者の発話意図を理解してい

\*1 ゲームの操作者には、ローパスフィルタを通すことで音韻情報を削除した教示音声を与えられた。

\*2 コミュニケーション実験に関する詳しい説明は、参考文献 [Komatsu 02, 小松 03] を参照されたい。

たとえられた。また、その際に、韻律情報における普遍的な機能の一つとみなすことができた警告韻律が、操作者に対して自己の行動への注意として解釈されていたことから、意味獲得プロセスの効率化の点で警告韻律が重要な役割を果たしていることが確認された。

### 3. 意味獲得モデルの構築

前節のコミュニケーション実験の結果から、コミュニケーションを成立させることのできた被験者ペアの操作者の行動を実現できるような意味獲得モデルを構築することで、ユーザの発話意味を理解する学習モデルが実現できると考えられた。具体的には、以下のような条件を満たす意味獲得モデルを提案した。

1. 発話を与えられると、その発話はなんらかの行動を示していると認識する。つまり、発話を与えられなければ、行動は起こさない。
2. 教示区別の決め手となる音響的・韻律的特徴を見つけ出す。
3. 警告韻律を抽出し、意味学習に利用する。

そして実際に、以下のような概要のモデルを構築した。

1. 自分の行動に対して正の報酬を受けたとき（行動が成功したとき-ボールを当てることができたとき）、自分の行動の直前に発せられた教示音声の意味は、自分のとった行動だと認識し、負の報酬を受けたとき（警告韻律を与えられたとき）、教示音声の意味は、自分の行動ではないと認識する。
2. 教示音声はある程度の誤差を持って発せられると仮定する。報酬を受けたときの教示音声・行動のセットは蓄積され（音声-行動データ）、その蓄積されたデータはいくつかのクラスタに分類される。一つのクラスタが一つの教示の意味に相当する。

本モデルでは、正規混合分布から音声-行動データが生成されると仮定したため、それぞれの正規分布のパラメータ（平均値・分散）を求めることが、教示の意味学習に相当する。本モデルでは、EMアルゴリズム [Dempster 77] を用いて、混合分布中の各分布のパラメータを推定した。しかし、従来のEMアルゴリズムでは負の報酬を得た際のデータを学習に利用することはできない。そこで本論文では、従来型のEMアルゴリズムのE stepを拡張することで、負の報酬を受けた音声-行動データを学習に利用できる拡張型EMアルゴリズムを新たに提案した。具体的には、音声データ  $i$  が与えられた時、混合分布中の分布  $j$  に属する行動を出力することで負の報酬を受けた場合（教示者から警告韻律を与えられた場合）、音声データ  $i$  が分布  $j$  に属する確率を示した値（隠れ値）を  $Z_{ij} \rightarrow 0$  とし、残りの分布の隠れ値を  $Z \rightarrow 1/(分布数 - 1)$  と修正することとした（成功例の場合には、従来の方法と同様に現在のパラメータから隠れ値を推定する）。この拡張型EMアルゴリズムを実装した、本意味獲得モデルのアルゴリズムを図3に示す。

### 4. 意味獲得モデルの評価

本節では、前節で提案された意味獲得モデルが、教示者とインタラクションしながら発話の意味を学習していく操作者のモデルとして適しているのかどうかを実験的に検討した。具体

1. 音声 (ピッチ), 行動 (ラケット位置), 報酬データ読み込み
2. 突然 0 になる, あるいは突然大きな値になっているピッチデータを補正
3. 補正データをさらに 0.05[s] 間隔幅の移動平均で円滑化  
→このデータに対して以下の処理を実行
4. 差分, 2 次差分, 極端な変化頻度のデータ, 有声率を計算
5. 教示の開始, 停止のポイントをチェック (1.0[s] の無音区間があれば別教示とみなす)
6. 報酬があった場合
  - 6-1 教示区間の行動をサーチして, 現在に近い行動に重みづけをしてラケットの左右行動値を計算
  - 6-2 それをもとに加重速度を計算
  - 6-3 教示区間の音声データの平均を計算 (6-2, 6-3 がクラスタリングの元データとなる)
  - 6-4 混合ガウス分布を仮定して拡張型 EM アルゴリズムの開始
  - 6-5 計算が一定回数を越える, または, 平均・分散データの変化が一定の値以下になるまで計算をする
  - 6-6 各パラメータを更新
7. 1. に戻る

図 3: 意味獲得モデルの学習アルゴリズム

的には, コミュニケーション実験でゲームの操作者が操作していたゲーム上のラケットに, 提案された意味学習モデルを実装し, 実際に一般のユーザからリアルタイムで教示を受けた際のモデルの学習状況を確認した。

近年, コンピュータ上のエージェントやコンピュータそのものに対して人間が伝言を行うという実験を行った研究 [原田 02] では, その際に人間が「恥ずかしさ」「違和感」を感じてしまうため, 機械に対して与える発話は人間に与えるものと比べて不自然なものになると報告している。ユーザが音声インターフェイスを使用する際に, この「不自然な発話」を使用し続けると, ユーザに精神的な負担が生じてしまうとも考えられるため, その不自然さを解消することは必要不可欠である。この原田の研究では, エージェントにロボットのような物理的実態を持たせることで, エージェントに対するユーザの自然な発話を誘発できると報告しているが, 人間と対峙するエージェント全てが物理的実態を持つことができるとは限らないと考えられるため, そのような物理的実態に依存せずに自然な教示を誘発できるような条件を考察することは重要な課題である。そこでこの評価実験では, その条件の一つとして, 被験者に対して教示対象をイメージさせるようなインストラクションを与えることとした。このインストラクションは「このモデルは, まるで○○のように扱うことができます」といった, キャッチフレーズのようなものに相当するものである。

この評価実験には 10 人の被験者が参加し, そのうち 4 人 (グループ 1) には「コンピュータが動かすラケットにボールがうまく当たるように, このコンピュータに対して教示を与えてください」, そして残りの 6 人 (グループ 2) には「コンピュータが動かすラケットにボールがうまく当たるように, このコンピュータに対して人間と接するように教示を与えてください」というインストラクションを与えた。つまり, グループ 1 の被験者には, 教示対象はコンピュータであるとのイメージが与えられ, グループ 2 の被験者には, 教示対象は単純なコンピュータではないとのイメージが与えられたことになる。この実験の結果からは, ほとんどの被験者が自分の意図どおりにラケットを操作できるようになったことが観察された。しかし, グループ 1 の被験者は, 警告韻律を使用することなく, 自分の用いた教示を一方向的にモデルに対して使用し続けていたのに対し, グループ 2 の被験者は, 警告韻律を使用しながら, 教示の種類を徐々に減少させるといった, コミュニケーション実験で観察された教示者の行動と同じような行動をしていたことが観察さ

れた\*3。ここから, このモデルは, グループ 2 のようなインストラクションを受けた一般的なユーザの適応を促し, かつ同時にそのユーザの適応を利用して, 自然に与えられる発話の意図通りに行動できたことが確認できた。また, その際に, 韻律情報における普遍的な機能の一つとみなすことができた警告韻律を, 意味学習において負の報酬として利用することができた。適応学習した意味獲得モデルに対してユーザがさらに適応していくことを繰り返す「適応のループ」を構成することで, この二者間に真の意味での相互適応が実現できると考えられるが, ここでは, ユーザの適応を促進し, その適応を利用した意味獲得モデルがユーザの発話意図を理解するという「適応のループ」の一部をユーザとモデルの間に実現することができた。

## 5. 議論

この意味獲得モデルによって構築することができた, ユーザとインターフェイス間における「適応のループ」の一部を, 継続的な「適応のループ」として発展させることで, 真の意味での相互適応を可能にするインターフェイスを実現するには, 今後, どのような技術が必要となるのであろうか。本研究で提案した意味獲得モデルでは実現できず, かつ今後も検討が必要だと思われた課題を以下に列挙する。

1. ユーザにとって不自然な動作と感じられるようなモデル出力の回避  
本意味獲得モデルでは, ユーザが不自然に感じるような行動が観察されており (例, 何を言っても一方向にしか動かない, 突然の教示方略の変化に対して異なる反応をすることができない), このことが, モデルに対するユーザの適応を阻害しかねない一要因になっていたと考えられる。これは, 発話の意味学習が EM アルゴリズムという確率的な学習手法に基づいていたことに起因すると考えられるため, 今後はこの不自然な行動を補完する機能を用意することが必要になると考えられる。
2. 行動を意味する教示以外の意味を獲得することができない  
本研究の意味獲得モデルは, 「与えられる発話は何らかの行動を意味している」との仮定に基づいて開発されたた

\*3 実験に関する詳しい説明は, 参考文献 [Komatsu 03] を参照されたい

め、評価(いいよ〜)や副詞的(もうちょっと)な意味を持つ教示の意味を理解することができないという問題があった。今後は、評価教示、行動教示の意味をそれぞれ同時に理解することができるアルゴリズム[鈴木 01]を応用することなどで、評価教示の意味学習も可能とすることが求められる。

### 3. より複雑な状況への適応能力

本意味獲得モデルの能力は、ユーザの意図通りにラケットを上下に動かせるかどうかという単純な動作により確認されたものであるため、本研究で使用したTVゲームよりも複雑な自由度をもつようなインターフェイスに対してこのモデルを応用できるかどうか疑問が残る。そこで、さらに高い自由度を持つ環境に対応するための対処法の一つとして、獲得した発話パターンをシンボル(記号)として表現することで、音韻情報に相当する情報をこのモデル自身が獲得していくようなシステムの導入が考えられる。

これら列挙したような問題点を克服できた場合、本意味獲得モデルは、試行錯誤的な学習状況が許されるような状況で用いられるインターフェイスのための基本的な適応学習モデルとして有効だと考えられる。具体的には、自動カスタマイゼーション機能を持つことで、ユーザの好みを学習していくようなカーナビゲーションシステムや、ペットロボットなど人間とコミュニケーションを構築できるロボットシステムなどといったアプリケーションへの応用が期待される。

## 6. おわりに

本稿では、実験者から与えられるインストラクションという条件付きではあるが、一般のユーザの適応を促し、かつ同時にその適応を利用して、ユーザから与えられる発話の意図を理解することができる意味獲得モデルを開発した。また、その際に、韻律情報における普遍的な機能の一つとみなすことができた警告韻律を、意味学習における負の報酬として利用することができた。このような意味獲得モデルは、インターフェイスの自動カスタマイゼーション技術や、人間とコミュニケーションを成立させることができるペットロボットへの基礎技術として将来的に応用可能であると考えられる。

## 参考文献

- [Dempster 77] Dempster, A., Laird, N. and Rubin, D.: Maximum Likelihood from Incomplete data via the EM Algorithm; *Journal of Royal Statistical Society B*, Vol.39, pp.1-38 (1977)
- [原田 02] 原田悦子: 音声インタフェースにおけるエイジェント性と社会的文脈の効果: 伝言実験による検討; 日本認知科学会第19回大会, pp.14-15 (2002)
- [林 98] 林康子: 感動詞「ええ」におけるピッチ曲線と感情認知: 信学技法, SP98-43, pp.65-72 (1998)
- [小松 02] 小松孝徳・開一夫・岡夏樹: 人間とロボットとの円滑なコミュニケーションを目指して; 人工知能学会誌, Vol.16(7), pp.679-686 (2002)
- [小松 03] 小松孝徳, 鈴木健太郎, 植田一博, 開一夫, 岡夏樹: パラ言語情報を利用した相互適応的な意味獲得プロセスの実験的分析; 認知科学, Vol.10(1), pp.121-138 (2003)

[Komatsu 02] Komatsu, T., Suzuki, K., Ueda, K., Hiraki, K. and Oka, N.: Mutual Adaptive Meaning Acquisition by Paralanguage Information: Experimental Analysis of Communication Establishing Process; in *Proceedings of the 24th Annual Conference of the Cognitive Science Society (CogSci2002)*, pp.548-453 (2002)

[Komatsu 03] Komatsu, T., Utsunomiya, A., Suzuki, K., Ueda, K., Hiraki, K. and Oka, N.: Toward a mutual adaptive interface: An interface and a user induce and utilize the partner's adaptation; in *Proceedings of the 7th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES2003)*, to appear

[Pirrehumbert 90] Pirrehumbert, J. B. and Hirschberg, J.: The meaning of intonational contours in the interpretation of discourse; In P. R. Cohen. and M .E .Pollack.(Eds.), *Intentions in Communication*, pp.271-311. Cambridge, MA: MIT Press (1990)

[鈴木 01] 鈴木健太郎, 植田一博, 開一夫: 評価教示の計算論的意味学習モデル; 認知科学, Vol.9(2), pp.200-212 (2002)