

論文の内容の要旨

論文題目 韻律情報を利用した意味獲得モデルの構築：
適応的インターフェイスの実現へ向けて

氏名 小松孝徳

本論文の目的は、ユーザとの自然で双方向的なインタラクションを通じ、ユーザの発話中の韻律情報と機能との結び付きを学習する音声インターフェイスの基礎技術となるような「意味獲得モデル」を提案することである。具体的には、発話の韻律的特徴の差異に基づいて教示種類を弁別し、その弁別された情報と機能との結び付きを学習することで発話の意味を理解する学習モデルの構築を目的とした。この際、本論文では、コミュニケーション成立時における人間の認知的特性を踏まえたうえで、それを学習モデル作成に生かすというアプローチを取る。

第1章では、人間とのコミュニケーションを目指したロボットなどのインターフェイス開発における技術的な課題について紹介した。その課題を解決すべく、人間のコミュニケーション成立時における認知的特性に注目した上で、人間とコミュニケーションできるような音声インターフェイスの構築を目指す、という本研究の目的が述べられている。その際に、本論文では、話者の意図が直接的に反映されやすい発話音声中の韻律情報の役割に注目した。

第2章では、発話中の韻律情報に関する研究、エージェントによる意味獲得研究など、本研究に関連した先行研究について紹介した。そして、それら先行研究との比較から、本研究には、1) 人間とエージェントとの相互適応的なインタラクションがコミュニケーションに及ぼす影響に注目した点、2) 発話中の普遍的な役割を持つ情報が意味学習プロセスに与える影響に注目した点、といった新規性があることを説明した。

第3章では、本研究の採用したアプローチについて述べた。具体的には、まず、発話中の音韻情報を使用できないような状況に置かれた人間同士が、どのようにして音韻情報を用いずに相手の発話を理解していくのかを観察するコミュニケーション実験を行うことで、コミュニケーション成立時における人間の認知的特性を明らかにする。そして、この実験の結果を基に、教示の種類を発話の韻律的特徴から弁別し、その弁別された情報と機能との結び付きを学習することで、発話の意

味を理解する学習モデルを構築する、という方針を示した。

第4章では、コミュニケーション成立時における人間の認知的特性を考察するために、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を設定し、そこで話し手の発話意味を聞き手がどのようにして理解していくのか解析するためのコミュニケーション実験について説明した。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のできた被験者ペアの意味獲得プロセスにおいて、以下の三点が観察された。

1. 韻律情報による注意喚起

聞き手は未知の音韻情報であっても、その音の「聞こえ方」によって教示の種類を理解していた。また、教示音声の韻律情報の一つである高いピッチ成分が、聞き手に対して注意を喚起していたことが観察された。

2. 複合報酬による強化学習的な意味獲得プロセス

聞き手は、与えられる音声教示と自分の行動とを対応させて教示の意味を獲得していた。ここから、本実験における意味学習プロセスは、行動が成功したことにより得られる正の報酬と、話し手の音声の高いピッチから与えられる負の報酬という、二種類の報酬を利用した強化学習的なプロセスであると考えられた。

3. 相互適応的な行動の変化

話し手・聞き手双方の行動において、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。

第5章では、第4章のコミュニケーション実験で観察された人間の発話理解プロセスを基に、発話中の韻律情報と聞き手の行動（インターフェイスで言えばその機能）とを結び付けることで、発話の意味理解を可能とする意味獲得モデルを提案した。具体的には、次のような事柄を満たす学習モデルを提案した。

1. 自分の行動に対して正の報酬を受けた時、自分の行動の直前に発せられた教示音声の意味は、自分のとった行動を指示していると認識し、負の報酬を受けたとき、教示音声の意味は、自分の行動を指示していないと認識する。
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たと仮定したため、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学習に使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。

第6.1節では、第5章で提案し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。まず、コミュ

ニケーション実験で人間の聞き手が操作していたラケットに、提案された意味学習モデルを実装し、モデルの学習能力を熟知した理想的な教示者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。その結果、意味学習モデルでは、以下の点が実現できていたことが確認された。

1. 与えられた発話の韻律的特徴から、その差異を見出すことによる教示種類の弁別。
2. 成功・失敗例を報酬として活用することによる、行動を通じた未知の教示意味の獲得。
3. 警告韻律でモデルに負の報酬を与えることで、モデルのパラメータを局所解(例えば、全てのデータが一つのクラスターで説明されてしまうような状態)から脱出させることができ、その結果、最適解に達するまでの学習が継続可能なこと。

ここから本モデルは、理想的な教示を与えられた際には、それらの教示の意味を学習できたといえる。

しかし、このような技術が実際のインターフェイスに実装された場合、それを使用するユーザは先のような理想的な話し手(教示者)であるとは考えにくい。そこで第6.2節では、一般的なユーザから発話を与えられた場合でも、この意味獲得モデルはその発話の意味を獲得できるのか確認する実験を行った。また、人間はコンピュータに対して自然に話しかけることが難しいと言われているが、このような状況においても自然な発話を誘発することのできる条件を考察することは重要である。本実験ではその条件として、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に注目した。その結果、以下のことが明らかになった。

1. 人間同士のコミュニケーション実験と同様の設定にした被験者、および「人間と接するように教示して下さい」とインストラクションを受けた被験者は、コミュニケーション実験でも観察されていたような「教示者の操作者に対する適応学習プロセス」を経ながら、教示の意味をモデルに獲得させていた。また、被験者はその際にストレスを感じないような自然な発話を使用していた。
2. 一方、「コンピュータに対して教示して下さい」とインストラクションを受けた被験者も、モデルに意味を獲得させることに成功していた。しかし、実験中において、二種類の行動教示を根気よく同じように与え続けた結果として、自分の行いたい教示を行えずにストレスを感じていたことが報告されていた。よって、このインストラクション条件は、「ユーザが自然に使用できる」といった観点からすると望ましいものではないと考えられる。
3. 「ペットに対して教示して下さい」「しつけるように教示して下さい」といったインストラクションを受けた被験者は、行動教示と共に評価教示を使用し続けていた。このため、全ての入力音声は何らかの行動を意味している、という仮定の基に構築された本意味学習モデルでは、このようなユーザの発話意味を学習することができなかった。

以上よりこのモデルは、一般的なユーザから自然に与えられる発話の意味を獲得する能力があることが確認された。また、ユーザが自然にコンピュータと接するためには、適切なインストラクションが与えられる必要があることが確認され、このような知見は、適応的インターフェイスの使用に当たってのキャッチフレーズ(例えば、「人間に対して話しかけるように!」)として有効に活用できよう。ただ、本意味獲得モデルは確率的な学習手法を使用しているため、被験者が今までと大き

く異なる教示を与えた場合に、直ぐにその教示に反応して今までと違う行動を取る、といったことが実現できていなかった。よって、人間同士のコミュニケーション実験で観察されたような、相手の行動に対して逐次反応するという密接な相互適応を実現するためには、突然の教示の変化にも対応できるような学習機能を追加するなどの拡張が必要であると考えられる。

本論文では、発話中の韻律情報が持つ普遍的機能を利用した確率的な学習手法を用いた意味獲得モデルを提案し、このモデルは実際のユーザから与えられる発話の意味をインタラクティブに理解できることを確認した。また、人間同士が行うような相互適応プロセスのように密接なインタラクションを人間-エージェント間を実現するには、開発された確率的な学習手法を用いた意味獲得手法だけでは十分ではないことも明らかにできた。ここから、本研究は人間とコミュニケーションを成立させることのできるロボットなどのインターフェイス実現への基礎的な知見として、HAI 技術の確立に大きく寄与すると考えられる。