階層ディリクレ過程隠れマルコフモデルを用いた正直シグナルのモデル化

Modeling of Honest Signals Using Hierarchical Dirichlet Process Hidden Markov Model

片上 祐介[1*] 阿部 香澄[1] アッタミミ ムハンマド[1] 長井 隆行[1] 中村 友昭[1]
Yusuke Katakami, [1] Kasumi Abe, [1] Attamimi Muhammad, [1] Takayuki Nagai, [1] Tomoaki Nakamura[1]

[1] 電気通信大学
[1] The University of Electro-Communications

Abstract: Recent studies have shown that we unconsciously use the signals that represent our purposes and/or intents to communicate with each other. We called these signals as "Honest Signals." In this study, a sociometer is used to measure the human interaction. Using captured data, we model the interaction based on multimodal hierarchical Dirichlet process hidden Markov model. We then implement the model to the robot. Thanks to the model, robots can generate "Honest Signals" which correspond to their partners; and conduct interaction in natural manners.

1 はじめに

多様なコミュニケーションスタイルが混在する現代において，相手の思考や目的を理解することは，他人とコミュニケーションをとる上で重要である．そのような相手の心的な部分を汲みとることは我々人間同士の対話だけでなく，近年ではロボットと人との対話においてもより重要視されつつある．それによってロボットがより人らしい対話を行うことが目指されてきた．従来の研究では，発言内容といった言語的な情報や，振る舞いや表情といった言語情報の単純な補助要素から対話相手の心的な変化や対話の意図を予測し，それらをロボットに応用してきた．

例えば，表情認識を用いて対話中の相手の感情を認識し，ロボットに同調させたり [1]，韻律などの非言語情報より相手の発言に隠された感情や意味合いを抽出してきた [2]．また楽しい，痛い，辛いなどの言葉そのものがもつ正負の印象を識別することで，発話からポジティブ，ネガティブの印象を抽出し，ロボットはこうした情報を用いて発話することでより自然で，共感的なコミュニケーションが実現してきた [3]．

しかしこれらの要素は「嘘をつく」,「作り笑顔をする」といった自らの意図的な行動で造り出せる要素であり，相手を真に理解する上で判断の難しい要素とも言える．これらの曖昧な判断をもとにした情報をロボットに用いてきたがゆえに，未だにロボットと人との対話の中には違和感や不自然さが残っていると，我々は考える．

そこで本研究では普段我々が対話を通して無意識に感じ取る部分に注目する．例えば対話相手の「高圧的な態度」や対話中の「重苦しい空気」といった，対話の中で自然と形成されていく部分にこそ，相手が伝えたい真の意図や目的が隠されているはずである．実際，人間は対話においてこれらの意図や目的を正直シグナル [4] と呼ばれるシグナルに変換し，無意識のうちにやりとりすることで円滑なコミュニケーションを実現するとされる．

そこで本稿ではこれらの無意識のシグナリングのやりとりを階層ディリクレ過程隠れマルコフモデルを用いてモデル化する．実際の対話において人間が放つシグナルをロボットが受け取り，それに応じてモデルからロボットが発するべきシグナルを抽出する．それらをロボットを通して発信することで，より人間とロボットのより自然なコミュニケーションの実現を目指す．

2 正直シグナル

対話において，人間は自分の考えや目的を無意識のうちにシグナル化し，互いにやりとりしている [4]．これらは単純な言語情報や非言語情報とは別であり，生物学的な部分に根ざしたものであるため，私達の行動に直接強い影響を与えるコミュニケーションツールで

*連絡先：電気通信大学情報理工学研究科知能機械工学専攻長井研究室
〒 182-8585 東京都調布市調布ヶ丘 1 丁目 5 – 1
E-mail: katakami@apple.ee.uec.ac.jp

表 1: 社会的役割の特徴

社会的役割	シグナルの組み合わせ
打診	高い活動レベル　+　低い一貫性 (高い変動性)
能動的傾聴	低い活動レベル　+ 低い一貫性 (高い変動性)
協調	高い影響力　+　豊富なミミクリ + 高い一貫性
主導	高い影響力　+ 高い活動レベル + 高い一貫性

ある．つまり相手の話す言語や，対話相手の非言語的な行動そのものの意味は分からずとも，相手の意図や目的を理解するための主体となるものである．例えば第 1 章で例として挙げた対話中に起こる「重苦しい空気」はこれらのシグナルのやりとりで生まれ，対話中の相手の感情や目的が見え隠れしている．

2.1　基礎シグナル

正直シグナルは単に 1 つのシグナルではなく，影響力，ミミクリ，活動レベル，一貫性と呼ばれる 4 つの基礎シグナルから構成されている．1 つ目の「影響力」は注意のシグナルであり，主に対話相手への関心や注意レベルを測定できる．これは対話相手の発話タイミングをいかにコントロールできるかで計測する．例えば自らの発言中，対話相手に喋らせる隙を与えない人は，その対話において影響力は高い人だと言える．2 つ目の「ミミクリ」は他人の行動や発言を反射的に模倣する度合いで計ることのできる共感のシグナルである．例えば日常の会話の中で見られる相手に OK?と聞かれ，反射的に “OK!” と答えることもミミクリの一種と言える．3 つ目の「活動レベル」は動きや声の大きさから，人間の興奮や関心の度合いを測るシグナルである．興味，関心あるものを目にしてはしゃぎ出す子供は，自律神経が興奮状態にあり，活動レベルが高い状態にあると言える．4 つ目のシグナルは「一貫性」と呼ばれるもので，発言に対する決意の強さと精神的集中を示す．これは声の高さや大きさなどに一貫性があるかどうかで計る．例えば相手を説得する時に自然と話し方にムラが無くなるのは，発言時に自分の思考や決意がしっかりしていることを示している．

2.2　社会的役割

2.1 節で説明した基礎シグナルは通常の対話において単独で使われることはなく，複数で組み合わされて使われることが多いとされる．そうすることで人間は状況に合わせて，4 つの役割を対話の中で無意識のうちに演じている (表 1). 1 つ目は「打診」である．相手の興味，関心，思考を探る態度であり，相手と相互作用を深めたい時に演じる役割である．初対面の人間と話す時，相手のことを詳しく知ろうと質問したり，それに対し相手の反応を伺うような態度がこの役割にあたる．2 つ目は「能動的傾聴」である．これはやりとりのほとんどを聞き手側に回る時に見られるもので，相手の情報をオープンに求める態度を指す．3 つ目は「協調」で，主に相手の話や意見への支援や共感を示す役割である．4 つ目は「主導」と呼ばれるもので，場の流れや話しの意見を自分の思うように導きたい場合に見られる．

このように対話中において，対話者間で複数のシグナルが行き交いすることで互いの意図や目的を対話に反映させている．相手から受け取るシグナルに対し，それに応じて複数のシグナルを出すことで人間特有の自然なコミュニケーションが成立する．

3　正直シグナルの検証

本研究において，ロボットは自律で相手のシグナルを感知することが求められる．またそれらのデータから相手の放つ複数のシグナルを理解し，それに応じて自分のとるべき振る舞いを決める必要がある．

そこでまずは人間同士で対話実験を行い，対話からシグナルを数値化して抽出する．そうすることで正直シグナルの有無とロボットが自律でシグナルを判別できるかを検証した．

3.1　対話実験

本実験では被験者として 2 人 1 組のペアを 4 組用意し，4〜5 分の短い対話を 3〜5 回行った．なお被験者はいずれも 21〜24 歳の大学生，大学院生である．また対話時は単純な雑談だけでなく，特定の社会的役割が見られるようなテーマも複数用意し，テーマに沿った対話も複数回行ってもらった．この時，被験者の胸につけた独自のセンサ (以下ソシオメーター) から体の動きの情報と音声情報を記録した．またカメラを用いて被験者の全体の様子を記録した (図 1).

3.2　解析・評価手法

対話実験後，評価者には採取した全 15 対話分のデータ (約 75 分) 中の対話の様子をカメラで確認しながらデータを切り取り，切り取った区間内の被験者の社会的役割を選択してもらった (図 2)．ラベルとして 2.2 節で述べた 4 つの社会的役割に加え,「判定不可」の計 5 つを用意した．なおデータを切り取るタイミングや長さは全て評価者に任せてある．ここで各被験者の発話

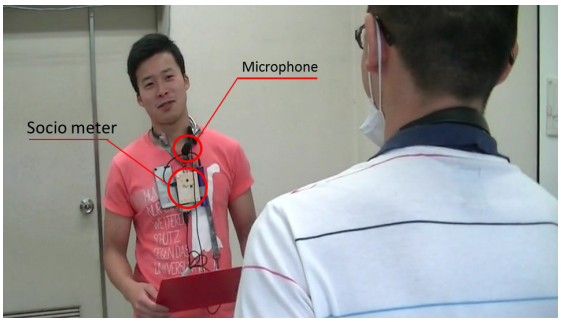

図 1: 対話実験における対話データ観測の様子

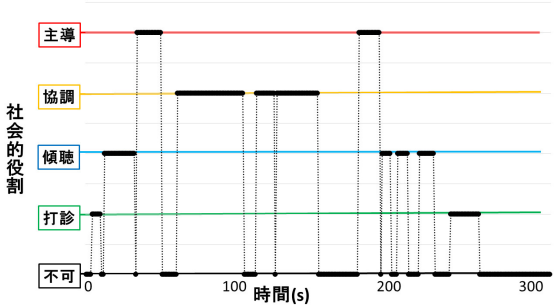

図 2: 対話中の被験者の役割遷移

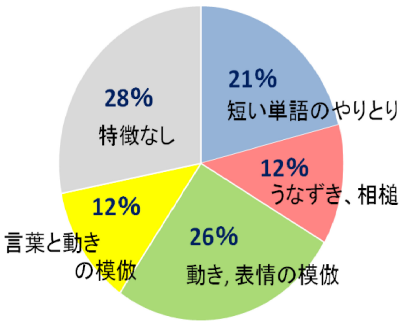

図 3: 「協調」時のミミクリ行動の分類

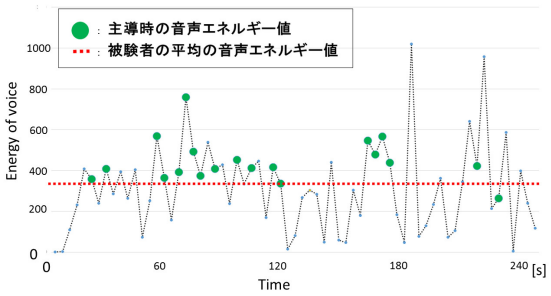

図 4: 対話中の時系列の音声エネルギー値 (主導)

の音量 (音声エネルギーの平均)，発話のばらつき (音声エネルギーの標準偏差)，発話時間，1 回の発話の平均長さを算出し，対話全体を通して見た時の各役割中の基礎シグナルの有無を検証する．

3.3 解析結果

対話データから基礎シグナルの有無を検証した結果を表 2 に示す．なお表 2 の括弧内の数値は基礎シグナルの特徴を確認できたデータの割合 (各役割の全データに対する) である．活動レベルは音声エネルギーの高低，一貫性は前 4 秒間の音声の分散値との比較，ミミクリはビデオ検証によるミミクリ行動の有無，影響力は前 4 秒間の自分と相手の発話量の増減よりその高低を検証した．主導，能動的傾聴では役割を構成する基礎シグナルを比較的はっきりと確認することができた．例えば主導時であれば，通常より音声エネルギーは高くなり，活動レベルの上昇が見て取れる (図 4)．打診では，活動レベルに関してはデータごとで多少のムラはあるものの，一貫性の特徴は 6 割以上のデータで確認することができた．

一方協調におけるミミクリに関しては音声から特徴的な値を抽出することができなかった．しかしビデオによる確認を行った際には，短い言葉の反復などのミミクリ動作を見て取れた．また音声だけでなく動作の模倣など動きの情報の有用性を見て取ることができた (図 3)．

以上の結果より，実対話において人間は複数のシグナルを発信していると言えるだろう．相手のシグナルを受け取り，自分の目的に合わせその時々でシグナルを変化させていると言える．また，今回の実験からそれらを数値化して捉えることができるとわかった．これは，対話における情報からシグナルに関する特徴選択を行うことで，ロボットが相手の出すシグナルを判別することができる可能性を示していると言えるだろう．

4 正直シグナルのモデル化

本章では 3 章で示した正直シグナルを実際にロボットが扱い，やりとりするために正直シグナルのモデル化を行う．そのために，各シグナルを特徴量として扱い，multimodal hierarchical Dirichlet process hidden Markov model（MHDP-HMM）(4.2 を参照) を用いてロボットが自律で自分の出すべきシグナルを生成する．そうすることで，ロボットが相手の出すシグナルを受けて， 自分の行うべき振る舞いを選択することを目指す．

表 2: 役割毎の特徴検証結果

社会的役割	各役割における基礎シグナルの有無
打診	高い活動レベル (5 割) ＋ 低い一貫性 (6 割)
能動的傾聴	低い活動レベル (8 割) ＋低い一貫性 (7 割)
協調	高い影響力 (6 割) ＋ 豊富なミミクリ (6 割) + 高い一貫性 (7 割)
主導	高い影響力 (9 割) ＋高い活動レベル (7 割) + 高い一貫性 (7 割)

4.1 シグナルの分類

本稿では，ロボットが人と対話を行うときに得られるマルチモーダル情報をカテゴリ分類することで，対話中におけるシグナリングをモデル化する．このモデルを用いることで，我々は相手のシグナルから自分の出すべきシグナルを予測することができる．

本研究でシグナルの分類を行う理由としては，実対話において人間が発信する正直シグナルの組み合わせの多様さにある．Pentland ら [4] は，対話におけるシグナルを社会的役割として 4 つのカテゴリに分類している．しかし実際の対話において全てのシグナルをそれら 4 つのカテゴリに単純に当てはめていくことは難しい．なぜなら実際の対話では同じシグナルの強弱の表現でも，相手の状態によって強弱そのものに若干の差が出てきてしまう．つまり，一般的には同じ「能動的傾聴」と呼ばれる態度でも，分類されるカテゴリは複数存在することになるため，相互的なシグナルを分類していく上でカテゴリ数を事前に指定することは難しい．よって無次元の潜在状態を仮定し，データに応じて状態数を決定していくような柔軟なモデルを用いる必要があると言えるだろう．

4.2 MHDP-HMM

4.2.1 生成モデル

Hidden Markov model（HMM）は，マルコフ過程によって遷移する状態と，各状態から独立に出力される観測によって構成される確率モデルである．この HMM にディレクレ過程を導入し，無限の状態を持つモデルへと拡張したものが HDP-HMM である．HDP-HMM の各状態から複数の観測を仮定したマルチモダル HDP-HMM（MHDP-HMM）のグラフィカルモデルは図 5 に示す．この図において，$(s_0, s_1, \cdots, s_T)$ は対話中における状態を表している．また，各状態から出力される観測値は，自分の動き（図中の m^1_*），音（図中の v^1_*）と，相手の動き（図中の m^2_*），音（図中の v^2_*）である．各状態 $s_t (t = 0, \cdots, T)$ は無限の状態 $k \in [0, \infty)$ をとることができ，π_k が状態 k から各状態間へ遷移する確率を表している．この π_k は，γ をパラメータとする GEM 分布によって生成された β と，α をパラメータとするディレクレ過程から生成される．

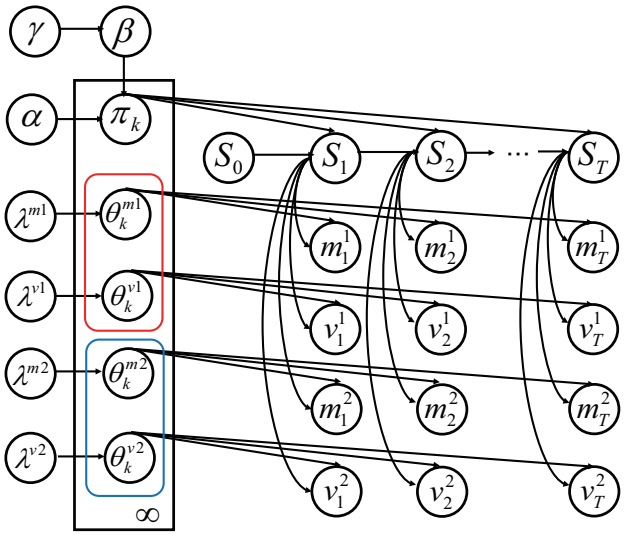

図 5: Multimodal Hierarchical Dirichlet Process Hidden Markov Model のグラフィカルモデル

$$\beta \sim GEM(\gamma) \quad (1)$$

$$\pi_k \sim DP(\alpha_0, \beta) \quad (2)$$

時刻 t の状態 s_t は，$t-1$ の状態 s_{t-1} と，遷移確率 π_k によって決定される．また，θ^*_* は，観測値を生成する確率分布のパラメータであり，本稿では正規分布の平均と分散を仮定している．さらに，正規分布の事前分布として，正規・ウィシャート分布を仮定しており，そのパラメータが λ^* である．

$$s_t \sim \mathcal{M}(\pi_{s_{t-1}}) \quad (3)$$

$$\theta^*_k \sim P(\theta^*_k|\lambda^*) \quad (4)$$

$$m^*_t \sim \mathcal{N}(m^*|\theta^{m*}_{s_{t-1}}) \quad (5)$$

$$v^*_t \sim \mathcal{N}(v^*|\theta^{v*}_{s_{t-1}}) \quad (6)$$

ここでは，遷移確率 π_k と正規分布のパラメータ θ^*_* を学習データから推定する．

4.2.2 モデルの学習

モデルの学習はギブスサンプリングにより，各時刻 t の状態 s_t をサンプリングすることで実現する．ギブスサンプリングでは，s_t を除いた残りを条件とした以下の条件付き確率から s_t をサンプリングする．

$$P(s_t|\boldsymbol{s}_{-t}, \beta, \boldsymbol{M}^1, \boldsymbol{V}^1, \boldsymbol{M}^2, \boldsymbol{V}^2, \alpha, \boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2) \propto$$
$$P(s_t|\boldsymbol{s}_{-t}, \beta, \alpha) \times$$
$$P(\boldsymbol{m}^1_t|s_t, \boldsymbol{s}_{-t}, \boldsymbol{M}^1_{-t}, \lambda^{m1}) \times$$

$$P(\boldsymbol{v}_t^1|s_t, \boldsymbol{s}_{-t}, \boldsymbol{V}_{-t}^1,\lambda^{v1}) \times$$
$$P(\boldsymbol{m}_t^2|s_t, \boldsymbol{s}_{-t}, \boldsymbol{M}_{-t}^2,\lambda^{m2}) \times$$
$$P(\boldsymbol{v}_t^2|s_t, \boldsymbol{s}_{-t}, \boldsymbol{V}_{-t}^2,\lambda^{v2}) \quad (7)$$

ただし，$\boldsymbol{M}^*$，$\boldsymbol{V}^*$ はそれぞれ，全観測データの集合であり，$\boldsymbol{\lambda}^* = (\lambda^{m*},\lambda^{v*})$ とする．また，負の添字は時刻 t の状態を除いた残りのデータ集合を意味しており，$\boldsymbol{s}_{-t}$ は s_t を除いた全時刻の状態，$\boldsymbol{M}_{-t}^*$，$\boldsymbol{V}_{-t}^*$ はそれぞれ $\boldsymbol{M}^*$，$\boldsymbol{V}^*$ から m_t^*，v_t^* を除いた残りのデータ集合を表現する．この式において，$P(\boldsymbol{m}_t^*|s_t, \boldsymbol{s}_{-t}, \boldsymbol{M}_{-t}^*,\lambda^{m*})$ と $P(\boldsymbol{v}_t^*|s_t, \boldsymbol{s}_{-t}, \boldsymbol{V}_{-t}^*,\lambda^{v*})$ は，ベイズ推定よりそれぞれ以下のように求められる．

$$P(\boldsymbol{m}_t^*|s_t, \boldsymbol{s}_{-t}, \boldsymbol{M}_{-t}^*,\lambda^{m*}) = \int P(\boldsymbol{m}_t^*|s_t,\theta_{s_t}^{m*})P(\theta_{s_t}^{m*}|\boldsymbol{s}_{-t}, \boldsymbol{M}_{-t}^*,\lambda^*)d\theta_{s_t}^{m*} \quad (8)$$

$$P(\boldsymbol{v}_t^*|s_t, \boldsymbol{s}_{-t}, \boldsymbol{V}_{-t}^*,\lambda^{v*}) = \int P(\boldsymbol{v}_t^*|s_t,\theta_{s_t}^{v*})P(\theta_{s_t}^{v*}|\boldsymbol{s}_{-t}, \boldsymbol{V}_{-t}^*,\lambda^*)d\theta_{s_t}^{v*} \quad (9)$$

また，状態遷移確率である $P(s_t|\boldsymbol{s}_{-t},\beta,\alpha)$ は，n_ij を状態 i から j へ遷移した回数とする．

学習は，ランダムな初期値から始め，式 (7) によるサンプリングを繰り返すことで，遷移確率 $P(s|\boldsymbol{s},\beta,\alpha)$ と，その状態と対応した観測値を出力する確率分布である $P(\boldsymbol{m}_t^*|s, \boldsymbol{M}_{-t}^*,\lambda^{m*})$ と $P(\boldsymbol{v}_t^*|s, \boldsymbol{V}_{-t}^*,\lambda^{v*})$ を得ることができる．また，本稿ではハイパーパラメータ α，β もサンプリングすることで推定を行なっている [6].

4.2.3 モデルを用いた予測

学習したモデルを用いることで，相手の振る舞いから自分が取るべき行動を予測することが可能となる．例えば，ある時刻 t において，相手の動き m_t^2 と音 v_t^2 が観測された場合，自分が取るべき動き m_t^1 と音 v_t^1 は次のように推定することができる．

$$m_t^1 \sim \sum_{s_t} P(m_t^1|\theta_{s_t}^{m1},\lambda^{m1},\alpha)P(s_t) \quad (10)$$

$$v_t^1 \sim \sum_{s_t} P(v_t^1|\theta_{s_t}^{v1},\lambda^{v1},\alpha)P(s_t) \quad (11)$$

ただし，$P(s_t)$ はギブスサンプリングより求められ，具体的には以下の式を用いて状態 s_t をサンプリングする．

$$P(s_t) \propto P(s_t|\beta,\alpha)P(m_t^2|s_t,\lambda^{m2})P(v_t^2|s_t,\lambda^{v2}) \quad (12)$$

4.3 特徴量の抽出

続いて本稿における正直シグナル検証に使用する特徴量について説明する．本稿では 3 章の対話実験で得られたデータを用いて，モデルに使用する特徴量を生成する．手順としては，3 章の対話実験で得られたデータを一定の時間幅で区切っていき，一定の情報量から特徴量を生成していく．今回はデータを 1 秒ずつずらしながら 30 秒間隔でデータを抜き出し，特徴量を生成した．特徴量を構成する要素としては，音声の情報として 30 秒間での音声エネルギーの平均値，音声エネルギーの分散値，発話時間，の 3 つを用いた．また動きの情報として，30 秒間での運動エネルギーの平均値と運動エネルギーの分散値の 2 つを用いる．

また 4.1 節で示したように正直シグナルは相手の情報も含めた相互的なシグナルであるため，対話相手の情報も必要となる．そこで対話相手の観測データからも上記の 5 つの要素を同様に生成していき，自分と相手の情報を合わせたものを 1 つの特徴ベクトルとして扱う．

4.4 提案モデルによる対話データの分類

ここでは提案モデルの分類精度について検証する．

まず 3 章で観測した対話データのうち 9 対話分 (約 40 分) のデータから特徴量を抽出し，提案モデルを用いてカテゴリ分類を行った．しかし提案モデルは教師なし学習を行っているため, 各カテゴリ内にどのようなデータが分類されているかわからない．そこで各カテゴリ内のデータがどのようなものか確認するために，各データに以前行ったデータのラベリング結果 (3.2 節参照) をもとにしたラベリングを行った．分類に使用するデータは自分と対話相手両方の情報を用いているため，ラベルは単一なものでなく,「主導と傾聴」といったように複数の役割の組み合わせになる．

次にこれらのラベリング結果を用いて，提案モデルによるデータ分類が正直シグナルによる分類とどの程度一致するかを検証した．データの一致率はカテゴリ内に最も多く含まれるラベルを正解ラベルとしたとき, 各カテゴリ内の正解ラベルのデータ数の総和を全体のデータ数で割った値で算出される．このとき不可のラベルを含む曖昧なラベルのデータは考えないものとした．

検証した結果を図 6 に示す．下のバーは分類に使用したデータを時系列で並べ，不可を含むラベルの付いたデータを除いた結果である．各データはラベル (役割の組み合わせ) 毎に色分けがしてある．上のバーは下のバーの各データの状態を示したものである．図は各状態 (カテゴリ内の正解ラベル) に合わせて色分けがしてある．

結果，提案モデルによるデータの一致率は約 7 割となった．つまり提案モデルでは対話における約 7 割のデータ (複数のシグナルの組み合わせ) を分類することが可能なモデルと言える．

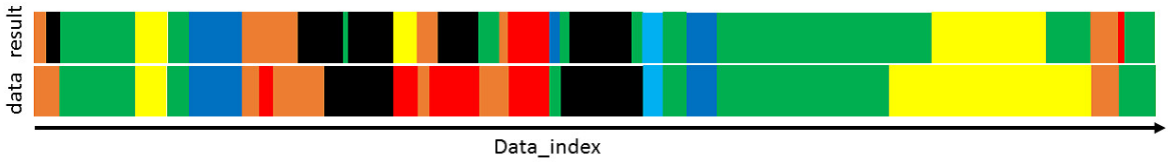

図 6: 実データとカテゴリ内データとの一致率

図 7: 正直シグナルロボットによる対話実験の様子

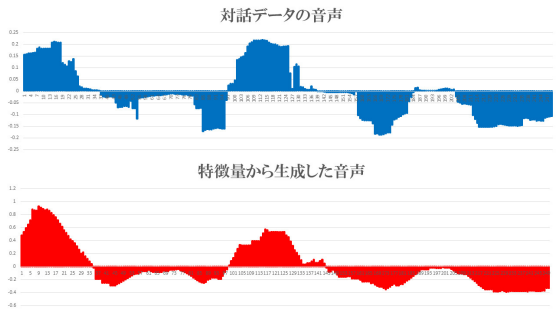

図 8: 時系列の音声エネルギー値の比較

5　実験

本章では 4 章で示した提案モデルを実際にロボットが扱い，シグナルのやりとりができるかを検証した．それと同時に正直シグナルが人間同様にロボットにとっても有用なものであるかを検証した．

5.1　正直シグナルロボットによる対話実験

本実験では 4 章で得られた特徴量をもとに，実際の人間同士の対話における正直シグナルのやりとりを 2 台のロボット同士で再現した (図 7)．そのやりとりの様子を評価者に観察してもらい，対話が終了するごとに各対話におけるロボットの印象評価を行った．対話の再現するにあたり，本稿では時系列に並んだ特徴量をロボットの音声や動きを決定するパラメータに変換し，それを 1 秒ごとに更新していくことでロボットの音声や動きを再現した．また実験で用いる特徴量を複数用意し，ロボットを通して複数の対話シーンを再現した．

5.2　ロボットによる正直シグナルの再現

実験を行うにあたり，特徴量をロボットの音声と動きに変換する必要がある．そこで，まずは音声を作成するために単一の周波数波形を出し続ける音源を 2 種類用意した．今回は 200Hz と 400Hz のビープ音を用意した．これらを特徴量に合わせ 1 秒ごとに波形を変化させていくことで，シグナルに応じた音声の変化を再現した．具体的な手順としては発話時間から，その 1 秒間での発音の有無を決定した．今回は前後 1 秒間の差が負であれば，音源の振幅を 0 とし，その 1 秒間では発音しないものとした．発話時間の差が正であれば，次に音声エネルギーの値に合わせて，音源の振幅の大きさを決定した．最後に音声エネルギーの分散値の値に合わせて，別の周波数の正弦波を音源に掛け合わせることで，人間の音声のムラを表現した．これらの手順で作成した音声は元の音声の変化をうまく表現していることがわかる (図 8)．

また動きに関しても，音声と同様に 1 秒ごとにロボットに入力する値を変え，体の回転で動きの変化を再現した．運動エネルギーの大きさから回転角の大きさを，また運動エネルギーの分散値から回転速度の調整を行うことで人間の動きの大きさや変化を再現した．

5.3　実験条件

今回，検証を行うにあたり対話を再現する特徴量を 5 パターンで生成した．なお特徴量生成する元となる対話は talk1〜5 まで全て同じものを使用した．

【talk1】: これは，3 章で観測した対話データから抽出した時系列の特徴量を両方のロボットにそのまま用いる．そうすることで，人間のシグナルの相互的なシグナルのやりとりをそのままロボットを介して表現した．

【talk2】: これは，talk1 で用いたデータの時系列をランダムに入れ替えたものである．対話相手に合わせ

た相互的なシグナルのやりとりでなく，適当なシグナルのやりとりの様子を再現した．

【talk3】: これは，互いの出す音が重なることなく，どちらか一方のみが発音する様子を表したものである．人間の対話においてターンテイキングの様子がはっきりした対話の様子を再現した．

【talk4】: 提案モデルを用いて一方の人 (A さん) のデータから，もう一方の人 (B さん) が取るべき行動を予測したものである．学習には 9 対話のうち 8 対話分のデータを使用し，talk1 で使用した対話を認識用として使用し，行動の予測に用いた．つまり，ロボット 1 には A さんの対話データから抽出した特徴量をそのまま使用し，ロボット 2 には提案モデルで生成した特徴量を用いた．

【talk5】: talk4 とは逆に提案モデルを用いて一方の人 (B さん) のデータから，もう一方の人 (A さん) が取るべき行動を予測したものである．行動予測の流れは talk4 と同じである．ロボット 1 には提案モデルで生成した特徴量を，ロボット 2 には B さんの対話データから抽出した特徴量をそのまま使用した．

5.4 評価アンケート

ロボットに対する印象評価のために，以下のような質問項目を用意した．各項目は 5 段階で評価するものとし (5:当てはまる，4:少し当てはまる，3:どちらともいえない，2:あまり当てはまらない，1:当てはまらない)，評価者である 22〜24 歳の大学生または大学院生 9 人が回答した．また以下の項目以外にも，事前アンケートとして，普段どの程度ロボットに触れ合っているかなどロボットの知的さへの理解度を調べるアンケートも行った．また実験終了後にもロボットの表現したシグナルをどのように捉えたのかを知るためにアンケートを行った．

【実験に関するアンケート項目：1 対話ごと】

Q1. 動画を見てロボットを単純なロボットと思えないと感じた．

Q2. このロボットに人間らしさを感じた．

Q3. ロボットに感情的な変化があると感じた．

Q4. 動画をみてロボットに幼さや大人っぽさを感じた．

Q5. 機械的な印象を感じる．

Q6. 知的なロボットだと感じた．

Q7. このロボットは自分の意思を持っていそうだ．

Q8. このロボットは何かしら意図的なコミュニケーションを行っている．

Q9. このロボットは人間と意思疎通ができそうだ．

Q10. このロボットは適当に動いているように見える．

Q11. このロボットは互いに双方向なコミュニケーションをしている．

Q12. 何かしら相互的なやりとりをしているように見える．

Q13. このロボットは相手の情報を汲み取って動いていそうだ．

Q14. 一方的なやりとりしか行われていない．

【実験に関するアンケート項目：対話実験終了後】

Q1. ロボットの出していた音について感じた項目に○をつけてください (複数回答可).

(単音だった，強弱があった，高低があった，発音間隔にリズムがあった，発音間隔に波があった)

Q2. その他にロボットの出していた音について何か感じたあれば記入をお願いします．

Q3. ロボットの動きについて感じた項目に○をつけてください (複数回答可).

(単調な動きだった，強弱があった，動きの間隔にリズムがあった，動きの間隔に波があった，相手の動きを真似ていた)

Q4. その他にロボットの動きについて何か感じた点があれば記入をお願いします．

5.5 アンケート結果

まずは，ロボットコミュニケーションにおける正直シグナルの有用性について検証を行う．比較に用いたのは talk1，talk2，talk3 のデータである．図 9 は talk1，2，3 の各質問項目における評価の平均値と標準偏差を表している．また検証を行うにあたり質問項目ごとで各対話に対し対応のある t 検定を行った．これらの図より Q8，Q12，Q14 の各項目において talk2 と talk3 の間に有意傾向を見られた．この結果より正直シグナルの相互性を考慮せずに動きや音を生成した talk2 が，より人間らしく意図的かつ相互的なコミュニケーションを行っているように感じることを示している．一方で対話全体を通してみると「全体的に単調な音が続き，機械的に感じる」という意見があった．このことから他の対話にない talk2 における単純な連続した短い音の切れ目が人間特有の声のムラに受け取られてしまったと考えられる．この結果より，言語情報を除いた人間特有の声のムラやゆらぎを今回の手法ではうまく表現しきれていないと言えるだろう．

逆にターンテイキングを意図的に加えたものの方が一方的なやりとりをしている印象を被験者に与えてしまった．このことから単純なターンテイキングを加えるだけでは，人間の対話のような相互的なやりとりを再現できないことがわかる．今回，対話中の人間の正直シグナルのやりとりを再現した talk1 のデータは他と何かしら有意な差は見られなかったが，影響力の強弱から生まれる時に互いに言葉が重なるような場面も相互的なやりとりを実現する上で必要な要素と考えられる．

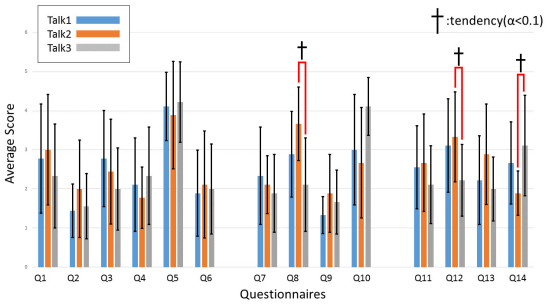

図 9: アンケート結果 (talk1， 2， 3 比較)

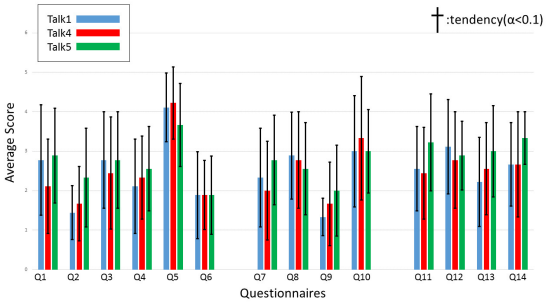

図 10: アンケート結果 (talk1， 4， 5 比較)

また各質問項目の平均スコアを見てみると，全体的に人間らしい知的さを示す値が小さく，機械的といった印象値のスコアが高いことから，今回実験においてロボットのコミュニケーションにおける正直シグナルの有用性を示すには難しい結果となった．

また提案モデルの評価を行うために，提案モデルで生成した特徴量をロボットに用いた talk4，talk5 と talk1 との評価を比較した．図 10 は talk1，4，5 の各質問項目における評価の平均値と標準偏差を表している．また検証には前述の検証と同様に質問項目ごとで各対話に対し対応のある t 検定を行った．検定を行った結果，こちらは特に各対話間で有意な傾向や差は見られなかった．talk1 との評価値との誤差も全体的に少なく，提案モデルによる行動の予測自体はうまくいっていると言えるだろう．

6 まとめ

本稿では正直シグナルと呼ばれる，人間の意図や目的が反映された無意識のシグナルを観測し，そのデータを階層ディリクレ過程隠れマルコフモデルでモデル化することを試みた．

結果として，実際に人間は相手の状態に応じ正直シグナルの相互的なやりとりを行っていることがわかった．またそれらを数値化し特徴量とすることで，ロボットが対話相手のシグナルを受け取り，提案したモデルを用いて自分のとるべき行動を選択することができるとわかった．

しかし，本稿においてロボットを正直シグナルを用いることの有用性を示すことはできなかった．正直シグナルを音声や動きで表現したロボットを用いた評価実験では，人間のような知的さを評価するスコアの値は小さく，逆に機械的な印象を示す値は高いままであった．アンケートの回答にも多く見られたが，評価者が一定のルールに基づき作成される音に対し単調さを感じてしまったことが要因として挙げられる．また評価者によっては本来同期すべき動きの音の連動が感じられないという意見もあり，動きと音で表現されるシグナルをうまく再現できていない可能性が考えられる．

今後の取り組みとしては，まずはロボットで正直シグナルで再現する方法をもう一度検討する必要があるだろう．また，それと同時にミミクリなど数値化し捉えることのできてないシグナルも特徴量として抽出していく必要もある．

その後に再びロボットを用いた対話実験を行い，正直シグナルのロボットへの有用性を示すとともに，人とロボットとの円滑な対話の実現を目指す．

参考文献

[1] 山野美咲，薄井達也，橋本稔：情動同調に基づく人間とロボットのインタラクション手法の提案，HAI シンポジウム 2008，2-D-4 (2008)

[2] 多田和彦，矢野良和，道木慎二，大熊茂：感情遷移における急激な韻律特徴変化の検出による感情遷移判別法，知能と情報 (日本知能情報ファジィ学会誌)，Vol.22，No.1，pp.90–101 (2010)

[3] 大竹裕也，萩原将文：評価表現による印象推定と傾聴型対話システムへの応用，知能と情報 (日本知能情報ファジィ学会誌)，Vol.26，No.2，pp.617–626 (2014)

[4] Alex (Sandy) Pentland: HONEST SIGNALS - How They Shape Our World-. The MIT Press (2008)

[5] 片上祐介，阿部香澄，アッタミミムハンマド，長井隆行：人とロボットの対話における正直シグナルの利用，第 14 回計測自動制御学会システムインテグレーション部門講演会，2G2-4 (2014)

[6] Y. W. Teh，M. I. Jordan，M. J. Beal，D. M. Blei: Hierarchical Dirichlet processes，Journal of the American Statistical Association，Vol.101，No.476，pp.1566–1581 (2006)