

第9章

ヒューマンロボットインタラクションへの応用

本章では、これまで構築したロボット聴覚システムのヒューマン・ロボットインタラクションへの応用を行う。主に6章で説明した実時間人物追跡システムを利用し、何に注意を向けるかを決定する注意制御モジュールに、インターパーソナル理論に基づいたパーソナリティを導入し、より豊かなインタラクションを目指す。

9.1 選択的注意とパーソナリティ

表情や声の調子を用いた感情認識や生成も、ソーシャルインタラクションの重要な要素である [23, 22, 76] が、本章では、選択的注意に注目したインタラクションに注目する。一般に、単に見たり聞いたりすることに対し、注視、傾聴といった選択的注意はソーシャルインタラクションで重要な役割を演じている [102, 123]。

選択的注意は、人間の聴覚では、例えばパーティのような騒がしいところでも、特定の会話に注意することができ、適宜、別の会話に参加することができるという能力が知られている。これは、カクテルパーティ効果 (*cocktail party effect*) と呼ばれ、1953年に Cherry により報告された [30]。このような機能をロボットで実現する際には、雑音下の認識ということだけではなく、どこに注意を向け、どのように注意を切り替えるのかということも大きな問題である。

こういった問題に対し、パーソナリティという概念を導入することは有効であろう。ソフトウェアエージェントを対象としたパーソナリティの研究は広く行われている。例えば、Bates らのグループは、適当なタイミングで、明確に感情を表現するような *believable agents* を提案している [19]。Loyall と Bates は、音声もしくは、テキストを介して、人間が不信感を抱かずに、楽しんでインタラクションすることができるキャラクタをコン

ピュータ上に構築した [72]. また, そのキャラクタは, 人間を介さないコンピュータ上のキャラクタ同士のインタラクションでも, 人を惹きつけることができる. Cassell は人間同士のインタラクションの研究に基づいて, 顔とジェスチャー認識を統合した会話エージェントを開発した [28]. Hayes-Roth は, 即興, もしくはある筋書きに沿って行動することができる知的で自律的なキャラクタを作り出す研究をするために, 仮想シアタープロジェクトを立ち上げている [42].

こうしたソフトウェアエージェントの研究の中で, Reeves と Nass は ソフトウェアエージェントのパーソナリティ解析に, 心理学で用いられる *Big Five* 理論における *Five-Factor* モデルを使用した [104]. *Big Five* 理論では, パーソナリティの次元は外向性 (*Dominance / Submissiveness*), 協調性 (*Friendliness*), 誠実性 (*Conscientiousness*), 情緒安定性 (*Emotional Stability*), 開放性 (*Openness*) で構成される. これらの 5 つの次元は, エージェントにも基本的なパーソナリティとして定義されている.

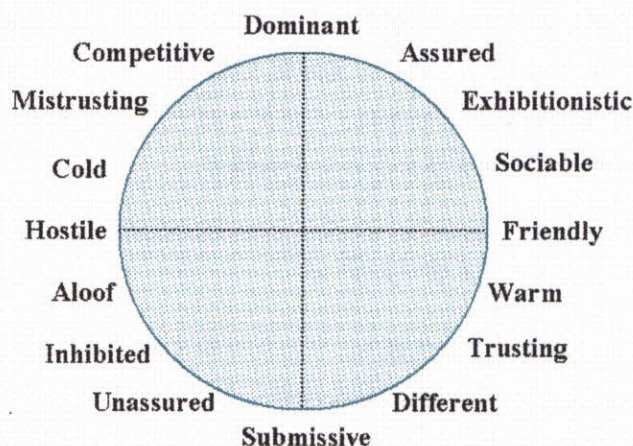


図 9.1: Interpersonal Circumplex によるパーソナリティの表現

これに対し, ロボットを対象としたパーソナリティの研究は少ない. 小野らは *Robovie* を用いて, ジェスチャーによる共同注意研究を行っている [100]. Breazeal は, 顔の表情や声の調子で, 感情の認識と生成を行う機能を *Kismet* に実装した [23, 22]. Waldherr らは, ポーズやジェスチャーを認識する *AMELLA* を開発している [120]. 松阪らは, 人間の視線の抽出と音声認識を用いて, 複数の人間とインタラクションを行うようなロボット *ROBITA* を開発している [74]. これらは, 人間とロボットのコミュニケーションを豊かにするという意味でパーソナリティについて暗に触れているだけであり, 現時点では, 物理的な身体を持つエージェント, もしくはヒューマノイドロボットには, これまで明確なパーソナリティに対する研究が行われていない.

一つの理由として、*SIG* のように多くのロボットは、表情の表出ができないため、ソフトウェアエージェントで利用されている *Big Five* 理論は、軸が多すぎて、適用が難しいことがあげられる。そこで、*SIG* では、同じく心理学で利用されるインターパーソナル理論 (*Interpersonal Theory*) を使用する。これは、人間のインタラクションのパターンを外向性 (*Dominance/Submissiveness*) と協調性 (*Friendness/Hostility*) という 2 つの軸を利用して扱う。図 9.1 は、これを *interpersonal circumplex* によって図示したものである。

図 9.1 では、外向性 (*dormant*) と協調性 (*friendly*) という 2 つの独立な軸を用いて *Dominant*, *Assured*, *Exhibitionistic*, *Sociable*, *Friendly*, *Warm*, *Trusting*, *Different*, *Submissive*, *Unassured*, *Inhibited*, *Aloof*, *Hostile*, *Cold*, *Mistrusting*, *Competitive* といった様々なパーソナリティを表現することができる [64]。

また、言語は、最も重要な人間のコミュニケーション手段の一つではあるが、非言語の感覚情報に基づく行動も重要である。非言語情報でもパーソナリティを表現することはでき、より豊かなインタラクションが可能であろう。

そこで、本章では、注意制御にインターパーソナル理論を導入して、人間とロボットの非言語インタラクションを行う場合について論じる。

9.2 パーソナリティに基づく注意制御

パーソナリティの導入は、6 章で説明した実時間人物追跡システムに対して、注意制御部を改良することによって行う。注意制御部では、聴覚ベース、視覚ベース、およびアソシエーションストリームから、1 本のストリームを選択し、そのストリームに注意を向けることによって行われる。そこで、パーソナリティに基づいたストリーム選択のルールを導入した。

パーソナリティは図 9.1 のように、一つの円 (*circumplex*) 上に表現されるので各パーソナリティは、円内の点 (r, θ) として表現できる。ただし、 $0 \leq r \leq 1$, $0 \leq \theta \leq 2\pi$ とする。従って、2 つの軸 (外向性と協調性) の値は、それぞれ、 $r \cos \theta$, $r \sin \theta$ として表現することができる。また、上述の各パーソナリティは中心角 $\pi/8$ の扇形として表現することができる。例えば、友好的なパーソナリティは $-\frac{\pi}{16} \sim \frac{\pi}{16}$ として表現でき、支配的なパーソナリティは $\frac{3\pi}{16} \sim \frac{5\pi}{16}$ として表現することができる。

ここで、興味という内部パラメータを導入し、注意制御モジュールは、最も興味の大きいストリームを選択するものとした。興味の絶対値の総和は、システム内で常に一定に保たれ、興味の更新は、以下のルールに従うものとした。

1. 新しいストリームへの興味 — 新しいストリームが作られた場合、他のストリームに対する興味は等しく、 $r \cos \theta$ 倍されることにより、興味が失われる。新しいスト

- リームは、他のストリームが失った興味の全量を与えられる。これは $r\cos\theta$ が大きい場合は、注意が、新しいストリームに移ることを意味している。ストリームがアソシエーションされた場合は、各ストリームの興味はすべて、アソシエーションストリームに与えられる。
- 2. 興味の減衰 — 注意を向けたストリームの興味は時定数 $1/k$ の指数関数 (e^{-kT}) に従って減衰する。ここで k は、協調性を示す軸の値を利用して $1 - r \sin \theta$ と定義する。失われた興味は他のストリームに等しく分配される。
 - 3. 確信度の減衰 — 一時的に沈黙していたり、視野外にはずれたため、センサ情報イベントが観測されなくなるストリームがある。そのようなストリームに対しては、確信度が減衰する。興味の減衰と同じ時定数をもつ指数関数として定義した。

これにより、 r, θ を指定することによって、様々なパーソナリティを持ったインタラクションを実現することができる。

9.3 パーソナリティ導入による様々なインタラクション

4 章で紹介した部屋を利用して、インタラクションの例を示す。以下の 4 つのケースについて実験を行った。ただし、実装したシステムでは、興味の減衰、確信度の減衰の 2 つについては用いず、新しいストリームへの興味のみについて実装した。

	想定ケース	パーソナリティ	(r, θ)
1	受付として	Dominant	$(1, \pi/2)$
2	ステレオ定位の追跡	Dominant	$(1, \pi/2)$
3	コンパニオンとして	Friendly	$(1, 0)$
4	そっぽを向く SIG	Hostile	$(1, \pi)$

9.3.1 受付ロボット

受付業務は、特定の筋書きにそって行動するため、パーソナリティを Dominant に設定した。 $(r, \theta) = (1, \pi/2)$ なので、既にストリームに注目している場合には、新しくできたストリームに注意を向けないようになっている。つまり、一度注意を向けてしまえば、そのストリームが消滅しない限り、一つのストリームを注目し続けるようになっている。

受付のシナリオは、以下の通りである。

- ケース 1：既知の訪問者
 - 1. ロボットが知っている (顔データベースに登録済み) 訪問者が受付にやってくる。

2. 参加者は SIG に“こんにちは”と話しかける.
3. SIG は“こんにちは, XXX さんですね?”と確認する.
4. “はい”と答える.
5. SIG は“XXX さん, どうぞお入りください.”と答え, 受付作業を終了する.

● ケース 2: 未知の訪問者

1. ロボットが知らない (顔データベースに未登録) 訪問者が受付にやってくる.
2. 参加者は SIG に“こんにちは”と話しかける.
3. SIG は“どちらさまですか?”と名前を尋ねる.
4. 訪問者は名前を答える.
5. SIG は, 顔と名前をデータベースに登録し, “こんにちは, XXX さん, どうぞお入りください.”と答え, 受付作業を終了する.

ケース 1, 2 のスナップショットを, それぞれ, 図 9.2, 9.3 に示す.

図 9.2 a) は訪問者がやってきて声をかけた状態を示している. この時点では, 訪問者が視野外にいるため, SIG はまだ気づいていない. SIG に話しかけたとき音モジュールによって生成された音源方向を伴う聴覚イベントがアソシエーションモジュールに送出され, 音ストリームが生成される. このストリームが引き金となり, 注意制御モジュールは SIG がそちらに向くように行動を制御する.

図 9.2 b) では, SIG は, 訪問者の顔を捉えたので, 顔ストリームが生成される. また, 同時に音ストリームとアソシエーションされ, アソシエーションストリームが生成される. 一方で, 顔認識を行った結果, 登録されている情報に該当者があったため, 名前の確認を行う. SIG の声は, 図中のスピーカから発生するようになっている.

図 9.2 c) では, 訪問者が名前が正しいことを告げ, それを確認した SIG が挨拶をする (図 9.2 d)).

SIG は, 図 9.2 e) で, 訪問者に部屋に入るように案内し, 図 9.2 f) で訪問者が部屋に入るにより, SIG の受付業務が終了する.

図 9.3 a) はケース 1 と同様, 訪問者がやってきて声をかけた状態である. 訪問者が SIG に話しかけた, 音モジュールによって音ストリームが生成され, SIG は音ストリームの方向を向く.

図 9.3 b) では, SIG は, 訪問者の顔を捉え, ケース 1 と同様にアソシエーションストリームが生成される. しかし, 顔認識を行った結果, 登録されている情報に該当者がいなかった (閾値以下) ため, 図 9.3 c) で, 訪問者に名前を尋ねる.

図 9.3 d) は, 訪問者の名前の音声認識を行い, SIG は名前と顔を登録する.

ケース 1 と同様, SIG は, 訪問者に部屋に入るように案内し, 受付業務を終了する (図 9.3 e), f)).

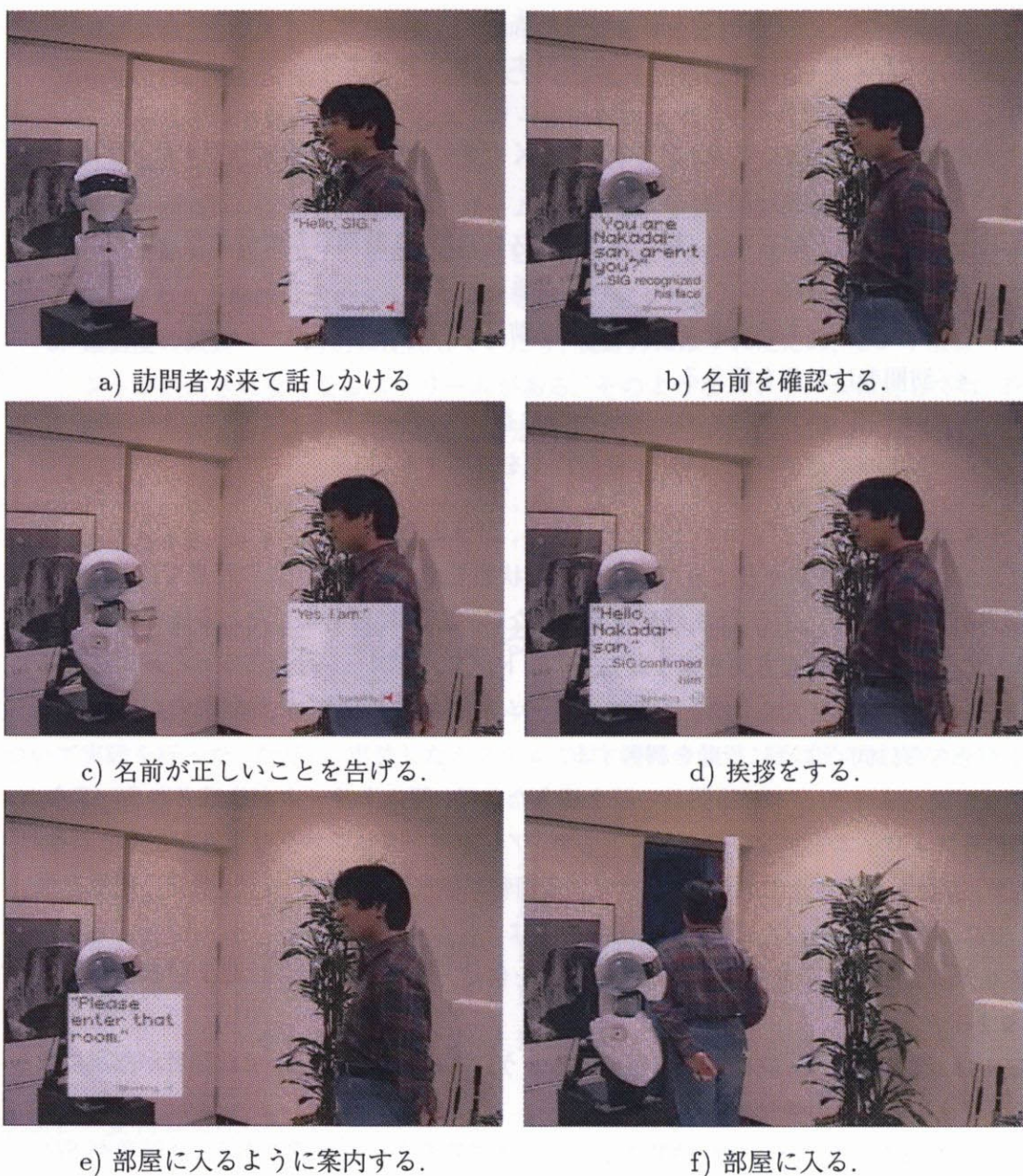


図 9.2: 受付ロボットの作業の流れ (既知の場合)

ケース 1, 2 とともに 相手の人間に注目し, 与えられたタスクを遂行することができている. 設定したパーソナリティが Dominant であるため SIG は自分の声 (スピーカ) によって作られる音ストリームには注意を向けないようになっていることがわかる. ケース 2 において, 名前の認識には, ロボット自身のマイクを使っており, また, 音声認識用のロボットの言語辞書には, 事前に名前情報が登録されている. 完全に未知の言葉 (*Out of Vocabulary, OoV*) の検出は難しく, 他の研究に委ね, 本研究では扱わない.

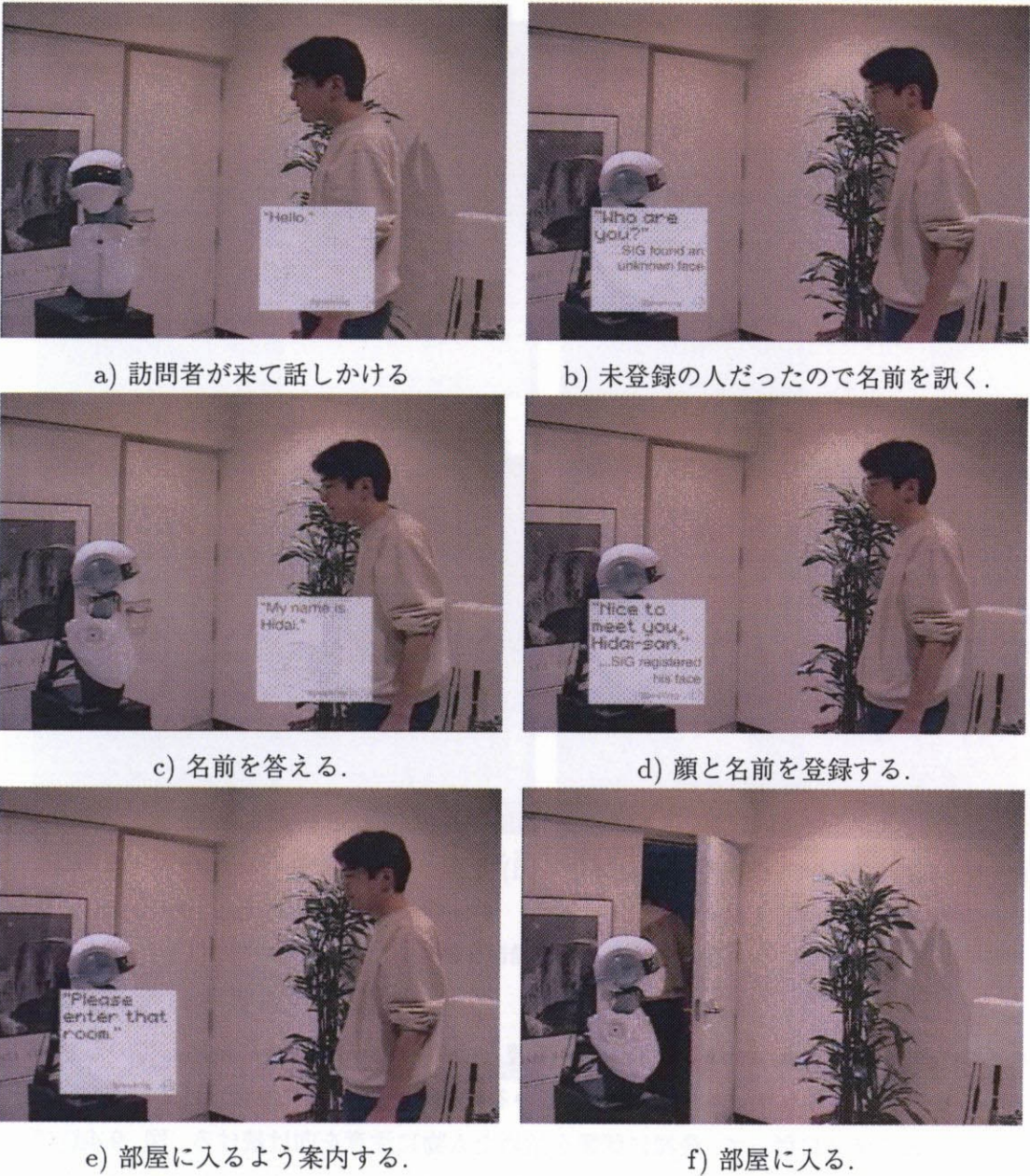
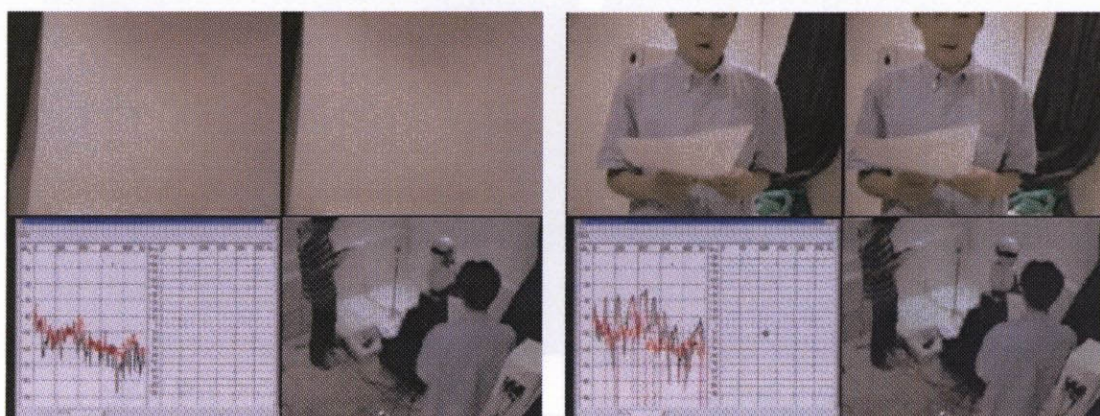


図 9.3: 受付ロボットの作業の流れ (未知の場合)

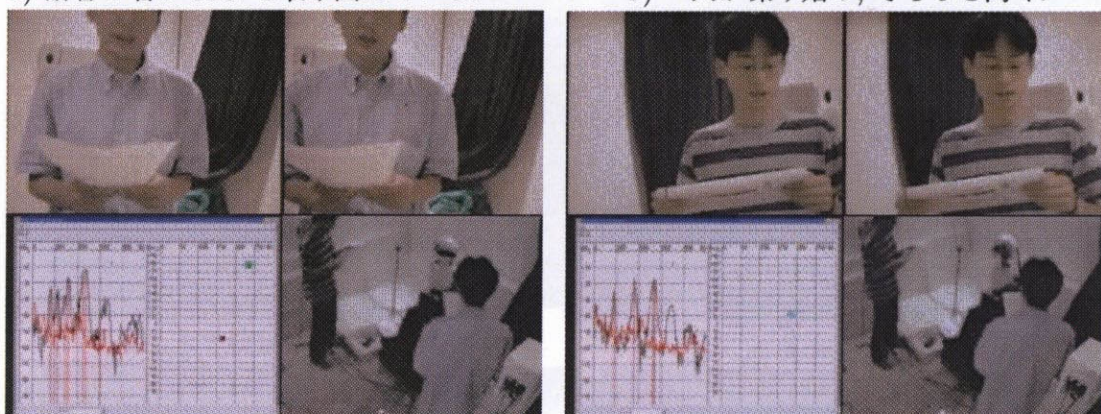
9.3.2 同時発話の追跡

パーソナリティを Dominant に設定し、2 話者が同時に発話する場合を紹介する。状況は、図 9.4a) のように、2 人の話者が SIG を取り囲んでいる。まず、図 9.4b) のように、うち一人が、朗読を始める。この話者に対する音ストリームが生成され、SIG は、音ストリー



a) 話者 2 名が SIG を取り囲んでいる。

b) 一人が喋り始め、そちらを向く。



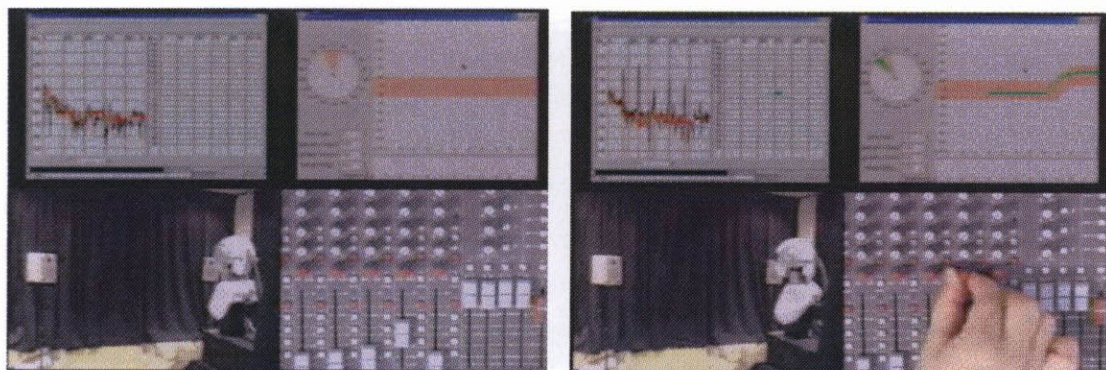
c) 2 人が同時に喋っているが、最初に話し始めた人に注目し続ける。

d) 正面の人が話をやめたため、もう一人に向きを変える。

図 9.4: 2 人の同時発話を聴く SIG

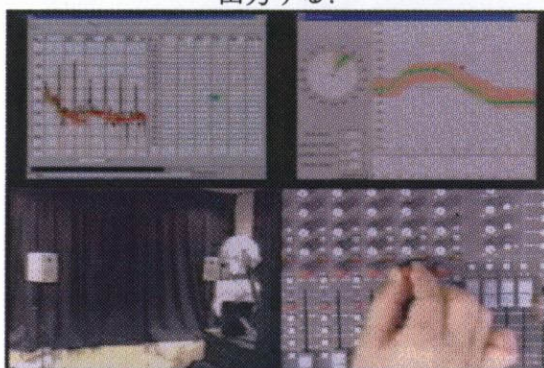
ムの方向を向く。図 9.4c) では、もう一人も朗読をはじめ、2 人が同時に発話している状態になる。つまり、音ストリームは 2 本できている状態となる。この場合、SIG は、設定されたパーソナリティに従って、最初に朗読を始めた人物に注意を向け続ける。図 9.4d) では、注意を向けていた人が朗読をやめたため、朗読を続けているもう一人の人物に向きを変える。

この例では、最初に現れたストリームに注目し続けることがわかり、Dominant のパーソナリティが顕著に現れている例といえる。また、同時に 2 人が発話している場合でも、SIG が適切に動作することができることも示している。



a) 2つのスピーカから、同じ音を等音量で出力する。

b) バランスを制御し、定位を変える。



c) SIG はステレオ定位を追跡する。

図 9.5: ステレオスピーカの定位を追跡する SIG

9.3.3 ステレオ定位の追跡

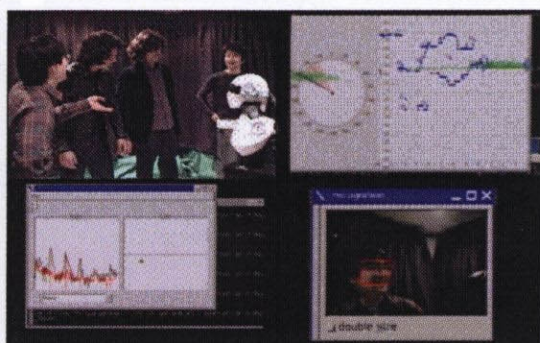
同じく、パーソナリティを Dominant に設定して、ステレオ定位を追跡する場合を紹介する。

状況は、図 9.5a) の左下のように、部屋の 2 点 (SIG に対し $\pm 45^\circ$ の位置) にスピーカを設置した。スピーカからは、同じ基本周波数の調波構造を有する音が、等音量で出力される。この場合、SIG は 2 つの音を 1 つの音と認識し、その定位は、人間と同様、正面方向となる。

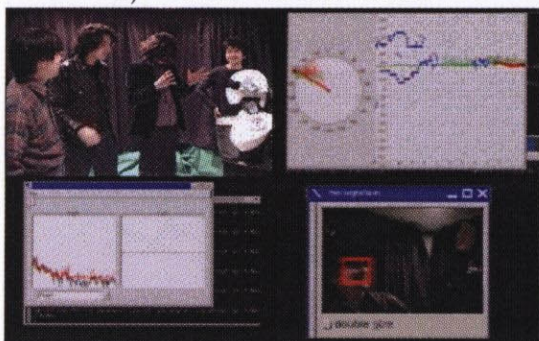
スピーカの音量は、図 9.5a) の右下のミキサーを介してバランスコントロールが可能になっている。従って、図 9.5b) 右下のようにミキサーのバランスを制御することによって 2 つのスピーカの音量を調節でき、定位を移動させることができる。SIG は、この定位の移動に対して、図 9.5c) の左下、および右上のストリームチャートに示されるようにパーソナリティに応じて、正確に、かつ、まじめに追跡を行っていることがわかる。



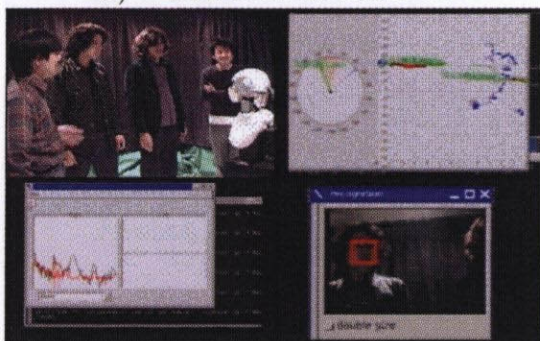
a) 一番左の人の顔を追跡.



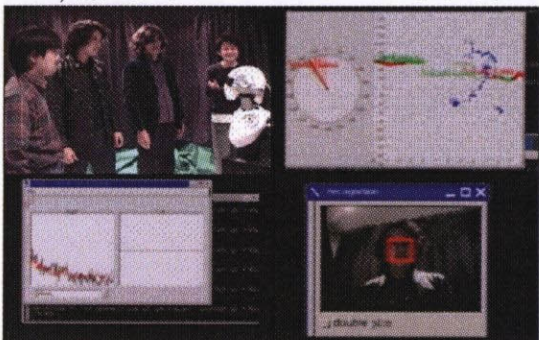
b) 一番左の人が話し始める.



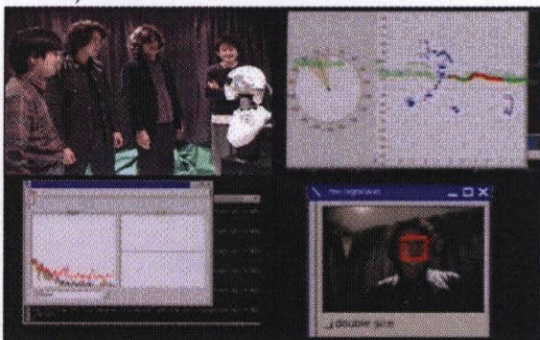
c) 顔と声がアソシエーションされる.



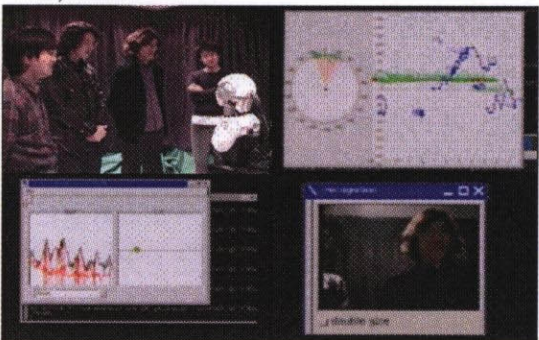
d) 左から2番目の人に注意が移る.



e) 顔と声がアソシエーションされる.



f) 左から2番目の人が話を止める.



g) 右から2番目の人が話し始める.



h) 一番左の人が話し始める.

図 9.6: 4 人の呼びかけに応じる SIG

ステレオのバランスコントロールによって、人間と同様の定位を行っていることは、本研究で使用した人間の音源定位モデルが、うまく工学的に実現できていることを示している。

インタラクションという点で言えば、このケースは、直接的に人間とインタラクションを行うというわけではないが、人間がステレオ定位を移動させ、これを *SIG* が追跡することによって間接的なインタラクションを行っているといえる。

9.3.4 コンパニオンとして

パーソナリティを *Friendly* に設定して、複数の人の会話を聞くようなコンパニオンを紹介する。

設定した状況では、図 9.6 のように、4 人が自発的に *SIG* に発話を行い、*SIG* がそのうちの一人に注意を払う。

4 分割された右上の図は、ストリームビューワ (左:レーダチャート、右:ストリームチャート) である。左下の図は、音モジュールビューワを示している。右下の図は、*SIG* の左カメラの画像情報を示している。検出された顔は四角で囲まれている。左上の図は、俯瞰カメラで録画した画像である。

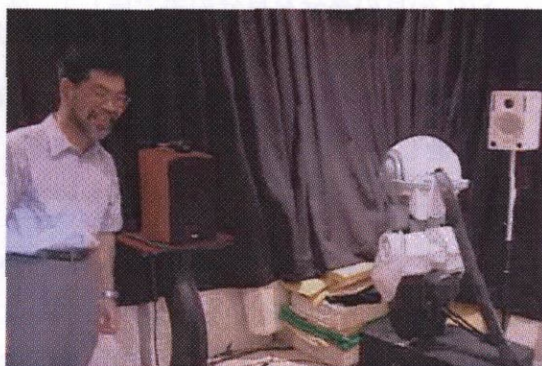
パーソナリティが $(r, \theta) = (1, 0)$ なので、『興味』は、新しいストリームに向けられる。結果、新しいストリームが発生する度に、新しいストリームを追跡し、あたかも *Friendly* な動作が達成されている。これは、2 話者同時発話の追跡と対照的であることがわかる。

この例では、パッシブなコンパニオンとしての振る舞いである。アクティブなコンパニオンは今後の課題である。

9.3.5 そっぱを向く *SIG*



a) 人が話しかける。



b) そっぱを向いてしまう。

図 9.7: そっぱを向く *SIG*

パーソナリティを Hostile に設定した。この場合、 $(r, \theta) = (1, \pi)$ なので、ストリームの興味が負の値となる。興味が負の場合、SIG は追跡を行わないのではなく、そのストリーム方向から逃れるような行動制御を行う。図 9.7a) のように、SIG に話しかけた人に対し、図 9.7b) のようにそこから逃れる方向へ体を向け、あたかも Hostile のパーソナリティを実現している。



図 9.8: SIG の目を隠す被験者



図 9.9: SIG の見えない位置から呼びかける被験者

9.4 考察と今後の課題

パーソナリティを導入した注意制御では、インタラクションを行った人間側にも興味深い行動が見られた。各人には、前もって、SIG のアソシエーション機構やストリームベースの追跡について説明を行った。このような状況で、人間側には以下のような行動が見られた。

1. SIG の目の部分を自分の手で隠して、話しながら、SIG の周りを歩き回って聴覚のパフォーマンスを調べた (図 9.8)。
2. 聴覚のパフォーマンスを調べるため、SIG から見えないように、床にはいつくばって、話しながら這いずり回った (図 9.9)。
3. SIG と鬼ごっこをして遊んだ人もいた。

このように本当に聴覚情報を使っているのか疑ってみたり、それを受け入れて遊んでみたりといった様々な行動が見られた。

現時点では、受付ロボットを除いては、SIG はパッシブであり、決して十分なインタラクションを行うことができるわけではないが、それでも、1 時間以上 SIG と遊んでいく人も見受けられた。このような結果は、言語情報を用いていないが、Eliza を連想させる [122]。

つまり、*SIG* は話者の方向を向くことによって、人々をインタラクションに誘ったり、その機能の原理を探求するように人々を促したりと人々の興味を喚起させることができる。

本章で紹介しなかったパーソナリティの実現や、パーソナリティに基づく感情の表出などが今後の課題である。また、今回は、基本的に、非言語でパッシブなインタラクションを扱ったが、言語によるインタラクションをあわせて用いることにより、一層の自然なインタラクションが期待できるだろう。

本章で扱ったパーソナリティの導入には、その他にも様々な課題が残されている。提案した人工的なパーソナリティが、実際に、目的のパーソナリティを表現できているのかという検証は、大きな課題である。幸運なことに、インターパーソナル理論の分野では、*circumplex* 相関行列を解析するためのソフトウェアが提供されている。今後、ユーザのインタラクションのデータを集め、パーソナリティに基づいた選択的注意アーキテクチャが、目的のパーソナリティを実現しているかを評価する予定である。これにより、将来的に、ロボットのパーソナリティの汎用的な理論の導出につながることを期待している。

9.5 まとめ

視聴覚を統合した複数人物追跡システムの応用として、人間とロボットのソーシャルインタラクションの例を挙げて、その有効性を示した。また、パーソナリティを導入して、柔軟な注意制御を実現した。パーソナリティに基づく注意制御機構は、現時点では、非常に簡単なものであるが、バラエティに富んだインタラクションを実現し、人々の興味を引くことができ、身体を持った非言語 *Eliza* ともいえるべきインタラクションを実現できた。これまで、パーソナリティの研究は、ソフトウェアエージェントの分野で行われているが、実際にロボットを用いた研究はほとんど報告されていない。しかし、本章での結果からもこれからの研究が期待される分野であり、そのような分野に一石を投じることができたと考えている。ただし、アクティブなインタラクションや目的のパーソナリティの検証など様々な課題が残されており、これらは今後の課題として取り組んでいく予定である。

第 10 章

考察

本章では、本研究で述べた技術の実用化に向けた課題を中心に議論を行う。提案したロボット聴覚システムは、従来のロボットでは見られなかった聴覚処理を実現し、自然な人間とのインタラクションが実現できることを様々な評価によって示したが、実用的な応用には、未だ課題を多く残している。

例えば、音源定位や分離は、正面方向で角度差が 30 度以上であることが要件である。実際には、3 から 4 音源が本研究で示した定位・分離の性能限界である。また、調波構造を有している音であっても、基本周波数が高い音では、定位の精度は悪化してしまう。反響の大きな部屋では、定位や分離そのものが難しい。このような点については、実環境での動作を考えると、さらなる改善が要求される。

さらに、水平方向、かつ前方向の定位に限られるため、上下方向への対応や前後問題といった問題に対応できていないことは、実環境では問題である。こういった問題には、アクティブな動作を利用することで、ある程度の対応が可能であると考えている。

情報の統合については、システムの定量的な評価が大きな課題である。静的なシステムでは、一度、マイクロホンやカメラで録音・録画してしまえば、何度でもオフラインで同じデータを用いた実験を行うことができるのに対し、実時間で動作するシステムの場合、刻々と周囲の状況が変化するため、実験を完全に再現すること自体が難しい。実時間システムの評価はその手法から検討する必要がある、今後の大きな課題である。

ロボット聴覚システムの応用として示した音声認識では、9 種類の音響モデルを用いた手法を紹介したが、話者数や音源方向数に対するスケーラビリティの検証は今後の課題である。また、話者方向および、発話のタイミングを固定にして実験を行っているため、これらのパラメータが変動するような場合にも対応が必要である。

パーソナリティを導入した注意制御については、パーソナリティに基づく感情の表出、アクティブなインタラクション、より多くのパーソナリティの実現という問題だけでなく、実現したパーソナリティを検証するという課題も残されている。このような課題には今後

対応していく予定である。

本研究で論じなかったが、この他にも考慮すべき点がいくつか挙げられる。

一つの課題は汎用性であろう。まず、どれだけのロボットに本研究の手法が適用できるかという点があげられる。この点については、現在、新しく開発中のロボット SIG2 (図 10.1) および、京都大学の Repliee (図 10.2) という 2 台のロボットを利用して、その汎用性を検証している。定量的な検証はできていないが、Repliee に関しては、モータ制御部分の変更のみで、他のパラメータを特に変更しなくても、数m 四方の部屋で、聴覚情報による音源追跡を達成したという報告を受けている。これは、提案した聴覚エピソード幾何だけでなく、複数の聴覚情報を統合する方法が実環境処理に本質的であることを示すよい例であろう。



図 10.1: SIG2

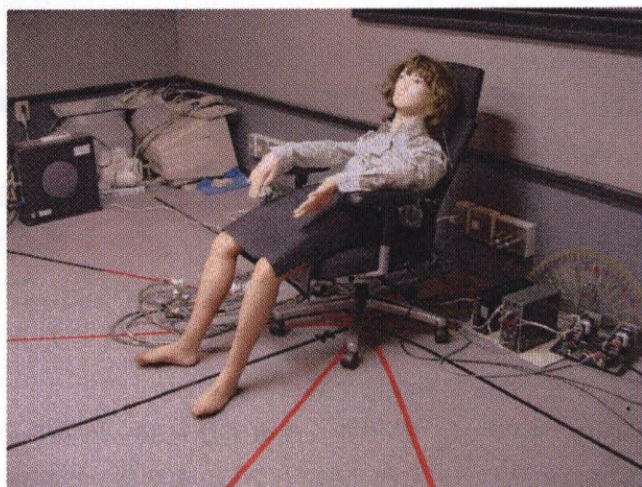


図 10.2: Repliee, Kyoto Univ.

もう一点として、環境に対するロバスト性、つまり、どの程度バラエティに富んだ環境で、動作が期待できるのかという点が挙げられる。この点では、様々な環境での検証が必要である。現時点では、無響室、および、残響時間が 0.2 秒から 0.3 秒程度の比較的反響の少ない部屋での検証を行っている。しかし、反響のある衝立をロボットの近辺においてしまうというような、音響環境的に特異なものが存在する場合には、期待した動作ができないことは確認している。従って、反響の強い、より一般的な部屋においては、エコーキャンセルなどの手法を導入する必要があるだろう。

また、ノイズが大きい展示会場のような場所での動作は、より多くの問題が生じることが予想される。音源定位・追跡に関しては、特徴的な (音量の大きな) 音を対象にした音源追跡実現の可能性は、アクティブオーディションに基づき、音源定位を向上することによ

り、十分にあり得ると考えている。

音声認識については、奈良先端大学の ASKA [90] が展示会場でヘッドセットを利用した音声認識に成功しているが、ロボットのマイクロホンを使った認識はできなかったという報告があるなど、課題が多い。一般に音源分離問題は不良設定問題であることもあり、分離の S/N 比の向上には限界があり、音声認識や話者認識などのフロントエンドとして、常に、実用に耐えうる音源分離能力を提供することは難しい。従って、音声認識については、音響モデルや言語モデルの改良だけではなく、音声認識エンジンそのものの改良も必要であろう。信頼できないような信号を含んだ音声信号に対する認識手法として multi conditioning, missing data, missing feature といった手法がある [18, 105]。こういった手法は、AURORA II などのノイズを含んだ数字データベースを使って、世界中で研究が行われている。また、リップリーディングによって得られる視覚情報と統合することによる手法も活発に研究されている。こういった方法を取り入れると共に、ディクテーションのようにすべての発話を認識するのではなく、word spotting のように、発話の重要な部分を取り出し、その部分のみを認識対象とするといった方法をあわせて利用することも効果的であろう。

特定の部屋でなく、音響環境の異なる複数の部屋で動作を可能にすることも、汎用性を考えた時の一つの課題であろう。同じ部屋内であっても位置が変われば、音響環境が変化するため、部屋内の移動によって生じる環境変化の問題もこの課題に含まれる。SIG の胴体は上半身だけであり、現時点では、このような課題を検証することはできない。このような課題に対しては、SIG2 で今後研究を行っていく予定であるが、人間のような素早い音響環境への適応を実現するために、音響環境を自動的に学習するような仕組みを導入する必要があると考えている。

汎用性と並んでもう一つの大きな課題は、システムの物理的な大きさである。SIG システムでは、ロボット本体に対して、5 台の PC が接続されており、持ち運びが難しいシステムとなっている。リモートブレイン [51] のように処理部をロボットと切り離してしまうという考え方もあり得るが、人間と共生するようなヒューマノイドサイズのロボットでは、安全面、セキュリティ面を考慮するとロボット内で一つの閉じた系として、自律的な判断を行うことが望ましい。SIG2 では、昨今流行のベアボーンマシンを 5 台搭載することにより、自律的な動作が可能なシステムを構築している。しかし、この手法にも限界があり、センサ情報を一つ増やすたびにコンピュータを増やしていたのでは、スペースの面からも電力の面からも限界があるのは明らかである。この問題への一つの解はハードウェア化であろう。各処理をチップ化して組込むことにより、省電力化、省スペース化を図ることができる。実際、一部の機能に対し、FPGA (Field Programmable Gate Array) を対象としたハードウェア化を共同研究として進めている。FPGA は DSP (Digital Signal Processor) と比べ、クロック数が低くても高速な処理が可能であり、消費電力が極端に小

さい。また, ASIC (*Application Specific Integrated Circuit*) 化による大量生産も容易である。このように, ロボットの知覚処理では, ソフトウェア, ハードウェア両面からのアプローチが必要であろう。